



Available online at [www.sciencedirect.com](http://www.sciencedirect.com)



Research Policy 34 (2005) 259–285

---

---

research  
policy

---

---

[www.elsevier.com/locate/econbase](http://www.elsevier.com/locate/econbase)

## Scientific teams and institutional collaborations: Evidence from U.S. universities, 1981–1999

James D. Adams <sup>a,\*</sup>, Grant C. Black <sup>b</sup>, J. Roger Clemmons <sup>c</sup>, Paula E. Stephan <sup>d</sup>

<sup>a</sup> Department of Economics, Rensselaer Polytechnic Institute and NBER, 3504 Russell Sage Laboratory, Troy, NY 12180-3590, USA

<sup>b</sup> Indiana University at South Bend, USA

<sup>c</sup> Institute for Child Health Policy, University of Florida, USA

<sup>d</sup> Georgia State University, USA

Received 6 August 2004; received in revised form 29 December 2004; accepted 16 January 2005

Available online 23 March 2005

---

### Abstract

This paper explores recent trends in the size of scientific teams and in institutional collaborations. The data derive from 2.4 million scientific papers written in 110 top U.S. research universities over the period 1981–1999. The top 110 account for a large share of published basic research conducted in the U.S. during this time.

We measure team size by the number of authors on a scientific paper. Using this measure we find that team size increases by 50% over the 19-year period. We supplement team size with measures of domestic and foreign institutional collaborations, which capture the geographic dispersion of team workers. The time series evidence suggests that the trend towards more geographically dispersed scientific teams accelerates beginning with papers published at the start of the 1990s. This acceleration suggests a sharp decline in the cost of collaboration. Our hypothesis is that the decline is due to the deployment of the National Science Foundation's NSFNET and its connection to networks in Europe and Japan after 1987.

Using a panel of top university departments we also find that private universities and departments whose scientists have earned prestigious awards participate in larger teams, as do departments that have larger amounts of federal funding. Placement of former graduate students is a key determinant of institutional collaborations, especially collaborations with firms and with foreign scientific institutions. Finally, the evidence suggests that scientific output and influence increase with team size and that influence rises along with institutional collaborations. Since increasing team size implies an increase in the division of labor, these results suggest that scientific productivity increases with the scientific division of labor.

© 2005 Elsevier B.V. All rights reserved.

*JEL classification:* L3; O3

*Keywords:* Science; Research and development; Collaboration; Teams

---

### 1. Introduction

---

\* Corresponding author. Tel.: +1 518 276 2523;  
fax: +1 518 276 2235.

E-mail address: [adamsj@rpi.edu](mailto:adamsj@rpi.edu) (J.D. Adams).

Over the past century teams of scientific specialists have largely replaced the independent scientist,

much as corporate R&D laboratories have largely replaced the independent inventor. This trend towards larger teams is strongly evident in nearly all data on scientific research, including our own.<sup>1</sup> Advancing instrumentation and the sheer quantity of what there is to know have pushed it, while improving transportation and communications have pulled it, to its present state of development. Purely pencil-and-paper research is the lone exception to the rule. But this forms a dwindling share of science and may give in to the same forces that have lead to large-scale research elsewhere.

The study of scientific teams is important for at least two reasons. First, it brings to light changes in the research production function that would otherwise stay hidden. In particular, scientific collaboration might increase the effectiveness of research, just as specialization increases general productive efficiency.<sup>2</sup> The evidence on efficiency is suggestive but not definitive. Collaborative research is more highly cited (Presser, 1980; Sauer, 1988), suggesting that collaboration does raise quality. But more able researchers also attract more coworkers (Zuckerman and Merton, 1973) so that separating efficiency from talent in a cross-section is not easy. In addition the higher citation rate could reflect hidden self-citations between team members. Still, the size of scientific teams has increased steadily with time, so that growth in capital, knowledge, transport, and communication could be responsible for the rising propensity to collaborate, even as talent has remained about the same.

A second reason why collaboration is important lies in its role as a channel of knowledge flows between scientists. And since collaborators are increasingly found in different institutions and countries, the entire subject is relevant to the tendency for knowledge to flow more readily and over greater distances than ever before and to the enlargement of the scale and scope of knowledge spillovers.

<sup>1</sup> See Zuckerman and Merton (1973), de Solla Price (1986), and Hicks and Katz (1996), for trends in the size of scientific teams since 1900. Wiener (1994), writing about mid-century, strongly disapproves of the notion that corporate R&D laboratories might supplant individual researchers and inventors. And yet Mowery and Rosenberg (1998, Ch. 2) describes exactly this process of replacement.

<sup>2</sup> The form of the research production function is central to the properties of growth models, as a comparison between Romer (1990) and Jones (1995) reveals. Thus, the findings of this paper could prove indirectly relevant to growth theory.

Our analysis begins with a simple model of team formation in which talent and team size increase with the stock of knowledge, the quality of an institution's research management, and prior success at grant-raising. In addition the model addresses the causes behind the geographic dispersion of research teams, also known as institutional collaboration. This is shown to increase when research budgets are more responsive to geographic dispersion and when coordination costs increase less rapidly with that dispersion.

The paper continues with extensive empirical work on scientific teams and institutional collaborations. This research relies on a database that covers most of U.S. science during the years 1981–1999. The empirical work that uses this data falls into three parts. The first describes time trends in team size and institutional collaboration across science using a series of figures. The second describes trends in teams and collaborations by field of science. Given the complexity of the field and time dimensions, we present the second set of findings using tables rather than figures. The third and final part of the empirical work consists of regression analysis of a panel of university-fields observed over time. This analysis examines the underlying determinants of team size and institutional collaborations. The regression findings are in general agreement with the implications of the model.

Descriptive findings include the following, which apply out overall and for most field. Team size increases by about 50% over the period 1981–1999. This trend towards larger teams is found to accelerate, rising from a 2.19% annual rate of growth in the 1980s to a 2.57% rate in the 1990s. This is an acceleration factor of 17% ( $2.57/2.19 - 1$ ) between the two decades. Using mileage indicators between the top 110 U.S. universities that form the core of our data, we study geographic dispersion directly. The annual rate of growth in mileage rises from 3.53% in the 1980s to 4.45% in the 1990s, an acceleration of 29%.

During the period 1981–1999 the rate of domestic institutional collaboration more than doubled between U.S. universities and between U.S. universities and U.S. firms. Foreign collaborations, while not as common, increased five-fold. Of all regions, collaboration with Asia increased most rapidly, followed by Europe, with all other regions trailing by a considerable margin. These differences reflect differences in the growth rates of scientific research by region of the world. As with the

other indicators foreign collaboration accelerates going into the 1990s. The foreign share in institutional collaborations rose annually by 5.11% during the 1980s, but by 7.41% during the 1990s. Thus, growth in foreign collaboration accelerates by 45% ( $7.41/5.11 - 1$ ) over the two decades. We conclude that institutional collaboration, especially in the international sphere, has undergone accelerating growth in recent decades.

In additional work we use regression analysis to investigate factors that might promote or hinder the formation of teams and collaborations. This analysis is carried out using a panel of 12 main sciences in the top 110 U.S. universities. Hence, the observations are at the level of university-fields over time. Our results include the following: a larger stock of federally funded R&D, private control of a university, and the number of prestigious awards increase team size. Collaboration with other U.S. universities is an increasing function of the stock of R&D, private control, and the number of recent Ph.D.'s placed in other top 110 schools. Collaboration with foreign institutions of science is found to increase with the stock of R&D, private control, the number of prestigious awards, and with placement of recent Ph.D.'s in leading research countries. Here the results are more fragile, and there is evidence that the stock of R&D and private control trade places with the number of prestigious awards. In addition we examine collaboration with U.S. firms, finding again that R&D and private control increase joint scientific research with firms. But in this case, prestigious awards decrease collaboration with firms, probably because universities earning such awards emphasize basic rather than applied research. As with foreign collaboration, placement of recent Ph.D.'s in industry is a significant factor in determining collaboration between universities and firms. The greater importance of Ph.D. placements in the case of foreign and corporate collaborations than in university collaborations is consistent with the scarcity of close substitutes for former Ph.D.'s as coworkers in foreign and corporate environments.

The empirical work concludes with a study of the role of team size and collaboration in the determination of research output as measured by papers and citations. The papers and citations are "fractionated," in that they consist of estimated proportions that a university and field contribute to both. We find that papers and especially citations increase with team size, but that the role of shares of institutional affiliations in the produc-

tion of papers is less clear. Universities that collaborate more with foreign institutions, and especially other top 110 schools, produce fewer papers, holding team size constant, in part because a fraction of the papers is "outsourced". On the other hand collaboration with foreign institutions and other top 110 schools is linked to an increase in total citations, so that a trade off of fewer papers in return for larger overall scientific influence may be taking place.

The rest of this paper consists of six sections. Section 2 reviews the literature and models the underlying determinants of scientific teams and institutional collaborations. Section 3 describes the database and the calculations that we have undertaken using it. Section 4 presents time series evidence on scientific teams and institutional collaborations. Section 5 compares trends in these indicators across scientific disciplines. Section 6 provides regression evidence on the determinants of team size and institutional collaborations using a panel of university-fields. Section 7 concludes.

## 2. Analytical framework

The economics literature on teams is both theoretical and empirical. One line of theoretical work examines the problem of free-riding and proposes incentive schemes that punish shirking (Holmstrom, 1982; Kandel and Lazear, 1992). Another line of research examines the relationship between specialization, team size, and the extent of the market (Becker, 1985; Becker and Murphy, 1992), while (Rosen, 1982) looks at the role of managerial talent in determining the size of the firm and its work force using an efficient supervision model.

Recent empirical research on teams in steel mills by Ichniowski et al. (1997) finds that innovative management practices that promote cooperation in teams and offer pay incentives increase productivity in steel mini-mills over traditional management practices that tend to limit worker responsibility. Other empirical research examines institutional collaborations in science and technology (Arora and Gambardella, 1990; Mowery, 1992; Powell, 1996; Stephan and Levin, 2000; Zucker et al., 2001; Adams et al., 2001; Adams, 2002; Adams et al., 2003; Adams and Marcu, 2004; Adams, 2005). These papers tend to find that institutional collaboration is related to the complementarity of skills, often abetted by policy and by the increasing complexity of scientific problems.

The present paper concerns small teams of co-workers in scientific research. However, owing to the location of team members in different scientific institutions, the paper is concerned with institutional collaborations as well. Unlike some previous literature this paper abstracts from free-riding and supervision, because the system of reward for priority in discovery severely punishes individual shirking by team members and rewards good work with publication, reputation, and income. Following Becker and Murphy (1992) we regard the analysis of teams as closely linked with specialization, the division of labor, and the location of team members, because these factors play large roles in the empirical analysis and influence the efficiency of scientific research.

In the subsequent exposition we assume that scientific research yields satisfaction to investigators, but we suppose that this gain in utility is insufficient for investigators to self-finance their research. We base this assumption on the rarity of scientific research in societies that lack government or private philanthropic support for science. But if self-finance is ruled out, then it follows that the quantity of research is subject to an externally imposed research budget constraint, even though the budget is responsive to grant-raising efforts.<sup>3</sup> In conformity with the empirical work and with the perception that skill, specialization, and the division of labor are inter-related and essential elements in the formation of scientific teams, we assume that the decision variables are the skill of team workers  $s$ , the size of the team  $n$ , and distance, or geographic dispersion between team workers  $D$ .<sup>4</sup> Geographic dispersion exceeds zero if and only if the team involves institutional collaboration so that we regard  $D$  as an indicator of such collaboration.

Research output  $Q$  is produced according to a Cobb–Douglas production function that depends on skill and team size. Therefore,

$$Q = As^\alpha n^\beta. \quad (1)$$

<sup>3</sup> In the analysis we assume that the research budget  $R$  is at a maximum with respect to grant-raising efforts.

<sup>4</sup> We abstract from the inter-relatedness of skill and distance, which implies that  $s = s(D)$ ,  $s' > s'' < 0$ . This constraint would be relevant if abilities were noticeably scarce within a given distance. The analysis in the paper could easily be extended to deal with this complicating factor.

The exponents obey the inequalities  $0 < \alpha < 1, 0 < \beta < 1, \alpha + \beta < 1$  so that production is subject to decreasing returns to scale. The parameter  $A$  represents total factor productivity. It could represent the university environment and the ability or eminence of faculties that are matched to these environments in advance of the formation of scientific teams.  $A$  also undergoes independent increases as knowledge expands, because knowledge is a factor of production, even though it is fixed with respect to individual researchers. In turn  $A$  tends to increase skill and team size. Thus, complexity of projects can be viewed as an indicator for the stock of knowledge and other sources of total factor productivity of the knowledge production function that give larger teams an increasing advantage. It has been suggested to us that sociological “norms” of scientific fields have changed in favor of larger teams. However, another interpretation is that larger and higher-skilled teams are more efficient as knowledge increases, so that norms are simply a reflection of efficiency rather than an independent causal factor that increases the division of scientific labor. And besides all the above, the increasing emphasis by funding agencies on team awards involving large grants and multiple scientific institutions is consistent with the advantages of larger teams as driven by  $A$  in this framework, in which the advantages are likely driving the policies.

