Final Report

"The Knowledge Creation Process in American Science: Basic Findings from the American Scientists Survey"

John P. Walsh, PI
School of Public Policy
Georgia Institute of Technology
Hsin-I Huang
School of Public Policy
Georgia Institute of Technology
Yeonji No
School of Public Policy
Georgia Institute of Technology

In Collaboration With
Sadao Nagaoka (Project Leader, JSPS Project)
Hitotsubashi University

Masatsura Igami
NISTEP

Final Report to Hitotsubashi University, Sadao Nagaoka, project leader for the research on the Knowledge Production Process in Science, sponsored by the Japan Society for the Promotion of Science.
March 15, 2011

Table of Contents

SUMMARY

This paper reports the initial findings from a survey of American scientists on the knowledge creation process in science. Topics surveyed include: Motivation and the other basic characteristics of the research project that yielded the paper, such as measures of Stokes quadrants, serendipity, and scientific competition; the knowledge production process, such as the respondent's roles in the project and uses of external knowledge and geographic location of those knowledge sources, and organization of research project; research inputs, including project duration, funding and sources of funds; composition of the research team, by rank, organization type, field, country of origin, research skill/specialty gender, and number of students trained and personnel hired by project as well as authorship rules; the outputs of the project, such as number of other papers, patents, licenses, startup firms; and scientists' demographics, family status (marital status, children), education and training, mobility, awards, and publication counts. The sample included an oversampling of the top 1% of most highly cited papers in each field (H papers, hereafter), and a random sample of all papers (excluding the top 1%, hereafter referred to as N papers). The survey covered all fields covered by Thompson/ISI's Web of Science, including clinical medicine and social sciences, stratified by field of science as well as by H v. N papers. The survey was conducted by a research team at Georgia Institute of Technology, School of Public Policy, in collaboration with Sadao Nagaoka of Hitotsubashi University and Masatsura Igami of NISTEP. The project was funded by a research contract from Hitotsubashi University, from funding from the Japan Society for the Promotion of Science.

Major Findings are as follows:

- About 80% of publications were by researchers in universities, and another 10% were in government labs. The distribution is fairly similar for H and N projects. The share of government research institutes is substantially higher in material science and environment/ecology/geosciences. Private firms are most prevalent among highly cited engineering papers, representing 12% of the total.
- About half of respondents were 45 or older when they submitted the focal paper. The average age was lowers in computer science/mathematics.
- About 30% of researchers had stayed abroad for one year or more for study or research, and about a third had changed jobs in the five years before the project. Mobility is lowest among those in government organizations and hospitals.
- The goal of the research fits in Pasteur's quadrant (emphasizing both advancing knowledge and addressing practical problems) in 32% of H and 25% of N projects. Yet, Bohr's quadrant (emphasizing advancing knowledge, without addressing practical problems) is the modal goal, representing 47% of those for H projects and 43% of those for N projects.
- Serendipitious findings were quite common. Furthermore, the share of serendipitous outcomes was higher for the H projects (48% of projects) than for the N projects (43% of projects). For H projects, computer & mathematics and agriculture science/plant and animal science were fields that were most likely to produce serendipitous output.
- Respondents were aware of scientific competition, with H projects facing greater competition and more concerned with competition. Priority loss is a big concern in fields such as computer science/mathematics, material science and basic life science.
- Published literature is the most important knowledge source (about half of respondents rated published literature as 5 out of 5). Other important knowledge sources include: researchers with different research skills, colleagues in their organization and past collaborators. The importance of different knowledge sources are relatively similar between H and N projects.

- The most commonly used project management strategy was setting ambitious project goals. This was also the project management characteristic that most distinguished H (80%) from N (66%) projects. The second most used was developing a research community outside the lab. This also had a substantial gap between H (59%) and N (50%) papers.
- The number of authors varies substantially across fields. Clinical medicine/psychiatry/psychology and basic life science have the largest median team size. On the other hand, computer science/mathematics and social science have fewer authors. Also, the number of the authors is greater for H projects than N projects in most fields.
- Junior researchers (PhD students and post-docs) make a disproportionate contribution to scientific research, as measured by their likelihood of appearing as first author when authorship reflects contribution. Fully half of first authors for life science papers from universities are post-docs.
- The time lags from beginning the project to submission of the paper is shorter for H papers than N papers. For example, in chemistry, the average time lag for H papers is 2 years but the time lag for N papers is almost 4 years.
- Project funding tends to come from multiple sources. University research is heavily dependent on outside funding, and that is especially true for the H projects. In contrast, government labs and industry researchers are much more likely to use only intramural funds.
- NIH and NSF are the dominant outside sources for university researchers
- DOD and DOE are major funders in some fields. Department of Energy is a major funder for H projects in physics (19%), chemistry (15%) and material science (14%). Department of Defense funds many H projects in material science (43%), engineering (34%) and chemistry (26%).
- About 10% of university researchers get some industry funding.
- We find that 7 percent of the N projects and 14% of the H projects resulted in at least one patent application. For the highly cited papers, we find high rates of patenting (around 30%) in chemistry, materials science and engineering, consistent with prior work. Life science related fields have moderate rates of patenting, with about 15% of projects resulting in a patent. Physical sciences, mathematics and computing and geosciences all have rates of patenting under 10% of project.
- We find that 4 percent of the N projects and 8% of the H projects resulted in at least one license. The majority of these licenses include providing know how. Only 1 percent of the N projects and 4% of the H projects resulted in a startup.
- Two-thirds of H projects including a foreign-born post-doctoral fellow. Projects are also more likely to have foreign-born PhD students than US born. However, US born master's students are more common than are foreign born.
- Over 40% of projects hired people specifically for the project, which suggests that another important impact of scientific research is job creation.

1. BACKGROUND AND PURPOSE OF THE RESEARCH

Scientific work is increasingly collaborative (Wuchty, Jones, and Uzzi 2007a). And, there is evidence that collaborative research has a greater impact on a field (Jones, Wuchty, and Uzzi 2008; Wuchty, Jones, and Uzzi 2007a). However, there is still substantial debate about the reasons for such impact (Valderas, Bentley, Buckley, and Wray 2007; Wuchty, Jones, and Uzzi 2007b). In part this controversy is due to a lack of large scale data on the details of scientific collaborations. Most work until now has dealt with modest sample sizes (Cummings and Kiesler 2005; Shrum, Chompalov, and Genuth 2001; Walsh and Maloney 2007; Walsh and Maloney 2002), or with large scale aggregate data on papers and their authors and institutions but with little data on the details of the projects that produced these collaborations (Jones, Wuchty, and Uzzi 2008; Wuchty, Jones, and Uzzi 2007a). One simple hypothesis is that more people produce better papers. However, a more nuanced hypothesis is that the composition of the collaboration (in terms of fields, senior/junior, institutions, and skills) is key to understanding the success and the impact of the collaboration. In particular, heterogeneous collaborations may have significant advantages (Shrum, Chompalov, and Genuth 2001; Walsh and Maloney 2002). On the other hand, such collaborations may suffer from problems of coordination and conflicting cultures (for example, difficulties talking cross disciplinary boundaries) (Cummings and Kiesler 2005; Walsh and Maloney 2007).

Similarly, commercialization of scientific output is becoming increasingly important (Roessner, Bond, Okubo, and Planting 2009; Rothaermel, Agung, and Jiang 2007). However, the project-level determinants of the translation of scientific research into commercial outputs are still not well understood. For example, do multi-institution collaborations benefit from access to broader networks of organizations that might be able to commercialize their results, or are they hampered by transaction costs and differing expectations about commercialization? How does commercialization vary across fields, and do inter-disciplinary collaborations benefit from greater visibility and more commercialization possibilities?

Policy makers, who view collaboration as a means of increasing research quality, have pushed for increased collaboration, especially of an inter-disciplinary nature, in both the US and Japan. However, the impact of inter-disciplinary collaboration is also poorly understood. In addition to the growing interest in collaboration generally, there is also a strong policy focus on new forms of collaboration and new forms of science. For example, Gibbons, et al., argue that contemporary science is characterized by a shift to Mode 2 science, which emphasizes inter-disciplinarity, cross-organizational and cross-sectoral collaboration, as well as greater engagement with societal issues. (Gibbons, Limoges, Nowotny, Schwartzman, Scott, and Trow 1994). Stokes (1997) also argues for the growing importance of science that both generates new knowledge and addresses practical concerns--what he calls Pasteur's Quadrant. Etzkowitz and Leydesdorff (2000) emphasizes cross-sectoral collaboration when they argue that we have entered the age of the triple helix of science that integrates universities, government labs and industry. At the same time, these authors also note the importance of Mode 1 or Bohr's Quadrant, where research is disciplinary-based and focused on acquiring fundamental understanding.

These typologies raise various questions, such as the composition of science, the extent to which it is integrating across fields, organizations, sectors and goals, and how this composition varies by research area and by organization type (for example, elite universities versus second-tier universities). Furthermore, how do these modes of science relate to traditional measures of research performance (e.g., citations)?

There is also growing interest in issues related to the scientific labor force. In particular, there is concern about how to more fully incorporate women and underrepresented minorities into science (COSEPUP 2007; MIT 1999). At the same time, we see significant differences in the participation of women across disciplines. While we know the basic demographic, such as the percent of women in different fields (National Science Board 2008) and how this varies across countries (OECD 2005), we know substantially less about the roles women play in scientific teams (Fox and Mohapatra 2007). In addition, there is growing interest in gender differences in scientists' participation in translating science into commercial application (Ding, Murray, and Stuart 2006). In order to further our understanding of the composition of research teams and the roles that women play in different fields in the U.S. and Japan, we will examine the potential for collecting detailed data on the demographics of scientific teams and analyze the composition of teams in different settings, as well as the relations between demographics (rank, gender) and performance in different contexts.

Finally, as science becomes increasingly internationalized, there is a strong need to understand research teams in a global context and for major research centers such as the US and Japan to learn from the experiences in the other country in order to develop best practices for guiding scientific collaboration. These comparisons are especially useful as both countries are facing strong challenges not only from each other, but from scientists in Korea, China and other Asian countries as well (National Science Board 2006). This comparison can be especially informative since, while both countries have well-developed science systems, they are organized quite differently in terms of funding, staffing, university-industry linkages, technology transfer systems, etc. (Kneller 1999). For example, in Japan, the share of institutional/internal funding tends to be greater, while in the US, the share of external, project-based funding tends to be greater (Kneller 1999). Also, there is much more circulation in the S&E labor market in the US than in Japan (Walsh and Nagaoka 2009). In addition, institutions to encouraging licensing of university technology are more recent in Japan than in the US, although informal links between universities and firms were strong before the recent policy reforms (Hicks 1993; Kneller 1999; Walsh and Saegusa 2003). One important linkage in Japan is the dispatch of young company engineers to the university for engaging in a joint research with a university professor. Differences in the patents systems in the two countries (such as the longer grace period in the US, as well as the availability of continuations in part) may also have significant effects on technology transfer from university to firms in each country (Hegde, Mowery and Graham 2007). We also see that women are significantly less represented among Japanese researchers, although recent policy changes in Japan have tried to increase the participation of women (Normile 2006).

Thus, both countries' science systems may benefit from detailed comparisons, controlling for field, of the structure of scientific collaborations and its relations to scientific and commercial performance. Both countries are currently debating their future science policies and how to simultaneously support more basic research and encourage the translation of that research into commercial innovations. Similarly, both countries are exploring programs that will increase the participation of women in science and, in the case of the US, underrepresented minorities. At the same time, both countries are taking a hard look at the productivity of their science systems and asking for evidence of performance in the face of increasingly tight fiscal constraints. Our project is a first step in exploring these various institutional differences and their implications for the organization of scientific work and the outcomes of scientific research.

The survey was designed to address the following basic questions:

1. What is the relations between, for example, team composition and performance of the project (Fox and Mohapatra 2007)?,
2. Are there gender differences in commercialization (Ding, Murray, and Stuart 2006)?,
3. Which types of science lead to university-based startups (Roessner, Bond, Okubo, and Planting 2009)?
4. What are the contribution to workforce development of different Federal funding sources?
5. What is the distribution within and across fields and countries of science projects by Stokes' quadrants (Bohr's, Pasteur's) and by Mode 1 and Mode 2 (Gibbons et al. 1994; Stokes 1997)?

6. What are the rates of serendipitious findings in different contexts (for example, by country, by funding source and by composition of the research team, controlling for field and other project characteristics)?
7. Is serendipity associated with more scientifically important and/or commercially successful research findings?

This report provides basic descriptive statistics on the data collected to answer these questions. These data will provide the foundation for a variety of future analytic papers that address these questions in detail. The following sections of the report give the details of the data collection and then summarize the results organized by theme.

2. OVERVIEW OF THE SURVEY METHOD

2-1 Generating a sample of scientific research projects

To address these and other questions, we conducted a large-scale survey of US scientists. As the first stage of this project, funded by Hitotsubashi University, we developed a sample of research papers, in collaboration with the Japanese research team. We began with sample of over 9000 journal articles with at least one author affiliated with an institution in the US, collected from Thompson/ISI Web of Science. This sample was stratified by field (22 strata) and by highly cited (top 1%) papers (H papers, representing 1/3 of the total) and normal papers (N papers, representing 2/3 of the total). In determining the H papers, we used the number of forward citations to a paper as of December 31, 2006.

For each strata for each year, the papers with the top 1% of citation counts was selected. In addition, a random sample of about twice that size was selected. This resulted in a list of 9428 papers, across 22 strata, which we then aggregated to 11 strata for presentation purposes (Table 1). Clinical medicine was the largest strata, with over 1900 papers in total (over 600 in the top 1% and over 1200 in the random sample). Materials science was the smallest strata, with about 100 in the top 1% and about 200 in the random sample. This sampling strategy gives us sufficient cases to be able to estimate field specific estimates for both the top and random groups, even if we get only a modest (25%) response rate.

The survey was targeted to a single author on the paper. Thus, we wanted to address the survey to an author that we feel is most likely to be able to respond to detailed questions on the organization and outcomes of the project. We will describe this person as the primary author. In addition, we needed up to date contact information (email and postal address) for this research. In order to determine the primary author, we followed the following protocol:

Check author's full name and institution:

We begin by searching the record in the Web of Science, and through the Web of Science, the original publication cover page. These sources will give us the number of authors and institutions on the paper, and will also often give the full name of the primary author and the email address.

From this list of authors, we need to determine the primary author. If one is available, we use the reprint author. If no eligible reprint author is available, then we use either the first author, or the last author, depending on the field. Based on prior work by the Japanese research team, we found the following fields tend to put the primary author last: Biology and biochemistry, Chemistry, Immunology, Microbiology, Molecular biology & Genomics, Multidisciplinary. Based on the survey data, we were able to confirm this result. For these fields, we take the last author as primary. For other fields, we take the first author as primary.

If the first choice for primary author is unavailable (either because he is not in the US, or he is deceased, or we cannot find current contact information for him), then we go to the second choice primary author, which would be the last/first author. If this author is also ineligible, then we go to the first/last but one, but two, etc. until we find an eligible author or we exhaust the list (at which point the paper becomes defined as out of the population).

Verifying Author Address

Once we have determined a primary author, we do a web search to find out if the author is still at the institution listed on the paper. If he is no longer there, then we use name and field information to find the current affiliation. The current address (email and post address) is recorded in the file. If an author becomes ineligible (e.g., he has moved overseas), we go back to the original list and find the next author based on the priority rule above.

Using this protocol, we were able to get contact information (either email, post-mail or both), for over 95% of the cases in the sample. To reduce respondent burden, we randomly sampled one paper from the set of papers by an author, giving priority to High papers. After removing these duplicates, we end up with 2912 High papers and 6011 Normal papers, for a total of 8923 papers in the final mail out sample. Some of these papers included authors who were no longer available (deceased, retired, ill, out of the country, etc.). We also found additional duplicate cases during the survey administration. When possible, we tried to contact another author, but in some cases, these papers had to be declared out of the population. In addition, some papers were not research papers (either review papers, letters to the editor, or similar non-research papers). These were also excluded from the population. Thus, the final response rate should be adjusted to account for these ineligible cases.

The report describes the research projects associated with highly cited papers as “H projects” and those from normal papers as “N projects”. The project is defined as the series of research activities that produced the focal paper and any closely related papers.

2-2 Implementation of the Survey

The survey questionnaire was conducted on the web. Each sampled author was sent a personalized email message explaining the survey and asking the respondent to complete the questionnaire. The email message included the URL of the survey, as well as an individual password token for logging on to the web site. The web survey included customized information on the focal paper (authors, title, journal and publication year, as well as the co-authors for the collaboration questions). In order to reduce the burden on the servers, the mail-outs were spread out over several weeks.

The time-line of the survey is as follows:
Initial mail-outs: September-November, 2010
Reminder emails: November-December, 2010
Second (final) reminders: January, 2011

2-3 Field classification for presentation and analysis

The original sample consisted of 22 ESI journal field strata. However, because of small Ns in some fields and to ease presentation, these were aggregated into 10 fields. For some purposes, these can be further aggregated into three broad fields: physical sciences, medicine and life sciences (with social sciences excluded). The relations among the 22 journal fields, the 10 analysis fields and the 3 broad fields are shown in Exhibit 1.

Exhibit 1. Relation between the 22 ESI Journal Fields, 10 Aggregate Fields and 3 Broad Fields.

22 ESI Journal fields	10 fields	Broad fields
Chemistry	Chemistry	Physical Science
Materials Science	Material Science	
Physics	Physics & Space Science	
Space Science		
Computer Science	Computer & Mathematics	
Mathematics		
Engineering	Engineering	
Environment/Ecology	Environmental/Ecology/ Geoscience	
Geosciences		
Clinical Medicine	Clinical Medicine/Psychology	Medicine
Psychiatry/Psychology		
Agricultural Sciences	Agricultural Science/Plant Animal Science	Life Science
Plant & Animal Science		
Biology & Biochemistry	Basic Life Science	
Microbiology		
Molecular Biology & Genetics		
Neuroscience & Behavior		
Pharmacology & Toxicology		
Immunology		
Multidisciplinary	Either 22 ESI journal fields were assigned based on the analysis of backward citation	Either 22 ESI journal fields were assigned based on the analysis of backward citation
Economics & Business	Social Science	Social Science
Social Sciences, general		

2-4 Sector Classification for the Affiliated Scientists

The survey asked respondents to identify the sector of the organization they were affiliated with at the time the focal paper was submitted for publication. The types of organizational affiliations were University (including joint university research institutions and higher technical schools), Public Research Organization (including national testing/research institutions, independent administrative institutions, special corporations and testing/research institutions of local governments), Private Firm, Private Non-Profit Organization, Hospital and Other. We report the sectoral distribution of the respondents below.

2-5-1 Response rate, overall and by field

Out of 8,856 (excluding duplicates) survey targets, we received 2,329 responses (as of Jan 28, 2011). The total response rate is 26.3% using total mail out as the denominator; the adjusted response rate is 28.1% excluding ineligible cases (i.e., out of scope papers, overseas authors only and deceased cases, see Appendix 1 for details). If we limit the response rate estimate to those cases that were actually contacted (i.e., also excluding those that were returned undeliverable or for whom we could not find any valid contact address), the adjusted response rate is 30.3%. We also received many partial responses (875 people completed part of the survey). However, for this report, we only report completed cases. The unadjusted response rate is 27.7% for H papers and 25.6% for N papers. Response rates by field are shown in Exhibit 2. While ranking the response rates by fields, fields over 30% response rate are environment/ecology & geosciences (37%), social science (32.4%) and agriculture, plant, and animal science (31%). The response rate in clinical medicine & psychiatry/psychology is the lowest (21%) among the 10 fields excluding the multidisciplinary[1] field. The response rate of H papers seems to be higher than or at least equal to the response rate in N papers in almost all fields, the exception is the computer science & mathematics (3% higher for the N papers).

Exhibit 2. Response Rate by Field.

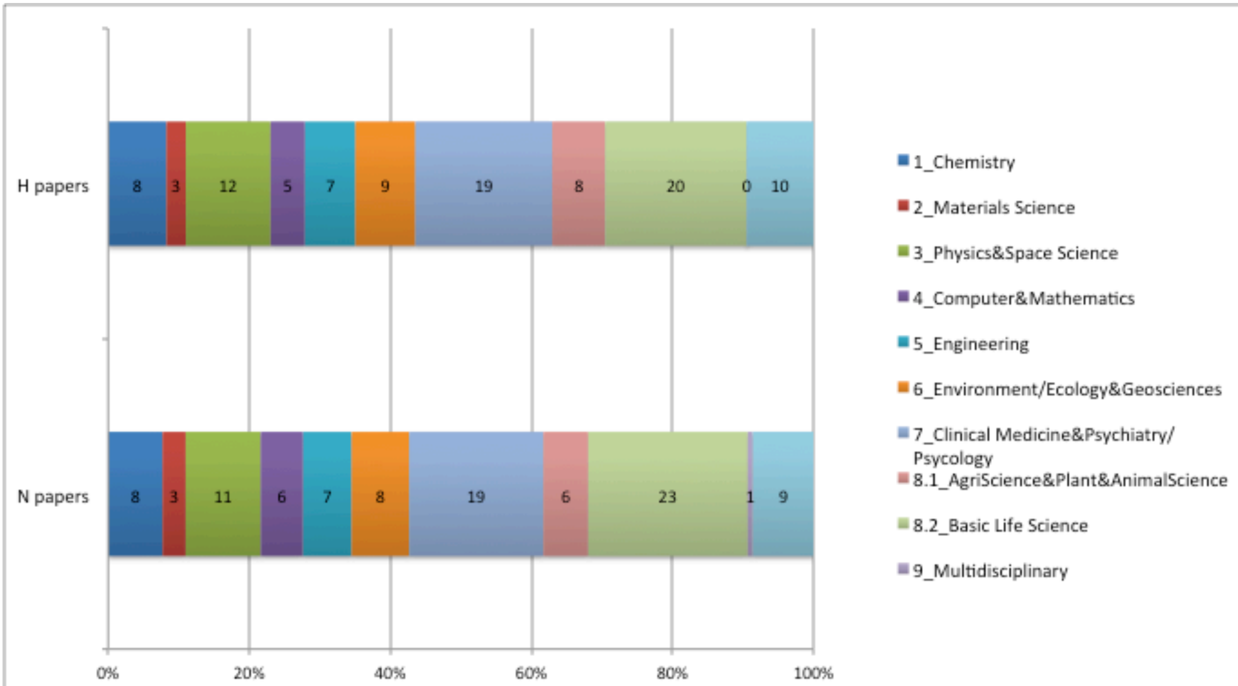
	All			H papers			N papers			
Fields	Survey targets	Responded	Response rate	Survey target	Responded	Response rate (A)	Survey target	Responded	Response rate (B)	(A) - (B)
1_Chemistry	663	184	27.8%	204	66	32.4%	459	118	25.7%	6.6
2_Materials science	261	72	27.6%	82	22	26.8%	179	50	27.9%	-1.1
3_Physics & Space science	993	259	26.1%	347	96	27.7%	646	163	25.2%	2.4
4_Computer Science & Mathematics	508	131	25.8%	165	39	23.6%	343	92	26.8%	-3.2
5_Engineering	571	162	28.4%	186	57	30.6%	385	105	27.3%	3.4
6_Environment/Ecology&Geosciences	522	193	37.0%	183	68	37.2%	339	125	36.9%	0.3
7_Clinical Medicine/Psychiatry /Psychology	2165	446	20.6%	718	155	21.6%	1447	290	20.0%	1.6
8.1_Agriculture Science & Plant & Animal Science	508	157	30.9%	181	60	33.1%	327	97	29.7%	3.5
8.2_Basic life science	1954	506	25.9%	602	159	26.6%	1352	348	25.7%	0.9
9_Multidisciplinary	78	11	14.1%	2	0	0.0%	76	11	14.5%	-
10_Social Science	641	208	32.4%	212	76	35.8%	429	132	30.8%	5.1
All	8864	2329	26.3%	2882	798	27.7%	5982	1531	25.6%	2.1

[1] In this survey, for the papers in the multidisciplinary field, (i.e., those papers published in the journals like Nature and Science) we re-reclassified into one of the 10 aggregate fields based on the references in the papers. However, there remains eleven papers could not be reclassified.

2-5-2 Distribution in fields of the respondents by two types of the papers

Exhibit 3 represents the distribution in fields between H papers and N papers. The distributions are almost identical between H papers and N papers. The only exception is in the basic life sciences, where the share of N paper is 3% more than that of the H papers.

Exhibit 3 Field compositions of the respondents



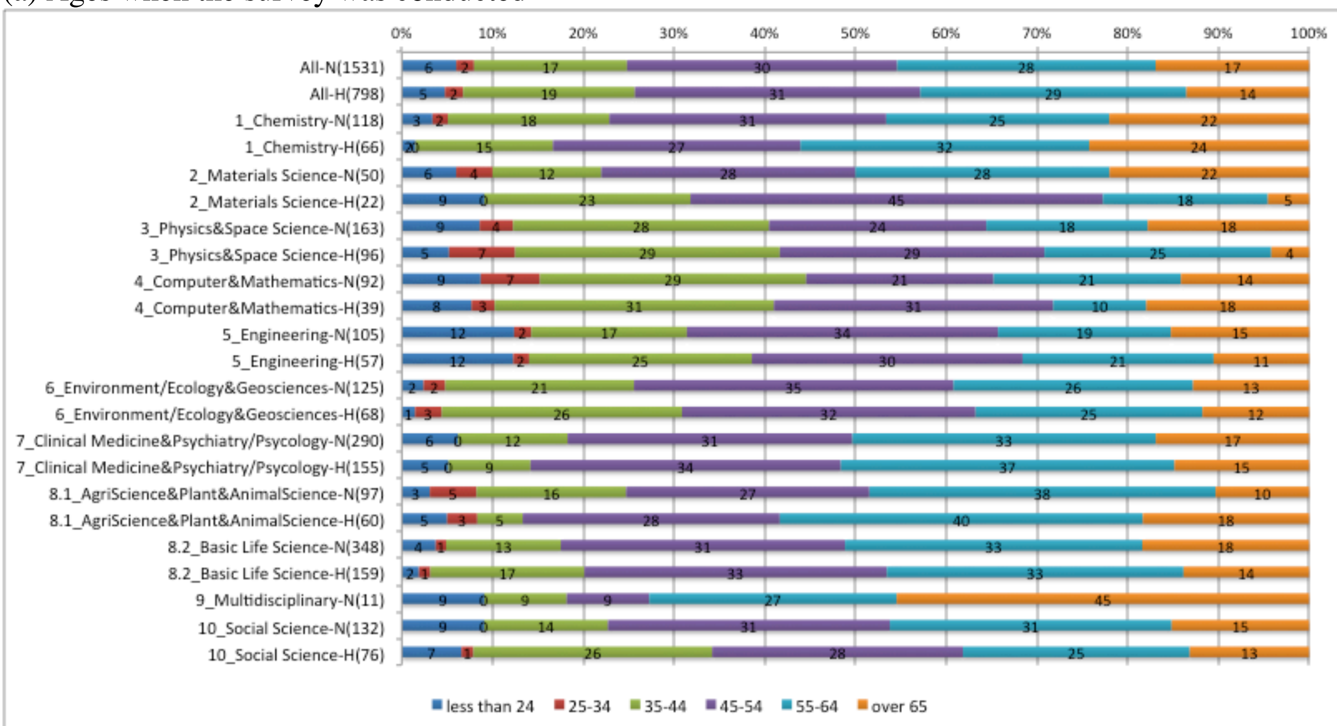
2-6 Characteristics of the respondents

2-6-1 Age

In Exhibit 4A, chart (a) summarizes the age distribution of the respondents, at the time when the survey was conducted (Sep 2010); chart (b) summarized the age distribution of the respondents, at the age when the focal paper was submitted. Exhibit 4B gives the mean (and standard deviation) of age at the time of the survey, by field. The average ages of respondents when the survey was conducted, across all fields in the year 2010 are 53 (SD = 10.5) for the H paper authors and 54 (SD = 11.1) for N paper authors.

Exhibit 4A. Age distributions of respondents

(a) Ages when the survey was conducted



(b) Age when the focal paper was submitted

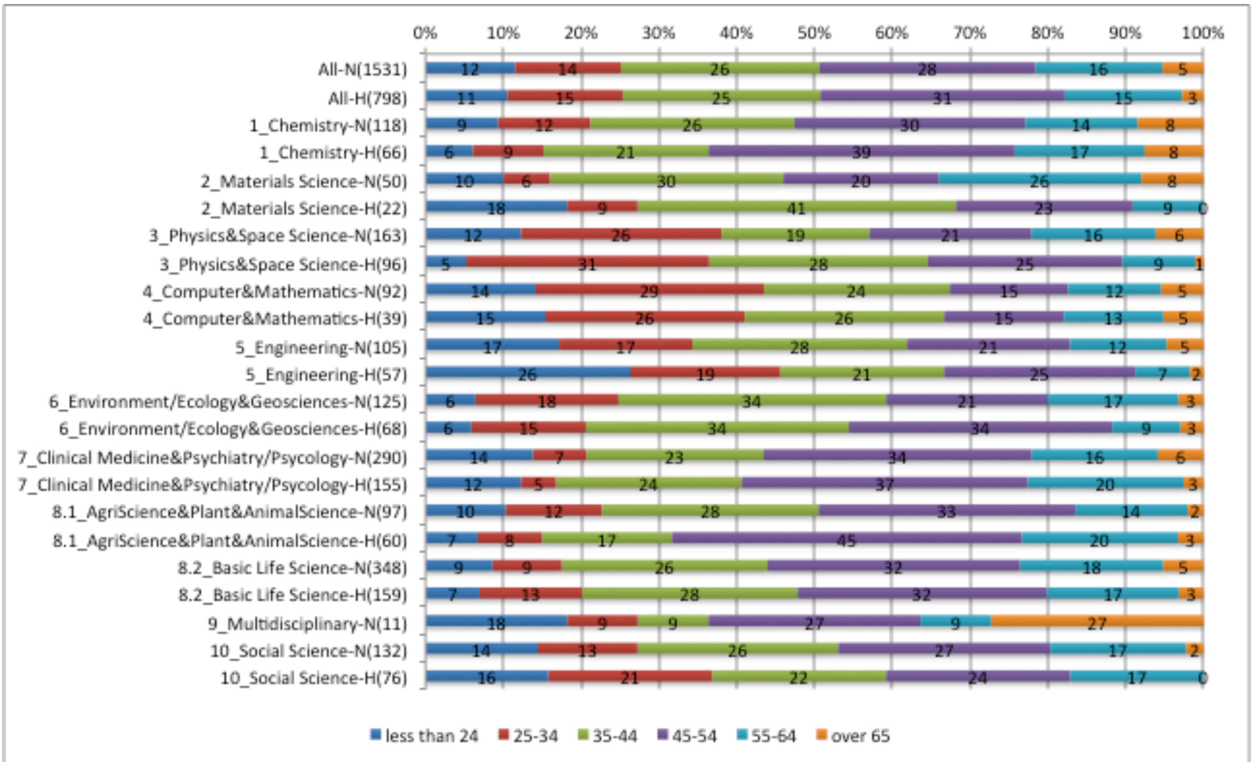


Exhibit 4B. The average age when the survey was conducted across fields by Top vs. Random papers

	H papers (798)		N papers (1531)	
	Mean	S.D.	Mean	SD
1_Chemistry	56.4	10.6	55.3	12.3
2_Materials science	49.7	8.0	55.9	12.1
3_Physics & Space science	48.2	10.5	51.8	12.8
4_Computer Science & Mathematics	50.4	12.6	50.7	12.6
5_Engineering	51.0	11.0	53.1	11.4
6_Environment/Ecology&Geosciences	51.3	11.0	52.1	10.6
7_Clinical Medicine	55.5	9.5	55.9	9.7
8.1_Agriculture Science&Plant&Animal Science	55.9	9.3	53.2	10.5
8.2_Basic life science	53.8	10	55.4	10.5
9_Multidisciplinary	-	-	63.7	13.7
10_Social Science	52.1	9.6	54.5	9.7
All	53	10.5	54.3	11.1

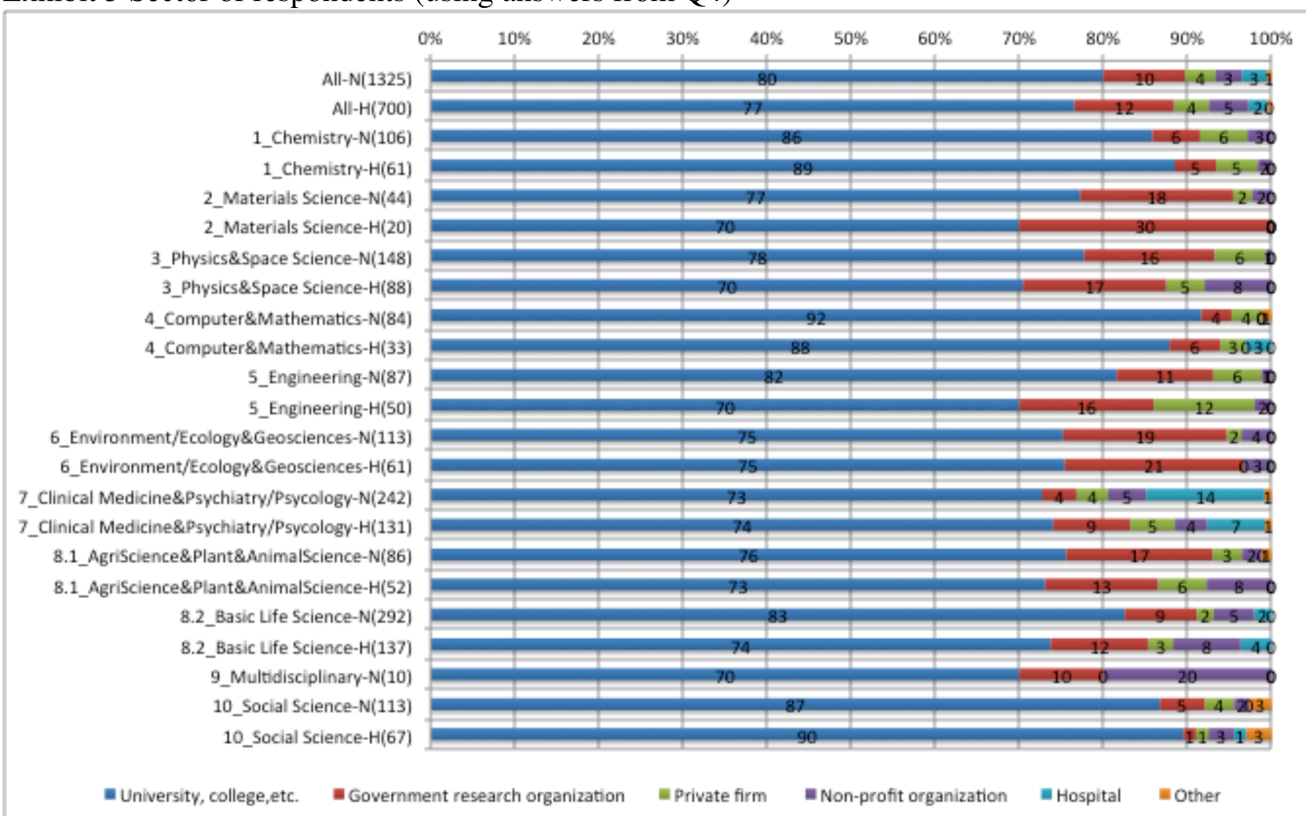
If we look at the age when the focal paper was submitted, overall, about half of respondents were age 45 or older when they submitted the focal paper. The average ages of respondents across all fields are 45.5 for H papers and 46.6 for N papers. In chemistry, clinical medicine and basic life science, over half of the respondents were age 45 or older when they submitted the paper, for both H and N papers (see Exhibit 4A(b)). In addition, for agricultural science H papers and materials science N papers, over half the respondents were age 45 and above. In contrast, in computer science/math, and H engineering papers, over 40% of respondents were less than 35 when they submitted the paper.

