DO FIRMS CHANGE CAPABILITIES BY HIRING NEW PEOPLE?

A STUDY OF THE ADOPTION OF SCIENCE-BASED DRUG DISCOVERY

Nicola Lacetera, MIT
Iain M. Cockburn, Boston University and NBER
Rebecca Henderson, MIT and NBER

DRAFT

February 20th, 2004

Financial support from the MIT Program on the Pharmaceutical Industry and the MIT Program on Innovation in Markets and Organizations is gratefully acknowledged. Michael Rees made an invaluable contribution in the construction of part of the data. We also thank Joel Baum, Anita McGahan, Neer Asherie, Ernst Berndt, Kevin Boudreau, Francesco Lissoni, Lorenzo Zirulia; and participants to the MIT Strategy Proseminar and the MIT/Harvard Economic Sociology Workshop for their comments and suggestions. All remaining errors are our own.

ABSTRACT:

Do firms build new capabilities by hiring new people? We explore this question in the context of the pharmaceutical industry's movement towards science-driven drug discovery. We focus particularly on the potential problem of endogeneity in interpreting correlation between hiring and changes in organizational outcomes as evidence of the impact of new hires on the firm, and on the more fundamental conceptual question of the conditions under which hiring might be a source of competitive advantage, given the well known objection that resources that are freely available through the market cannot be a source of differential capabilities. Using data on the movement and publication of "star" scientists, we find that the adoption of science based drug discovery within the firm is closely correlated with the hiring of star scientists. This correlation appears to be reasonably robust to a number of controls for endogeneity. We also show that the hiring of highly talented scientists appears to have a significant impact on the behavior of scientists already working within the firm. We interpret this as consistent with the idea that hiring may change organizational capabilities through the interaction of new talent with the policies, routines and people already in place within the firm.

1. Introduction

Recent work in strategic management and in organizational theory more broadly has highlighted the critical question of how organizational capabilities change. Work by Levinthal (1997), Milgrom and Roberts (1990), and Rivkin (1998, 2000) has highlighted the degree to which complementarities between organizational practices make changes in "activity systems" a particularly perilous activity – an observation shared by a number of scholars writing in the tradition of population ecology (see, for example, Barnett and Carroll (1995)).

Nonetheless, we have some evidence that firms do change organizational capabilities, and that these changes may significantly enhance long term performance (Cockburn et al. (2000)). Given the critical role of organizational capabilities in shaping long term strategic advantage, building a better understanding of the mechanisms that firms use to change them is thus of central interest to the study of strategic management.

In this paper we explore the degree to which the hiring of particularly skilled employees contributes to the evolution of a firm's organizational capabilities. The belief that hiring may play an important role in the building of new capabilities has a long history in the scholarly literature, and is a staple of the popular business press. For example, movement of key individuals from universities to firms appears to be amongst the most effective mechanisms of knowledge transfer between these types of organizations (Dasgupta and David (1994), Zucker and Darby (1997)). Several writers have claimed that particularly skilled employees are critical to a firm's ability to integrate different, diffused forms of knowledge and capabilities (Henderson (1994), Grant (1996)), as well as to the productivity of the other members of the teams with whom they work (Hamilton et al. (2003)). Similarly Almeida and Kogut (1999), Song et al. (2003), and Breschi and Lissoni (2003) use patent citation analyses to show that knowledge transfer amongst organizations is closely correlated with the movements of key people.

Work in this tradition commonly models the contribution of skilled workers to the development of new capabilities as occurring essentially through the transfer of knowledge "embedded" in individuals, that is to say, acquisition of human capital by the firm. But a number of studies have explored the broader question of the degree to which hiring key individuals may also bring about more fundamental change in the organizational capabilities of the firm.[1] McKelvey (1982) argues that key workers may act as "carriers" of organizational capabilities or as "genes," and Penrose (1959) hints at a similar point. In a sequence of important empirical papers, Zucker, Darby and collaborators find that the birth and development of new biotechnology firms are positively correlated with the presence of academic star scientists in the same geographic area; that the enhancement in performance comes from those stars who

[1] A long tradition of research has analyzed the characteristics of senior management teams, suggesting that the composition and background of top managers has a significant impact on the policies and performance of the firm. See, among others, Barker and Mueller (2002); Bertrand and Schoar (2002) and Murmann and Tushman (1997).

actually have some form of ties with the firms; and that these ties positively affect the market value of biotech firms (see Zucker et al. (1998), Darby et al. (1999) and related papers). In a case study on the adoption of biotechnology by a large incumbent firm, Zucker and Darby (1997) stress that hiring highly skilled scientific personnel played a critical role in the firm's organizational, strategic and technological transformation.

Studies of this nature raise two critical methodological issues. The first is that of simple endogeneity or selection bias. Of course, a correlation between the hiring of new employees and a change in the organizational capabilities of the firm does not necessarily imply that the former *causes* the latter. It may be the case, for example, that the adoption of a new strategy leads simultaneously both to a deliberate investment in new organizational processes and routines and to the decision to hire new people with new skills. Firms which have adopted such a strategy will "select into" hiring new people, with the result that both hiring rates and organizational capabilities may change, without hiring having caused the change in organizational capabilities.

The second question is more fundamentally conceptual. As work in the resource-based tradition has long stressed, if firms are homogeneous in all respects, and factor markets are perfectly competitive, highly skilled workers are equally valuable to all firms, competing bids from firms will result in any potential rents accruing to the workers rather than to the firms that seek to hire them (Barney (1986), Peteraf (1993)). Since labor markets, even for quite specialized workers, would seem to be prototypically competitive this raises the question: How can human capital resources hired in a competitive factor market be a long term source of competitive advantage?

One potential answer, of course, is that particular types of workers might be "complementary", in the technical sense, to particular systems of organizational practice, including incentive systems, formal structures and informal cultural norms and ways of working (a systemic perspective shared by Athey and Stern (1998), Ichniowscki, Shaw and Prennushi (1997)). From this perspective hiring may "cause" a change in organizational capability, but only to the degree that hiring is accompanied by simultaneous changes in other aspects of the firm's ways of working. In this sense, endogeneity problems are also a consequence of the presence of important complementarities.

A number of recent studies have highlighted the importance of these issues. For example Groysberg (2001) finds that job-switching of star financial analysts reduces the short term performance of both the individual worker and their new employer. The effect is stronger when an analyst moves to lower ranked firms; this suggests that some set of complementary factors is required for stars to improve performance. Rao and Drazin (2002) explicitly recognize that other characteristics of a firm may drive the decision to hire, and control for this through a two-step sample selection analysis. They find that under-performing firms are more likely to hire from competitors and that such hiring has a positive impact on the ability of such laggards to introduce new financial products in the US mutual fund industry.

