

**Location of birth unknown: What we can learn by studying CVs**

Monica Gaughan  
University of Georgia

Presented November 7, 2007 at:

**CAREER PATTERNS OF FOREIGN-BORN SCIENTISTS AND ENGINEERS,  
TRAINED AND / OR WORKING IN THE U.S.**

**The Science and Engineering Workforce Project**  
of the National Bureau of Economic Research

**Sponsored by the Alfred P. Sloan Foundation**

November 7, 2007

Richard Freeman, Paula Stephan, and John Trumbour: Organizers

**The National Bureau of Economic Research**  
1050 Massachusetts Avenue, 3<sup>rd</sup> floor, Cambridge, MA

## **Theoretical Underpinnings**

The use of the curriculum vita to study scientific careers was developed in the late 1990s by the Research Value Mapping Program at Georgia Institute of Technology. It was originally developed to incorporate longitudinal data into survey and interview-based research on scientific careers and the organization of science. Together, these complex efforts are designed to operationalize and test scientific and technical human capital models (Bozeman et al. 2001). Briefly, scientific and technical human capital is comprised of the individual and social factors that enable scientists to create their work. Human capital theory is well elaborated with respect to scientific career study, but incorporation of social effects is still relatively recent. In our conceptualization, the social world is captured in the institution through which and in which scientists work as their careers develop over time (Figure 1). These experiences become direct and indirect inputs into the development of scientific capacity.

When RVM began work on CV analysis in the late 1990's, the RVM team was able to locate only two peer reviewed articles referring to the use of CVs for the analysis of scientific careers. There was, therefore, very little guidance available to assist us in the development of coding, data management, and analytic strategies. The data reported here come from the third major effort in CV collection, coding and analysis. I begin with a brief history of research using the curricula vita of academic scientists and engineers before detailing the approach used in this analysis.

## **Earlier collections of curricula vita**

The first effort to evaluate the use of CVs occurred in a Department of Energy/National Science Foundation funded study of scientists and engineers affiliated with Engineering Research and Science and Technology Centers. The basic design concept was to use the multidisciplinary science center as the focal cluster for collecting the CVs of scientists and engineers affiliated with them (Dietz et al. 2001). The data collection included structured questionnaires, curricula vita, and in some cases semi-structured interviews and site visits. A great deal of work has been published from that effort,<sup>1</sup> but only two of the publications address the issue of foreign-born scientists and engineers (Lee 2004; Lee and Bozeman 2006), neither of which relied on the CV itself for the identification of whether or not the person was foreign-born. The data were further limited in the selected nature of the study population: scientists and engineers already affiliated with multidisciplinary science and engineering centers.

The second collection of curricula vita was funded by the National Institute of Child Health and Human Development to evaluate the impacts of a Congressionally-earmarked program to support translational research and training in human infertility. This research was specifically designed to compare center affiliates of the earmarked program with other scientists working on human infertility, but who were not affiliated with the earmarked program. To date, this CV collection has not been used to study foreign-born scientists (Gaughan 2007; Gaughan and Ponomariov 2007).

---

<sup>1</sup> Corley et al. (2003), Dietz (2004), Dietz and Bozeman (2005), Gaughan and Bozeman (2002), Gaughan and Robin (2004), Lin and Bozeman (2006).

### **Sampling and data collection of results reported here**

The most recent CV collection was designed to overcome some of the study limitations of the first two efforts by explicitly designing the sample to be broadly representative of academic scientists and engineers in NSF-defined disciplines. As with the other efforts, the design is not without its limitations. First, we limited the development of our sample frame to scientists and engineers working in Research Extensive (Carnegie 2000) universities. Our rationale for targeting this population of institutions is that the majority of academic research in science and engineering takes place in these 150 universities. Furthermore, since training of graduate students and postdoctoral fellows was an important concern of the study, we sampled only tenured and tenure-track scientists and engineers working in academic departments actively training Ph.D. students (NSF). We further excluded economists and health scientists from the scientist sample, and focused our engineering recruitment on faculty members in 5 disciplines in engineering: electrical, chemical, mechanical, civil and materials. The result was a sample frame of 36,874. From this frame, we sampled 4516 scientists and engineers in the disciplines studied in this paper. We explicitly over-sampled women, who comprise half of the sampled and studied population.