2-6-2 Sector compositions of the respondents when the focal paper was submitted

Of all the respondents in the US, 79% (N= 1596) are university researchers, 10.5% (N = 212) are government lab researchers and 4% (N = 77) are from private firms. Exhibit 5 shows the sector composition of the respondents when they submitted the paper. Among all the respondents with H papers, 77% were in universities or colleges, and 12% were in the government research institutes. For respondents with N papers, the sector distribution is quite similar, 80% of them were in universities and colleges, and 10% were in the government research institutes.

By fields, the share of government research institutes is substantially higher in material science (18% for N papers & 30% for H papers) and environment/ecology and geosciences (19% for N papers and 21% for H papers). Private firms are most prevalent among highly cited engineering papers, representing 12% of the total.

Exhibit 5 Sector of respondents (using answers from Q4)



2-6-3 Roles of the respondents in the research projects

This section summarizes 1) the role of respondents in the research management; 2) the role of respondents in the research implementation of the research project that produced the focal paper.

As shown in Exhibit 6, for those respondents with H papers, 83% of them said they played the leading role in the research management, i.e., the design of the research project, administration of the research project, and application for the research grant. Including those respondents who reported that they were a member of the research management, but less than a leading role (7%), in total 9% of respondents played at least some role in management. Similarly, 87% of respondent with N papers also reported that they played some role in the research management. This suggests that our sample consists primarily of authors that key members of the research team and hence should be competent to answer the other questions on the survey.

Although about 90% of the respondents were in the management roles, there are some fields reported that “management was not necessary” in a substantial proportion. For example, among all the computer & mathematics respondents, 21% said that management was not necessary for them, followed by physics & space science, (16% for N papers and 11% for H papers). This suggests that the structure of projects may differ significantly across fields.

In term of the role of respondents in the research implementation, Exhibit 7 shows that over 60% of respondents in both groups said that they were involved in the central part of the research and contributed the most to the research output, and another 20% said they took part in the central part of the research. Only about 4% of our N respondents and 2% of H respondents were those who provided material, data, equipment or facilities.

Exhibit 6. Role of Respondents in the Management of the Project.

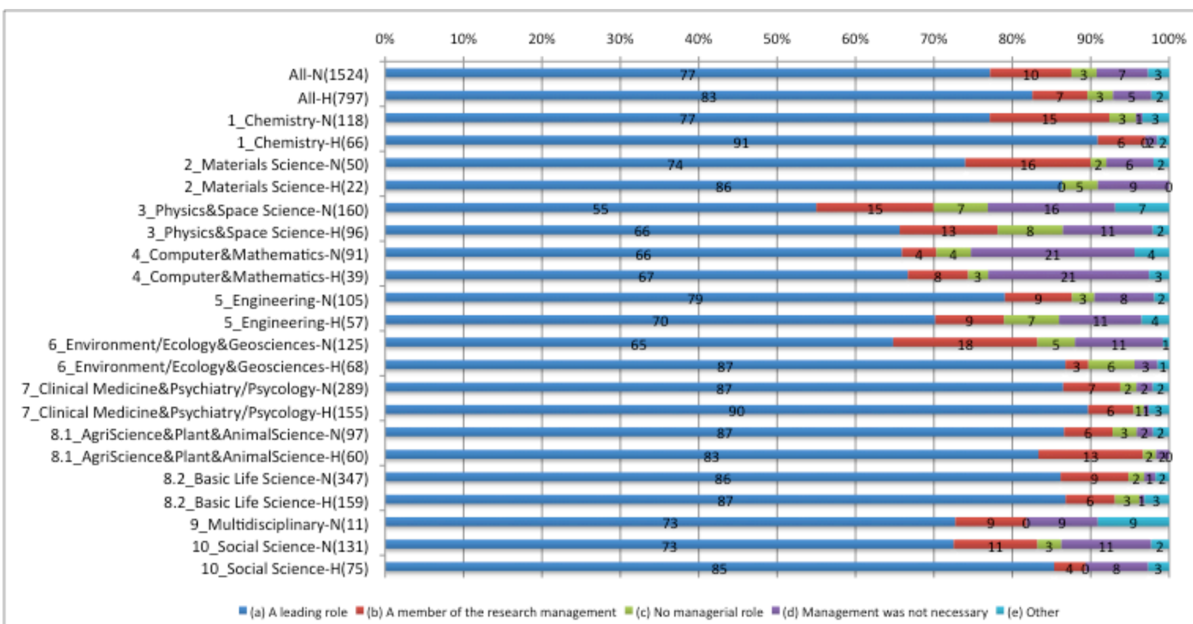
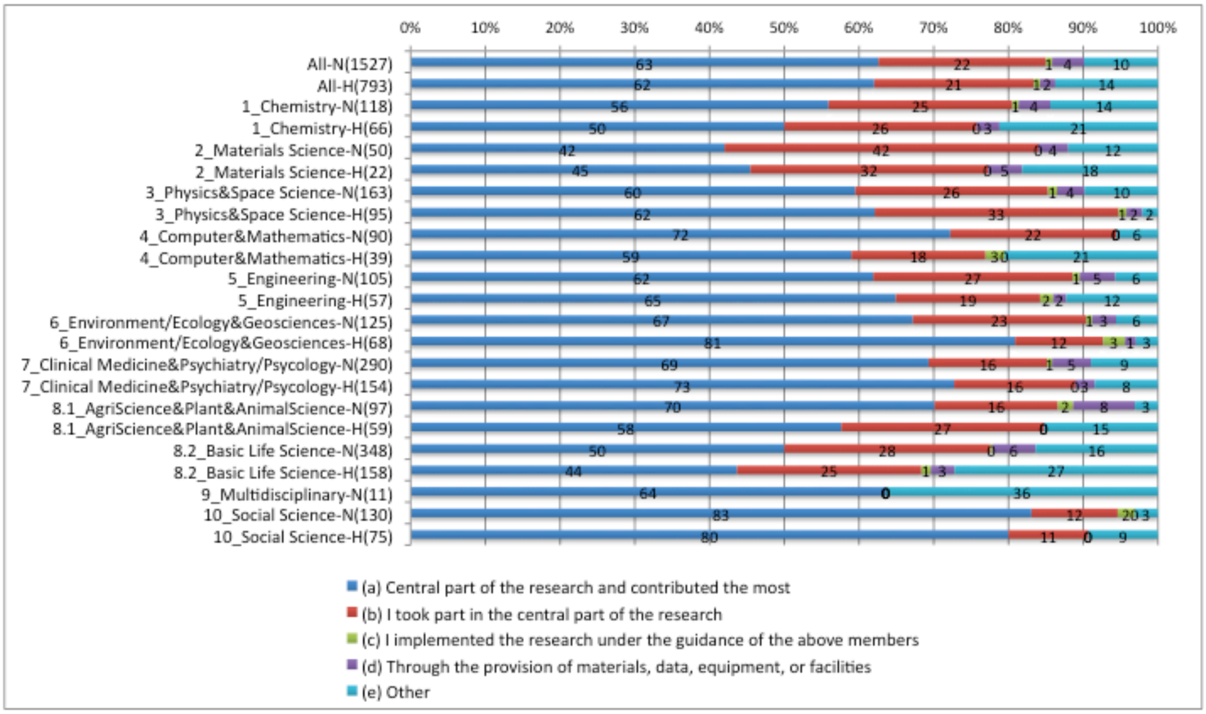


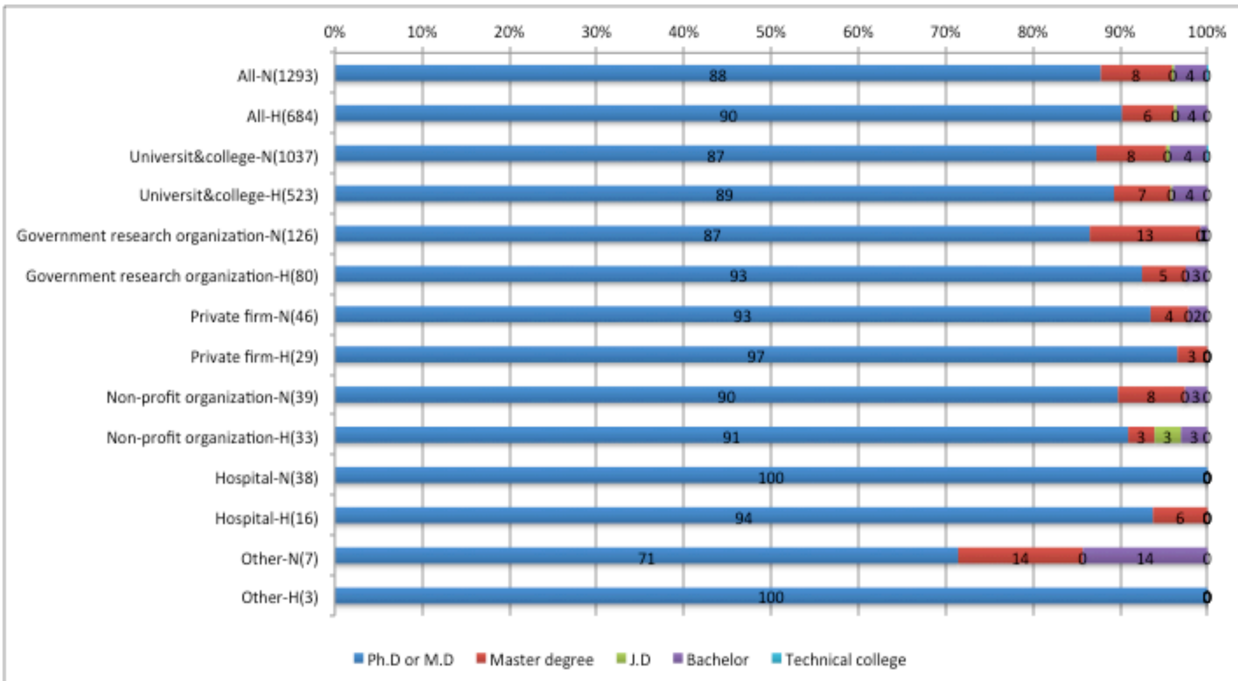
Exhibit 7. Role of Respondents in the Research Implementation.



2-6-4 Research careers of the respondents

Exhibit 8 shows the distribution of the highest degree of the respondents when the research project was launched, broken out by organization. Among all the respondents, 90% of respondents for H papers and 88% of respondents for N papers had a Ph.D or M.D. About 7% had a master's degree (many of whom may have been PhD students during the surveyed research project). The proportion of Ph.D or M.D respondents is quite high across all sectors, even industry respondents, where over 90% had a PhD or MD.

Exhibit 8. Highest Degree of Respondents when the Project was Launched, by Sector.



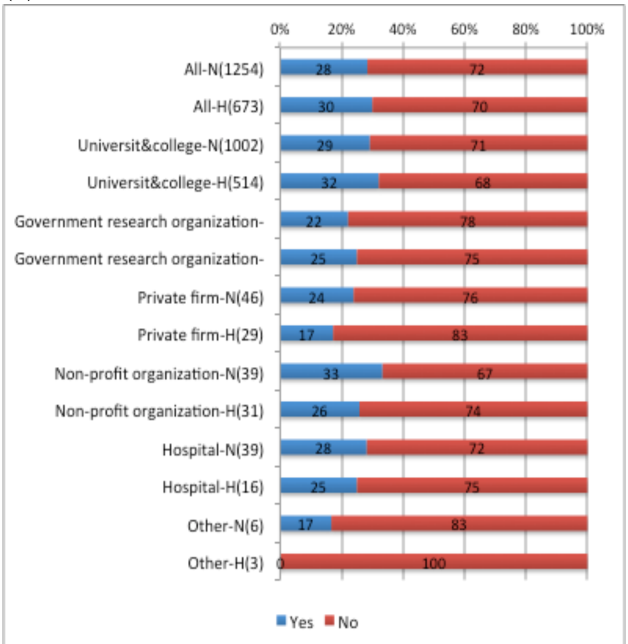
We also asked about the research accomplishments at the time the research project was launched (see Exhibit 9). Around 28% of respondents said they had won a distinguished paper/conference award, there is no different between H papers and N papers. Forty percent of respondents said they had served on an editorial board of an international journal. Authors with H papers (46%) are more likely to serve on an editorial board of an international journal than authors with N papers (37%).

We find that 30% had stayed abroad for one year or more for study or research, with no difference between H and N. Overseas experience was especially high for H papers by firm respondents. We also see that about a third of our respondents had changed organization in the previous five years. Mobility was highest among those in private firms (with over half of those in the H group having changed organizations in the last 5 years). In contrast, mobility is lowest among those in government organizations and hospitals.

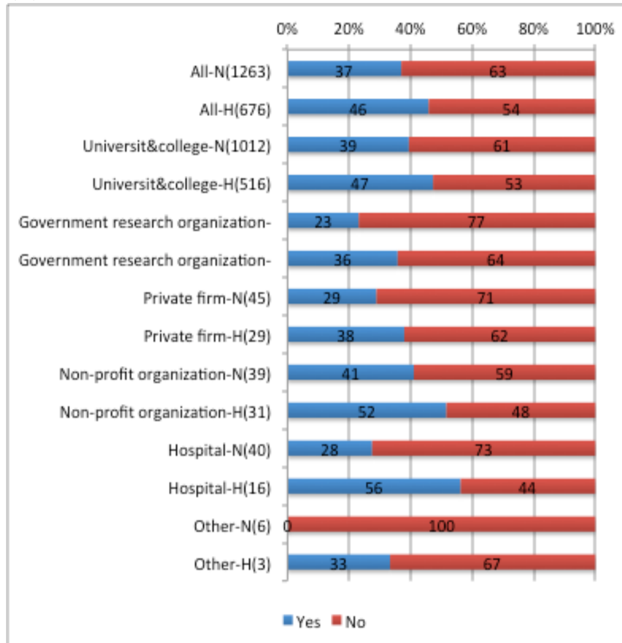
Exhibit 9 Research accomplishments of the respondents when the research project was launched (1)

(a) Had won a distinguished paper award or a conference award from an academic society
(b) Had served on an editorial board of an international journal

(a)

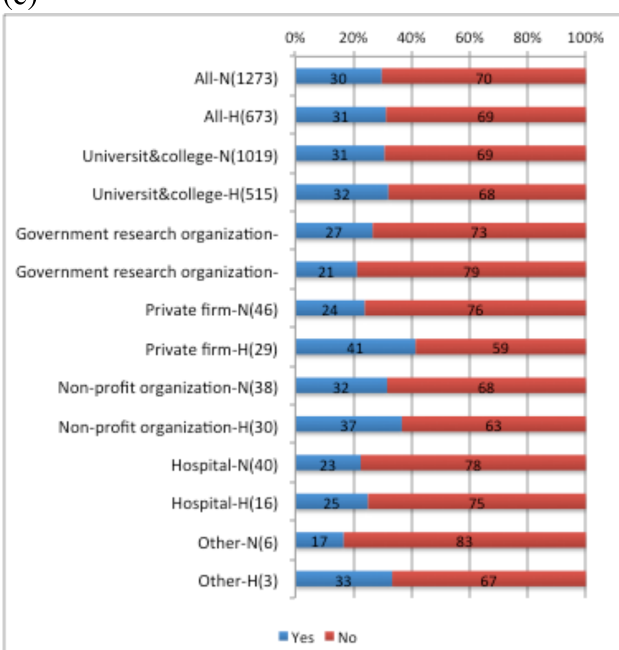


(b)

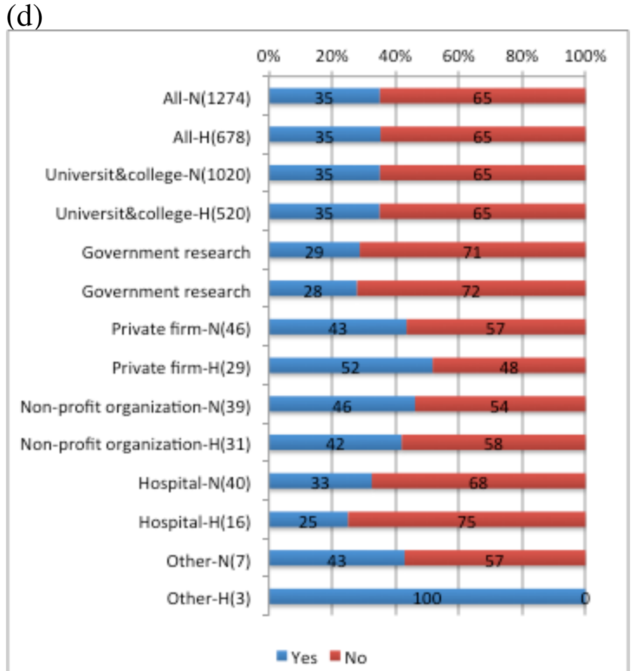


(c) Had stayed abroad for one year or more for study or research
(d) Had changed academic or research positions across organizations in the preceding five years

(c)



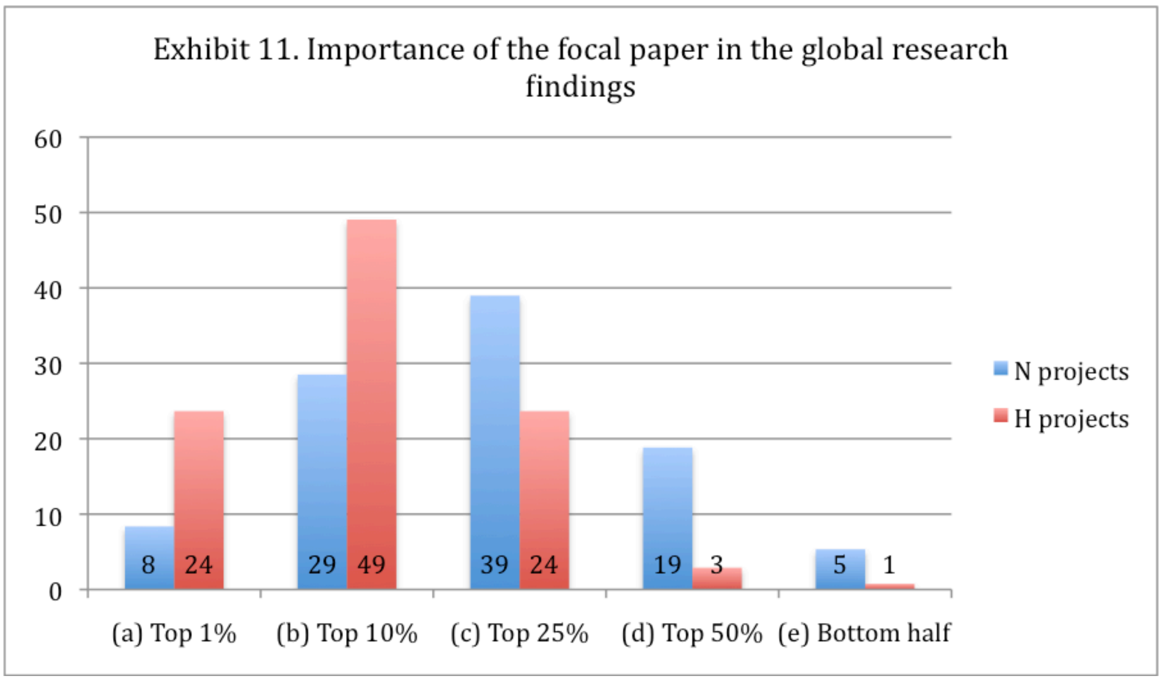
(d)



3 CHARACTERISTICS OF THE FOCAL PAPERS

3-1 Importance of the focal paper in the field.

For exhibit 11, we have asked respondents to evaluate the importance of their paper compared to the global research findings in the same field. Highly cited papers are significantly more likely to assess themselves in the top 1% and top 10%, compared to N papers. With regards to H papers, 24% of the respondent ranked their paper to be in the top 1%, and 73% evaluated their paper to be in the top 10% in the world. On the other hand, 37% of N project papers assessed themselves to be in the top 10% of the research findings in the world. These results suggest some caution in using this self-reported measure when estimating the impact of the paper, since, by construction, all of the H papers were chosen from the top 1% of the citation distribution and none of the N papers were from this part of the distribution. Of course, this could be due to problems with citations as a measure of the research importance of the paper, in addition to biases in self-reporting.



4. MOTIVATIONS FOR THE RESEARCH PROJECT AND UNCERTAINTIES IN THE RESEARCH

4-1 Motivations for the research project

Building off of Stokes' work on the motivations for research, we asked our respondents to describe the motivations of their projects on two dimensions: finding fundamental principles or understanding and solving specific issues in real life. For papers H projects, 79% responded that finding fundamental principles or understanding is a very important research motivation (5 on a 5 point scale). At the same time, 44% of the H papers also consider that solving specific issues in real life is an important research motivation. Respective ratios for the N paper are 68% and 40%. The projects for which both motivations were very important, what Stokes describes as "Pasteur's quadrant" were 32% of H and 25% of N projects. These results suggest that a significant fraction of research sits in Pasteur's quadrant, if it we limit our definition of "high" as 5 out of 5. Furthermore, H projects are more likely to be in Pasteur's quadrant, suggesting that much important research in advancing fundamental knowledge is also geared toward addressing practical problems.

Stokes defined "Bohr's quadrant" as those who were high on "pursuit of fundamental principles/understanding" but not especially interested in addressing practical problems. Bohr's quadrant represented 47% of those for H projects and 43% of those for N projects. Thus, despite calls for more Mode 2 or Pasteur's quadrant research, we see that a significant share of contemporary science is primarily focused on advancing knowledge, without regard for the application of the results. In contrast, "Edison's quadrant", defined as those who are focused on "solving specific issues in real life", represents only 15% of N projects and 12% of H projects. Exhibit 15B gives the percent of those in Pasteur's quadrant, and those emphasizing applied research ("solving issues in real life"), by field, excluding those respondents working for private firms. Clinical medicine, engineering and agricultural sciences are all high on applied research motives, especially for N papers, which is consistent with the missions of these fields, which are to translate scientific findings into practical application. Similar, clinical medicine high on Pasteur's quadrant. On the other hand, physics & space science is low on Pasteur's quadrant.

Exhibit 16 shows how the research motivations vary by field. Those who consider 'pursuit of fundamental principle/ understanding' as very important account for more than 80% of projects for chemistry, materials science, physics/space science, computer/mathematics and basic life science. On the other hand, respondents tend to put less emphasis on "solving issues in real life" across the sample. However, it was a very important motivation for fields such as Engineering and Clinical medicine & Psychiatry/psychology, accounting for more than 55% of projects.

Exhibit 15A. Distribution of the projects by a quadrant model

(a) Motivation for the H projects

		Solving specific issues in real life		
		Other	Very important	Total
fundamental principles/	Very important	47%	32%	79%
	Other	9%	12%	21%
Total		56%	44%	100%

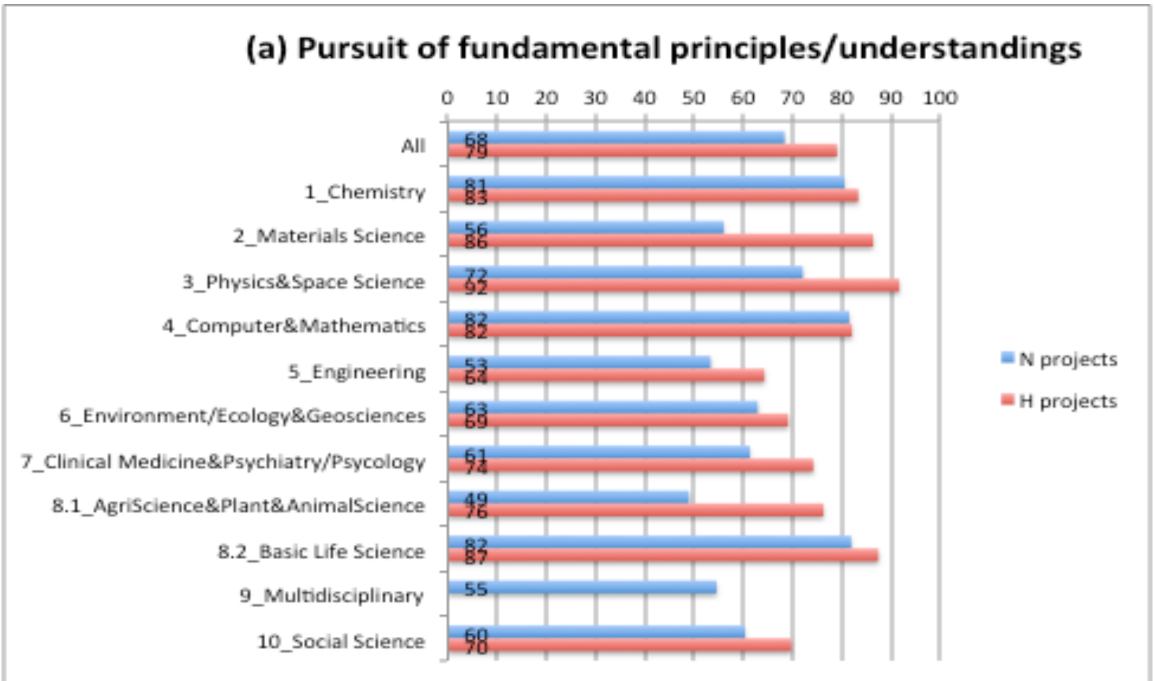
(b) Motivations for the N projects

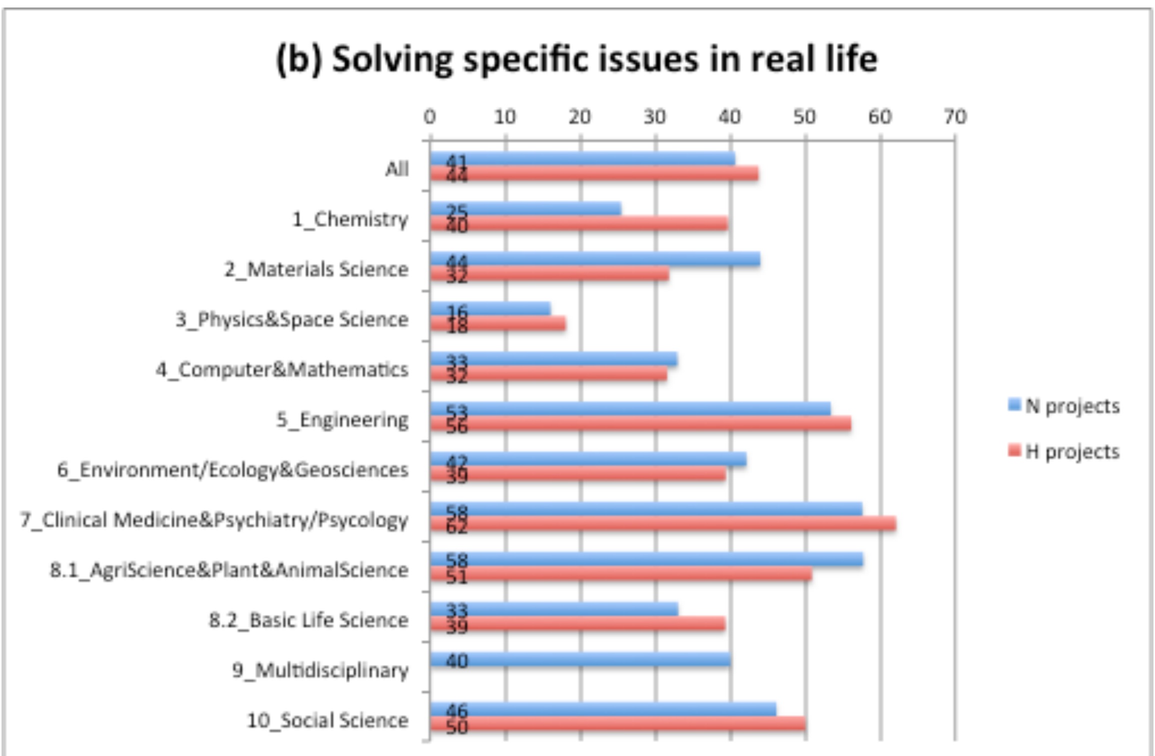
		Solving specific issues in real life		
		Other	Very important	Total
fundamental principles/	Very important	43%	25%	68%
	Other	17%	15%	32%
Total		59%	40%	100%

Exhibit 15B. Percent of Applied Projects and Pasteur's Quadrant Projects, by field, excluding firm respondents.