Song et al. (2003) find that inter-firm knowledge transfer through the mobility of engineers is more likely to occur when the hiring firm is less exploitative of its accumulated knowledge[2]. However none of these studies explore the degree to which hiring of new workers affects the performance of employees already employed at the firm.

In this paper we explore the degree to which the hiring of "star" scientists was instrumental in changing the ability of pharmaceutical firms to adopt science-based drug discovery – a fundamental change in the nature and organization of pharmaceutical research. Science based drug discovery (sometimes known as "rational" drug design) is a method for discovering drugs that relies on a deep knowledge of the mechanisms underlying disease which replaced the so-called "random screening" method of finding new drugs that dominated the industry for much of the 20th century. [3] The adoption of science-based drug discovery provides a particularly interesting setting in which to study questions about the role of hiring new workers in the process through which organizational capabilities are developed and transformed. The move to a more science-driven research process appears to have been a difficult one that required the adoption of a range of new organizational practices, took a surprisingly long period of time, and acted as a significant source of competitive advantage for those firms who were able to adopt it first (Gambardella (1995); Cockburn et al. (2000)). Furthermore, during this period of technological upheaval, industry participants often suggested that the successful adoption of the new techniques required the infusion of new blood and claimed that they were aggressively attempting to hire new researchers (see, for example, Whaldholz, 1985).

There are a number of pathways through which newly hired scientists might plausibly have been able to assist firms in making the switch to science driven drug discovery. One of these is simply through the addition of their own knowledge and skills. A firm might hire an expert in the mechanisms of diabetes, for example, to spearhead a new diabetes program. This individual could be expected not only to start doing science-driven drug discovery in the area of diabetes, but also to transfer their knowledge of state-of-the-art research methods to new colleagues.[4] More subtly, the presence of "stars" may change behavioral norms, or legitimize or otherwise enable behavior by existing employees that is an important aspect of doing science-driven drug discovery. For example, researchers engaged only in applied science may get involved in more basic, long-term oriented activities, particularly publishing and attending conferences if guided by some outstanding colleagues. Stars may also influence the hiring of other scientists, thanks to their ability to better screen young talents, and/or to their ties with the scientific

[2] Long and McGinnis (1981) and Huckman and Pisano (2003) offer evidence of similar phenomena.

[3] It is important to note that science-based drug discovery is not the same thing as the adoption of "biotechnology", or the use of large molecular weight molecules as drugs. In modern firms the line between "biotechnology" and "science driven drug discovery" has blurred as researchers increasingly rely on the tools of genetics and proteomics to find new drugs, but during the period covered by our data the distinction between them was quite clear. Gambardella (1995); Galambos and Sewell (1995); Henderson (1994); Cockburn, Henderson and Stern (1999a).

community (Cockburn et al., 1999b). Ultimately, hiring such individuals may therefore affect the total research output of the firm, above and beyond the projects that they themselves pursue.

We use a novel dataset composed of the "star," or most productive, scientists who joined one of 21 major pharmaceutical firms over the period 1980-1994. Several studies have noted that the distribution of scientific productivity is highly skewed, and that it is the highly productive scientists at the right tail of the distribution who are likely to be most effective in changing the firms that they join (Zucker, Darby and Torero (2002)). By following their publication records, we are able to see both when these top researchers joined one of the firms, and where they came from.

We begin by exploring the correlation between hiring and the adoption of the techniques of science-driven drug discovery at the firm level, using both instrumental variables and two-step estimation in an attempt to control for possible endogeneity. We then turn to an analysis at the level of the individual, and explore the degree to which the hiring of star scientists changed the behavior of scientists already employed by the firm. We reason that if star scientists change the capabilities of a firm merely by increasing its stock of high quality human capital, then the publication rates of the existing scientific labor force will be unaffected by their presence. Instead, we find that the individual productivity of existing scientists is positively affected by both the adoption of "science-driven" policies and by the presence of an internal community of star scientists. We interpret this result as consistent with the hypothesis that star scientists change the capabilities of organizations by complementing or combining with the unique set of individuals and organizational policies in place at the firm, and suggest that it highlights the importance of exploring the nature of this interaction in more depth.

The rest of the paper is organized as follows. Section 2 describes our data, introduces our empirical strategy and outlines the implementation of the statistical analysis. The results are presented and discussed in Section 3. Section 4 concludes.

2. Empirical analysis

We base our analysis on a sample of 21 major pharmaceutical companies,[5] observed over the period 1980-1994. Though not a comprehensive sample, we believe these firms to be reasonably representative of the research-based "core" of the industry. Between them they accounted for roughly 50% of US pharmaceutical sales over the period. We chose 1980-1994 as our sampling frame since this was the period during which the techniques of science-based drug discovery transformed the industry. Captopril, the first major drug commonly acknowledged to have been discovered using the new

[4] As Zucker et al. (1998) point out, at least in the case of biotechnology, much of the relevant knowledge of new techniques was possessed by a surprisingly small number of skilled individuals.

[5] The 21 firms are Abbott, Beecham, Bristol-Myers, BMS, Ciba-Geigy, Eli Lilly, Fuijsawa, Glaxo, Hoechst, Hoffman La Roche, Merck, Pfizer, Sankyo, Searle, SKB, Smithkline, Squibb, Takeda, Upjohn, Burroughs Wellcome, and Yamanouchi.

techniques, was first marketed in 1981, and by 1994 the transformation of the industry was widely acknowledged to be complete (Henderson et al. (1999)).

The data: Identifying star scientists

We identified the "star" scientists employed by these firms through a two step procedure. Using publication data from the ISI Science Citation Index (SCI), we first identified every publishing scientist who could unambiguously be determined to have been affiliated with one of the firms in our sample between 1980-1994. This is not straightforward: papers can be attributed to organizations because the SCI lists addresses for authors, however most papers have multiple authors, and the SCI does not directly link authors to addresses (or necessarily provide addresses for all of the authors.) We therefore determined the affiliation of authors by searching the global set of publications for each firm for papers with only a single address.[6] The authors on these publications could then be unambiguously assigned to a particular firm at a given point in time. For each of the scientists thus identified, we then re-searched the SCI database for all the publications in which he or she was one of the authors. This search resulted in the identification of 36,314 scientists and 191,288 papers. We then defined "star" scientists as those whose three-year moving average of annual publications was greater than 5 for at least one year. This procedure identified 936 stars. Notice that we define stars according to any such episode of above-average publishing. While there may be a difference between those scientists who already have an outstanding publication record before joining a firm, and those who gain their star status once they are inside a firm, in the majority of cases, the scientists in our sample appear to have already shown their potential.[7] We plan to explore this issue in more depth in future work.[8]

Figures (1), (2) and (3) show, respectively, the total number of star scientists employed by the firms in our data set over time, those hired in each period, and the average number of publications per head. About 320 star scientists were on our firms' payrolls in 1980, while by 1992 there were more than 800. The average firm hired about 3.7 top scientists each year, and these scientists were active in a firm for, on average, 8.4 years.