The sampled scientists and engineers were asked to complete a formal structured questionnaire and to provide their CV in one of three ways: a hard copy included with the returned questionnaire, by email, or by providing an internet link to an online version. The response rates have been low, but characteristic for this busy population. The overall questionnaire response number is 1647, or 36%. To increase the number of usable CVs, trained undergraduate students searched the internet to find the CVs of respondents who did not provide them with the questionnaire.

The length, complexity, and lack of standardization of academic CVs are astonishing. Even the CV of the same person can include formatting inconsistencies that make coding extremely difficult. In the first CV project, we used a coding and data management scheme that resulted in an unwieldy proliferation of variables. In effect, trained graduate and undergraduate students coded every possible piece of information into a person-level flat file with potentially thousands of variables. For example, one respondent had three variables to code each of his 600 publications. As already noted, a large body of work has been published from these data. I think each researcher who has worked with the data would agree: we collected too much!

In the NICHD coding, we used the lessons learned from the ERC/STC study to focus our CV coding on constructs of theoretical interest (rather than the sheer force of empirical will exerted in the first effort), and to change our data management approach from an exclusively person-centered approach to a relational approach supported by ACCESS. In this way, we were able to link individuals with their educational and job histories and scientific productivity.

This most recent coding effort builds on the NICHD project by further limiting the amount of information collected about publications because of redundancy with other databases such as ISI Web of Science. We abandoned the collection of grant information because the diversity of reporting is so great that acceptable levels of reliability cannot be obtained. The most important coding change for the results reported here, we incorporated two new coding sub-protocols to capture information about whether or not the person is foreign born. Appendix A shows the first interface for entering data.

Coders were asked to check the box “Click here if there is any evidence the individual is a non-US born citizen.” The coder was then to type in all of the evidence that supported that conclusion—excluding evidence that comes from higher education degrees. Appendix B shows the interface used to code data from educational histories. The box “Check if this is an international institution” allows us to know if each type of degree an individual has was conferred by a foreign institution.

### **Design of this analysis**

This study is organized to answer the question, what can we know about foreign born scientists using academic CVs? Prior work provides a good idea of the diversity of substantive areas that can be explored using the CVs, whether they are from foreign born scientists or not. The much more specific question I seek to answer here is if the only data we have is the academic CV, could such data be used to understand if the person is foreign-born? Recall that CVs were collected as part of a formal survey, and constitute a subset of the final data. The survey questionnaire included two questions addressing nativity, shown in Appendix C. For this analysis, I use the 500 respondents who indicated they were not US citizens. This figure represents 30% of the total survey respondents (n=1647), and is consistent with national estimates of foreign-born scientists and engineers in academic science.

We take as our starting point 500 scientists we know are foreign-born by their self- report. Of these, approximately 350 CVs associated with these respondents have been located, an overall yield rate of about 70%.<sup>2</sup> To date, 211 CVs have been coded from the following disciplines: Biology, Earth and Atmospheric Science, Agriculture, Mathematics, Chemistry and Computer Science. Some engineering CVs are included in this analysis, but the bulk of those remain to be coded.

### **Findings**

Data from the “Click here if there is any evidence the individual is a non-US born citizen” sub-protocol (Appendix A) are sufficiently uninteresting that a table is unwarranted. Of 211 coded CVs, we were able to infer the person was foreign-born on the basis of non-educational information only 11 times, or five percent of known cases.

More promising are results when educational histories are considered. Figure 2 shows the distribution of degree types by whether or not the degree was conferred by a foreign institution of higher education. Overall, 222 (46%) of degrees were conferred by foreign institutions. However, the distribution is skewed toward bachelor’s degrees, with 77% of the bachelor’s degrees earned at foreign institutions. In contrast only 23% of doctorates are conferred by foreign institutions.

