The Impact of Research Collaboration on Scientific Productivity

Sooho Lee and Barry Bozeman

Social Studies of Science 2005; 35; 673

DOI: 10.1177/0306312705052359

S|S|S

ABSTRACT Based on the curricula vitae and survey responses of 443 academic scientists affiliated with university research centers in the USA, we examine the long-standing assumption that research collaboration has a positive effect on publishing productivity. Since characteristics of the individual and the work environment are endogenously related to both collaboration and productivity, this study focuses on the mediating effect of collaboration on publishing productivity. By using the two-stage least squares analysis, the findings indicate that in the presence of moderating variables such as age, rank, grant, gender, marital status, family relations, citizenship, job satisfaction, perceived discrimination, and collaboration strategy, the simple number ('normal count') of peer-reviewed journal papers is strongly and significantly associated with the number of collaborators. However, the net impacts of collaboration are less clear. When we apply the same model and examine productivity by 'fractional count', dividing the number of publications by the number of authors, we find that number of collaborators is not a significant predictor of publishing productivity. In both cases, 'normal count' and 'fractional count', we find significant effects of research grants, citizenship, collaboration strategy, and scientific field. We believe that it is important to understand the effects of the individual and environmental factors for developing effective strategies to exploit the potential benefits of collaboration. We note that our focus is entirely at the individual level, and some of the most important benefits of collaboration may accrue to groups, institutions, and scientific fields.

Keywords normal and fractional publication counts, research collaboration, scientific productivity

The Impact of Research Collaboration on Scientific Productivity

Sooho Lee and Barry Bozeman

The collaboration of scientists[1] in research activity has become the norm (Beaver & Rosen, 1979b). The increasingly interdisciplinary, complex, and costly characteristics of modern science encourage scientists to get involved in collaborative research. Funding agencies, particularly government agencies, facilitate active research collaboration as part of their funding conditions.[2] Despite the ubiquitous nature of collaboration in science, the benefits of collaboration are more often assumed than investigated. Our interest is in the impacts of research collaboration on publication productivity. Do those who collaborate more tend to have more publications? Most studies of collaboration include an underlying assumption that collaborative activity increases research productivity (Lotka,

Social Studies of Science **35/5(October 2005) 673–702**

ISSN 0306-3127 DOI: 10.1177/0306312705052359
www.sagepublications.com

1926; Price & Beaver, 1966; Zuckerman, 1967; Godin & Gingras, 2000). Surprisingly few studies have actually tested that proposition. Certainly, there are good reasons to think that collaboration may enhance research productivity. Many collaborations center on the joint use of expensive or unique equipment without which research would be not only less productive but also impossible (Meadows, 1974; Thorsteinsdottir, 2000). In an age of '*Mode 2*' science,[3] some research seems to require collaboration to bring special expertise and knowledge not otherwise available but crucial to research outcomes (Thornsteinsdottir, 2000). Often, tacit knowledge and knowledge of technique are best conveyed through collaboration (Beaver & Rosen, 1978, 1979a). In many cases, collaboration is the key mechanism for mentoring graduate students and postdoctoral researchers (Bozeman & Corley, 2004), and enhancing the productivity of individual scientists (Melin, 2000).

Despite these good reasons to expect that scientific collaboration will enhance productivity, the relationship between the two is not obvious. The fact that researchers and policy-makers *perceive* that collaboration increases productivity does not make it so. Indeed, there are some arguments as to why collaboration may undermine productivity. Transaction costs are usually an unavoidable consequence of working with others (Landry & Amara, 1998). Staying in touch by various media, engaging in social ingratiation, waiting for others to comment, respond, or do their part of the research – these are just some of the factors taking time and energy even in the best collaborative relationships. Not all collaborations are ideal. Most active collaborators have had projects that were never finished or that had disappointing results because one or more of the collaborators did not live up to expectations. Many researchers, especially senior researchers, collaborate not so much to increase their own productivity as to mentor graduate students and postdoctoral researchers (Bozeman & Corley, 2004). While such collaboration is likely to enhance the productivity of some parties, others are likely to be a drag on the productivity of the more experienced researchers; to the latter, therefore, this may represent a 'tithe' given voluntarily.

The conceptualization and measurement of collaboration present difficulties. In particular, does one focus on the productivity increments related to particular scientific outputs, such as publications, or take a much broader view of increments to scientific capacity? And if one examines increments in the capacity to do scientific work, does one focus on the individual, the research group, or some concepts of a scientific field? Elsewhere we have considered the impact of collaboration strategies on 'scientific and technical human capital' (S&T human capital) (Bozeman et al., 2001; Bozeman & Rogers, 2002). S&T human capital is the sum of scientific and technical and social knowledge, skills and resources embodied in a particular individual. It is both human capital endowments, such as formal education and training, and social relations and network ties that bind scientists and the users of science together. S&T human capital is the unique set of resources the individual brings to his or her own work and to collaborative efforts. S&T human capital can be understood at the level of the individual and it is possible to measure the individual scientist's training, skills and even tacit knowledge. It also is possible to measure the individual's ties to networks and transactions with others in those networks. In focusing on the individual, it seems very useful to think of S&T human capital in terms of the scientist's professional life cycle.

Examining collaboration from the standpoint of a multi-level S&T human capital model shows that productivity implications are part and parcel of the analytic focus. Thus, for example, any particular collaboration may be a productivity decrement for specific individuals but a productivity increment for a field, educational cohort, or 'knowledge value collective'.[4] A senior researcher choosing to collaborate with a graduate student may, from one perspective, not be making the most productive use of her time. Working alone or with another senior scholar would perhaps result in equal or higher quality achieved in less time. But the same activity may be quite productive from the standpoint of the work group or the scientific field, because the collaboration is likely to lead to a greater increment in S&T human capital than would work performed alone.

Based on the data from 443 academic scientists, our research examines the effects of collaboration on scientists' productivity, measured in terms of the number of journal paper publications. We examine publication productivity by two measures, a simple count of peer-reviewed journal papers (a normal count) and a fractional count in which co-authored papers are divided by the number of coauthors. We are only measuring individual productivity, not group productivity.

Does Collaboration Affect Research Productivity?

Since Lotka's pioneering works on the productivity of scientists (Lotka, 1926), many subsequent studies have confirmed a strong relationship between collaboration and scientific productivity. Analyzing 592 scientists' publications and collaborative activities, Price & Beaver (1966: 1014) found that 'there is a good correlation between the productivities and the amount of collaboration of the authors. The most prolific [person] is also by far the most collaborating, and three of the four next most prolific are also among the next most frequently collaborating.'

With interviews of 41 Nobel laureates in science, Zuckerman (1967) identified a strong relationship between collaboration and productivity: laureates published more and were more apt to collaborate than a matched sample of scientists. In a study of collaboration in musicology, Pao (1982) also identified a strong relationship between collaboration and productivity. Although only 15% of the literature of musicology was the result of collaborative authorship, the musicologists who collaborated the most were also the most productive. Applying a normalized diversity measure to study the productivity of authors, Pao found a high degree of correlation between productivity and collaboration in computational musicology.

Pravdic & Oluic-Vukovic (1986) analyzed collaborative patterns in chemistry at both the individual and the group level. They found that scientific output as measured by publications is closely dependent on the frequency of collaboration among authors. They argued that productivity is affected by the type of collaborative links: while collaboration with high-productivity scientists tends to increase personal productivity, collaboration with low-productivity scientists generally decreases it.

Given the strong relationship between collaboration and productivity, what elements in collaboration can affect productivity? Despite the lack of direct causal explanations, several elements have been identified from the literature:[5] division of labor, complementary skills, time efficiency, intellectual stimulus, intellectual renewal or new skills learned from collaborator, companionship and a sounding board to discussion of research, access to equipment, communication of new information, and new publishing opportunities.

Do motivations for collaboration matter to productivity? In an early publication about motives for collaboration, Beaver & Rosen (1978) identified 18 motives: access to special equipment and facilities, access to special skills, access to unique materials, access to visibility, efficiency in use of time, efficiency of use of labor, to gain experience, to train researchers, to sponsor a protégé, to increase productivity, to multiply proficiencies, to avoid competition, to surmount intellectual isolation, need for additional confirmation of evaluation of a problem, need for stimulation of cross-fertilization, spatial propinquity, and accident or serendipity. The authors provided a conceptual analysis, but no data about motives or their impacts.

Recently Melin (2000) surveyed 195 university professors about their motives for collaboration and the chief benefits of collaboration. In their answers to open-ended questions, the respondents' most often-reported (41%) motive for collaboration is that the 'co-author has special competence'. Other common motives included 'co-author has special data or equipment (20%)', 'social reasons: old friends, past collaboration (16%)', 'supervisor–student relation (14%)', and 'development and testing of new methods (9%)'. With regard to the benefits of collaboration, the respondents pointed to 'increased knowledge (38%)', 'higher scientific quality (30%)', 'contact and connections for future work (25%)', and 'generation of new ideas (17%)'. Melin concluded that scientists collaborate for strong pragmatic reasons.

