Using the Curriculum Vita to Study the Career Paths of Scientists and Engineers: An Exploratory Assessment

James S. Dietz

Ivan Chompalov

Barry Bozeman

Eliesh O'Neil Lane

and Jongwon Park

Research Value Mapping Program
School of Public Policy
Georgia Institute of Technology
GCATT Building, Room 120
250 14th Street, NW
Atlanta, Georgia 30318-0490
http://rvm.pp.gatech.edu
gt7631a@prism.gatech.edu

June 26, 2000

A version of this paper was presented at the American Evaluation Association meeting in Orlando, FL, November 4, 1999.

Abstract

In this paper we assess the utility of the curriculum vita (CV) as a data source for examining the career paths of scientists and engineers. CVs were obtained in response to an email message sent to researchers working in the areas of biotechnology and microelectronics. In addition, a number of CVs were obtained "passively" from a search of the Internet. We discuss the methodological issues and problems of this data collection strategy and the results from an exploratory analysis using OLS regression and event history analysis. In sum, despite difficulties with coding and variation in CV formats, this collection strategy seems to us to hold much promise.

Using the Curriculum Vita to Study the Career Paths of Scientists and Engineers: An Exploratory Assessment

Introduction

Scientists' and engineers' career trajectories have much in common with other professional paths. Motivational factors are not so very different, including income, need for achievement and recognition, and desire for "interesting" work. Scientists and engineers face many of the same constraints as others, choosing jobs because of a spouse's opportunities, the quality of schools available to children, distance from family, and so forth. Thus, the standard models available to labor economists can tell us much about scientists and engineers.

But there are some respects in which scientists and engineers differ dramatically from dentists or attorneys or airline pilots. Some are obvious like the peculiar formal assets, such as patents and publications, that scientists and engineers bring with them. Other assets are more subtle, less formal, but perhaps even more important. Each scientist and engineer can be thought of as a unique embodiment of "knowledge value"—a walking set of knowledge, skills, technical know-how and, just as important, a set of sustained network communications, often dense in pattern and international in scope. In previous work,[1,2] we outlined a knowledge value (which we then referred to as "scientific and technical human capital") model as an alternative model for research evaluation, originating in response to the limitations of traditional economic and state-of-the-art models. The knowledge value model puts more weight on the sustained ability of scientists and engineers to enhance their own capabilities and those with whom they work than do traditional models.[3] Knowledge value includes not only the researcher's human capital but the social capital he or she draws upon in creating knowledge and interacting in various social and professional contexts. It includes not just the educational credentials normally recognized in traditional human capital models[4-6] but the researchers' tacit knowledge,[7,8] craft knowledge, and know-how. And, essential to the effective exploitation of all of these human capital endowments is the social capital[9-12] that scientists continually exercise in engaging their interests.

These endowments not only make the study of scientists' and engineers' career trajectories more difficult (e.g., less amenable to standard labor models) but more nuanced and more challenging. When a dentist changes jobs it is of interest chiefly to old and new clients. When a scientist or engineer changes jobs the implications are often profound: the movement of knowledge value is, arguably, a vital element of scientific discovery, technological innovation, and even economic development. For knowledge value transcends the intellect of any one individual. Individual migration patterns of scientists and engineers can be likened more appropriately to the movement of the "web" of knowledge value that they possess—a web that continually takes on new shapes and patterns.

If the career trajectories of scientists and engineers are often a bit more complicated and less predictable than dentists (or airline pilots or attorneys), they also leave more marks along the trail. One of the great, albeit largely unexploited, advantages of studying the careers of scientists and engineers is the near universal reliance on the curriculum vita (CV). The utility of CV data for study of knowledge value is striking, at least at first blush, because of its tremendous richness—richness that must be tapped if one is to move away from traditional labor and productivity models in favor of knowledge value models.

The CV provides not only a clear-cut indicator of movement from one work setting to the next but is, in a sense, a representation of certain aspects of knowledge value. The CV, unlike other data sources, often recounts the entire career of the scholar in some detail. Thus, it is not simply a list of credentials, but a historical document that evolves over time capturing changes in interests, jobs, and collaborations. Whether viewed as a historical record, a marketing tool, or a scientific resource, it is a potentially valuable datum for persons interested in career trajectories, research evaluation, or, more generally, science and technology studies. Not only is the CV nearly universal, it is in some respects standard, and it is relatively easily obtained (sometimes even from the public domain). Most important, the CV contains useful, concrete information on the timing, sequence, and duration of jobs, work products (e.g., articles, patents, papers), collaborative patterns, and scholarly lineage. The CV is, indeed, a rich source of longitudinal data, which lends itself especially well to the study of phenomena associated with careers and labor flows—precisely the target of knowledge value.

In addition to its value as a stand-alone source of data, a great advantage of the CV is that it can be used in conjunction with other sources of data. The availability of a wide array of citation data through the Science Citation Index is extremely valuable. These same databases also include information on the "power index" (i.e., the likelihood of citation) of journals. Similarly, the aggregate data provided by US National Science Foundation (NSF) databases, such as SESTAT, also serves as a potentially fruitful linkage. The problem, of course, is that the decision to use such benchmarks and cognate data requires making significant "up front" decisions on data collection strategies.

On the other hand, this approach is not without its limitations or problems. In fact, several of the advantages of using the CV as a data source can also be viewed as disadvantages. First, because the information is self-reported, it is subject to being portrayed in a favorable light or even completely fabricated. But so too is any self-reported questionnaire or interview data. Second, the semi-structured format falls short of a purely standardized template, thus risking the elimination of valuable information or the inclusion of extraneous non-relevant data. Perhaps most significant, however, is the enormous work involved in coding the CV for subsequent data analysis. Not only is the coding time-consuming, but it is tedious and runs the risk of introducing error due to coder fatigue. In some cases it is possible to have as many as 2,000 variables for one CV.

Despite its limitations, the potential of the CV as a research tool is enormous. Yet it has been used only sparingly—and sometimes incidentally—as a research device. We seek to address this neglect, to explain it, and to assess the promise and obstacles to a research agenda employing the CV as primary data. The development of such a methodology undoubtedly would provide a unique and potentially useful alternative for evaluating scientists' and engineers' careers.