	Applied (rated 5)		Pasteur	
	High	Normal	High	Normal
1_Chemistry	40.0%	21.6%	29.1%	14.4%
2_Materials Science	30.0%	44.2%	30.0%	30.2%
3_Physics&Space Science	15.9%	11.1%	12.2%	6.8%
4_Computer & Mathematics	31.3%	32.5%	25.0%	22.1%
5_Engineering	45.5%	56.3%	30.2%	28.8%
6_Environment/Ecology & Geosciences	39.0%	42.6%	28.8%	19.4%
7_Clinical Medicine & Psychiatry/Psychology	59.8%	54.4%	45.9%	34.5%
8.1_AgriScience & Plant & Animal Science	45.7%	55.4%	30.4%	26.5%
8.2_Basic Life Science	38.9%	30.3%	31.3%	24.2%

Exhibit 16. Motivations for initiating the research projects by field (% very important)



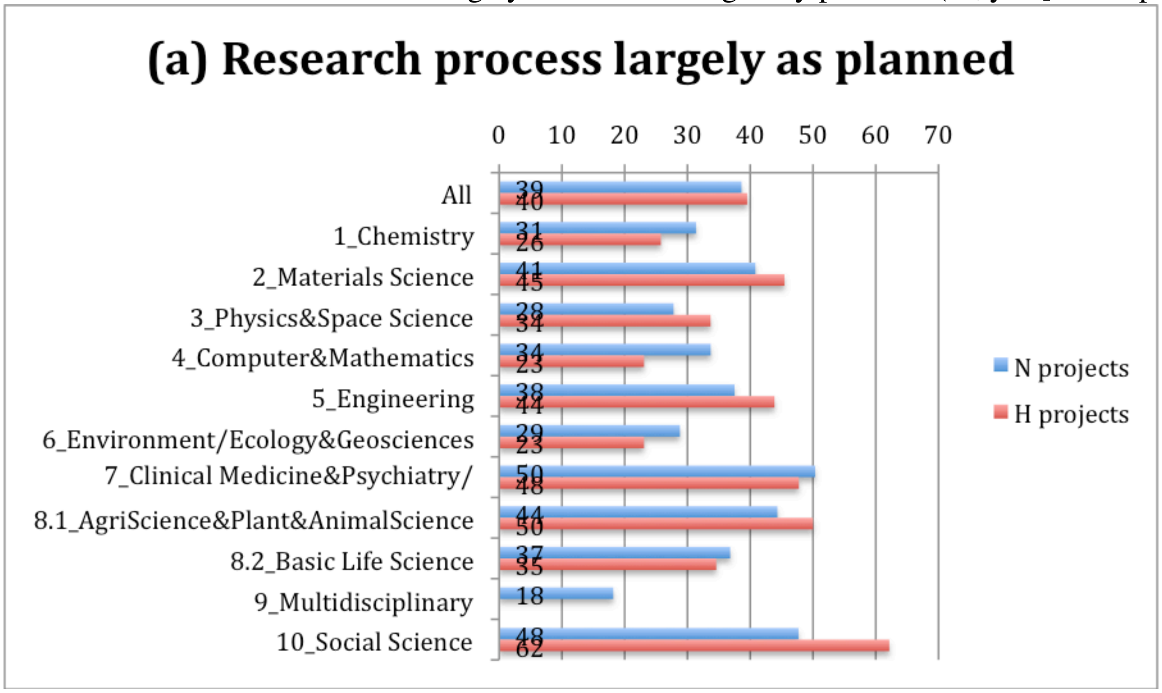


4-2 Uncertainties in the research process and research outcomes

In addition to asking about the motivations of the research, we also explored the extent to which the research process and research results varied from what was initially planned. Because science is an uncertain activity, we would expect that projects would often have to change their research process as they gain more knowledge. Furthermore, we expect that the research results may vary from what the project was initially targeting. This uncertainty can take two forms. The project results may turn out more, or less, significant that initially anticipated. In addition, the research may produce answers to questions that were not initially posed when the project began, what we call serendipity (Stephan 2010). This serendipity may be critical for advancing science. According to Stephan, one potential source of serendipity may be access to novel equipment, which may provide unexpected capabilities and allow answering questions that were not posed as part of the initial project. As one example, during a large scale project on neutrino detection, the physicists built a large array of photon detectors buried in Antarctic ice. The resolution of the array depends on the regularity of the ice formations. Based on contemporary geoscience understanding, the array was designed to have a particular expected resolution. In fact, once data collection started, the resolution was much greater than expected, because it turned out the ice was more regular than current geoscience knowledge predicted. The results was a significant geoscience paper based on this neutrino physics experiment, a serendipitous finding (Halzen 2010).

Exhibit 17 summarizes the results for on uncertainties in the research process. The research proceeded as initially planned for 39% of the N projects and 40% of the H projects. Also, the research outcomes came out as initially planned for 38% N projects and 26% for H projects. Thus, while the research process is equally certain in the two cases, there is more uncertainty in the results in the case of H projects (which may be why they are highly cited). Social science research tends to have very high rates of the research process proceeding as originally planned. In addition, clinical medicine and agricultural science also have high rates of coming out as originally planned. According to one of our interviews with a clinical researcher, clinical medicine may be especially constrained from changing because of the need to get protocol modifications approved by human subjects review boards.

Exhibit 17. Incidence of research largely the same as originally planned (%, yes [1 on 5 point scale])



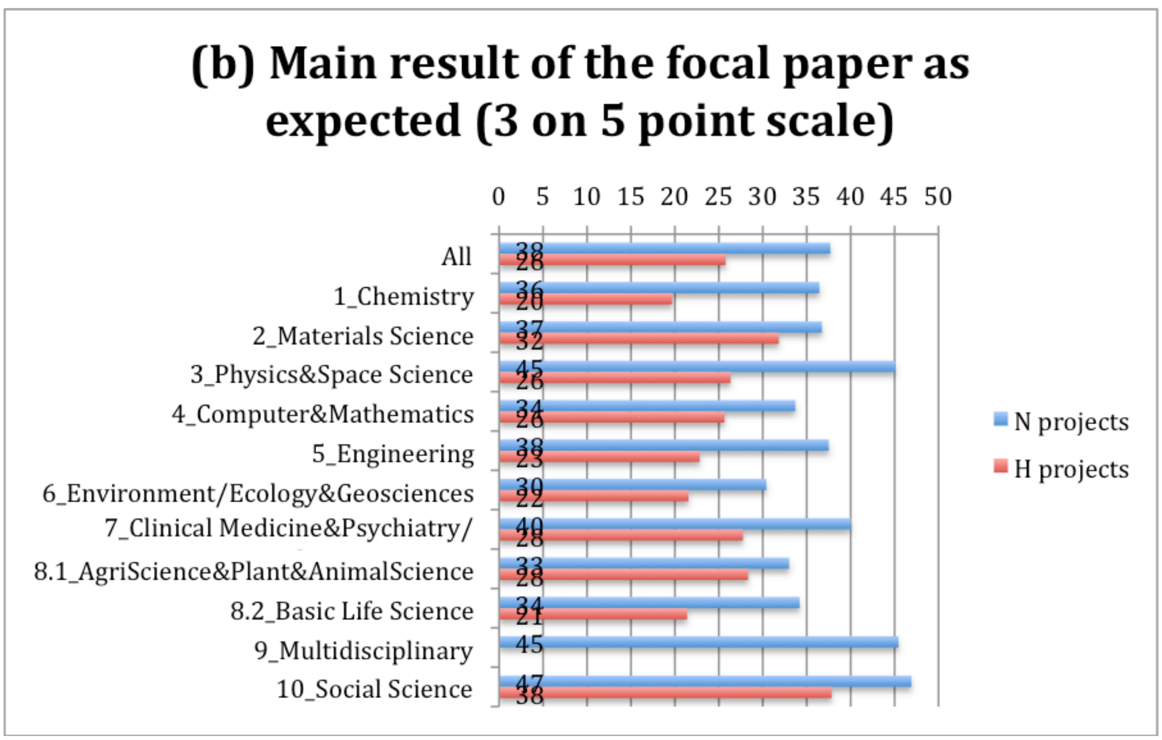
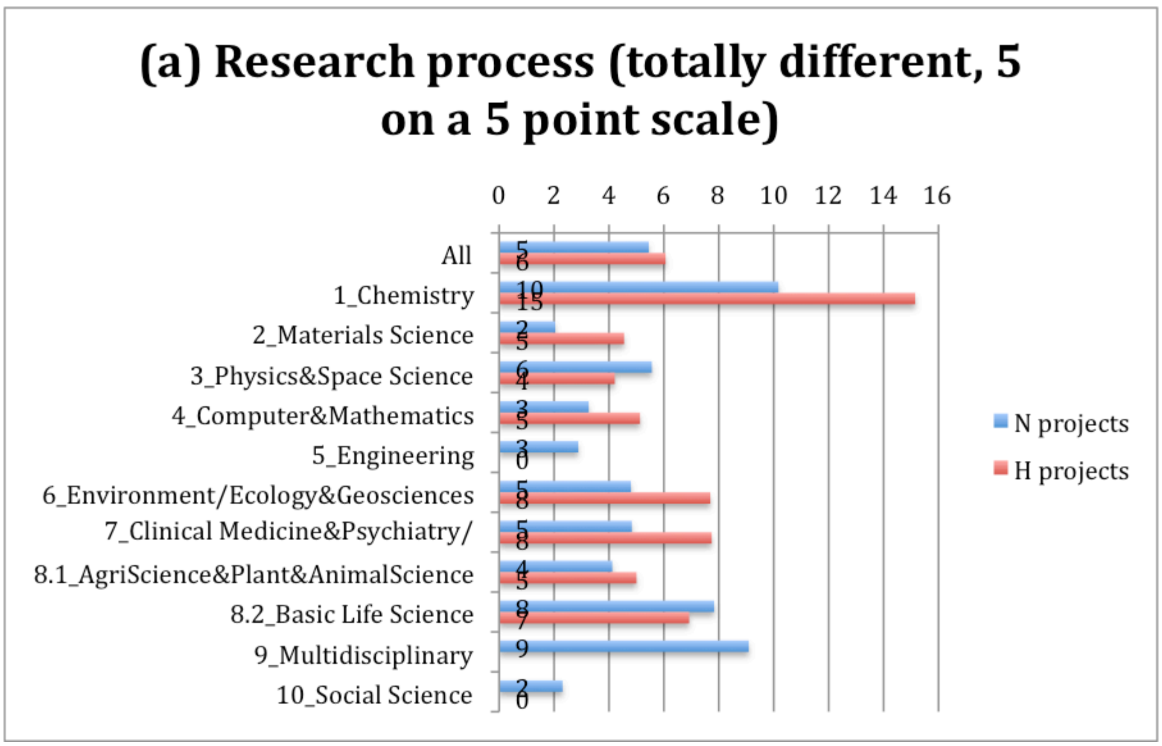
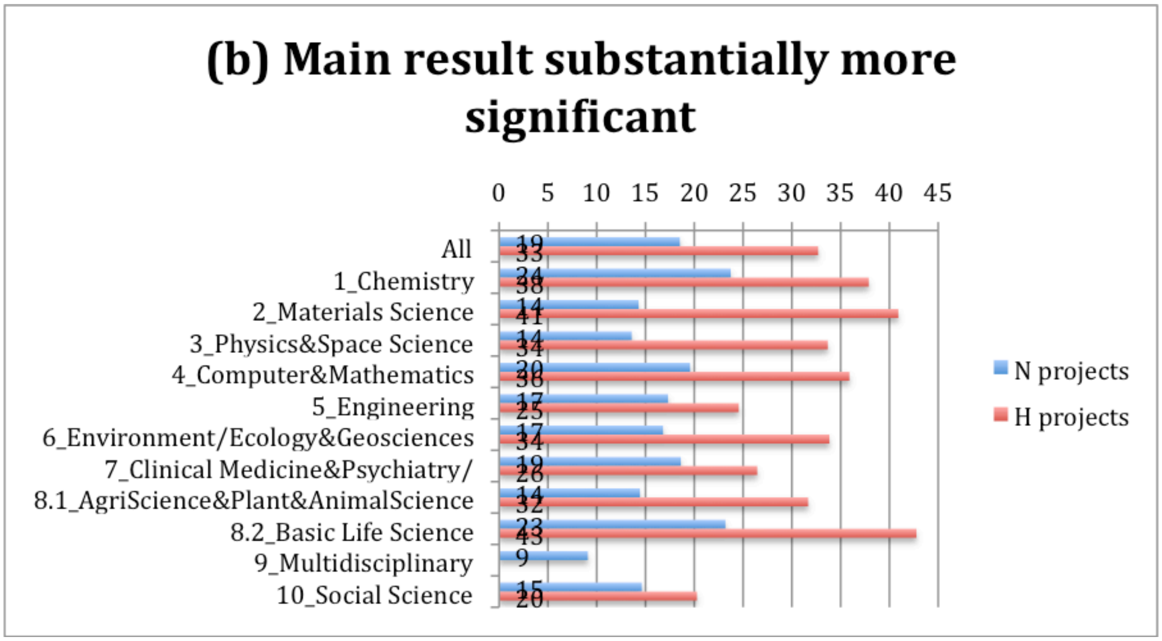


Exhibit 18 gives the summary results for the case where the research process and results were significantly different the original plan. Using a measure of a 5 on a 5 point scale, only 5% of the N projects and 6% of the H projects turn out to be very different compared to the original research plan. With regards to the research output, 19% of the N project and 33% of the H projects came out to be significantly better than researchers' original plan (5 on a 5 point scale).

Overall, N projects have relatively low uncertainties in research outcome as opposed to the H projects, which are more likely to have had an unexpectedly good outcome. H projects in life science, chemistry and material science were especially likely to generate substantially more significant outcomes than originally planned. Also, chemistry involves relatively high research process uncertainties, especially for H projects. These results suggest that chemistry may be the most flexible in terms of designing new experiments in the middle of project to take advantage of new information and new ideas.

Exhibit 18. Highly unexpected research process or results (% 5 on 5 point scale).



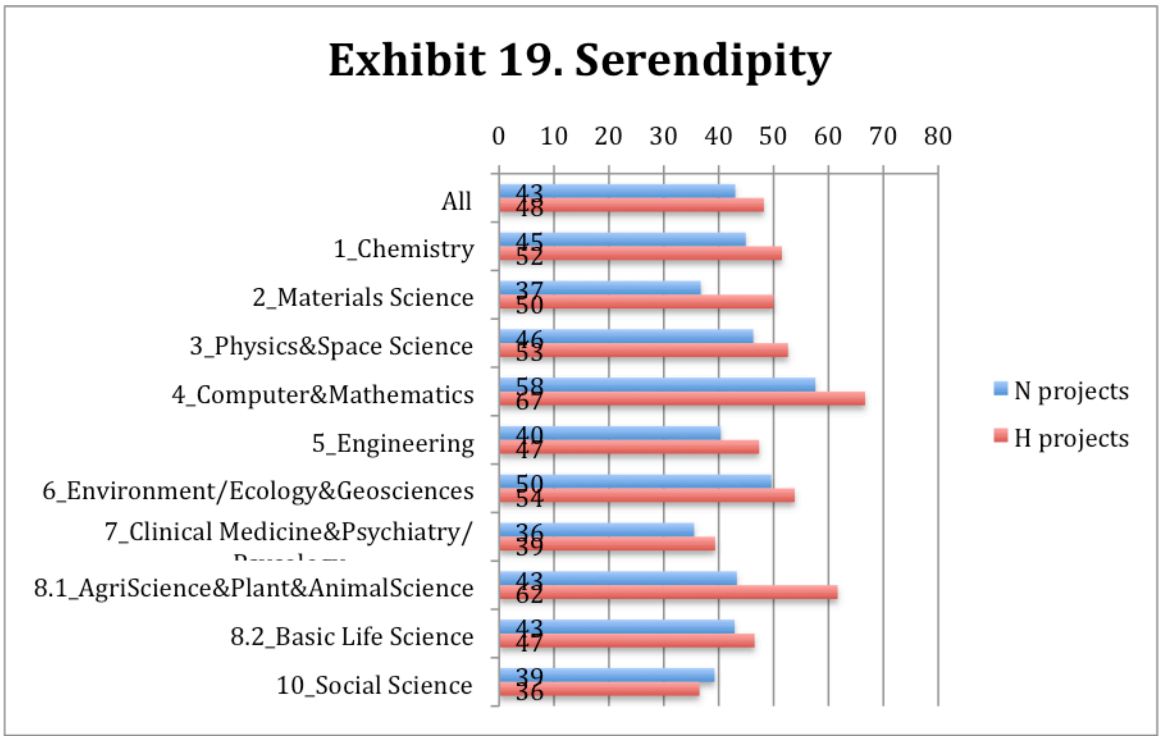

(b) Main result substantially more significant
0 5 10 15 20 25 30 35 40 45
All
1_Chemistry
2_Materials Science
3_Physics&Space Science
4_Computer&Mathematics
5_Engineering
6_Environment/Ecology&Geosciences
7_Clinical Medicine&Psychiatry/
8.1_AgriScience&Plant&AnimalScience
8.2_Basic Life Science
9_Multidisciplinary
10_Social Science
N projects
H projects

4-3 Serendipity

As we note above, uncertainty in research can also take the form of a serendipitous outcome, by which we mean that the research produced answers to the research question that is not originally posed.

Exhibit 19 shows the percent of projects with serendipitous findings, overall and by field, for N and H projects. The share of serendipitous outcomes was slightly higher for the H projects (48% of projects, across all fields) than for the N projects (43% of projects). For H projects, computer & mathematics and agriculture science/plant and animal science were fields that were most likely to produce serendipitous output. Unlike prior work in Japan, however, we do not see a strong field-level relationship between research output uncertainties and serendipity. Further work will look at the drivers of serendipity in order to understand what kinds of projects are most likely to generate serendipitous findings. We are also interested in the relations between serendipity and commercialization of research. One hypothesis is that many of the commercialized projects are those that were not originally looking for a commercial application of their research, but were where the research serendipitously produced a commercially viable technology.

Exhibit 19. Serendipity



5 RESEARCH COMPETITION

One important characteristic of science is the priority reward system (Merton 1973; Stephan 2010). Under such a system, competition is a key characteristics of the environment of science, and scientists are keenly aware of the pressure from competitors, although this competition varies significantly across scientific fields (Hagstrom 1974; Hong and Walsh 2009). As shown in Exhibit 20, most of the respondents were able to indicate the number of their domestic and international competitor even if "unknown" was available. Only 7% of the H projects and 13% of the N projects indicated that they don't know if there is any competitor in the US. Also, 19% of the H projects and 24% of the N projects suggest that there is no competitor in the US. Put differently, highly cited papers have more US competitors, with almost 22% reporting 6 or more competitors, compared to 15% of N papers with at least 6 competitors. On the other hand, scientists are aware of fewer competitors internationally across the types of projects.

Exhibit 21 shows whether respondents recognize 5 or more global competitors, *ex-ante*. Also, it indicates if researchers were concerned over the priority issues. In general, international competitors are more likely to be recognized in H projects than in N projects. The number of global competitors is largely recognized in fields such as physics, engineering and basic life science. However, no foreign competitors were recognized *ex-ante* in fields like environment/ecology & geoscience and Agriculture/Plant/Animal science.

With regards to the priority loss, 41% of the H projects and 25% of the N projects were concerned about losing priority. Among them, 7% of H and 4% of N projects were very much (5 on a 5 point scale) concerned about priority loss much. In particular, priority loss is a big concern in fields like Computer science/Mathematics, Material science and Basic life science.

Exhibit 20. Number of competitors (competing teams) recognized *ex-ante*
(at the stage of project initiation)

		Percent					
		None	1	2-5	6-10	More than 10	Unknown
USA	H projects	18.5	11.2	41.2	13.0	8.8	7.4
	N projects	23.9	11.4	37.0	9.4	5.6	12.8
Foreign	H projects	21.9	10.2	33.4	12.1	11.2	11.1
(Outside of US)	N projects	25.4	9.8	30.8	9.5	7.9	16.6

Exhibit 21. Number of foreign competitors recognized ex-ante and concern over priority loss

	%, 5 or more major foreign competitors		%, no foreign competitors		Very much concerned over priority loss		Concerned over priority loss (including very much)	
	H	N	H	N	H	N	H	N
All	55.8	50.2	18.4	23.9	7.2	3.9	41.4	25.1
1_Chemistry	60.0	57.0	12.3	21.9	6.2	4.4	49.2	22.8
2_Materials Science	54.6	55.6	9.1	28.9		4.4	27.3	15.6
3_Physics&Space Science	72.6	61.8	9.5	16.6	13.7	7.0	52.6	37.0
4_Computer&Mathematics	42.1	46.1	18.4	15.7	13.2	3.4	36.8	31.5
5_Engineering	64.3	52.0	12.5	20.6	3.6	3.9	44.6	19.6
6_Environment/Ecology& Geosciences	56.7	36.7	22.4	30.8	9.0	2.5	37.3	19.2
7_Clinical Medicine & Psychiatry/Psychology	50.7	45.9	22.0	27.1	4.0	4.0	34.7	19.5
8.1_Agriculture Science & Plant &Animal Science	50.9	47.9	27.3	33.0	7.3	1.1	43.6	18.1
8.2_Basic Life Science	62.8	61.1	11.8	17.8	9.2	5.0	47.7	35.9
10_Social Science	30.1	25.4	38.4	34.9	2.7	0.8	26.0	11.1

6 KNOWLEDGE SOURCES AND RESEARCH MANAGEMENT

6-1 External knowledge sources that inspired the research project

Scientific research is a cumulative process, the new advances in science building on existing knowledge from a variety of sources (e.g., textbooks, published literature, experts, formal collaboration or informal collaborations). Thus, an important question in the scientific work process is the role of various sources for generating new scientific research projects. To capture this dimension of the project, our survey asked respondents a series of 11 questions about knowledge sources that were useful for conceiving the research project, on a 6 point scale, including 0 for did not use, and ranging from 1 (not important) to 5 (very important). We identified 5 categories of knowledge sources that inspired the research project, including (1) open literatures (scientific literature and patent literature); (2) forums (e.g. conferences); (3) internal or past collaborators (e.g. colleagues); (4) external experts (e.g. competitors or partners in industry) and (5) experts with different skills.

Exhibit 22 summarizes the results, giving the percent responding 5: Very Important for each item. For H papers, the most important knowledge sources that inspired the research project is published literatures (49%), followed by researchers with different research skills (15%), colleagues in your organization (14%), and past collaborators (12%). For N papers, the most important knowledge sources that inspired the research project is also published literatures (53%), followed by colleagues in your organization (14%), and researchers with different research skills (11%). There is little difference between H projects and N projects in most of the knowledge source questions; however, we find H projects report being somewhat more likely to use knowledge from researchers with different skills than N projects (15% for H projects vs. 11% for N projects) and somewhat less likely to rely on published literature (although it is still the most important source).

In the survey, we also asked the respondents to identify the country location of the knowledge source (for example, the location of the key researcher for a published paper, country of origin for visiting researchers, conference venue for conferences, etc.). The question about location of the external knowledge sources is a follow-up question after the question of the importance of the external knowledge source for suggesting the research project. Respondents who answered 4 (important) or 5 (very important) were also asked where the key external knowledge sources were located.

For the US scientists, most of the key knowledge sources (from open publications, conferences, colleagues, to researchers/partners with different trainings and skills) were domestic. On the other hand, the source of visiting researchers or post-doctoral researchers tends to be very international. Around 60% of the respondents reported that the key visiting researchers or postdoctoral researchers came from another country (28% from EU, 8% from Japan, and 8 % from China). Thus, the circulation of scholars from overseas to the US seems to be an important conduit for knowledge flows.

Exhibit 22. External knowledge sources used in the research projects: Importance of various knowledge sources for suggesting the research project (% 5:very important).

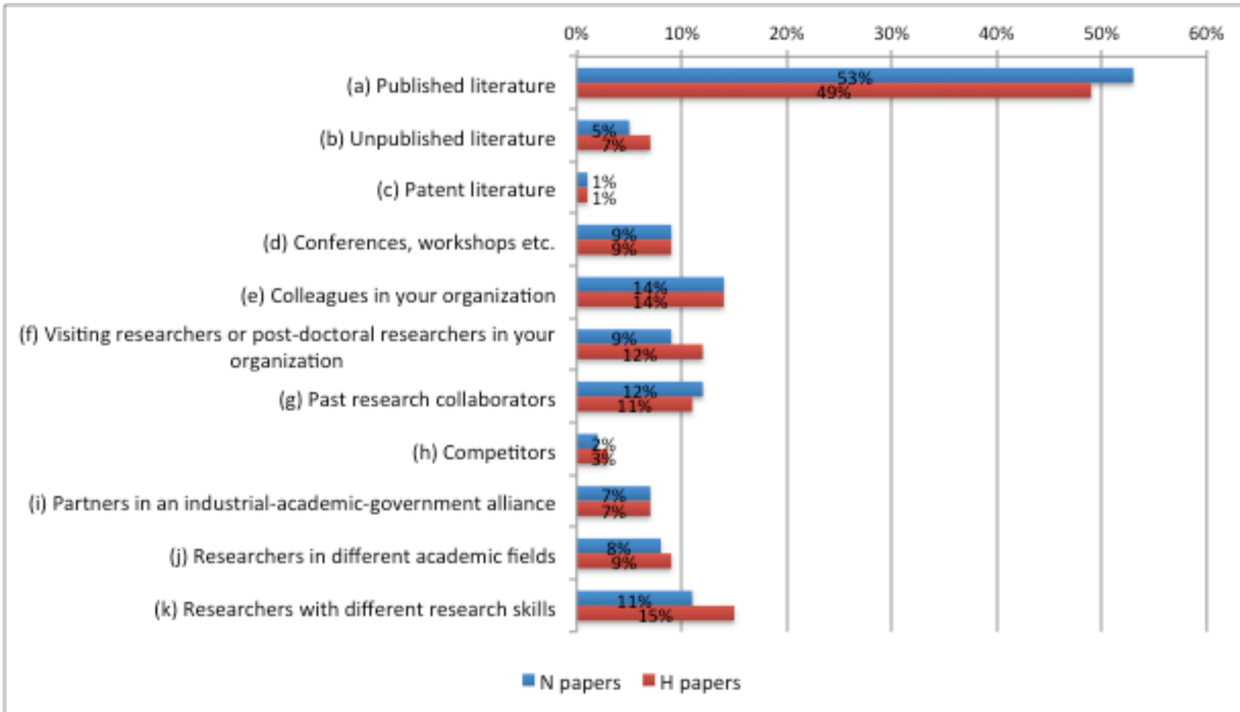
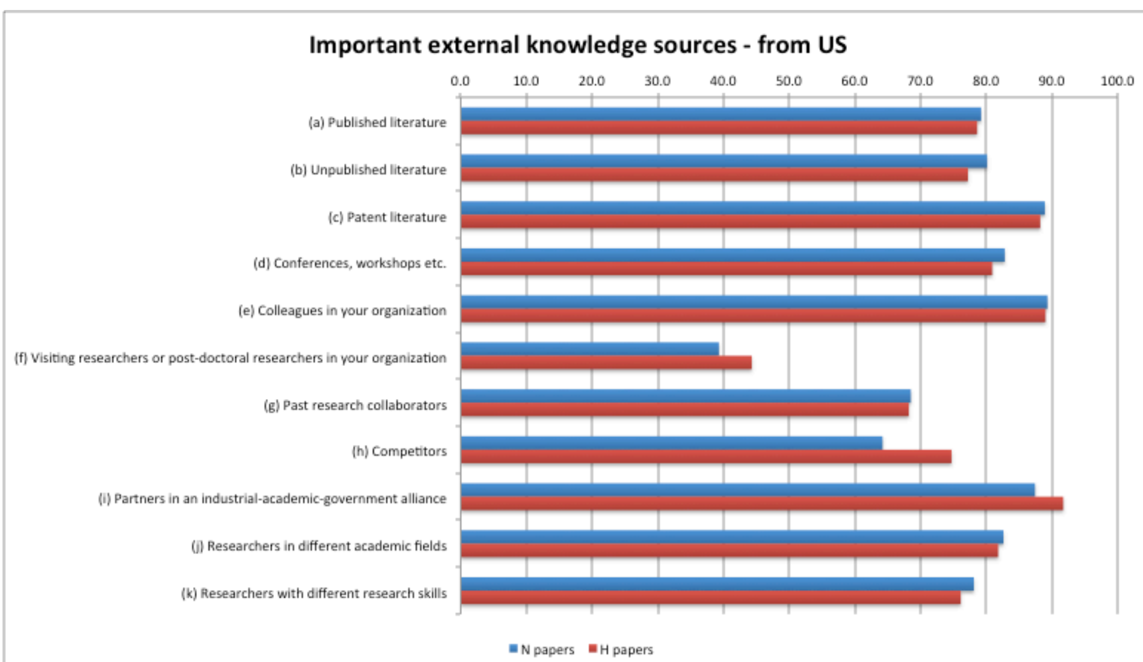
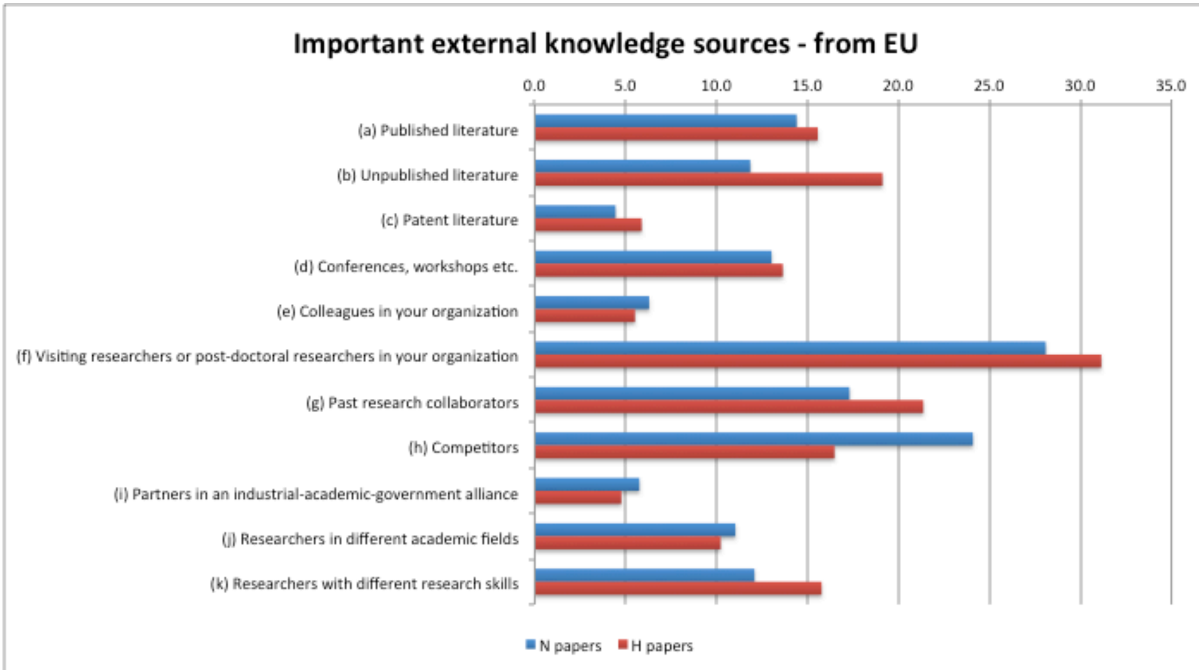
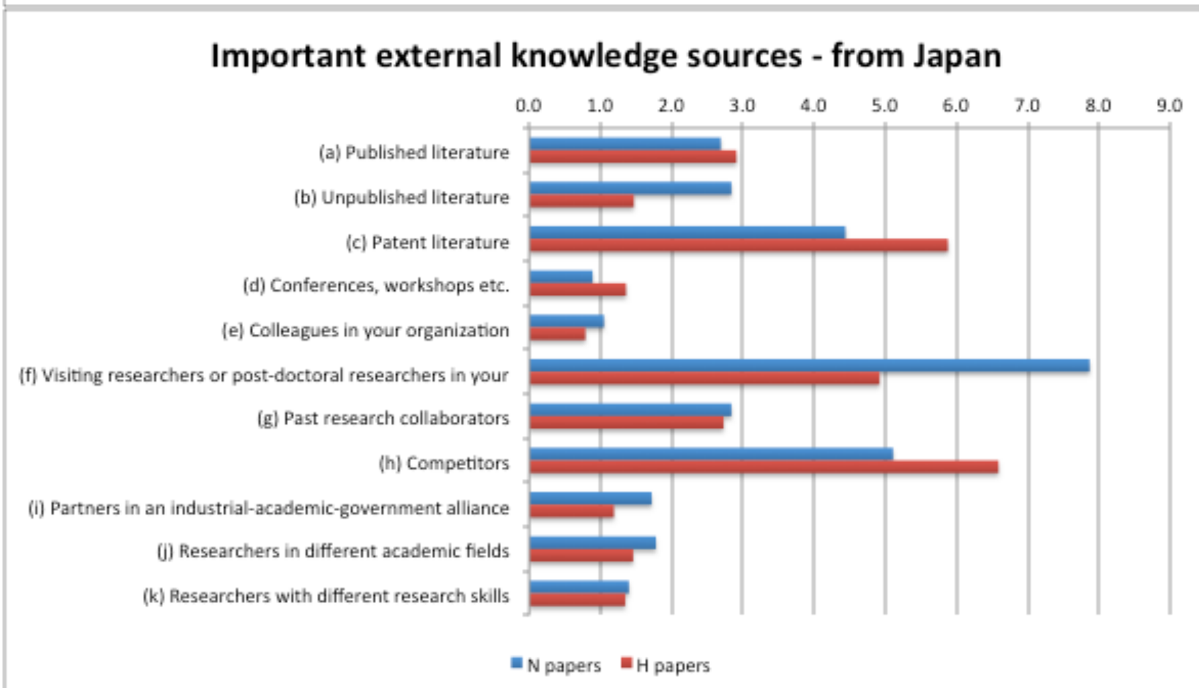
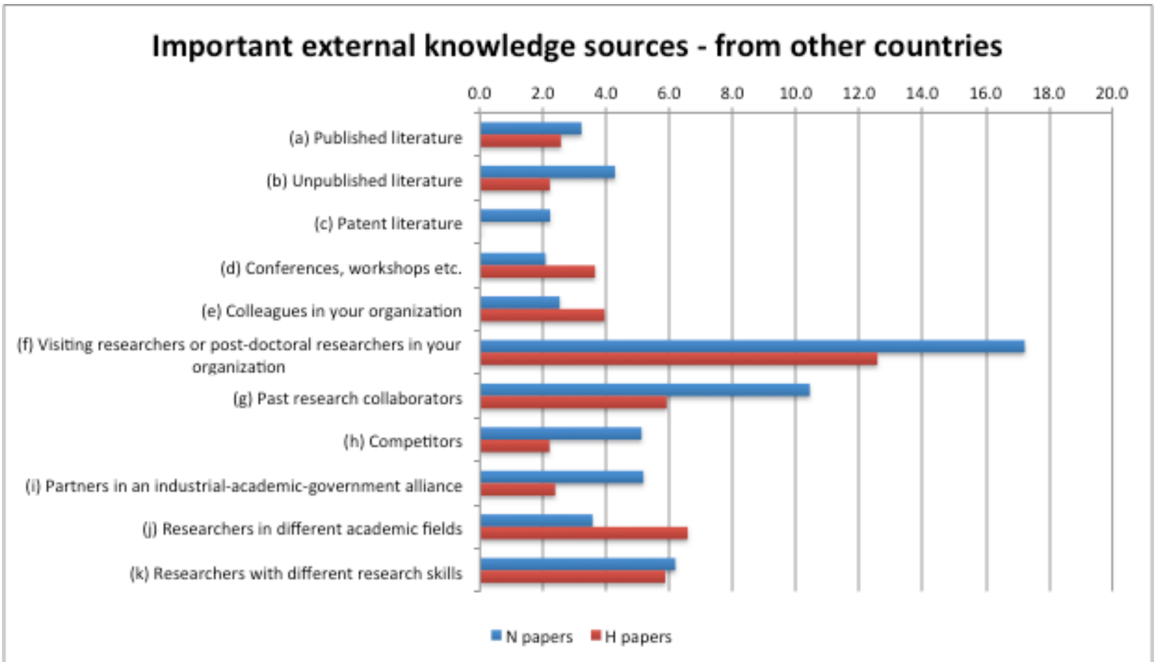


Exhibit 22(b) Percent of respondents choosing each country as the key knowledge source for each source, US, Europe, Japan, other.