In order to explore the degree to which it is reasonable to assume that the first year in which a scientist publishes with a particular firm is the first year in which they are employed there, we attempted

[6] The criterion is similar to that used in other studies (Zucker, Darby and Torero, 2002; Cockburn and Henderson (1998)). A potential shortcoming of this method of identifying the institutional affiliations of individuals is that it eliminates those scientists who never published with a team composed only of researchers with the same affiliation. However, this problem is likely to be of limited importance here since we consider only scientists with high publication rates.

[7] Even "rookie" recruits have generally held positions as post-docs before joining a firm and arrive with a track record of publications. Our controls for endogeneity, which we discuss later, also indirectly account for this problem.

[8] We could use other measures of scientific productivity, such as weighting publications by the relevance of the journal or the number of citations received. However these methods are not free of problems. For example, it is well known that citation is often highly ritualized, and many citations, per se, may mean a negative as well as a positive opinion about an article. The use of impact factors has also been criticized since, among other reasons it is based on how many citations the average article in a journal receives in a relatively short period of time after publication.

to obtain information on the previous affiliation of each scientist. We tried to be as careful as possible in considering the last names of each scientist, his or her first and middle name initials, and discipline, location and coauthoring patterns, in order to maximize the probability of observing actual changes in affiliations for a given researcher. Using this procedure we were able to obtain information on previous affiliations for about 420 of the 936 scientists in our data base. These scientists were all publishing with a different address in the two years immediately before the year in which they started publishing within one of our firms, so we are reasonably confident that for those individuals for whom we cannot identify prior activity, we can take first publication date as the hiring date.

Of those scientists for whom we can identify a previous affiliation before they were hired by one of our sample firms, about 91% held a position in a non-business organization such as a university, research institute, or hospital. There is an increase in hiring from business organizations over time. None of the scientists hired in 1981 came from a firm, whereas in 1992 approximately 17% had previously been employed in a for-profit organization. This tracks the overall trend in the share of PhDs in life sciences working in business organizations: according to the National Academy of Science (NAS, 1998), the percentage of Life Science PhDs employed in industry rose from 10% in 1973 to around 19% in 1995 [9].

Movements between firms within the sample are relatively rare. 84 scientists, or about 9% of the sample, held a job position in more than one of the sample firms for a total of 96 changes of employer (74 scientists switch once, eight scientists switch twice, and two scientists switch three times). Thus notwithstanding a fairly high overall level of hiring activity throughout the period, movements between major firms are a relatively marginal phenomenon.[10] In what follows, we explore whether scientists coming from competitors appear to have a different impact on the firm than those coming from "other" organizations.

Despite these small numbers, there does seem to be some evidence that firms differ in their relative position in the larger network within which human capital is exchanged. Some firms hire many scientists but "send" only a few to their competitors. Others send many scientists to their rivals. Some firms, like Merck, are active both as recruiters and as sources for long periods of time (Table 1).

For the purposes of our statistical analysis we omit those scientists who were already employed by a firm in 1980, and those who start publishing in 1993 or 1994, in order to address the truncation problem of our data. This reduces the relevant sample to 580 researchers.

Empirical strategy

[9] We have some detailed information (87 observations) on affiliations of these scientists after they exit from the set of firms in our sample. The majority of these scientists become affiliated with a non-business organization.

[10] This finding is consistent with results on coauthoring behavior of researchers affiliated to business organizations reported in Cockburn and Henderson (1998): while such practice is present and growing, it mainly concerns relations between firms and research oriented and/or publicly funded organizations like universities and hospitals; the share of papers coauthored by scientists of two or more different business firms is small.

We build our analysis of the effect of hiring talented scientists on the adoption of science-driven R&D at the firm level using a basic adoption equation:

$$y_{it} = \alpha + \beta t + \gamma h_{it} + \delta x_{it} + \varepsilon_{it} \quad (1)$$

where y_{it} is our measure of the degree of adoption of science-driven discovery for firm *i* in period *t*; *t* is a time trend; h_{it} represent our hiring variables; x_{it} includes control variables; ε_{it} is an error term and $\alpha, \beta, \gamma,$ and δ are parameters to be estimated. An estimate of $\gamma > 0$ would be consistent with the adoption of science-driven drug discovery being at least partially driven by the hiring of star scientists.

This reduced-form specification immediately raises the problem of endogeneity discussed in the introduction. To address this issue, as a first step we include a number of control variables in the regression, and lag the explanatory variables. Since this may not fully control for either potential omitted variable problems or for possible selection bias, we take a number of additional steps.

We first exploit the panel structure of the data. We can rewrite the error term in the basic empirical model (1) as follows:

$$\varepsilon_{it} = u_i + \xi_{it} \quad (2)$$

That is, we decompose ε_{it} into a time varying component (ξ_{it}) and into a time invariant, firm specific part (u_i). If we assume that the unobserved characteristics that are correlated with the hiring variable are firm-specific and time-invariant then the term u_i will account for them, and a fixed effect regression, taking deviations from the mean (at firm level) will correct the potential bias. Second, we estimate (1) using instruments, and third we attempt to control for potential self-selection bias by using two stage estimation.

Our disaggregated analysis of productivity at the level of the individual scientist is based on a production function for publishing activity:

$$PUBS_{ijt} = f(y_{it}, h_{it}, z_{ijt}; \varphi; \zeta_{ijt}) \quad (3)$$

where *j* is an index for each single scientist, y_{jt} is our measure of the firm's adoption of science driven drug discovery, h_{it} is a measure of the presence of other skilled scientists, z_{ijt} is a vector of controls, φ is a vector of parameters to be estimated and ζ_{ijt} is an error term. By carefully constructing our vector of control variables we attempt to control for the base rate propensity of each scientist to publish. We then interpret significant coefficients on the degree to which the firm has adopted science driven drug discovery and on the number of other star scientists present in the firm as evidence that each star's behavior both shapes and is shaped by his or her organizational context.

Statistical implementation: dependent and independent variables, controls, and functional forms

Firm level analysis

Our measure of the degree of adoption of science-driven drug discovery is PUBFRAC: the fraction of those individuals whose names appear on a patent application in a given year who also appear

as authors on papers published within two years of the patent application. Details of the construction of this variable are given in Cockburn et al. (2000). Briefly, it attempts to incorporate the degree to which, inside a firm, those researchers who are directly involved in the drug discovery process are also participating in scientific publication. As it is measured as a share, in principle it captures the propensity of the firm's researchers to engage in basic science, independent of the scale of the firm.[11] PUBFRAC is correlated with a number of other measures of the degree to which a firm has adopted the tools of science-based drug discovery, including the number of papers published by the firm and qualitative, interview-based measures of the same concept (Cockburn et al. (2000)).