Organization of the paper

Our interest in knowledge value, and the potential of CVs as a research tool, stems from a general interest in assessing the impacts of government-financed research projects. The *Research Value Mapping (RVM) Program* within the School of Public Policy at the Georgia Institute of Technology began in 1996, using 30 intensive case studies of research projects as sources of both qualitative and quantitative information about the nature and intensity of the projects' scientific and socioeconomic impacts.[13]

The Phase I work, sponsored by the US Department of Energy's (DOE) Office of Science, focused entirely on DOE-sponsored projects in government and university labs. We are now in Phase II based on continued funding from DOE and with new funding from NSF. Whereas Phase I focused on information developed in the case studies, Phase II is focusing on knowledge value impacts, using the CV as one research tool to examine labor flows and career trajectories. The core hypothesis of Phase II is that many of the impacts of projects are not easily confined within normal project boundaries but occur over considerable time as knowledge value diffuses into other settings.

We will ultimately test several hypotheses about the connection between the characteristics of team-oriented R&D projects and the diffusion of knowledge value via the projects' affiliates. A preliminary study (begun for this paper) of scientists and engineers provides the opportunity to explore the use of the curriculum vita as a methodological tool for garnering such information.

A major objective of this paper, then, is simply to determine the extent to which it is possible to obtain useful CV data and to assess the utility of various approaches to collecting CVs. In next [third] section, we present a brief overview of the literature on scientific careers. In the fourth section, methodological issues are presented and discussed with specific attention to several coding and data-related issues. In the fifth section, we address issues of validity and reliability, data consistency and quality, and CV accessibility. After examining the preliminary findings (sixth section), we reflect more broadly (final section) on our assessment of the utility of a CV-based methodology, including possible strategies for improving the quality and consistency of data.

Scientific productivity and knowledge value

Studies of innovations have established the importance of close coupling for knowledge transfer and the diffusion of innovations in the economy.[14] The flow of people from one organization, firm, or group to another is key in the process of knowledge exchange. But, despite some good attempts,[15, 16] the extant literature has not managed to fully capture the dynamic nature of these flows over time and across research contexts. Careers are inherently dynamic—evolving and intersecting in planned and unplanned ways, but traditional research evaluation models view them as static or, at best, additive and cumulative over time.

Life cycle models view the careers of scientists as a longitudinal function of the individual's skill levels and his or her incentives to act productively.[17, 18] The concept originated in human capital theory from an economics tradition[19] that sought to relate investments in human beings (education, training, job and life experiences, and personal health) to an individual's earnings trajectory. At earlier stages of career building, productivity incentives are strong while skills are growing. At the early to middle stages, both incentives and skills are strong as productivity peaks. And at middle to later stages, both wane, as does productivity.[20, 21] Although there is plenty of empirical evidence to support this notion of diminishing marginal rates of productivity, such models have failed to explain much variation in productivity.[22] Moreover, as Stephan and Levin[23] have pointed out, many of these life-cycle models lack sufficient attention to the research process and the institutional setting of the process—something, incidentally, more akin to our concept of knowledge value.

Researchers have also called attention to the role of early career collaboration and mentoring as spurs to longer-term scientific productivity. Long and McGinnis[24] found that predoctoral collaboration with mentors had significant and lasting effects on the careers of biochemists. The productivity of the mentor was positively and strongly related to the biochemists' own publication productivity six years later. For students who had not collaborated with their mentor, there was no relationship. Similarly, Reskin,[25] studying chemists who obtained their Ph.D. in the late 1950s, found graduates from higher "caliber" departments were more likely to have collaborated with their doctoral mentor and showed higher productivity after their first postdoctoral decade than graduates from lesser-prestige departments. Zuckerman[26] revealed that Nobel prize winners viewed their doctoral apprenticeship as crucial to their later success. Specifically, they pointed to its role in building broad skills such as knowledge of proper standards of achievement, tastes in choice of research problems, and confidence in their work and abilities.

Life *course* models can be thought of as an enhancement or conceptual expansion of life *cycle* models. The most important contribution of life course theories to the understanding of the scientific careers and knowledge value, is the notion that human lives are linked, or interdependent with each other, and—not just statically—but dynamically over time. Merton[27] recognized this in titling his book, *On The Shoulders of*

Giants, in which he illustrates how Newton made his intellectual advances using the contributions of his scientific peers and forefathers. Elaborated by Elder,[28] the life course paradigm views individual lives as affected by the historical period in which events occur, the developmental timing and sequence of events, and the involvement of the individual in relevant social relationships. Elder refers to the concept of human agency—which, as applied to science, can be thought of as the unique set of abilities that each scientist uses to translate his or her training and skills into scientific outputs. In a sense, human agency is a recognition that individuals vary in the predispositions (both strengths and weaknesses) they bring to the construction of a life course.

The life course model, with all of its subtleties about the sequencing and timing of career events, seems to be amenable to examination through the coding and analysis of CV data. However, very few studies have employed CVs as data sources about trends in job mobility in science. Typically, when CVs are used in career studies, they are used as a supplemental source of information that serves to fill in gaps left from other research modalities.[29,30] Even when CVs[31] are utilized as the primary or only data source,[32] their advantages or disadvantages are rarely discussed. The notion of using CVs as a research tool is hardly a novel idea. But the actual utility of CVs lies in answers to some quite practical questions and the resolution of some fundamental methodological issues.

Research issues and concerns

A wealth of information is provided in most CVs but the coding of the information and its entry into a database is not at all straightforward. When one considers that some CVs include hundreds of publications and conference papers, many with multiple authors, the costs of labor become apparent. The options are few. To capture almost all the information in a CV into standard databases would require a small army of labor, well trained and perhaps not all "low end" inexperienced data entry personnel. The other option—limiting data capture—must, almost inescapably, be pursued. Absent prodigious data entry resources, the only option is to forgo much data or to categorize data at a relatively high level of abstraction (e.g., simplify authorship order). Unfortunately, however, too high a level of data aggregation is incongruent with the theoretical complexities of the knowledge value model. The trick of the trade, then, is to optimize time, data capture, and labor. The hazards include insufficient culling and poorly predicted labor requirements. Can one develop heuristics or some empirical base for making such decisions? That is one of our concerns in the paper and the overall project.

With anything less than complete data capture, the particular operationalization of CV data becomes vital. Even after whole sections of data are dismissed (in our case, such sections as conference papers, courses taught, internal working papers), one still must grapple with measuring the remainder. Is it important, for example, to capture not only article publications but also author numbers and author order? How does one represent data in more economical indices and, more to the point, how does one know which indices are most useful without sufficient original data to employ in indices? Are data best represented in arrays, across time, or in cross-sectional detail? To be sure, many of these answers depend upon specific hypotheses of specific studies, but if the CV database is to serve as a general resource for multiple research objectives, specific hypotheses provide little relief.