Important external knowledge sources - from EU
0.0
5.0
10.0
15.0
20.0
25.0
30.0
35.0
(a) Published literature
(b) Unpublished literature
(c) Patent literature
(d) Conferences, workshops etc.
(e) Colleagues in your organization
(f) Visiting researchers or post-doctoral researchers in your organization
(g) Past research collaborators
(h) Competitors
(i) Partners in an industrial-academic-government alliance
(j) Researchers in different academic fields
(k) Researchers with different research skills
N papers
H papers


Important external knowledge sources - from Japan
0.0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
9.0
(a) Published literature
(b) Unpublished literature
(c) Patent literature
(d) Conferences, workshops etc.
(e) Colleagues in your organization
(f) Visiting researchers or post-doctoral researchers in your
(g) Past research collaborators
(h) Competitors
(i) Partners in an industrial-academic-government alliance
(j) Researchers in different academic fields
(k) Researchers with different research skills
N papers
H papers



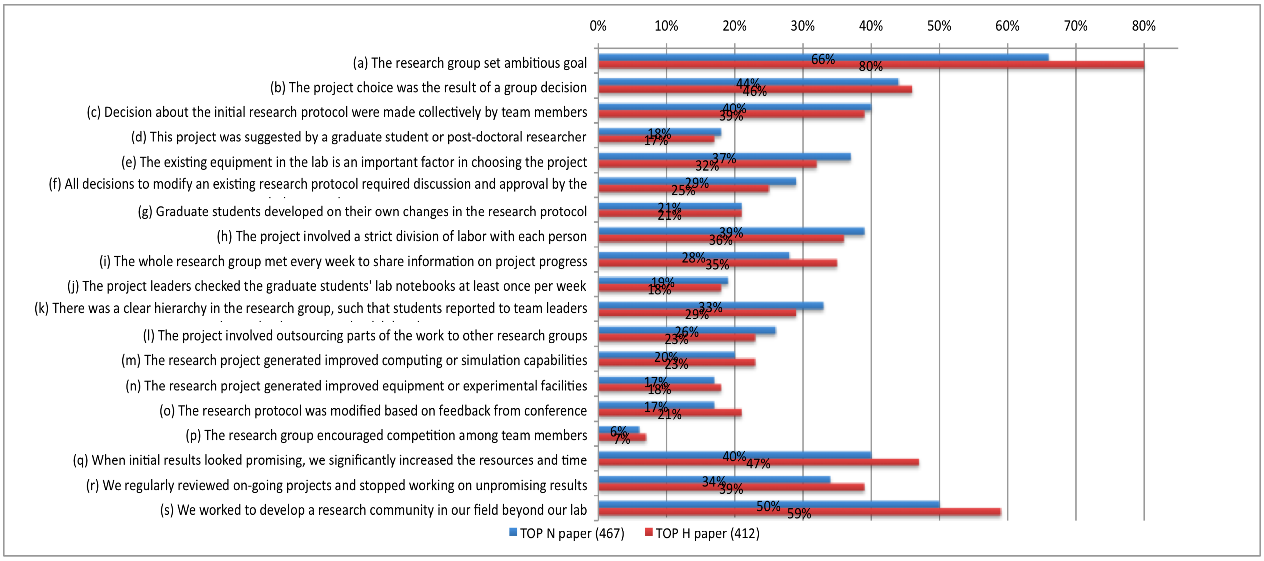
6-2 Research management

A key aspect of contemporary team science is the need to organize the activities of potentially many researchers, possibly spread across several sites. Even for smaller research teams, coordination of research efforts may be key to success. Based on prior work from the study of organizations, we suspect that systematic differences in lab structures may be a form of dynamic capabilities that are associated with sustained high performance (Teece 1986; Teece 2007).

The questionnaire asked about the following characteristics of lab management: setting ambitious project goals; participation of group members in project choice and protocol; graduate student originated projects; whether existing equipment guided project choice; decision-making process for modification of protocols; graduate student autonomy; division of labor; information sharing; supervision of graduate students (checking lab books); hierarchy; outsourcing; creation of new computer programs or simulations; creation of new equipment; feedback from conferences; competition within the team; adjusting of project mix based on early results and development of a research community (outside the lab).

While the questions were asked of all respondents, the problem of lab organization is likely to be most acute for larger labs. So, for this analysis, we report the responses only for those cases where the publication included at least 5 authors. This leaves us with over 800 cases (about half from H and half from N projects).

Exhibit 23. Research management (papers with 5 or more authors).



We find that the most widely used (mostly likely to be a 4 or 5 on the 5 point scale from used "not at all" to "very much so") was setting ambitious project goals. This was also the project management characteristic that most distinguished H (80%) from N (66%) projects. The second most used was developing a research community outside the lab. This also had a substantial gap between H (59%) and N (50%) papers. Other commonly used processes included adjusting the project mix depending on whether the early results are promising or not promising, again with the H projects being most likely to do this. We also find that about 40% of projects had developed the project choice and the protocol as group decisions, with little difference between H and N projects. A strict division of labor and a strict hierarchy were somewhat more common in N compared to H projects, while H projects were more likely to hold weekly information meetings of the whole research group. Just under 20% of projects said that the project choice was suggested by a graduate student (and little difference by strata). We find that competition within the group is quite rare, with only 6-7% of groups reporting this as describing their group.

Further work is needed to see if these practices are associated with differences in performance, once we control for differences in inputs. For example are lab management practices a form of dynamic capabilities that affect performance (unexpected discovery, serendipity, citations or commercialization)?

Research requires access to equipment. In many fields having access to research equipment (e.g. an electron microscope or a nuclear magnetic resonance) is the necessary condition to do research (Stephan and Levin, 2002). Running a lab is costly as well, according to Stephan's (1996) study, the average purchase price of the equipment in a lab is $3.25 million. Collaborative research is not only for obtaining diverse human capital from scientists, but also a mean to access research facilities through collaborators. For example, some advanced equipment is shared by the whole department or only owned in certain institutions, such as a high-energy particle accelerator or a super computer. In addition, as science becomes global, scholars suggest that Internet is able to lessen the communication difficulty among distant researchers in a collaborative team (Walsh and Maloney, 2007).

To understand the importance of accessing advanced research equipment and remote researchers for scientific research, we asked our survey targets: "To what extent did you have difficulty accessing potentially important research facility, databases and remote researchers for your project?" on a Likert-scale from 0 to 5 where 0 is did not need, 1 is no difficulty and 5 is great difficulty getting access. Firstly, we divide respondents into non-users and some level of users, and Exhibit 24 shows the result. H papers (29%) are slightly higher than N papers (26%) in the use of advanced external experimental equipment and research facility. Future analyses will break these results out by field to see which fields are most equipment dependent. Not surprisingly, about 90% of the respondents reported that they used databases of journals and published papers, consistent with the prior findings that published literature is one of the most important sources of information. The comparison between H papers and N papers shows that authors with H papers were more likely to access to research tool databases (54% vs. 47% for N papers), communicate with remote researchers using Internet (56% vs. 45% for N papers) and access to the latest (unpublished) research information (68% vs. 58% for N paper).

Exhibit 24. Use of advanced research facilities, databases, and the Internet for distant collaborators.

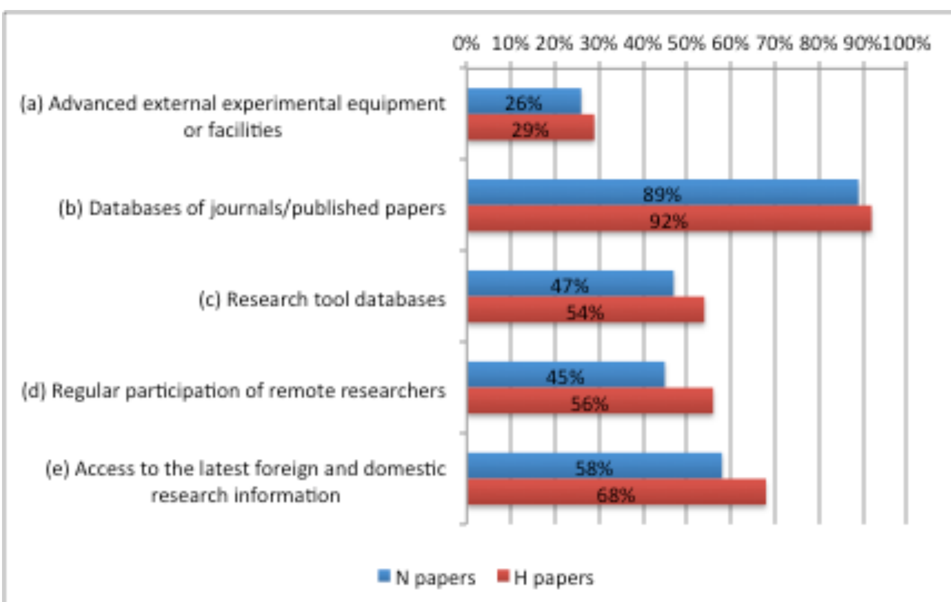


Exhibit 25. Percent of respondents reporting difficulty accessing advanced facilities, databases, and the Internet for distant collaborators (answered 4 or 5).

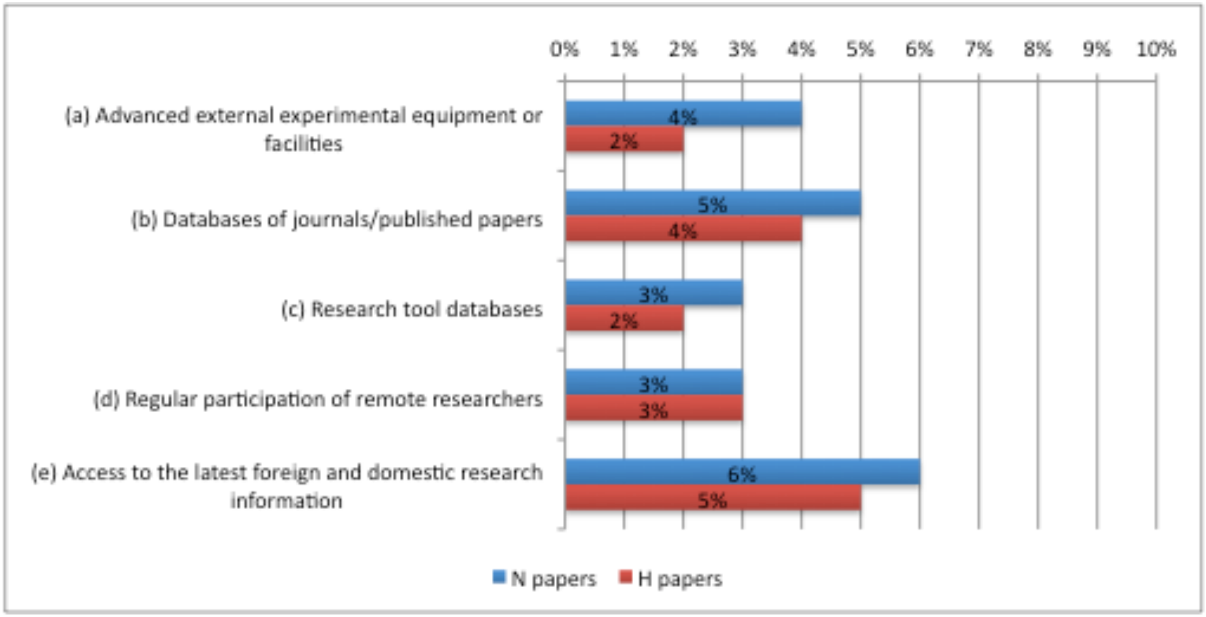


Exhibit 25 gives the percentage of respondents reporting difficulties in access scientific resources that they needed. In almost every case, the N papers report more difficulties. In particular, access to equipment is particularly problematic for N papers, where 4% report difficulty (out of 26% who needed such equipment), compared to 2% out of 29% for H papers. Put differently, 15% of N projects that needed access to equipment had trouble access it, while only 7% of H papers had such difficulties. Similarly, for unpublished papers, 6% of N projects had trouble access unpublished information (out of the 58% who needed such information), versus 5% of 68% for H papers. In other words, about 10% of N projects had trouble accessing unpublished information, compared to 7% of H projects. Access to invisible colleges may be a critical source of information and may explain some of the difference between H and N projects (Crane 1972).

7. RESEARCH TEAMS AND COLLABORATION

7-1 Number of authors

Prior work has shown that team research is becoming the norm in science (Wuchty, Jones, and Uzzi 2007a). While this prior work has provided detailed analyses of the changing number of authors and institutions in science, further work is needed to explore the details of collaboration structures, including non-co-author team members, and the impact of these structures on group performance. Furthermore, these structures vary significantly by field (Walsh and Bayma 1996). As a first step in understanding the structure of scientific groups, we present results on the number of authors and the number and types of non-author team members. We begin by describing the number of authors (Exhibit 26). Overall, the median and mean number of authors is 5 and 7 persons respectively for H projects and 3 and 4 persons respectively for N projects. Thus, H projects tend to be larger than N projects, consistent with prior work by Wuchty and colleagues (Wuchty, Jones, and Uzzi 2007b).

Exhibit 26. Distribution of number of authors by field.

		N	Min	Q1	Median	Q3	Max	Mean
All	H projects	798	1	3	5	8	374	7.08
	N projects	1531	1	2	3	5	74	4.02
1_Chemistry	H projects	66	1	2	4	6	12	4.59
	N projects	118	1	2	3	5	14	3.72
2_Materials Science	H projects	22	2	3	5	6	13	5.27
	N projects	50	1	2	3.5	5	9	3.68
3_Physics&Space Science	H projects	96	1	3	4.5	7	374	13.03
	N projects	163	1	2	3	4	50	4.03
4_Computer& Mathematics	H projects	39	1	2	2	5	11	3.36
	N projects	92	1	1	2	3	15	2.33
5_Engineering	H projects	57	1	2	3	6	78	5.87
	N projects	105	1	2	2	3.5	22	3.03
6_Environment/Ecology &Geosciences	H projects	68	1	3	4	7	49	6.42
	N projects	125	1	2	3	5	19	3.99
7_Clinical Medicine &Psychiatry/Psychology	H projects	155	1	5	8	11	47	8.60
	N projects	290	1	3	4	6	16	4.79
8.1_AgriScience&Plant &AnimalScience	H projects	60	1	4	5	7	37	6.41
	N projects	97	1	2	3	4	8	3.40
8.2_Basic Life Science	H projects	159	1	4	6	9	58	7.36
	N projects	348	1	3	4	6	74	4.99
10_Social Science	H projects	76	1	1	2	3	9	2.71
	N projects	132	1	2	2	3	13	2.76

The number of authors varies substantially across fields. Using median as a measure for comparison, we can observe that Clinical medicine & Psychiatry/Psychology and Basic life science have large team size. On the other hand, Computer science/Mathematics and Social science has fewer numbers of authors. Also, the size of the authors is larger for H projects than N projects in most fields. In particular, Clinical medicine & Psychiatry/Psychology has the largest difference in number of authors between H projects and N projects. Exhibit 27 shows distribution of number of authors by sector. The mean or median number of author does not differ substantially across sectors. This suggest that it is field-level characteristics that drive the team size, perhaps related to the need for specialized equipment, or access to distributed data source.

Exhibit 27. Distribution of number of authors by sector.

		N	Min	Q1	Median	Q3	Max	Mean
All	H projects	798	1	3	5	8	374	7.1
	N projects	1531	1	2	3	5	74	4.0
University	H projects	617	1	3	4	7	374	7.1
	N projects	1193	1	2	3	5	74	4.0
Government Lab	H projects	82	1	3	5.5	8	77	7.4
	N projects	135	1	2	3	5	24	4.1
Private Firm	H projects	33	1	3	4	7	26	5.6
	N projects	109	1	2	3	5	22	4.3

7-2 Number of collaborating researchers, students and technicians, who are not co-authors on the paper

Not all members of the research team are co-authors, and the co-authorship rules may vary by field and by sector. The sum of the number of collaborating researchers, students and technicians are shown in exhibit 28. It is a skewed distribution that has median of 0 non-author research team members. The number of non-coauthor research team members, broken down by sector, is shown in exhibit 29. The distribution is similar to the previous table. Thus, unlike in Japan, non-author collaborators are not so common in the US (Nagaoka, Igam, Eto, and Ijichi 2010).

Exhibit 28. Number of collaborating researchers, students and technicians, who are not co-authors on the paper by field

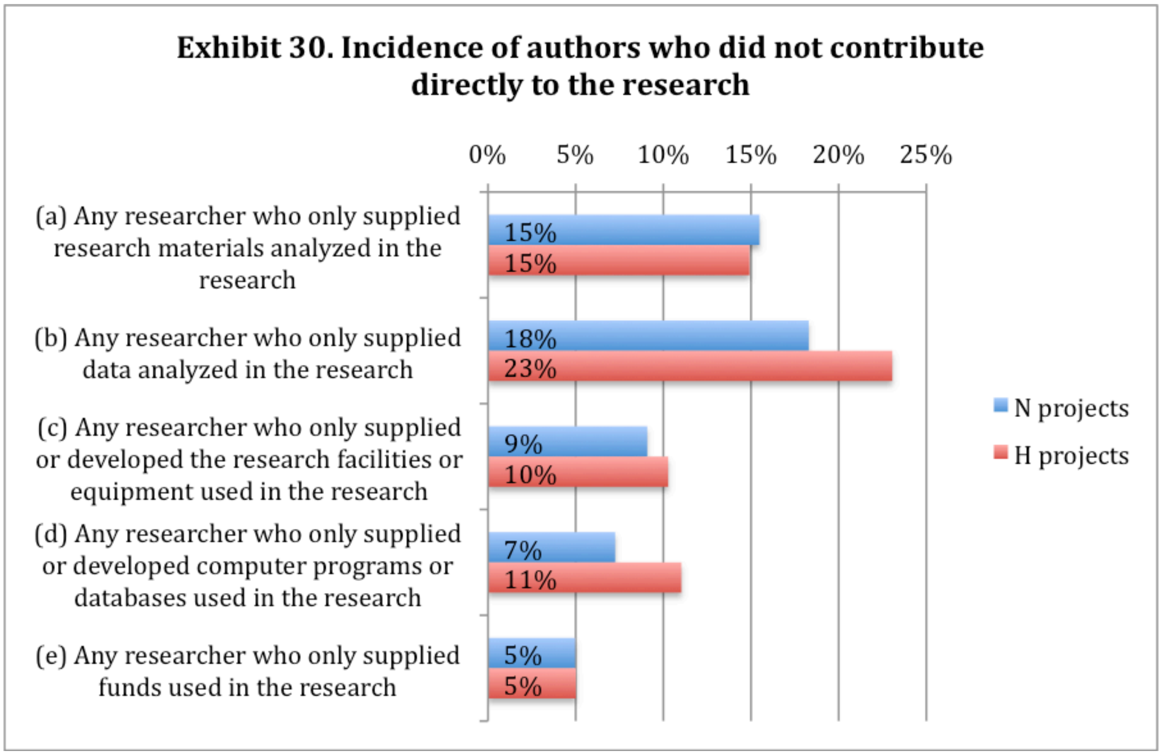
		N	Min	Q1	Median	Q3	Max	Mean
All	H projects	798	0	0	0	2	330	4.1
	N projects	1531	0	0	0	2	170	2.2
1_Chemistry	H projects	66	0	0	0	2	20	1.8
	N projects	118	0	0	0	1	16	1.3
2_Materials Science	H projects	22	0	0	0	1	5	1.0
	N projects	50	0	0	0	3	19	2.2
3_Physics&Space Science	H projects	96	0	0	0	1	330	6.6
	N projects	163	0	0	0	1	22	1.3
4_Computer & Mathematics	H projects	39	0	0	0	0	3	0.3
	N projects	92	0	0	0	0.5	20	1.5
5_Engineering	H projects	57	0	0	0	3	33	3.8
	N projects	105	0	0	0	3	15	2.0
6_Environment/Ecology & Geosciences	H projects	68	0	0	0	5	220	12.1
	N projects	125	0	0	1	3	28	3.1
7_Clinical Medicine &Psychiatry/Psychology	H projects	155	0	0	0.5	5	60	5.3
	N projects	290	0	0	0	2	170	3.5
8.1_AgriScience & Plant & Animal Science	H projects	60	0	0	1	2	18	2.8
	N projects	97	0	0	1	4	33	3.7
8.2_Basic Life Science	H projects	159	0	0	0	3	17	2.0
	N projects	348	0	0	0	1	18	1.4
10_Social Science	H projects	76	0	0	0	2	26	2.4
	N projects	132	0	0	0	1	16	1.3

Exhibit 29. Number of collaborating researchers, students and technicians, who are not co-authors on the paper, by sector.

		N	Min	Q1	Median	Q3	Max	Mean
All	H projects	798	0	0	0	2	330	4.1
	N projects	1531	0	0	0	2	170	2.2
University	H projects	617	0	0	0	3	330	4.6
	N projects	1193	0	0	0	2	110	2.1
Government Lab	H projects	82	0	0	0	2	60	4.0
	N projects	135	0	0	0	2	18	1.9
Private Firm	H projects	33	0	0	1	2	16	1.9
	N projects	109	0	0	0	2	170	4.5

7-3 Scope of authors: who are the authors?

Authorship may be given for a variety of reasons, including those who provide data, provide funding and provide research materials. Our survey examined the types of authors that appeared on each paper (Exhibit 30). About 15% of the papers included researchers who only provided research materials both for the N and H projects. Other than that, H projects are more likely to include researchers as co-authors who only supplied data (23%), facilities (10%) or computer programs (11%). Only a few projects report giving co-authorship for providing research funds (5% for both H and N projects). Future work will explore the differences by field, and the drivers of authorship. For example, so called "gift authorship" may be less common when commercialization is more likely, since concerns over property rights may limit sharing authorship.



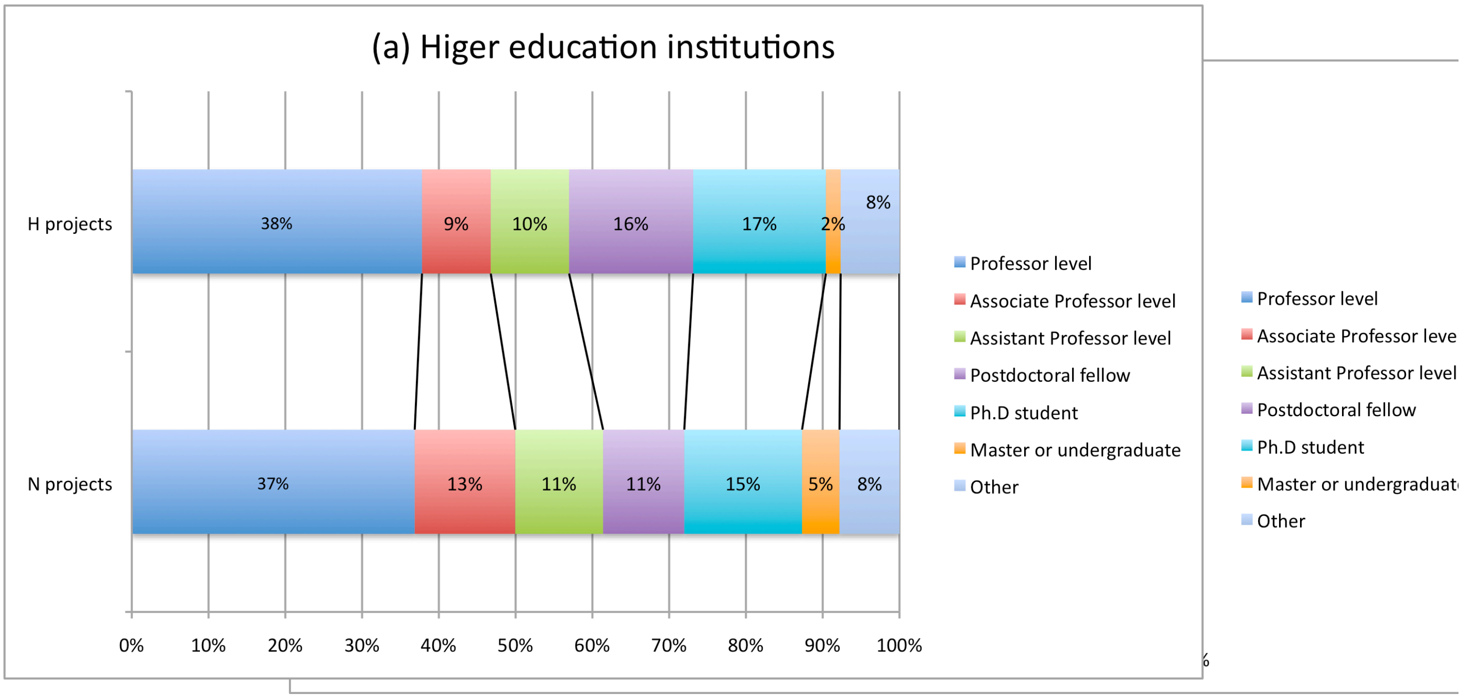
7-4 Combination of authors in academic/professional positions

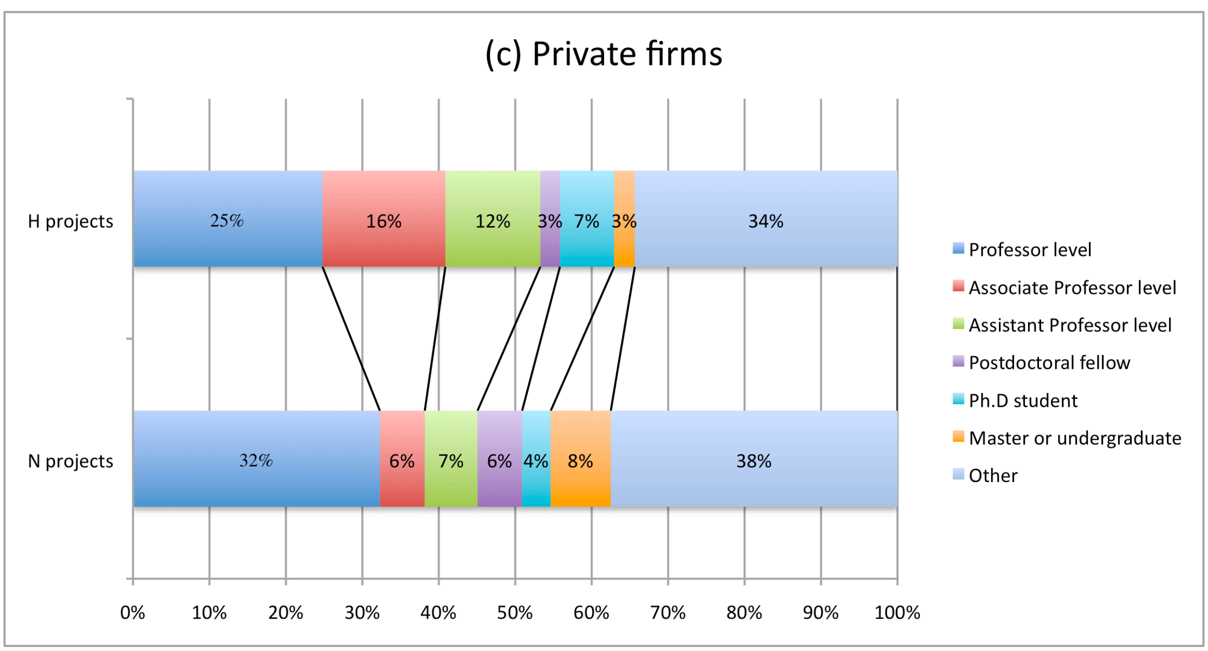
To further explore the division of labor in scientific teams, we asked our respondents to describe the characteristics of the authors, including themselves. For these items, we gave the respondents a matrix with the first author, last author, contact author, respondent, and additional authors (up to 6 total), and asked for the position (as well as additional information such as field, skill type, organization and gender) of each co-author. For cases where the paper had more than 6 authors, we randomly drew the additional authors (beyond first, last, contact and respondent). In the few cases where the respondent was not the first, last, or contact author and there were more than six authors total, we added a seventh author to the list (82 cases). However, for purposes of presentation, we limit the responses to those with 6 or fewer authors, which represent over 80% of our sample.

Exhibit 31 gives details on the composition of the research team by rank. Panel (a) gives the results for university respondents, panel (b) gives the results for government lab respondents, and panel

(c) gives the results for firm-centered projects. For university-centered projects, professors are the modal category, just under 40% for both H and N papers. The distribution by position is fairly similar across both groups, with the exception that post-docs are more common in H papers than in N (16% v. 11%), and master's students are more common in N than in H (5% v. 2%). For government-lab centered projects the distribution is similar. Again, we see post-docs are more common in H projects. For the firm-centered projects, the H projects are less likely to have senior level people than the N projects. We also see that many respondents chose "other" for firm respondents, suggesting that the categories may not have mapped well into the division of labor for industrial labs.

Exhibit 31. Composition of authors in academic/professional positions (a paper basis, by sector).



