International Collaboration in Scientific Research

I. Abstract and Methodology

An increasing proportion of the teams that dominate scientific research (Wuchty, Jones and Uzzi, 2007) include researchers from more than one country. The share of articles with a US-based author that have at least one researcher outside the country increased from 23% in 2000 to 32% in 2010 (NSF, S&E Indicators 2012, table 5-19). What factors have led to increased globalization in the production of scientific knowledge? How do collaborators connect with one another? What is the division of work among collaborators? Do papers based on international collaborations have greater or less impact than other papers in their field?

In this paper we analyze international collaboration of research in three "hot" scientific fields: nanotechnology, biotechnology, and particle physics. Our analysis uses data from publications in these fields, a survey of the researchers on such teams, and interviews with collaborating researchers. We focus on collaborations in which at least one researcher is US-based.

Our publication data comes from the Thomson Reuters Web of Science database, where we identify articles in the fields using information from the subject categories assigned to each journal. We identify teams by the names of co-authors and international teams by the location of the authors. Our survey data come from our own on-line survey of the authors of a sample of papers published in the 2000s. We have conducted a pilot survey of 100 corresponding authors in nanotechnology and will launch the survey for the entire sample in May.

The shift towards greater US collaboration with scientists overseas presumably reflects the increased global dispersion of the scientific workforce; the large number of international students and post-docs who have returned home from the US but continued working with US-based researchers; improvements in communication; reduced costs of travel; and the geographic distribution of scientific equipment and materials, particularly large machinery. We have selected our fields to cover a range of rapidly growing scientific areas with different knowledge production functions, and changes in labor supplies and capital equipment that we anticipate would affect collaborations between US-based researchers and overseas researchers.

Nanotechnology is a general interdisciplinary applied technology, where engineers often collaborate with material scientists and where the electron microscope is a pivotal research tool. The US and other countries have begun sizable investments at different times. Biotechnology is lab-based, where life scientists largely funded by NIH dominate research and where researchers often have close links to big pharma firms. The most important change in research technology has been the US-sponsored Human Genome Project, the result of international efforts, which now provides data freely on the web. Particle physics has two parts – theoretical and experimental, with much leading edge research requiring massive investments in accelerators and colliders. The recent development of the Large Hadron Collider (LHC) has changed the geographic locus of empirical research while the US decision to forgo the proposed large collider in Texas arguably spurred the greater growth of string theory in the US than in Europe.

We begin our analysis with a brief review of the trends in international collaborations in total, and then focus on US international collaborations in the three fields of concern. We estimate the share of multi-authored US papers with international co-authors, and the international proportion and national composition of those authors for papers with differing numbers of authors. For a sub sample we examine the ethnic link between the US-based author(s) and the non-US based author(s). We create a cross-country and field database over time that allows us to link the proportion of foreign-based co-authors by country to each country's proportion of authors in the

field. This allows us to estimate the effect of the geographic distribution of researchers on US collaborations. To determine the importance of US-specific doctoral training, we calculate each country's share of US PhDs granted in the fields, where the biggest change is the increased proportion of US PhDs granted to Chinese and others from developing countries. Unfortunately NSF data on post-docs by field and year does not provide information on the country of non-US post-docs, so we are limited to examining international post-docs in the aggregate.

Using the Web of Science data, we will contrast the impact factors of the journals in which papers appeared and the number of forward citations to papers with/without international collaborators. For instance, we will compare a three-authored paper with only US researchers to a paper with one or two overseas researchers. We will differentiate among collaborating countries.

Our survey allows us to go below publication data and aggregate measures of research activity and labor supplies to investigate the micro-relations involved in collaborations in each field. We ask about the rank of team members and whether each coauthor had access to data, equipment, knowledge/ skills or other inputs that the other authors did not have; about how the collaboration formed (*How did you first come in contact with each of your coauthors?*); and about the corresponding author's assessment of the benefits of the collaboration (*Which of the following would you say were the 3 most important benefits of doing this research collaboratively?* Possible answers include: *access to data, access to equipment, access to specific skills or knowledge, the research experience was more pleasant, etc.).* We also collect data on funding that supported the research, including funding for cross-country collaborations and on the frequency of communication among coauthors and the communication technology used. Ideally, the survey data will allow us to assess the motives and rewards to researchers from the international collaborations. We will interview a random sample of respondents to get a deeper understanding of the benefits and costs of the collaborations.

Finally, to examine how research collaborations change as the locus of equipment and machinery changes, we will examine the impact on international collaborations by CERN's Large Hadron Collider. Our hypothesis is that the US-based proportion of researchers on empirical particle physics papers will drop while US international collaborations with Europeans will rise as American scientists move their work to CERN. But we suspect that the opposite may have occurred in theoretical particle physics, with the US concentration on string theory reducing international collaborations. Since the US, EU, Japanese, Chinese, and Russian nanotechnology funding initiatives target work in the country or area covered, but encourage jointly funded work, we are uncertain whether the net effect is to reduce or increase international collaborations. We will calculate the number of international collaborations in nanotechnology that acknowledge funding support from more than one country. We are also uncertain how the availability of data from the Human Genome Project and digital data in biosciences (or elsewhere) that allow persons who did not collect data access to analyze it has or will affect collaborations in the future. We will discuss these topics with researchers in the fields before designing and undertaking empirical analysis.

II. Current Status of Work

We are in the midst of creating the database of publications across fields and linking the country-year-level data to aggregate data on the distribution of researchers by country across fields and the distribution of US PhDs granted by country in the fields. We have completed our pilot survey and will be sending the survey to the full sample of authors in May. We are also creating a timeline of shocks that might affect international collaborations in each field: the decision to build CERN's LHC, the timing, size, and terms of nanotech funding programs, and availability of digitalized biological data over the Internet.