On the cost side we assume that the research budget  $R$  must cover wage costs of all team members, as well as a fixed cost  $F$  that depends on dispersion of team workers, representing coordination costs. Of course, geographic dispersion entails benefits as well as costs. In part the gain derives from the additional funding that can be secured, as the evidence presented below on international and firm–university collaboration suggests. This is the productive role of dispersion in the present analysis. The research budget constraint is:

$$R(D) = w(s)n + F(D). \quad (2)$$

The amount of funding is assumed to be a concave function of distance, so that

$$R' > 0, \quad R'' < 0. \quad (3)$$

Thus, the returns to dispersion are diminishing. Below we allow for both a shift in funding and also of the sensitivity of funding to distance or dispersion, which are plausible causes of institutional collaborations. For example, funding could increase as a result of prior

awards. Alternatively, an increase in the urban density of universities would tend to increase the sensitivity of funding to distance. In (2) the wage rate  $w$  is an increasing and convex function of skill, as in the phenomenon of “superstars” (Rosen, 1981). This yields the following properties of the wage function:

$$w' > 0, \quad w'' > 0. \quad (4)$$

We suppose that the fixed cost  $F$ , which represents co-ordination costs, is an increasing function of distance  $D$  owing to the difficulty of meeting and communicating which  $D$  imposes. Therefore, the properties of the fixed cost function are

$$F' > 0, \quad F'' > 0. \quad (5)$$

In (5) coordination costs are an increasing function of distance. As in the case of funding  $R$ , changes in the fixed costs of scientific teams are plausible and realistic features of cross-sectional and time series data. Fixed costs and their sensitivity to geographic dispersion tend to decline over time, with improvements in transport and telecommunications.  $F$  also tends to decrease with prior investments in team workers, especially graduate students, which make working at a distance less costly. This is likely to be a potent factor in both cross-sectional data, where universities with larger and more highly ranked graduate programs are more likely to engage in institutional collaborations; and in time series data, due to the growth of graduate programs over time.

The problem is one of maximizing output (1) subject to the research budget constraint (2) and their properties as expressed in (3)–(5), where the control variables are skill  $s$ , team size  $n$ , and distance  $D$ . The Lagrangian function for this problem is

$$L = As^\alpha n^\beta + \lambda[R(D) - w(s)n - F(D)] \quad (6)$$

First order conditions for (6) are

$$\begin{aligned} \frac{\partial L}{\partial s} &= \frac{\alpha Q}{s} - \lambda w'n = 0 \\ \frac{\partial L}{\partial n} &= \frac{\beta Q}{n} - \lambda w = 0 \\ \frac{\partial L}{\partial D} &= \lambda(R' - F') \leq 0 \\ \frac{\partial L}{\partial \lambda} &= R - wn - F = 0 \end{aligned} \quad (7)$$

Optimal amounts of skill and team size are assumed to exceed zero, so that the first two expressions are

equalities. The third expression of (7) is for the moment left as an inequality, to suggest that if funding is sufficiently unresponsive to dispersion, then  $D$  equals zero and institutional collaboration does not occur.

However, consider the case where  $D > 0$ , so that variations in all three controls are allowed. The second order conditions that ensure a maximum for this problem are that the determinants of the bordered Hessian of the Lagrangian alternate in sign:

$$\begin{vmatrix} L_{ss} & L_{sn} & L_{s\lambda} \\ L_{sn} & L_{nn} & L_{n\lambda} \\ L_{s\lambda} & L_{n\lambda} & 0 \end{vmatrix} > 0, \quad \begin{vmatrix} L_{ss} & L_{sn} & 0 & L_{s\lambda} \\ L_{sn} & L_{nn} & 0 & L_{n\lambda} \\ 0 & 0 & L_{DD} & 0 \\ L_{s\lambda} & L_{n\lambda} & 0 & 0 \end{vmatrix} < 0 \quad (8)$$

The specific form of (8) reflects the separability of the first order conditions for  $n$ ,  $s$  from the first order condition for  $D$ . For exceptionally clear statements of the general conditions see Chiang (1974, Section 12.3) or Dixit (1990, Ch. 7). Using this information and the method of comparative statics we can show that skill  $s$  and team size  $n$  tend to increase with productivity  $A$ . To see this form the displacement system of (7),

$$\begin{aligned} [H] \begin{bmatrix} ds \\ dn \\ dD \\ d\lambda \end{bmatrix} &= \begin{bmatrix} -\frac{1}{A} Q_s dA \\ -\frac{1}{A} Q_n dA \\ 0 \\ 0 \end{bmatrix}, \quad \text{where} \\ [H] &= \begin{bmatrix} L_{ss} & L_{sn} & 0 & L_{s\lambda} \\ L_{sn} & L_{nn} & 0 & L_{n\lambda} \\ 0 & 0 & L_{DD} & 0 \\ L_{s\lambda} & L_{n\lambda} & 0 & 0 \end{bmatrix}. \end{aligned} \quad (9)$$

Solving (9) for changes in the control variables and pre-multiplying by the transposed shift vector yields

$$\begin{aligned} &\left[ \begin{array}{ccc|c} -\frac{1}{A} Q_s dA & -\frac{1}{A} Q_n dA & 0 & 0 \end{array} \right] \begin{bmatrix} ds \\ dn \\ dD \\ d\lambda \end{bmatrix} \\ &= \left[ \begin{array}{cccc} \frac{1}{A} Q_s dA & \frac{1}{A} Q_n dA & 0 & 0 \end{array} \right] [H]^{-1} \\ &\times \begin{bmatrix} \frac{1}{A} Q_s dA \\ \frac{1}{A} Q_n dA \\ 0 \\ 0 \end{bmatrix} < 0 \end{aligned} \quad (10)$$

We have pre- and post-multiplied the right hand side of the equality (10) by  $-1$ . Since by (8)  $[H]$  and  $[H]^{-1}$  are negative definite, the expression on the right of (10) is strictly negative as shown so that a weighted combination of  $s$  and  $n$  increases as  $A$  increases. Likewise we can show that an increase in sensitivity of funding to distance  $D$ , or a decrease in sensitivity of fixed costs  $F$  both cause an increase in  $D$  and institutional collaboration. Finally, an exogenous increase in funding  $R$ , perhaps due to past awards, will tend to increase team size and skill. These implications fit the regression results in Section 6 rather well.

### 3. Database of scientific papers

The data set consists of 2.4 million scientific papers, published during 1981–1999, that have at least one author from a set of leading U.S. universities. These “top 110” universities account for most of U.S. academic research. All papers belong to a standard set of communications consisting of articles, reviews, notes, and proceedings. The source of the data is the Institute for Scientific Information (ISI) in Philadelphia, specifically ISI’s Current Contents data base.<sup>5</sup>

The papers are assigned to fields according to the classification of the journal in which they appear. This classification system consists of 88 academic fields. In order to link the paper data to the 12 main sciences in the National Science Foundation (NSF) CASPAR database, we assign each of the 88 sub-fields to one of the 12 main fields.<sup>6</sup> The Appendix lists the 110 universities, ranked by their R&D funding in 1998. Table 1 shows the 12 main NSF sciences and their 88 sub-field components.

As noted in the introduction we use the data both at the paper level and at the level of university-fields. At the paper level we compile time trends and cross-sectional patterns by field and year. The data record date of publication, scientific fields of journals in which the papers appear, institutional affiliation of authors, ad-

dress information on city, state, and country; and author names as well as number of authors.<sup>7</sup>

It is important to see that the address information is completely separate from author information, so that a name cannot easily be assigned to a location at this time. The address information is nevertheless useful in its own right. Besides the top 110 universities the addresses identify U.S. and foreign institutions consisting of other universities and colleges; governments and government research institutes; medical centers; corporations; and all other institutions.<sup>8</sup> Among other uses we employ the addresses to construct fractions of scientific papers written in one or more of the top 110 schools.<sup>9</sup>

We also construct numbers and shares of different types of institutions as estimates of the location of team workers. Within the U.S. the institutional types consist of: (a) U.S. Government; (b) Other U.S. Universities; (c) U.S. Corporations; (d) U.S. Medical Centers; and (e) All Other U.S. Institutions. Outside the U.S. the institutional types consist of: (a) Foreign Governments; (b) Foreign Universities; and (c) All Other Foreign Institutions, including by country. As we have seen, this information allows us to assign fractions of papers to different institutional classes as well as to provide indicators of the proportional contribution by each class.

Table 2 reports the distribution of scientific papers by the 12 main science fields. The table includes the years 1981, 1990, and 1999 and all years, showing which fields gain share and which lose share.

<sup>7</sup> There is no limit at this time on the number of authors in the ISI data. The maximum number in our sample is 210, while the mean number in the paper-level data is 2.36. Notice that the number of authors underestimates the number of team members when it excludes contributors, such as research assistants. It is an overestimate when it includes honorific authors. In short, the number of authors measures the size of scientific teams with error. This error is unavoidable since we lack any other measure of team size.

<sup>8</sup> About 5% of the addresses could not be assigned.

<sup>9</sup> The fractions are  $1/2$ ,  $1/2$  in the case of two institutions,  $1/3$ ,  $1/3$ ,  $1/3$  for three institutions, and so on. The cumulative distribution of the number of top 110 institutions per paper is as follows, with number of institutions in parentheses: 79.6% (1 institution), 96.8% (2 institutions or less), 98.3% (3 institutions or less), and 99.5% (4 institutions or less). Of course, these are extremely crude indicators of contributions because they do not include time and effort by team members, nor do they differentiate among types of effort. In short, the institutional address fractions do not measure labor input, even though we use them to attribute output to scientific institutions.

<sup>5</sup> The journal set consists of approximately 5500 journals that were active in 1999, as well as about 1600 inactive journals that were cited by the active journals.

<sup>6</sup> The 12 fields are: agriculture, astronomy, biology, chemistry, computer science, earth sciences, economics and business, engineering, mathematics and statistics, medicine, physics, and psychology.

Table 1

Composition of 12 main science fields, papers and citations of the top 110 U.S. universities

Main science field	Sub-fields within main science fields
Agriculture	General agriculture and agronomy; aquatic sciences; animal sciences; plant sciences; agricultural chemistry; entomology and pest control; food science and nutrition; veterinary medicine and animal health
Astronomy	Astronomy and astrophysics
Biology	General biological sciences; biochemistry and biophysics; cell and developmental biology; ecology and environment; molecular biology and genetics; biotechnology and applied microbiology; microbiology; experimental biology; immunology; neurosciences and behavior; pharmacology and toxicology; physiology; oncogenesis and cancer research
Chemistry	General chemistry; analytical chemistry; inorganic and nuclear chemistry; organic chemistry and polymer science; physical chemistry and chemical physics; spectroscopy, instrumentation, and analytical science
Computer science	Computer science and engineering; information technology and communications systems
Earth sciences	Atmospheric sciences; geology and other earth sciences; geological, petroleum, and mining engineering; oceanography
Economics and business	Economics; accounting; decision and information sciences; finance, insurance, and real estate; management; marketing
Engineering	Aeronautical engineering; biomedical engineering; chemical engineering; civil engineering; electrical and electronics engineering; engineering mathematics; environmental engineering and energy; industrial engineering; materials science; mechanical engineering; metallurgy; nuclear engineering
Mathematics and statistics	Mathematics; biostatistics and statistics
Medicine	General and internal medicine; anesthesia and intensive care; cardiovascular and hematology research; cardiovascular and respiratory systems; clinical immunology and infectious disease; clinical psychology and psychiatry; dentistry and oral surgery; dermatology; endocrinology, metabolism, and nutrition; environmental medicine and public health; gastroenterology and hepatology; health care sciences and services; hematology; medical research, diagnosis, and treatment; medical research, general topics; medical research, organs and systems; neurology; oncology; ophthalmology; orthopedics, rehabilitation, and sports medicine; otolaryngology; pediatrics; radiology, nuclear medicine, and imaging; reproductive medicine; research, laboratory medicine, and medical technology; rheumatology; surgery; urology and nephrology
Physics	General physics; applied physics, condensed matter, and materials science; optics and acoustics
Psychology	Psychology and psychiatry

Source: Institute for Scientific Information.

Table 2

Distribution of papers by field, of the top 110 U.S. universities 1981, 1990, 1999, and all years

Field of science	Number of papers			
	1981	1990	1999	All years
Agriculture	8,697 (9.0%)	10,714 (8.5%)	9,341 (6.1%)	189,004 (7.8%)
Astronomy	1,688 (1.7%)	1,581 (1.3%)	2,913 (1.9%)	35,508 (1.5%)
Biology	24,928 (25.7%)	32,495 (25.7%)	41,742 (27.4%)	634,737 (26.3%)
Chemistry	7,951 (8.2%)	10,432 (8.3%)	12,205 (8.0%)	194,798 (8.1%)
Computer science	872 (0.9%)	1,611 (1.3%)	2,045 (1.3%)	28,037 (1.2%)
Earth sciences	2,802 (2.9%)	3,818 (3.0%)	4,956 (3.3%)	72,920 (3.0%)
Economics	1,758 (1.8%)	2,600 (2.1%)	2,363 (1.6%)	43,540 (1.8%)
Engineering	5,334 (5.5%)	9,204 (7.3%)	11,689 (7.7%)	170,147 (7.1%)
Mathematics	3,086 (3.2%)	3,127 (2.5%)	3,623 (2.4%)	60,710 (2.5%)
Medicine	26,791 (27.6%)	33,154 (26.3%)	41,199 (27.0%)	648,704 (26.9%)
Physics	7,289 (7.5%)	11,521 (9.1%)	13,840 (9.1%)	215,942 (9.0%)
Psychology	5,770 (6.0%)	6,017 (4.8%)	6,538 (4.3%)	115,482 (4.8%)
Total	96,966	126,274	152,454	2,409,529

Source: Institute for Scientific Information and authors' calculations.

Among the life sciences, which dominate the data, biology gains while agriculture loses. Among the physical sciences astronomy and physics gain share. Not surprisingly, among the mathematical sciences, computer science gains and mathematics and statistics lose. Engineering increases its share, and finally, the social and behavioral sciences perceptibly lose share.