Since hiring star scientists will, almost by definition, increase a firm's measured level of PUBFRAC, we exclude their publication and patenting behavior from its construction. Our regression thus explores the impact of the hiring of star scientists on the propensity of *other* scientists within the firm to publish and patent simultaneously. Figure (4) illustrates trends in PUBFRAC over time, showing the gradual adoption of the techniques of science-driven drug discovery across the sample.

Following the classic diffusion literature (see e.g. Griliches (1957)), we transform the dependent variable to yield a continuous variable on the whole $(-\infty,+\infty)$ interval with a sigmoid form imposed on the degree of adoption. The simplest of these transformations is the so called log odds: $log[y_{it}/(1 - y_{it})]$, where log(.) is the natural logarithm.

We measure hiring behavior using several measures. Our primary measure is NEWSTAR: the number of stars hired in a given year by a particular firm. We include NEWSTAR both in levels and as a fraction of the firm's number of employees, in an attempt to control for the possibility that the effect of new hires is proportional to the size of the existing firm. In order to explore the degree to which star scientists hired from competitors have a different effect from those hired from other employers, in some specifications we decompose the overall number of hired stars into those coming from one of the other firms of our sample (NEWSTAR_COMP) and those coming from "other" organizations (NEWSTAR_OTH). Since newly hired stars might not have an immediate impact on organizational capabilities, in some specifications we include the 2-year cumulative sum of NEWSTAR.[12]

Some simple mean comparisons (not reported here) show that those firms that hire more star scientists than the median level (i.e., more than 2) have a level of PUBFRAC 7.5 percentage points higher

[11] Gittelman and Kogut (2003) elaborate a similar measure and interpret it as capturing the ability of a firm to translate research into invention and the extent to which the communities of researchers and inventors overlap.

[12] One could imagine using a longer time span to cumulate new hires, or a "stock" of stars as an independent variable. However these variables will inevitably be strongly correlated with other regressors, such as controls for a time trend and for size, and it would be more difficult to think of instrumental variables for such cumulative measures. Moreover, our immediate focus is on the impact of "new blood" on the organization.

than the their competitors in the next period. This difference is statistically significant.[13] This corresponds to roughly 7.5% more patent authors publishing than would have otherwise have published.

We follow Cockburn et al. (2000) in our choice of control variables. In that study the adoption of science-driven drug discovery was found to be significantly correlated with each firm's initial orientation to the practice. However the effect of initial conditions, expressed by the pre-sample value of PUBFRAC, fades over time: there is a significant negative coefficient on the interactive term (initial PUBFRAC * time), suggesting that laggards adopt more aggressively, so that there is convergence among firms in their rate of adoption. A major insight from this analysis, therefore, is that it is important to control for a firm's historical propensity towards science-oriented research of a firm when studying the contemporaneous determinants of the degree of adoption. We also include a time trend, [14] and control for the size of the firm by including the (log of) total assets, since economies of scope or scale may enable larger firms to capture a larger share of the "spillovers" generated by their rivals and by the public sector (Nelson (1959) and Arrow (1962)). We omit a number of other factors explored in the original study, since they appeared to have little significant effect on the rate of adoption. Our model can be therefore restated as follows:

$$Y_{it} = log\,[y_{it}/(1-y_{it})] = \alpha + \beta t + \theta(y_{i0}*t) + \zeta y_{i0} + \gamma h_{it} + \delta x_{it} + \varepsilon_{it} \qquad (4)$$

We allow for (unmodeled) correlation in the residuals at the firm level, by clustering them.

For the instrumental variable estimations, we explore the use of a number of different instruments. Our initial thought was to use measures of the size of the market for academic scientists, reasoning that it would influence both the demand and supply of labor, but that it would not influence any particular firm's orientation to science-based drug discovery. We were able to find data on the median salary of life science PhDs in academia and on the average salary of full Professors in Medical Schools, but unfortunately these variables had limited explanatory and identification power, also because they only varied (rather slowly) over time and not cross-sectionally. Moreover, we were unable to construct them for the non-US firms in our sample. We therefore use two alternative instruments: the short term financial position of the firm and a dummy variable for the "nationality" of the firm, as revealed by the location of its headquarters. We believe that the short term financial position of a firm is likely to affect hiring decisions but is unlikely to directly shape a firm's propensity to move to science based drug discovery. We proxy for short term financial position with each firm's net income lagged by one year.[15] We use dummy variables for the nationality of the firm to capture idiosyncratic characteristics specific to the institutional context of the firm's head office, which might affect the hiring behavior. Since it is well known that, in small samples, the finite sample bias of the estimates increases with the number of

[13] Significant differences are obtained also when we split the sample according to the mean level or the 75% percentile (5) of NEWSTAR.
[14] We have also tried to use time dummies instead of the time trend. All results described below are unchanged.
[15] All financial measures are expressed in constant dollars. These data, as well as data on employees, were obtained from Compustat files, the Global Access database and Kresge Fiches.

instruments (Hahn and Hausman (2002)), we use only net income as an instrument when NEWSTAR is included in absolute value, and use the country dummies when it is scaled by the number of employees, reasoning that financial performance may change both the number of scientists hired and the total number of employees hired. Newstar takes non-negative integer values, with a non-negligible amount of zeroes. A standard 2sls estimation would imply a linear regression as first stage, which may not be appropriate given the count nature of our suspected endogenous variable. For this reason, in some exercises we substitute the first stage linear regression of newstar on all the predetermined variables with an "equivalent" Poisson regression, and use the predicted value of newstar as an instrument in (4).

Finally, we try to control for the fact that firms may self-select into different hiring strategies using a two-step approach that models each firm's decision to hire explicitly. Shaver (1998) and Hamilton and Nickerson (2003) have explored the application of self-selection models to strategy research in the case of dichotomous independent variables. Since the hiring variable cannot easily be reduced to a dichotomous choice we follow Wooldridge (2001), and first consider hiring as a dichotomous choice (hiring vs. not hiring in a given period) and, in a probit regression, predict the probability of hiring at all as a function of the regressors in equation (4) and of our instruments. In a second step, we take only observations where NEWSTAR is not zero, and run a 2SLS regression on the basis of equation (4). We test the significance of NEWSTAR on this selected sample, and check for the presence of a selection bias by looking at the significance of the coefficient on the correction term: the inverse Mills ratio calculated from the probit first step.[16]

We also explore a more "direct" way to assess the presence of complementarities between hiring and other organizational features of a firm. We interact our hiring variables with the lagged value of PUBFRAC in some exercises, and with the initial value in others. The presence of complementarities should imply a positive coefficient on these variables.

