論　文

Agglomeration and Networking in Academic Research

Hiroki INOUE[a], Hiromi SAITO[b], Shigeru SUGIHARA[c,*],
Shigeru HIROTA[a], Taisuke KAMEDA[c]

Abstract

Research activity is crucial to economic growth. To effectively enhance science research, this paper investigates the mechanisms of producing high quality research by analyzing the determinants of research outcomes. These analyzed determinants include externalities, such as the benefits of agglomeration and network effects; the effects of the structure and nature of research networks, such as density and variety; and attributes of coauthors, such as publishing a certain quantity of papers or quality papers. This paper also contributes to the measurement of the quality of academic research by introducing several new indicators of quality. Our results show that agglomeration and the network effect are observed with diminishing effects, as the range of externalities goes farther away from a researcher, the exception being the collaboration with foreign researchers. Concerning network characteristics, linkage with prolific coauthors or organizers significantly contributes to a quantitative increase in research outcomes, but very little to qualitative outcomes. As such, it is the linkage with quality coauthors that enhance research quality.

JEL Classification Code: I23, O39

Keywords: Research Productivity, Research Quality, Externality,
Network Characteristics

[a] Kyoto University
[b] Chiba University
[c] Cabinet Office, Government of Japan
*Corresponding author: Economic and Social Reseach Institute, Cabinet Office, Government of Japan, 1-6-1 Nagata-cho, Chiyoda-ku, Tokyo 100-8914, Japan

学術研究における集積とネットワーク

井上寛規・齋藤裕美・杉原　茂・広田　茂・亀田泰佑

〈要 旨〉

研究活動は経済成長の源泉である。その研究活動を効率的に推進するため、本稿では科学研究の成果の決定要因を分析することにより、質の高い研究が生み出されるメカニズムを明らかにした。分析で取り上げた要因としては、集積の利益やネットワーク効果といった外部性、稠密性や多様性といった研究ネットワークの構造と特性、発表論文の量や質で測った共著者の属性等が挙げられる。さらに本稿は、研究の質についての新たな指標を複数導入しており、学術研究の質の評価尺度についても新たな貢献を行っている。分析結果によれば、研究活動の集積及びネットワーク効果が観察された。この外部性は当該研究者から離れると共に逓減の度合いが大きくなるが、海外の研究者との共同研究においてはその限りでない。また、ネットワークの特徴に関して、多くの論文を執筆した研究者や研究プロジェクトのリーダー経験のある研究者との共同研究は、研究者の業績の量を増加させるが、質の向上には結びついていない。研究の質を向上させるのに有効なのは、質の高い研究を行っている研究者との共同研究であることが示唆された。

JEL 分類コード: I23, O39
キーワード：研究の生産性、研究の質、外部性、ネットワーク特性

1. Introduction

Research activity is crucial to technological progress and, hence, to economic growth. Japan, in particular, has to rely on research and technological progress for growth with its declining labor force. Thus, the Japanese government has tried to strengthen science research activities after the Science and Technology Basic Law in 1995.

However, even if the government aggressively assists science research financially, it does not necessarily lead to enhanced research outcomes. Knowledge of the mechanisms of producing research outcomes is needed to effectively enhance science research.

The promotion of academic research poses many questions about the nature of research activities. Does research activity benefit from externalities from agglomeration or the research network? To what distance does research activity exert spillover effects? Do the structure and nature of the research network have prominent effects on research outcomes?

Moreover, in promoting academic research, it should be noted that simply increasing the quantity of research activities is not sufficient to raising the economic welfare of a nation. As is the case with general economic activities, research activity should generate more outputs with the same inputs, or the same outputs with fewer inputs. It is imperative to raise the efficiency of academic research. In particular, to enhance the quality of research, the quality should be explicitly taken into account. The difficulty here is to accurately measure the research quality. Conventional measures of quality include the number of published papers and number of citations. This paper takes advantage of institutional factors in medicine to measure the quality of research and uses some new measures of quality in the analysis, as will be explained below.

NISTEP (2015) quantitatively describes Japan's science and technology (S&T) activities based on paper data, etc. This data source is also based on *Web of Science* (WoS) by Thomson Reuter. This shows that the volume of scientific papers has increased in the world. About 400,000 papers were published in the early 1980s, while over 1,200,000 papers were published in 2013. On the other hand, the volume of scientific papers produced in Japan has remained at the same level over the last 10 years. However, Japan's share has declined, accounting for 8% in the early 1990s, 8.6% in early 2000, and 5.2% in early 2010, based on the basis of fractional counting.[1]

Particularly, the share of papers with high impact, adjusted top 10% papers (or adjusted top 1% papers), was 5.8% (or 4.7%) in the early 1990s, 5.9% (or 4.7%) in early 2000, and 3.6% (or 2.9%) in early 2010. This shows that the share of papers in Japan declined in both quantity and quality.

When we review publishing trends in the world, papers in life science, basic life science, or clinical medicine account for half of all papers. However, the ratio of these papers has slightly decreased during the past 30 years. In Japan, papers on basic life science have slightly decreased, by 2.5 points, from 1981 to 2013, while papers on clinical medicine have increased, by 12 points.

In addition, the style of research activities has recently changed. The number of internationally coauthored papers has increased in the world. In 2013, the share of domestic papers was 76%, while that of internationally coauthored papers was 24%. Moreover, Japan is no exception; the number of internationally coauthored papers in Japan has increased, those focusing on clinical medicine accounting for 19.2%.

This paper investigates the mechanisms of producing high quality research and sheds light on the way to promote academic research by analyzing the determinants of research outcomes.

On the mechanics of producing high quality research outcomes, the analyzed determinants include: (i) externalities, such as the benefits of agglomeration and network effects from researchers in the same institution, in nearby areas, or overseas, and (ii) the effects of the structure of research networks, such as density, flexibility, and variety, as well as the nature of links, such as a link with researchers who publish a lot of papers or quality papers. The beneficial effects of the geographical concentration of research activities and interactions among them include economies of scale, knowledge spillover, and diversity. It is not easy to clearly separate these effects in empirical analyses; in this paper, agglomeration denotes economies of scale at various distances, network effect refers specifically to knowledge spillover through coauthorship, and diversity is contained as one of the elements of the structure and nature of networks, together with the other properties of networks listed below.

Among the first set of determinants, externalities, we start with the analysis of conventional economies of scale (agglomeration), which assert that a researcher profits from other researchers within the same institute (or same city), so that he or

[1] Please refer to Ch.4 in NISTEP (2015) regarding the basis of fractional counting.

she is more productive when the affiliated institution embraces more researchers, or more papers written by the researchers belonging to the same institute. Then, we move on to the analysis of spillover effects, which investigates whether the beneficial effects of other researchers mentioned above extend to other institutions, nearby areas, and overseas. A noble feature of our analysis is that it is based on individual researchers/papers, as is explained below, and, therefore, the actual linkages among researchers can be used to shed light on the range of network effects.

Then, we examine the second set of determinants, the structure and nature of the network. As for the structure, we examine how densely congregated the coauthors within the network are, and whether a variety of institutions are represented in the network.

A densely populated or concentrated network may have an advantage of transmitting information very rapidly. A network with fixed membership could facilitate a deep exchange of ideas, while a network with changing membership may be favorable for absorbing new ideas. Similar to the latter, a network that is comprised of members from multiple institutions possibly embraces a variety of ideas and, hence, may contribute to breakthroughs in academic research; however, medicine networks among multiple institutions often represent participation in multi-center randomized controlled trials with uniform protocols (i.e., the same objectives and methods).

As for the nature of the research network, it may be that collaborators with specific attributes exert a favorable effect on the quantity and quality of the research outcome. We will examine a wide variety of characteristics. These attributes may be viewed as either quantitative or qualitative. A typical quantitative attribute is whether a researcher's coauthors publish a lot of papers, while typical qualitative attributes include whether coauthors publish papers in core journals or if papers are cited in clinical guidelines. Linkage to the first type, massive producer, may contribute to the quantity of published papers, but may not contribute to the quality of research. The second and third types may have the opposite effects. The full set of attributes examined will be presented later.

Our unit of analysis is individual papers/researchers rather than research institutes, as was the case in previous analyses. This type of analysis permits us to entangle intricate relationships among researchers, and our analyses of the effects of the range and structure of research networks are very much substantiated by this choice of unit.

Furthermore, distinction between individual researchers and institutions enables policymakers to draw up a promotional plan that is grounded on the real source of quality research. When contribution from individual researchers is independent of institutional contribution, policymakers should target these individual researchers. If a study confounds the contribution from individual researchers with institutional contribution, it could wrongly propose a promotional plan that covers all researchers in the targeted institutions.