We use paper-level statistics for the descriptive work in Sections 4 and 5, because this retains the means and standard deviations of the original data. In Section 6 we carry out regression analysis of the determinants of team size, domestic and foreign collaboration, and research “output”. For this purpose we construct a panel at the level of universities, fields and years. The reason is that the panel allows us to combine the ISI papers and citations data with information on lagged university-field level R&D and characteristics of doctoral programs. The NSF CASPAR database of universities, a compendium of NSF surveys, is the source of the data on university R&D. The National Research Council (NRC) 1993 Survey of Doctoral Programs (NRC, 1995) includes characteristics of graduate programs, especially counts of Nobel prizes and other prestigious awards, as well as rankings of quality of Ph.D. programs in 1993. Finally, microdata from the NSF Survey of Earned Doctorates (SED) provide us with estimates of the migration of Ph.D. students to the academic and industrial sectors of the U.S. economy, as well as to other countries.<sup>10</sup>

We impose one other constraint on the panel, which does not apply to the paper-level data. We consider only leading departments out of the top 110. All other schools form a remainder cell within each field. More precisely, we include the top 25 universities in astronomy plus a remainder, the top 50 universities in agriculture, chemistry, computer science, economics and business, earth sciences, mathematics and statistics, physics, and psychology, plus one remainder each. And finally we include the top 75 universities in biology, medicine, and engineering plus remainders for each

of these three fields. Summing across fields, and accounting for the fact that only 48 schools of agriculture formally exist, the panel data consist of 660 university-fields in any given year. Our purpose in breaking out fewer individual schools in smaller fields, and more in larger fields, is to avoid large numbers of empty cells for universities in which fields (and doctoral programs) are small or non-existent.<sup>11</sup> The result is a panel of 12,540 observations, before bad or missing data are excluded, that approximates teams and institutional collaborations of 660 university-fields in 12 main sciences over the 19-year period. This panel includes an array of variables that are likely to drive teams, collaborations, and research output. We describe the major variables in the panel data and sources of these variables in Section 6, where we consider determinants of team size and collaboration.

#### 4. Time trends in scientific collaborations

This section documents upward trends in scientific teams and institutional collaborations. Figs. 1–10 display time series of scientific research, team size, and institutional collaborations by type of partner and by domestic–international classification. All graphs refer to the period 1981–1999. Fig. 1 shows trends in the output of U.S. scientific papers. The upper line is total papers having at least one author from a top 110 U.S. university. The middle line consists of U.S. equivalent papers: the fraction of U.S. institutions in all institutions on each paper summed over all papers. The lower line consists of top 110 paper equivalents. This is the fraction of top 110 U.S. institutions in all institutions summed over papers. U.S. and top 110 equivalent papers grow more slowly than total papers. Top 110 equivalents decline by 2.5% as of 1999 compared with the 1995 peak. The growing spread between total papers and U.S. equivalents shows the growing contribution of foreign institutions. This could be seen as beneficial: foreign institutions produce more of the research and transfer more of

<sup>10</sup> The migration data used in this paper represents flows of new Ph.D.’s with definite plans at the time of graduation, so that the destinations that we use are projected as within a few months of graduation. The data are undercounts, since one-third of new Ph.D.’s do not have definite plans. Moreover, the data represent even a greater undercount to industry, since many new Ph.D.’s go to industry only after completing their postdoctoral training.

<sup>11</sup> The size of the remainder of the top 110 equals an average university-field in the individual top 25, 50, or 75 schools in a field, showing that the residual aggregation procedure misses rather little. This finding reflects the positive skew of academic programs. For more on this see Adams and Griliches (1998).

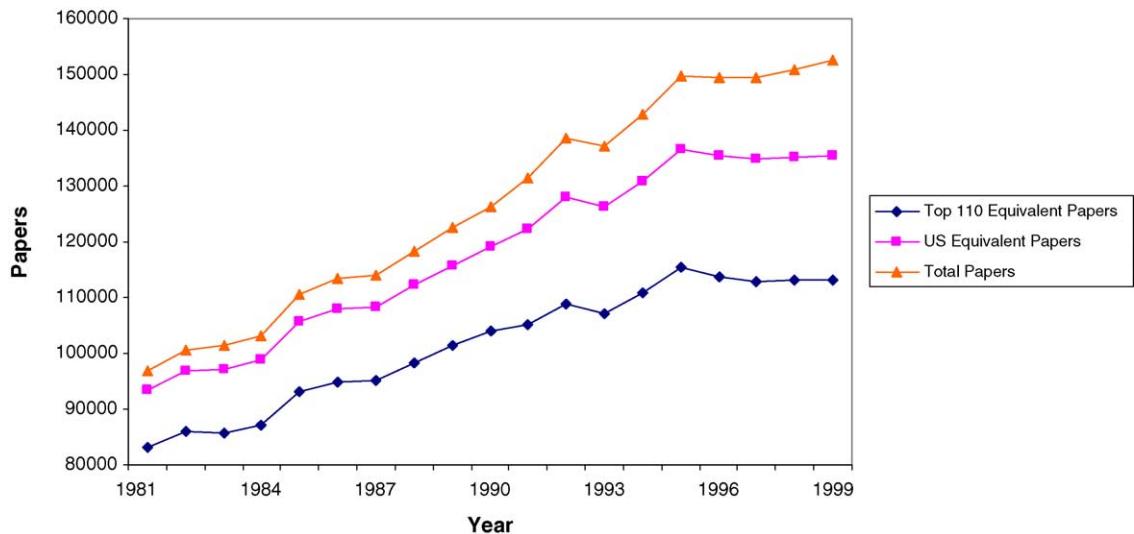


Fig. 1. Total papers, top 110 equivalents, and U.S. equivalents of the top 110 U.S. universities, 1981–1999.

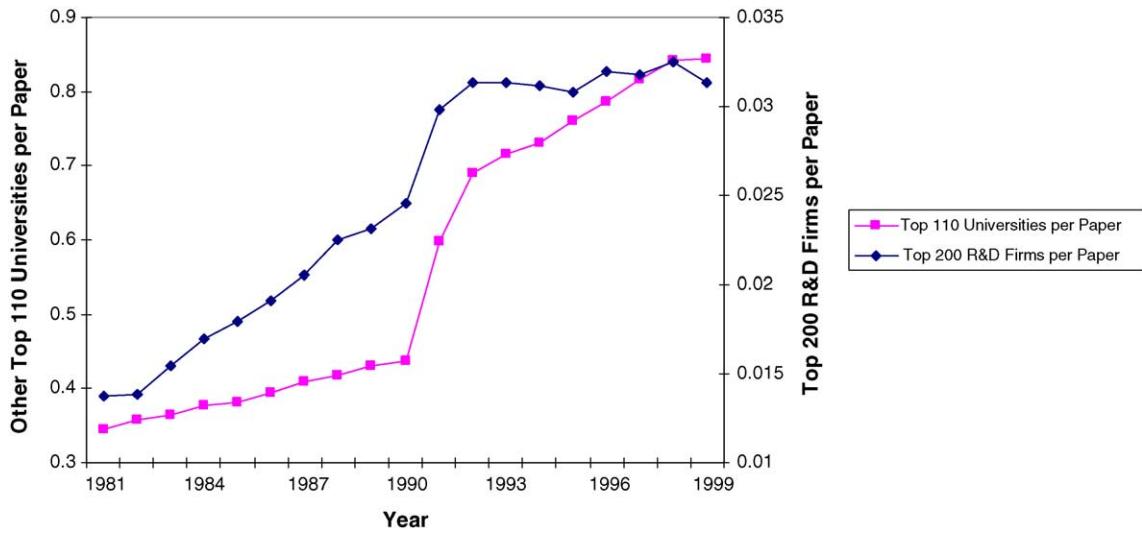


Fig. 2. Top 110 U.S. universities and top 200 U.S. R&D firms, per paper, 1981–1999.

their knowledge to the U.S. Or it could be grounds for pessimism: remaining in the same place after 1995 required an increase in the foreign contribution.

Fig. 2 graphs domestic collaboration with *other* top 110 schools (left scale) and with the top 200 U.S. R&D firms (right scale).<sup>12</sup> While the number of schools and

firms per paper grow at about the same rate, growth in collaboration with firms is less rapid in the 1980s and more rapid in the 1990s. This acceleration in university–firm joint research poses an issue of attribution that will be a recurring theme of this paper. It could represent increasing placement of graduate stu-

<sup>12</sup> We define the number of *other* top 110 schools as the number of top 110 schools minus one, which represents the “home” institution.

In this way we account for conditioning of the data on membership in the top 110 universities.

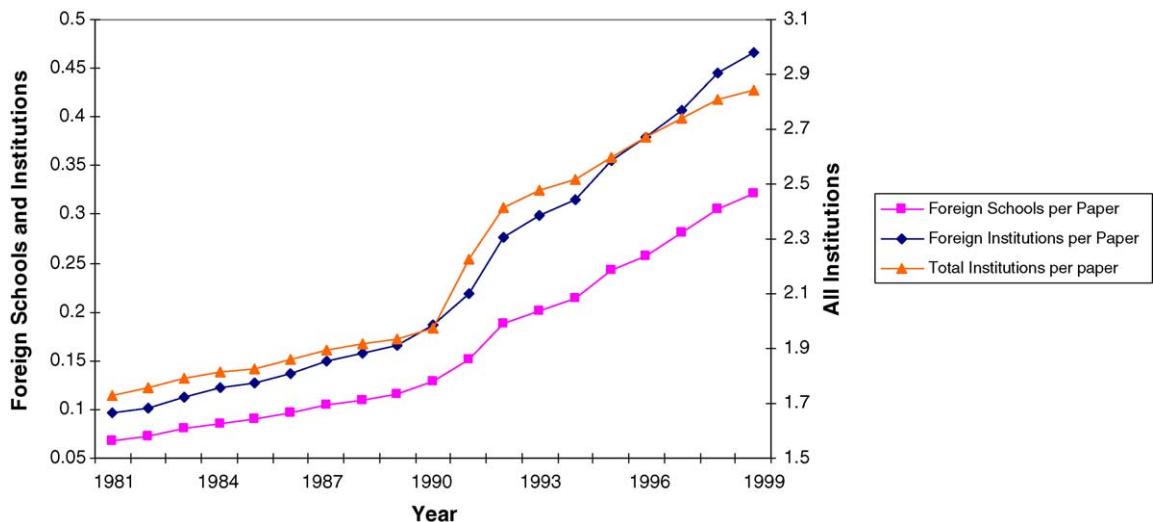


Fig. 3. Foreign schools, foreign institutions, and all institutions, per paper, 1981–1999.

dents with firms rather than schools, it could stand for the success of federal programs designed to promote joint research, it could be a by-product of decelerating growth in government support for academic research, or it could signify a slowdown in industrial support for basic research leading to increasing reliance on university collaborators. This last point is consistent with the decline in scientific papers published in industry since 1991. On this point see National Science Foundation, Science and Engineering Indicators (2004, Volume II, Table 5–36).

Comparative trends in foreign and total collaborations are the subject of Fig. 3. Trends in foreign universities and foreign institutions per paper are on the left scale. The right scale indicates the total of all institutions. Collaborations with foreign universities and institutions grow more rapidly than institutional collaboration as a whole, consistent with Fig. 1.<sup>13</sup>

Fig. 4 finds that growth rates in team size (authors per paper) and institutional collaborations are about the same. Given that foreign collaborations are growing more rapidly than all collaborations, domestic collaborations have to grow *more* slowly than team size. Thus,

scientific teams are becoming more internationalized over time.

The next two figures examine trends in foreign collaboration by region of the world.<sup>14</sup> Fig. 5 reports trends in counts of foreign addresses per paper. Europe is the dominant region for collaboration, reflecting the size of its scientific sector. The countries of East and South Asia come in second by the end of the period, followed by the rest of the Americas. The rest of the world, composed of Africa, the Middle East, and Oceania, runs a distant fourth. Fig. 6 brings out more clearly the differences in growth by region. The figure normalizes each of the series in Fig. 5 by its 1981 value. Growth is more rapid in Asia and Europe and slower in the Americas and the rest of the world. This difference clearly reflects the faster growth of scientific resources in Asia and Europe than in the rest of the world.

Figs. 7 and 8 consider interactions between team size and the foreign share of institutional affiliations.

<sup>13</sup> Correcting the total for conditioning on a top 110 university, the total institutions series increases from 0.8 to 2.0, or 2.5 times. Foreign schools per paper increase from 0.07 to 0.32, or more than 4.5 times, while all foreign affiliations increase from 0.1 to 0.46, which is again more than 4.5 times.

<sup>14</sup> In this figure Europe consists of Western and Eastern Europe as well as the European Soviet Socialist Republics of the former Soviet Union. Asian countries include Japan, India, China, and other countries of East and South Asia, such as Malaysia, Indonesia, South Korea, Singapore, and Taiwan. Other countries in the Americas include Canada, Central America, the Caribbean, and South America. Africa, the Middle East, and Oceania includes Israel (Middle East) and Australia and New Zealand (Oceania), and thus contains the developed countries in each region.

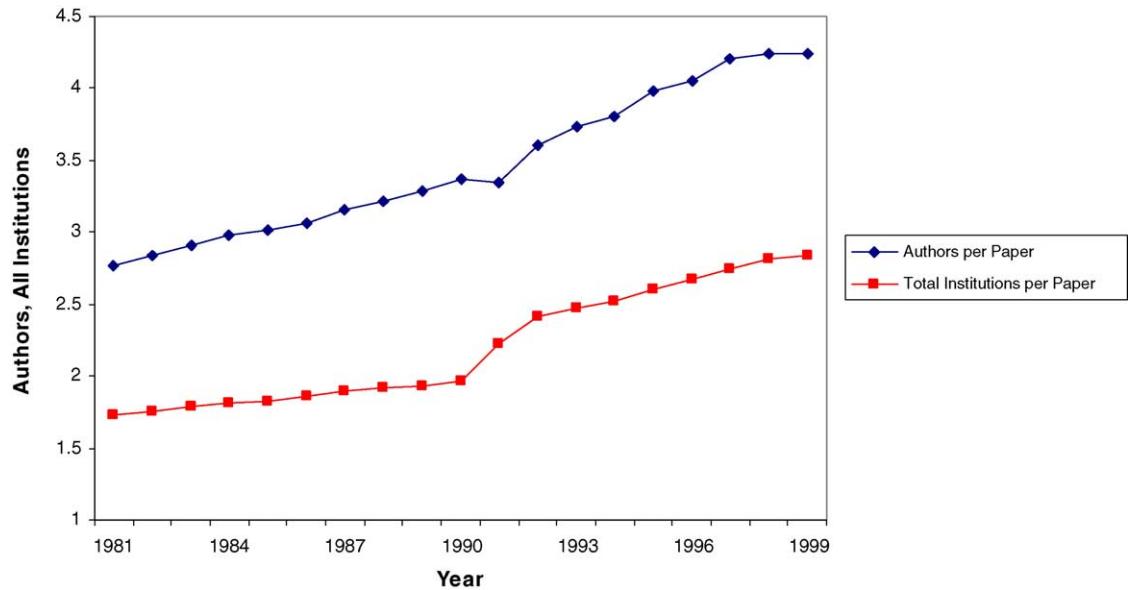


Fig. 4. Mean authors and institutions, per paper, 1981–1999.