A Model of Collaboration's Effects on Productivity

Based on our literature review of collaboration and productivity, this study hypothesizes that collaboration tends to have positive effects on research productivity. Policy-makers apparently assume that collaboration has positive effects on research productivity, otherwise one would be hard pressed to account for the diverse programs that in one way or another either encourage or mandate research collaboration. Similarly, researchers themselves seem to have accepted the idea that collaboration results in greater productivity (Beaver, 2001). There are many good reasons to believe that collaboration has salutary effects on research productivity. Research in many fields is more complex and requires more specialized knowledge, more than any single individual can expect to have. Collaboration permits individuals to play to their strong suits, contributing their strongest skills and deepest knowledge, relying on others to contribute other skills and knowledge. Research groups sometimes seem to be more than the sum of their parts, producing a synergistic effect.

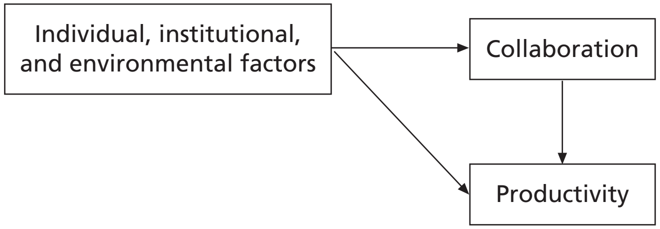
Even if collaboration does have an effect on productivity, the relationship is not so straightforward, since the various individual, institutional, and environmental factors in research activity are often endogenous to collaboration and productivity. As shown in Figure 1, our modeling strategy focuses on examining whether any observed relationship between collaboration and productivity is a direct one, or perhaps a function of interactions with other variables correlated with both collaboration and productivity.

Age, Rank, and Status

Is the effect of collaboration on productivity chiefly a function of the scientists' age? Since senior scientists, or at least those who have had longer careers, have had more time to develop S&T human capital and to build up their professional networks, any productivity increment from collaboration could be confounded with age.

In the sociological study of scientific productivity, age has long been an important issue. Lehman (1953) argued that scientists' major contributions are most likely to occur in their late 30s or early 40s, and thereafter decline in frequency. He emphasized that the age peak occurred earlier in abstract and theoretical disciplines such as theoretical physics, and later in more empirically based fields such as biology. Pelz & Andrews (1966) found two productivity peaks: the first one in scientists' late 30s and early 40s and the second one at age 50 years. Stephen Cole (1976) reported a slightly curvilinear relationship between age and quality of publications for a cross-section of academics in six scientific fields. In a more recent study of age and productivity, Levin & Stephan (1991) found that life cycle

FIGURE 1
Basic relationship between collaboration and productivity



effects are present in a fully specified model of publishing productivity that, among other things, controls for individual fixed effects such as motivation and ability. Using data from 903 natural scientists, they found evidence that, on average, scientists become less productive as they age and that the age effect is attributed to age per se and not to the possibility that older scientists in the sample have different attributes, values, or access to resources than younger members in the sample.

The relationship between collaboration and productivity might be explained by one's rank or tenure. It seems reasonable to expect that collaboration would be a different experience for tenured senior faculty and research group leaders than for untenured junior faculty, postdoctoral researchers, or graduate students. The specter of exploitation is always in the background when there are status differences among collaborators. At the same time, a collaboration that is quite productive for an experienced junior researcher may prove 'inefficient' for the mentor. Finally, there are important learning effects that may make collaboration more productive for more senior scholars. Not only do they have time to acquire greater knowledge and scientific and technical human capital, but they also have more experience with the collaboration process itself and, all else being equal, may have more productive returns from collaboration.

Research Grants and Contracts

Both collaboration and productivity may be influenced by success with grants and contracts. In the first place, most grants are for teams of researchers, and those who are working on grants often have commitments to devoting a certain percentage of their time to collaborative or team-based goals, projects, and publications. Second, the principal investigator (PI) of the grant often has an extended set of collaborations, not only because of formal contractual commitments, but also due to norms for crediting the PI in publications using the PI's data or experimental apparatus. In general, we expect those with grants, especially larger grants (in funding dollar terms), to collaborate more and to have more publications. We do not expect that the dollar amount of the grant will be nearly so important as simply having been awarded grants or contracts. In the first place, dollar amounts are often related to field- and discipline-specific dynamics, such as the expense of equipment. In the second place, earlier research has shown that research productivity is not monotonic in its relationship to magnitude of funding (Gaughan & Bozeman, 2002; Godin, 2003).

We consider not only the dollar amount of grants, but also the scientist's 'batting average': the percentage of submitted proposals that actually are funded. This is a rough quality indicator (though a messy one, given the many confounding factors related to the funding agency, institution, and discipline) and, perhaps, an indication of greater cumulative advantage.

Gender and Family Relations

One of the most consistent findings in the literature on research productivity is that women tend to have somewhat lower publication rates than men (Cole & Zuckerman, 1984; Fox & Faver, 1985; Long, 1987; Bellas & Toutkoushian, 1999). This may partly be due to the fact that women collaborate less than men and have less-developed collaboration networks. Or it could be due to systematic discrimination that may make it more difficult for women to obtain resources, and this may, in turn, limit their ability to publish. It is also possible that women collaborate less and produce fewer scientific papers because, compared with men, they are less likely to have a full-time homemaking spouse, and more likely to have a prominent role in child-rearing (Astin, 1978; Kyvik & Teigen, 1996). Marital status also seems to interact with gender and productivity, with married men being most productive and unmarried women the least. We anticipate that gender, and family and marital status, will moderate the relationship of collaboration and productivity.

However, contrary to the general perception and findings that women are less productive, Clemente (1973) argued that sex differences in publication productivity are negligible. He used a sample of 1899 male and 306 female sociologists and examined types of publications. Similarly, Wanner et al. (1981) found that gender does not affect productivity in terms of papers published. They used a sample of 17,399 university faculty members from almost all fields and disciplines.

In a more recent study, using the longitudinal productivity data of 556 male and 603 female biochemists, Long (1992) found that sex differences in the number of publications and citations increase during the first decade of the career, but are reversed later in the career. He also found that papers by women on average receive more citations than those by men. As revealed in Long's study, lifetime differences in productivity might be negligible or small, but the difference in the early stage of careers seems more visible. Xie & Shauman (1998) again confirmed a decline in the effects of gender on scientific productivity, attributing this in part to the increasing ratio of women in scientific jobs.

Citizenship

With increasing numbers[6] of foreign nationals working in US research institutions, factors related to nationality, culture, and language are likely to affect collaboration and, in turn, productivity. Collaboration as a social interaction cannot be understood without considering culture, language, and 'particularistic characteristics'[7] that are embedded in a scientist's research activity. In all likelihood, researchers prefer to work with others who are fluent in their own language, and those who are not fluent in English are less likely to be solicited for collaboration. At the same time, scientists who are not fluent in English may be even more strongly motivated to collaborate than those who are (Bozeman & Corley, 2004). We expect that nationality factors will have a direct effect on collaboration and an indirect effect on productivity through collaboration.

Job Satisfaction

Few group behaviors are immune from influence by individuals' sense of personal esteem and job or life satisfaction. We expect that personal and job satisfaction, including satisfaction with pay and with colleagues' perceptions of research contributions, will relate to both collaboration and productivity. Babu & Singh (1998) identified 'work satisfaction' as one of the determinants of research productivity. But the relationship is surely a complicated one. Invitations to collaborate are an indication of the respect and esteem of colleagues, but they may also be motivated by a rational calculation of past contributions. Thus, any relation to productivity is necessarily a complicated one. We hypothesize that job satisfaction will be an important factor in collaboration, but will be more an effect than a cause of research productivity.

Perceived Discrimination

A special case of job dissatisfaction is perceived discrimination. A researcher who believes she or he is being discriminated against on the basis of sex, religion, or national origin (or probably any basis) is likely to be less active in seeking collaborators, at least local ones, and this may have negative effects on productivity. We hypothesize that scientists who perceive discrimination (that is, believe that they are being discriminated against) will be less productive and have fewer collaborators.

Collaboration Strategies

As we noted earlier, any calculation of the apparent costs and benefits of collaboration for productivity should consider the motive for collaboration. Someone collaborating as a mentor to inexperienced students may have a 'service' motive. On the other hand, someone collaborating on the basis of the other parties' reputation may have either a 'quality' motive or a 'social capital' motive. These are just a few of the possibilities. Some choose collaborators for their business-like demeanor, others just because collaborations are fun or entertaining. One of the most powerful predictors, of course, is proximity.

We hypothesize that the relationship between collaboration and productivity will be moderated by the researchers' strategies for collaboration: those seeking collaborators with complementary skills or strong scientific reputations will have the greatest productivity gains from collaboration and those seeking primarily to help students or junior colleagues will have fewer productivity gains.

Data and Method

Data and Sample

The data for this study were collected in three different stages. First, in 2000, we and other researchers associated with the Research Value Mapping (RVM) Program at Georgia Institute of Technology collected 1370 curricula vitae (CV) from a complete list[8] of university professors and researchers affiliated with National Science Foundation (NSF) or Department of Energy (DOE) research centers at US universities. The CV data include 3000 variables on demographic information, degree, job, publication, patent, professional affiliation, and grants.