Our analyses try to incorporate the quality of research. In doing so, we exploit an institutional feature in medicine for the identification of the research quality. In medicine, evidence-based medicine is systematically encouraged and clinical guidelines are compiled by professional associations. Citations in guidelines could be a good indicator of research quality. In addition, the quality of research is evaluated by professional panels in the process of compilation. Such evaluation may also be used as a quality indicator. In the analyses below, we use the former guideline-related indicator, citation in guidelines, as well as other conventional quality indicators, such as the number of citations and publications in prestigious journals. The identification of prestigious journals is also facilitated by the institutional feature of medicine. The National Institute of Health (NIH) selects the PubMed Core Clinical Journals, which may be regarded as "prestigious" or mainstream journals.

The organization of the paper is as follows. Section 2 reviews the related literature. Section 3 explains the data. Section 4 analyzes the determinants of research efficiency in a descriptive way, while Section 5 formally estimates quantitative models. Finally, Section 6 concludes.

2. Literature Review

First, we review which determinants of research outcomes were applied by previous works. Particularly, we focus on networks on research.

With regard to the network effect, Banerjee *et al.* (2013) examine how participation in a microfinance program diffuses through social networks. They estimate structural models of diffusion that allow us to (i) determine the relative roles of basic information transmission versus other forms of peer influence, and (ii) distinguish information passing by participants and non-participants.

As for the network characteristics, Azoulay, Zivin, and Wang (2010) analyze the spillover effects between researchers, using coauthorship along with those eminent life scientists who suddenly died, or superstars. They estimate the magnitude of spillovers generated by 112 academic superstars who died prematurely and unexpectedly, thus providing an exogenous source of variation in the structure of their collaborators' coauthorship networks. Following the death of a superstar, they find that collaborators experience, on average, a lasting 5 to 8% decline in their quality-adjusted publication rates.

Newman (2001a) constructs a bibliographical database in which two scientists are considered connected if they have coauthored one or more papers together. They study a variety of statistical properties of the networks, including the number of papers written by authors, authors per paper, collaborators of scientists, existence and size of a giant component of connected scientists, and degree of clustering in the networks. Further, they highlight some apparent differences in collaboration patterns between the subjects studied. In the following paper, we study a number of measures of centrality and connectedness in the same networks.

When considering the effects of networks on innovation, we have to also scrutinize the diversity, as the networks can be regarded as media of the diversity between different groups. As for the concept of diversity, though many disciplines use their own characterization, Stirling (1998) organizes the idea into three subordinates: variety, balance, and disparity. Variety refers to the number of categories; the higher the number of categories, the greater the variety. This is the most common aspect of diversity, and some papers use the term "diversity" in reference to this subordinate. For example, Lay and Dreher (1992) examine the determinant of variety using the different best practices in the modes of usage of numerically controlled machine tools. Moreover, balance refers to the pattern in the apportionment of that quality across the relevant categories. The more equal the fractions, the more even the balance and greater the diversity. This aspect is commonly used in such contexts as income inequality and market structure. Further, disparity refers to the nature and degree to which the categories themselves are different from one another. As achieved by Nguyen *et al.* (2005) with regard to oil refining technology, this is usually captured by some form of distance measure. Stirling (2007) proposes some quantitative heuristics that try to capture all of these subordinate factors of diversity.

This characterization of diversity has a significant influence on subsequent studies. Applying the three aspects, Yegros-Yegros, Rafols, and D'Este (2015) analyzes the effect of the degree of interdisciplinarity on the citation impact of individual publications. They operationalize interdisciplinarity as disciplinary diversity in the references of a publication. Interestingly, they find that variety has a positive effect on impact, whereas balance and disparity have a negative effect. Another example is van Rijnsoever *et al.* (2015), which investigates the influence of network position and composition of innovation projects on the creation diversity of an emerging technology, using variety and balance among the three components.

However, previous works do not necessarily consider all of these various aspects of the network. Some focus on the structure of networks such as distances and centrality (e.g., Newman (2001b)). Others mainly deal with the nature of the network such as the quantity and quality of collaborators' outcomes (e.g., Azoulay, Ding, and Stuart (2007)). Our paper analyzes both the structure and nature of the network using a variety of indicators.

Second, we review which measurements have been applied by previous works as outcomes of innovative activities. This is important when we analyze the efficiency of innovation.

Carlino and Kerr (2015) marshal the measurement of technological innovation in three ways: (i) by the inputs used in the innovation process, such as R&D expenditures or venture capital (VC) investment; (ii) by intermediate outputs of the innovation effort, such as the number of patents; or (iii) by some final measure of innovative work, such as the number of new product announcements.

The earlier studies often use R&D expenditures as a measurement of output because of the scarcity of available data. Needless to say, such measurements are not appropriate to examine the productivity and efficiency of the innovation because the ratio of outcomes to input does matter in measuring the productivity.

As for the intermediate output of the innovation, Azoulay, Ding, and Stuart (2007) analyzes the impact of agglomeration on patenting. They examine the individual, contextual, and institutional determinants of faculty patenting behavior in a panel dataset spanning the careers of 3,884 academic life scientists. They conclude that the patent stock of the university affects the individual researcher's patenting behavior. Azoulay, Ding, and Stuart (2009) examine the influence of faculty patenting activity on the rate, quality, and content of public research outputs in a panel dataset. In the process of the estimation, they also use the patent stock of the affiliate. Moreover, Lichtenberg (2013) examines the relationship across diseases between the long-run growth in the number of publications about a disease, and change in the age-adjusted mortality rate from the disease. Further, Duranton and Kerr (2015) present a review on agglomeration in the context of geoeconomics.

An example of research that adopts the third measurement is Feldman and Audretsch (1999). It uses the United States Small Business Administration's Innovation Data Base (SBIDB), which consists of new product introductions compiled from the new product announcement sections of over 100 technology, engineering, and trade journals, spanning every manufacturing industry. They show considerable support for the thesis that diversity, rather than specialization, better promotes technological innovation. Although most previous works in the areas of biomedicine measure the outcome using such indicators as the quantity and quality of patents, some studies adopt more direct indicators of the research outcome. For example, Lichtenberg (2013) uses the mortality rate of the disease concerned. On the other hand, we analyze using the information from the clinical guidelines as the direct measure of the research outcome in the medical field, as well as the bibliometrical data. We regard clinical research as an index to show how much academic research is actually utilized for medical practice.

Regarding the literature in Japan, there are some papers that deal with the mechanism of scientific research. Saka, Igami, and Tomizawa (2015) relate the WoS with the Database of Grants-in-Aid for Scientific Research of Japan, and examine the relationship between the grants-in-aid and papers within the WoS. Their research covers not only the medical field, but also entire disciplines. On the one hand, their scope is limited to Japan. Yonetani, Ikeuchi, and Kuwahara (2013) use the WoS and Survey of Research and Development by the Japanese Statistics Bureau to investigate the relationship between the number of published papers and deterministic variables such as the number of researchers and research funds. The unit of this research is university, and the researcher basis analysis is not carried out. On the other hand, we focus on researcher basis analysis. Using researcher as the unit of analysis enables us not only to examine the relationship between research outcomes, agglomeration, and networking in more detail, but also to shed light on the problem that the policymakers should target, individual researchers or institutions.

3. Data and Variables

One of the focuses of our analysis is to utilize the rich information of individual paper/researcher basis data, rather than institute basis data. We construct a paper-basis dataset and researcher-basis dataset, mainly using the Thomson Reuters' *WoS Core Collection*,[2] with some additional data sources, such as the NIH grant database and references of clinical guidelines. WoS is a citation database that covers over 12,000 journals in more than 250 disciplines.

In the paper-basis dataset, we collect all of journal articles published in the journals categorized as Cardiac and Cardiovascular Systems in the WoS between 2008 and 2014. The data covers not only journals in specific countries, but also highly ranked journals all over the world, which is measured by the impact factor. The total number of the articles is 160,355. This dataset contains each paper's authors, author affiliations, each affiliation's address, and number of citations. Using the affiliation's address, we aggregate the number of papers to the faculty, university, city, and country levels to observe the extent of agglomeration. In addition, we examine whether each paper is cited by the clinical guidelines in the field, that is, the 2014 AHA/ACC Guideline for the Management of Patients With Non-ST-Elevation Acute Coronary Syndromes (Amsterdam *et al.* (2014)), 2013 ACCF/AHA Guideline for the Management of Heart Failure (Yancy *et al.* (2013)), and 2013 ACCF/AHA Guideline for the Management of ST-Elevation Myocardial Infarction (O'Gara *et al.* (2013)).