Fig. 7 displays raw data on the foreign share classified by intervals of team size. The foreign share is greater in larger teams. International cost-sharing of large-scale projects could lie behind this relationship, for example in instrument time-sharing, in the Human Genome project, in large-scale space missions, and so on. Fig. 8 brings out comparative growth

more clearly again by normalizing each of the series in Fig. 7 by its 1981 value. The graph shows that smaller teams are becoming more internationalized at a faster rate. Since larger teams are more internationalized in 1981, this pattern of comparative growth implies convergence in the foreign share by team size.

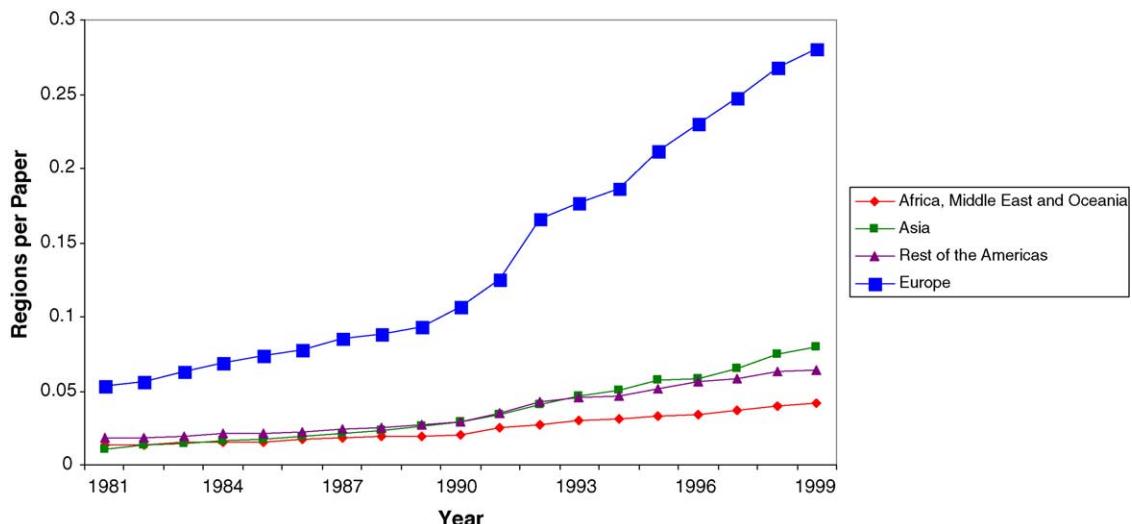


Fig. 5. Foreign address counts per paper, by region of the world, 1981–1999.

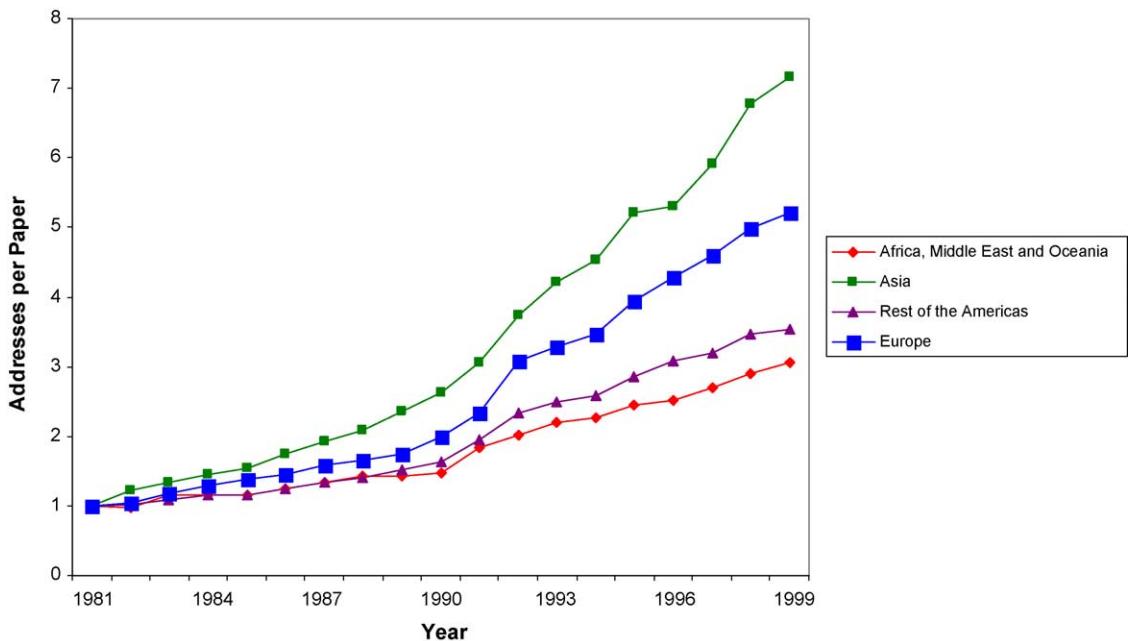


Fig. 6. Normalized foreign address counts per paper, by region of the world, 1981–1999 (1981 = 1.0).

Figs. 9 and 10 indicate comparative trends in the foreign share by groups of fields. Fig. 9 reports time series of the foreign share in which fields are grouped by their initial foreign share. The three most interna-

tional fields are astronomy, mathematics and statistics, and physics. The three least international fields are agriculture, biology, and medicine. Foreign shares of remaining fields (chemistry, computer science, earth

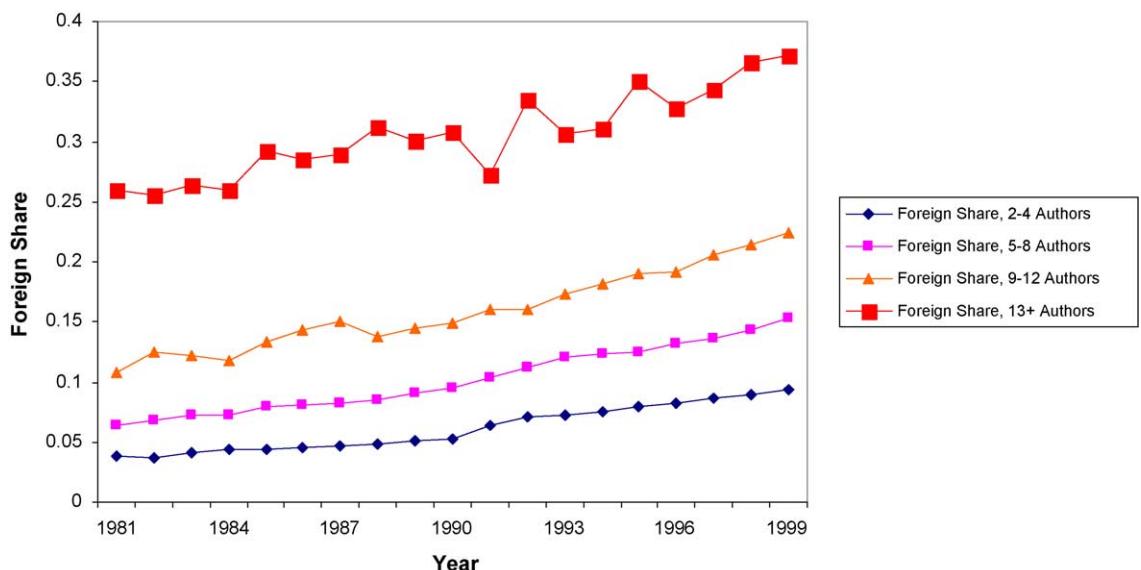


Fig. 7. Foreign address share per paper, by team size, 1981–1999.

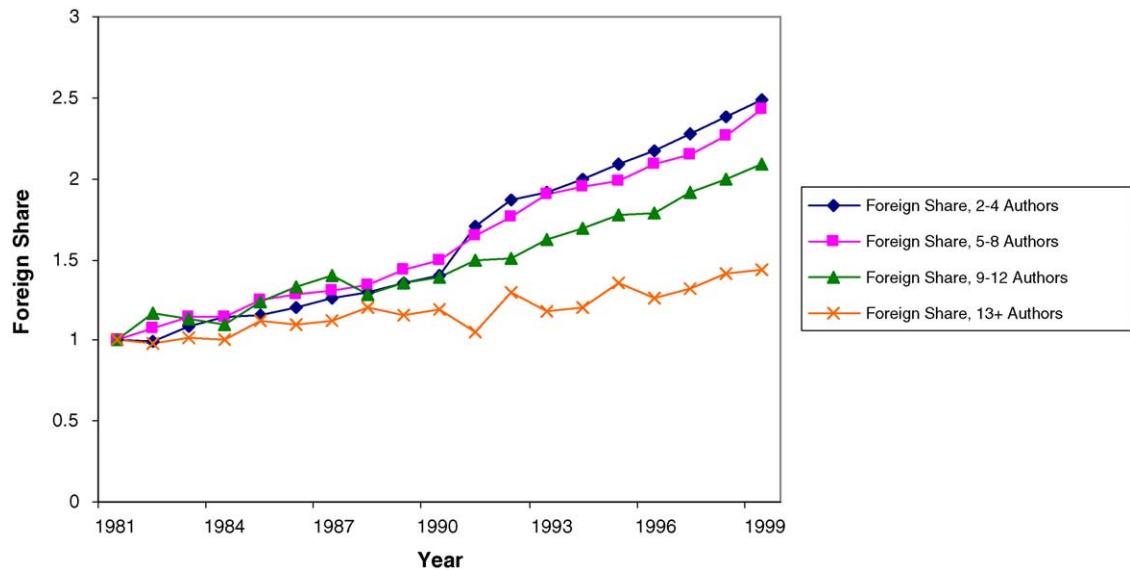


Fig. 8. Normalized foreign address share per paper, by team size, 1981–1999 (1981 = 1.0).

sciences, economics, engineering, and psychology) lie in the middle. Fig. 10 brings out comparative growth more clearly by normalizing each of the series in Fig. 9 by its 1981 value. The figure shows that the life sciences are becoming internationalized at the most rapid rate. Since these fields are the least internationalized at the start, this result suggests mild convergence in the foreign shares.

## 5. Findings on collaborative behavior by field of science

We turn now to descriptive findings by field of science. The purpose of the section is to reveal differences by field of science in the growth of team size and institutional collaboration. We display the results in a series of tables given the number of fields that each entails.

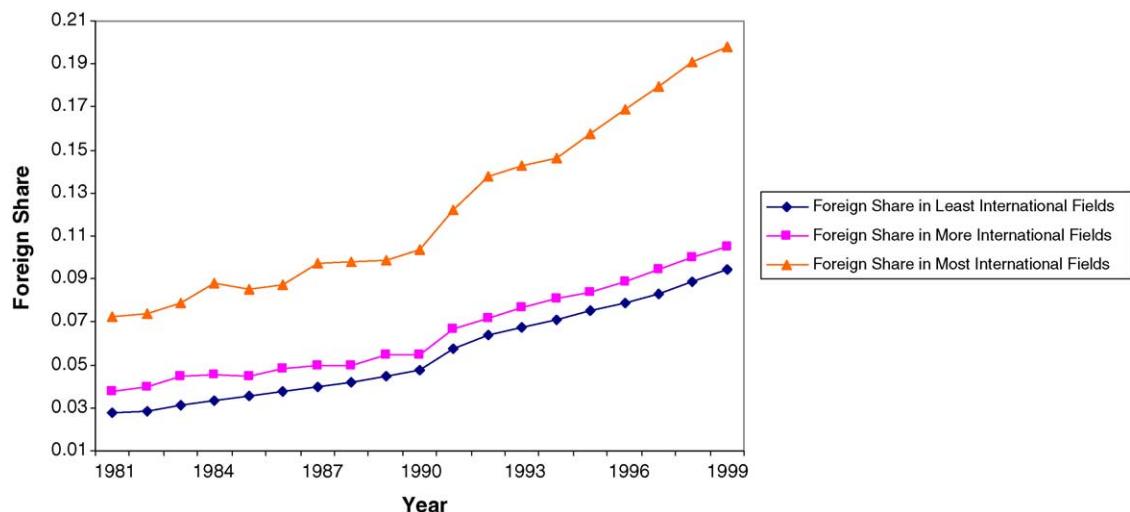


Fig. 9. Foreign address share per paper, by degree of internationalization, 1981–1999.

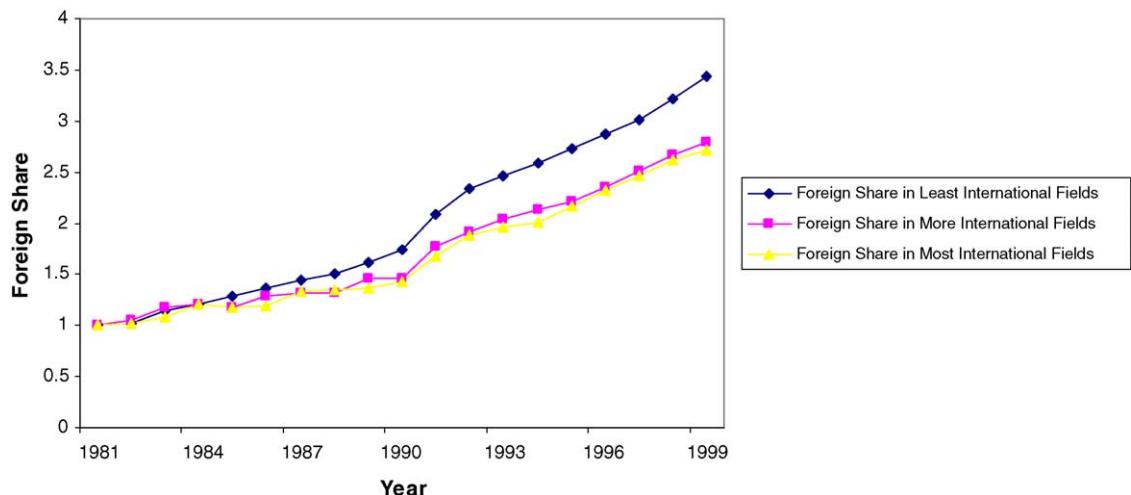


Fig. 10. Normalized foreign address share per paper, by degree of internationalization, 1981–1999 (1981 = 1.0).