Individual level analysis of scientific productivity

Recall our basic equation:

$$PUBS_{ijt} = f(y_{ijt}, h_{ijt} z_{ijt}; \varphi, \zeta_{ijt}) \quad (3)$$

To implement equation (3), we assume that, if a scientist publishes with one firm in year [t] and in year [$t+n$], then he or she was employed by the same firm for all n years, unless he or she publishes with another firm in the intervening period. Scientists who cease publication and who never publish again are assumed to have left the firm the year after their last publication.

We measure research productivity by publication counts. Since our dependent variable is discrete and values are non-negative and concentrated in relatively small numbers (between 0 and 10 in about

[16] In this case, we use both the net income and the country dummies as instruments. Using only one instrument for both the probit analysis and the IV regression would create multicollinearity problems. See Wooldridge (2001).

90% of the cases), we use Poisson regressions. We explore the robustness of our results to the Negative Binomial specifications to allow for over-dispersion, and to Random Effect (at the individual level) Poisson specifications (Hausman et al. (1984), Cameron and Trivedi (1998)). As before, we measure the degree of a firm's adoption of science-based drug discovery, as PUBFRAC. We operationalize the presence of other star scientists as STARTOT, the total number of stars employed by the firm in that year.

A central focus of our analysis is controlling for the "base rate" propensity of the scientist to publish. Recall that we wish to argue that evidence that PUBFRAC or STARTOT are significantly correlated with publication behavior is consistent with the hypothesis that there is an interactive, potentially complementary relationship between star scientists and the firms that hire them.

We include a number of variables that might explain "base rate" propensity to publish. We include TENURE, the number of years the scientist has been with the firm, and TENURE^2. In an attempt to control for individual effects we include three "cohort" variables (see Levin and Stephan (1991)): one (COHORT1) for those scientists who are publishing in 1973 (the earliest year for which publications are recorded), one (COHORT2) for those who first publish before the beginning of our period, i.e. from 1974 to 1980, and one for the youngest researchers ((COHORT3, the reference group).[17] We also include INPUB, the number of publications in the first year in which the scientist appears in the Science Citation Index, as an observable measure of the native ability of a scientist (see also Long and McGinnis (1981) for a similar choice), and the time elapsed from the first publication to the current period as a measure of overall experience, EXPER and its square EXPER^2 (Levin and Stephan, 1991). A time trend is added to account for the average increase of publication activities over time. Since there is a drop in the average number of publications in 1989, we also include a dummy variable equal to 1 in 1989 and zero otherwise (DUM89) in several specifications. Finally, to control for the "Matthew effect" we include the number of publications in the preceding year (Merton (1968)).[18] Firm dummies are included in all specifications. We therefore have firm- and scientist- specific, unobservable and observable effects in addition to time varying covariates.[19]

3. Results

Table (2) presents variable descriptions and summary statistics. Correlations at both the firm and individual scientist level are presented in Table (3). The firm-level sample has 162 observations for 17

[17] The age-vintage-time problem is discussed extensively in e.g. Berndt, Griliches, and Rappaport (1995).
[18] "For whosoever hath, to him shall be given, and he shall have more abundance." Matthew, 12:13.
[19] We are using two of the three methods to deal with dynamic panel data, as identified by Cameron and Trivedi (1998): (a) "ignore" the panel structure, and exploit the use of several observable characteristics; and (b) adopt a random effect specification, with appropriate assumptions about the conditional likelihood and the "starting" values of the dependent variable, with exploration of the use of the lagged value of publications (see also Wooldridge (2001)). Random-effect Poisson Models allow for heterogeneity in the variance and, in this respect, are similar to negative binomial specifications. FE regressions would not be consistent when not strictly exogenous regressors are added (such as the lagged publications). A third method would include

firms[20], except for those cases when EMPL is included, in which case only 154 observations are available.[21] The restricted sample for which NEWSTAR>0 contains 140 data points. In the case of the scientist level data set, we were able to find complete information for only a sub-sample of scientists: the final sample includes 367 researchers for a total of 2693 observations.[22]

Table (4) shows the core results for our firm level analysis. The estimated coefficient on NEWSTAR is positive and significant in all but one case, and it reduces the impact of initial conditions by up to about 10% of the value of the associated parameter. The implied increase in PUBFRAC from an increase in the yearly number of star hires by one standard deviation is estimated to be about .04, i.e. the ratio of inventors who also publish papers increases by 4. %points.[23] This is a non-negligible amount, and one that is consistent with organizational changes involving those scientists already employed by the firm.[24]

In models (5) and (6) we explore the degree to which star scientists hired from competitors have a significantly different effect from those hired from other sources. The estimated coefficients are significant, and their magnitude seems to imply that hiring from direct competitors has a slightly larger effect, The difference between in the estimates of the two parameters is statistically significant above the 90% level. Recall, however, that the number of scientists hired from competitors is relatively small.

Table (5) presents our efforts to control for possible endogeneity. Specifications (1) and (2) include firm fixed effects. Hausman tests cannot reject the null hypothesis of equality between the random effects and fixed effects results at any statistically meaningful level of significance, and, not surprisingly our core result is largely unchanged. This is consistent with our observable and time-invariant firm effect, the initial condition on PUBFRAC (which we can include in the random effect regressions but not, of course, in the fixed effects exercises), capturing a good deal of the time-invariant firm heterogeneity, or at least the part of this heterogeneity that is relevant for our study.

Models (3), (4) and (5) present our instrumented results. Our core findings seem quite robust to the use of IV techniques, although we lose some precision in some of the specifications, and are very much aware that the problem of potential endogeneity is a difficult one that is not easily solved. Model (6) presents the two-step results. NEWSTAR is still significant, and its estimated coefficient increases in

---

scientist fixed effects and perform a semi-differenced regression with generalized method-of-moments techniques. We plan to explore this further route in future analysis.

[20] Abbott, Beecham, Bristol-Myers, BMS, Burroughs Wellcome. Ciba-Geigy, Eli Lilly, Glaxo, Hoechst, Hoffman La Roche, Merck, Pfizer, Searle, SKB, Smithkline, Squibb, Upjohn,

[21] Restricting the sample to 154 observations, in the cases in which we have 162, does not significantly change the results.

[22] We were often unable to determine the date and number of papers of first publication.

[23] A linear model for the log-odd transformation implies that $pubfrac = \frac{e^{X\beta}}{1+e^{X\beta}}$. Therefore, the derivative of pubfrac with respect to newstar is given by: (1-pubfrac)*pubfrac*[estimated coefficient on newstar]. We set pubfrac at the mean, and multiplied this derivative by the standard deviation of newstar. The estimates are from table 4, model 2. The relatively small number of observations implies that these results should be treated with caution.