An interesting problem is in identifying long versus short CV versions. A great many CVs get condensed or truncated (e.g., "most recent publications") and the information that may seem important in an early career (e.g., all conference presentations) may be unimportant to the scientists later and may disappear entirely from the CV. The opposite problem is CV "embellishment" where every career detail gets recorded with seemingly equal weight. We are certainly able to delete extraneous information, but information not included is information lost. Aside from the database problems such variation can cause, there are significant validity issues and issues regarding coding accuracy, coding labor and duration, and coder fatigue. And, despite explicitly asking for long version CVs, about one in five of the CVs submitted for this study, were short versions.

Interestingly, the availability of CVs on the web has been helpful and hurtful to those interested in the CV as data. The popularity of the web has meant that a great many more CVs are accessible, but the institutionalization of websites has led to a stylistic conformance of CVs, which is not itself a problem, and, typically, significant abridgment, which can be a great problem. If the CV on the web is typically an institutional rather than individual marketing resource, the rational marketing approach is succinct information about more people, rather than detailed information about particular people.

By most any standard, the coding of more than a few CVs is a daunting task. We know that coding error rates from relatively tractable survey data range from about 5 to 10 percent.[33,34] What about more difficult, less obvious CV data? While coding errors can at least be determined with some ease, it is not clear even which standard is best for coding reliability.[35,36] Moreover, good measures of intercoder reliability require a good number of coders, again accelerating costs. Most important of all, however, is coding validity. Except for the most straightforward issues, CV coding is almost always sure to cause problems for any but the best-trained eye. Explaining to a coder how to deal with visiting professors working at (apparently) three different places, in two sites, with three ambiguous titles requires time, patience, and imagination. For example, the difference between a postdoc and a fellow may be vital in some instances, not others. And how does one determine if a proceedings publication is consequential when working in a number of very different fields. Is it possible to conveniently detail such matters for coders in anything less than a 50-page codebook?

Data collection and coding

CV Collection

To test the practicality of using CVs to study the careers of scientists and engineers, four data collection approaches were used: an NSF database search, an industry search, an Internet search, and a collection of CVs from a multi-institutional, microelectronics research center headquartered at the Georgia Institute of Technology. We purposely limited our collection to the broad fields of biotechnology (and related fields such as biochemistry and bioengineering) and microelectronics-related areas[37] to make the task a manageable one. The goal was to obtain an expected sample of 300-400 CVs from researchers with diverse backgrounds, working in diverse research contexts.

We expected that these methods would not only ensure that most, if not all, of the potential problems would surface (at least in these fields), but that they would provide an opportunity to test various data analysis methods, including descriptive statistics, OLS regression, and event history analysis. Thus, CVs were not collected to be representative of any particular group, but to allow us to develop and test coding procedures and data analysis approaches and to evaluate various methods of obtaining CVs.

For the NSF, industry, and research center collections, an email message was sent directly to potential respondents who were asked to submit a full CV via email. In contrast, for the Internet search, various search engines and search phrases were tested to identify a subgroup of web-posted CVs. Of the sample group, 50 CVs were solicited from industry scientists and engineers, 200 from NSF-funded academic researchers, 100 from the web, and all faculty and graduate students affiliated with the multi-institutional research center and its primary research program (210).

A sample of 200 researchers funded by NSF's biotechnology program and working at US institutions was obtained from the NSF awards database. This strategy has the main advantage of identifying a group of active researchers (in any given field(s)) whose email addresses are provided by NSF. An email was sent inviting the researchers to submit their full CV. Approximately 20 percent of the email addresses taken from the database were erroneous or obsolete. The researchers attempted to obtain a current address for all of undelivered emails using institutional directories via the worldwide web. No follow-up was done on the nonrespondents. Fifty-five CVs resulted. Four respondents formally refused and one claimed not to have been funded by NSF. The effective response rate was approximately 28 percent.

Prior to conducting the industry search, the Science Citation Index (SCI) was explored for the years 1995-99 using the key word "biotechnology." Over 3,100 titles were returned. Because searching each title for industry affiliation would have required extensive time and would not necessarily have yielded useful results, we opted to identify five journals[38] that were likely to draw readers and authors from industry. Two[39] of these were not available on the SCI and were substituted by two additional journals.[40] Fifty-nine email notes were mailed and 19 responses were received—a response rate of approximately 32 percent. Based on their email addresses, thirteen of the industry recipients did not reside in the U.S., seven appeared to work in government agencies (which we considered non-academic for our purposes), and two worked with university-affiliated hospitals.

Our third group of CVs was acquired via a passive search strategy over the Internet using popular web search engines. This approach has several advantages: it is non-intrusive, it utilizes CVs already available in the public domain, it is cost effective, and there is virtually no wait time. On the other hand, two major disadvantages were identified: the sample includes only people who posted their CV on the web, and web-posted CVs tended to lack detail.

Web searches were conducted using the keywords "curriculum AND vita*[41] AND biotechnology." We tested ten search engines: Alta Vista, Excite, Google, Goto, Hotbot, Infoseek, Look smart, Netscape, Snap, and Yahoo. The initial results using standard keywords were disappointing since these search engines turned up too many irrelevant web pages such as job search lists, academic newsletters, job announcements or résumé writing guide pages—"noise." Advanced search features were used to reduce the noise level dramatically.[42] Ninety-four CVs were retrieved from the web.

Our fourth approach to CV collection was through a pilot study we conducted of the Microelectronics Research Center (MIRC) at the Georgia Institute of Technology and its major multi-institutional collaborative research program, the Interconnect Focus Center (IFC). We sought to collect all of the CVs of researchers and graduate students affiliated with MIRC (146) as well as researchers affiliated with the IFC (64). Like the NSF and industry targeted collections, we sent email messages to the center's affiliates asking them to submit their CVs. Two follow up emails were sent to non-respondents. Our overall response rate was 41 percent for MIRC (52 percent for faculty and 37 percent for graduate students[44]) and 73 percent for the IFC component.

Coding Methodology

To develop a preliminary coding protocol, we reviewed a subset of the CVs from each of the four respondent groups to identify problems and potential solutions. Over 30 potentially useful variable "sets" were identified. However, many of these variable sets included multiple (i.e., up to 10) degrees received, multiple (i.e., up to 600) publications, multiple (i.e., up to 50) patents, and so forth. The number of variables for each respondent depends, unlike a questionnaire, on the length of the CV. Junior researchers could have as few as 25 variables per CV; seasoned veterans could have as many as 2,000. Several practice coding exercises were conducted to obtain information on intercoder reliability, to improve the coding protocol and process, and to minimize coding time. After these preliminary steps, the coding protocol was revised, retraining was implemented, and the coders proceeded to code the 281 CVs collected from the four respondent groups. The goal was to design the coding process so that a work-study student could be trained to code the typical CV with minimal reliability problems in 30 minutes or less.