Table 3

Team size and its rate of growth, by field, of the top 110 U.S. universities 1981, 1990, and 1999

Field of science	Mean authors per paper				
	1981	Percentage of annual growth rate (1981–1990)	1990	Percentage of annual growth rate (1990–1999)	1999
Agriculture	2.407	1.55	2.768	2.00	3.314
Astronomy	2.654	2.36	3.283	4.57	4.952
Biology	2.810	2.11	3.398	2.55	4.274
Chemistry	2.816	1.04	3.093	1.68	3.597
Computer science	1.861	1.47	2.124	2.42	2.640
Earth sciences	2.288	2.30	2.814	2.78	3.615
Economics	1.572	1.04	1.727	1.29	1.939
Engineering	2.289	1.14	2.537	1.80	2.984
Mathematics	1.531	0.97	1.671	1.47	1.907
Medicine	3.259	1.79	3.828	1.99	4.580
Physics	3.091	5.46	5.053	4.03	7.264
Psychology	2.209	1.66	2.565	2.24	3.138
Total	2.766	2.19	3.368	2.57	4.244

Source: Institute for Scientific Information and authors' calculations.

**Table 3** reports team size in 1981, 1990, and 1999, as well as growth in team size across the decades of the 1980s and 1990s. In 10 of 12 fields, growth occurs more rapidly in the 1990s. This pattern dominates the grand average in the bottom row. Acceleration in the growth of team size is the rule.

**Table 4** explores geographic dispersion of team members within the top 110 U.S. universities. We assume that the highest ranked university-field on each paper is the “head” institution and proceed to calcu-

late mileages to other top 110 institutions on that paper based on latitude and longitude coordinates.<sup>15</sup> If only one top 110 institution participates in a paper then the mileage is zero. Therefore, changes in the mileage statistics depend on changes in the tendency to work

<sup>15</sup> The calculation assumes that the earth is a sphere and calculates distance using the geodesic or shortest distance between two points on that sphere. For more, see Adams and Jaffe (1996) and Adams (2002).

Table 4  
Distance between team workers in the top 110 U.S. universities, by field, 1981, 1990, and 1999

Field of science	Mean distance in miles				
	1981	Percentage of annual growth rate (1981–1990)	1990	Percentage of annual growth rate (1990–1999)	1999
Agriculture	50.9	2.63	64.5	7.15	122.8
Astronomy	264.6	2.76	339.2	1.82	399.5
Biology	76.8	4.45	114.6	4.46	171.2
Chemistry	50.8	0.58	53.5	4.65	81.3
Computer science	138.2	0.45	143.9	0.12	145.4
Earth sciences	145.7	2.29	179.1	5.95	306.0
Economics	124.8	4.28	183.5	0.81	197.3
Engineering	61.2	1.76	71.7	2.94	93.4
Mathematics	128.3	1.68	149.2	1.50	170.8
Medicine	62.2	5.34	100.6	4.84	155.5
Physics	92.6	2.51	116.1	3.40	157.7
Psychology	88.8	4.18	129.4	4.55	194.8
Total	77.7	3.53	106.8	4.45	159.4

Source: Institute for Scientific Information and authors' calculations.

with other top 110 schools. The average mileage on a paper directly captures geographic dispersion within the system of top 110 schools.

The table reports mean distances in 1981, 1990, and 1999 and compares growth rates across decades. Growth in geographic dispersion is quite clear but *acceleration* in growth is less obvious. Six of 12 fields show evidence of acceleration (agriculture, chemistry, earth sciences, engineering, physics, and psychology), the growth of one (biology) is constant, and the remaining five (astronomy, computer science, economics and business, mathematics and statistics, and medicine) reveal mild deceleration. And yet the overall pattern is one of growth acceleration. Table 4 reveals the expanding geographic scope of collaboration within the top 110 universities: mean distances double from 78 to 159 miles over the period. And despite some mixed results, overall growth accelerates across the two decades. On average the rate of growth in mileage increases from 3.5% in the 1980s to 4.5% in the 1990s.

Table 5 considers patterns of domestic institutional collaboration by field. The table is specifically concerned with collaborations with other top 110 schools and with top 200 R&D firms. The table reports levels of collaboration of both kinds in 1981 and 1999 and reports growth over the full period. The table reveals changes in collaboration within academia, as well as between academia and indus-

try. Growth in collaboration within the university sector is most rapid in agriculture, biology, chemistry, and psychology. The collaboration between universities and firms follows a slightly different pattern. Collaboration in the life sciences (agriculture, biology, and medicine) and psychology expands the most rapidly of all. Industry–university collaboration in more established industrial–scientific fields like chemistry, computer science, engineering, and physics grows less rapidly. Of course, the level of collaboration is far higher in these fields than in the life and behavioral sciences. Astronomy and economics are the only fields where university–firm collaboration declines. All of these patterns are likely the result of changes in the population of industrial scientists in different disciplines.

Table 5 allows us to compare growth rates in domestic collaboration across sectors and fields. Define a relative increase in the “outward” dimension of a field as an excess in growth of collaboration with industry over growth in collaboration with universities. Likewise an increase in a field’s “inward” dimension takes place when the growth rate with industry falls short of the growth rate with universities. Based on this criterion, agriculture, biology, computer science, medicine, and psychology are becoming more outward disciplines. By the same token astronomy, chemistry, earth science, economics and business, engineering, mathematics and statistics, and physics are becoming more inward.

Table 5

Indicators of U.S. institutional collaboration, by field, papers of the top 110 universities, 1981 and 1999

Field of science	Other top 110 universities per paper <sup>a</sup>			Top 200 firms per paper		
	1981	Percentage of annual growth rate (1981–1999)	1999	1981	Percentage of annual growth rate (1981–1999)	1999
Agriculture	0.224	5.44	0.630	0.003	8.11	0.014
Astronomy	0.387	2.72	0.649	0.016	-0.34	0.015
Biology	0.403	5.45	1.135	0.007	6.03	0.022
Chemistry	0.195	7.50	0.811	0.023	2.78	0.039
Computer science	0.197	2.55	0.320	0.078	4.25	0.175
Earth sciences	0.250	3.71	0.506	0.021	0.24	0.022
Economics	0.177	3.47	0.342	0.009	-3.09	0.005
Engineering	0.175	4.73	0.430	0.046	2.91	0.080
Mathematics	0.166	3.11	0.300	0.012	2.69	0.020
Medicine	0.500	3.67	1.005	0.005	8.26	0.024
Physics	0.280	3.98	0.596	0.048	0.11	0.049
Psychology	0.214	5.34	0.590	0.002	6.59	0.007
Total	0.345	4.70	0.843	0.014	4.18	0.031

Source: Institute for Scientific Information and authors' calculations.

<sup>a</sup> This is the number of top 110 universities per paper minus one. This measure maintains symmetry with top 200 firms per paper, which is the number of "other" institutions as well, in this case, top R&D firms.

**Table 6** concludes the descriptive findings by examining trends in foreign collaboration across the sciences. The table computes the foreign share in all institutions for papers published in 1981, 1990, and 1999. It also examines growth in the share across the 1980s and 1990s. For comparison we include mean foreign

and total institutional affiliations in brackets. Almost without exception growth in the foreign share is more rapid than domestic growth as shown in **Table 5**. Moreover, growth accelerates in every field. Average growth is 5.11% in the 1980s but it is 7.41% in the 1990s, so that the acceleration factor here is 0.45 (7.41/5.11 – 1).

Table 6

Measures of foreign affiliation, by field, papers of the top 110 universities, 1981, 1990, and 1999

Field of science	Mean share of foreign affiliations [mean foreign affiliations, mean total affiliations]				
	Foreign institutional affiliations in 1981	Percentage of annual growth in foreign share (1981–1990)	Foreign institutional affiliations in 1990	Percentage of annual growth in foreign share (1990–1999)	Foreign institutional affiliations in 1999
Agriculture	0.028 [0.067, 1.482]	4.77	0.043 [0.113, 1.638]	9.81	0.104 [0.340, 2.397]
Astronomy	0.086 [0.247, 1.963]	5.49	0.141 [0.509, 2.479]	6.14	0.245 [1.144, 3.436]
Biology	0.034 [0.092, 1.732]	6.31	0.060 [0.189, 2.046]	6.73	0.110 [0.455, 3.085]
Chemistry	0.046 [0.111, 1.456]	3.14	0.061 [0.161, 1.573]	6.35	0.108 [0.363, 2.441]
Computer science	0.043 [0.102, 1.479]	3.70	0.060 [0.151, 1.680]	7.03	0.113 [0.317, 2.057]
Earth sciences	0.052 [0.141, 1.627]	5.33	0.084 [0.267, 1.950]	7.23	0.161 [0.559, 2.626]
Economics	0.041 [0.092, 1.432]	3.06	0.054 [0.130, 1.634]	6.16	0.094 [0.255, 1.926]
Engineering	0.040 [0.095, 1.506]	3.13	0.053 [0.131, 1.574]	7.60	0.105 [0.302, 2.111]
Mathematics	0.071 [0.161, 1.438]	4.45	0.106 [0.248, 1.588]	5.12	0.168 [0.422, 1.901]
Medicine	0.021 [0.067, 2.035]	6.29	0.037 [0.132, 2.265]	8.14	0.077 [0.345, 3.140]
Physics	0.070 [0.205, 1.721]	3.74	0.098 [0.456, 2.169]	7.70	0.196 [1.194, 3.187]
Psychology	0.016 [0.042, 1.559]	6.22	0.028 [0.075, 1.775]	8.28	0.059 [0.188, 2.329]
Total	0.036 [0.097, 1.731]	5.11	0.057 [0.186, 1.971]	7.41	0.111 [0.466, 2.840]

Source: Institute for Scientific Information and authors' calculations.

Table 7

Means and standard deviations of principal regression variables, top 110 U.S. university data

Variable	Mean (S.D.)
Indicators of teamwork and research “output”	
Number of authors in a university-field per paper	2.65 (0.96)
Number of authors per paper	4.26 (6.43)
Other top 110 U.S. universities per paper	0.41 (0.57)
Foreign share per paper	0.07 (0.05)
U.S. corporate share per paper	0.02 (0.02)
Number of papers by a university-field	149.28 (158.28)
Number of citations received by a university-field, this year and the next four years	708.75 (1575.74)
Characteristics of university-fields	
Stock of federally funded R&D in a university and field (in thousands of 1992 dollars)	58,277.65 (81,955.07)
Stock of federally funded R&D in a university and field per paper (in thousands of 1992 dollars)	485.71 (746.57)
Private university	0.35 (0.48)
Number of awards <sup>a</sup>	0.23 (0.62)
Local university R&D ratio <sup>b</sup>	0.08 (0.16)
Equipment expenditure/R&D, previous 3 years	0.07 (0.05)
Share of graduate students placed in top 40% schools, lagged 2 years	0.16 (0.18)
Share of graduate students placed in U.S. firms, lagged 2 years	0.18 (0.20)
Share of graduate students placed in 12 top research countries, lagged 2 years <sup>c</sup>	0.05 (0.10)

Notes: Sources for the data are the Institute for Scientific Information, National Science Foundation, National Research Council, and authors' calculations.

<sup>a</sup> Awards include the Fields Medal, MacArthur Awards, the National Medal of Science, the National Medal of Technology, Fellow of the National Academy of Science, and the Nobel Prize. See the text for a further discussion.

<sup>b</sup> The local university R&D ratio equals R&D in the same field, but in other universities, within 25 miles, divided by R&D in the same field, but in other universities, within 200 miles.

<sup>c</sup> The top 12 research countries are Australia, Canada, France, Germany, Japan, the United Kingdom, Italy, the Netherlands, Israel, New Zealand, Sweden, and Switzerland.

## 6. Regression findings

We now take up the third part of the empirical work. Following up on themes introduced by the analytical framework of Section 2, this concerns the problem of explaining team size, institutional collaboration by sector and country, and research “output”. For this purpose we have constructed a panel of universities, fields and years, in order to match team size, various measures of institutional collaboration, and research “output” with data from the National Science Foundation and the National Research Council that are reported at this level, as explained in Section 3. Table 7 contains descriptive statistics from the panel. The statistics show that the average university field-year observation has a team size of 4.26 authors per paper, of which 2.65, or almost two-thirds, are estimated to reside within a university-field.<sup>16</sup> The average university-field article

involves 0.41 other top 110 universities. On average foreign institutions contribute a 5.2% share of all institutional affiliations, while U.S. firms contribute 2.0%. The average number of papers in a university-field is 149, for which 709 citations are received from other top 110 universities during its first 5 years including the year in which the paper is published.

The average stock of deflated R&D is about 58 million dollars of 1992. This is the 8-year stock depreciated at a 15% rate and lagged 1 year. Thus, the R&D stock in 1981 is the sum of deflated and depreciated R&D over the years 1973–1980 and likewise for all other years.<sup>17</sup>

---

the university-field observations weight large teams more heavily than do individual papers.

<sup>17</sup> The choice of a 8-year (and thus incomplete) R&D stock is dictated by the 1973 start date for flows of R&D in the CASPAR data base, as this interacts with the 1981 start date of the ISI data. The 8-year stock is thus the longest history that we have. We should say that the CASPAR R&D data, while they represent a major achievement in data collection on universities, also contain substantial respondent errors. We have tried to flag these errors and to remove all suspicious observations on R&D from our analysis.