Concerning the measurement of the quality of research, we utilize new indicators that arise from institutional features of medicine along with traditional indicators, such as the number of citations. These new indicators include the number of papers published in "Core Clinical Journals" and those cited in clinical guidelines. The list of Core Clinical Journals is selected and provided in PubMed by the National Library of Medicine, and the guidelines are compiled by medical associations. As such, we exploit such official and semi-official judgment about the quality of published papers.

In the paper-basis dataset, the major research output variable is the number of citations of each article, including self-citations. In the WoS, the highly cited papers are defined as papers whose number of citations is within the top 1%.

The reputation of the journal where the article is published is also an important measure of the quality of the article. In order to specify such top journals, we choose journals on the PubMED Core Clinical Journals list.[3] In the WoS's Cardiac and

---

[2] http: //thomsonreuters.com/en/products-services/scholarly-scientific-research/scholarly-search-and-discovery/web-of-science-core-collection.html.

[3] https: //www.nlm.nih.gov/bsd/aim.html.

Cardiovascular Systems area, the following journals are on the list: *American Heart Journal, American Journal of Cardiology, Circulation, Annals of Thoracic Surgery, Heart, Heart & Lung, Journal of the American College of Cardiology, Journal of Thoracic and Cardiovascular Surgery,* and *Progress in Cardiovascular Diseases*. Among these journals, we exclude two, *Annals of Thoracic Surgery* and *Journal of Thoracic and Cardiovascular Surgery*, because they focus on surgery. We defined the "core journals" as the remaining seven journals.

As the Literature Review indicates, the effects of the structure and nature of networks can be classified into three categories: variety, balance, and disparity. Variety can be captured by the number of authors in a network, balance connotes no dominant coauthor that can be captured by equal frequency of coauthorship within a network, and disparity within a network means that the network contains coauthors coming from different fields, nationalities, and cultures. Balance could be captured by the fixedness of the network, but we are unable to calculate such an index due to our limited capability. Disparity is captured by colleague density and the number of participating universities. As an indicator that reflects the structure of the coauthor network of each paper, we compute colleague density. This is the share of colleagues in coauthors at each scope, that is, within the same faculty, within the same university but different faculty, within the same city but different university, and within the same country but different city. It is calculated by dividing the number of colleague relationships with total number of possible pairs of authors. The higher this density, the more colleague relations exist in the coauthorship of the paper.

In calculating the number of universities to which the authors belong, when an author belongs to multiple universities, we count them all for this index.

The researcher-basis dataset covers all authors of the articles contained in the above-mentioned paper-basis dataset. However, we exclude 27,536 researchers whose faculty we cannot identify. As such, the total number of the researchers is 322,748, and this dataset contains each researcher's list of published articles.

In this researcher-basis dataset, we have to identify researchers across papers. As the WoS does not have unique IDs for researchers, we must make a judgement as to whether a researcher in a paper is identical to another researcher with the same name in a different paper; therefore, we regard researchers with the same name in the same city as an identical. Note that we also regard researchers with the same name, but in a different institute, in the same city as identical. This is because researchers often belong to multiple institutes, such as both a medical school and hospital. However, we have to also note that we can count the same researcher twice if he or she moves to a different city. In our dataset, we have 31,413 identical researcher names in different cities, which amounts to around 9.6% of the total researchers, and possibly includes researchers who moved between cities or who belong to multiple institutes in different cities. Therefore, we need to make further efforts to identify such researchers using other data sources.

The outcome variables are the number of papers each researcher publishes, total number of citations of his or her papers (including self-citations), maximum number of citations of his or her papers, number of papers published in the top journals (as defined above), and number of papers cited in the clinical guidelines. The fractional counting method could be applied to the calculation of the number of published papers of each researcher to account for the existence of coauthors: if two researchers coauthor two papers, it could be considered one for each researcher. Further, we checked the robustness of our naïve counting and found that both methods give similar results. Therefore, we simply count the number of papers for each researcher.

The determinants of the dataset can be classified into two categories: (i) benefits of agglomeration or network effects and (ii) network characteristics. To measure the extent of the agglomeration, we compute the number of active researchers who belong to the same faculties, same universities less the same faculty, same city less the same university, and same country less the same city. By "active researcher," we mean researchers who have published at least one paper in the WoS database. Alternatively, we can use the number of papers instead of researchers.

These kinds of multi-layered agglomeration effects are transmitted to the individual researcher through the network. To observe these network effects, we use the number of coauthors from the same faculty, university, city, and country.

In order to capture the network characteristics, we specify what kind of and how many coauthors each researcher has. As for these attributes of the coauthors, we prepare the following variables. The total number of coauthors is a simple summation of the coauthors of his or her papers. For computational convenience, we do not omit duplication when a researcher coauthored with the same author multiple times. The total number of papers and citations are also simple summations of those of coauthors. As for the number of prolific coauthors, we define it as the researchers who published (and are contained in the WoS) more than 20 papers. This criterion corresponds to the

99th percentile of our dataset. We also compute the number of highly cited coauthors, the definition of which is that the citations of their papers are above the 99th percentile in Cardiac and Cardiovascular Systems. In the medical field, the last researcher of the author list plays a special role in many cases. The last author is the leader of the laboratory and usually plays an important role in obtaining the research funds. Thus, to obtain the quality factors of coauthors, we also count the number of coauthors whose papers are in the core journals and are cited by the clinical guidelines.

4. Descriptive Statistics and Data Analysis

In this section, we describe the basic properties of the data and conduct a data analysis on the relationship between the outcome variables and determinants of research productivity. The overall structure of this section is as follows. We start with outcome variables, such as the number of published papers, citations, and papers published in top journals, and then we present some simple descriptive statistics and their distributions.

Then, we treat with the determinants of research productivity. First, simple descriptive statistics are presented and, then, their relationships with the outcome variables are investigated. These determinants can be grouped into two categories: (i) externalities and (ii) structure and nature of network with coauthors.

The first category, externalities, includes agglomeration and network effect. Agglomeration is the usual economies of scale, while network effect is the beneficial effect of being linked with coauthors. After presenting descriptive statistics and distributions, we examine how outcomes change as agglomeration or the coauthor network change.

The next category of determinants includes the structure of the network and attributes of coauthors. Networks possess various kinds of structural characteristics, such as whether interactions within the network being dense or sparse, whether the coauthor relationship is fixed or flexible, whether the network is centralized or decentralized, and whether the network exhibits the diversity or homogeneity of participants. These structural characteristics are expected to affect research productivity. The attributes of coauthors, which may also be expected to influence research productivity, include whether they are prolific, highly cited, or able to produce papers that are published in core journals or cited in guidelines. Just as in the case of externalities, first, simple descriptive statistics are presented and, then, their relationships with outcome variables are investigated. It should be noted in advance that the relationships between the outcomes and determinants are not necessarily causal here. To identify the causal effect, we will have to examine changes in the outcomes when the determinants are exogenously altered. Such examination requires very detailed information on the historical and institutional dynamics of academic research and research institutions. We would like to conduct a more precise analysis of these relationships by incorporating historical and institutional information in the future.

(1) Outcome variables

Let us start with the outcome variables. Conceptually, research outcomes can be classified into two categories: quantity and quality. For quantity variables, we use the number of published papers, citations, and highly cited papers, while quality outcome variables include the number of papers published in top journals and papers cited in clinical guidelines.

The dataset includes 160,355 papers and 322,748 researchers in total. Panel (a) of Table 1 shows the descriptive statistics based on individual papers.

The average number of authors per paper is seven. On average, each paper is cited 12 times by other papers, and around 8.2% of papers are published in the core journals.

Panel (b) exhibits the descriptive statistics aggregated to researchers. Note that this dataset contains only researchers who published at least one paper. A vast majority of researchers published no papers at all, we suppose, and, thus, are excluded from the dataset. An average researcher in this dataset publishes 2.4 papers. The total number of citations of each researcher's papers is, on average, 31. Only 0.22 of each researcher's papers are published in the core journals, and 0.02 are cited in the guidelines.

In addition, the distributions of the outcome variables are very skewed. An average researcher publishes very few papers (perhaps none) and, even if he or she publishes any papers, they are not very often cited by other papers. Even fewer researchers publish papers that are published in the core journals or cited in the guidelines. To represent such skewed distributions, we draw Gini diagrams in Figure 1.