[24] Recall that PUBFRAC is calculated excluding the publication and patenting activity of the star scientists themselves.

magnitude. The estimated coefficient on the inverse Mills Ratio is positive and significant, showing some evidence of self-selection.

Model (7) of Table (4) reports an exercise with an interaction term between the hiring variable and the lagged value of PUBFRAC included in the regression. The high level of multicollinearity between NEWSTAR and the interaction between NEWSTAR and lagged PUBFRAC should make us cautious in the interpretation of the results, and we should expect some difficulty in obtaining precise estimates. Nonetheless when the interactive variable is added, its associated coefficient is positive and significant, while NEWSTAR loses statistical significance. When only the interactive variable is added, the associated coefficient is significant[25] as well. These results are consistent with the hypothesis that the effect of newly hired scientists on the firm is conditioned by the organizational structures that are already in place.

Given the limited size of our sample, these results must be treated with caution. The fact that there is some robustness in results obtained even in such a small sample is encouraging, but without a fully specified structural model (which would include, for example, an explicit model of the labor market), we cannot address the issue of endogeneity definitively. It is appropriate, therefore, to interpret the significant coefficients on our measures of external hiring as being consistent with the hypothesis that bringing new scientists into the firm plays an important role in assisting the firm to develop new organizational capabilities, rather than as a definitive structural test of the idea.

Table (6) presents our results at the individual scientist level. Our control variables are generally significant and show the expected signs, which is reassuring. In particular, we can see in some cases the presence of a job-tenure effect as well as an overall experience or life cycle, with the latter showing an inverted U-shaped form.[26] Moreover, the coefficient on the number of publications in the very first year of scientific activity has a positive and significant estimate, suggesting the appropriateness of controlling for unobserved ability.

Most interestingly, despite all these controls and the use of firm dummies, the coefficients of both PUBFRAC, our measure of organizational orientation, and STARTOT, the number of star scientists at the firm, are positive and significant. These results suggest a significant effect of the firm's broader organizational orientation and of the presence of other star scientists on the individual performance of

[25] Similar results obtain when NEWstar is interacted with the initial value of pubfrac, though the multicollinearity is even greater and so we have more noise. Also, results are similar if NEWSTAR is divided by EMPL.

[26] Since we use only publication data after a scientist joins one of the firms of our sample, we typically do not capture the first part of the experience cycle at the level of the single scientist, but we may capture it cross sectionally. Moreover, in order to make the sample homogeneous and the results comparable among different specifications, we excluded the first year in which a scientist publishes in a firm (this was necessary in the specification where the lagged publications are added.) This reduces our ability to capture a job tenure effect, because our time series is not very long and even one year (and in particular the first one) of data may make a difference. The impact of this choice is less important for experience, which has a wider (and more distributed) range of values. Results not reported here show that an inverted U relationship with tenure, is indeed much stronger and more robust when the first year of publication is added to the data, and that it "co-exists" with the experience cycle effect. The results

those stars already employed by the firm. We interpret them as consistent with the hypothesis that the performance of star scientists is conditioned by the nature of the firm that they join, suggesting that there may well be important complementarities between the hiring of "new blood" and the existing organizational competencies of the firm.

4. Conclusions

In this paper, we offered theoretical insights and empirical findings designed to clarify the degree to which hiring skilled workers can provide an organizational and strategic resource for the firm. We stressed the danger of interpreting simple correlations between hiring and improvement in firm performance as evidence that hiring builds organizational capabilities, and speculated that for hiring to build capability, new hires must be in some way complementary to the existing assets, routines or procedures of the firm.

In statistical terms, we dealt with these issues by moving from the analysis of simple correlations to addressing the problems of unobserved heterogeneity, causality ,endogeneity, and self-selection. We also explored whether the performance of star workers is shaped by the organizational capabilities of the firm — particularly the presence of other stars — in order to understand how the individual capabilities of newly hired skilled workers are translated into durable organizational capabilities.

Our results suggest that in the pharmaceutical industry of the eighties and early nineties there was a significant relation between the adoption of science-driven research and the hiring of new scientists, and that external hiring had a significant effect on internal organizational capabilities. Our findings also strongly imply that the evolution toward science-driven R&D was not solely driven by the output and work practices of "new blood." Our study of the determinants of individual publication rates is consistent with the idea that there is a complex and intriguing interaction between individual and organizational capabilities – just the kind of interaction that would be required to translate the ability to hire in a freely available market into the building of a unique organizational capability.

We see a number of potential extensions of our work. On the empirical and statistical front, a larger sample size and more powerful controls and instrumental variables could increase the robustness of our results and allow us to build further, more elaborate tests. For example we plan to explore the patterns of coauthoring and/or co-patenting between stars and non-stars; the time lag between the start of the patenting activity and the publication activity of each scientist inside a firm; and the participation of the same star in the work of different teams.

---

on all the other variables are almost unchanged when larger samples are used. This joint finding is not broadly present in the existing literature, as far as we are aware.

Comparing the dynamics of the hiring policy of star scientists with the overall hiring strategies of firms may also be a source of insight. It would be interesting to know, for example, if less highly published scientists switch firms with a different frequency. Additional information on the previous and subsequent occupations of the star scientists might enable us to determine the degree to which scientists with different backgrounds have different impacts on the firm they join, and more closely tracking the movements of stars in and out of our sample firms would allow us to explore whether the departure of top scientists weakens the scientific capabilities of an organization. The impact of stars may also vary within a given firm, in different research projects.

In addition, it might be important to study the effect of a firm's position in the network of skilled labor market on its ability to change.

Most fundamentally, we believe that this paper highlights both the importance of new hires to the development of new organizational capability and the complexity of the process through which this occurs. The hiring of particularly productive scientists was almost certainly critically important to the development of new capabilities in pharmaceutical research – but these scientists were not a freely transferable resource that could be safely left in a laboratory to produce new drugs. They became, instead, an integral part of their new employer's research effort, with impacts on research productivity extending beyond their own immediate areas of activity. Theoretical research in strategy has increasingly pointed towards the importance of complementarities between people, procedures and routines as a long term source of competitive advantage. This paper gives renewed impetus to this belief, and highlights the importance of viewing particular individuals as one element of the complex, interlocking system that is the successful firm.

References

Agrawal, A. and R. Henderson (2002) "Putting Patents in Context: Exploring Knowledge Transfer From MIT." *Management Science*, 48(1):44-60.

Almeida, P and B. Kogut (1999) "Localization of Knowledge and the Mobility of Engineers in Regional Networks." *Management Science* 45:905-917

Arndt, M. and A. Weintraub (2003) "Feeding the Pipeline. How the Drug Giants Are Reinventing their Labs to Boost Productivity." *Business Week*, May 2003.