---

<sup>16</sup> The average number of authors is greater in the panel data consisting of university-fields than in the original paper level data, because

Table 8

Determinants of team size, dependent variable: log(authors/paper) (*t*-statistics in parentheses)

Variable or statistic	Eq. (8.1)	Eq. (8.2)	Eq. (8.3)	Eq. (8.4)
Time period	1983–1999	1983–1999	1983–1999	1983–1999
Restrictions on team size	None	Less than 10 workers	None	Less than 10 workers
Fields included	All 12 main fields	All 12 main fields	10 fields <sup>a</sup>	10 fields <sup>a</sup>
Year dummies included	Yes, significant	Yes, significant	Yes, significant	Yes, significant
Field dummies included	Yes, significant	Yes, significant	Yes, significant	Yes, significant
Log (stock of federally funded R&D, divided by papers lagged 2 years)	0.012 (3.0) <sup>**</sup>	0.014 (6.4) <sup>**</sup>	0.007 (1.5)	0.007 (2.6) <sup>**</sup>
Private university (1, yes; 0, no)	0.044 (7.5) <sup>**</sup>	0.035 (10.6) <sup>**</sup>	0.049 (6.3) <sup>**</sup>	0.031 (7.3) <sup>**</sup>
Number of awards <sup>a,b</sup>			0.021 (3.1) <sup>**</sup>	0.019 (5.4) <sup>**</sup>
Local university R&D ratio <sup>c</sup>			0.020 (1.0)	0.026 (2.3)*
Equipment expenditure/R&D, previous 3 years			-0.166 (-2.5) <sup>**</sup>	-0.010 (-0.3)
Top 20% in field (1, yes; 0, no)			-0.029 (-3.0) <sup>**</sup>	0.001 (0.3)
Root M.S.E.	0.267	0.147	0.295	0.155
Adjusted <i>R</i> <sup>2</sup>	0.76	0.81	0.76	0.78
Number of observations	9,638	9,228	7,371	6,979

*Notes:* Estimation method is OLS. Sources for the data are the Institute for Scientific Information, the National Science Foundation, the National Research Council, and authors' calculations. The dependent variable is the logarithm of the mean number of authors per paper in a university-field, and year.

<sup>a</sup> Agriculture and medicine lack data on prizes and awards.

<sup>b</sup> Awards include the Fields Medal, MacArthur Awards, the National Medal of Science, the National Medal of Technology, Fellow of the National Academy of Science, and the Nobel Prize. See the text for a further discussion.

<sup>c</sup> The local university R&D ratio equals R&D in the same field but in other universities within 25 miles, divided by R&D in the same field but in other universities within 200 miles.

\* Parameter is significantly different from zero at the 5% level for a one-tailed test.

\*\* Parameter is significantly different from zero at the 1% level for a one-tailed test.

The average stock of lagged R&D per paper, a measure of resources per unit of "output", is a half million dollars of 1992. Private universities account for 35% of the sample. On average there are 0.23 prestigious awards per university-field, as taken from [National Research Council \(1995\)](#).

The local university R&D ratio captures the geographic concentration of research in the vicinity of a university-field, which could measure the ease of institutional collaboration. The local R&D ratio is defined as the ratio of other universities' R&D within 25 miles to the same R&D within 200 miles. In this way it tries to express the convenience of nearby collaborators. The ratio of equipment expenditure to R&D over the previous 3 years may suggest capital-labor substitution, as well as replacement of institutional collaborations by internal funding in a university.

Sixteen percent of former graduate students go to a school that ranks in the top 40% of the top 110 universities. Eighteen percent go to U.S. firms, and

5% go to 12 countries that are highly active in scientific research. These variables measure pools of collaborators that are likely to have funded research, and as such could drive shares of other top 110 schools, U.S. firms, and foreign institutions in the research of a given university-field. In the empirical work we lag the graduate student shares by 2 years in order to take publication lags into account, and this restricts the regression samples to the years 1983–1999.

[Table 8](#) reports regressions that explain a measure of team size, the logarithm of the total authors per paper. Since authors per paper are university-field means, the data are continuous and the estimation method is OLS. All the regressions remove bad or missing data on the R&D stocks. All include dummy variables for year and field, in which 1981 and chemistry are the omitted categories. The dummies absorb trend and field effects, which are highly significant and resemble those depicted in the preceding figures. The general form of

the team size regressions is

$$\begin{aligned} \ln \text{Size}_t = & \alpha_0 + \alpha'_t D_t + \alpha'_F D_F + \beta_R \ln(\text{R&D Stock}_t) \\ & + \beta_C \text{Private university} + \beta_A \text{Awards} \\ & + \beta_L \text{Local R&D}_t + \beta_E \text{Equipment ratio}_{t-1} \\ & + u_t \end{aligned} \quad (11)$$

in which the subscript for the individual university-field is suppressed. Here  $D_t$  is a vector of time dummies and  $D_F$  is a vector of field dummies, while Private equals 1 if a university is privately controlled, and zero otherwise, and the other variables are explained in conjunction with Table 7.

Eq. (8.1) is a baseline regression that includes the logarithm of the stock of federally funded R&D in thousands of 1992 dollars and the private university indicator. We find that the stock of federally funded R&D per paper to an extent increases team size.<sup>18</sup> This suggests that larger projects entail greater specialization. Private universities also form significantly larger teams, for which there are several possible explanations. Private universities may obtain more R&D funding from private foundations and wealthy donors, which we are not able to measure. Another possible reason for the finding is that better pay, better start-up packages, and better working conditions in top private institutions attract more talented faculty (Ehrenberg, 2003; Ehrenberg et al., 2005). This talent advantage, which is related to salary and governance advantages of top private institutions, could help to pull together a larger pool of coworkers (Zuckerman and Merton, 1973).

Eq. (8.2) repeats (8.1) but restricts team size to less than 10 workers. The idea behind this restriction is that university-field R&D is increasingly mismatched with team size as size increases, because an escalating share of funding is external and is not captured by university-field R&D. Thus, the error in the R&D stock rises with team size. Consistent with this idea, the coefficient on R&D stock increases slightly and is more significant in (8.2) than (8.1).

Eqs. (8.3) and (8.4) add a battery of variables to (8.1) and (8.2). As a whole these variables reduce the regression coefficient of the stock of R&D per paper. The bat-

tery includes the number of prestigious awards. Since awards data are missing for agriculture and medicine, (8.3) and (8.4) exclude these two fields. Awards increase team size, consistent with the notion that funding and talent attract coworkers. The local R&D ratio, which tries to capture local concentration of potential team members, has a small positive effect on team size. This is not always significant, perhaps because team workers in the same school substitute for team workers elsewhere. The equipment intensity of R&D spending in the most recent 3 years could signify capital-labor substitution. Consistent with this its coefficient is negative, but again this is not always significant. Finally, the dummy indicator of whether a university-field is in the top 20% in field decreases team size, possibly indicating the availability of graduate student assistance within an institution.<sup>19</sup>

Tables 9–11 consist of Grouped Logit equations. Their general form is

$$\begin{aligned} \ln \left( \frac{s_t}{1 - s_t} \right) = & \alpha_0 + \alpha'_t D_t + \alpha'_F D_F \\ & + \beta_R \ln(\text{R&D Stock}_t) \\ & + \beta_C \text{Private university} \\ & + \beta_P \text{Placement}_t + \beta_L \text{Local R&D}_t \\ & + \beta_E \text{Equipment ratio}_{t-1} + e_t \end{aligned} \quad (12)$$

We again suppress the university-field subscript for notational simplicity. The idea of (12) is to study determinants of the relative proportion  $s_t$  of team affiliations on the left of (12) that consists of other U.S. universities, foreign institutions or U.S. corporations. The share  $s_t$  is a probability. If this probability follows the Logistic distribution, then the logarithm of the odds ratio forms a regression function which (12) exemplifies.<sup>20</sup> Unlike (11) a key variable in this equation is Placement<sub>t</sub>, which measures the fraction of recent Ph.D.'s that are employed in top 40% departments of the top 110, in the top 12 research countries, and in U.S. industry, as noted in the discussion of Table 7.

<sup>18</sup> We divide federally funded R&D by papers lagged 2 years, in order to avoid division error bias with the logarithm of authors per paper on the left hand side of the regression.

<sup>19</sup> The dummy indicator for whether a university-field is in the top 20% of a discipline, and below, the proportion of its graduate students that place in the top 40% of schools, need to be carefully distinguished. They are not the same thing.

<sup>20</sup> See Greene (2000), Ch. 19, Section 19.4.6 for a discussion, or see another survey of the econometrics of limited dependent variables.

Table 9

Determinants of the relative contribution of U.S. universities, dependent variable: log (other top 110 share/(1 – other top 110 share)) (*t*-statistics in parentheses)

Variable or statistic	Eq. (9.1)	Eq. (9.2)	Eq. (9.3)	Eq. (9.4)
Time period	1983–1999	1983–1999	1983–1999	1983–1999
Fields included	All 12 main fields	All 12 main fields	10 fields <sup>a</sup>	10 fields <sup>a</sup>
Year dummies included	Yes, significant	Yes, significant	Yes, significant	Yes, significant
Field dummies included	Yes, significant	Yes, significant	Yes, significant	Yes, significant
Log (stock of federally funded R&D, per paper lagged 2 years)	0.056 (11.8)**	0.067 (12.7)**	0.068 (12.8)**	0.067 (12.2)**
Private university (1, yes; 0, no)	0.076 (13.1)**	0.057 (8.9)**	0.101 (15.1)**	0.078 (10.5)**
Number of awards <sup>b</sup>			0.021 (4.9)**	0.019 (4.3)**
Fraction of Ph.D.'s placed in top 40% departments, lagged 2 years		0.053 (2.6)**		0.103 (4.5)**
Local university R&D ratio <sup>c</sup>		0.218 (12.0)**		0.204 (10.4)**
Equipment expenditure/R&D, previous 3 years		0.074 (1.0)		0.005 (0.1)
Top 20% in field (1 if yes; 0 if no)		−0.034 (−5.1)**		−0.022 (−2.8)**
Root M.S.E.	0.268	0.267	0.270	0.266
Adjusted <i>R</i> <sup>2</sup>	0.57	0.57	0.60	0.61
Number of observations	9,613	8,621	7,726	7,182

Notes: Estimation method is grouped logit. Sources for the data are the Institute for Scientific Information, National Science Foundation, National Research Council, and authors' calculations.

<sup>a</sup> Data on prizes and awards are missing for agriculture and medicine.

<sup>b</sup> Awards include the Fields Medal, MacArthur Awards, the National Medal of Science, the National Medal of Technology, Fellow of the National Academy of Science, and the Nobel Prize. See the text for further discussion.

<sup>c</sup> The local university R&D ratio equals R&D in the same field but in other universities within 25 miles, divided by R&D in the same field but in other universities within 200 miles.

\*Parameter is significantly different from zero at the 5% level for a one-tailed test.

\*\* Parameter is significantly different from zero at the 1% level for a one-tailed test.

In Table 9 the dependent variable is the logarithm of the share of other top 110 schools in the research of a given university-field divided by one minus the share. While the relative share seems a useful way to explore the substitution of research by other top 110 schools for internal research, the method does have one limitation. Observations for which the relative share equals zero cannot be included in the estimation procedure because zero values rule out any finite value for the regression function. The same point applies to Tables 10 and 11, which also use Grouped Logit. However, these zero share observations are few and far between.

All the equations include year and field dummies, which are highly significant. In Eqs. (9.1) and (9.3) the logarithm of the stock of federally funded R&D per paper significantly increases the share of other top 110 schools. Private universities collaborate to a larger extent with other top 110 schools, as do schools where the faculty have earned a larger number of awards. Again we attribute the greater reach of institutions with

more research dollars, prestige, and awards to greater resources and talent.

Eqs. (9.2) and (9.4) include an array of new variables. The fraction of former Ph.D. students placed in the top 40% of schools in a field is a significant factor in collaboration. Again top 20% status in a field deters collaborations with other top 110 schools, perhaps because of the availability of graduate assistance within a university-field.

Table 10 is similar to Table 9, except that the dependent variable is the relative share of foreign institutions in the research of a university-field. For this reason the Ph.D. placement indicator is the fraction of Ph.D.'s who have located to top research countries. The role of the university-field R&D stock per paper is not as strong in Table 10 as it was before in Table 9, perhaps because the availability of foreign R&D in part drives the collaboration. And yet additional R&D and private control do increase the foreign share in (10.1) and (10.2).

The foreign placement indicator contributes strongly to foreign collaboration. It is of some interest

Table 10

Determinants of the relative foreign contribution, dependent variable:  $\log(\text{foreign share}/(1 - \text{foreign share}))$  (*t*-statistics in parentheses)

Variable or statistic	Eq. (10.1)	Eq. (10.2)	Eq. (10.3)	Eq. (10.4)
Time period	1983–1999	1983–1999	1983–1999	1983–1999
Fields included	All 12 main fields	All 12 main fields	10 fields <sup>a</sup>	10 fields <sup>a</sup>
Year dummies included	Yes, significant	Yes, significant	Yes, significant	Yes, significant
Field dummies included	Yes, significant	Yes, significant	Yes, significant	Yes, significant
Log (stock of federally funded R&D divided by papers lagged 2 years)	0.026 (4.7) <sup>**</sup>	0.014 (2.4) <sup>*</sup>	0.008 (1.2)	0.005 (0.7)
Private university (1, yes; 0, no)	0.035 (5.2) <sup>**</sup>	0.030 (4.2) <sup>**</sup>	0.007 (0.8)	0.008 (1.0)
Number of awards <sup>b</sup>			0.021 (4.2) <sup>**</sup>	0.018 (3.2) <sup>**</sup>
Fraction of Ph.D.'s placed in the top 12 <sup>c</sup> research countries, lagged 2 years		0.278 (7.1) <sup>**</sup>		0.194 (4.4) <sup>**</sup>
Equipment expenditure/R&D, previous 3 years		−0.510 (−6.7) <sup>**</sup>		−0.378 (−5.0) <sup>**</sup>
Top 20% in field (1 if yes; 0 if no)		0.035 (5.0) <sup>**</sup>		0.009 (1.0)
Root M.S.E.	0.312	0.309	0.312	0.310
Adjusted <i>R</i> <sup>2</sup>	0.72	0.73	0.67	0.67
Number of observations	9,509	9,169	7,629	7,461

Notes: Estimation method is grouped logit. Sources for the data are the Institute for Scientific Information, National Science Foundation, National Research Council, and authors' calculations.