Second, the *RVM Survey of Careers of Scientists and Engineers* was conducted in October 2001. A mailed questionnaire was sent to the 997 university faculty members from the RVM CV data (retired professors and one industrial researcher were deleted). We received 443 returns for a 44% response rate. The survey included questions about research collaboration, grants and contracts, job selection and work environment, and demographic information. The respondents included: engineering faculty (*n* 181, 41%); bioscience faculty (*n* 66, 15%); computer science faculty (*n* 25, 6%); chemistry faculty (*n* 47, 11%); physics faculty (*n* 43, 10%); and faculty from other science fields (*n* 57, 13%). Among the respondents are the following: tenured faculty (*n* 278, 63%); untenured faculty (*n* 165, 37%); men (*n* 383, 87%); women (*n* 58, 13%); US-born scientists (*n* 303, 68%); and foreign-born scientists (*n* 139, 32%). The average age of the respondents was 46 years in the year 2000. The sample had a larger proportion of foreign-born scientists, but a smaller proportion of women when compared with the national population of scientists (but the results for science and engineering centers were more representative).[9]

Third, after completing the survey, we regularly updated the respondents' affiliation information by consulting their institutional websites, and checking whether or not they stayed in the same center. Five respondents in the sample moved from one institution to another. This study excludes those respondents from further analysis. The most recent update on publications was performed in April 2004, and our analysis includes these updated data.

Measuring Research Productivity

We rely on two measures of publication productivity: a normal count and a fractional count of peer-reviewed journal papers for three years (2001 through 2003) of the post-survey period. The publication records of each respondent in the sample were traced back in *Science Citation Index Expanded (SCI-EXPANDED)* through the *ISI Web of Science*.[10] *SCI-EXPANDED* covers more than 3300 journals from more than 100 scientific disciplines. The authors were identified by matching the name, department, and institution from the CV-survey data and the SCI data. SCI provides the name, department, institution, and address of each co-author.

For the normal count,[11] all the peer-reviewed journal papers were counted for each respondent between 2001 and 2003. For the fractional count, each paper was divided by the number of co-authors. However, the data did not allow use of a weighted measure of publication, since the sample came from several disciplines rather than one specific discipline. Nor do the data permit quality comparisons among the journals or their impact ratings. We contemplated including quality-based indicators (such as citations and journal impact factors), but a pilot study indicated that doing so for the thousands of journal publications in our data set would be prohibitively expensive in time and resources.[12]

Major Variables

The Appendix shows the descriptive statistics for the variables in this study.

Descriptive Findings

Collaboration

Our questionnaire asked respondents to indicate the number of persons, by category, with whom they had engaged in 'research collaborations' within the past 12 months. The categories included male university faculty, male graduate students, male researchers who are not university faculty or students, female university faculty, female graduate students, and female researchers who are not university faculty or students.

While self-reported collaboration has some disadvantages in terms of the stability of the construct, we agree with Duque et al. (2005) that it has significant advantages over the conventional operationalization of collaboration as co-authorship. We focused on research collaborations rather than publications because: (1) we wished to include important collaborations that did not involve publication (including research that had not yet been submitted for publication); (2) we wished to exclude co-authors who achieved that status not by virtue of collaboration but because of position (for example, head of a laboratory or project). We focused only on the past 12 months because we expected that a limited time frame would both improve recall and reduce the response difficulty.

While we believe that using self-reported collaborations has many advantages, chiefly that it relies on the researchers' idea of a significant collaboration rather than an externally imposed concept, it is possible that the approach results in some degree of response bias in favor of social desirability. However, we also believe that such bias is likely to be limited, because the value of having many collaborators does not approach the social desirability of, say, having many publications. Moreover, our semi-structured interviews that accompanied the questionnaire suggest that university center researchers are not, in general, impressed with numbers of collaborators, at least not independent of numbers of publications (Bozeman & Boardman, 2003).

FIGURE 2
Total number of collaborators. Mean value 13.8, median value 12.0, valid *N* 360.

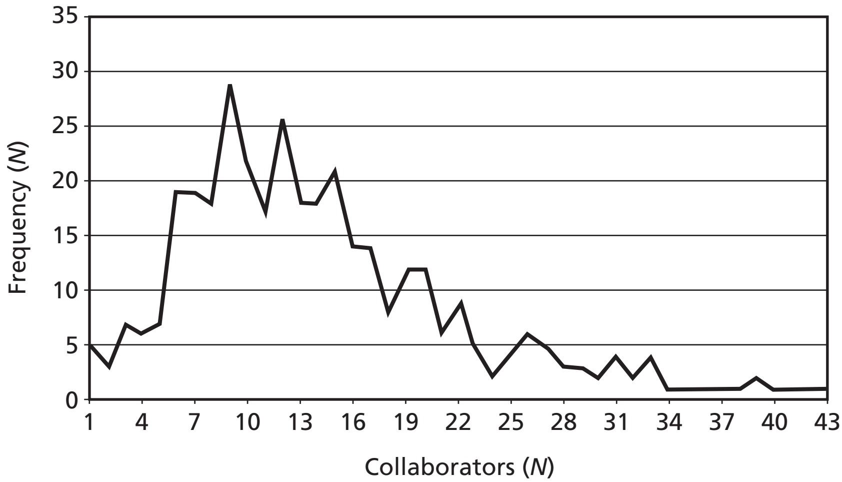


Figure 2 presents data for the total number of collaborations for a single year. Table 1 presents the distribution of numbers of collaborators, with the mean value being 13.8 and the median value being 12.0. Among this mean of 13.8 collaborators, the averages by category are 10.2 men, 3.6 women, 5.7 faculty, 5.9 graduate students, and 2.2 non-university scientists. In sum, the data seem to show active collaboration, chiefly with other academics, with men making up a much higher percentage of collaborators, and with faculty and graduate students having about equal likelihood of being chosen as collaborators.

One might expect considerable difference in disciplines' collaboration patterns. As shown in Table 1, engineering disciplines generally have more collaborators than other disciplines. Particularly, electrical engineers are the most active collaborators whereas biology/life sciences and physics researchers are well below the mean value. These variations could be interpreted, to some extent, by referring to differences between experimental scientists and theoretical scientists. The former tend to collaborate more than do the latter, since experimental scientists often use large and costly instrumentation that requires a large number of collaborators (Meadows & O'Connor, 1971; Gordon, 1980).

Another of our questionnaire items sought to determine the extent to which researchers collaborated with persons geographically close or faraway. We asked respondents to indicate the percentage of research time spent: working; alone; with researchers in the immediate work group or laboratory; with researchers in their own university but not in the immediate work group; with researchers in other US universities; with researchers in other nations' universities; or with researchers in industry and researchers in government laboratories. Table 2 shows that more than half (51.1%) of research time is spent with colleagues in the immediate

TABLE 1
Disciplinary difference in the number of collaborators

Field	Valid *N*	Mean	Median	SD
All	360	13.80	12	9.97
Chemical engineering	39	14.77	13	7.58
Civil engineering	14	11.21	12	6.33
Electrical engineering	41	17.34	14	13.83
Mechanical engineering	19	14.42	11	10.69
Other engineering	41	17.31	14	10.13
Biological/Life sciences	58	9.74	8	7.08
Computer sciences	18	14.78	15	3.80
Chemistry	39	14.95	13	14.89
Physics	35	12.74	10	8.95
Other natural sciences	39	13.28	11	7.14

work group, with the next largest amount of time (15.9%) devoted to working alone. Thus, researchers spent about one-third of their research time collaborating with persons outside their immediate work group, and they spent only about one-quarter of their time with those outside their university.[13]

We created a 'collaboration cosmopolitanism scale'[14] in order to indicate the extent to which researchers tended to be more or less 'cosmopolitan' (collaborating with those outside the proximate work environment). The scale ranged from 0 to 5 with 0 being the least cosmopolitan and 5 the most. There was no large variance among the disciplines with regard to cosmopolitanism. Physicists had the highest cosmopolitanism scale, chiefly because they were somewhat more likely to collaborate with researchers in

TABLE 2
Research time

Work setting	*N*	Mean percentage of research time	SD
Research time working alone	405	15.93	20.01
Research time working with researchers and graduate students in my immediate work group	405	51.10	23.85
Research time working with researchers in my university, but outside my immediate work group	405	11.44	12.66
Research time working with researchers who reside in nations other than the USA	405	5.11	7.73
Research time working with researchers in US universities other than my own	405	8.21	10.63
Research time working with researchers in US industry	405	5.23	7.94
Research time working with researchers in US government laboratories	405	2.98	6.53

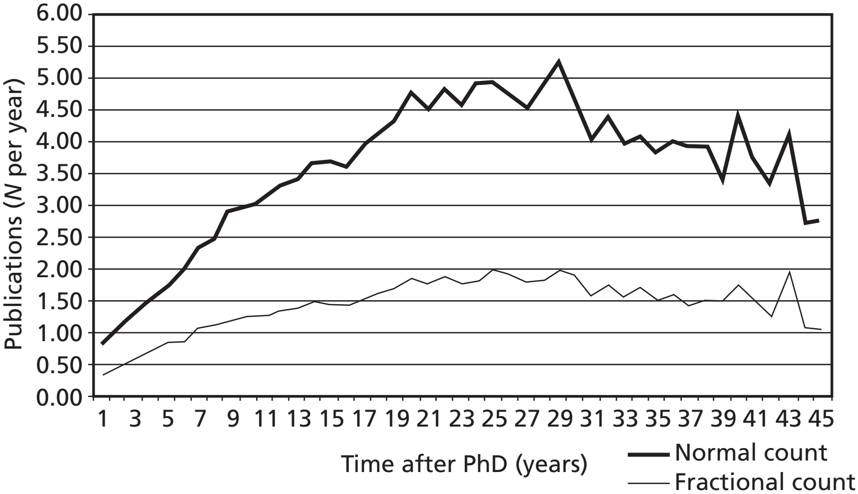
other nations. Mechanical engineers and biologists were less likely to be cosmopolitan. Among the various demographic variables, a few are associated with greater cosmopolitanism. Significant correlates include gender ($>.03$) (men score higher), tenure ($>.01$) (tenured faculty score higher), status as a principal investigator ($>.001$), and total dollar amount of current grants and contracts ($>.001$).