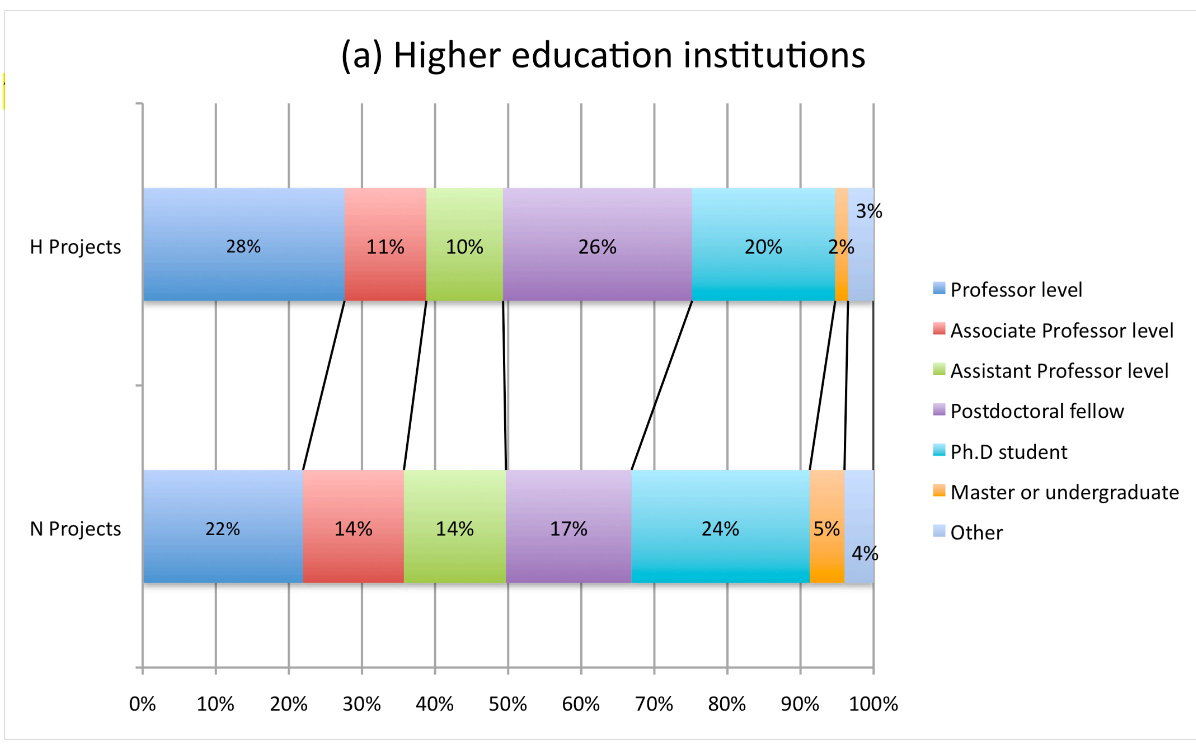


7-5 Who are the first authors when the authors are listed in order of contribution?

Name ordering conventions vary by field, by research group and by other factors related to the reward system in science. However, in general, the first author is considered an important position in the author list, except in cases where the names are listed alphabetically. The last author is also an important position, often reserved the lab head. In general, the first author is considered the one who made the major contribution to the design and execution of the study, and is often the one credited with the finding (as the Smith, et al. convention of citation implicitly implies). Thus, we want to see who the first authors are, and, in particular, what is the role of junior researchers (PhD students and post-docs) in H and N papers. Our survey asked the respondent to tell us how names were ordered in their paper, for example, alphabetically, by order of contribution, senior author first, senior author last. We selected the papers that listed authors by contribution (N=1396).

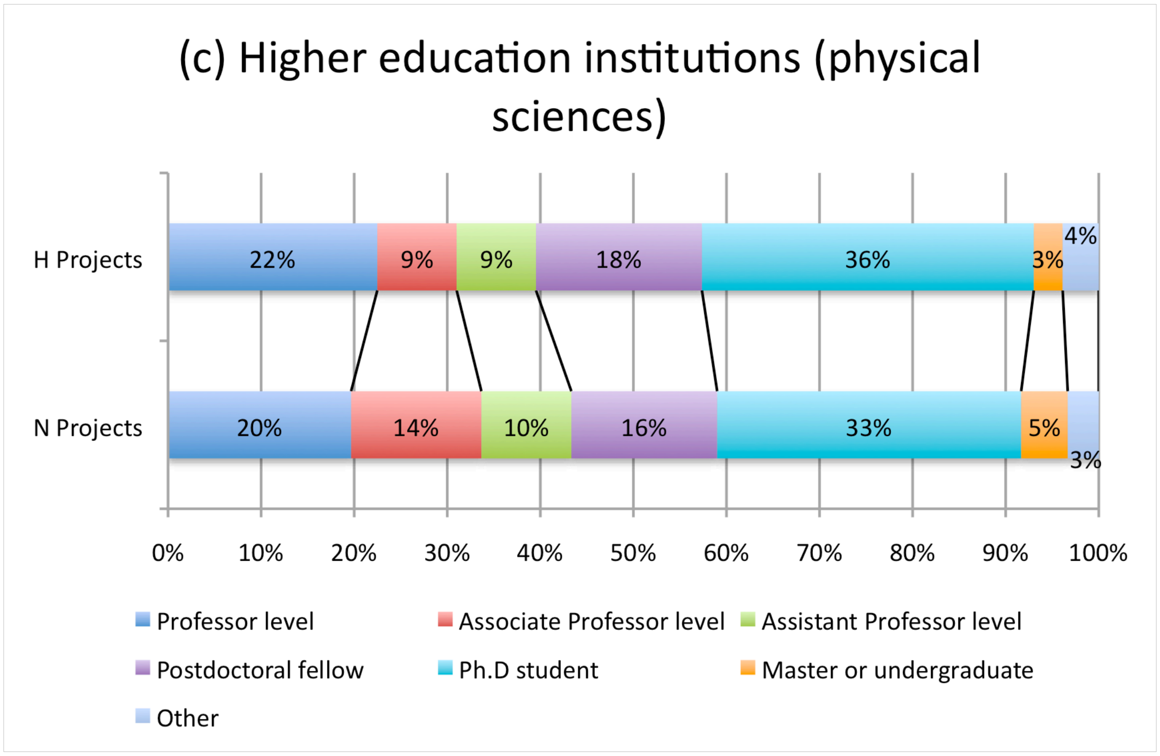
Exhibit 32(a) gives the results for university-centered projects. We can see that junior researchers (post-docs and PhD students) are over-represented as first authors (panel (b)). For example, for H papers, post-docs represent 26% of the first authors, v. 16% of all authors. Thus, junior researchers make a disproportionate contribution to scientific research, as measured by their likelihood of appearing as first author when authorship reflects contribution. Post-docs are especially prominent as first authors in life science fields compared to physical sciences (panels c and d). Fully half of first authors for life science papers from universities are post-docs. Post-docs are also heavily represented in public research institutions, representing 36% of first authors in H papers.

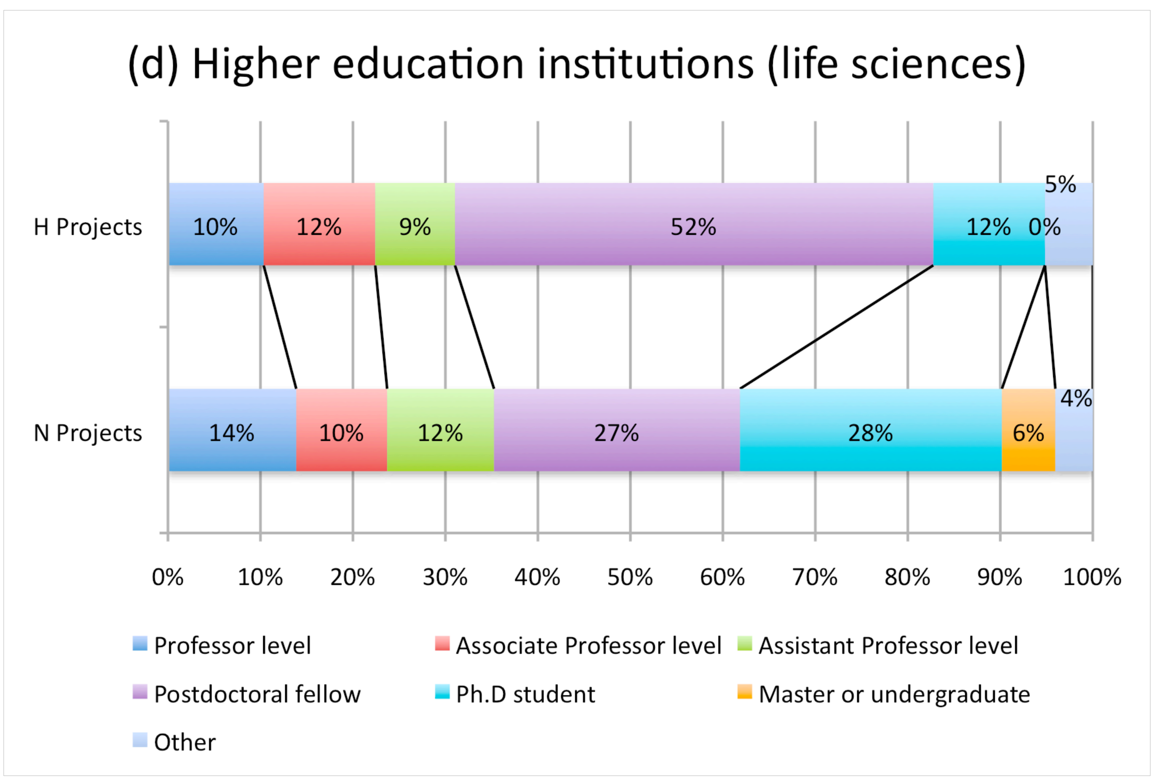
Exhibit 32 Academic/professional positions of the first authors in the focal papers whose authors are listed in order of their degree of contributions (by sector)

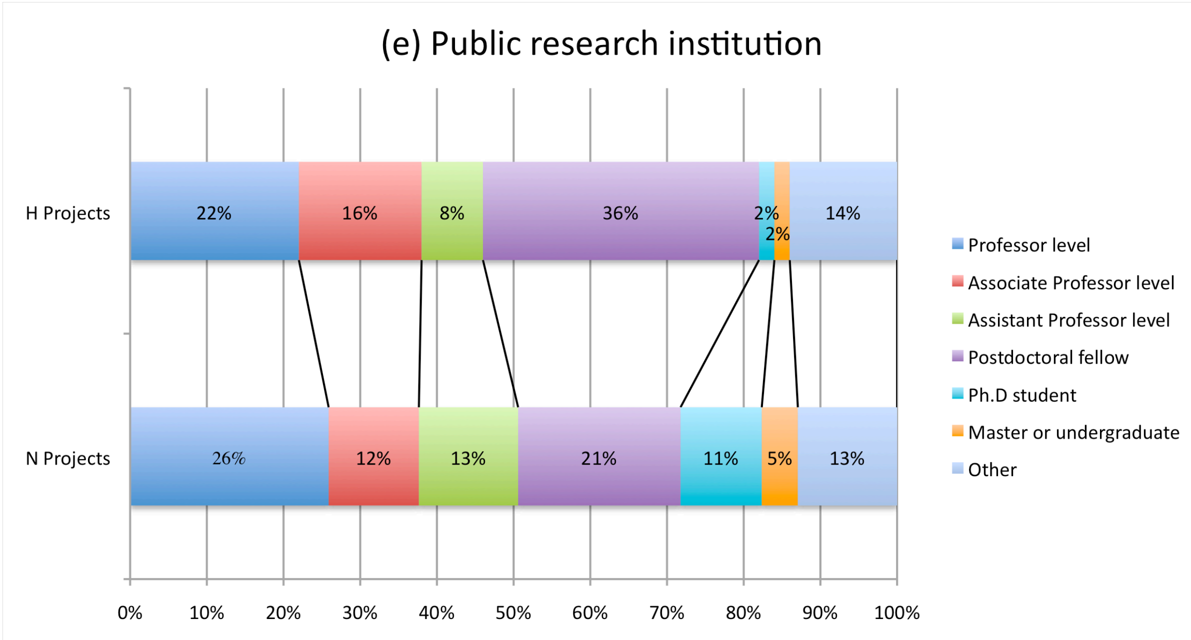


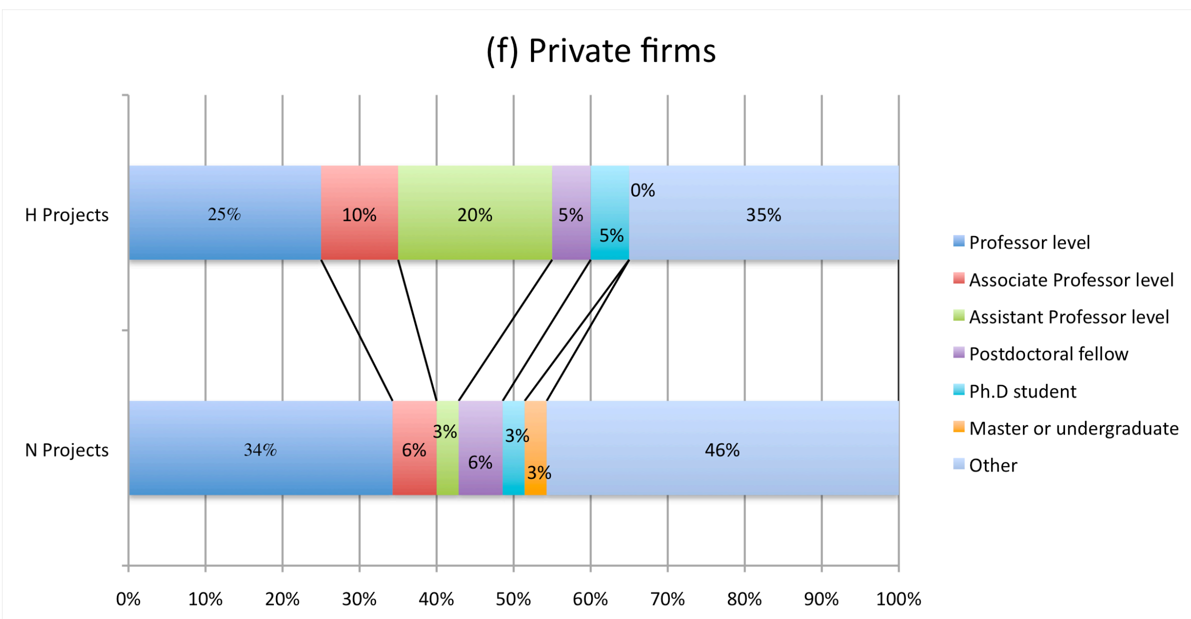
(b) PhD students and Post-doctoral fellow as a share of all and first authors, university publications.

		All			First Author	
	PhD	Post-doc	Sum	PhD	Post-doc	Sum
H papers	17%	16%	33%	20%	26%	46%
N papers	15%	11%	26%	24%	17%	41%





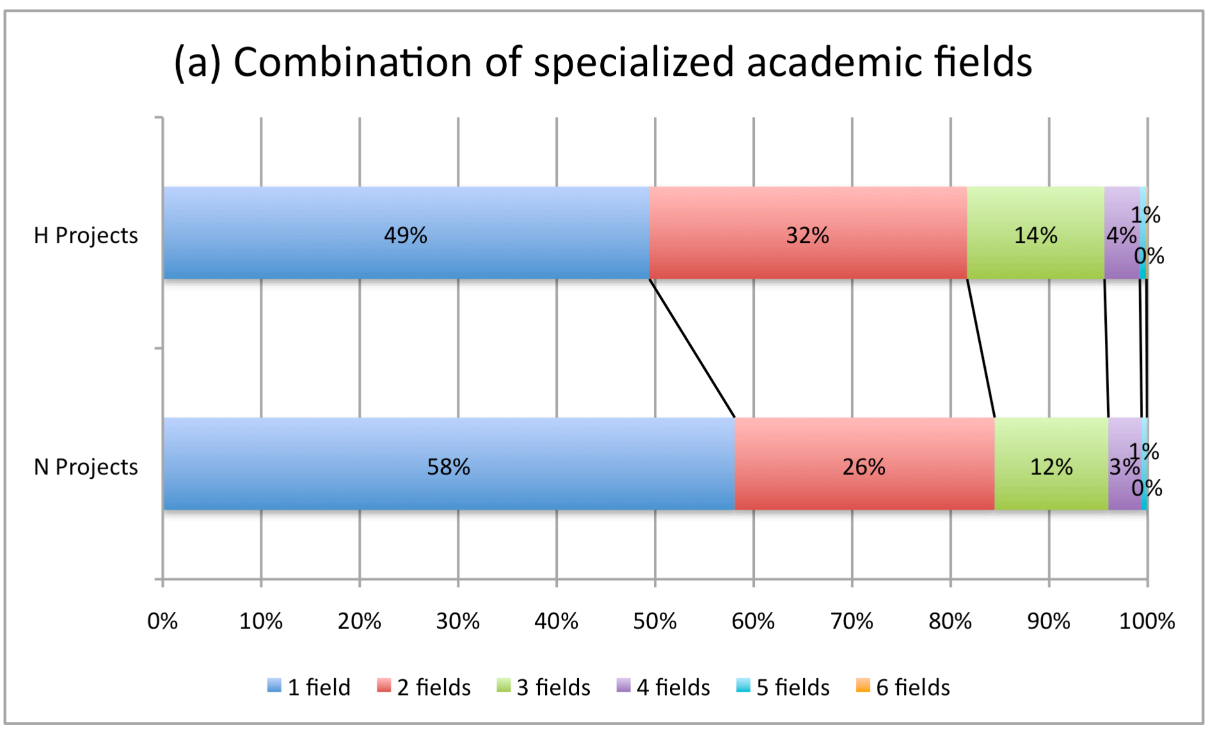

(e) Public research institution
H Projects
N Projects
22%
16%
8%
36%
2%
2%
14%
26%
12%
13%
21%
11%
5%
13%
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
Professor level
Associate Professor level
Assistant Professor level
Postdoctoral fellow
Ph.D student
Master or undergraduate
Other

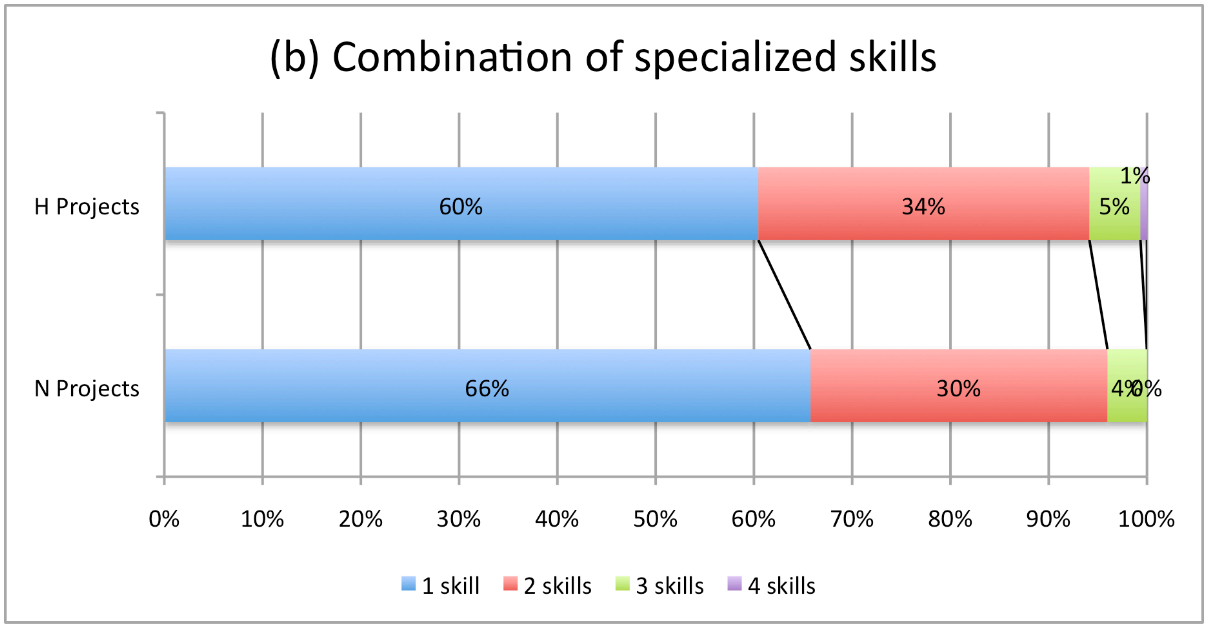

(f) Private firms
H Projects
N Projects
25%
10%
20%
5%
5%
0%
35%
34%
6%
3%
6%
3%
3%
46%
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
Professor level
Associate Professor level
Assistant Professor level
Postdoctoral fellow
Ph.D student
Master or undergraduate
Other

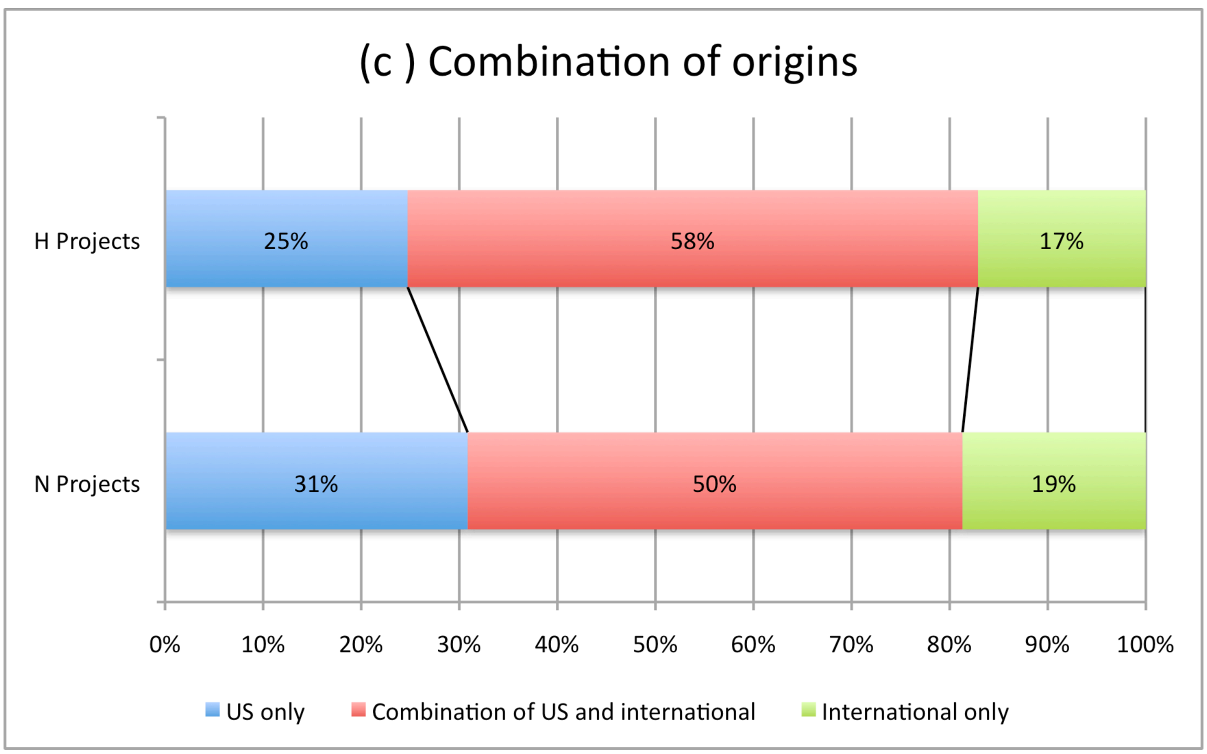
Diversity may be a key component of team performance (Hong and Page 2004). Teams may benefit from having access to a variety of problem solvers. We use our data on team composition to explore several dimensions of team diversity, including academic field, skills (theory, experimental, clinical), country of origin, and sector.

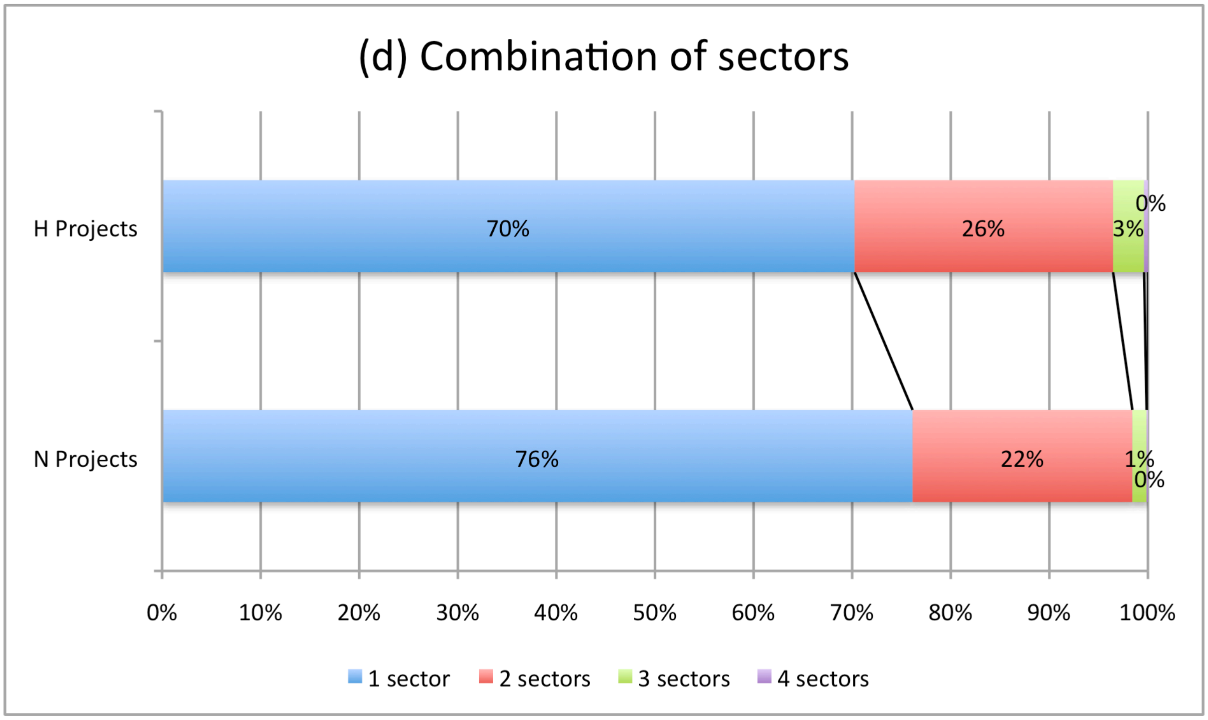
Exhibit 33 gives the results. We see that about half of H papers are multidisciplinary, as are just under half of N papers. In fact, almost 20% of H papers involved researchers representing 3 or more specialties. Thus, interdisciplinary research is common, and more common among the most highly cited papers. We also see that H papers are more likely to combine multiple skills (like theory and experimental skills). We also see that US science is embedded in a global network of scientists. In panel (c) we see that 70% of N papers and 75% of H papers have at least one international collaborators (either foreign born US-based scientists or overseas scientists). Finally, we see that the sector diversity is relatively low, compared to other dimensions of diversity, with 70% of H papers and 76% of N papers having authors all from the same sector.

Overall, across all dimensions of diversity, we find that H projects are more diverse than N projects, consistent with prior work suggesting that diversity is important for innovation (Hong and Page 2004; Walsh and Maloney 2002).




(b) Combination of specialized skills
H Projects
N Projects
60%
34%
5%
1%
66%
30%
4%
0%
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
1 skill
2 skills
3 skills
4 skills


(c) Combination of origins
H Projects
N Projects
25%
58%
17%
31%
50%
19%
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
US only
Combination of US and international
International only


(d) Combination of sectors
H Projects
N Projects
70%
26%
3%
0%
76%
22%
1%
0%
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
1 sector
2 sectors
3 sectors
4 sectors

8 INPUTS FOR RESEARCH PROJECTS

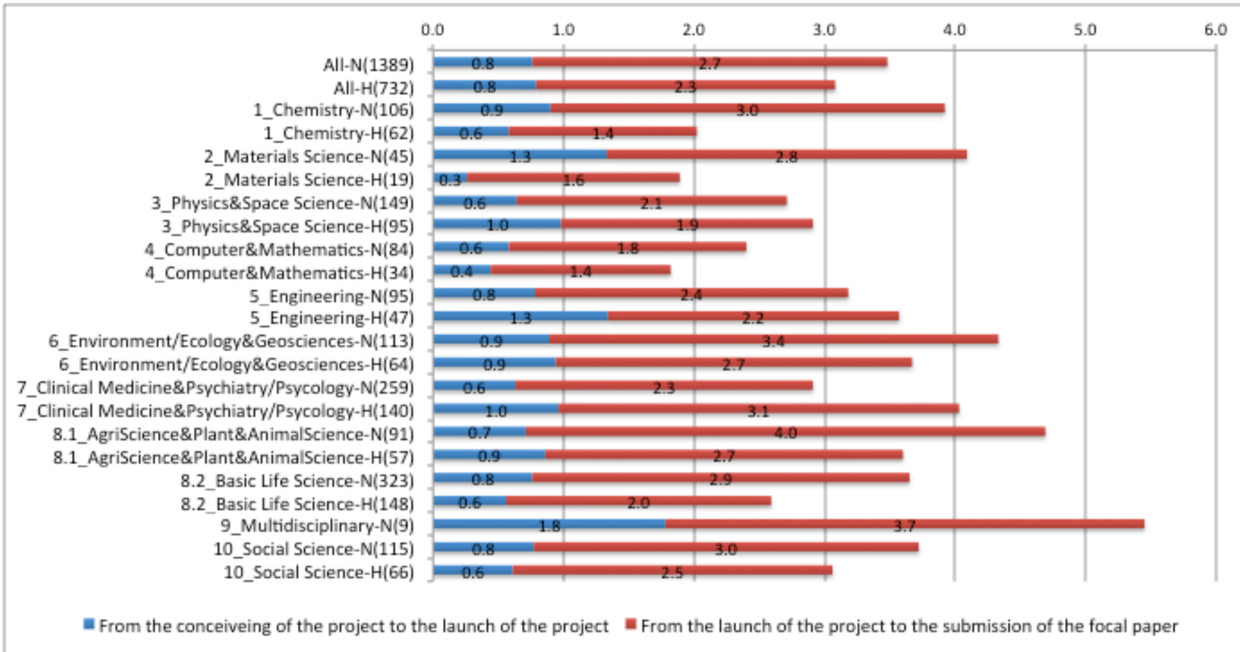
As Stephan points out in her review of the economics of science, scientific activity is generally conditions on access to resources (Stephan 2010). These resources include time, money, equipment and research assistants. In this section, we discuss the inputs to the research process and it it varies between H and N projects.

8-1 Time lag between research project conception and the submission of the focal paper

We begin with a discussion of the time devoted to a research project. The peer review system in the scientific community provides an invisible race among researchers to achieve priority in discoveries (Audretsch et al., 2002), encouraging rapid production and dissemination of research results. In this survey, we asked our respondents how many years it takes from the conception of the research project through the launch of the research project to the submission of the focal paper. Exhibit 34 shows that on average it takes around three years for H projects from having a project idea to finally submitting the paper, while for N papers, on average it takes three and half years, about half a year longer than H papers. However, when we break down to fields, for physics & space science and engineering, H papers took slightly longer than N paper between the year of conception of research project and the year of the submission of the focal paper.

The lag is especially short for projects in computer & mathematics, on average it takes half a year from having a project idea to actually launch the project and they submit the paper in another year and a half. However, when we compare between H and N projects across fields, we find that in many fields, the time lags of H papers are shorter than N papers. For example, in chemistry, the average time lag for H papers is 2 years but the time lag for N papers is almost 4 years. Basic life science is similar. Although the focal paper in this survey may not be the earliest paper from the research project, we still find some evidence that H papers are produced on a shorter time schedule. This is consistent with the results above showing that H papers have higher concern about competition. The only exceptions are physics & space science, engineering and clinical medicine, where we find that H papers took somewhat longer time than N paper between the year of conception of research project and the year of the submission of the focal paper.

Exhibit 34 Time lags between the year when the project was conceived and the year when the focal paper was submitted



0.0
1.0
2.0
3.0
4.0
5.0
6.0
All-N(1389) 0.8 2.7
All-H(732) 0.8 2.3
1_Chemistry-N(106) 0.9 3.0
1_Chemistry-H(62) 0.6 1.4
2_Materials Science-N(45) 1.3 2.8
2_Materials Science-H(19) 0.3 1.6
3_Physics&Space Science-N(149) 0.6 2.1
3_Physics&Space Science-H(95) 1.0 1.9
4_Computer&Mathematics-N(84) 0.6 1.8
4_Computer&Mathematics-H(34) 0.4 1.4
5_Engineering-N(95) 0.8 2.4
5_Engineering-H(47) 1.3 2.2
6_Environment/Ecology&Geosciences-N(113) 0.9 3.4
6_Environment/Ecology&Geosciences-H(64) 0.9 2.7
7_Clinical Medicine&Psychiatry/Psycology-N(259) 0.6 2.3
7_Clinical Medicine&Psychiatry/Psycology-H(140) 1.0 3.1
8.1_AgriScience&Plant&AnimalScience-N(91) 0.7 4.0
8.1_AgriScience&Plant&AnimalScience-H(57) 0.9 2.7
8.2_Basic Life Science-N(323) 0.8 2.9
8.2_Basic Life Science-H(148) 0.6 2.0
9_Multidisciplinary-N(9) 1.8 3.7
10_Social Science-N(115) 0.8 3.0
10_Social Science-H(66) 0.6 2.5
From the conceiveing of the project to the launch of the project
From the launch of the project to the submission of the focal paper

Labor input is one of the indicators describing the scale of the research project. We asked respondents to identify the approximate man-months that the entire research team spent, from the point when the research project began to the point when the most recent research finding were submitted for publication. Exhibit 35 shows the descriptive statistics of the labor input by (a) fields and (b) sector.