To test intercoder reliability, we examined coding decisions of five coders on a subset of 37 variables from two sets of 10 CVs. Various measures of intercoder reliability have been proposed,[45-51] but for our purposes we found Crittenden and Hill's measure of intercoder reliability (Rs) the most useful and relevant. Tables 1 and 2 summarize the results from our first preliminary coding tests. Overall, the average reliability coefficient value of .766 on the first round of coding shows that we would be wise to heed the advice to further refine the instrument and coding scheme. While there is no widely accepted "threshold level" of intercoder reliability, for this particular coefficient anything below .850 should probably be considered problematic and anything below .600, outright unacceptable. Only 16 out of 37 items satisfied the .850 requirement. However, 7 out of 37 items fell below .600. The principle coding problems stemmed from the limited standardization in CV formats (possible international effects compound this problem), missing information, and coder error or the misinterpretation of data.

Insert Tables 1 and 2 approximately here

A closer inspection of the errors, however, demonstrated that many of them were due to coders coding information out of order (e.g., coding the second publication in the third publication variable spot) which compounded to ensure future errors.[52] In addition, coders had significant problems with the original codebook which was revised and resubmitted to a second test (which we label coding experiment 2 in Tables 1 and 2) using the same coders but new CVs. We expected there would be improvements due to the enhanced codebook as well as learning effects, and we were correct. Tables 1 and 2 show that the mean intercoder reliability rate increased to .805. There were 15 items that scored above the .850 level and three that scored below .600. While, in general, we felt that these intercoder reliability rates were still unacceptably low, we recognize that the complexity of the coding task at hand is such that we should expect rates below that achieved in more typical questionnaire coding. But to address the problem directly, we put in place a more elaborate coder training program and worked more closely with coders during the actual coding operations than we originally expected. We also relied more extensively on post hoc data cleaning than may be typical.

Interestingly, we found in both experiments that coding time was unrelated to error rate suggesting that our reliability problems were due either to the complexity of the CVs, the complexity of our coding scheme, the motor skills and knowledge of our coders, or all three. The average time to code a CV was 18 and 24 minutes on the two, small-scale coding experiments conducted on a subset of the variables and about 30 minutes (which ranged from 1 minute to 260 minutes) on average for the 281 CVs coded for the data analysis that is presented in the next section.

Findings and discussion

Aside from testing the quality of data in terms of its validity and reliability we sought to test the "usability" of the data in various modes of analysis. Three data analysis approaches were used: basic descriptive statistics, OLS regression, and event history analysis. The point of this analysis is not to record empirical findings about the careers of scientists and engineers, but to demonstrate that CV data can successfully be collected, coded, cleaned, and used in common statistical analyses.

Insert Table 3 approximately here

Table 3 shows the results of the basic descriptive statistics including the variable names and types, number of variables to a set, and number of coding categories for categorical data.[53] One data limitation that was identified through this analysis concerned the issue of missing data. In coding CVs the distinction between missing and null or zero data is not always clear. For example, if no patents or professional society memberships or grants appear on a CV does this mean the respondent has zero patents, memberships, or grants or has simply failed to list them? In addition, the means and ranges for various variables such as publications, awards, and patents show that the data are highly skewed, making statistical analyses based on normality assumptions questionable.

Second, we ran an OLS regression[54] with publication rate (i.e., number of publications per year after receipt of doctoral degree) as the dependent variable and number of publications during Ph.D. studies, total number of jobs, time in assistant professorship rank, number of professional society memberships, and a patent dummy[55] (0 = no patents, 1 = patents) as independent variables.

Insert Table 4 approximately here

Because the knowledge value model places emphasis on capacity building and human and social capital, we chose to include variables that reflect those issues (at least for preliminary demonstration). Number of Ph.D. publications is a good indicator not only of early ability but of early opportunity, accumulative advantage, and perhaps good mentoring. Total number of jobs may be thought of as a measure of professional job instability or even incompetence. Alternatively, it may be an indicator of strong job marketability or the possession of skills in short supply. However, from a human resources or knowledge value standpoint it may indicate diversity of professional experience and exposure to a large number of professional contacts and cultural environments. As a result, we chose to enter this variable in quadratic form, suggesting that those in the extreme may be professionally ill-adept, while those with some diversity of experience may be more productive. We wanted to test at least one time-dependent variable, duration of assistant professorship rank, as a measure of professional ability. Finally, number of professional society memberships was entered as a crude measure of the size of the professional network and the patent dummy as another measure of professional diversity, that is, commercial applications of research.

The results indicate that the coefficients on Ph.D. publications, duration of time in assistant professor rank, and having one or more patents were statistically significant.

The coefficient on duration of time in assistant professor rank is negative indicating the more time spent in that position the lower the publication rate of the researcher. Having patents and refereed publications during Ph.D. studies were positively associated with career publication rate. The adjusted R^2 for this model was a respectable .28.

Third, we modeled the rate of promotion of rank to full professor through methods known as event history analysis, dynamic analysis, or survival analysis.[56-59] While most past research has focused on how productivity affects advancement in academic rank, it is certainly not the only determinant. Indeed, a number of studies have underscored the explanatory potential of variables such as age,[60-64] gender,[65-70] field of research,[71] and prestige of doctoral department.[72-74]

Most of these studies employ cross-sectional or limited time-varying approaches. In contrast, CV data allow us to examine career advancement in a knowledge value framework as an inherently dynamic process. We estimated two dynamic models[75] that focus on the rate of advancement from either associate or assistant professor[76] to full professor status as a function of time and several covariates. The state of *not* achieving full professor rank given that the respondent had achieved assistant professor rank was coded as 0, while the state of *having* achieved full professor given a previous assistant professor job was coded as 1.

Model One tests the effects of the following covariates on the rate of promotion to full professor: publication rate (calculated as number of publications from the beginning of the first assistant professor job divided by the number of years since first assistant professor job); age cohort;[77] and disciplinary field of highest degree.[78] Model Two introduces two additional variables: gender (female=0, male=1) and a dummy variable denoting whether the respondent had registered at least one patent during the period of interest.

Insert Table 5 approximately here

In Model One (see Table 5), three covariates have a significant effect on the hazard function. The partial correlation R is highest for the field dummy 'engineering'. The risk ratio (exp(B)) for this indicator variable shows that, partialing out all other covariates, the hazard of promotion to full professor for engineers is almost 200 percent higher than the hazard for promotion of biologists and biomedical researchers. As would be expected, publication rate also affects positively the transition rate to full professor. With every additional publication per year, the hazard rate increases by 7 percent. Cohort 1976-85 impacts the hazard negatively. Controlling for other covariates in the model, the hazard of attaining a full professor status for this cohort is only 49 percent of the hazard for the cohort who received their first degree prior to 1961. It is not clear to us why this, rather than the most recent, cohort impacts the hazard rate negatively. It may be an artifact of our data.