Productivity

In order to examine life cycle effects we included a description of career trends and annual average number of publications. Figure 3 shows the mean number of publications in the period after researchers received their doctoral degrees, with the lower line representing a fractional count and the upper line a normal count. A zero ('0') means that less than one year has passed since receiving the doctoral degree and a '44' means that 44 years have passed since the individual received the doctoral degree. Thus, the table gives insight into productivity levels during the course of a researcher's career.

The normal count data show that productivity peaks between years 23 and 28, averaging nearly five publications per year during that period (discerned from more detailed tables than provided here). After that period the researcher has four publications per year for about five years or so, and then the average drops to a little more than two per year after 40 years. Interestingly, the average is less than three publications for the first eight years, the time during which many researchers are struggling to be awarded tenure. Naturally there are some cohort effects after six to eight years due to 'drop outs' among persons who did not receive tenure: the zero to eight-year cohorts presumably include some people who will not receive

FIGURE 3
Career publication productivity



tenure and the cohorts after eight years probably include very few people who did not receive tenure.

With regard to the fractional count data, there are fewer peaks and the curve is somewhat smoother. The effect of using a fractional count is to make the data more closely to approximate a normal distribution, perhaps indicating that later years' productivity is related to the S&T human capital and the collaborative arrangements that develop. As before, the early years and the later years are less productive, but there is less of a sharp peak from years 8 to 40, though the most productive years appear to be from about 19 to 29 years after the doctoral degree.

There are considerable disciplinary differences in numbers of publications during the researchers' life course. Chemistry has the highest number of publications and computer science the lowest. The data also indicate that, whereas chemistry researchers peak between 28 and 30 years after the doctoral degree, physics researchers peak at 37 years after the degree, much later than one might expect. In terms of gender difference in the career productivity, in the normal count of productivity, men have a higher rate than women until the 18th year, at which time women have a somewhat higher productivity rate. The data must be treated with caution as the relatively small percentage of women (*n* 58, 13.1%) in the sample makes the trend data highly subject to individual cases and small cohorts.

Focusing on the productivity of the post-survey period, Table 3 shows that there are significant differences among the categories of rank, marital status, citizenship, and gender. A Tukey HSD test for rank indicates that, in terms of the normal count, full professors have significantly higher productivity than associate and assistant professors, while associate and assistant professors do not significantly differ. In the fractional count, a significant difference is found only between full professors and associate professors. The *t* tests provided in Table 3 show that married, foreign-born, and male scientists are more productive in terms of both fractional and normal count.

Findings: Collaboration's Effect on Productivity

We hypothesize that collaboration is positively related to productivity, measured by both normal counts and fractional counts of publications. The analysis deals with two fundamental questions:

- Even if collaboration is correlated with normal count productivity, is there a positive correlation when the number of co-authors is factored in (that is, fractional count)?
- If collaboration and productivity are correlated, does research collaboration affect researchers' publishing productivity, or is any observed relationship an artifact of co-variation with other factors? To put it another way, does the relationship hold with a properly specified model?

TABLE 3
Publication productivity by category

Category		NC	FC	Difference
Rank	Full professor	4.20	1.27	Tukey HSD test:
	Associate professor	2.62	0.80	NC: all the differences are significant ($p < 0.001$) except between associate and assistant professors.
	Assistant professor	2.98	1.00	FC: only the differences between full and associate professor are significant ($p < 0.001$).
Marital status	Married	3.23	0.99	NC: sig. ($p = 0.02$)
	Single	2.25	0.73	FC: sig. ($p = 0.08$)
Citizenship	US-born	2.84	0.87	NC: sig ($p < 0.001$)
	Foreign-born	3.83	1.21	FC: sig. ($p < 0.001$)
Gender	Male	3.27	1.01	NC: sig. ($p = 0.05$)
	Female	2.30	0.72	FC: sig. ($p = 0.05$)

Notes: NC, normal count; FC, fractional count; sig., significant.

We address this issue by first considering the relationship between collaboration and productivity, and then determining if the relationship vanishes with alternative specifications of the model.

Zero-Order Correlations for Collaboration and Productivity

Table 4 shows the zero-order correlations between collaboration and productivity. As one would expect, the correlation between normal count and fractional count productivity is quite strong (correlation coefficient .94). Since normal count is the numerator for the fractional count recoded variable, one would expect a strong correlation, but not necessarily one as strong as 0.9. The table also shows that there is a significant positive relationship between the number of collaborators and both normal count productivity (.26, $p < .001$) and fractional count (.22, $p < .006$). While these correlations seem not to explain a great deal of variation between collaboration and productivity, one would expect from the literature that collaboration would be only one factor determining productivity (Babu & Singh, 1998). To determine if collaboration has a real impact on productivity, we need to consider the moderating effect of several relevant factors.

Two-Stage Least Squares Results for the Collaboration Effect

As specified in the previous section, collaboration in our model is a mediating variable through which the real effect might be modified by some other factors such as individual characteristics and work environment variables. One of our concerns is to deal with any possible interaction between collaboration and productivity. In other words, if productivity inversely influences collaboration or if collaboration is correlated with the error term of productivity, the ordinary least squares (OLS) technique is

TABLE 4
Correlations of collaboration and productivity

	Collaboration	Productivity	
		Normal	Fractional
Collaboration	1		
Productivity			
Normal	0.26**	1	
Fractional	0.22**	0.94**	1

Note:
** Correlation is significant at the 0.001 level (two-tailed).

not appropriate and, perhaps, yields biased results. This study relies on a two-stage least squares (2SLS) test of the hypotheses. When using 2SLS, an appropriate instrumental variable should be identified with at least two conditions (Wooldridge, 2000): the instrumental variable should be correlated with the original variable (collaboration), and it should not be correlated with the error term of the dependent variable (productivity). We use the cosmopolitanism scale as the instrumental variable in the equation, since it meets the two conditions, and it is theoretically and practically relevant to collaboration (Glaser, 1964). It has a correlation of .30 ($p < .001$) with collaboration, but .00 ($p > .3$) with productivity (both normal and fractional) residuals. In the first stage, the number of collaborators was regressed on all the variables including the cosmopolitan scale. Then, the predicted value of collaboration was created. In the second stage, we regressed productivity on the predicted collaboration and all other variables. We also dealt with the identification problem when using 2SLS, which should meet two conditions such as order and rank conditions (Kline, 1998). First, our model satisfies the order condition, since the number of excluded variables is at least as large as the number of included endogenous variables. Second, the rank condition is also satisfied since the rank matrix (the rank is one in the model) is at least one less than the number of endogenous variables, such as collaboration and productivity.

As Table 5 shows, the age effect does not have significant effects on collaboration but moderates both measures of productivity. The effect is only significant at the alpha level of 0.1. We use career age as a single proxy variable for physical age, rank, and status, because using these variables together in the same equation would bring in a multi-collinearity problem. On average, the scientists who have had longer careers, mostly tenured faculty, are more productive in both normal and fractional productivity. This is not surprising, because the tenure process at most universities, in part, selects on the basis of research productivity. Furthermore, this is consistent with cumulative advantage theories, and with the idea that S&T human capital requires many years of development.