The median of labor input of the research project is 36 man-months for H projects and 24 man-months for N projects. Thus, H projects tend to be larger than N projects. The average labor input is 222 man-months for H projects and 192 man-months for N projects. If we break down by fields, we see that H projects in physics & space science and in clinical medicine &psychiatry/psychology and N projects in basic life science, the average labor inputs exceed 300 man-months. These data contain some outliers and so some caution should be used when interpreting means.

Exhibit 35. Total research man-months spent on the research project.
(a) By fields

Unit: Man-month

	N	Min	Q1	Median	Q3	Max	Mean
All-N	1531	0	12	24	55	120000	192.1
All-H	798	0	12	36	67	37500	221.7
1_Chemistry-N	118	1	10	24	50	1000	57.6
1_Chemistry-H	66	1.5	9	24	37	300	32.7
2_Materials Science-N	50	2	17	36	60	1000	87.4
2_Materials Science-H	22	6	15	45	60	336	59.2
3_Physics&Space Science-N	163	1	6	14.5	36	1200	54.4
3_Physics&Space Science-H	96	1	12	33	96	24000	339.3
4_Computer&Mathematics-N	92	0	7	12	48	960	49.0
4_Computer&Mathematics-H	39	3	8	20	40	100	28.2
5_Engineering-N	105	1	12	30	60	500	48.6
5_Engineering-H	57	3	20	36	150	2400	172.2
6_Environment/Ecology&Geosciences-N	125	2	15	34	60	600	61.4
6_Environment/Ecology&Geosciences-H	68	4	12	30	60	1200	94.3
7_Clinical Medicine & Psychiatry/Psychology-N	290	1	10	24	48	50000	278.7
7_Clinical Medicine & Psychiatry/Psychology-H	155	0	24	48	80	37500	570.7
8.1_AgriScience & Plant & Animal Science-N	97	1	12	27.5	60	10000	153.2
8.1_ AgriScience & Plant & Animal Science-H	60	0	24	40	81	1200	96.5
8.2_Basic Life Science-N	348	1	18	36	72	120000	448.8
8.2_Basic Life Science-H	159	4	20	36	84	3000	125.2
10_Social Science-N	132	1	6	18	48	800	42.1
10_Social Science-H	76	0	6	18	36	8568	178.6

(b) By sector

Unit: Man-month

	N	Min	Q1	Median	Q3	Max	Mean
All-N	1325	0	12	24	56	120000	209.2
All-H	700	0	12	36	66	37500	194.7
University&college-N	1060	0	12	24	56	10000	73.3
University&college-H	536	0	12	30	60	15000	145.0
Government research organization-N	129	1	12	24	60	1800	89.1
Government research organization-H	83	1	20	48	100	1200	133.6
Private firm-N	48	1	6	12	42.5	50000	1184.3
Private firm-H	29	1	13.5	32	55	730	77.8
Non-profit organization-N	41	12	27	40	87.5	120000	3415.6
Non-profit organization-H	33	2	19	48	100	1200	106.7

8-3 Research project budgets

We also measured the size of the project by the research budget. In the US questionnaire of the survey, we asked our respondents to report the approximately amount of funds directly used for the research project. The research fund should include personnel costs for researchers and supporting personnel, if it was part of the grant budget. The question was designed as a multiple-choice question, with 15 options ranging from less than $10,000 to more than $100,000,000. We recoded the categorical answers to the median of each range (using $125,000,000 for the top category). Using the recoded new variable, we ran the descriptive statistics as shown in Exhibit 36. Again, because of outliers, we should be cautious when interpreting means.

On average, the amount of research funds spent for H projects is about $4,400,000, which is significantly higher than the amount of research funds spent for N papers (about $1,800,000). Similarly, the median H project had a budget of $175,000 while the median N project had a budget of only $55,000. We also see the mean of research funds vary across fields. H projects in physics and space science and in clinical medicine had average budgets over $7 million, about four times more than N projects in those fields. The median H project in material science had a budget of $375,000.

By sector, we find that government research organizations spent the most research funds, especially for H projects. In addition, among all the H projects, universities and colleges spent the least research funds with an average of about $1,700,000.

Exhibit 36 Amount of money directly used for the research project.
(a) By field

Unit: $1,000

	N	Min	P25	Median	P75	Max	Mean
All-N	1531	5	55	55	175	125000	1801.06
All-H	798	5	55	175	625	125000	4352
1_Chemistry-N	118	5	55	55	175	125000	2486.92
1_Chemistry-H	66	5	55	175	375	62500	1261.38
2_Materials Science-N	50	5	55	175	375	8750	540.65
2_Materials Science-H	22	55	175	375	625	6250	740.68
3_Physics&Space Science-N	163	5	55	55	175	125000	1852.64
3_Physics&Space Science-H	96	5	55	175	375	125000	7627.37
4_Computer&Mathematics-N	92	5	5	55	175	62500	1015.17
4_Computer&Mathematics-H	39	5	55	55	175	875	156.35
5_Engineering-N	105	5	55	55	175	125000	2360.74
5_Engineering-H	57	5	55	175	1750	125000	6228.77
6_Environment/Ecology&Geosciences-N	125	5	55	175	375	87500	2111.06
6_Environment/Ecology&Geosciences-H	68	5	55	175	375	125000	4673.66
7_Clinical Medicine&Psychiatry/Psycology-N	290	5	5	55	175	125000	1418.53
7_Clinical Medicine&Psychiatry/Psycology-H	155	5	55	175	1750	125000	7342.29
8.1_AgriScience&Plant&AnimalScience-N	97	5	55	55	175	87500	1280
8.1_AgriScience&Plant&AnimalScience-H	60	5	55	175	625	6250	672.76
8.2_Basic Life Science-N	348	5	55	175	375	125000	2593.18
8.2_Basic Life Science-H	159	5	55	175	625	125000	3863.32
9_Multidisciplinary-N	11	5	5	30	375	3750	488.5
10_Social Science-N	132	5	5	55	175	37500	637.64
10_Social Science-H	76	5	5	55	175	125000	2358.71

(b) By sector

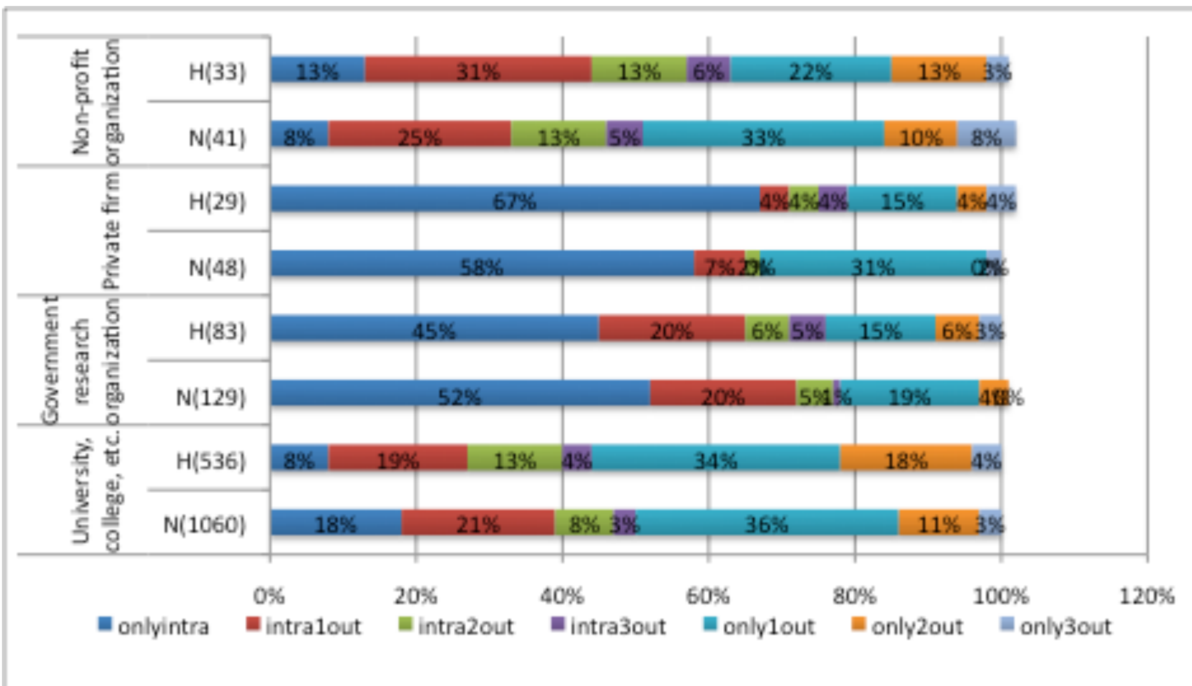
Unit: $1,000

	N	Min	P25	Median	P75	Max	Mean
All-N	1325	5	55	55	175	125000	1741.74
All-H	700	5	55	175	625	125000	4036.81
Universit&college-N	1060	5	55	55	175	125000	1746.83
Universit&college-H	536	5	55	175	375	125000	2723.18
Government research organization-N	129	5	55	175	375	17500	807.94
Government research organization-H	83	5	55	375	1750	125000	12169.94
Private firm-N	48	5	55	55	625	17500	1201.06
Private firm-H	29	5	55	175	1750	62500	4174.82
Non-profit organization-N	41	55	175	175	875	125000	7010.13
Non-profit organization-H	33	55	55	175	625	125000	4906.06
Hospital-N	40	5	5	55	175	3750	215.26
Hospital-N	16	55	55	175	1187.5	37500	4076.88

8-4-1 Combination of multiple sources of funds

In this survey, we asked respondents to indicate the approximate percentage of total project funding that came from different sources. One major finding is that project funding tends to come from multiple sources. Exhibit 37 shows the frequency of the combinations of multiple sources of funds for research projects. We divide sources of funding into intramural[2] funds and extramural funds. For university researchers, 8% of H projects use only the intramural funds and 18% of N projects use only the intramural funds. On the other hand, about 50% of N projects and 56% of H projects only use extramural funds. Thus, we can see that university research is heavily dependent on outside funding, and that is especially true for the H projects. In contrast, for government labs researchers, 45% of H projects use only intramural funds and 52% of H projects use only the intramural funds. Around 60% of research projects from private firms use only intramural funds.

Exhibit 37. Combination of sources of funds



8-4-2

Disaggregated sources of funds

To further understand the research funding picture in the US, we asked respondents to tell us the sources of funding by agency and/or organization type (domestic firm, foreign firm, foundation, etc.). Exhibit 38 summarized the results. Here, we show the percent with any internal funding, any industry funding, and any funding from National Institutes of Health (NIH), National Science Foundation (NSF), Department of Defense (DOD) and Department of Energy (DOE). For this analysis, we limit the results to university researchers.

We see that NIH and NSF are the dominant outside sources for university researchers, with approximately 30% of H projects getting some funding from each source. We also see that a large percent of projects get some internal funds, although as we note above, the number with only internal funding is quite small. About 10% of university researchers get some industry funding. We also see that

[2] Intramural funds means that the funding comes the institutions that the research team members belong to.

DOD and DOE are major funders in some fields. Department of Energy is a major funder for H projects in physics (19%), chemistry (15%) and material science (14%). Department of Defense funds many H projects in material science (43%), engineering (34%) and chemistry (26%).

Exhibit 38. Percent with any funding from various sources, by field.

	Some Internal		Any Industry		NIH		NSF		DOE		DOD	
	H	N	H	N	H	N	H	N	H	N	H	N
1_Chemistry	35.2	45.1	16.7	13.2	16.7	18.7	51.9	36.3	14.8	11.0	25.9	7.7
2_Materials Science	35.7	32.4	14.3	26.5	0.0	2.9	50.0	20.6	14.3	2.9	42.9	17.6
3_Physics&Space Science	40.3	35.7	3.2	7.0	1.6	5.2	41.9	40.0	19.4	9.6	12.9	10.4
4_Computer & Mathematics	37.9	33.8	3.4	7.8	31.0	3.9	34.5	28.6	3.4	2.6	3.4	10.4
5_Engineering	34.3	49.3	17.1	8.5	5.7	5.6	25.7	11.3	11.4	5.6	34.3	5.6
6_Environment/Ecology & Geosciences	47.8	40.0	2.2	5.9	8.7	3.5	63.0	36.5	8.7	8.2	2.2	2.4
7_Clinical Medicine & Psychiatry/Psychology	40.2	52.3	19.6	10.2	57.7	42.0	1.0	4.5	0.0	0.6	1.0	2.8
8.1_AgriScience & Plant & Animal Science	55.3	69.2	5.3	21.5	21.1	1.5	47.4	9.2	13.2	1.5	0.0	0.0
8.2_Basic Life Science	31.7	36.9	5.9	8.3	72.3	59.8	10.9	12.4	0.0	2.1	1.0	4.6
ALL	40.1	44.4	9.7	9.8	31.7	24.5	27.6	19.0	6.7	4.1	8.4	5.2

9 OUTPUTS AND IMPACTS OF THE RESEARCH PROJECTS

9-1 Number of refereed papers from the research project

While the survey focused on a specific publication, we also asked about how many other papers the research project produced. As shown in Exhibit 39, for US-based authors, nearly all of the refereed papers from research projects are written in English regardless of the type of projects (96% for H projects and 97% for the N projects).

When we look at the number of papers produced by the projects, we see that H projects produced a median of 7.5 papers and 24.1 papers on average across all fields (Exhibit 40). N projects produced fewer papers, with a median of 3 papers and mean of 5.8 papers. We can observe that H projects not only produced a higher impact paper, but also produced more refereed papers compared to N projects. Environmental science H projects produced the most papers, with a median of 20.5 papers.

Exhibit 39. Distribution of the refereed papers from a research project by language

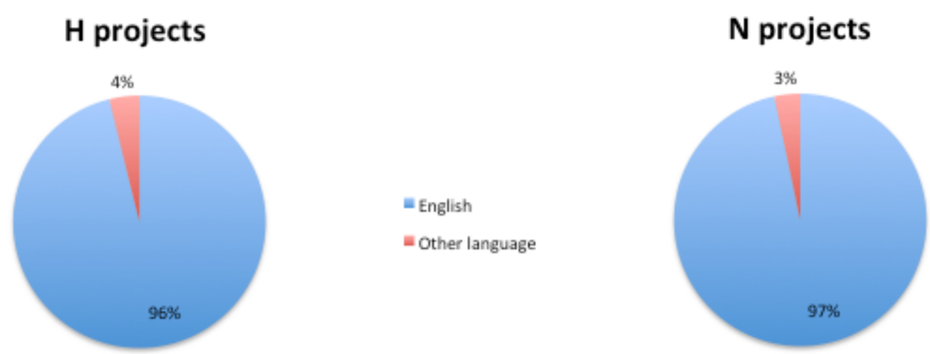


Exhibit 40. Distribution of the number of refereed papers yielded from research project

(a) Distribution of the number of refereed paper by field

		N	Min	Q1	Median	Q3	Max	Mean
All	H projects	798	1	2	7.5	20	400	24.1
	N projects	1531	1	1	3	6	85	5.8
1_Chemistry	H projects	66	1	2	3.5	6	25	6.8
	N projects	118	1	1	4.5	8	35	7.8
2_Materials Science	H projects	22	2	2	2	2	2	2.0
	N projects	50	1	3	5.5	7	26	7.1
3_Physics&Space Science	H projects	96	1	4	9	13	400	36.9
	N projects	163	1	2	3	8	85	8.1
4_Computer&Mathematics	H projects	39	1	1	4	10	30	7.9
	N projects	92	1	1	2	3	20	3.5
5_Engineering	H projects	57	4	6	10	17	53	15.4
	N projects	105	1	2.5	4	9.5	37	7.8
6_Environment/Ecology & Geosciences	H projects	68	2	10	20.5	80	300	66.6
	N projects	125	1	2	4.5	13	34	8.5
7_Clinical Medicine & Psychiatry/Psychology	H projects	155	1	2	6	20	101	14.7
	N projects	290	1	1	1	3	40	4.4
8.1_Agriculture Science & Plant & Animal Science	H projects	60	1	1	8.5	16	16	8.5
	N projects	97	1	1	2	4	11	3.5
8.2_Basic Life Science	H projects	159	1	4	7	17	200	22.9
	N projects	348	1	1	4	6	40	6.4
9_Multidisciplinary	N projects	11	8	8	9.5	11	11	9.5
10_Social Science	H projects	76	1	1	3	20	200	23.9
	N projects	132	1	1	1.5	4	20	3.0

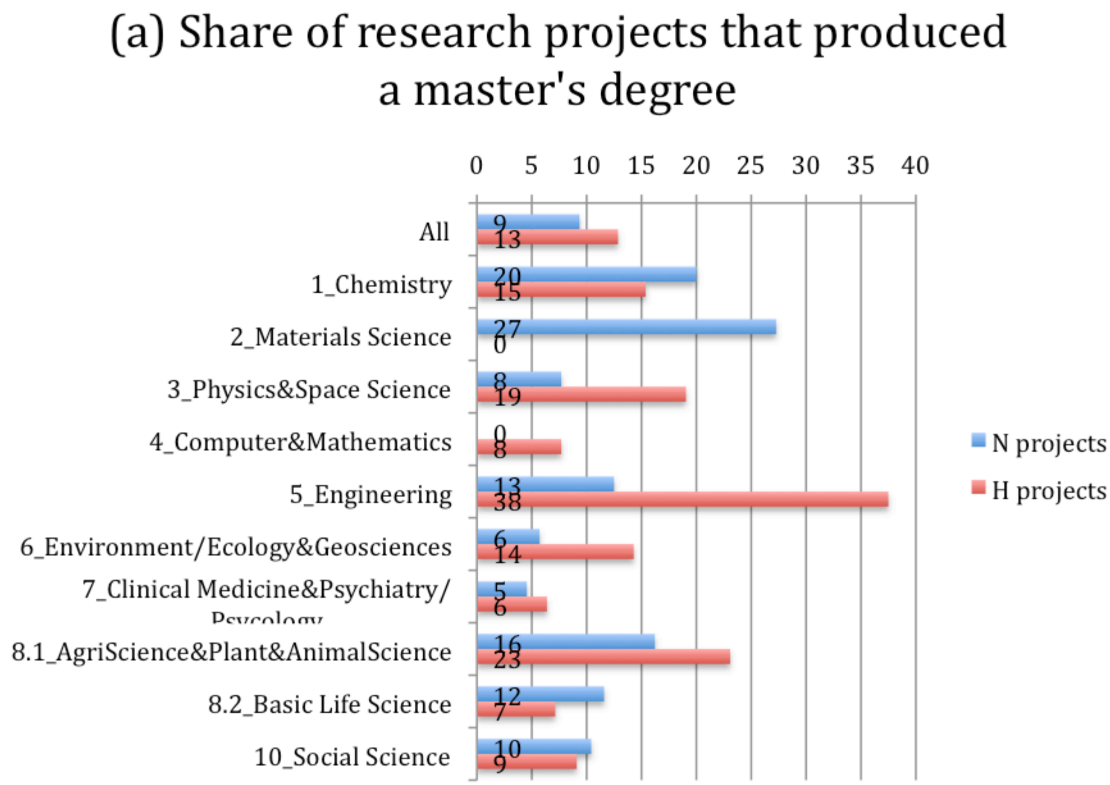
(b) Distribution of the number of refereed paper by sector

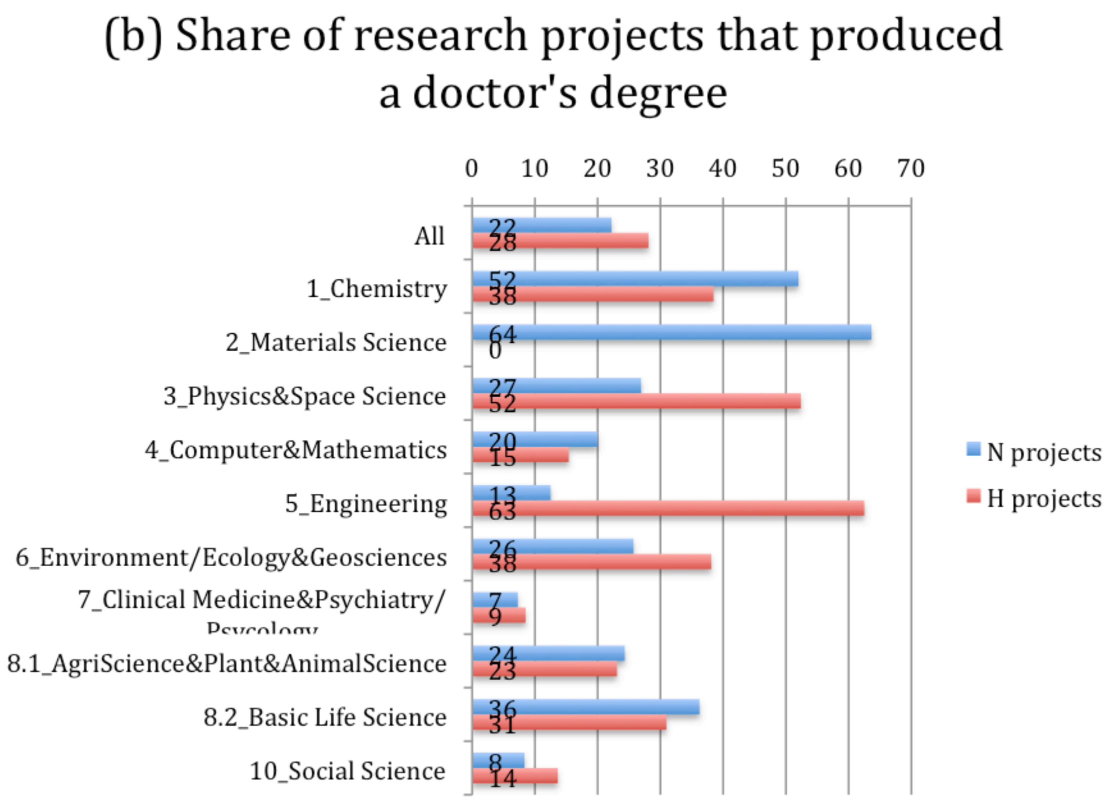
		N	Min	Q1	Median	Q3	Max	Mean
All	H projects	798	1	2	7.5	20	400	24.1
	N projects	1531	1	1	3	6	85	5.8
University	H projects	617	1	2	6	20	300	23.1
	N projects	1193	1	1	3	6	40	5.2
Government Lab	H projects	82	1	3	8	24	400	45.6
	N projects	135	1	2	4	6	37	7.6
Private Firm	H projects	33	1	10	10	30	40	18.2
	N projects	109	1	1	1	8	35	6.3

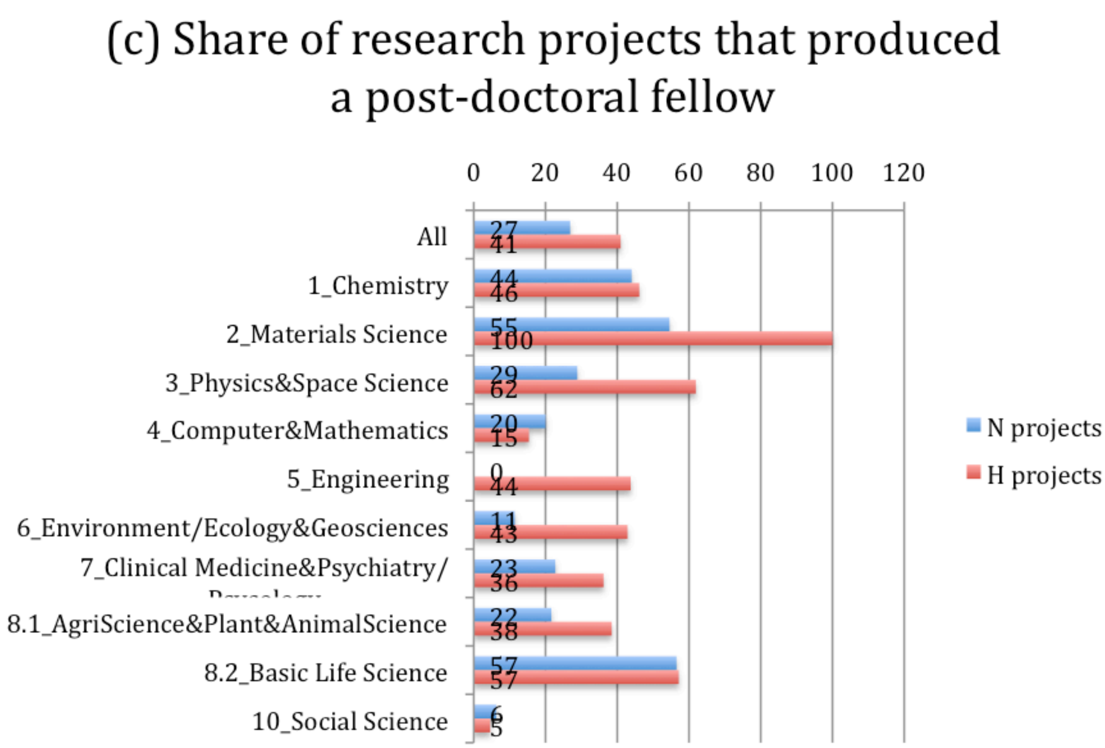
9-2 Graduate students trained through the research project

A key output of research is trained personnel. Therefore, we asked our respondents how many students received their master's degrees, Ph.D. or post-doctoral training through the research project (Exhibit 41). We can observe that the type of training associated with a project varies significantly across fields and project types. For example, N projects produce more master's and doctoral degrees in fields like chemistry and material science; on the other hand, H projects are more likely to produce all types of degrees and post-doctoral fellows in fields like physics and engineering. Materials science H projects in our sample produced no master's students, but all of them included post-doctoral fellows. In contrast, in engineering, none of the N projects in our sample generated post-doctoral fellows.

Exhibit 41. Students and post-doctoral fellows trained as part of the research project.




(b) Share of research projects that produced a doctor's degree
0
10
20
30
40
50
60
70
All
1_Chemistry
2_Materials Science
3_Physics&Space Science
4_Computer&Mathematics
5_Engineering
6_Environment/Ecology&Geosciences
7_Clinical Medicine&Psychiatry/
8.1_AgriScience&Plant&AnimalScience
8.2_Basic Life Science
10_Social Science
22
28
52
38
64
0
27
52
20
15
13
63
26
38
7
9
24
23
36
31
8
14
N projects
H projects


(c) Share of research projects that produced a post-doctoral fellow
0
20
40
60
80
100
120
All
1_Chemistry
2_Materials Science
3_Physics&Space Science
4_Computer&Mathematics
5_Engineering
6_Environment/Ecology&Geosciences
7_Clinical Medicine&Psychiatry/
8.1_AgriScience&Plant&AnimalScience
8.2_Basic Life Science
10_Social Science
27
41
44
46
55
100
29
62
20
15
0
44
11
43
23
36
22
38
57
57
6
5
N projects
H projects

9-3 Patent applications and license agreements.

In addition to scientific outputs and trained personnel, universities and government labs are increasingly called on to commercialize their research. While there is substantial research on the rates of commercialization by universities (such as in the annual AUTM reports), and individuals (Bercovitz and Feldman 2008; Ding, Murray, and Stuart 2006; Stuart and Ding 2006), there is little systematic data on the likelihood of commercialization at the project level. Do address this gap, we collected detailed information on the commercialization of the results of the surveyed projects, including patenting, licensing and startups.

We begin with a discussion of patenting. We collected project-level data on patenting and licensing of inventions related to the research project. We begin by asking respondents if the findings from the research project lead to a patent application. Exhibit 42 below gives the results. We find that 7 percent of the N projects and 14% of the H projects resulted in at least one patent application. The totals are similar if we only consider respondents from universities. Papers from firms are especially likely to lead to patents, not surprisingly. However, perhaps more surprising is that over half of the highly cited papers by firms are patented, while only 20% of the random papers are also patented. Thus, paper-patent pairs are especially likely to involve highly cited papers, particularly in the case of firms. The only exception is papers by NPOs, where the random sample papers are more likely to be patented.

Exhibit 42. Patent applications by sector.

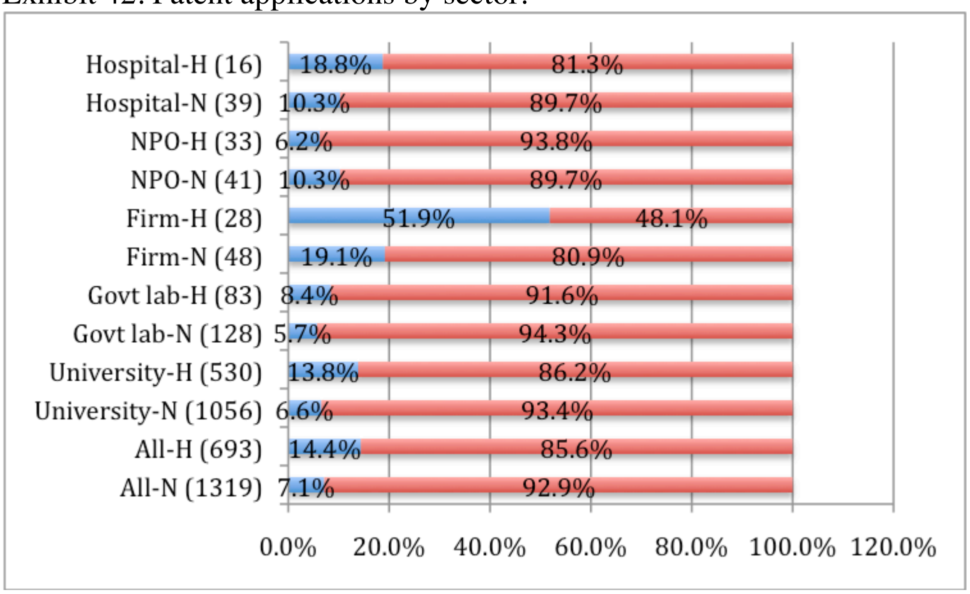
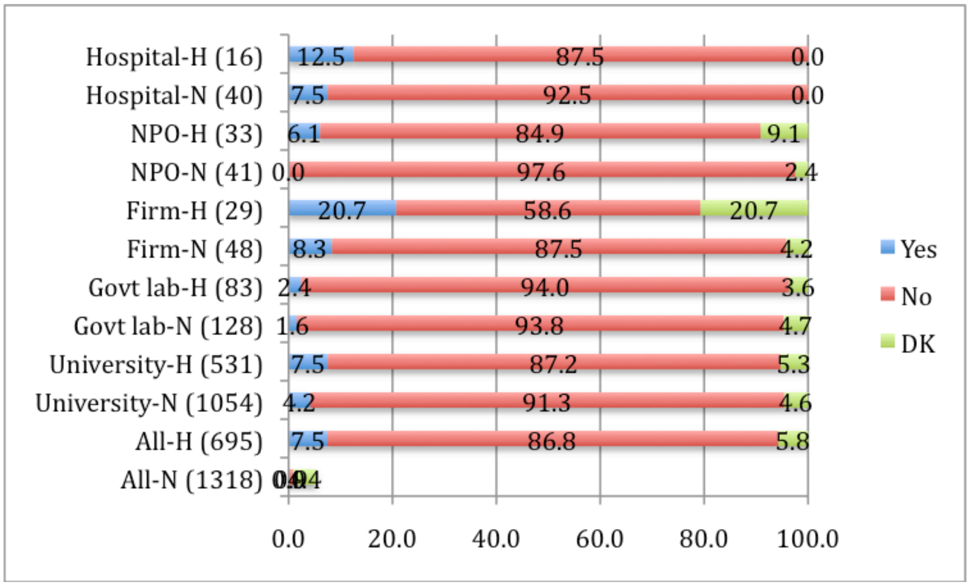


Exhibit 43. Licensing of patents by sector.



Note: All includes “other” (but not missing on sector).