Using degrees as the unit of analysis can be misleading, however, as many individuals have multiple degrees of the same type. Categorizing degrees with the individual scientist in mind, a different picture emerges. Figure 3 shows individuals categorized into six possible types: not enough evidence, all degrees are foreign, BS and

---

<sup>2</sup> Since we have not completed coding all of the available CVs, the final figure and yield rate is likely to change somewhat.

MS are foreign, Bachelor's is foreign, no higher education outside the US, and no education on CV.

To determine that a person had no foreign institution as a degree, then all degrees had to be from a US institution. What we find is that 33 individuals (16%) had no higher education outside the US. Not enough evidence is present in 25 (12%) cases. For example, if the person only listed a US PhD, with no other degrees listed, he or she is categorized as having insufficient information to evaluate the institutions conferring degrees. A final methodological concern is that 18 (9%) of the CVs had no educational information at all; these cases are the result of getting CVs unobtrusively via the internet rather than directly from the scientist.

Overall, we are able to identify known foreign-born scientists by using educational histories only 64% of the time. If we exclude the CVs collected unobtrusively, that figure goes up to 70% of the time. By either indicator, this is a very low rate of positive identification of foreign-born scholars using CVs. The prospects of using CVs to identify foreign-born scholars are further complicated by other limitations of the methodology, to which I now turn.

### **Sources of Bias**

There are a number of bases of selection that need to be studied further. First, the response rate for the questionnaire survey was a low 38%. Furthermore, only about 70% of respondents provided (or we were able to find unobtrusively) their CVs. Finally—and most important for this work—CVs accurately identified 64% of people who are known to be foreign-born by their own report.

The issue of selection bias is one that affects the ability of researchers to generalize. Ongoing work is attempting to specify whether there is bias, its sources, and potential remedies. For example, a comparison of respondents and non-respondents to the survey revealed that there was no field or rank differences, but that women were slightly more likely to respond to the survey. We have not yet compared those who provided CVs to those who did not, but such an analysis is possible.

### **Conclusions**

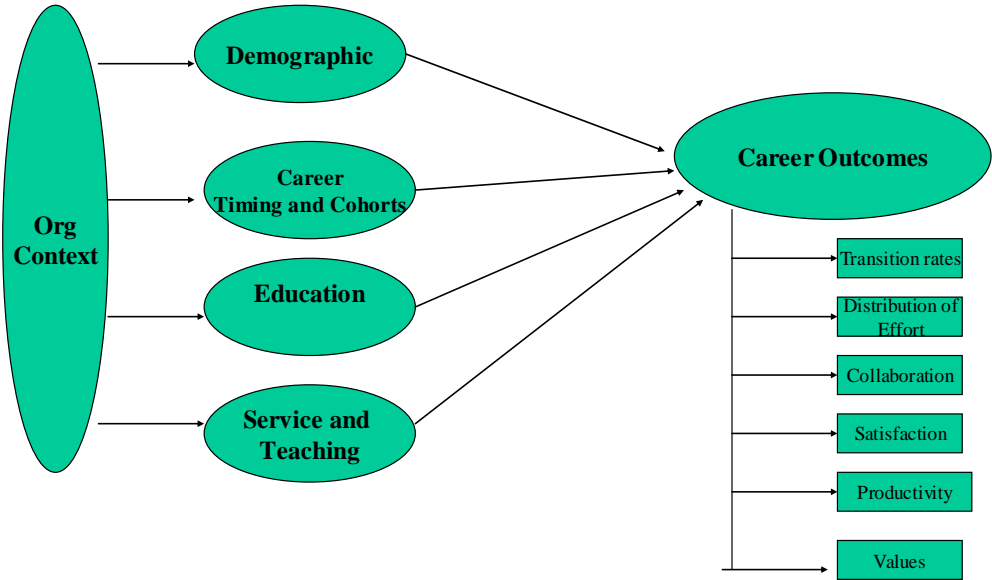
In about two-thirds of cases, the full academic CV can be used to make inferences about whether or not a person is foreign-born. With such a low rate of positive identification, CVs should not be used exclusively to identify foreign-born scholars. However, in combination with other types of data, CVs can be used to develop detailed educational and employment trajectories which may be used to understand how foreign-born scientists enter the US and develop as students and scholars.