Since one of the chief purposes of research grants and contracts is to enhance collaboration and research productivity, we might expect a strong

TABLE 5
The results of two-stage least squares analysis

Dependent variable / Independent variable	First stage: Collaboration	Second stage: Productivity Normal	Second stage: Productivity Fractional
1. Age effect			
Career age	–.001 (–.02)	.007 (.11)*	.004 (.11)*
2. Grant effect			
Log of current grants	.17 (.11)**	.02 (.19)***	.02 (.23)***
Batting average		–.05 (–.02)	–.02 (–.01)
3. Family relations effect			
Gender	.90 (.03)	.09 (.05)	.04 (.04)
Marital status	–2.03 (–.07)	.04 (.02)	.02 (.01)
Spouse – job	.58 (.03)	–.02 (–.01)	.04 (.04)
Children	.03 (.01)	.05 (.08)	.02 (.05)
4. Citizenship effect			
Foreign-born	–.26 (–.01)	.26 (.17)***	.17 (.18)***
5. Job satisfaction effect			
Job satisfaction	1.84 (.10)**	.07 (.05)	.07 (.09)
6. Discrimination effect			
Discrimination	.39 (.04)	.06 (.08)	.03 (.08)
7. Collaboration strategies			
Taskmaster	.69 (.07)	–.04 (–.05)	–.02 (–.04)
Nationalist	–1.15 (–.13)**	.03 (.04)	.08 (.02)
Mentor	1.47 (.16)***	–.01 (–.01)	.01 (.01)
Follower	.67 (.07)	–.02 (–.04)	–.02 (–.04)
Buddy	–.13 (–.01)	.03 (.04)	.02 (.04)
Tactician	.56 (.06)	.08 (.11)***	.05 (.12)**
8. Field effect			
Basic vs applied	–1.57 (–.08)*	.14 (.09)**	.01 (.01)
9. Collaboration effect			
Number of collaborators		.03 (.14)**	.01 (.03)
Cosmopolitan scale	4.30 (.24)***		
R^2	.17	.17	.16
F	4.87	4.76	4.67
Significance	.00	.00	.00

Notes:
*$p < .10$, **$p < .05$, ***$p < .001$.
Values in parentheses are beta weights.
The unstandardized predicted values for the number of collaborators are used in the second stage of the two-stage least squares analysis.

relationship with collaboration and productivity. The result shows that the relationship is robust in both the collaboration and productivity equations. Due to being skewed with regard to the distribution of grants, the data were transformed into a natural logarithm. In the equation, grants have relatively stronger explanatory power, in terms of the beta weight measure, than other variables. This indicates that collaboration often depends on the resource availability provided via grants. In particular, grants show the strongest effect on productivity, with beta weights of .19 and .23. However, batting average (percentage of submissions funded) has no significant effect on research productivity.

In this model we examine not only gender itself, but also certain variables presumed to be closely related, such as number of children and whether the spouse is a full-time homemaker or family caregiver. The table shows that when we control for all other variables, the gender and family relations variables are not significant.

One might reasonably expect country of birth to interact with collaboration, and perhaps productivity as well. In terms of collaboration, being foreign-born does not have any significant coefficient in the presence of all other variables. But the sign of the coefficient indicates that US-born scientists have more collaborators (although the difference is not statistically significant). The low level of collaboration might be explained by cultural and language problems (DiTomaso et al., 1993; Choi, 1995). However, on average, foreign-born scientists are more productive than US-born scientists. In the sample, foreign-born scientists have a higher productivity by about 26% (in terms of a normal count of publications) and by about 17% (in terms of a fractional count). Like tenured researchers, foreign-born scientists have been, to some extent, selected for productivity. Had they not been more productive than average, their likelihood of remaining employed in the US as scientists and, thus, becoming naturalized citizens would be sharply reduced (Bauer et al., 2000).

The relationships between job satisfaction and productivity have been investigated in a wide variety of settings, ranging from factory workers to submarine crews to sports teams. The results of these studies vary greatly, with some finding that job satisfaction causes greater productivity, others that productivity causes satisfaction, and still others that there is no relation between the two. Some studies have examined the reciprocal effects between satisfaction and productivity while others have examined threshold effects. It is, perhaps, not surprising that the findings are so unstable. In addition to the usual problems of inconsistent measures, it seems plausible that highly varied work settings have an important bearing on the relationship between satisfaction and productivity. The preponderance of studies of job satisfaction and productivity in research and development organizations suggests a positive relationship between the two (Pfeffer & Langton, 1993). We created a composite index of job satisfaction by using three questionnaire items such as 'I am satisfied with my job', 'My colleagues in my department appreciate my research contributions', and 'I think I am paid about what I am worth in the academic market'. The internal reliability of these items is moderately dependable: the Cronbach alpha is .54. Our results show that job satisfaction seems to have little impact on scientists' productivity, at least holding the other variables constant. However, job satisfaction has a significant positive effect on collaboration. Sorting out the relationships among satisfaction, work environment, and productivity requires more research and, in all likelihood, more detailed and nuanced measures of satisfaction.

Regarding discrimination, we asked respondents about the extent to which perceived discrimination interacts with productivity and collaboration. We also used a composite index of discrimination by using two discrimination items: 'sex-based discrimination' and 'discrimination due to other reasons'. The Cronbach alpha (.66) of these items is relatively strong. The results indicate that perceived discrimination on the basis of sex, race, religion, and national origin does not have any moderating effect on collaboration and productivity. We also tested the interaction effect of gender and sex-based discrimination, and citizenship and nationality discrimination. But we did not find any significant effect on collaboration and productivity in the full model. It is worth noting that perceived discrimination does not appear to be strongly related to research productivity. It is probably even more important to note that the findings must be treated with care because there is relatively little variance in the two discrimination variables. Only 4.7% of the sample perceives that they are discriminated against by race, ethnicity, religion, or national origin. Only 5.4% of the respondents reported perceiving discrimination on the basis of sex (24 individuals: 14 women and 10 men).

Previous studies have shown that scientists' collaboration choices are based on coherent strategies and a wide variety of motives (Bozeman & Corley, 2004). Our concern in this model was to determine if the particular motives for collaboration were actually more important than the total number of collaborators for affecting research productivity. In this scheme, we first employed a factor analysis of 13 questionnaire items about collaboration motives. Through a varimax rotation, we identified six factors as shown in Table 6. Despite a relatively low Cronbach alpha, we have relatively high loadings for each factor. Factor scores were created and saved for the 2SLS analysis.

As shown in Table 5, mentor type of collaboration ('helping graduate students' and 'helping junior faculty') is significantly related to the number of collaborators. Likewise, nationalistic motivation in collaboration ('collaborator and respondent are of same nationality' and 'collaborator is fluent in respondent's language') is significantly related to the collaboration. This implies that those who pursue collaboration with persons of the same nationality and native language collaborate less. But the mentor and nationalist motives for collaboration do not have significant effects on productivity. By contrast, the tactician ('respondent and collaborator have complementary skills') has a strong impact on productivity. The finding recalls Melin's (2000) assertion about the pragmatic reasons for collaboration and their positive impact on productivity.

Field is one of the most important control variables in science studies. Different disciplines often have different research cultures and environments, which influence collaboration and productivity patterns. As an extreme example, astrophysicists are often involved with projects that include a great many collaborators, sometimes more than 100. Computer engineering is also distinctive, because peer-reviewed conference proceedings are often held to be more important than peer-reviewed journal

TABLE 6
Factor analysis of collaboration strategy

Factors . . . Items	Taskmaster	Nationalist	Mentor	Follower	Buddy	Tactician
Collaborator sticks to the schedule	.83					
Collaborator has strong work ethic	.82					
Collaborator and respondent are of same nationality		.87				
Collaborator is fluent in respondent's language		.80				
Collaborate to help junior colleagues			.83			
Collaborate to help graduate students			.80			
Collaborator has strong scientific reputation				.82		
Someone in administration requested the collaboration				.65		
Practices for assigning credit (for example, order of authorship)				.43		
Quality and value of my previous collaborations with the person					.67	
Length of time that respondent has known the person					.65	
Collaborator is fun or entertaining					.64	
Respondent and collaborator have complementary skills						.84
Internal reliability (Cronbach alpha)	.60	.57	.57	.42	.32	

Note: Rotation method: varimax.

publications. For simplicity's sake (and to preserve degrees of freedom), we divided the sample into 'basic' and 'applied' disciplines. Basic sciences include physics, chemistry, and biology, whereas applied sciences includes all the engineering fields. Although this is a rough indicator of field effect, we intend for it to moderate the collaboration effect on productivity. According to the results, applied sciences are generally more collaborative; that is, researchers in those fields have more collaborators, on average. However, basic sciences are more productive in the normal count, although there is no significant relationship with the fractional count.

The above analysis presents the context for understanding the impact of collaboration on publishing productivity. When this series of moderating variables is included, collaboration turns out to have a significant impact on normal count productivity ($p < .05$), but not on fractional count. We expected that collaboration might have a consistently positive effect on both normal and fractional productivities. The finding indicates that the simple number of publications significantly depends on the number of collaborators, but the *net* impacts of collaboration (as revealed by fractional count data) are less clear.

Like so many cases in the social sciences, the research outcome is rife with complexity. In some cases, collaboration has a positive effect on productivity; in other cases, it has little discernible effect on weighted publication productivity; and, in still others, it may even have a *suppressing* effect (for all the reasons discussed earlier, including transactions costs). Indeed, Duque et al. (2005) show that suppressing effects may be the most likely outcome in some cases in developing nations. Our own micro-examination of the CVs in our database suggests that there is considerable variance in the relationship between collaboration and productivity, but that the benefits of collaboration for productivity are apparently balanced by the disadvantages (thereby resulting in no strong statistical relationship between the two). This suggests that it is especially important to understand the factors influencing the effect of collaboration on productivity.