<sup>a</sup> Agriculture and medicine lack data on prizes and awards.

<sup>b</sup> Awards include the Fields Medal, MacArthur Awards, the National Medal of Science, the National Medal of Technology, Fellow of the National Academy of Science, and the Nobel Prize. See the text for a further discussion.

<sup>c</sup> The top 12 research countries are Australia, Canada, France, Germany, Israel, Italy, Japan, the Netherlands, New Zealand, Sweden, Switzerland, and the United Kingdom. See the text for further details.

\* Parameter is significantly different from zero at the 5% level for a one-tailed test.

\*\* Parameter is significantly different from zero at the 1% level for a one-tailed test.

to note that the share of equipment expenditures in the recent past discourages foreign collaboration. This result suggests the role of the funding motive (see Section 2) for foreign and other collaborations, especially in equipment-intensive fields such as astronomy and experimental physics, which are well represented in these data (see Figs. 9 and 10 and Table 6). Foreign collaborations could amortize fixed costs of expensive equipment across countries in such fields.

Table 11 repeats the exercise of Tables 9 and 10, this time using the logarithm of the relative U.S. corporate share in the research of a university-field as the dependent variable. While the stock of R&D per paper has rather weak effects on the corporate share, private control of a university increases U.S. corporate collaboration at a high level of statistical significance. In contrast university-fields that have earned many prestigious awards collaborate less with firms. What these results suggest to us is that corporate R&D support is much sought after by private universities. This is true, except in the case of departments with a strong basic science focus that tend to win prestigious awards and

extensive federal support. Notice that because firms are the primary supporters of research collaborations with universities, it is their R&D that is the likely driver of collaboration, not the university-field's R&D.

Also in Table 11, the fraction of former Ph.D.'s placed in industry strongly drives collaboration with firms. Recent equipment-intensity also appears to substitute for firm collaborations, again suggesting the role of outside research partners in underwriting research and equipment expense.

Fig. 11 assembles time effects  $\alpha'_t$  from the regressions in Tables 8–11. These effects hold constant the science field, R&D stock, private control dummy, Ph.D. placements, equipment intensity of R&D, as well as other variables. For this reason time effects from the regression tables should lie closer to "pure" effects of technological change on collaboration than trends in the raw data. To show what these effects look like, Fig. 11 graphs the previously omitted regression coefficients for time dummies in Eqs. (8.1), (9.1), (10.1), and (11.1). A comparison of the various line graphs confirms what we have already seen—that the foreign share increases more rapidly than the other indicators,

Table 11

Determinants of the relative U.S. corporate contribution, dependent variable:  $\log(\text{U.S. corporate share}/(1 - \text{U.S. corporate share}))$  ( $t$ -statistics in parentheses)

Variable or statistic	Eq. (11.1)	Eq. (11.2)	Eq. (11.3)	Eq. (11.4)
Time period	1983–1999	1983–1999	1983–1999	1983–1999
Fields included	All 12 main fields	All 12 main fields	10 fields <sup>a</sup>	10 fields <sup>a</sup>
Year dummies included	Yes, significant	Yes, significant	Yes, significant	Yes, significant
Field dummies included	Yes, significant	Yes, significant	Yes, significant	Yes, significant
Log (stock of federally funded R&D divided by papers lagged 2 years)	0.024 (2.7) <sup>**</sup>	0.026 (2.8) <sup>**</sup>	0.003 (0.3)	0.010 (1.0)
Private university (1, yes; 0, no)	0.070 (6.3) <sup>**</sup>	0.067 (5.9) <sup>**</sup>	0.144 (11.3) <sup>**</sup>	0.136 (10.2) <sup>**</sup>
Number of awards <sup>b</sup>			-0.060 (-6.8) <sup>**</sup>	-0.063 (-6.6) <sup>**</sup>
Fraction of Ph.D.'s placed in U.S. industry, lagged 2 years		0.532 (14.3) <sup>**</sup>		0.482 (11.5) <sup>**</sup>
Equipment expenditure/R&D, previous 3 years		-0.263 (-2.4)*		-0.316 (-2.7) <sup>**</sup>
Top 20% in field (1 if yes; 0 if no)		-0.005 (-0.4)		0.010 (0.8)
Root M.S.E.	0.461	0.453	0.460	0.452
Adjusted $R^2$	0.69	0.70	0.63	0.64
Number of observations	8,378	8,103	6,597	6,472

Notes: Estimation method is grouped logit. Sources for the data are the Institute for Scientific Information, National Science Foundation, National Research Council, and authors' calculations.

<sup>a</sup> Agriculture and medicine lack data on prizes and awards.

<sup>b</sup> Awards include the Fields Medal, MacArthur Awards, the National Medal of Science, the National Medal of Technology, Fellow of the National Academy of Science, and the Nobel Prize. See the text for a further discussion.

\* Parameter is significantly different from zero at the 5% level for a one-tailed test.

\*\* Parameter is significantly different from zero at the 1% level for a one-tailed test.

and so on. However, there is an interesting jump in the time series of relative shares contributed by foreign institutions, U.S. universities, and U.S. firms, which occurs for papers published between 1990 and 1991. This jump applies primarily to papers conceived in the

late 1980s given various delays that we discuss immediately below.

Our hypothesis is that the 1990–1991 jump results from improvements in the National Science Foundation's NSFNET from 1987 onward. National Research

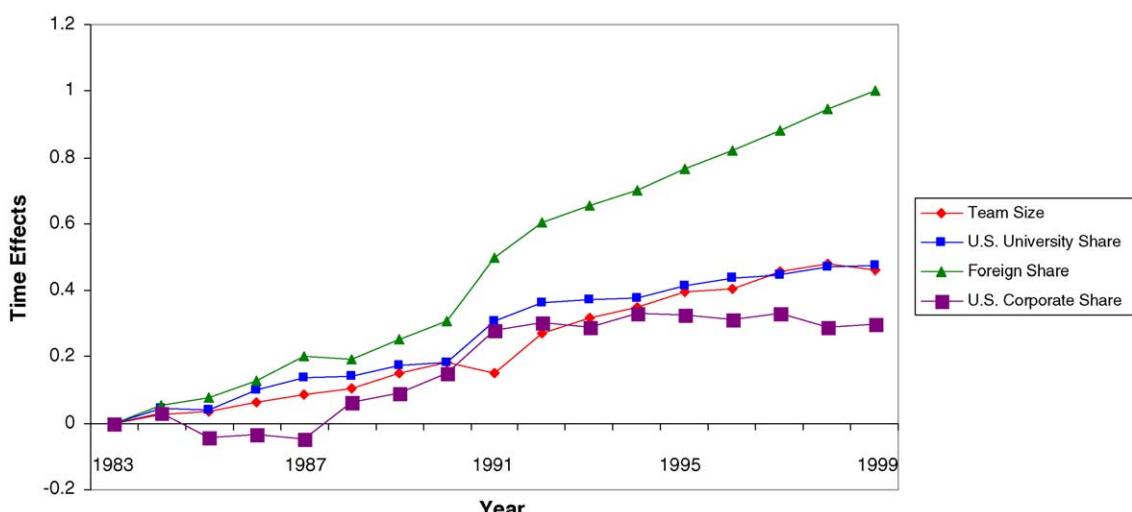


Fig. 11. Trends in team size and institutional collaborations, time effects from regression analysis, 1983–1999 (1983 = 0.0).

Council (1999, Ch. 7) discusses the role of the National Science Foundation (NSF) in building a national “network of networks” during 1986–1991. It is worth noting that the speed of the NSFNET backbone was an anemic 56 KB/s as of 1986 (National Research Council, 1999, p. 78). Only after Merit Computer Network Incorporated, IBM and MCI were placed under contract to expand NSFNET in 1987 did the speed and connectivity of NSFNET begin to approach modern standards (National Research Council, 1999, p. 179). The pattern of international linkages to NSFNET is also consistent with this story. Canada, Denmark, Finland, France, Iceland, Norway, and Sweden connected in 1988, while Australia, Germany, Israel, Italy, Japan, Mexico, Netherlands, New Zealand, and the United Kingdom connected in 1989 (Mowery and Simcoe, 2002, p. 1376). The domestic expansion of NSFNET, and slightly later its international expansion were followed ultimately in 1994 by the merger of NSFNET with the commercial internet known as CIX. Noam (2001, Ch. 3, pp. 63–65) documents the merger and the disputes that led up to it. The provision to the research community of a high-speed NSFNET after 1987 and its interconnectivity to the rest of the research world after 1988 enabled scientific teams to collaborate more cheaply and effectively at a distance at precisely this point in history. The change applies to papers published in the 1990s and afterwards, especially given construction lags after 1987, followed by gestation and publication lags of 2–3 years. This interpretation seems reasonable, especially given that there is no clear jump in team size in 1990–1991 in Fig. 11: instead, “external” team workers are replacing “internal” ones. Moreover, the unfolding of interconnectivity in the early 1990s and the development of the World Wide Web stimulated the deployment of the web browser from 1995 going forward (National Research Council, 1999; Mowery and Simcoe, 2002). These complementary innovations helped to sustain the growth of institutional collaboration throughout the 1990s as observed in Fig. 11. Of course, we do not deny the importance of other factors driving the growth of institutional collaborations. These include the secular growth of graduate programs, the growth of Big Science Projects, and the expansion of science in Europe and Japan. But such factors account for long-term growth rather than the once-and-for-all jump in collaboration that one observes in 1990–1991. The most likely candidate for the jump is

the deployment of NSFNET after 1987, a key technology that lowered the cost of joint research between teams in different institutions and countries.

The empirical work concludes with Table 12, which is concerned with the explanation of research “output” measured by the sum of fractions of papers and citations to those papers by a university-field.<sup>21</sup> The estimation method is Ordinary Least Squares. As before, year and field dummies are included throughout the table and are highly significant. Eqs. (12.1)–(12.3) use the logarithm of the fractional number of papers as the dependent variable, while (12.4) to (12.6) use the logarithm of fractional 5-year citations. By “fractional” of course, we mean the sum of the internal paper and citation fractions within a university-field.

In Eq. (12.1) and the others the coefficient of the logarithm of the lagged stock of R&D is positive and highly significant. Consistent with the analytical framework in Section 2, it is also significantly less than 1.0. This is also the finding of Adams and Griliches (1998), who suggest that diminishing returns to the stock of R&D apply at the university-field level. In (12.1) the logarithm of all authors per paper decreases the output of *fractional* papers. However, this result picks up movement of authorship outside the university-field in larger teams. The negative sign is spurious. It represents mismatching of “total” teams with “inside” papers: larger teams involve more institutional collaboration, a relatively smaller number of “inside” authors, and thus a smaller number of internal, fractional papers. To see this, notice that when the fractional number of authors inside a university-field is used instead, as in (12.2), the coefficient on the logarithm of authors reverses and becomes positive and highly significant. That is, “inside” authors increase “inside” papers.

Eq. (12.3) adds shares of other U.S. universities, foreign institutions, and U.S. firms to (12.1). Shares of outside institutions reduce the fractional number of papers, in part because authorship moves away from the university-field. Another possibility, though, which is raised by the citation results, is that a quantity-quality

<sup>21</sup> Recall that if a university-field contributes half a paper, the fraction assigned is 1/2; if it contributes a third, then the fraction is 1/3, and so forth (see footnote 7). For present purposes this fractionation of papers and citations avoids multiple counting of the papers and citations across universities and limits “output” to the estimated portion contributed by a university-field.

Table 12

Determinants of research “output”, dependent variables: log (papers), log (citations over 5 years) (*t*-statistics in parentheses)

Variable or statistic	Log (papers)			Log (Citations over 5 years)		
	Eq. (12.1)	Eq. (12.2)	Eq. (12.3)	Eq. (12.4)	Eq. (12.5)	Eq. (12.6)
Time period	1981–1999	1981–1999	1981–1999	1981–1995	1981–1995	1981–1995
Fields included	All 12 main fields	All 12 main fields	All 12 main fields			
Year dummies included	Yes, significant	Yes, significant	Yes, significant	Yes, significant	Yes, significant	Yes, significant
Field dummies included	Yes, significant	Yes, significant	Yes, significant	Yes, significant	Yes, significant	Yes, significant
Log (stock of federally funded R&D)	0.457 (89.0)**	0.450 (87.1)**	0.443 (89.3)**	0.553 (69.6)**	0.546 (68.0)**	0.557 (69.6)**
Log (authors per paper)	−0.085 (−4.8)**		0.019 (1.1)	0.312 (10.2)**		0.264 (8.4)**
Log (university-field authors per paper)		0.286 (9.0)**			0.548 (10.8)**	
Top 110 U.S. university share per paper			−5.973 (−30.1)**			1.276 (4.0)**
Foreign share per paper				−1.059 (−7.6)**		1.237 (5.2)**
U.S. corporate share per paper				−.378 (−1.4)		0.094 (0.2)
Root M.S.E.	0.497	0.495	0.453	0.688	0.687	0.686
Adjusted <i>R</i> <sup>2</sup>	0.80	0.81	0.80	0.82	0.82	0.82
Number of observations	10,772	10,772	10,772	8,504	8,504	8,504

Notes: Estimation method is OLS. Sources for the data are the Institute for Scientific Information, National Science Foundation, National Research Council, and authors' calculations.<sup>b</sup> Awards include the Fields Medal, MacArthur Awards, the National Medal of Science, the National Medal of Technology, Fellow of the National Academy of Science, and the Nobel Prize. See the text for a further discussion. \*Parameter is significantly different from zero at the 5% level for a one-tailed test.

\*\* Parameter is significantly different from zero at the 1% level for a one-tailed test.

tradeoff exists in the data. An increase in the foreign share may genuinely imply that fewer but better papers are written within a university-field.