In Model Two, where gender and patents are added, the publication rate is no longer significant. Cohort 1961-65 joins cohort 1976-85 in significantly affecting the estimated hazard in a negative direction as compared to the reference group (the earliest cohort). The effect size for the two cohorts is about the same. Finally, having one or more patents prior to rank advancement also significantly affects the hazard rate—the transition rate to highest academic rank is 133 percent higher for those with patents than those without, holding other covariates constant. Since these models are hierarchically nested, we can compare the fit of the two models by performing a χ^2 difference test, which in our case suggests that Model Two leads to only a minor improvement in goodness of fit over Model One.

Finally, the assumption of proportionality of the hazard which Cox regression is based upon holds with respect to one source of non-proportionality—population heterogeneity—but may not in the case of the second source (time dependence). We ran diagnostic plots (not shown here) stratifying the logarithmic survivor function separately by gender, cohort, and field. All three plots showed approximate parallelism between the strata lines which upholds the assumption for proportionality of the hazards. However, when we look at the shape of the survivor function there is some evidence for non-linearity, which may suggest further exploration and consideration of alternative survival analysis models. Overall, the two models demonstrate the utility of using CV data to dynamically model scientific productivity in a knowledge value framework.

Conclusion

Results from this preliminary assessment of the use of CVs as data for research on scientists' and engineers' career trajectories shows that the potential of this approach is mitigated by several practical problems, some easily remedied, some not. Our study seems at least to provide some notion of the likely magnitude of problems.

One of the most basic issues is the availability of CVs and best approaches to obtaining them. We found that obtaining CVs was sometimes more difficult than expected. Since CVs are routinely requested for all sorts of purposes, we expected the routine nature of our request would yield considerable returns. The results show that developing an adequate response likely requires considerable effort to ensure the accuracy of the address data. Allowing more time for response, in connection with a follow up, will likely yield better results as scientists are highly mobile and sometimes do not receive email quickly. Further, those receiving large quantities of email in all likelihood have some heuristic for triage and, quite likely, our request would not survive the triage.

The results for the web search for CVs were interesting and, despite disappointing returns, the learning curve was such that greater familiarity with particular search engines is likely to yield improvement. At this point, the returns for the web search entail considerable selection effects since there is no reason to believe that the persons with readily available web-based CVs are representative of the entire population of scientists and engineers. It seems likely that the number of CVs on the web will continue to increase for some time and, along with the increase, will tend to be more representative. However, if the web search has many flaws, the non-obtrusiveness of the web CVs has great appeal and, at least, this approach seems useful for preliminary studies, pretests, methodological tests, and, especially, hypothesis development. It provides an inexpensive means of getting started on CV-based research.

If the acquisition of the CVs was somewhat more difficult than we expected, coding was somewhat less of a time burden than expected. Coders were able to code CVs in less than 30 minutes each, despite a considerable volume of data and a relatively sophisticated codebook. The intercoder reliabilty levels were not acceptable in many cases. The coding challenge is considerable, especially if one anticipates using student coders. We found that the reliability and validity of CV data depends crucially on extensive coder training and a significant investment of time in data cleaning and quality checks. With these safeguards, we believe respectable reliability rates can be achieved.

It is clearly possible to perform relatively sophisticated statistical analyses, such as we did for this study, using the CV as a sole source of data. Both the OLS regression and event history analyses show significant and interesting potential findings that confirm the effects of age cohort, disciplinary field, and early publication productivity in advancement in rank and in overall publication output. What surprised us was the relationship between having patents and academic productivity and advancement. However, these two interdisciplinary fields defy the traditional categories of basic versus applied research and it is possible that their is more than a fair share of comingling between scholarly and commercial cues to knowledge value and productivity.

A less obvious, but potentially significant, issue in assessing the quality and utility of CV data is phenomenological in nature. Arguably, the CV means very different things to different scientists and engineers and the respective constructions of the CV may have implications for study objectives. This is redolent of the point made by Latour and Woolgar.[79] For example, the CV of the recent graduate is perhaps best thought of as a marketing tool whereby the CV's author seeks to maximize credibility[80] to potential employers, using such artifices as seem likely to achieve employment objectives. Similarly, late career scientists may well view the CV more as a historical record, focusing chiefly on the chronicle of output and activities. These are just two of the constructions one might envision. Each construction may well embody different motives, different communication strategies (including variance in communicated content), and different CV revision strategies. In all likelihood, purveying a CV on the web can be understood in part as a reflection of a particular set of constructions. The fact that so many scientists have multiple versions of CVs (some more suitable for obtaining grants, others more suitable for employment, still others more suitable for consulting or service) seems to reinforce the notion of multiple roles and constructs for CVs.

In sum, the use of CVs as serious data for social inquiry seems to us to have much potential. But despite the familiarity of this everyday artifact, knowledge of its social meaning, its research utility and the attendant practical problems in its research use is just beginning to accumulate. Using CVs for research is not exactly a brave new world, but an old world seen a new way.

Notes and references

1. B. Bozeman, J.S. Dietz, M. Gaughan, Scientific and Technical Human Capital: An Alternative Model for Research Evaluation, International Journal of Technology Management (forthcoming).

2. J.S. Dietz, Building a Social Capital Model of Research Development: The Case of the Experimental Program to Stimulate Competitive Research, Science and Public Policy, (forthcoming).

3. J.D. Rogers, B. Bozeman, Knowledge Value Alliances: An Alternative to R&D Project Evaluation, (in submission).

4. J. Mincer, Investment in Human Capital and Personal Income Distribution, Journal of Political Economy, (1958) Aug.

5. G. Becker, Investment in Human Capital: A Theoretical Analysis, Journal of Political Economics, 70 (1962) No. 5, S9-S49.

6. T.W. Schultz, The Economic Value of Education, Columbia University Press, New York, 1963.

7. M. Polanyi, The Tacit Dimension, Cox & Wyman, London, 1967.

8. M. Polanyi, The Logic of Tacit Inference. Knowing And Being, Routledge & Kegan Paul, London, 1969.

9. P. Bourdieu, The Forms of Social Capital, In J. G. Richardson (Ed.), Handbook Of Theory And Research For The Sociology Of Education, Greenwood, New York, 1986.