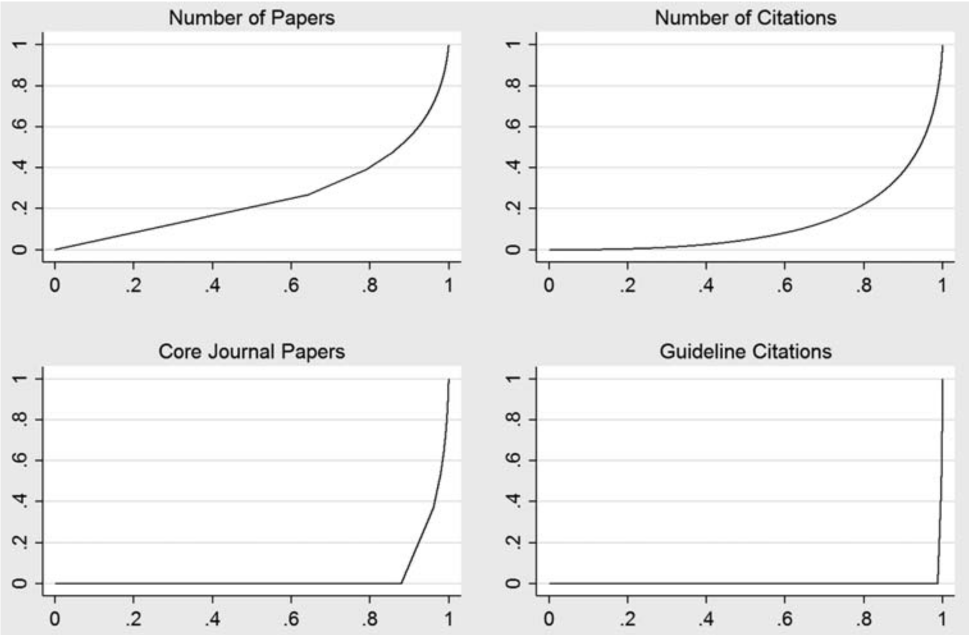
In the upper left panel, the horizontal axis indicates the cumulative proportions of researchers and the vertical axis indicates the cumulative proportions of the number of papers. If a 45-degree line is observed in a Gini graph, the proportion of papers is

Table 1：Descriptive Statistics

(a) Paper-basis Dataset

Variable	Obs.	Mean	Std. Dev.	Min.	Max.
Number of Authors	160,355	7.053	4.117	1	70
Number of Citations	160,355	12.038	27.788	0	2,575
Core Journal Dummy	160,355	0.082	0.275	0	1
Number of Citations by the Guidelines	160,355	0.003	0.060	0	3
Number of Affiliates of Authors	158,831	3.338	2.859	1	71
Share of Colleagues of Faculty in Coauthors	150,076	0.527	0.358	0	1
Share of Colleagues of Univ. less Faculty	150,076	0.080	0.156	0	1
Share of Colleagues of City less Univ.	150,076	0.041	0.113	0	1
Share of Colleagues of Country less Faculty	150,076	0.112	0.207	0	1
Share of Foreign Coauthors	150,076	0.240	0.283	0	1
(b) Researcher-basis Dataset					
Attributes of Researcher					
Number of Papers	322,748	2.399	3.777	1	121
Total Number of Citations	322,748	31.083	80.478	0	6,672
Number of Papers in the Core Journals	322,748	0.221	0.914	0	38
Number of Papers Cited by the Guidelines	322,748	0.019	0.241	0	23
Agglomeration Measurements					
Number of Researchers of the Same Faculty	322,748	19.803	41.470	0.2	1,133
Number of Researchers of the Same Univ. less the Same Faculty	322,748	303.443	437.040	0	3,686
Number of Researchers of the Same City less the Same Univ.	322,748	913.423	1,282.317	0	6,328
Number of Researchers of the Same Country less the Same City	322,748	29,823.540	31,708.870	0	84,552
Coauthorship Network					
Number of Coauthors from the Same Faculty	322,748	4.523	3.046	0	51
Number of Coauthors from the Same Univ. less the Same Faculty	322,748	1.493	2.245	0	29
Number of Coauthors from the Same City less the Same Univ.	322,748	0.558	1.408	0	42
Number of Coauthors from the Same Country less the Same City	322,748	1.204	2.593	0	61
Number of Coauthors from Abroad	322,748	0.843	2.434	0	63
Attributes of Coauthors					
Total Number of Coauthors	322,748	19.272	36.371	1	1,350
Total Number of Papers	322,748	134.487	685.188	0	55,110
Total Number of Citations	322,748	2,128.344	12,658.690	0	1,314,506
Number of Prolific Coauthors	322,748	1.782	9.690	0	520
Number of Highly Cited Coauthors	322,748	1.644	14.014	0	1,356
Number of the Last Author	322,748	16.549	77.494	0	4,280
Number of the Core Journal Papers	322,748	15.023	96.144	0	8,324
Number of Papers Cited by the Guidelines	322,748	0.977	8.660	0	734
Institutional Capabilities					
Univ. Fund	1,119	1,594,117	3,601,130	3,183	31,500,000
NIH Grants for the Univ.	2,865	181,000,000	518,000,000	1,834.74	4,350,000,000

Figure 1：Inequalities of Outcomes



commensurate with the proportion of the outcome, which implies that the distribution of the outcome is "equal." On the other hand, if the curve is strongly convex to below, the distribution is "unequal," or skewed. The graph in the upper left panel is very much skewed. Up to around 20% of the cumulative proportion of the citation is zero, which means that around 20% of papers are not cited at all. A little more than 70% of papers represent only 20% of citations. The very steep rise of the curve at the end of the horizontal line shows that only a very small number of papers account for the vast majority of citations. The rest of the graph is drawn in terms of individual researchers. Again, the distribution of the number of papers among researchers is skewed, and is more so in the case of the number and maximum number of citations. Further, the distributions become extremely skewed for the number of core journal papers and guideline-cited papers.

（2） Externalities: agglomeration and network effect

Let us move on to the first category of the determinants of research productivity, namely externalities. Among these, the variables representing agglomeration are the number of papers written by researchers in the institution as well as the number of researchers in the institution. To investigate the range of economies of scale, the above variables are aggregated for the faculty to which a researcher belongs; for the institution, but outside of the faculty, to which a researcher belongs; for the city to which the institution belongs, excluding the institution to which a researcher belongs; and for the nation to which the institution belongs, excluding the institution to which a researcher belongs.

Referring back to Table 1, faculties embrace around 20 researchers, on average. The average number of researchers in the same university, but outside of the faculty to which a researcher belongs, is 303. Moreover, for the same city, but outside of the university, this number is 913, and for the same country, but outside the city, it is 29,823.

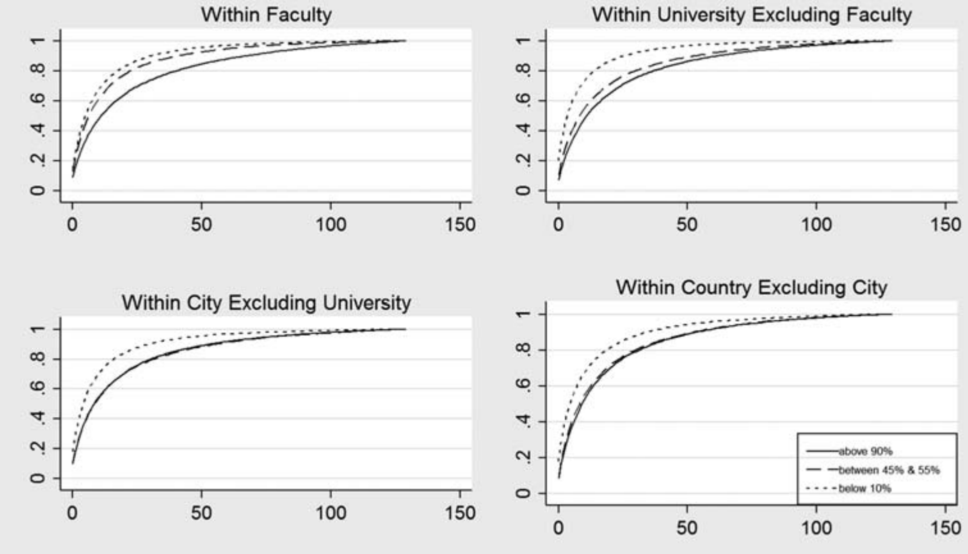
The other externality is network effect, which is assumed to be embodied in the number of coauthors with whom a researcher works. Again, to examine the range of network effects, we count the number of coauthors in the faculty to which a researcher belongs; in the institution, but outside of the faculty, to which a researcher belongs; in the city to which the institution belongs, excluding the institution to which a researcher belongs; and in the nation to which the institution belongs, excluding the institution to which a researcher belongs.

An average researcher collaborates with 4.5 researchers within the same faculty. The number of coauthors is 1.5 within the same university, but different faculties; 0.6 within the same city, but outside the university; 1.2 within the same country, but outside the city; and 0.8 for coauthors abroad.