Discussion and Conclusion

Our research question is, on its face, a simple one: 'To what extent, if any, do scientists' collaborations affect their publishing productivity?' Our answers proved less simple. When publishing productivity is measured by 'normal count' (a scientist's total number of publications), collaboration is a strong predictor of publishing productivity. When publishing productivity is measured by 'fractional count' (dividing credit by the number of co-authors), collaboration and publishing productivity are not significantly related, at least not in a model controlling for moderating variables. These findings suggest the need for more extensive research on the impact of collaboration, in all its forms, on research productivity, in all its meanings.

Our 'simple question' is important for a number of reasons. As we stated in the introduction, there is a strong belief among policy-makers and apparently most scientists that scientific collaboration has positive effects on scientific productivity. The collaboration-as-synergy assumption affects not only particular research awards, but also entire programs of research policy. The assumption even affects the move toward interdisciplinary science centers, institutions built in part to promote collaboration, inter-disciplinary work and inter-sectoral cooperation (Behrens & Gray, 2000; Bozeman & Boardman, 2003).

A second reason for examining the impact of collaboration on scientific productivity is that there may be spillovers beyond the publication of papers. Elsewhere (Bozeman & Corley, 2004), we have argued that collaboration seems to be a major factor in promoting and transmitting 'scientific and technical human capital'. Our data show that senior faculty members are more likely to engage in mentor-oriented collaboration. Early career researchers may substantially benefit from these junior–senior collaborations, learning craft knowledge not easily transmitted by any other means.

Finally, examining the collaboration–productivity relationship is interesting, because anything more than the most superficial model requires us to examine a host of alternative explanations. Our focus, then, has been on examining collaboration's effects on productivity, but also on seeking to understand the way that a host of other potentially relevant factors affect the relationship. We believe this point is key. The impacts of collaboration on productivity are best viewed in terms of a contingency model.

Despite the interesting case presented by the fractional count analysis, one should not take lightly the strong and remarkably robust findings for the normal count analysis. The relationship is not simply an artifact of rank, gender, grants, or even the 'cosmopolitanism' of collaboration. It is important to remember that many studies simply could not be performed by individual scientists acting alone – not in a research environment in which science is increasingly interdisciplinary, equipment-dependent, and project-based.

The normal count case is important inasmuch as it shows that a research environment that demands increased collaboration does not undermine productivity. However, the fractional count case is important as a reminder that scientists and policy-makers need to have something more than a knee-jerk reaction to the presumed benefits of collaboration. The overall finding that collaboration is not significantly related to fractional count productivity masks a great deal of variance. Not all collaborations are created equal; some collaborations greatly enhance productivity, even by fractional count, whereas others inhibit it. The strategic question, one not fully addressed here, is 'under what circumstances can one expect that collaboration will be shown by either measure to be effective?'

Our focus is on the number of collaborators. We have no measures of quality of collaboration and it is certainly the case that not all collaborations are equally fruitful. The number of collaborators is not the same as the number of collaborations. If one collaborates ten times with one person, how does that differ, in productivity effects, from collaborating ten times with ten different persons?

Another question we pondered during our analysis was the nature of the collaboration-seeking dynamics. Our data are about collaborations, and we do not address (and neither does any other large-scale study of which we are aware) the dynamics of collaboration-seeking. Who seeks whom? For example, do collaboration dynamics mimic communication dynamics, with lower-status individuals seeking to collaborate with higher-status individuals? Or do such dynamics get overshadowed by mentoring motivations or, more mundanely, by the effects of proximity?

While we believe that the impact of collaboration on publishing productivity is an important research question, it is imperative that this single question should not monopolize collaboration studies. One must consider how collaboration affects the *composition* of research, not just the resulting productivity. It seems likely that at least some of the content of the work done collaboratively differs *because* it is done collaboratively. Is the work different? Is it better? Would the work have even been possible were it not collaborative?

One way to think of collaboration is in terms of the extent to which resources fit research needs. Presumably, the resources brought together in collaborations are different than those mobilized in solo research. But what is the value added and does that value offset the transactions costs? In many cases the crucial question is not enhanced productivity, but a more optimal fit among resources. One has skills one wishes to match with others, but not all skills are equal. Even in the mentoring-motivated collaboration, the complementary relationship seems to be maintained by helping the mentored researcher develop skills, and at the same time by sharing their new ideas. How are such relationships mediated? Our study suggests that most people collaborate with others in their immediate environment. This seems sensible in many ways. Such local collaborations may reduce transactions costs and may promote work that could not easily be done through more remote communications. And people are not drawn together entirely by accident. People work in the same setting because, among other factors, they share similar interests or at least compatible ones. But it seems highly unlikely that a system that is so strongly dictated by proximity (with more than half of collaborators being from one's work group) is the most efficient collaboration mechanism.

From a strategic standpoint, collaboration studies need greater emphasis on capacity, especially the development of S&T human capital. Even if one understands the relationship between collaboration and individual researchers' publishing productivity in all its richness and complexity, the health and well-being of scientific fields will continue to depend, critically, on the ability to replicate and extend research skills across generations. Collaboration may be the key element in S&T human capital development, and the implications of collaboration for career development are likely to be quite different than for discrete measures of individual publishing productivity.

Notes

The authors gratefully acknowledge the support of the National Science Foundation and the Office of Basic Energy Sciences, US Department of Energy. This work was performed as part of the project 'Assessing Economic and Social Impacts of Basic Research Sponsored by the Office of Basic Energy Sciences', under contract DE-FG02-96ER45562, and 'Assessing R&D Projects' Impacts on Scientific and Technical Human Capital Development' (SBR 98-18229). The opinions expressed in the paper are the authors' and do not necessarily reflect the views of the National Science Foundation or the Department of Energy. We are especially grateful to Juan Rogers, James Dietz, Jongwon Park, Min-wei Lin, Monica Gaughan, Ivan Chompalov, and three anonymous reviewers.

1. Our concern is with physical and natural scientists as well as engineers. For convenience, we use the term 'scientists' to encompass all.
2. In the USA, a series of technology transfer policies initiated in the 1980s (Bayh-Dole Act, Stevenson-Wydler Act, and Cooperative Research Act) enhanced interaction

among researchers throughout research and development organizations. In particular, some technology programs such as Advanced Technology Program (ATP) require inter-organizational collaboration for funding and research.

3. According to Gibbons et al. (1994), the new production of knowledge (mode 2) is not confined to a single discipline. Rather it is transdisciplinary, reaching across multiple disciplines. It involves the close interaction of many actors throughout the (more reflexive) process of knowledge production, resulting in a more socially accountable production of knowledge (p. vii).
4. Rogers & Bozeman (2001) define 'knowledge value collective' as the set of individuals who interact in the demand, production, technical evaluation, and application of scientific and technical knowledge.
5. Major studies include Maanten (1970), Meadows & O'Connor (1971), Crane (1972), Meadows (1974), Beaver & Rosen (1978), Goffman & Warren (1980), Heffner (1981), Fox & Faver (1984), Katz & Martin (1997), and Melin (2000).
6. According to National Science Foundation's *Science and Engineering Indicators 2004*, in 2001, the foreign-born scientists and engineers accounted for 37.6% of the doctorate holders in science and engineering, and 20.9% of the science and engineering faculty in US universities. In the areas of natural sciences and engineering, about 30% of the university faculties were foreign-born. In particular, among the engineering faculty, 35.5% were foreign-born. In the years between 1973 and 1999, foreign-born doctoral scientists' academic employment in the USA increased by more than four times from 13,531 to 73,268.
7. As the counternorm to universalism, particularism involves the consideration of 'functionally irrelevant characteristics' in the allocation of resources and rewards (Cole & Cole, 1973). Examples of particularism range from cronyism in the review of grant proposals; to racism and sexism in the hiring, tenure, and promotion processes; and to favoritism or personal opposition in allocating awards, honor, and research fellowships (Tang, 2000).
8. After identifying a total of 97 NSF and two DOE university research centers in 2000, the RVM team collected the list of scientists who were affiliated with the research centers. Fortunately, almost all the websites of the centers provided a list of the affiliated scientists with basic information such as email address, telephone number, affiliation, and mailing address. For a few centers, the RVM team contacted the director or managing scientist to obtain the list of affiliated scientists. The team finally collected basic information of 3814 affiliated scientists. The team requested CVs to all the 3814 scientists through emails, and received 1370 CVs.
9. Female scientists and foreign-born scientists account for 24.0% and 20.9% respectively, of the total academic doctorate researchers (except social sciences) in the USA (National Science Foundation, 2004).
10. *ISI Web of Science* is accessible through the URL <www.isinet.com>.
11. In cases of multi-authored publications, how much credit should be given to each co-author? There are three kinds of counting found in the research literature. First, 'straight count' gives credit to only the first authors. The great advantage of this procedure is that the exact number of papers is completely preserved. Cole & Cole (1973) claimed 'the omission of collaborative citations to papers on which the author was not the first among collaborators does not affect substantive conclusions.' They therefore recommend such omission. According to Lindsey (1980), this strategy has two results. First, it solves the problem of distributing credit for multiple-authored work by disregarding all but the first author who receives all the credit. Second, it greatly reduces the work required to collect data on any sample of scientists. Straight count assumes that the order of authors' names listed on a given paper reflects the level of their contributions, with the greatest contributor listed first, and so on in descending order. But a problem with a straight count is that it may discriminate against those scientists whose names appears late in an alphabetic listing. Rudd (1977) found a greater percentage of first authors among those with last names beginning with A to F

compared with G to M, and with G to M compared with N to Z (the percentage of first authors in the three groups were 56.8, 29.9, and 13.3 respectively).