10. P. Bourdieu, L. Wacquant, An Invitation To Reflexive Sociology, University of Chicago Press, Chicago, 1992.

11. J.C. Coleman, Social Capital in the Creation Human Capital, American Journal of Sociology, 94 (1988) S95-S120.

12. J.C. Coleman, Foundations of Social Theory. The Belknap Press of Harvard University Press, Cambridge, 1990.

13. B. Bozeman, J.D. Rogers, J.S. Dietz, J. Park, The Research Value Mapping Project: Qualitative—Quantitative Case Studies of Research Projects Funded by the Office of Basic Energy Sciences, Final Project Report to the US Department of Energy, (1999) rvm.pp.gatech.edu.

14. E.M. Rogers, Diffusion of Innovations. Fourth Ed., The Free Press, New York, 1995.

15. P.E. Stephan, S.G. Levin, The Critical Importance of Careers in Collaborative Scientific Research. Revue d'Economie Industrielle, 79 (1997) 45-61.

16. D.K. Simonton, Creative Productivity: A Predictive And Explanatory Model Of Career Trajectories And Landmarks, Psychological Review, 104 (1997) No. 1, 66-89.

17. A.M. Diamond, An Economic Model of the Life-Cycle Research Productivity of Scientists, Scientometrics, 6 (1984) No. 3, 189-196.

18. A.M. Diamond, The Life-Cycle Research Productivity of Mathematicians and Scientists, Journal of Gerontology, 41 (1986) No. 4, 520-525.

19. G. Becker, Human Capital: A Theoretical and Empirical Analysis, with a Special Reference to Education, University of Chicago Press, Chicago, 1964.

20. S.G. Levin, P.E. Stephan, Research Productivity Over The Life Cycle: Evidence For Academic Scientists, American Economic Review, 81 (1991) No. 1, 114-132.

21. Stephan and Levin, op. cit.

22. P.E. Stephan, The Economics of Science, Journal of Economic Literature, 34 (1996) 1199-1235.

23. Stephan and Levin, op. cit.

24. J.S. Long, R. McGinnis, The Effects of The Mentor on The Academic Career, Scientometrics, 7 (1985) Nos. 3-6, 255-280.

25. B.F. Reskin, Scientific Productivity and the Reward Structure of Science, American Sociological Review, 42 (1977) June, 491-504.

26. H.A. Zuckerman, The Scientific Elite, Free Press, New York, 1977.

27. R.K. Merton, On The Shoulders Of Giants: A Shandean Postscript, University of Chicago Press, Chicago, 1993.

28. G.H. Elder Jr., Time, Human Agency, And Social Change: Perspectives On The Life Course, Social Psychology Quarterly, 57 (1994) 4-15.

29. J.S. Long, P. Allison, R. McGinnis, Rank Advancement in Academic Careers: Sex Differences and the Effects of Productivity, American Sociological Review, 58 (1993) 703-722.

30. L. Gomez-Mejia, D. Balkin, Determinants of Faculty Pay: An Agency Theory Perspective, Academy of Management Journal, 35 (1992) No. 5, 921-955.

31. Methodologically, CVs have been found to closely match information from other secondary sources such as the American Psychological Association's directory. However, these other secondary sources have been shown to undercount the number of published journal articles as compared to CVs. See J. Heinsler, R. Rosenfeld, Charting Academic Careers, Social Studies of Science, 17 (1987) 135-144.

32. S. Bonzi, Trends in Research Productivity among Senior Faculty, Information Processing & Management, 28 (1992) No. 1, 111-120.

33. F.J. Fowler, Survey Research Methods, revised edition, Sage, Newbury Park, 1988.

34. R.A. Singleton, Jr., B. Straits, M. Straits, Approaches to Social Research, Second Ed., Oxford University Press, New York, 1993.

35. G. Bohrnstedt, D. Knoke, Statistics for Social Data Analysis, Second Ed., F.E. Peacock, Itasca, IL, 1988.

36. E. Babbie, The Practice of Social Research, Sixth Ed., Belmont, Wadsworth Publishing, CA, 1992.

37. Biotechnology and microelectronics were chosen because they represent interdisciplinary fields where research is conducted both in academia and industry. This was expected to result in high variance among CV contents and formats which was ideal for this exploratory phase of the research.

38. Journal of Heart Valve Disease, Annals of Biomedical Engineering, Journal of Biomechanics, Echocardiography, and American Society of Echocardiography.

39. Journal of Echocardiography and Echocardiography.

40. Biotechnology and Applied Biochemistry and Biotechnology and Bioengineering (an electronic journal).

41. We used the asterisk as wildcard since both curriculum vita and curriculum vitae were used. Some web search engines did not allow using asterisk as wildcard and turned up zero hits. We used Boolean expression "vita or vitae" in those cases.

42. Two tactics were used to remedy noise problems. First, we employed various keywords in different combinations. Using Yahoo, the string "curriculum AND vita* AND biotechnology AND PhD" turned up about 800 hits, approximately half of which were relevant. Second, we used various advanced search features, such as Boolean expressions, to reduce the noise level. For example, for the Alta Vista search we added "NOT job NOT jobs" since most of the noise was related to job-

related websites. Refining the key words dramatically reduced the noise level. Recently developed search engines such as GOTO and HOTBOT seemed to employ different search algorithms; their noise levels were extremely low while they turned up relatively small numbers of hits. For example, with refined keyword combinations, "curriculum vita* + biotechnology - job - jobs," GOTO produced 197 hits with noise level around 10 percent. HOTBOT produced similar results with a similar level of noise. See S. Lawrence, C.L. Giles, Accessibility of Information on the Web, Nature, 400 (1999) No. 6, 107, for a discussion of internet searches and noise levels.

43. We believe that the low response rate for graduate students is due to the fact that many may not yet have CVs or résumés even though we used both term in our email solicitation.

44. W. Scott, Reliability of Content Analysis: The Case of Nominal Scale Coding, Public Opinion Quarterly, 19 (1955) 321-325.

45. R. Craig, Generalization of Scott's Index of Intercoder Agreement, Public Opinion Quarterly, 45 (1981) No. 2, 260-264.

46. G. Funkhouser, E. Parker, Analyzing Coding Reliability: The Random-Systematic-Error Coefficient, Public Opinion Quarterly, 32 (1968) No. 1, 122-128.

47. J. Fleiss, Measuring Nominal Scale Agreement Among Many Raters, Psychological Bulletin, 76 (1971), 378-382.

48. J. Landis, G. Koch, The Measurement of Observer Agreement for Categorical Data, Biometrics, 33 (1977), 159-174.

49. K. Crittenden, R. Hill, Coding Reliability and Validity of Interview Data, American Sociological Review, 36 (1971) No. 6, 1073-1080.