Arrow, K. (1962) "Economic Welfare and the Allocation Of Resources For Invention." In R. Nelson (ed.) *The Rate and Direction of Inventive Activity*. National Bureau of Economic Research, Princeton University Press, Princeton (NJ). Pages 609-619.

Athey, S. and S. Stern (1998) "An Empirical Framework for Testing Theories About Complementarity in Organizational Design." NBER Working Paper No. 6600.

Barnett, W. P. and G. R. Carroll (1995) "Modeling Internal Organizational Change." *Annual Review of Sociology*, 21:217-236.

Barney, J. B. (1986) "Strategic Factor Markets," *Management Science* 32:1231-1241.

Bertrand, M. and A. Schoar (2002) "Managing with Style: The Effect of Managers on Firm Policy." MIT Sloan School of Management Working Paper No. 4280-02.

Breschi, S, and F. Lissoni, (2001) "Knowledge spillovers and local innovation systems: a critical survey." LIUC Papers in Economics 84, Cattaneo University (LIUC).

Cameron, A.C. and P. K. Trivedi (1998) *Regression Analysis of Count Data*, Cambridge University Press.

Cockburn, I.M. and R. Henderson (1998) "Absorptive capacity, coauthoring behavior, and the organization of research in drug discovery." *Journal of Industrial Economics*, 46(2):157-182.

Cockburn, I.M., R. Henderson, and S. Stern (1999a) "Balancing Incentives: The Tension Between Basic and Applied Research." NBER Working Paper No. 6882.

Cockburn, I.M., R. Henderson, and S. Stern (1999b) "The Diffusion of Science-Driven Drug Discovery: Organizational Change in Pharmaceutical Research." NBER Working Paper No. 7359.

Cockburn, I.M., R. Henderson, and S. Stern, (2000) "Untangling the Origins of Competitive Advantage", *Strategic Management Journal*, 21(10-11):1123-1145.

Darby, M., Q. Liu, and L. Zucker (1999) "Stakes and Stars: The Effect of Intellectual Human Capital on the Level and Variability of High-Tech Firms' Market Values." NBER Working Paper No. 7201.

Dasgupta, P. and P. David (1994) "Towards a New Economics of Science." *Research Policy*, 23:487-521.

Galambos, L. and J. Eliot Sewell (1996) *Networks of Innovation: Vaccine Development at Merck, Sharp and Dohme, and Mulford, 1895-1995*. Cambridge University Press,.

Gambardella, A. (1995) *Science and Innovation - The US Pharmaceutical Industry During the 1980s*, Cambridge University Press.

Gittelman, M. and B. Kogut (2003) "Does Good Science Lead to Valuable Knowledge? Biotechnology Firms and the Evolutionary Logic of Citation Patterns," *Management Science*, 49(4):366-382.

Grant, R. (1996) "Prospering in Dynamically-Competitive Environments: Organizational Capability as Knowledge Integration" *Organization Science*, 7(4):375-387.

Griliches, Zvi, (1957) "Hybrid Corn: An Exploration in the Economics of technical Change." *Econometrica*, 25(4):501-522.

Groysberg, B. (2001) "Can They Take it with Them? The Portability of Star Knowledge Workers' Performance: Myth or Reality*?*" Mimeo, Harvard Business School.

Hahn, J. and Hausman, J..(2002) "Notes on Bias in Estimators for Simultaneous Equation Models", *Economics Letters,* 75, 237-241.

Hamilton, B. and J. Nickerson (2003) "Correcting for Endogeneity in Strategic Management Research." *Strategic Organization*, 1(1):51-78.

Hamilton, B. J. Nickerson, and H. Owan (2003) "Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation," *Journal of Political Economy*, 111(3):465-497.

Hausman, J., B.H. Hall and Z. Griliches (1984) "Econometric Models of Count Data with an Application of the Patents R&D Relationship." *Econometrica* , 52(4):909-938.

Henderson, R. (1994) "The Evolution of Integrative Competence: Innovation in Cardiovascular Drug Discovery." *Industrial and Corporate Change*, 3(3):607-630.

Henderson, R., L. Orsenigo, and G. Pisano (1999) "The Pharmaceutical Industry and the Revolution in Molecular Biology: Exploring the Interaction between Scientific, Institutional and Organizational Change", in Mowery, D. and Nelson, R. (eds.), *The Sources of Industrial Leadership: Studies of Seven Industries*, Cambridge University Press.

Huckman, R. and G. Pisano, (2003) "The effect of Organizational Context on Individual Performance." NBER Working Paper No. 1027.

Ichniowscki, C., K. Shaw, and G. Prennushi (1997) "The Effects of Human Resource Management on Productivity." *American Economic Review*, 87(3):291-313.

Levin, S. and P. Stephan (1991) "Research Productivity Over the Life Cycle: Evidence for Academic Scientists", *American Economic Review*, 81(1):114-132.

Levinthal, D. (1997) "Adaptation on Rugged Landscapes." *Management Science* 43:934-950.

Long, J.S., and R. McGinnis, (1981) "Organizational Context and Scientific Productivity." *American Sociological Review*, 46(4):422-442.

McKelvey, B. (1982) *Organizational Systematics*, University of California Press.

Merton, R. (1968) “The Matthew Effect in Science: The Reward and Communication Systems of Science Are Considered.” *Science*, 159(3810):56-63.

Milgrom, P. and J. Roberts (1990) “The Economics of Modern Manufacturing: Technology, Strategy and Organization.” *American Economic Review*, 80(3):511-528.

Murmann, P. and M. Tushman, (1997) “Impact on Executive Team characteristics and organization context on organization responsiveness to environmental shock.” In R. Garud, P. Nayyar, and Z. Shapira (eds.), *Technological Innovation: Oversights and Foresights*, Cambridge University Press.

Nelson, R. (1959) “The Simple Economics Of Basic Scientific Research.” *Journal of Political Economy,* 67(2):297-306.

Penrose, E. (1959) *The Theory of the Growth of the Firm,* John Wiley & Sons.

Peteraf, M. (1993) “The Cornerstones of Competitive Advantage: A Resource-Based View”, *Strategic Management Journal*, 14(3), 179-191.

Rao, H. and R. Drazin (2002) “Overcoming Resource Constraints on Product Innovation by recruiting Talent from Rivals: A Study of the Mutual Fund Industry, 1986-1994.” *Academy of Management Journal*, 45(3):491-508.

Rivkin, J. (1998) “Optimally Local Search on Rugged Landscapes.” HBS Working paper 98-067.

Rivkin, J. (2000) “Imitation of Complex Strategies” Management Science, 46(6):824-844.

Shaver, J.M., (1998) “Accounting for Endogeneity When Assessing Strategy Performance: Does Entry Mode Choice Affect FDI Survival?” *Management Science*, 44(4):571-585.