The second counting method is 'fractional count', which gives data on fractional authorship. This approach has some of the same advantages as the straight count. Each item in a multiple-authored paper is divided by the number of authors and then summed to one (Pravdic & Oluic-Vukovic, 1986). Narin (1976) argued that there does not seem to be any reasonable way to deal with the attribution problem, except to attribute a fraction of a publication to each of the authors. Lindsey (1980) vindicated the advantage of fractional count, pointing out that it can control for the bias in overestimating production when the full value of a co-authored paper is awarded to all contributors. The main weakness, however, is that the procedure is tedious.

The third one is 'normal count', which is most frequently applied. It gives full credit to all contributors. In other words, it gives equal treatment to each author, which results in giving a full credit to each of them regardless of who happens to be the first or the last author. The correct number of papers in a given sample cannot be expressed by summing up the authors' data. The inaccessibility of the actual number of papers is the major drawback of the normal count procedure (Pravdic & Oluic-Vukovic, 1986). Another problem is that in most cases there is no reason to expect that co-authors contribute equally. Hagstrom (1965) found evidence that some publications listed authors for purely social reasons. More recently, LaFollette (1992) argued that the practice of making colleagues 'honorary co-authors' has become quite common.

12. For example, checking on the citations of any single paper often took as long as 40 minutes and rarely less than 5 minutes. There are more than 20,000 publications in our database. According to our rough calculation, including impact factors and citations would have required almost as much time as did the collection and coding of the original data.
13. We expect that these numbers are higher than the average for all university researchers because the respondents are affiliated with centers, many of which were established with a collaboration mandate or strategy.
14. The cosmopolitan scale is a measure of how close or far away a participant's collaborators are (that is, a participant with more collaborators in foreign countries would rank higher on the cosmopolitan scale than a participant with collaborators only in the USA). This is not, of course, a true physical distance scale since, for example, a collaborator in a foreign country may be closer than a collaborator in another part of the USA. The scale was calculated by multiplying the fraction of their time each participant spent working with a type of collaborator by the cosmopolitan rank of that variable (measured on a 0 to 5 scale). 'Research time spent working alone' is given a value of 0 on the cosmopolitan scale. Similarly, 'research time spent working with members of the same work group' is assigned a 1 and 'time spent working with others in the same university, but a different work group' is assigned a value of 2. 'Working with researchers at a different university' counts as a 3 on the cosmopolitan scale and 'working with others in industry or government laboratories' are both assigned a value of 4. Lastly, 'working with researchers in other nations' counts as a 5 on the cosmopolitan scale. For instance, if I work alone 10% of the time, within my own work group 20% of the time, with scholars at other universities 30% of the time, with industry 10% of the time, government 10% of the time and with scholars at other nations 20% of the time, my cosmopolitan score would be 2.9 [that is, 0.1(0) + 0.2(1) + 0.3(3) + 0.1(4) + 0.1(4) + 0.2(5)].

References

Astin, Helen S. (1978) 'Factors Affecting Women's Scholarly Productivity', in Helen S. Astin & Werner Z. Hirsch (eds), *The Higher Education of Women: Essays in Honor of Rosemary Park* (New York: Praeger): 139–57.

Babu, A. Ramesh & Y.P. Singh (1998) 'Determinants of Research Productivity', *Scientometrics* 43(3): 309–29.

Bauer, Thomas, Magnus Lofstrom & Klaus F. Zimmermann (2000) 'Immigration Policy, Assimilation of Immigrants and Natives' Sentiments towards Immigrants: Evidence from 12 OECD-Countries', Discussion Paper No. 187, August 2000. Bonn: The Institute for the Study of Labor (IZA).
Beaver, Donald (2001) 'Reflections on Scientific Collaboration (and its Study): Past, Present, and Future – Feature Report', *Scientometrics* 52(3): 365–77.
Beaver, D. & R. Rosen (1978) 'Studies in Scientific Collaboration: Part I – The Professional Origins of Scientific Co-authorship', *Scientometrics* 1(1): 65–84.
Beaver, D. & R. Rosen (1979a) 'Studies in Scientific Collaboration: Part II – Scientific Co-Authorship, Research Productivity and Visibility in the French Scientific Elite, 1799–1830', *Scientometrics* 1(3): 133–49.
Beaver, D. & R. Rosen (1979b) 'Studies in Scientific Collaboration: Part III – Professionalization and the Natural History of Modern Scientific Co-authorship', *Scientometrics* 1(3): 231–45.
Behrens, Teresa R. & Denis O. Gray (2000) 'Unintended Consequences of Cooperative Research: Impact of Industry Sponsorship on the Climate for Academic Freedom and Other Graduate Student Outcomes', *Research Policy* 30(2): 179–99.
Bellas, M.L. & R.K. Toutkoushian (1999) 'Faculty Time Allocation and Research Productivity: Gender, Race, and Family Effects', *Review of Higher Education* 22: 367–90.
Bozeman, Barry & Craig Boardman (2003) *Managing the New Multidiscipline, Multipurpose University Research Centers* (Washington, DC: IBM Endowment for the Business of Government).
Bozeman, Barry & Elizabeth Corley (2004) 'Scientists' Collaboration Strategies: Implications for Scientific and Technical Human Capital', *Research Policy* 33(4): 599–616.
Bozeman, Barry & Juan Rogers (2002) 'A Churn Model of Scientific Knowledge Value: Internet Researchers As a Knowledge Value Collective', *Research Policy* 31(5): 769–94.
Bozeman, Barry, James Dietz & Monica Gaughan (2001) 'Scientific and Technical Human Capital: An Alternative Model for Research Evaluation', *International Journal of Technology Management* 22(7): 636–55.
Choi, Hyaeweol (1995) *An International Scientific Community: Asian Scholars in the United States* (Westport, CT: Praeger).
Clemente, Frank (1973) 'Early Career Determinants of Research Productivity', *American Journal of Sociology* 79(2): 409–19.
Cole, Jonathan & Stephen Cole (1973) *Social Stratification in Science* (Chicago, IL: The University of Chicago Press).
Cole, Jonathan R. & Harriet Zuckerman (1984) 'The Productivity Puzzle: Persistence and Change in Patterns of Publication of Men and Women Scientists', in M.W. Steinkamp & M.L. Maehr (eds), *Advances in Motivation and Achievement* (Greenwich, CT: JAI): 217–56.
Cole, Stephen (1976) 'Age and Scientific Performance', *American Journal of Sociology* 84(4): 958–77.
Crane, Diana (1972) *Invisible Colleges: Diffusion of Knowledge in Scientific Communities* (Chicago, IL: University of Chicago Press).
DiTomaso, Nancy, George F. Farris & Rene Cordero (1993) 'Diversity in the Technical Work Force: Rethinking the Management of Scientists and Engineers', *Journal of Engineering and Technology Management* 10(1): 101–27.
Duque, Ricardo B., Marcus Ynalvez, R. Sooryamoorthy, Paul Mbatia, Dan-Bright Dzorgbo & Wesley Shrum (2005) 'The Collaboration Paradox: Scientific Productivity, the Internet, and Problems of Research in Developing Areas', *Social Studies of Science* 35(5): 755–85.
Fox, Mary F. & Catherine A. Faver (1984) 'Independence and Cooperation in Research: The Motivations and Costs of Collaboration', *Journal of Higher Education* 55(3): 347–59.

Fox, Mary F. & Catherine A. Faver (1985) 'Men, Women, and Publication Productivity: Patterns among Social Work Academics', *Sociological Quarterly* 26(4): 537–49.

Gaughan, Monica & Barry Bozeman (2002) 'Using Curriculum Vitae to Compare some Impacts of NSF Research Grants with Research Center Funding', *Research Evaluation* 11(1): 17–26.

Gibbons, Michael, Camille Limoges, Helga Nowotny, Simon Schwartzman, Peter Scott & Martin Trow (1994) *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies* (London & Thousand Oaks, CA: SAGE Publications).

Glaser, Barney G. (1964) *Organizational Scientists: Their Professional Careers* (New York: The Bobbs-Merrill Company, Inc).

Godin, Benoit (2003) 'The Impact of Research Grants on the Productivity and Quality of Scientific Research', INRS Working Paper 2003, National Sciences and Engineering Research Council of Canada (NSERC).

Godin, Benoit & Yves Gingras (2000) 'Impact of Collaborative Research on Academic Science', *Science and Public Policy* 27(1): 65–73.

Goffman, W. & K.S. Warren (1980) *Scientific Information Systems and the Principle of Selectivity* (New York: Praeger).

Gordon, Michael (1980) 'A Critical Reassessment of Inferred Relations Between Multiple Authorship, Scientific Collaboration, the Production of Papers and Their Acceptance for Publication', *Scientometrics* 2: 193–201.

Hagstrom, Warren. O. (1965) *The Scientific Community* (New York: Basic Books).