So far, we have presented basic descriptive statistics for both agglomeration and network effect. Now, we will examine their relationship with the outcomes. Figure 2-1 shows the cumulative distribution functions used to investigate what happens to the distribution of outcomes when supposed determinants change.

The case of the number of papers is very similar to the case of the number of citations; hence, we refrain from reporting the former. The solid line in the upper left panel, for example, depicts the cumulative distribution of faculties as the number of citations changes, in the case where agglomeration measured by the number of researchers is the highest 10%. The solid line shows the share of faculties with the number of citations, on the horizontal axis, at this level of agglomeration. Very roughly, the share is around 80% for faculties with five or less papers, which researchers in the faculties publish, and nearly 100% for faculties with 10 or less papers. The dashed line does the same exercise in the case where agglomeration is the middle 45 to 55%, and

Figure 2-1：Shapes of Cumulative Distribution Functions of the Number of Citatioons by the Degree of Agglomeration

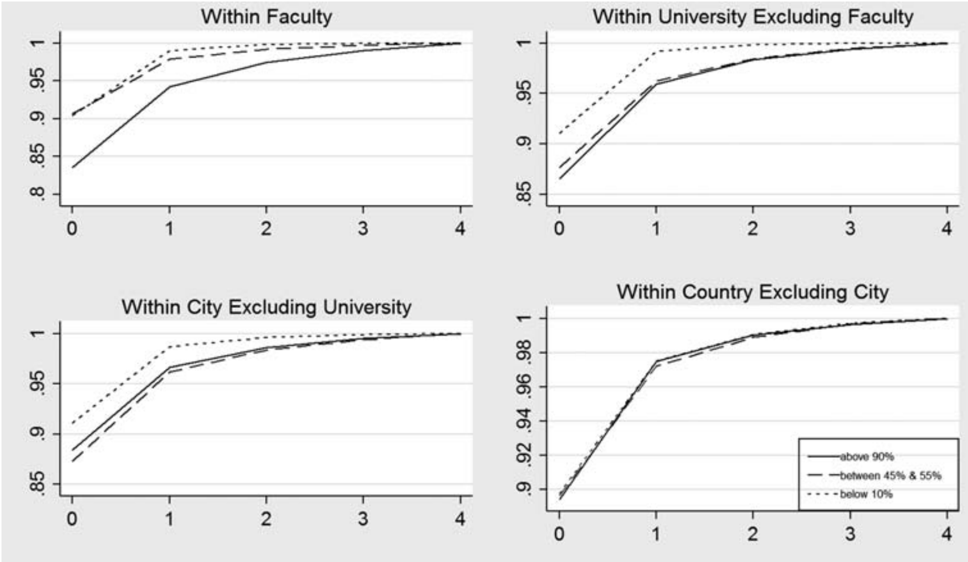


Note: To make the difference of distributions easy at which to look, they are drawn without highest 5%.

the dotted line in the case the lowest 10%. By comparing the three lines, one can see the change in the distribution when the agglomeration changes. In this panel, the solid line for the highest 10% runs is located Southeast, the dotted line for the lowest 10% Northwest, and the dashed line in the middle. This means that as the agglomeration increases from the lowest (dotted line) to the highest (solid line), the share of the faculties with a larger number of papers increases. This, in turn, implies that larger agglomeration corresponds to a higher outcome. Along the first row, one can proceed to the right to find out that the agglomeration of a university outside of faculty, as well as that of a city outside of a university, generally contributes to an increase in the number of papers published by the researcher. In these cases, however, the shift from the dotted to dashed line is much larger than that from the dashed to solid line, implying that marginal increases in the number of papers become less prominent as we move farther away from the faculty to which a researcher belongs. In other words, diminishing marginal returns to scale are observed. The lower right panel depicts the case of agglomeration of a country, excluding the city to which a researcher belongs. Agglomeration at this level does not contribute to an increased number of papers.

These trends change slightly when we use the number of publications in core journals as the outcome variable in Figure 2-2.

Figure 2-2 : Shapes of Cumulative Distribution Functions of the Number of Core Journal Papers by the Degree of Agglomeration



Note: To make the difference of distributions easy at which to look, they are drawn without highest 1%.

In the upper left panel, the solid line is located Southeast, and the dotted line Northwest, and this tendency is preserved in the upper right graph. In the left graph on the lower row, the solid line tends to be located above the dashed line and, in the right panel on the second row, three lines overlap. Therefore, the benefits of agglomeration are rather limited to the faculty and university.

With respect to network effect, overall trends are the same as the case of agglomeration; hence, we again refrain from reporting the results.

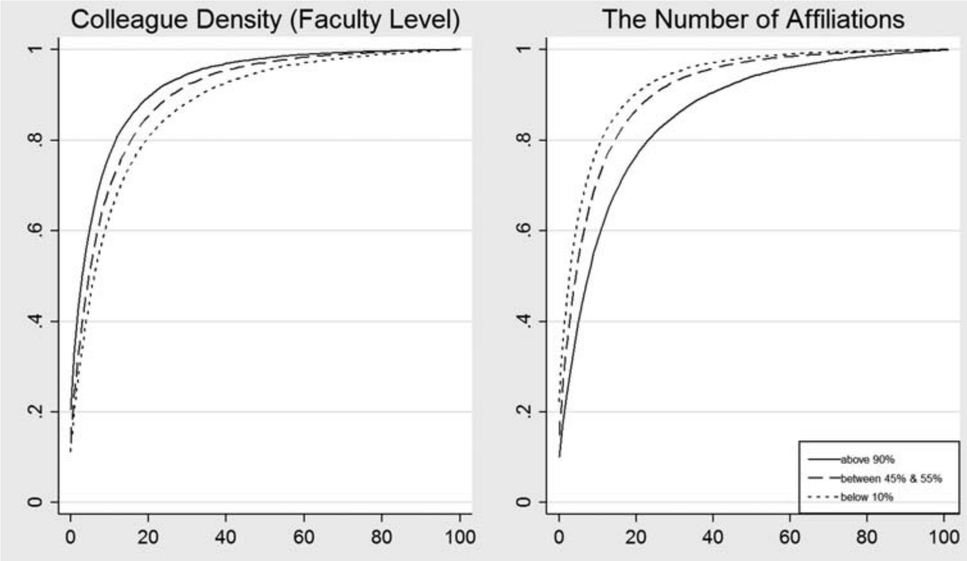
In sum, positive externalities, both agglomeration and network effect, are observed with diminishing effects, as the range of externalities goes farther away from a researcher, the exception being the collaboration with foreign researchers.

(3) Structure and nature of network

Next, we examine the structure and nature of the network. As for the structure, we examine how densely congregated the coauthors within the network are, and whether a variety of institutions are represented.

First, we examine the effect of coauthor density on the number of citations. This density is calculated only for individual papers, so that outcomes, such as the number of papers or core journal papers, are meaningless. In the left panel of Figure 3, as is usual,

Figure 3 : Shapes of Cumulative Distribution Functions of the Number of Citations



Note: To make the difference of distributions easy at which to look, they are drawn without highest 1%.

the horizontal axis indicates the number of citations, while the cumulative share is shown on the vertical axis.

The upper left graph exhibits that the solid line runs from North to West of the dashed and dotted lines, implying that the denser the coauthor network at the faculty level, the smaller the number of citations. This result is contrary to our hypothesis that intensive interaction among collaborators entails better outcomes. The hypothesis is also rejected for other levels of coauthorship, except for foreign coauthors (the graphs are omitted here). Thus, collaboration with foreign researchers is beneficial to quality research. The rejection of the density hypothesis is disappointing, but there may be another interpretation. This measure could be a measure of diversity rather than of the density of the network. The hypothesis is that the versatility of opinions and methods in research is beneficial, and that diversity is measured by the share of colleagues of the same institution within the coauthors of a researcher. Under this interpretation, the negative relationship between outcomes and the above measure of "density" (actually a measure of diversity) represents a beneficial effect of collaborating with researchers from other institutions. The left panel of Figure 3, as explained above, shows that collaboration with researchers from other faculties within the same university does not exhibit any beneficial effect, which contradicts the diversity hypothesis.

Another measure of diversity is the number of universities from which coauthors participate. If coauthors come from a larger number of universities, it may imply that a variety of opinions and research methods are incorporated into the research, enhancing its quality. This measure of diversity is calculated for individual papers so that it is only meaningful for the number of citations among outcomes. In the right panel of Figure 3, the solid line lies to the Southeast of the dashed and dotted lines, which implies that the number of citations of a paper increases as the number of universities increases; hence, the diversity of the coauthors increases. We also note a complication coming from endogeneity: in medicine, multiple institutions participate in a clinical trial to facilitate an accumulation of observations. A large number of observations naturally imply the higher quality of the clinical trial. Therefore, multiple participation may represent a larger number of observations, rather than diversity of opinions or methods.