## References

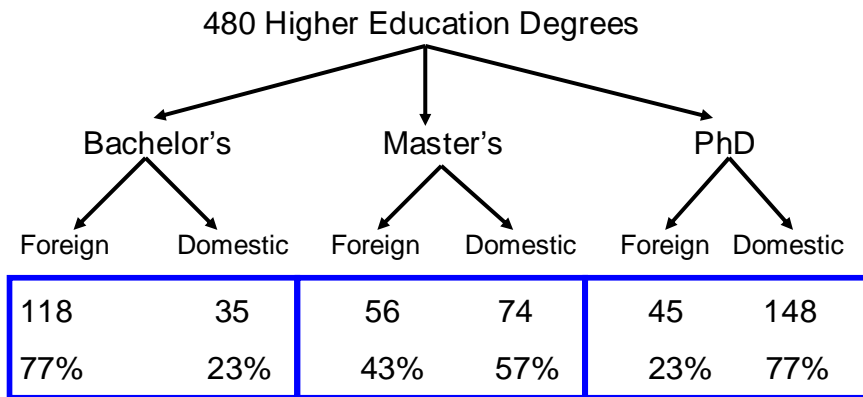
- Bozeman, B., Dietz, J. & Gaughan, M. (2001). Scientific and Technical Human Capital: An Alternative Model for Research Evaluation. *International Journal of Technology Management*, 22, 716-740.
- Carnegie Foundation. (2000). Classification of higher education. Retrieved on March 22, 2005, from <http://www.carnegiefoundation.org/Classification/index.htm>
- Corley, E., Bozeman, B. & Gaughan, M. (2003). Evaluating the impacts of grants on women scientists' careers: The curriculum vita as a tool for research assessment. In P. Shapira and S. Kuhlmann (Eds.), *Learning from Science and Technology Policy Evaluation: Experiences from the U.S. and Europe*. Cheltenham, UK: Edward Elgar Publishing.
- Dietz, J. & Bozeman, B. (2005). Academic careers, patents, and productivity: Industry experience as scientific and technical human capital. *Research Policy*, 34, 349-367.
- Dietz, J. (2004). Scientists and Engineers in Academic Research Centers: An Examination of Career Patterns and Productivity. Dissertation at Georgia Institute of Technology.
- Dietz, J., Chompolov, I., Bozeman, B., Lane, E. & Park, J. (2001). Using the curriculum vita to study the career paths of scientists and engineers: An exploratory assessment. *Scientometrics*, 49, 419-442.
- Gaughan, M. (2007). Using the curriculum vitae and survival analysis to assess impacts of multidisciplinary center funding on career trajectories of new researchers. International Conference of the International Society for Scientometrics and Informetrics. Madrid, Spain. June 25-28, 2007.
- Gaughan, M. & Bozeman, B. (2002). Using curricula vitae to compare some impacts of NSF research center grants with research center funding. *Research Evaluation*, 11, 17-26.
- Gaughan, M. & Ponomariov, B. (Provisional acceptance). Center impacts on faculty publication productivity, collaboration, and grants velocity: Using curricula vita in a matched control design. *Research Evaluation*.
- Gaughan, M. & Robin, S. (2004). National science training policy and early scientific careers in France and the United States. *Research Policy*, 33, 569-581.
- Lee, S. & Bozeman, B. (2006). The impact of research collaboration on scientific productivity. *Social Studies of Science*, 35, 673-702.
- Lee, S. (2004). Foreign-born scientists in the United States: Do they perform differently than native-born scientists? Dissertation at Georgia Institute of Technology.
- Lin, M. & Bozeman, B. (2006). Researchers' industry experience and productivity in university-industry research centers: A scientific and technical human capital explanation. *Journal of Technology Transfer*, 31, 269-290.