Heffner, A.G. (1981) 'Funded Research, Multiple Authorship, and Sub-authorship Collaboration in Four Disciplines', *Scientometrics* 3(1): 5–12.

Katz, Sylvan J. & Ben R. Martin (1997) 'What is Research Collaboration?', *Research Policy* 26(1): 1–18.

Kline, Rex B. (1998) *Principles and Practice of Structural Equation Modeling* (New York: The Guilford Press).

Kyvik, Svein & Mari Teigen (1996) 'Child Care, Research Collaboration, and Gender Differences in Scientific Productivity', *Science, Technology, & Human Values* 21(1): 54–71.

LaFollette, Marcel C. (1992) *Stealing into Print* (Berkeley, CA: University of California Press).

Landry, Rejean & Nabil Amara (1998) 'The Impact of Transaction Costs on the Institutional Structuration of Collaborative Academic Research', *Research Policy* 27(9): 901–13.

Lehman, Harvey C. (1953) *Age and Achievement* (Princeton, NJ: Princeton University Press).

Levin, Sharon & Paula Stephan (1991) 'Research Productivity over the Life Cycle: Evidence for Academic Scientists', *American Economic Review* 81(1): 114–32.

Lindsey, Duncan (1980) 'Production and Citation Measures in the Sociology of Science: The Problems of Multiple Authorship', *Social Studies of Science* 10(2): 145–62.

Long, J. Scott (1987) 'Problems and Prospects for Research on Sex Differences in the Scientific Career,' in L.S. Dix (ed.), *Women: Their Underrepresentation and Career Differentials in Science and Engineering* (Washington, DC: National Academy Press): 157–69.

Long, J. Scott (1992) 'Measure of Sex Differences in Scientific Productivity', *Social Forces* 71(1): 159–78.

Lotka, Alfred J. (1926) 'The Frequency Distribution of Scientific Productivity', *Journal of the Washington Academy of Science* 16: 317–23.

Maanten, A.A. (1970) 'Statistical Analysis of a Scientific Discipline: Palynology', *Earth Science Reviews* 6: 181–218.

Meadows, Arthur J. (1974) *Communication in Science* (London: Butterworths).

Meadows, A.J. & J.G. O'Connor (1971) 'Bibliographic Statistics as a Guide to Growth Points in Science', *Science Studies* 1(1): 95–99.

Melin, Goran (2000) 'Pragmatism and Self-organization: Research Collaboration on the Individual Level', *Research Policy* 29(1): 31–40.
Narin, Francis (1976) *Evaluative Bibliometrics: The Use of Publication and Citation Analysis in the Evaluation of Scientific Activity* (Cherry Hill: Computer Horizons).
National Science Foundation (2004) *Science and Engineering Indicators 2004* (Washington, DC: NSF).
Pao, M.L. (1982) 'Collaboration in Computational Musicology', *Journal of the American Society for Information Science* 33(1): 38–43.
Pelz, Donald C. & Frank M. Andrews (1966) *Scientists in Organizations: Productive Climate for Research and Development* (New York: John Wiley and Sons, Inc.).
Pfeffer, Jeffrey & Nancy Langton (1993) 'The Effect of Wage Dispersion on Satisfaction, Productivity, and Working Collaboratively: Evidence from College and University Faculty', *Administrative Science Quarterly* 38(2): 382–407.
Pravdic, Nevenka & Vesna Oluic-Vukovic (1986) 'Dual Approach to Multiple Authorship in the Study of Collaborator and Scientific Output Relationship', *Scientometrics* 10(5/6): 259–80.
Price, Derek J. de Solla & Donald Beaver (1966) 'Collaboration in an Invisible College', *American Psychologist* 21: 1011–18.
Rogers, Juan & Barry Bozeman (2001) 'Knowledge Value Alliances: An Alternative to R&D Project Evaluation', *Science, Technology, & Human Values* 26(1): 23–55.
Rudd, Ernest (1977) 'The Effect of Alphabetic Order of Author Listing on the Careers of Scientists', *Social Studies of Science* 7(2): 268–69.
Tang, Joyce (2000) *Doing Engineering: The Career Attainment and Mobility of Caucasian, Black, and Asian-American Engineers* (Lanham, MD: Rowman & Littlefield Publishers, Inc.).
Thorsteinsdottir, O. Halla (2000) 'External Research Collaboration in Two Small Science Systems', *Scientometrics* 49(1): 145–60.
Wanner, Richard A., Lionel S. Lewis & David I. Gregorio (1981) 'Research Productivity in Academia: A Comparative Study of the Sciences, Social Sciences, and Humanities', *Sociology of Education* 54(4): 238–53.
Wooldridge, Jeffrey M. (2000) *Introductory Econometrics: A Modern Approach* (Mason, OH: South-Western College Publishing).
Xie, Yu & Kimerlee A. Shauman (1998) 'Sex Differences in Research Productivity: New Evidence about an Old Puzzle', *American Sociological Review* 63(6): 847–70.
Zuckerman, Harriet (1967) 'Nobel Laureates in Science: Patterns of Productivity, Collaboration, and Authorship', *American Sociological Review* 32(3): 391–403.

Sooho Lee is a postdoctoral fellow in the School of Public Policy at Georgia Institute of Technology. He has recently worked on the issues of immigrant scientists in the USA. In addition to being interested in science and technology policy, he is currently involved in several public management research projects.

Address: School of Public Policy, Georgia Institute of Technology, 685 Cherry Street, DM Smith Building, Atlanta, GA 30332-0345, USA; fax +1 404 385 5104; email: sooho.lee@pubpolicy.gatech.edu

Barry Bozeman is Regents' Professor of Public Policy, School of Public Policy, Georgia Institute of Technology. His research focuses on science and technology policy.

Address: School of Public Policy, Georgia Institute of Technology, 685 Cherry Street, DM Smith Building, Atlanta, GA 30332-0345, USA; fax +1 404 385 5104; email barry.bozeman@pubpolicy.gatech.edu

APPENDIX
Variable description

Variables		*N*	Minimum	Maximum	Mean	SD
Rank	Tenured faculty (yes = 1)	438	.00	1.00	.63	.48
	Rank (3 = full professor, 2 = associate, 1 = assistant)	360	1.00	3.00	2.14	.88
Motivations for collaboration (4 = Very important, 1 = not important)	Time known person	438	1.00	4.00	2.76	.74
	Administrative request	435	1.00	4.00	2.12	.92
	Helping junior colleagues	433	1.00	4.00	2.86	.85
	Strong scientific reputation	438	1.00	4.00	3.27	.73
	Complementary skills	438	2.00	4.00	3.79	.45
	Quality other collaboration	438	1.00	4.00	3.73	.54
	Helping graduate students	438	1.00	4.00	3.15	.78
	Fun or entertaining	438	1.00	4.00	2.92	.80
	Fluent my language	438	1.00	4.00	2.40	.81
	Same nationality	438	1.00	4.00	1.77	.55
	Strong work ethic	438	1.00	4.00	3.46	.61
	Sticks to schedule	438	1.00	4.00	3.23	.60
	How assign credit	438	1.00	4.00	2.69	.78
Collaborators –	Total collaborators	360	.00	85.00	13.82	9.97
	Male faculty	360	.00	50.00	4.48	4.58
	Male graduate students	360	.00	26.00	4.07	3.76
	Male not university	360	.00	20.00	1.69	2.27
	Female faculty	360	.00	10.00	1.21	1.54
	Female graduate students	360	.00	12.00	1.84	1.78
	Female not university	360	.00	6.00	.54	.97
Collaboration scale cosmopolitan scale		405	.00	4.20	1.60	.58

APPENDIX Continued

Variables		*N*	Minimum	Maximum	Mean	SD
Grants	Principal investigator of a research grant or contract	436	.00	1.00	.89	.31
	Total US dollar amount of current grant or contract as principal investigator	339	6000	100000000	2261782.9	8109476.9
	Proposals awarded US proposals submitted (–)	379	.00	1.00	.56	.22
Job satisfaction (Strongly agree = 4 . . . Strongly disagree = 1) (Cronbach α .54)	I am satisfied with my job	435	1.00	4.00	3.31	.70
	My colleagues in this department appreciate my research contribution	435	1.00	4.00	3.16	.72
	I think I am paid about what I am worth in the academic market	435	1.00	4.00	2.69	.84
Discrimination (Strongly agree = 4 . . . Strongly disagree = 1) (Cronbach α.66)	At my current institution, I am discriminated against on the basis of my sex	434	1.00	4.00	1.23	.57
	At my current institution, I am discriminated against on the basis of my race, ethnicity, religion, or national origin	433	1.00	4.00	1.20	.52
Individual characteristics	Gender (male = 1, female = 0)	438	.00	1.00	.87	.34
	Marital status (married = 1)	438	.00	1.00	.90	.30
	Spouse is full-time homemaker or family caregiver (yes = 1)	438	.00	1.00	.28	.45
	Year born	418	1926	1974	195	10.00
	Foreign-born (yes = 1)	438	0	1	.31	.46
Field	Basic (basic = 1)	438	0	1	.36	.48
Publishing productivity	Normal count	430	0	23	3.14	3.44
	Fractional count	430	0	7.84	.97	1.03