As for the attributes of coauthors, we examine a wide variety of characteristics. These attributes may be viewed as either quantitative or qualitative. A typical quantitative attribute is whether a researcher's coauthors publish many papers, while typical qualitative attributes include whether his or her coauthors publish papers in core journals or that are cited in clinical guidelines.

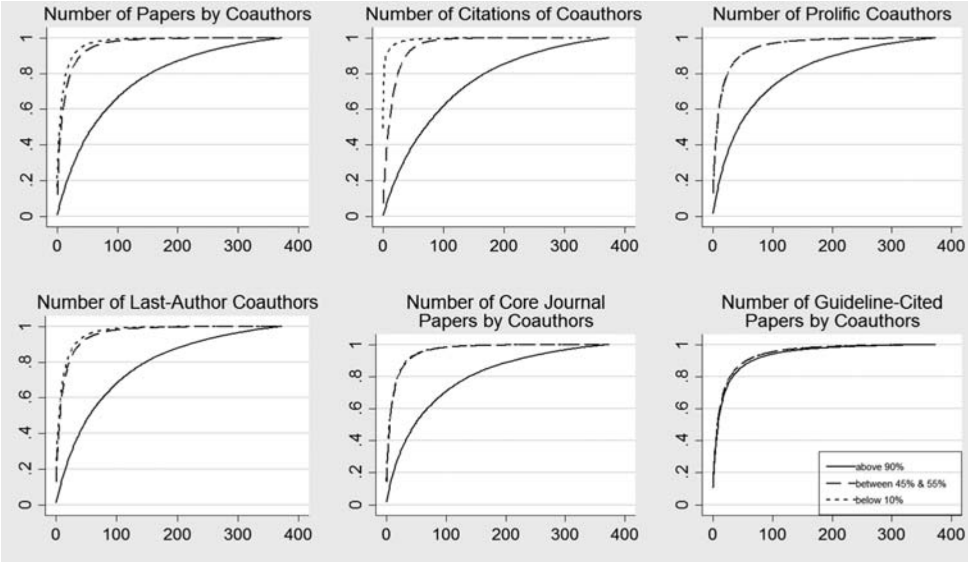
The full range of outcome variables, listed from the most quantitative to most qualitative, is as follows: number of papers that coauthors published and citations of the papers published by coauthors, number of coauthors who publish a very large number of papers or highly cited papers, number of coauthors who are the last authors in their papers, and number of papers published in core journals or cited in guidelines.

Referring back to Table 1, each researcher collaborates with 19 coauthors, on average, and the number of papers published by coauthors, averaged over researchers, is 134, which are cited 2,128 times in total. The most cited coauthors' paper is cited 818 times. On average, a little less than two coauthors are prolific or highly cited researchers (defined in the data section). Further, more than 16 coauthors have experienced last authorship. The coauthors of each researcher published 15 papers in core journals and had roughly one paper cited in guidelines.

Figures 4-1 and 4-2 show the cumulative distribution functions of researchers whose coauthors possess indicated attributes.

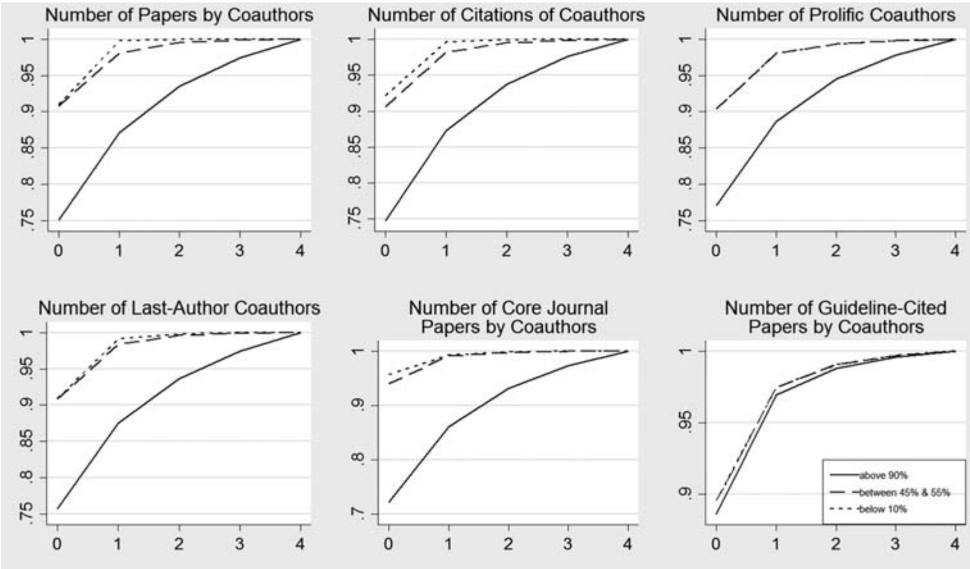
The solid line in the upper left panel of Figure 4-1, for example, depicts the cumulative distribution of researchers, with the number of citations indicated by the

Figure 4-1：Shapes of Cumulative Distribution Functions of the Number of Citations by the Degree of Link Attributes



Note: To make the difference of distributions easy at which to look, they are drawn without highest 1%.

Figure 4-2：Shapes of Cumulative Distribution Functions of the Number of Core Journal Papers by the Degree of Link Attributes



Note: To make the difference of distributions easy at which to look, they are drawn without highest 1%.

horizontal axis in the case where the number of papers that coauthors published is the highest 10%. Just as in the case of agglomeration, the solid line located to the Southeast implies that a larger number of coauthors' papers will raise the number of citations of a researcher's published papers. As we go from the left to right and, then, to the second and third rows, this trend is preserved. In the left panel in the third row, researchers are divided into two categories: those whose coauthors published guideline-cited papers and those whose coauthors did not. This is because coauthors with guideline-cited papers are rare; thus, even most of the researchers in the highest 10% do not publish guideline-cited papers. The attributes are arranged so that going from left to right and from top to bottom corresponds to the pecking order of outcomes from quantitative to qualitative outcomes. Therefore, quantitative coauthor attributes, as well as qualitative attributes, contribute to larger numbers of papers published by researchers.

Figure 4-2 shows the effects of link attributes on the number of papers published in core journals. A similar trend as before is observed. However, since papers published in core journals are rare, the difference between the case of the highest 10% (solid line) and lowest 10% (dotted line) seems to be smaller than the difference, when the outcome is taken to be the number of citations (Figure 4-1).

In summary, quantitative coauthors are good for quantitative outcomes, while not as beneficial to qualitative outcomes. On the other hand, qualitative coauthors are similarly good both for quantitative and qualitative outcomes.

5. Econometric Analysis

This section presents a couple of results of the formal estimation of the determinants of research productivity. This estimation is based on data for individual researchers, and we regress various outcome measures on the determinants of research productivity. The outcome variables are, in the order from quantitative to qualitative, the number of published papers, number of citations and highly cited papers, number of papers published in top journals, and number of papers cited in clinical guidelines.

The first set of determinants, externalities, consists of two groups: agglomeration and network effects. The variables representing agglomeration are the number of papers published by researchers in the institution as well as number of researchers in the institution. Network effects are assumed embodied in the number of coauthors with whom a researcher works. As was explained in Section 4, to investigate the range of agglomeration/network effects, the above variables are calculated for various units/areas at different distances.

Among the second set of determinants, the structure and nature of the research network, we are unable to incorporate structural characteristics in the regression models because the variables representing them are calculated on the basis of individual papers. Moreover, concerning the nature of the research network, we specifically use the attributes of coauthors, of which a typical quantitative attribute is whether a researcher's coauthors publish many papers and typical qualitative attributes include whether his or her coauthors publish papers in core journals or they are cited in clinical guidelines.

Since quantitative analysis in this paper is preliminary, we utilize simple linear regression with a technical correction to the dominant number of zeros in the data. Even though the outcome variables are continuous, the Tobit model is employed to account for the fact that the value zero is dominant in the data (i.e., the data is censored).[4] One group of regression models includes agglomeration at various distances as explanatory variables, while another group includes the network effects of different ranges.

First, we estimate models with agglomeration or network effect variables together with the number of papers each researcher published. Of course, when the number of published papers is the dependent variable, the last explanatory variable is omitted. The number of observations is large (322,748) resulting in precise estimates. This enables us to discuss the magnitude of agglomeration/network effects at various distances.

Then, various network characteristics are added to the above models one at a time. The number of observations is the same as the above estimation, again leading to precise estimates.

Table 2 shows the results for models with the number of researchers as agglomeration variables.