For the highly cited papers, when we compare across fields, we find high rates of patenting (around 30%) in chemistry, materials science and engineering, consistent with prior work. Life science related fields have moderate rates of patenting, with about 15% of projects resulting in a patent. Physical sciences, mathematics and computing and geosciences all have rates of patenting under 10% of project. These results are consistent with prior work showing that materials science and chemistry were the most broadly applicable fields of science for industrial R&D (Cohen, Nelson, and Walsh 2002). In every case, the percent of patents generated by the N papers is substantially lower. For the N papers, the life sciences do relatively better, with patenting rates equivalent to chemistry and engineering, at around 10%. This suggests that life science faculty (or their universities) are more willing to patent less important findings in life sciences than in chemistry or engineering, perhaps because patents (and the associated inventions) may have more value in these fields, even if the underlying science is not significant (cf. (Cohen, Nelson, and Walsh 2000). About a quarter of patents associated with H papers were applied for overseas, as were about 20% of N paper patents. Basic life science patents were especially likely to be filed oversees, with about one third of both kinds of projects' patents including an overseas patents. Again, since many of these are biotech or pharmaceutical related patents, global protection may be seen as especially critical to ensure favorable licensing terms.

When we look at patent licensing, we see that H papers are also more likely to have their patents licensed. We also see that licensing rates are somewhat higher in universities than in government labs. Firms, of course, have higher rates of licensing. We also see that university, government lab and NPO respondents are more likely to know if their patents were licensed than were industrial scientists. This suggests that licensing may be a more significant event for a public researcher, and/or they may be more involved in the licensing process.

Exhibit 44. Patent applications and foreign applications, by field.

	Number of projects that resulted in patent applications		Percent of projects that resulted in a patent application		Percent of applications that include a foreign application	
	H	N	H	N	H	N
ALL	107	104	13.9	7.0	26.2	20.2
1_Chemistry	18	11	28.6	9.6	16.7	0.0
2_Materials Science	7	7	31.8	15.2	28.6	14.3
3_Physics&Space Science	8	11	8.5	7.0	37.5	18.2
4_Computer&Mathematics	3	3	7.7	3.4	33.3	0.0
5_Engineering	16	7	30.8	6.9	18.8	14.3
6_Environment/Ecology&Geosciences	2	2	3.1	1.7	0.0	0.0
7_Clinical Medicine & Psychiatry/Psychology	19	23	12.8	8.3	31.6	17.4
8.1_AgriScience&Plant&AnimalScience	10	6	16.7	6.3	20.0	16.7
8.2_Basic Life Science	24	34	15.8	10.0	33.3	35.3
N("Yes")	107	104	107	104	28	21
N	771	1480	771	1480	107	104

Exhibit 45. License agreements and the provisions of know-how, by field.

	Number of projects that included licensed technology		Percent of projects that included licensed technology		Percent of licenses that included providing know-how	
	H	N	H	N	H	N
ALL	58	58	7.8	4.0	77.4	55.1
1_Chemistry	9	3	14.8	2.8	62.5	66.7
2_Materials Science	3	3	15.8	7.3	100.0	66.7
3_Physics&Space Science	2	4	2.2	2.7	50.0	33.3
4_Computer&Mathematics	3	5	8.1	5.7	33.3	50.0
5_Engineering	7	4	14.9	4.1	85.7	50.0
6_Environment/Ecology&Geosciences	0	2	0.0	1.7	--	100.0
7_Clinical Medicine &Psychiatry/Psychology	12	16	8.2	5.8	90.9	64.3
8.1_AgriScience&Plant&AnimalScience	5	6	8.3	6.3	20.0	40.0
8.2_Basic Life Science	15	15	10.0	4.5	100.0	50.0
N("Yes"_	58	58	58	58	41	27
N	747	1442	747	1442	53	49

We can see that license agreements are also more common among the H papers, although the yield rate (ratio of licenses to patents) is similar across the two groups (just under 60%). This high yield rate suggests that university patenting may be heavily conditioned by the existence of a licensee, especially among the random sample (where the yield rate for university patents is 64%), (Pressman, Burgess, Cook-Deegan, McCormack, Nami-Wolk, Soucy, and Walters 2006). Licensing is most common in materials science, chemistry and engineering (which also had high rates of patenting), with basic life science also having above average licensing. Consistent with prior work, the majority of these licenses (especially among the top papers) including providing know how (Thursby and Thursby 1999).

We also asked our respondents if they had founded a startup based on the research results. This question was not conditioned on patenting or licensing. In addition, we asked if they seriously considered founding a startup, a broader measure of commercial interests.

Exhibit 46. Percent of respondents who had founded a startup based on the research project, by sector

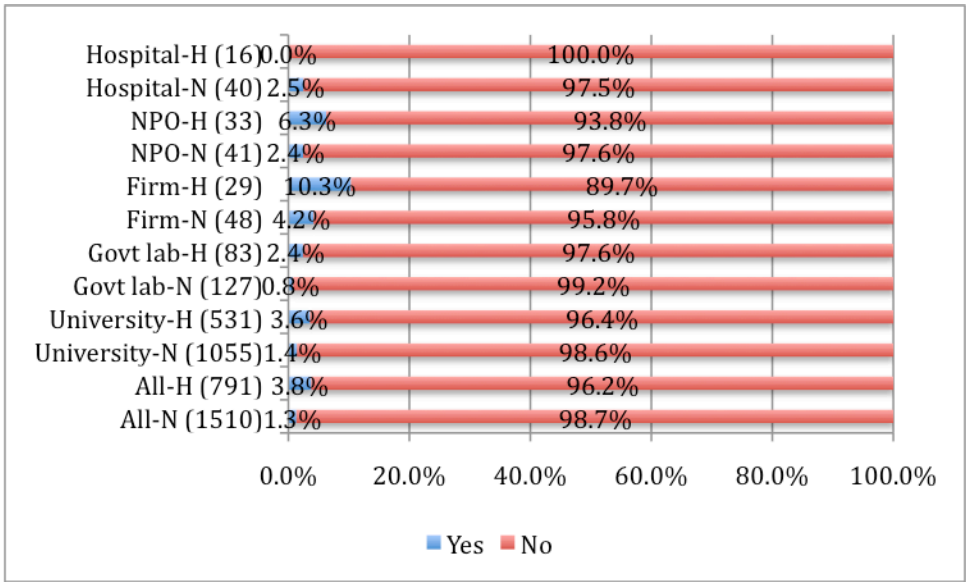
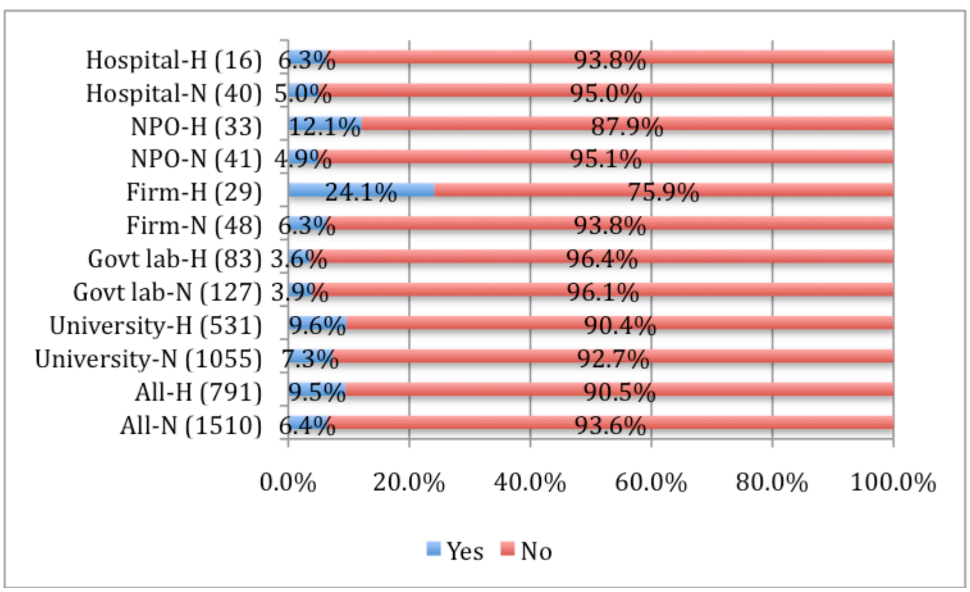
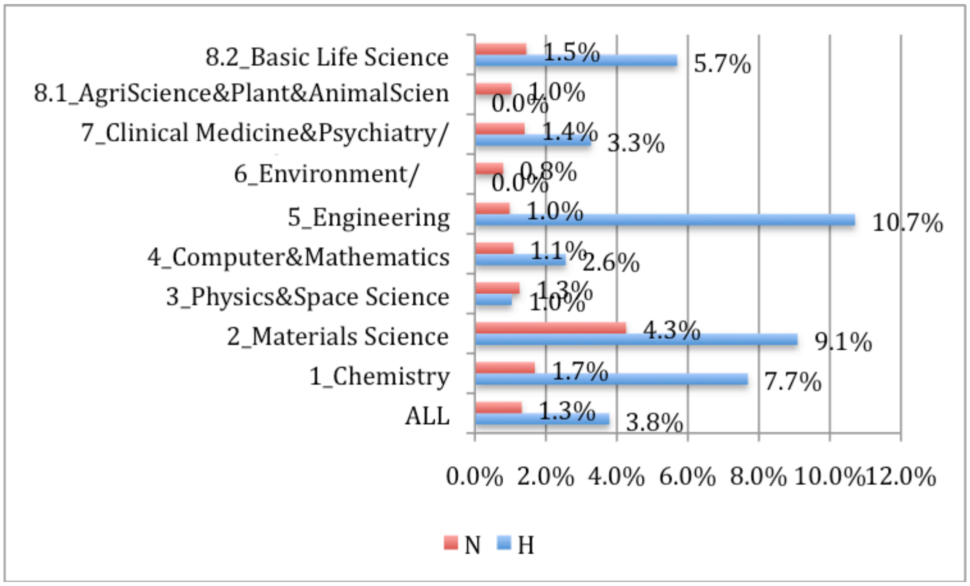


Exhibit 47. Percent of respondents who had either founded or seriously considered founding a startup based on the research project, by sector



We can see that startups are fairly rare, on a project-level, with about 4% of H papers and 1% of N papers producing a startup, overall, and for university respondents (Exhibit 46). Firm respondents have higher rates, with over 10% of H papers by firms associated with a startup. If we broaden the definition of startup activity to include those who seriously considered a startup (Exhibit 47), we find that almost 10% of H projects and 6% of N projects either did or seriously considered a startup, the rates similar for university respondents.

Exhibit 48. Startups founded by field



We can see that, for H papers, startups are most common in engineering, materials science and chemistry, with basic life science also have above average rates of startups among top papers. For the N sample, only materials science has significantly higher rates of startups.

Exhibit 49 summarizes the outputs of the research projects in our sample. First, we can see (by the last column of the table) that H projects are almost always more productive than N projects, on nearly every dimension, with MS degrees being the only exception. We can also see that training researchers is the major output of the research (besides the publication itself). We also see that foreign born personnel are widely participating in US science. In particular, projects are especially likely to train foreign-born (compared to US born) post-doctoral fellows, and that this is even more true for the H projects. Two-thirds of H projects including a foreign-born post-doctoral fellow. Projects are also more likely to have foreign-born PhD students than US born. However, US born master's students are more common than are foreign born. In addition, we see that a significant fraction (over 40%) of projects hired people specifically for the project, which suggests that another important impact of scientific research is job creation. In fact, one justification for substantial research funding in the recent economic stimulus package in the US was that spending money on research would generate jobs. The gap between H and N projects is especially sharp for commercialization, with H projects producing patents, licenses and startups at 2-3 times the rate of N projects.

Exhibit 49. Summary of research outputs

	H Project			N Projects			
	Projects producing this output	N	Ratio(a)	Projects producing this output	N	Ratio(b)	(a)/(b)
Refereed papers	798	798	100%	1531	1531	100%	1.0
PhD recipients (All)	390	584	67%	656	1115	59%	1.1
PhD recipients (US born)	247	457	54%	367	861	43%	1.3
PhD recipients (Foreign born)	232	409	57%	398	788	51%	1.1
Post-doc fellows (All)	384	550	70%	543	981	55%	1.3
Post-doc fellows (US born)	186	398	47%	260	754	34%	1.4
Post-doc fellows (Foreign born)	286	433	66%	378	732	52%	1.3
MS recipients (All)	124	369	34%	289	812	36%	0.9
MS recipients (US born)	84	331	25%	201	726	28%	0.9
MS recipients (Foreign born)	53	261	20%	125	548	23%	0.9
Hired researcher	317	699	45%	576	1389	41%	1.1
Patent applications	107	771	14%	104	1480	7%	2.0
Licensing	58	747	8%	58	1442	4%	1.9
Startup firm	30	791	4%	20	1510	1%	2.9

10 CONCLUSIONS

Based on a survey of scientists associated with highly cited and normal papers, we have collected a broad range of indicators of the research process and outcomes associated with scientific research.

We find that about 80% of publications were by researchers in universities, and another 10% were in government labs. The distribution is fairly similar for H and N projects. The share of government research institutes is substantially higher in material science and environment/ecology/geosciences. Private firms are most prevalent among highly cited engineering papers, representing 12% of the total. About half of respondents were 45 or older when they submitted the focal paper. The average age was lowers in computer science/mathematics. About 30% of researchers had stayed abroad for one year or more for study or research, and about a third had changed jobs in the five years before the project. Mobility is lowest among those in government organizations and hospitals.

Building on Stokes description of the dual goals of advancing knowledge and solving practical problems, we find that scientific research is distributed across the quadrants, and that the distribution varies between H and N projects. The goal of the research fits in Pasteur's quadrant (emphasizing both advancing knowledge and addressing practical problems) in 32% of H and 25% of N projects. Yet, Bohr's quadrant (emphasizing advancing knowledge, without addressing practical problems) is the modal goal, representing 47% of those for H projects and 43% of those for N projects. We also find that serendipitious findings were quite common. Furthermore, the share of serendipitous outcomes was higher for the H projects (48% of projects) than for the N projects (43% of projects). For H projects, computer & mathematics and agriculture science/plant and animal science were fields that were most likely to produce serendipitous output. Future work will explore the characteristics of projects that produce serendipitous findings and the importance of serendipity for generating high impact science. Furthermore, we hope to examine the relations between serendipity and commercialization of science. One conjecture is that the commercial application is a serendipitous outcome of a more science focused projects.

Respondents were aware of scientific competition, with H projects facing greater competition and more concerned with competition. Priority loss is a big concern in fields such as computer science/mathematics, material science and basic life science. Prior work suggests that this competition may have important positive and negative effects on science (Hagstrom 1974; Hong and Walsh 2009; Merton 1957). Future work will explore the effects of competition on scientific productivity, as well as on authorship rules and on training of researchers.

Scientific work builds significantly on prior knowledge. Published literature is the most important knowledge source (about half of respondents rated published literature as 5 out of 5). Other important knowledge sources include: researchers with different research skills, colleagues in their organization and past collaborators. The importance of different knowledge sources is relatively similar between H and N projects.

We find that the most widely used was setting ambitious project goals. This was also the project management characteristic that most distinguished H (80%) from N (66%) projects. The second most used was developing a research community outside the lab. This also had a substantial gap between H (59%) and N (50%) papers. Other commonly used processes included adjusting the project mix depending on whether the early results are promising or not promising, again with the H projects being most likely to do this. We also find that about 40% of projects had developed the project choice and the protocol as group decisions, with little difference between H and N projects. A strict division of labor and a strict hierarchy were somewhat more common in N compared to H projects, while H projects were more likely to hold weekly information meetings of the whole research group. Just under 20% of projects said that the project choice was suggested by a graduate student (and little difference by strata).

We find that competition within the group is quite rare, with only 6-7% of groups reporting this as describing their group. It would be interesting to see the extent to which these lab management strategies can be considered dynamic capabilities that might contribute to consistently high research performance.

About 90% of the respondents reported that they used databases of journals and published papers, consistent with the prior findings that published literature is one of the most important sources of information. The comparison between H papers and N papers shows that authors with H papers were more likely to access to research tool databases (54% vs. 47% for N papers), communicate with remote researchers using Internet (56% vs. 45% for N papers) and access to the latest (unpublished) research information (68% vs. 58% for N paper). H papers (29%) are slightly higher than N papers (26%) in the use of advanced external experimental equipment and research facility. It is likely that there is a relationship between access to advanced equipment and serendipity (Stephan 2010).

In addition to equipment and knowledge, science depends heavily on manpower. Science is increasingly becoming a team activity (Wuchty, Jones, and Uzzi 2007a). The number of authors varies substantially across fields. Clinical medicine/psychiatry/psychology and basic life science have the largest median team size. On the other hand, computer science/mathematics and social science have fewer authors. Also, the number of the authors is greater for H projects than N projects in most fields, consistent with the work of Wuchty and his colleagues, that shows that collaborative projects have higher impact.

Projects also depend heavily on research funding (Stephan 2010). Project funding tends to come from multiple sources. University research is heavily dependent on outside funding, and that is especially true for the H projects. In contrast, government labs and industry researchers are much more likely to use only intramural funds. NIH and NSF are the dominant outside sources for university researchers. However, in some fields, DOD and DOE are major funders. For example, Department of Energy is a major funder for H projects in physics, chemistry, and material science. Department of Defense funds many H projects in material science, engineering and chemistry. About 10% of university researchers get some industry funding.

Two-thirds of H projects including a foreign-born post-doctoral fellow. Projects are also more likely to have foreign-born PhD students than US born. However, US born master's students are more common than are foreign born. Over 40% of projects hired people specifically for the project, which suggests that another important impact of scientific research is job creation. The recent stimulus package funding in the US recognized this link between research funding and jobs and directed significant funding to research institutions. At the same time, the recipients of this funding had to document the job creation associated with these projects.

We also see that scientific publications are also associated with commercial activity. We find that 7 percent of the N projects and 14% of the H projects resulted in at least one patent application. For the highly cited papers, we find high rates of patenting (around 30%) in chemistry, materials science and engineering, consistent with prior work. Life science related fields have moderate rates of patenting, with about 15% of projects resulting in a patent. Physical sciences, mathematics and computing and geosciences all have rates of patenting under 10% of project. We find that 4 percent of the N projects and 8% of the H projects resulted in at least one license. The majority of these licenses include providing know how. Only 1 percent of the N projects and 4% of the H projects resulted in a startup. Future work will explore the drivers of commercialization, including individual, project, field and institutional characteristics. Two interesting questions are: are serendipitous findings especially likely to be commercialized?; and are collaborative projects more or less likely to be commercialized?

Our results suggest that science is a team activity that depends heavily on access to outside knowledge, funding, equipment and skilled personnel. In addition, the is substantial heterogeneity in project goals and in project outcomes. Furthermore, there are significant differences in the goals and organization of projects across fields. Finally, we see a large number of differences between H and N projects, both in terms of their inputs and structures and their various outcomes (in addition to citations). These leads to the conjecture that the drivers of very high impact science and more routine science may be quite different (Kelchtermans and Veugelers 2011). Because of the importance of scientific output for economic development, and the major role of public funding in supporting scientific research, it is imperative that we develop our understanding of the drivers of scientific productivity, and, in particular, what distinguishes high impact science from normal science. This report provides some initial clues. Further work is needed to explore these and similar data more systematically in order to guide policymakers and researchers to help increase the likelihood of producing high impact science.

REFERENCES

Bercovitz, Janet and Maryann Feldman. 2008. "Academic Entrepreneurs: Organizational Change at the Individual Level." *Organization Science* 19: 69-89.

Cohen, Wesley M, Richard R. Nelson, and John P. Walsh. 2002. "Links and Impacts: Survey Results on the Influence of Public Research on Industrial R&D." *Management Science* 48:1-23.

Cohen, Wesley M., Richard R. Nelson, and John P. Walsh. 2000. "Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (Or Not)." in *NBER Working Paper #7552*.

COSEPUP. 2007. "Beyond Bias and Barriers." National Academy of Sciences, Washington, DC.

Crane, Diana. 1972. *Invisible Colleges*. Chicago: University of Chicago Press.

Cummings, J. N. and S. Kiesler. 2005. "Collaborative research across disciplinary and organizational boundaries." *Social Studies of Science* 35:703-722.

Ding, Waverly W., Fiona Murray, and Toby Stuart. 2006. "Gender differences in patenting in the academic life sciences." *Science* 313:665-667.

Etzkowitz, H and L Leydesdorff. 2000. "The dynamics of innovation: from National Systems and ''Mode2'' to a Triple Helix of university–industry–government relations." *Research Policy* 29:109-123

Fox, M. F. and S. Mohapatra. 2007. "Social-Organizational Characteristics of Work and Publication Productivity Among Academic Scientists in Doctoral-Granting Departments." *The Journal of Higher Education* 78:542-571.

Gibbons, M, Camille Limoges, Helga Nowotny, Simon Schwartzman, Peter Scott, and Martin Trow. 1994. *The New Production of Knowledge*. London: Sage.

Hagstrom, W.O. 1974. "Competition in Science." *American Sociological Review* 39:1-18.

Halzen, Francis. 2010. "IceCube Project." in *US-Japan Workshop on Scientific Collaboration and Productivity*. Tokyo, Japan.

Hicks, Diana. 1993. "University-Industry Research Links in Japan." *Policy Sciences* 26:361-396.

Hong, Lu and Scott E. Page. 2004. "Groups of diverse problem solvers can outperform groups of high0aility problem solvers." *Proceedings of the National Academy of Sciences* 101:16385-16389.

Hong, Wei and John P. Walsh. 2009. "For money or glory?: commercialization, competition and secrecy in the entrepreneurial university." *Sociological Quarterly* 50:145-171.

Jones, Ben, Stefan Wuchty, and Brian Uzzi. 2008. "Multi-university Research Teams: Shifting Impact, Geography, and Stratification in Science." *Science* 322:1259-1262.

Kelchtermans, Stijn and Reinhilde Veugelers. 2011. "The great divide in scientific productivity: why the average scientist does not exist." *Industrial and Corporate Change* 20:295-336.

Kneller, R. 1999. "Intellectual Property Rights and University-Industry Technology Transfer in Japan." Pp. 307-347 in *Industrializing Knowledge: University-Industry Linkages in Japan and the United States*, edited by L. Branscomb, F. Kodama, and R. Florida. Cambridge, MA: MIT Press.

Merton, R.K. 1957. "Priorities in Scientific Discovery." *American Sociological Review* 22:635-659.

Merton, Robert K. 1973. *The Sociology of Science*. Chicago: The University of Chicago Press.

MIT. 1999. "A Study on the Status of Women Faculty in Science at MIT."

Nagaoka, Sadao, Masatsura Igam, Manabu Eto, and Tomohiro Ijichi. 2010. "Knowledge Creation Process in Science: Basic findings from a large-scale survey of researchers in Japan." Hitotsubashi University, Tokyo.

National Science Board. 2006. *Science and Engineering Indicators*. Washington, DC: GPO.

—. 2008. *Science and Engineering Indicators*. Washington, DC: GPO.

Normile, Dennis. 2006. "WOMEN IN SCIENCE: Getting Women Scientists Back on the Career Track in Japan." *Science* 311:1235-1236.

OECD. 2005. *Main Science and Technology Indicators*. Paris: OECD.

Pressman, Lori, Richard Burgess, Robert M Cook-Deegan, Stephen J McCormack, Io Nami-Wolk, Melissa Soucy, and LeRoy Walters. 2006. "The licensing of DNA patents by US academic institutions: an empirical survey." *Nature Biotechnology 24, 31 - 39 (2006)* 24:31-39.

Roessner, David, Jennifer Bond, Sumiye Okubo, and Mark Planting. 2009. "The Economic Impact of Licensed Commercialized Inventions Originating in University Research, 1996-2007." Biotechnology Industry Organization, Washington, DC.

Rothaermel, F.T., S.D. Agung, and Lin Jiang. 2007. "University Entrepreneurship: a taxonomy of the literature." *Industrial and Corporate Change* 16:691-791.

Shrum, W., van Chompalov, and Joel Genuth. 2001. "Trust, Conflict, and Performance in Scientific Collaborations." *Social Studies of Science* 5:681-730.

Stephan, P. E. 2010. "The Economics of Science." in *Handbook of the Economics of Innovation*, vol. 1. Amsterdam, Netherlands: North-Holland.

Stokes, Donald E. 1997. *Pasteur's Quadrant*. Washington: Brookings Institution.

Stuart, Toby E. and Waverly W. Ding. 2006. "When Do Scientists Become Entrepreneurs?: The Social Structural Antecedents of Commercial Activity in the Academic Life Sciences." *American Journal of Sociology* 112:97-144.

Teece, David J. 1986. "Profiting from technological innovation: implications for integrations, collaboration, licensing and public policy." *Research Policy* 15:285-305.

—. 2007. "Explicating dynamic capabilities: the nature and microfoundations of (sustainable) enterprise performance." *Strategic Management Journal* 28:1319-1350.

Thursby, J.G and M.C. Thursby. 1999. "Purdue Licensing Survey: A Summary of Results." Unpl. ms. Thesis, Krannert Graduate School of Management, Purdue University.

Valderas, Jose M., R. Alexander Bentley, Ralf Buckley, and K. Brad Wray. 2007. "Why Do Team-Authored Papers Get Cited More?" *Science* 317:1496.

Walsh, J.P. and M. Saegusa. 2003. "Broadening 'Academic' Research: Adapting to University Reforms in Japan." *Journal of Science Policy and Research Management* 18:47-58.

Walsh, J.P. and N.G. Maloney. 2007. "Collaboration Structure, Communication Media and Problems in Scientific Work Teams." *Journal of Computer Mediated Communication* 12:19.

Walsh, J.P. and Sadao Nagaoka. 2009. "Who Invents?: Evidence from the Japan-US Inventor Survey." in *RIETI Discussion Papers*. Tokyo: Research Institute of Economy, Trade and Industry.

Walsh, John P and Todd Bayma. 1996. "Computer Networks and Scientific Work." *Social Studies of Science* 26:661-703.

Walsh, John P. and Nancy G. Maloney. 2002. "Computer Network Use, Collaboration Structures and Productivity." Pp. 433-458 in *Distributed Work*, edited by P. Hinds and S. Kiesler. Cambridge, MA: MIT Press.

Wuchty, Stefan, Ben Jones, and Brian Uzzi. 2007a. "The Increasing Dominance of Teams in the Production of Knowledge." *Science* 316:1036-1039.

Wuchty, Stefan, Benjamin F. Jones, and Brian Uzzi. 2007b. "Why Do Team-Authored Papers Get Cited More?: Response." *Science* 317:1496.

APPENDIX 1 SURVEY METHODOLOGY

1-1 Response rate

There are multiple ways to calculate response rates. The simplest is the number of surveys completed divided by the number mailed out (by post or email). As seen in Appendix Table 1, this is 2329 completes divided by 8923 mailed out, for a raw response rate of 26.1%. However, many of the 8923 were not actually in the target population, either because they were duplicates of existing cases, they were deceased, they were not in the US, or it was not a research paper. If we exclude these cases (277+280+22+67=646) we are left with 8277 cases in the denominator. Thus, the adjusted response rate is 28.1%. If we limit our response rate calculation to those 7693 cases where a potential respondent was actually contacted (i.e., exclude those with no confirmed address), we get an effective response rate of 30.3%. We also had 875 partial responses, although these are not included in this report.

Appendix Table 1. A summary of the survey responses

		Count	**Total %**
Types of responses	Completed on the Web	2329	26.1%
	Partials (>= 25 pages)	67	0.8%
	Partials (3 – 24 pages)	808	9.1%
	Immediate quit (<= 2 pages)	371	4.2%
Non-responses	Non-response (silent)	3858	43.2%
	Declined	260	2.9%
	Undelivered	338	3.8%
	Wrong author	86	1.0%
	Unable to find or confirm address	160	1.8%
Ineligible cases	Review paper	277	3.1%
	Only international authors	280	3.1%
	Deceased	22	0.2%
	Additional Duplications	67	0.8%
	Total mail-outs	**8923**	

1-2 Targeted Number of Focal papers by Science Field of Journals

Exhibits 53 to 55 show the targeted number of focal paper candidates by journal fields. We have 9428 papers, including 3142 H papers and 6011 N papers. The sampling was designed to produce papers from 2000 to 2006 publication year, although a few papers are outside this range.

Exhibit 53. Targeted number of focal paper candidates by journal fields

Row Labels	1999	2000	2001	2002	2003	2004	2005	2006	2007	Grand Total
Agricultural Sciences		2	35	48	22	35	27	43		212
Biology & Biochemistry		5	93	102	97	107	116	86		606
Chemistry		9	118	132	131	146	148	101		785
Clinical Medicine		21	338	315	339	343	349	331	1	2037
Computer Science		2	39	36	60	65	63	41	1	307
Economics & Business		11	38	37	33	37	34	21		211
Engineering		19	98	102	102	104	92	66	2	585
Environment/Ecology		2	35	35	34	38	50	47		241
Geosciences		4	48	47	58	52	48	34		291
Immunology		2	46	43	36	37	40	46		250
Materials Science		6	40	51	48	46	54	43		288
Mathematics		4	41	37	45	35	37	15		214
Microbiology		3	32	44	51	56	39	47	1	273
Molecular Biology & Genetics		5	73	86	82	84	77	53		460
Multidisciplinary		1	18	23	16	12	6	10		86
Neuroscience & Behavior		5	45	49	54	59	60	55		327
Pharmacology & Toxicology		2	36	36	42	25	35	37		213
Physics		10	137	131	136	153	148	108		823
Plant & Animal Science		6	51	51	49	54	65	44		320
Psychiatry/Psychology		2	39	36	38	38	39	30		222
Social Sciences, general	2	14	62	76	58	78	72	85		447
Space Science		5	40	37	42	34	39	33		230
Grand Total	2	140	1502	1554	1573	1638	1638	1376	5	9428

Exhibit 54. Targeted number of focal paper candidates by journal fields (H papers)

	2000	2001	2002	2003	2004	2005	2006	Grand Total
Agricultural Sciences		12	15	9	12	9	13	70
Biology & Biochemistry	1	29	30	28	32	45	20	185
Chemistry	3	43	47	46	54	47	33	273
Clinical Medicine	5	114	105	119	123	123	94	683
Computer Science	1	13	11	20	24	22	11	102
Economics & Business	4	13	13	9	12	13	6	70
Engineering	4	35	34	32	39	31	19	194
Environment/Ecology	1	12	13	10	16	17	15	84
Geosciences	1	16	16	21	21	18	10	103
Immunology	1	12	14	13	14	13	15	82
Materials Science	1	13	19	15	16	21	11	96
Mathematics	1	14	14	12	14	13	3	71
Microbiology		11	14	16	17	19	12	89
Molecular Biology & Genetics	1	25	28	23	28	19	18	142
Multidisciplinary		1	1	1				3
Neuroscience & Behavior	2	12	15	16	18	19	19	101
Pharmacology & Toxicology		14	10	14	9	16	9	72

Physics	5	50	44	58	59	50	32	298
Plant & Animal Science	2	24	18	21	20	23	15	123
Psychiatry/Psychology		14	12	12	13	14	8	73
Social Sciences, general	5	20	24	19	32	28	20	148
Space Science		14	11	15	12	17	11	80
Grand Total	38	511	508	529	585	577	394	3142

Exhibit 54. Targeted number of focal paper candidates by journal fields (N papers)

	1999	2000	2001	2002	2003	2004	2005	2006	2007	Grand Total
Agricultural Sciences		2	23	33	13	23	18	30		142
Biology & Biochemistry		4	64	72	69	75	71	66		421
Chemistry		6	75	85	85	92	101	68		512
Clinical Medicine		16	224	210	220	220	226	237	1	1354
Computer Science		1	26	25	40	41	41	30		205
Economics & Business		7	25	24	24	25	21	15		141
Engineering		15	63	68	70	65	61	47	2	391
Environment/Ecology		1	23	22	24	22	33	32		157
Geosciences		3	32	31	37	31	30	24		188
Immunology		1	34	29	23	23	27	31		168
Materials Science		5	27	32	33	30	33	32		192
Mathematics		3	27	23	33	21	24	12		143
Microbiology		3	21	30	35	39	20	35	1	184
Molecular Biology & Genetics		4	48	58	59	56	58	35		318
Multidisciplinary		1	17	22	15	12	6	10		83
Neuroscience & Behavior		3	33	34	38	41	41	36		226
Pharmacology & Toxicology		2	22	26	28	16	19	28		141
Physics		5	87	87	78	94	98	76		525
Plant & Animal Science		4	27	33	28	34	42	29		197
Psychiatry/Psychology		2	25	24	26	25	25	22		149
Social Sciences, general	2	9	42	52	39	46	44	65		299
Space Science		5	26	26	27	22	22	22		150
Grand Total	2	102	991	1046	1044	1053	1061	982	4	6286

1-3 Affiliation of survey targets

Exhibit 59 shows number of survey targets by sector on the basis of information gathered from Web of Science and additional archival sources. Approximately three-quarters of the targets are located in universities, although government labs and firms also have significant numbers of target respondents. Exhibit 60 shows the top 30 institutions in terms of number of survey targets. Harvard, University of California, MIT and Stanford top the list.