If Eqs. (12.1)–(12.3) seem to suggest that collaboration reduces the number of papers for reasons that are mostly spurious, then (12.4)–(12.6) indicate that institutional collaboration increases citation and thus total scientific impact. The counterpart to the spurious negative effect of all team members on fractional papers in (12.1) is found in (12.4). In that equation the elasticity of citations with respect to all authors is much smaller compared with that of “inside” authors in (12.5). In (12.6) we see that that institutional collaboration increases fractional citations. Overall, the evidence of Table 12 suggests that the scientific division of labor increases research “output” as measured by production and citation influence of a university-field.<sup>22</sup> But more

work is needed to verify this hypothesis, beyond the evidence in this paper.

## 7. Discussion and conclusion

This paper has presented evidence on patterns of research collaboration in U.S. universities over the final two decades of the 20th century. The evidence on the size of scientific teams, as measured by authors per paper, suggests that specialization and the division of labor have increased markedly over this period, especially during the 1990s. Our findings on collaboration between institutions suggest a similar pattern of developments, but with some new twists. Collaboration with foreign universities increases more rapidly over time than team size, while domestic collaboration increases less rapidly. We take this as evidence that the location of team members is shifting and is becoming more geographically dispersed, perhaps because of funding advantages coupled with telecommunications improvements. However, we lack complete information on the causal factors directing this dispersion. It

<sup>22</sup> This is not a perfect test. Some citations could involve hidden self-citations to previous collaborations by the same research team. We cannot address this upward bias with the data that we have, because we cannot link names and addresses of researchers, including across papers.

seems plausible to say that domestic collaboration has for a long time been more feasible than international collaboration, and that only recently have modern communications technologies made international science viable for researchers on projects of normal size. This interpretation receives support from Figs. 7 and 8 where it is the *smaller* teams that are becoming internationalized at a faster rate. But in addition, an increasing emphasis on large databases, as in biology and medicine, and on massive instrumentation, as in astronomy and physics, may have also played important roles in these trends towards greater internationalization.

The growth of collaboration as observed in this article could be viewed as consistent with the increasing efficiency of the research enterprise. Collaboration at a distance permits a combination of complementary capabilities that leads to the execution of more and hopefully better research. In this way it is likely welfare-improving. However, a more somber interpretation is that lagging public funding of scientific research in the U.S. compels universities to engage in institutional collaborations, especially with firms and foreign institutions, as a substitute means of support. Additional investigations into the push and pull behind institutional collaboration seem well warranted,

especially given the implications of expanded teamwork for the productive efficiency of scientific research.

## Acknowledgements

The Andrew W. Mellon Foundation has generously supported this research. We thank Nancy Bayers and Henry Small of the Institute for Scientific Information (ISI) for substantial assistance with the ISI data used in this paper. We are also indebted to Manuel Trajtenberg, Cherisa Yarkin, and Ulrike Malmendier for thoughtful comments on early drafts. Finally, this paper has benefited from presentations at the Georgia Institute of Technology, the National Bureau of Economic Research, the American Association for the Advancement of Science meetings in Denver; February 2003; and the American Economic Association meetings in San Diego, January 2004. We take full responsibility for any remaining errors in the paper.

## Appendix A. The top 110 U.S. universities

See Table A.1.

Table A.1

The top 110 U.S. universities in the Institute for Scientific Information (ISI) database ranked by 1998 federal R&D

University name (rank)	1998 Federal R&D expenditures	University name (rank)	1998 Federal R&D expenditures
Johns Hopkins University (1)	752.983 <sup>a</sup>	Georgetown University (56)	84.801
Stanford University (2)	342.426	University of New Mexico, all campuses (57)	84.365
University of Washington-Seattle (3)	336.748	Virginia Polytechnic Institute and State University (58)	82.734
University of Michigan, all campuses (4)	311.450	Oregon State University (59)	82.416
Massachusetts Institute of Technology (5)	310.741	Michigan State University (60)	81.146
University of California-San Diego (6)	262.303	Colorado State University (61)	80.451
Harvard University (7)	251.876	Yeshiva University (62)	80.000
University of Pennsylvania (8)	247.914	North Carolina State University at Raleigh (63)	79.533
University of Wisconsin-Madison (9)	240.513	University of Maryland at Baltimore (64)	78.037
University of California-Los Angeles (10)	233.702	SUNY at Buffalo, all campuses (65)	76.037
Columbia University, all campuses (11)	229.723	University of Illinois at Chicago (66)	73.797
University of Colorado, all campuses (12)	228.342	Oregon Health Sciences University (67)	71.054
University of California-San Francisco (13)	219.912	University of Texas Health Science Center Houston (68)	70.446
University of Alabama, all campuses (14)	205.511	Rutgers the State University of NJ, all campuses (69)	69.829
Yale University (15)	205.046	University of Tennessee, all campuses (70)	69.793
University of Minnesota, all campuses (16)	204.741	Princeton University (71)	69.005
Cornell University, all campuses (17)	204.187	University of California-Santa Barbara (72)	68.408
University of Southern California (18)	190.547	Woods Hole Oceanographic Institution (73)	64.765
Washington University (19)	187.173	University of Missouri, all campuses (74)	63.556

Table A.1(Continued)

University name (rank)	1998 Federal R&D expenditures	University name (rank)	1998 Federal R&D expenditures
Pennsylvania State University, all campuses (20)	186.274	Tufts University (75)	61.167
California Institute of Technology (21)	177.748	University of Kentucky, all campuses (76)	60.760
Duke University (22)	172.532	University of Nebraska, all campuses (77)	58.482
University of North Carolina at Chapel Hill (23)	171.505	Wayne State University (78)	57.646
University of California-Berkeley (24)	171.135	Wake Forest University (79)	56.705
University of Illinois at Urbana-Champaign (25)	168.871	New Mexico State University, all campuses (80)	56.587
University of Pittsburgh, all campuses (26)	168.511	University of Texas Health Science Center San Antonio (81)	55.004
University of Texas at Austin (27)	165.082	Utah State University (82)	54.903
University of Arizona (28)	161.999	University of Georgia (83)	54.712
Texas A&M University, all campuses (29)	144.938	University of Connecticut, all campuses (84)	53.189
Case Western Reserve University (30)	132.274	Tulane University (85)	52.924
University of Rochester (31)	130.773	Iowa State University (86)	51.196
University of Maryland at College Park (32)	129.198	University of Kansas, all campuses (87)	50.567
Northwestern University (33)	127.911	Florida State University (88)	50.451
University of Chicago (34)	125.982	Virginia Commonwealth University (89)	48.167
Ohio State University, all campuses (35)	124.177	Dartmouth College (90)	45.053
Emory University (36)	118.045	Louisiana State University, all campuses (91)	47.090
University of Iowa (37)	115.312	University of California-Irvine (92)	45.902
University of California-Davis (38)	114.912	Washington State University (93)	44.510
Georgia Institute of Technology, all campuses (39)	113.643	Brown University (94)	44.412
Baylor College of Medicine (40)	110.610	Rockefeller University (95)	43.845
University of Florida (41)	106.510	Arizona State University Main (96)	41.359
Vanderbilt University (42)	106.325	Rice University (97)	34.772
Boston University (43)	104.428	University of Delaware (98)	33.688
University of Miami (44)	101.492	CUNY, all campuses (99)	32.412
New York University (45)	101.426	University of AK Fairbanks, all campuses (100)	31.505
University of Utah (46)	100.722	University of Vermont (101)	31.460
University of Massachusetts, all campuses (47)	100.122	University of California-Santa Cruz (102)	29.849
University of Texas Southwestern Med Center Dallas (48)	97.200	Syracuse University, all campuses (103)	29.200
Indiana University, all campuses (49)	95.840	Brandeis University (104)	28.098
Carnegie Mellon University (50)	95.046	University of Oregon (105)	27.041
University of Virginia, all campuses (51)	93.328	University of New Hampshire (106)	25.913
Purdue University, all campuses (52)	92.844	West Virginia University (107)	24.985
SUNY at Stony Brook, all campuses (53)	91.531	University of California-Riverside (108)	22.988
University of Cincinnati, all campuses (54)	90.307	Loyola University of Chicago (109)	17.685
University of Hawaii at Manoa (55)	86.886	Lehigh University (110)	13.019

Notes: Federal R&D is taken from the CASPAR database of the National Science Foundation.

<sup>a</sup> The data for Johns Hopkins University includes R&D expense for the Applied Physics Laboratory.

## References

- Adams, J.D., 2002. Comparative localization of academic and industrial spillovers. *Journal of Economic Geography* 2, 253–278.
- Adams, J.D., 2005. Learning, internal research, and spillovers: evidence from a sample of R&D laboratories. *The Economics of Innovation and New Technology*, in press.
- Adams, J.D., Jaffe, A.B., 1996. Bounding the effect of R&D: an analysis using matched establishment-firm data. *Rand Journal of Economics* 27, 700–721.
- Adams, J.D., Griliches, Z., 1998. Research productivity in a system of universities. *Annales D'Economie et de Statistique* 49/50, 127–162.
- Adams, J.D., Chiang, E.P., Starkey, K., 2001. Industry–university cooperative research centers. *Journal of Technology Transfer* 26, 73–86.
- Adams, J.D., Chiang, E.P., Jensen, J.L., 2003. The influence of federal laboratory R&D on industrial research. *The Review of Economics and Statistics* 85, 1003–1020.

- Adams, J.D., Marcu, M., 2004. R&D Sourcing, Joint Ventures, and Innovation: A Multiple Indicators Approach. NBER Working Paper 10474, Cambridge, MA.
- Arora, A., Gambardella, A., 1990. Complementarity and external linkages: the strategies of the large firms in biotechnology. *Journal of Industrial Economics* 38, 361–379.
- Becker, G.S., 1985. Human capital, effort, and the sexual division of labor. Part 2: Trends in women's work, education, and family building. *Journal of Labor Economics* 3, S33–S58.
- Becker, G.S., Murphy, K.M., 1992. The division of labor, coordination costs, and knowledge. *Quarterly Journal of Economics* 107, 1137–1160.
- Chiang, A.C., 1974. Fundamental Methods of Mathematical Economics, second ed. McGraw-Hill, New York.
- Dixit, A.K., 1990. Optimization in Economic Theory, second ed. Oxford University Press, Oxford, UK.
- Ehrenberg, R.G., 2003. Studying ourselves: the academic labor market. *Journal of Labor Economics* 21, 267–288.
- Ehrenberg, R.G., Rizzo, M.J., Jakubson, G.H., 2005. Who bears the growing cost of science at the university. In: Ehrenberg, R.G., Stephan, P.E. (Eds.), *Science and the University*. University of Wisconsin Press, Madison, WI.
- Greene, W.H., 2000. *Econometric Analysis*, fourth ed. Prentice-Hall, Upper Saddle River, NJ.
- Hicks, D., Katz, J.S., 1996. Science policy for a highly collaborative science system. *Science and Public Policy* 23, 39–44.
- Holmstrom, B., 1982. Moral hazard in teams. *Bell Journal of Economics* 13, 324–340.
- Ichniowski, C., Shaw, K., Prennushi, G., 1997. The effects of human resource management practices on productivity: a study of steel finishing lines. *The American Economic Review* 87, 291–313.
- Jones, C.I., 1995. R&D-based models of economic growth. *Journal of Political Economy* 103, 759–784.
- Kandel, E., Lazear, E.P., 1992. Peer pressure and partnerships. *Journal of Political Economy* 100, 801–817.
- Mowery, D.C., 1992. International collaborative ventures and the commercialization of new technologies. In: Rosenberg, N., Landau, R., Mowery, D.C. (Eds.), *Technology and the Wealth of Nations*. Stanford University Press, Stanford, CA.
- Mowery, D.C., Rosenberg, N., 1998. Paths of Innovation: Technological Change in 20th Century America. Cambridge University Press, Cambridge, UK.
- Mowery, D.C., Simcoe, T., 2002. Is the Internet a US invention?—An economic and technological history of computer networking. *Research Policy* 31, 1369–1387.
- National Research Council, 1995. *Research-Doctorate Programs in the United States: Continuity and Change*. National Academy Press, Washington, DC.
- National Research Council, 1999. *Funding a Revolution: Government Support for Computing Research*. National Academy Press, Washington, DC.
- National Science Foundation, 2004. *Science and Engineering Indicators 2004*. National Science Board, Arlington, VA.
- Noam, E.M., 2001. *Interconnecting the Network of Networks*. MIT Press, Cambridge, MA.
- Powell, W.W., 1996. Inter-organizational collaboration in the biotechnology industry. *Journal of Institutional and Theoretical Economics* 152, 197–215.
- Presser, S., 1980. Collaboration and the quality of research. *Social Studies of Science* 10, 95–101.
- Romer, P.M., 1990. Endogenous technological change. *Journal of Political Economy* 98 (Part 2), S71–S102.
- Rosen, S., 1981. The economics of superstars. *The American Economic Review* 71, 845–858.
- Rosen, S., 1982. Authority, control, and the distribution of earnings. *Bell Journal of Economics* 13, 311–323.
- Sauer, R.D., 1988. Estimates of the returns to quality and coauthorship in economic academia. *Journal of Political Economy* 96, 855–866.
- de Solla Price, D.J., 1986. *Big Science, Little Science, and Beyond*, revised edition. Columbia University Press, New York.
- Stephan, P.E., Levin, S.G., 2000. The importance of implicit contracts in collaborative scientific research. In: Mirowski, P., Sent, E.-M. (Eds.), *The Economics of Science*. University of Chicago Press.
- Wiener, N., 1994. *Invention: The Care and Feeding of Ideas*. MIT Press, Boston.
- Zucker, L.G., Darby, M.R., Armstrong, J.S., 2001. Commercializing Knowledge: University Science, Knowledge Capture, and Firm Performance. NBER Working Paper 8499, Cambridge, MA.
- Zuckerman, H., Merton, R.K., 1973. Age, aging, and age structure in science. In: Merton, R.K., Storer, N.W. (Eds.), *The Sociology of Science: Theoretical and Empirical Investigations*. University of Chicago Press, Chicago.