We also estimate models with the number of papers as agglomeration variables with similar results; hence, only the results of estimation using the number of researchers

---

[4] In the case of the number of papers, researchers who published at least one paper are included in the data. In this case, no censoring occurs in the estimation, so that the Tobit estimation is just the ordinary least squares estimation.

Table 2：Agglomeration Effect

Explanatory variables	Outcomes Number of papers Coefficient	Standard error	t-value	Number of citations Coefficient	Standard error	t-value
Economies of scale						
Faculty	0.00746	0.00017	44.55	0.03501	0.00257	13.64
Institution less faculty	0.00037	0.00002	22.43	0.00819	0.00026	32.12
City less institution	0.00006	0.00001	10.53	0.00001	0.00008	0.07
Nation less city	0.00000	0.00000	-8.11	0.00018	0.00000	53.71
Number of papers	n.a.	n.a.	n.a.	16.26369	0.02660	611.43
Constant	2.13928	0.01042	205.28	-21.29274	0.17143	-124.20
Number of observations	322,748			322,748		
Log likelihood	-885,011			-1,593,771		

Explanatory variables	Outcomes Core journal Coefficient	Standard error	t-value	Guidelines Coefficient	Standard error	t-value
Economies of scale						
Faculty	0.00391	0.00020	19.30	0.00339	0.00061	5.60
Institution less faculty	0.00012	0.00002	5.08	-0.00079	0.00008	-9.83
City less institution	-0.00002	0.00001	-2.22	0.00002	0.00002	0.92
Nation less city	0.00000	0.00000	2.69	0.00001	0.00000	9.24
Number of papers	0.29067	0.00186	156.39	0.17045	0.00449	38.00
Constant	-4.81178	0.02604	-184.81	-10.58591	0.16256	-65.12
Number of observations	322,748			322,748		
Log likelihood	-169,360			-25,809		

are reported.

The first model is the case where the number of published papers is the outcome variable. The number of researchers in the same faculty of each researcher positively increases the number of papers. When the distance of agglomeration is enlarged to the same university, but different faculty, it still exerts a positive influence on the number of papers, but by a much smaller magnitude. When the range is further expanded, the coefficient on the number of researchers in the same city, but different university, becomes smaller still, and that on the number of researchers in the same country, but different city, even change to negative. The same trend is observed in models with other outcome indicators, including the number of papers published in top journals and those cited in guidelines.

In sum, the agglomeration effect is limited to nearby locations. It is positive at the faculty level, but diminishes very rapidly to university, outer city, and country levels. It might be marginally beneficial to have a large agglomeration of researchers in other faculties in the same university, but agglomeration in the outer city or other parts of the country is scarcely helpful. On the other hand, even if agglomeration is not large at the city or country level, economies of scale could be created within a faculty, like an island in the ocean.

Table 3：Network effect

Explanatory variables	Outcomes Number of papers Coefficient	Standard error	t-value	Number of citations Coefficient	Standard error	t-value
Network effect						
Faculty	0.10008	0.00242	41.37	1.05041	0.03697	28.41
Institution less faculty	0.01084	0.00318	3.40	1.44378	0.04857	29.73
City less institution	-0.00127	0.00476	-0.27	1.14055	0.07244	15.74
Nation less city	0.03822	0.00263	14.54	2.33343	0.03970	58.78
Overseas	0.06050	0.00276	21.90	3.04400	0.04155	73.27
Number of papers	n.a.	n.a.	n.a.	16.25688	0.02639	616.06
Constant	1.83387	0.01679	109.25	-25.68246	0.26230	-97.91
Number of observations	322,748			322,748		
Log likelihood	-885,701			-1,591,849		

Explanatory variables	Outcomes Core journal Coefficient	Standard error	t-value	Guidelines Coefficient	Standard error	t-value
Network effect						
Faculty	-0.02403	0.00355	-6.76	-0.09400	0.01127	-8.34
Institution less faculty	-0.03034	0.00475	-6.39	-0.05423	0.01451	-3.74
City less institution	-0.00399	0.00678	-0.59	0.00636	0.01893	0.34
Nation less city	0.10842	0.00312	34.73	0.09061	0.00804	11.27
Overseas	0.15768	0.00301	52.37	0.11504	0.00740	15.55
Number of papers	0.29206	0.00184	158.82	0.17072	0.00447	38.23
Constant	-4.80490	0.03205	-149.93	-10.21562	0.16679	-61.25
Number of observations	322,748			322,748		
Log likelihood	-167,215			-25,606		

Next, we estimate models with the number of coauthors at different distances as an agglomeration variable. Table 3 shows these results.

When the number of published papers is the outcome variable, the number of coauthors in the same faculty of each researcher positively increases the outcome. When the distance of agglomeration is enlarged, just as the case of the agglomeration effect, the positive effect diminishes rapidly. The coefficient on city less university is very small and that on country less city is negative.

If we use the number of citations or maximum citations as outcome variables, the declining trend reverses. In the case of the number of citations, for example, the coefficient on faculty is significantly positive, and that on university less faculty increases. After a small dip at city less university, the coefficient strongly increases at country less city. Remarkably, a network with overseas researchers further increases the number of citations.

When the outcome variable is the number of papers published in top journals, the coefficients on faculty, university less faculty, and city less university are surprisingly negative, although the last one is not significant. Again, remarkably, coauthor networks at the country level and overseas exert positive effects on publication in top journals. Moreover, similar results are obtained if the number of papers cited in guidelines is taken as the outcome variable.

To sum up, with respect to the quantitative outcome, a network effect is very limited to nearby locations, just as the agglomeration effect. Once the indicators of high quality research are used as the outcome variables, neighborhood networks are not effective or even harmful. However, distant networks, especially those overseas, are very effective in conducting quality research.

In Tables 2 and 3, we can also see the effect of the number of papers published by each researcher. An interesting question might be whether producing a lot of papers implies that the researcher is conducting high quality research. Moreover, the results in these tables are virtually identical; as the number of published papers increases, research outcomes increase. This is true for all outcome measures, including the number of published papers, number of citations, number of papers published in top journals, and number of papers cited in clinical guidelines. Therefore, producing many papers would enhance both the quantity and quality of research.

The magnitude of impact, however, is diminishing; it is highest for the number of citations, decreases with the maximum citations and core journal papers, and is lowest with guideline citations. Of course, this is natural considering the scarcity of core journal or guideline-cited papers. However, it has an important implication that, even if the number of citations strongly increases with the number of published papers, the numbers of core journal papers and guideline citations increase only modestly. A quantitative increase in the number of papers may not be an efficient way to raise the quality of research.

Table 4 shows the results of the estimation of the models with network characteristics.

We added each attribute of coauthors into the model one at a time, and also included agglomeration or network effect variables as well as the number of papers of the researcher, which is naturally excluded when the outcome variable is the number of papers. Moreover, when the number of papers is the outcome variable, all network characteristics, both quantitative and qualitative, are significantly positive. The number of papers that coauthors publish and number of citations of the coauthors' papers increase the number of papers of each researcher, although the effect seems marginal. Further, this effect is much stronger when coauthors are prolific authors or authors of highly cited papers. High quality coauthors, such as those publishing in core journals or cited in guidelines, also exert strong network effects, while the effect of

Table 4：Network Attributes（with agglomeration variables）

	Outcome variables					
		Number of papers			**Number of citations**	
Attributes of coauthors	Coefficient	Standard error	t-value	Coefficient	Standard error	t-value
Number of papers coauthors publish	0.00335	0.00001	431.03	0.00633	0.00018	34.43
Number of citations of coauthors' papers	0.00016	0.00000	365.66	0.00148	0.00001	163.73
Number of prolific coauthors	0.24715	0.00053	463.56	0.43364	0.01336	32.45
Number of highly cited coauthors	0.10126	0.00044	229.79	0.91517	0.00753	121.48
Number of last authors	0.03091	0.00007	462.81	0.05338	0.00167	31.93
Number of core journal papers	0.01997	0.00006	333.28	0.06661	0.00121	55.18
Number of guideline-cited papers	0.14503	0.00072	200.81	0.11697	0.01225	9.55

	Outcome variables					
	Core journal papers			**Guideline-cited papers**		
Attributes of coauthors	Coefficient	Standard error	t-value	Coefficient	Standard error	t-value
Number of papers coauthors publish	-0.00011	0.00001	-9.91	-0.00010	0.00003	-3.67
Number of citations of coauthors' papers	0.00000	0.00000	-4.54	0.00000	0.00000	-2.96
Number of prolific coauthors	-0.01077	0.00085	-12.70	-0.01007	0.00204	-4.94
Number of highly cited coauthors	0.00059	0.00048	1.25	-0.00216	0.00109	-1.98
Number of last authors	-0.00109	0.00011	-10.39	-0.00062	0.00023	-2.70
Number of core journal papers	0.00101	0.00007	13.81	-0.00001	0.00015	-0.09
Number of guideline-cited papers	0.00277	0.00076	3.63	0.02632	0.00125	21.07

coauthors who have experienced last authorship is modest.