Exhibit 59. Number of survey targets by sector on the basis of the searched information

Sector	Number of survey targets	Share
University	6840	76.7%
Government Lab	757	8.5%
Private Firms	850	9.5%
Non Profit	222	2.5%
Others	61	0.7%
Unknown/ Overseas	193	2.2%
Total	8923	100.0%

Exhibit 60. Top 30 institutions in terms of the number of survey targets

Name of institution	Number of survey targets
Harvard University	151
University of California	132
Massachusetts Institute of Technology	96
Stanford University	89
University of Michigan	77
University of Washington	75
Harvard Medical School	64
Columbia University	63
University of Minnesota	62
University of Pennsylvania	59
University of Texas	58
Cornell University	53
University of Florida	52
University of Maryland	52
University of Illinois	50
Yale University	50
Duke University	49
Johns Hopkins University	47
Northwestern University	47
California Institute of Technology	44
Pennsylvania State University	42
University of Chicago	42
University of Wisconsin	41
National Institutes of Health	40
Massachusetts General Hospital	39
Ohio State University	39
Princeton University	39
University of Pittsburgh	39
Purdue University	37
University of Southern California	35

Questionnaire: We developed a questionnaire that covers the details of the research project that produced the paper and the outcomes of that project (see below). The questionnaire was designed to be administered via the web (we have experience with web-based surveys). The final survey instrument is included as supplemental material (Appendix 2).

Survey topics: The survey covers the following topics: Motivation and the other basic characteristics of the research project that yielded the paper, such as measures of Stokes quadrants, serendipity, and scientific competition; the knowledge production process, such as the respondent's roles in the project and uses of external knowledge and geographic location of those knowledge sources, and organization of research project; research inputs, including project duration, funding and sources of funds; composition of the research team, by rank, organization type, field, country of origin, research skill/specialty gender, and number of students trained and personnel hired by project as well as authorship rules; the outputs of the project, such as number of other papers, patents, licenses, startup firms; and scientists' demographics, family status (marital status, children), education and training, mobility, awards, and publication counts.

3-1 Method of implementing the survey

There are two phases of the survey, one is the contacting phase, and the other is the responding phase. For the contacting phase, for those survey targets are able to find the email addresses, we sent them an email invitation asking his/her participation to the survey and along with user token and the URL to access to the survey site. For those survey targets (N = 741) with only post mail address, we sent them the invitation letter by post mail instead.

The Web survey was designed to secure personal information of respondents. We provided unique tokens for each survey target. After a respondent logs in, the corresponding information of the publication will show on the second page of the website. Once the survey target clicked that he/she is the right authors and the focal paper is an outcome of a research project, then the page of the consent form is displayed, as well as the instruction for the survey. If survey targets checked "I agree to participate", then they are able to proceed to the questionnaire and answer the survey.

We put up one group of questions on each screen with Next and Back buttons provided for browsing. For some branching questions, we provided automatic branching to the linked question. We did not enforce answers nor implement input validation. The survey was hosted on a secure Georgia Tech server.

3-2 Basic time-line of the survey

The time line of the survey is listed below.

- Survey launch: September 7, 2010
- Two reminders: (November 5, 2010; Jan 10, 2011)
- Final due date: Jan 28, 2011

We made the following data cleaning steps to correct the inaccurate responses or data inconsistencies.

1) Examining consistency across questions on the time line of the research project, and recoding those years with strange symbols or missing digits as missing.
2) Cleaning most of the open text questions, such as numbers of man-month, amount of research funds over $100,000,000, and the percentage of the total research budget accounted for PI or Co-PI's salaries. Replacing negative numbers to missing value. Some values for budgets were out of range (less then $100 million on the question of budgets over $100 million). These were set to $125,000,000.
3) Checking the consistency in the question of the composition of the authors, to make sure the respondent checked the right box indicating, "If you are the author, check Yes", correctly.

5-1-1. Classification of scientific field

Exhibit 63 Relation between the ESI journal fields, the 10 fields and the large fields

22 ESI Journal fields	10 fields	Large fields
Chemistry	Chemistry	Physical Science
Materials Science	Material Science	
Physics	Physics & Space Science	
Space Science		
Computer Science	Computer & Mathematics	
Mathematics		
Engineering	Engineering	
Environment/Ecology	Environmental/Ecology/Geoscience	
Geosciences		
Clinical Medicine	Clinical Medicine/Psychology	Medicine
Psychiatry/Psychology		
Agricultural Sciences	Agricultural Science/Plant Animal Science	Life Science
Plant & Animal Science		
Biology & Biochemistry	Basic Life Science	
Microbiology		
Molecular Biology & Genetics		
Neuroscience & Behavior		
Pharmacology & Toxicology		
Immunology		
Multidisciplinary	Either 22 ESI journal fields were assigned based on the analysis of backward citation	Either 22 ESI journal fields were assigned based on the analysis of backward citation
Economics & Business	Social Science	Social Science
Social Sciences, general		

5-2. Field classification of the multidisciplinary papers

Exhibit 64 Journals in the multidisciplinary field and the focal papers

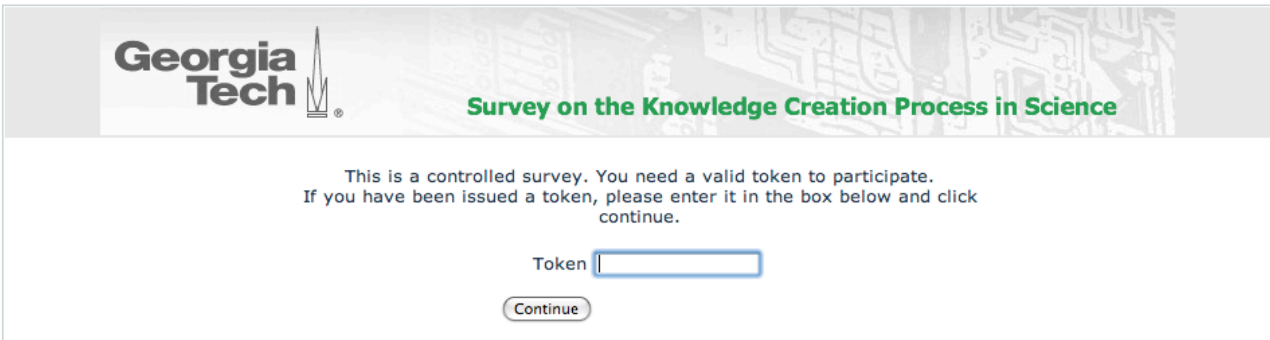
Journal Name	Focal papers
PNAS	614
SCIENCE	270
NATURE	243
Physical Review Letters	79
Scientific American	16
American Scientist	11
Chinese Science Bulletin	9
New Scientist	8
Synthese	8
Naturwissenschaften	7
Current Science	6
Total	1271

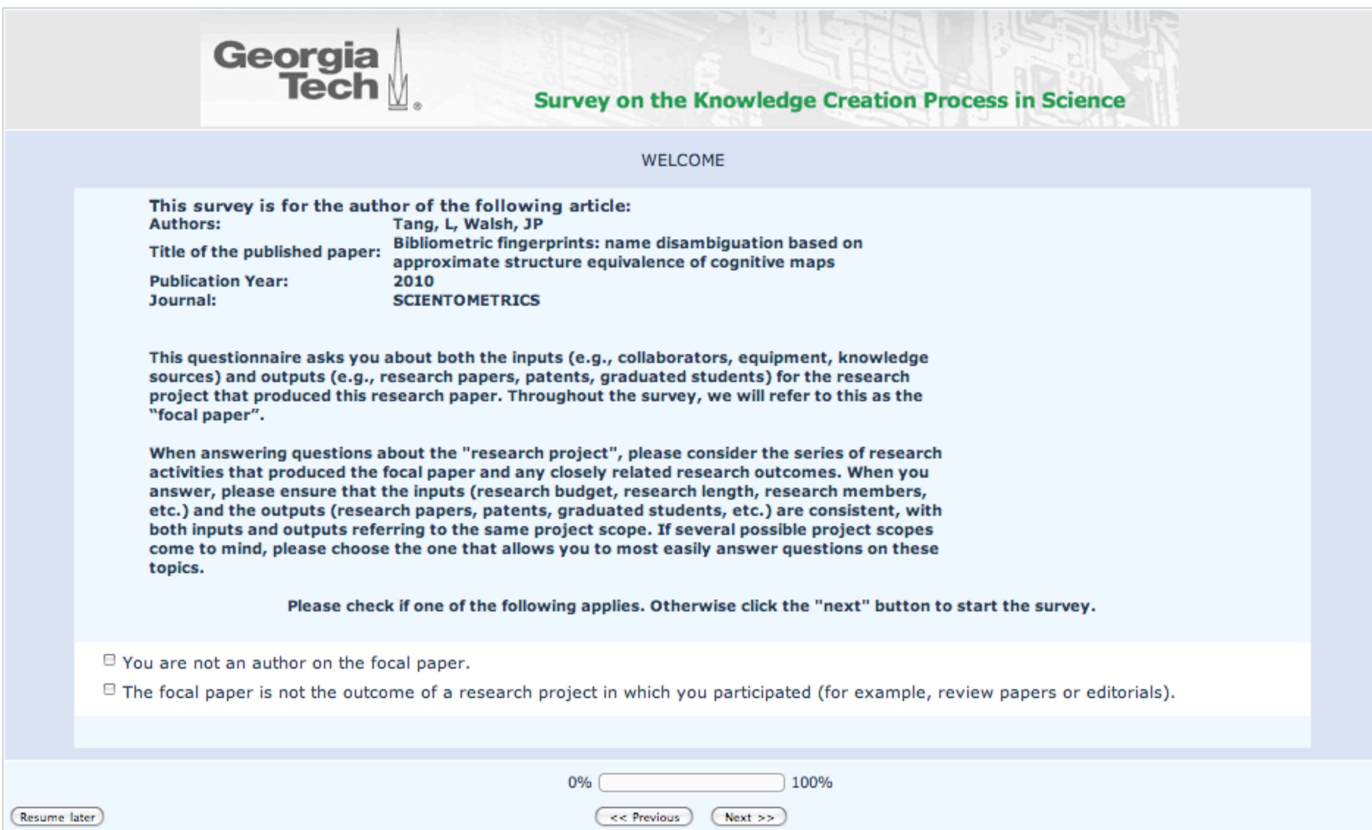
The papers from these “multi-disciplinary” journals were reclassified according to the journal classification of the references in the papers. We thank Masatsura Igami for providing these reclassification data. Using this method left 11 cases as multidisciplinary. Future work will hand classify these cases.

APPENDIX 2 FINAL QUESTIONAIRE

Academic Scientist Survey – Questionnaire

September 2010
Georgia Institute of Technology
School of Public Policy





*

Informed Consent

(1) Your responses to this questionnaire shall be kept under the strict control of the principal investigators. Individual responses will remain confidential and will not be disclosed to the public.

(2) It takes approximately 20 minutes to complete the questionnaire.

(3) The summary results of the survey will be made public by December, 2010, on the project website (see URL below). The results can be emailed to you if you would like. Follow the instructions at the end of the survey to request a copy of the report.

(4) Although your response is very important to us, your participation is voluntary. You will not benefit or be compensated for joining this study. The risks involved are no greater than those involved in daily activities. To make sure that this research is being carried out in the proper way, the Georgia Institute of Technology IRB may review study records. The Office of Human Research Protections may also look at study records. If you have any questions about your rights as a research subject, you may contact Ms. Melanie Clark, Georgia Institute of Technology at (404) 894-6942.

(5) Please do not hesitate to ask us if you have any questions by sending email to Professor John Walsh Georgia Institute of Technology, School of Public Policy (john.walsh@pubpolicy.gatech.edu). We have also posted more information on the project, including a Frequently Asked Questions sheet, at (http://www.prism.gatech.edu/~jwalsh6/SciSurvey.html).
We will also post summary results from the survey when they are available.

I have read the information contained above and would like to be a volunteer in this research study.

☐ I agree to participate

0% 100%

Resume later

<< Previous Next >>

1-1 Motivation for the research project

How important were the following motivations for initiating the research project that yielded the focal paper?

Please rate on a scale of 1 to 5 (1: Not important; 5: Very important).

	Not important [1]	[2]	[3]	[4]	Very important [5]
(1) Pursuit of fundamental principles/understandings					
(2) Solving specific issues in real life					

1-2 Research process for the focal paper

	Largely the same as originally planned [1]	[2]	[3]	[4]	Quite different than originally planned [5]
(1) Did the research project that yielded the focal paper proceed as initially planned?					

1-3 Research process for the focal paper

	Substantially LESS significant than expected [1]	[2]	[3]	[4]	Substantially MORE significant than expected [5]
2) Was the main result of the focal paper more or less significant than your initial expectations?					

1-4 Research process for the focal paper

	Yes	No
3) Has the research output found the answers to questions not originally posed (in other words, was the research output serendipitous)?		

1-5 Research competition

Approximately how many major research teams did you recognize as your potential competitors when you began the research project? Indicate the number of potential competitors in the US (i.e., a competing team with its leader located in the US) and outside of the US.

	None	1	2-5	5-10	More than 10	Unknown
Number of potential competitors **in the US**	○	○	○	○	○	○
Number of potential competitors **outside of the US**	○	○	○	○	○	○

1-6 Threat from competition

How concerned were you about the possibility that your competitors would have priority over your research results (in other words, concern about being "scooped")?

Rate on a scale of 1 to 5 (1: Not at all concerned; 5: Very concerned).

	Not at all concerned [1]	[2]	[3]	[4]	Very concerned [5]
Level of concern	○	○	○	○	○

1-7 Importance of the focal paper in the field

From your perspective, how important is the focal paper compared to the global research findings in the same field during the same period (published within a year before or after the focal paper was published). Please select the answer below that best describes your evaluation.

- ○ (a) It is one of the most important papers, ranking in the top 1%
- ○ (b) It is a very important paper, ranking in the top 10%
- ○ (c) It is a relatively important paper, ranking in the top 25%
- ○ (d) It ranks in the top 50%
- ○ (e) It ranks in the bottom half among papers published around that time

2-1 Your role in the research project

1) Please indicate which of the following best describes your role in the management of the research project.

- (a) A leading role in the research management, designing the research project, organizing the research team, and/or acquiring research funds (Principal Investigator or Co-PI)
- (b) A member of the research management but less than that of the leader
- (c) No managerial role
- (d) Management was not necessary
- (e) Other (please briefly describe your specific role in managing the project)________________

2-1 Your role in the research project

2) Please indicate which of the following best describes your role in the research implementation

- (a) I executed the central part of the research and contributed the most to the research output
- (b) I took part in the central part of the research but my contribution was not as substantial as the central researcher
- (c) I implemented the research under the guidance of the above members
- (d) I contributed to the research through the provision of materials, data, equipment, or facilities
- (e) Other (please briefly describe your specific role in implementing the project)

2-2 External knowledge sources that inspired the research project

1) How important were each of the following external knowledge sources (excluding members of the research team) for conceiving the research project?

Rate on a scale of 1 to 5 (1: Not important; 5: Very important), or 0: Did not use.

	Did not use [0]	Not important [1]	[2]	[3]	[4]	Very important [5]
(a) Published literature (articles in journals, etc.)	○	○	○	○	○	○
(b) Unpublished literature (preprints, information on websites, etc.)	○	○	○	○	○	○
(c) Patent literature	○	○	○	○	○	○
(d) Conferences, workshops or academic meetings	○	○	○	○	○	○
(e) Colleagues in your organization (university, government lab, firm, etc.)	○	○	○	○	○	○
(f) Visiting researchers or post-doctoral researchers in your organization	○	○	○	○	○	○
(g) Past research collaborators	○	○	○	○	○	○
(h) Competitors	○	○	○	○	○	○
(i) Partners in an industrial-academic-government alliance	○	○	○	○	○	○
(j) Researchers in different academic fields	○	○	○	○	○	○
(k) Researchers with different research skills (for example, experimental researchers for theorists)	○	○	○	○	○	○

(2) For the following highly rated sources (which you rated "4" or "5"), please specify the country where the key knowledge source was located (for example, location of the key researcher for a published paper, country of origin for visiting researchers, conference venue for conferences, etc.).

(Please choose the one most important location for each information source.)

	USA	EU[1]	Japan	China	Other country
(a) Published literature (articles in journals, etc.)	○	○	○	○	○

	USA	EU[1]	Japan	China	Other country
(b) Unpublished literature (preprints, information on websites, etc.)	○	○	○	○	○

	USA	EU[1]	Japan	China	Other country
(c) Patent literature	○	○	○	○	○

	USA	EU[1]	Japan	China	Other country
(d) Conferences, workshops, or academic meetings	○	○	○	○	○

	USA	EU[1]	Japan	China	Other country
(e) Colleagues in your organization (university, government lab, firm, etc.)	○	○	○	○	○

	USA	EU[1]	Japan	China	Other country
(f) Visiting researchers or post-doctoral researchers in your organization	○	○	○	○	○

	USA	EU[1]	Japan	China	Other country
(g) Past research collaborators	○	○	○	○	○

	USA	EU[1]	Japan	China	Other country
(h) Competitors	○	○	○	○	○

	USA	EU[1]	Japan	China	Other country
(i) Partners in an industrial-academic-government alliance	○	○	○	○	○

	USA	EU[1]	Japan	China	Other country
(j) Researchers in different academic fields	○	○	○	○	○

	USA	EU[1]	Japan	China	Other country
(k) Researchers with different research skills (e.g., experimental researchers for theorists)	○	○	○	○	○

[1] EU countries: Austria, Belgium, Bulgaria, Cypress, Czech Republic, Denmark, Estonia, France, Finland, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, United Kingdom

2-3 Research management

To what extent did your research group do each of the following?
Rate on a scale of 1 to 5 (1: Not at all; 5: Very much so).

[If any of the following items do not apply for your publication, please skip that item.]

	Not al all [1]	[2]	[3]	[4]	Very much so [5]
(a) The research group set ambitious project goals	○	○	○	○	○
(b) The project choice was the result of a group decision among all the members of the research group	○	○	○	○	○
(c) Decisions about the initial research protocol were made collectively by the whole research group	○	○	○	○	○
(d) This project was suggested by a graduate student or post-doctoral researcher	○	○	○	○	○
(e) The existing equipment in our lab was an important factor in choosing this project	○	○	○	○	○
(f) All decisions to modify an existing research protocol required discussion and approval by the whole research team	○	○	○	○	○
(g) During the course of the project, graduate students developed on their own changes in the research protocol	○	○	○	○	○
(h) The project involved a strict division of labor with each person responsible for a specific part of the research	○	○	○	○	○
(i) The whole research group met every week to share information on project progress	○	○	○	○	○
(j) The project leaders checked the graduate students' lab notebooks at least once per week	○	○	○	○	○

	Not al all [1]	[2]	[3]	[4]	Very much so [5]
(k) There was a clear hierarchy in the research group, such that students reported to team leaders and team leaders reported to lab heads	○	○	○	○	○
(l) The project involved outsourcing parts of the work to other research groups	○	○	○	○	○
(m) The research project generated improved computing or simulation capabilities	○	○	○	○	○
(n) The research project generated improved equipment or experimental facilities	○	○	○	○	○
(o) The research protocol was modified or additional experiments or analyses conducted based on feedback from conference presentations	○	○	○	○	○

(p) The research group encouraged competition among team members	○	○	○	○	○
(q) In our group, when initial results of a project looked promising, we significantly increased the resources and time devoted to that project	○	○	○	○	○
(r) In our group, we regularly reviewed on-going projects and stopped working on those that were not producing promising results	○	○	○	○	○
(s) We worked to develop a research community in our field beyond our lab	○	○	○	○	○

2-4 Access to advanced research facilities, databases and remote researchers

To what extent did you have difficulty accessing potentially important research facilities, databases and remote researchers for your project?

Please rate from 0 to 5
(0: Did not need; 1: No difficulty getting access; 5: Great difficulty getting access).

	Did not need [0]	No difficulty [1]	[2]	[3]	[4]	Great difficulty [5]
(a) Advanced external experimental equipment or facilities (such as accelerator, super computer, observatory, etc.)	○	○	○	○	○	○
(b) Databases of journals/published papers	○	○	○	○	○	○
(c) Research tool databases (genomes, materials, etc.)	○	○	○	○	○	○
(d) Regular participation of remote researchers (such as through the internet, teleconferencing, etc.)	○	○	○	○	○	○
(e) Access to the latest foreign and domestic research information (information available before it is published in journals)	○	○	○	○	○	○

3-1 History of the research project

Please indicate the time lines of the research project, as follows

	1) Year when the project was conceived	2) Year when the project was initiated	3) Year when the focal paper was first submitted for publication	4) Year when the most recent research paper was submitted (if not focal paper)
Year				

5) Has the project been completed?

	If Yes, when?	If no (on-going project), expected year of completion
Year		

3-2 Total research man-months expended on the research project

Please indicated the approximate man-months that the entire research team spent, from the point when the research project began to the point when the most recent research findings were submitted for publication.

(#Example: If three team members worked for 24 months, 18 months, and 6 months respectively in a two-year project, then the total is 48 man-months. Please round your answer to an integer.)

[ ] man-months

3-3 Research funds

Please tell us the approximate amount of funds directly used for the research project
The research funds should include personnel costs for researchers and research support personnel when they are supported by this pro
the PI or other faculty, if part of the grant budget).

Regarding the costs for large equipment, if the equipment was purchased solely for the project, include it as part of the research costs; c

- (a) Less than $10,000
- (b) More than $10,000 but not more than $100,000
- (c) More than $100,000 but not more than $250,000
- (d) More than $250,000 but not more than $500,000
- (e) More than $500,000 but not more than $750,000
- (f) More than $750,000 but not more than $1,000,000
- (g) More than $1,000,000 but not more than $2,500,000
- (h) More than $2,500,000 but not more than $5,000,000
- (i) More than $5,000,000 but not more than $7,500,000
- (j) More than $7,500,000 but not more than $10,000,000
- (k) More than $10,000,000 but not more than $25,000,000
- (l) More than $25,000,000 but not more than $50,000,000
- (m) More than $50,000,000 but not more than $75,000,000
- (n) More than $75,000,000 but not more than $100,000,000
- (o) More than $100,000,000

3-3 If the project budget was over $100,000,000, please tell us the approximate total budget for the project (in millions of dollars):

$ ____ ,000,000

Investigator salaries

	%
Of this total research budget, approximately what percent of the total budget was for PI or co-PI salaries (summer salary, release time, etc.)?	____ __________

3-4 Sources of research funds

Please indicate the approximate percent of total project funding that came from each of the following sources. If you have questions about how to allocate particular funding sources, please see the notes at the bottom of the table.

Only numbers may be entered in these fields

Total of all entries must not exceed 100

- (a) Internal funds - Funds of the institution that the research team members belong to (domestic or foreign) ____ %
- (b) Center grants (such as ERCs) from the US government --------------------------- ____ %
- (c) National Institutes of Health competitive research grants -------------------------- ____ %
- (d) National Science Foundation competitive research grants -------------------------- ____ %
- (e) Department of Energy competitive research grants --------------------------------- ____ %
- (f) Department of Defense competitive research grants -------------------------------- ____ %
- (g) Other competitive project grants from the Federal government -------------------- ____ %
- (h) Non-competitive project grants (such as a national project led by the government) ____ %

- (i) External funds from domestic state and local governments ------------------------ %
- (j) External funds from foreign governments --- %
- (k) Commissioned research from US firms -- %
- (l) Collaborative research with US firms -- %
- (m) Donations from US firms -- %
- (n) Other funding from US firms -- %
- (o) External funds from foreign firms -- %
- (p) Other (such as Foundations) -- %
- Total: %

Regarding the research money that state universities and public research institutions received from the government (excluding competitive research funds), if you cannot determine whether the funds are internal to your institution or are other external funds (government), choose non-competitive research grant if the funds are tied to specific research subjects, and choose internal funds of your institution if they do not target a specific subject.
When national research funds were allocated via a foundation, choose (competitive or non-competitive) external funds from US Federal Government. # If the headquarters of the firm is located in the US (abroad), please identify the fund as "External funds from US (foreign) firm."

4-1 Composition of the authors

Please identify the job position and type of organization (at the time when the focal paper was submitted for publication); field of expertise; skill/specialty; country of birth; and gender of each author.
When there are more than six authors, the ones in the list have been randomly selected (although the first and last authors are always included).

Co-author List	Job position	Type of organization	Field of expertise	Skill	Country of birth	Gender	If the author is you, check Yes
Tang, L	Please choose...	Please choose...	Please choose...	Please choose...	Please choose...	Please cho	☐ Yes
Walsh, JP	Please choose...	Please choose...	Please choose...	Please choose...	Please choose...	Please cho	☐ Yes
No, Y	Please choose...	Please choose...	Please choose...	Please choose...	Please choose...	Please cho	☐ Yes
Huang, H	Please choose...	Please choose...	Please choose...	Please choose...	Please choose...	Please cho	☐ Yes
Shibayama, S	Please choose...	Please choose...	Please choose...	Please choose...	Please choose...	Please cho	☐ Yes
Nagaoka, S	Please choose...	Please choose...	Please choose...	Please choose...	Please choose...	Please cho	☐ Yes
	Please choose...	Please choose...	Please choose...	Please choose...	Please choose...	Please cho	☐ Yes

4-2 Scope of authors

Please indicate whether any of the following types of researchers are included among the authors.

- ☐ (a) Any researcher who only supplied research materials analyzed in the research
- ☐ (b) Any researcher who only supplied data analyzed in the research
- ☐ (c) Any researcher who only supplied or developed the research facilities or equipment used in the research
- ☐ (d) Any researcher who only supplied or developed computer programs or databases used in the research
- ☐ (e) Any researcher who only supplied funds used in the research
- ☐ (f) Don't know
- ☐ (g) Other ☐

4-3 Order of authors

Which of the following best describes the name order of the authors on the focal paper?

- (a) Ordered by degree of the contribution of authors
- (b) Alphabetical order
- (c) Seniority (Senior author first)
- (d) Seniority (Senior author last)
- Other (Please provide specific type of the contribution)

4-4 Other members of the research team who are NOT coauthors on the paper

Please indicate the numbers of PhD-level researchers, students and technicians who played a significant role in the implementation of the project but are not co-authors of the focal paper.

Technicians provide technical service for the research under the guidance and direction of researchers. This does not include those who were involved in general affairs, accounting and miscellaneous duties in research-supporting work.

	Number	Don't know
(a) Collaborating researchers		
(b) Graduate students		
(c) Undergraduate students		
(d) Technicians		

4-5 The number of R&D personnel specifically hired for this project

Please identify the number of R&D personnel (authors of the paper as well as cooperating researchers, students and technicians) specifically hired for this project, i.e., whose personnel costs were covered in Q3-3 above. Approximate numbers will do if the exact count is difficult.

Only numbers may be entered in this field

5-1 Number of papers produced by the research project

1) Approximately how many refereed papers (including refereed conference proceedings) did the research project lead to, including the focal paper itself?

	English	Other languages
Number of published papers (refereed)		

5-2 Training of researchers

Please tell us how many people received a master's or a PhD degree or received post-doctoral training through the research project. Approximately numbers are sufficient.

	Born in the US	Born outside the US
(a) Received a master's degree		
(b) Received a PhD degree		
(c) Post-doctoral fellows		

5-3 Application for patents

1) Did the findings from the research project lead to a patent application?

- ○ Yes
- ○ No [skip to 5-4]
- ○ Do not know

5-3 Application for patents

2) How many patent applications were filed based on this project, in the US and internationally? For PCT international applications or those not to the US Patent Office (USPTO), all applications from the same invention should be counted as one.

	USPTO applications	**Non-USPTO applications (including PCT international applications)**
Patent applications		

3) Please inform us of the most important patent from the project by indicating its application (or publication or grant) number below. If it is not a patent from the US Patent Office, please indicate the name of the patent office.

	(Application, publication or patent number)	**Name of the patent office (e.g. USPTO, EPO, JPO, WO etc.)**
The most important patent from the project		

Examples:
Publication number: "EP2345678(A1)", "WO2010/0123456", "US2010/0123456", or just a number
Grant number: "EP2345678(B1)", "US7345678", or just a number
quot;Application number: "EP20101234567", "PCT/JP/2010/123456", "US11/123456", or just a number

4) Was any research team member or the organization with which he was affiliated the assignee (or co-assignee) of the above patent?

- Yes
- No

5-4 Technology transfer

1) Were any research results from the research project licensed, sold or assigned to an outside firm?

- (a) Yes
- (b) No, neither licensed, sold nor assigned to an outside firm [skip to 5-5]
- (c) Don't know

5-4 Technology transfer

2) If some research results were licensed, sold or assigned to an outside firm, how large were the recipients of the technology (numbers of employees)? When there was technology transfer to multiple firms, indicate all that apply. Also inform us whether they include a start-up firm (a firm five years old or younger).

Check all that apply.

	Size		Start-up (five years old or younger)	
	Yes	No	Yes	No
(1) 250 employees or more				
(2) Less than 250 employees but more than 50 employees				
(3) Less than 50 employees but more than 10 employees				
(4) less than 10 employees				

3) Did the technology transfer agreement include the research team providing know-how to the receiving firm?

- (a) Yes, it included providing know-how
- (b) No, it did not include providing know-how
- (c) Don't know

5-5 Start-up companies

	Yes	No
1) Did the findings from the research project lead to a start-up company? (A start-up company here means a new company established based on the findings of the research project, and does not include an existing company that is granted a license).		

5-5 Start-up companies

	Yes	No
2) Did you seriously consider the possibility of forming a startup company?		

5-5 Start-up companies

3) Please tell us the following about the start-up.

- Name of company
- Year established
- City
- State

4) How were the members of the research team involved in the start-up company? Check all that apply.

One or more research team members...:

- ☐ (a) founded the company
- ☐ (b) assumed executive positions
- ☐ (c) were involved as a member of the scientific advisory board
- ☐ (d) consulted for the start-up (technical guidance, etc.)
- ☐ (e) worked as employees on a part-time basis
- ☐ (f) worked as employees on a full-time basis
- ☐ (g) Other (Please describe their involvement) ☐

5) Why was the start-up company formed as a channel for commercialization? (Please choose the most important reason.)

- (f) Other reason

6-1 Your background

1) Please provide the following information about yourself.

	Year
1) Year of birth	

2) Please indicate your organizational affiliation during the project period.

If you moved during the course of the project, list the organization you were affiliated with when you began working on the project.

If the name of the organization changed, please indicate its current name.

- Name of university, company, research institute, etc.
- Name of school, department, group, division, etc.

2 Family situation

	Yes	No
(a) Were you married at the time when the research projects started?		
(b) Did you have children (including adult children) at the time when the research project started?		

6-2 Family situation

(c) How old were your children at the time that the research project started?

	Fill in the number of children in the following age brackets
(a) up to age 5	☐
(b) age 6-18	☐
(c) age 19 or above	☐

6-3 Educational background

1) What was your highest degree at the time you initiated the research project?

- ○ (a) Ph.D. or M.D.
- ○ (b) Master's degree (including partial completion of Ph.D)
- ○ (c) J.D.
- ○ (d) Bachelor's degree
- ○ (e) Technical college or junior college
- ○ (f) Other ☐

6-3 Educational background

2)
a) With respect to your highest degree (master's or Ph.D.), please answer the following questions.

(a) Year you received the degree	Year	☐

(b) Name of the university where you received your degree and your major

University	
Major	

	Year
3) In what year did you first submit a paper to a refereed journal? Write the year of the submission, regardless of whether it was accepted or not.	

6-4 Research career

Please inform us about your research experience at the time when you started the research project.
Had you....?

	Yes	No
(a) Stayed abroad for one year or more for graduate study or research (#This includes international students studying in the US)	○	○
(b) Changed academic or research positions across organizations in the preceding five years (#This excludes taking a job after your graduation)	○	○
(c) Had a visiting position or secondment to another institution in the preceding five years	○	○
(d) Served on the editorial board of an international journal	○	○
(e) Won a distinguished paper award or a conference award from an academic society	○	○

6-5 Publications

Please tell us how many refereed papers (including refereed conference proceedings) you published during the period 2007-2009. Please include co-authored papers.

	English	Other languages
Number of published papers (refereed)		

Effects on industry and on society, or other comments

If you have any comments on the impact of your research, or any other issues related to your research project, please use the space below.

Would you like to have the summary results from this survey sent to you when they are available?

- Yes
- No

If you would like to have a copy of the research summary, please include your email address below.

Email address:

THANKS FOR YOUR COOPERATION