Figure 1

### Scientific and Technical Human Capital



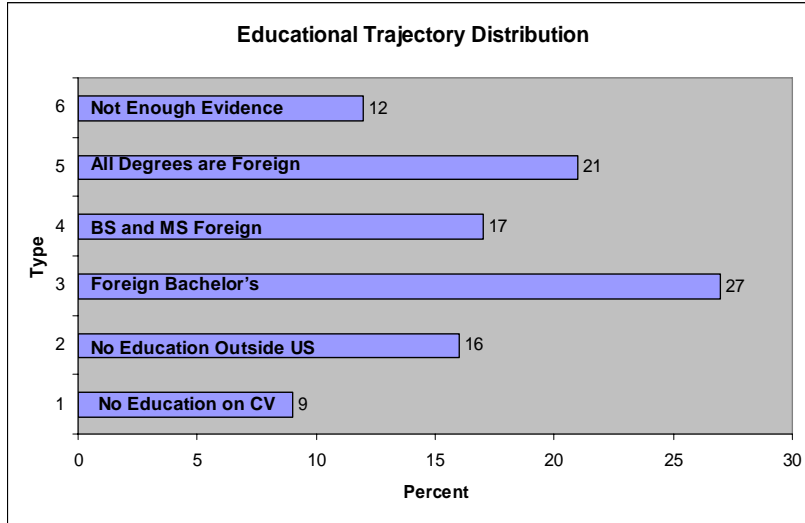
**Figure 2**  
**Distribution of degrees among 211 foreign-born scientists.**





**Figure 3**

**Educational Trajectories by Individual  
N=211**



## APPENDIX A

Access GUI interface for coding information about national origin

frmIndividuals

RVMID	<input type="text" value="999999"/>
Coder:	<input type="text" value="9"/> Castleman
Coding Date:	<input type="text" value="11/3/2007"/> MM/DD/YYYY
Time coding started:	<input type="text" value=""/> : <input type="text" value="0"/> any format
<b>CV Full or Partial?</b>	
<input checked="" type="radio"/> Full	
<input type="radio"/> Partial	
<hr/>	
First Name:	<input type="text" value="Monica"/>
Middle Name:	<input type="text"/>
Last Name:	<input type="text" value="Gaudhan"/>
Click here if there is any evidence the individual is a non-US born citizen: <input checked="" type="checkbox"/>	
If you checked the above box cite your evidence here <input type="text" value="indicates citizenship"/>	
1 if Marital Status	<input type="text" value="1"/>
1 if Children	<input type="text" value="0"/>

## APPENDIX B

Access GUI interface for coding information about location of higher education

frmDegrees

ID

**Degree Type**

- BA/BS
- MA/MS
- PhD
- MD
- JD
- Other

**Degree Year**

Check if this is an international institution

*CODER: Select University from the Drop-Down list or type in the name of the institution if it is not on the list. All institutions on the list are US based*

**Degree Institution:**

**Degree Field:**

APPENDIX C

2004 Survey of Academic Researchers  
Questionnaire Excerpt to Identify CVs of Foreign-born  
Research Value Mapping Program  
Barry Bozeman, PI

21. What is your current citizenship status?

- Native born U.S. citizen **USCITZ**
- Naturalized U.S. citizen **NATUSCIT**
- Non U.S. citizen with a permanent U.S. resident visa **PERMVISA**
- Non U.S. citizen with a temporary U.S. resident visa **TEMPVISA**

22. *[IF U.S. NATURALIZED CITIZEN OR NON U.S. CITIZEN]*, of which country are (were) you a citizen?

\_\_\_\_\_ **COUNTRY**