50. A. Montgomery, K. Crittenden, Improving Coding Reliability for Open-Ended Questions, Public Opinion Quarterly, 41 (1977) No. 2, 235-243.

51. It turns out that this is not at all a serious problem for data integrity. The data was subsequently sorted into chronological order to correct the problem.

52. For a complete list of variables and codes, visit rvm.pp.gatech.edu.

53. For the OLS regression, we selected only those cases where we had a full and complete CV from the respondent, n = 190.

54. We chose to enter this variable in dichotomous form because it was highly skewed and violated the normality assumption of OLS.

55. G. Carroll, Dynamic Analysis of Discrete Dependent Variables, Quality and Quantity, 17 (1983) 425-460.

56. N. Tuma, M. Hannan, Social Dynamics, Academic Press, Orlando, FL, 1984.

57. H. Blossfeld, A. Hamerle, K. Mayer, Event History Analysis, Lawrence Erlbaum Associates, Hillsdale, NJ, 1989.

58. K. Yamaguchi, Event History Analysis, Sage, Newbury Park, CA, 1991.

59. J. Cole, S. Cole, Social Stratification in Science, University of Chicago Press, Chicago, 1973.

60. S. Cole, Age and Scientific Performance, American Journal of Sociology, 84 (1979) No. 4, 958-977.

61. J. Hurlbert, R. Rosenfeld, Getting a Good Job: Rank and Institutional Prestige in Academic Psychologists' Careers, Sociology of Education, 65 (1992) No.3, 188-207.

62. Stephan and Levin, op. cit.

63. H. Zuckerman, R.K. Merton, Age, Aging, and Age Structure in Science, In R.K. Merton ed., The Sociology of Science: Theoretical and Empirical Investigations, University of Chicago Press, Chicago, 1973, 497-559.

64. J. Cole, H. Zuckerman, The Productivity Puzzle: Persistence and Change in Patterns of Publication of Men and Women Scientists, Advances in Motivation and Achievement, 2 (1984) 217-258.

65. M. Fox, C. Faver, Men, Women, and Publication Productivity: Patterns among Social Work Academics, The Sociological Quarterly, 26 (1985) 537-549.

66. Long, et al., op. cit.

67. M. Fox, Women and Scientific Careers, In S. Jasanoff, G. Markle, J. Petersen, T. Pinch, eds., Handbook of Science and Technology Studies, Sage, Thousand Oaks, CA, 1995, 205-223.

68. J.S. Long, M.F. Fox, Scientific Careers: Universalism and Particularism, Annual Review of Sociology, 21(1995) 45-71.

69. Y. Xie, K. Shauman, Sex Differences in Research Productivity: New Evidence About an Old Puzzle, American Sociological Review, 63 (1998) 847-870.

70. L. Hargens, W. Hagstrom, Scientific Consensus and Academic Status Attainment Patterns, Sociology of Education, 55 (1982) No. 4, 183-196.

71. Ibid.

72. J. Hurlbert, R. Rosenfeld, Getting a Good Job: Rank and Institutional Prestige in Academic Psychologists' Careers, Sociology of Education, 65 (1992) No. 3, 188-207.

73. Long et al., op. cit.

74. The most general model that we employ is known as Cox regression and has the following functional form: $h(t) = [h_0(t)]\exp\{b_1X_1 + \ldots..b_kX_k\}$, where h(t) is the hazard function (which in our case will be the same as the transition rate r(t) since we have only one destination in a two-state model), $h_0(t)$ is the baseline hazard function (when the Xs are set to 0), the bs are the regression coefficients, and the Xs are the covariates. Alternatively, we can take the natural logarithm of both sides and rewrite the equation as: $\ln h(t) = \ln h_0(t) + b_1X_1 + \ldots..b_kX_k$ The logarithm form of Cox regression makes sure that the hazard will not show a negative value (a theoretical impossibility).

75. We chose a status variable "promotion from either associate or assistant professor to full professor" to incorporate about a dozen cases in which scientists from older cohorts advanced directly from assistant to full professor rank.

76. Bachelor's degree received: 1 = 1960 or earlier, 2 = 1961-65, 3 = 1966-75, 4 = 1976-85, 5 = 1986 or after.

77. 1 = biology, 2 = engineering, 3 = physical sciences.

78. B. Latour, S. Woolgar, Laboratory Life: The Construction of Scientific Facts, Princeton University Press, Princeton, NJ, 1986.

79. B. Bozeman, The Credibility of Policy Analysis: Between Method And Use, Policy Studies Journal, 14 (1986) No. 4, 519-539.

Acknowledgements

The authors gratefully acknowledge the support of the U.S. Department of Energy (DE-FG02-96ER45562) and the U.S. National Science Foundation (SBR 98-18229). The opinions expressed in the paper are the authors' and do not necessarily reflect the views of the Department of Energy or the National Science Foundation. This work was performed under the *Research Value Mapping Program* [http://rvm.pp.gatech.edu] of the Georgia Institute of Technology. The authors wish to thank Jeff Bournes, Marie Chesnut, Jenna Dubois, Kathrin Jackson, Zahi Karam, Jungki Kim, Andy McNeil, Seth Sobel, Ryoung Song, and Larry Wilson for their assistance. The authors also thank Monica Gaughan and Juan Rogers for their helpful advice.

Table 1. Intercoder Reliability and Time of Coding for 10 CVs

Curriculum Vita	**Rs(i) for Coding Experiment 1**	**Coding Time (in minutes) Experiment 1**	**Rs(i) for Coding Experiment 2†**	**Coding Time (in minutes) Experiment 2**
1	.897	23.0	.938	18.0
2	.797	31.0	.765	21.0
3	.651	27.8	.881	18.4
4	.839	23.4	.709	14.4
5	.868	24.0	.792	15.6
6	.608	15.0	.830	16.0
7	.756	30.4	.830	13.8
8	.800	19.4	.630	21.8
9	.728	19.8	.832	18.0
10	.714	22.2	.849	10.0
Mean	.766	23.6	.805	16.7
Std. Dev	.090	5.02	.088	3.51

Note: Rs(i) stands for resume intercoder reliability.

† Coding experiment 2 used the same coders with different CVs and an improved coding protocol.