When the outcome variable is the number of citations, the overall trend is similar to the case of the number of papers, although the effect of each characteristic is generally larger, reflecting the fact that citations are 15 times larger than the number of publications.

If we take publication in core journals as an outcome measure, the picture changes drastically. Quantitative indicators generally lose effectiveness, while qualitative indicators remain positive. The coefficients on the number of papers or citations, prolific or last authors become negative, and those on core journal publications or guideline citations retain effectiveness. The difference between quantitative and qualitative indicators looms large in the case of guideline citation. Almost all quantitative attributes are negative. High quality coauthors who published core journal papers or guideline cited papers strongly increase the number of core journal papers and guideline-cited papers by each researcher. However, the effect of the number of papers published in core journals by coauthors on the number of guideline-cited papers is an exception, which is puzzling.

In summary, the quantitative attributes of coauthors are beneficial to quantitative outcomes, but not to qualitative outcomes. On the other hand, the qualitative attributes of coauthors are beneficial for both quantitative and qualitative outcomes. Yet, fostering collaboration to improve the quality of research requires more than encouraging prolific or highly cited coauthors; high quality researchers are essential to high quality collaboration.

Before ending this section, some remarks are in order concerning the endogeneity of the explanatory variables. Establishing collaboration and building networks with other researchers are not exogenously imposed actions. One chooses to collaborate and enter into a network because to do so is beneficial for research. If this is the case, we cannot simply interpret the correlation between research outcomes and links with excellent coauthors as indication of the causal effect of links with excellent coauthors on research outcomes. Instead of links generating outcomes, it may be that expected outcomes entail links with excellent collaborators. In other words, the effect on outcomes of links with distant researchers are, at least partially, not a cause, but a consequence, of expected outcomes.

In these situations, it would be better to control endogeneity by, for example, resorting to the instrumental variable (IV) method. However, the difficulty of finding adequate instruments prevents us to do so. An example of a candidate IV is the distance of a researcher from other researchers. The researchers with whom to collaborate are often those who are near to the researcher (in the sense of geography, languages, research styles, preferences, psychological tendencies, etc.). Thus, aggregate distance (in a suitable sense) from other researchers could be used as an IV, and the controlling of endogeneity is left to future research.

6. Conclusion

This paper investigates the mechanisms of the production of high quality research and sheds light on the way in which to promote academic research, by analyzing the determinants of research outcomes. In addition, this paper contributes to the measurement of the quality of academic research by introducing several new indicators of quality.

The analyzed determinants include the benefits of agglomeration and network effects, and effects of the structure of research networks as well as the nature of links.

The agglomeration effect is limited to nearby locations. To view this in an opposite direction, even if agglomeration is not large at the city or country level, economies of scale could be created within a faculty, like an island in the ocean.

With respect to the quantitative outcome, the network effect is very limited to nearby locations, just as the agglomeration effect. However, distant networks, especially those overseas, are very effective in conducting quality research.

Concerning the network characteristics, the linkage with prolific coauthors and/or organizers greatly contributes to a quantitative increase in research outcomes, but very little to qualitative outcomes. Instead, the linkage with quality coauthors enhances quality research. In other words, fostering collaboration to improve the quality of research requires more than encouraging prolific or highly cited coauthors.

References

Amsterdam, Ezra A., Nanette K. Wenger, Ralph G. Brindis, Donald E. Casey, Jr., Theodore G. Ganiats, David R. Holmes, Jr., *et al.* (2014). 2014 AHA/ACC guideline for the management of patients with non-ST-elevation acute coronary syndromes: A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation,* 130 (25), e344-426. doi: 10.1161/CIR.0000000000000134

Azoulay, Pierre, Waverly Ding, and Toby Stuart. (2007). The determinants of faculty patenting behavior: Demographics or opportunities? *Journal of Economic Behavior & Organization,* 63 (4), 599-623. doi: 10.1016/j.jebo.2006.05.015

Azoulay, Pierre, Waverly Ding, and Toby Stuart. (2009). The impact of academic patenting on the rate, quality and direction of (public) research output. *Journal of Industrial Economics,* 57 (4), 637-676.

Azoulay, Pierre, Joshua S. Graff Zivin, and Jialan L. Wang. (2010). Superstar extinction. *Quarterly Journal of Economics,* 125 (2), 549-589.

Banerjee, Abhijit, Arun G. Chandrasekhar, Esther Duflo, and Matthew O. Jackson. (2013). The diffusion of microfinance. *Science,* 341 (6144), 363-+. doi: 10.1126/science.1236498

Carlino, Gerald, and William R. Kerr. (2015). Chapter 6 - Agglomeration and Innovation. in J. V. H. Gilles Duranton & C. S. William (Eds.), *Handbook of Regional and Urban Economics* (Vol. 5, pp. 349-404): Elsevier.

Duranton, Gilles, and William R. Kerr. (2015). The logic of agglomeration. *National Bureau of Economic Research Working Paper Series,* No. 21452. doi: 10.3386/w21452

Feldman, Maryann P., and David B. Audretsch. (1999). Innovation in cities: Science-based diversity, specialization and localized competition. *European Economic Review,* 43 (2), 409-429. doi: 10.1016/S0014-2921 (98) 00047-6

Lay, Gunter, and Carsten Dreher. (1992). Diversity on firm level - different modes of usage of NC/CNC-machine tools in German mechanical engineering. *Revue d'Économie Industrielle,* 59 (1), 86-98.

Lichtenberg, Frank R. (2013). The impact of biomedical knowledge accumulation on mortality: A bibliometric analysis of cancer data. *National Bureau of Economic Research Working Paper Series,* No. 19593. doi: 10.3386/w19593

Newman, M. E. (2001a). Scientific collaboration networks. I. Network construction and fundamental results. *Physical Review E,* 64 (1 Pt 2), 016131. doi: 10.1103/PhysRevE.64.016131

Newman, M. E. (2001b). Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review E,* 64 (1 Pt 2), 016132. doi: 10.1103/PhysRevE.64.016132

Nguyen, Phuong, Pier-Paolo Saviotti, Pier-Paolo Trommetter, and Bernard Bourgeois. (2005). Variety and the evolution of refinery processing. *Industrial and Corporate Change,* 14 (3), 469-500. doi: 10.1093/icc/dth059

NISTEP. (2015). *Japanese Science and Technology Indicators 2015.*

O'Gara, Patrick T., Frederick G. Kushner, Deborah D. Ascheim, Donald E. Casey, Jr., Mina K. Chung, James A. de Lemos, *et al.* (2013). 2013 ACCF/AHA guideline for the management of ST-elevation myocardial infarction: A report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *Circulation,* 127 (4), e362-425. doi: 10.1161/CIR.0b013e3182742cf6

Saka, Ayaka, Masatsura Igami, and Hiroyuki Tomizawa. (2015). Analysis of the structure of Japan's scientific publication production using linkage data of bibliographic database (Web of Science) and database of Grants-in-Aid for Scientific Research (KAKEN). *Research Material,* 237.

Stirling, Andrew. (1998). *On the economics and analysis of diversity.* Retrieved from Brighton: http: //sro.sussex.ac.uk/28823/

Stirling, Andy. (2007). A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society Interface,* 4 (15), 707-719. doi: 10.1098/rsif.2007.0213

van Rijnsoever, Frank J., Jesse van den Berg, Joost Koch, and Marko P. Hekkert. (2015). Smart innovation policy: How network position and project composition affect the diversity of an emerging technology. *Research Policy,* 44 (5), 1094-1107. doi: 10.1016/j.respol.2014.12.004

Yancy, Clyde W., Mariell Jessup, Biykem Bozkurt, Javed Butler, Donald E. Casey, Jr., Mark H. Drazner, *et al.* (2013). 2013 ACCF/AHA guideline for the management of heart failure: A report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *Journal of the American College of Cardiology,* 62 (16), e147-239. doi: 10.1016/j.jacc.2013.05.019

Yegros-Yegros, A., I. Rafols, and P. D'Este. (2015). Does interdisciplinary research lead to higher citation impact? The different effect of proximal and distal interdisciplinarity.

Plos One, 10 (8). doi: 10.1371/journal.pone.0135095
Yonetani, Yutaka, Kenta Ikeuchi, and Terutaka Kuwahara. (2013). *An analysis on the relationship between input and output in the production of articles in universities* – An approach using Web of Science and Survey of R&D. *NISTEP Discussion Paper*, 89 (in *Japanese*)