Table 2. Intercoder Reliability for Coding 37 Items

Item #	Rs(i) Coding Experiment 1	Rs(i) Coding Experiment 2†	Item Name
1	.933	.682	CV version is full or partial
2	.933	1.000	Sex of respondent
3	.960	1.000	Year of birth
4	.920	.800	National origin
5	.920	.780	Citizenship
6	.880	1.000	Degree type of first degree
7	.690	.840	Degree field of first degree
8	.860	.880	Degree type of second degree
9	.880	.670	Degree field of second degree
10	.960	.960	Degree type of third degree
11	.810	.800	Degree field of third degree
12	1.000	1.000	Degree type of fourth degree
13	1.000	1.000	Degree field of fourth degree
14	1.000	1.000	Degree type of fifth degree
15	1.000	1.000	Degree field of fifth degree
16	* .560	* .570	Job title of first job
17	* .490	* .490	Job title of second job
18	.610	.770	Job title of third job
19	.680	.680	Job title of fourth job
20	.600	.410	Job title of fifth job
21	.740	1.000	Publication type of most recent pub.
22	.790	.960	Publication type of second most recent pub.
23	.630	.860	Publication type of third most recent pub.
24	* .550	.860	Publication type of fourth most recent pub.
25	.880	.760	Publication type of fifth most recent pub.
26	.620	.780	Dollar amount of first grant or contract
27	* .570	* .520	Funding source of first grant or contract
28	.630	.760	Dollar amount of second grant or contract
29	* .580	.680	Funding source of second grant or contract
30	.690	.760	Dollar amount of third grant or contract
31	* .560	.710	Funding source of third grant or contract
32	.690	.760	Dollar amount of fourth grant or contract
33	.660	.620	Funding source of fourth grant or contract
34	.660	.780	Dollar amount of fifth grant or contract
35	* .580	.690	Funding source of fifth grant or contract
36	.893	.960	Year of first patent
37	.933	1.000	Was first patent licensed or sold?
Mean	.766	.805	
Std. Dev.	.164	.163	

Notes: Rs(i) stands for item intercoder reliability.
† Coding experiment 2 used the same coders with different CVs and an improved coding protocol.

Table 3. List of variables with descriptive statistics.

Variable Name or Variable Set Name (number of variables in a set)	Variable Type (number of codes for categorical variables)	Range	Mean	Number of Missing Cases (N = 281) +
Coding time (in minutes)	Continuous	(1,260)	29.6	8
CV version	Dichotomous (2: not full=0, full=1)	(0,1)	.676	1
Sex	Dichotomous (2: female=0, male=1)	(0, 1)	.797	27
Year of birth	Continuous	(1917, 1976)	1951	199
Citizenship	Categorical (100)			212
National origin	Categorical (100)			210
Year of bachelor's degree	Continuous	(1938, 2000)	1977.6	36
Bachelor's degree institution	String			33
Bachelor's degree field	Categorical (150)			42
Year of master's degree	Continuous	(1939, 2000)	1980.5	118
Master's degree institution	String			
Master's degree field	Categorical (150)			126
Year of doctoral degree	Continuous	(1941, 1999)	1980.7	68
Doctoral degree institution	String			
Doctoral degree field	Categorical (150)			75
Publication type (600)	Categorical (9)			
Publication year (600)	Continuous			
Total number of publications	Continuous	(0, 553)	41.4	
No. single-author, refereed journal articles	Continuous	(0, 59)	2.8	
No. first author on multi-author refereed journal articles	Continuous	(0, 134)	7.3	
Number of multi-author (not first)refereed journal articles	Continuous	(0, 403)	24.9	
Total number of books or book chapters	Continuous	(0, 115)	4.5	
Total number of other refereed publications	Continuous	(0, 92)	0.4	
Job type (30)	Categorical (46)			
Job Institution (30)	String			
Total number of jobs	Continuous	(0, 23)	5.5	
Year became assistant professor	Continuous	(1955, 1999)	1982.6	157
Year became associate professor	Continuous	(1955, 1999)	1984.1	171
Year became full professor	Continuous	(1958, 2000)	1985.4	184
Dollar amount of grant or contract (50)	Continuous			
Funding source of grant or contract (50)	Categorical (22)			
Total number of grants and contracts	Continuous	(0, 50)	3.3	190
Avg. grant/contract size (those w/ awards)	Continuous	(129, 5325000)	359,749	218
No. professional society memberships (25)	Continuous	(1, 25)	4.5	46
Number of patents (200)	Continuous	(0, 199)	8.7	124

Note: Variables in italics are derived variables. + Missing data are problematic with some variables in that it is not clear if the data are missing or if they represent null or zero data (e.g., missing grant award information may indicate respondent no grants or missing information).

Table 4. OLS Regression Coefficients.

Independent Variable	Unstandardized Coefficient	Standardized Coefficient
Number of Ph.D. publications	.401** (.135)	.352
Number of professional society memberships	.231 (.173)	.170
Duration in rank of assistant professor	-.476* (.204)	.204
Number of jobs (squared)	-.005 (.004)	-.155
Has any patents?	2.063* (.877)	.276
Constant	3.582*	1.554
Adjusted R^2	.278	

Notes: Numbers in parentheses are standard errors.
Dependent variable is publication rate.
* $p < .05$ ** $p < .01$

Table 5. Event History Analysis for Promotion to Full Professor: Two Cox Regression Proportional Hazards Models

	Model 1				Model 2			
Independent Variables	B	Wald	R	Exp(B)	B	Wald	R	Exp(B)
Publication rate	.07*	3.26	.05	1.07	.051	1.76	.00	1.05
Cohort 1961-65	-.73	2.32	-.03	.48	-.88*	3.39	-.06	.41
Cohort 1966-75	.19	.27	.00	1.21	.13	.13	.00	1.14
Cohort 1976-85	-.71*	3.36	-.06	.49	-.78*	3.78	-.06	.46
Cohort 1986-00	-.14	.02	.00	.87	.13	.01	.00	1.14
Engineering	1.09***	9.73	.13	2.99	.95**	6.49	.10	2.59
Physical sciences	.56	2.64	.04	1.75	.45	1.48	.00	1.57
Gender (male=1)	-----	-----	-----	-----	.34	.53	.00	1.41
Patent (yes=1)	-----	-----	-----	-----	.85**	5.96	.09	2.33
-2Log Likelihood	457.76				433.85			
N of observations	124				124			

Note: For Model 1 the likelihood ratio $\chi2$=18.54,df=7,p<.01 and the overall $\chi2$=18.03,df=7,p<.05. For Model 2 the likelihood ratio $\chi2$=22.85,df=9,p<.01 and the overall $\chi2$=23.12,df=9,p<.01. The reference (omitted) category for the cohort dummies is cohort<1961; the reference category for engineering and physical sciences is 'biology'.

*p<.10 **p<.05 ***p<.01 (two-tailed tests)