Song, J., P. Almeida, P. and G. Wu, (2003) “Learning-by-Hiring: When Is Mobility More Likely to Facilitate Interfirm Knowledge Transfer?” *Management Science*, 49(4):351-365

Whaldholz, M., (1985) “Merck Picks Vagelos, Its Top Scientist, To Be Chief Executive and President.” *Wall Street Journal*, April 24, 1985.

Wooldridge, J. (2001) *Econometric Analysis of Cross Section and Panel Data*, MIT Press.

Zucker, L. and M. Darby (1997) “Present at the Revolution: Transformation of Technical Identity for a Large Incumbent Pharmaceutical Firm after the Biotechnology Breakthrough”, *Research Policy*, 26:429-426.

Zucker, L. M. Darby, and M. Brewer (1998) “Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises.” *American Economic Review*, 88(1):290-306.

Zucker, L. M. Darby, and M. Torero (2002) “Labor Mobility from Academe to Commerce.” *Journal of Labor Economics*, 20(3):629-660.

Figures

Figure (1): Total number of star scientists employed in firms, 1980-1992

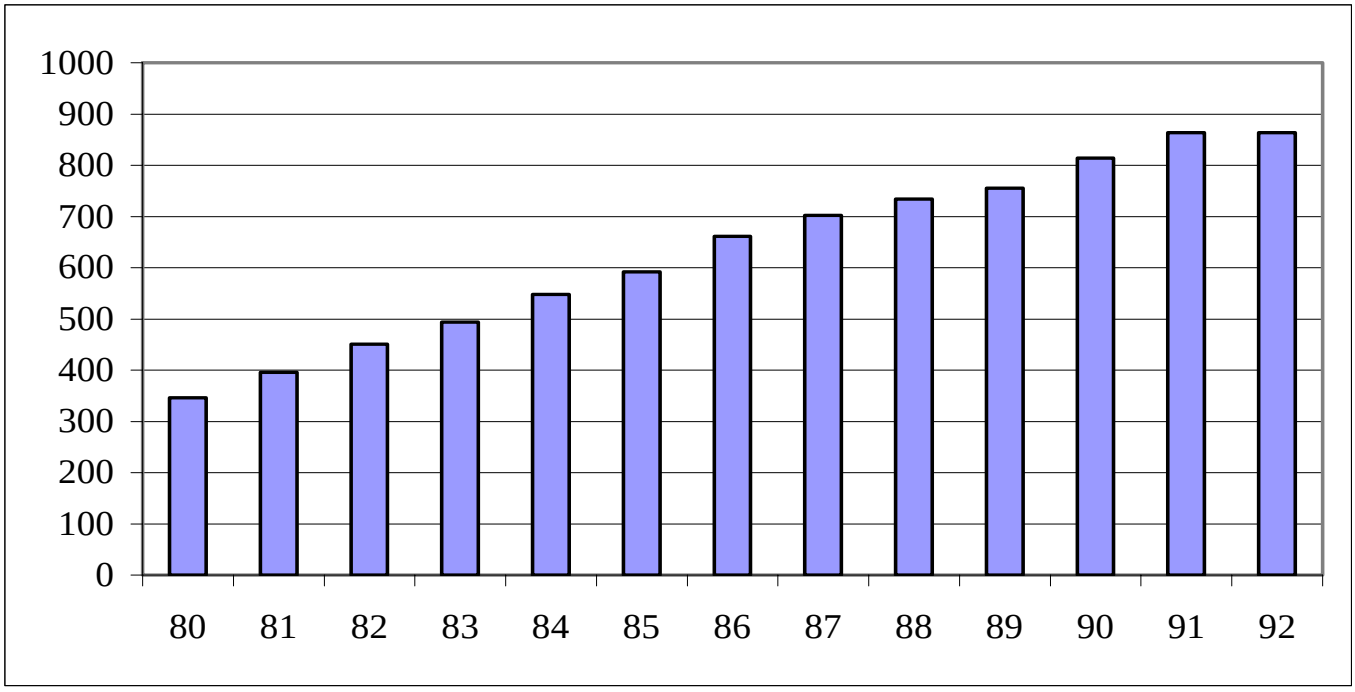
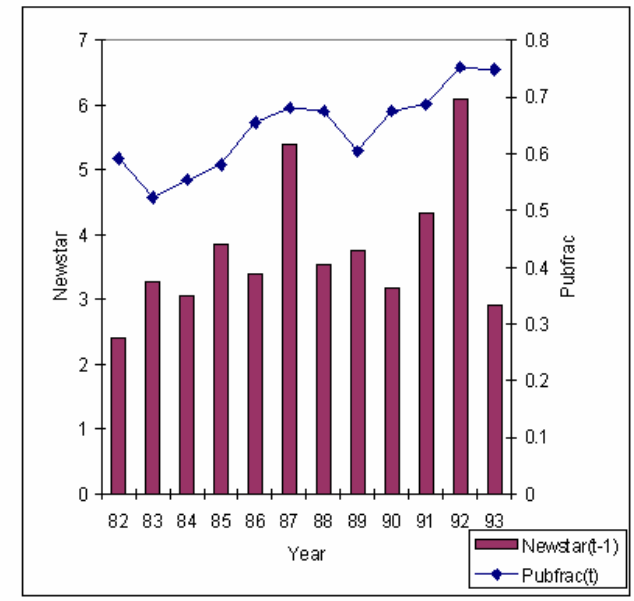


Figure (2): Mean Newstar over time vs Pubfrac



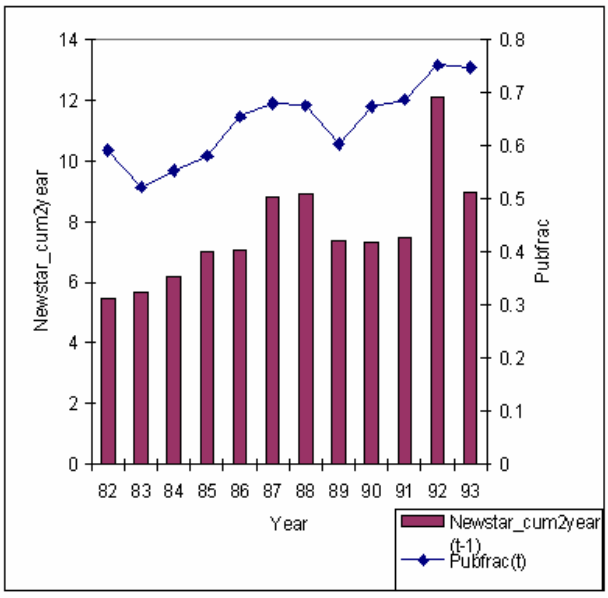


Figure (3) Mean papers per year, and papers per year vs pubfrac

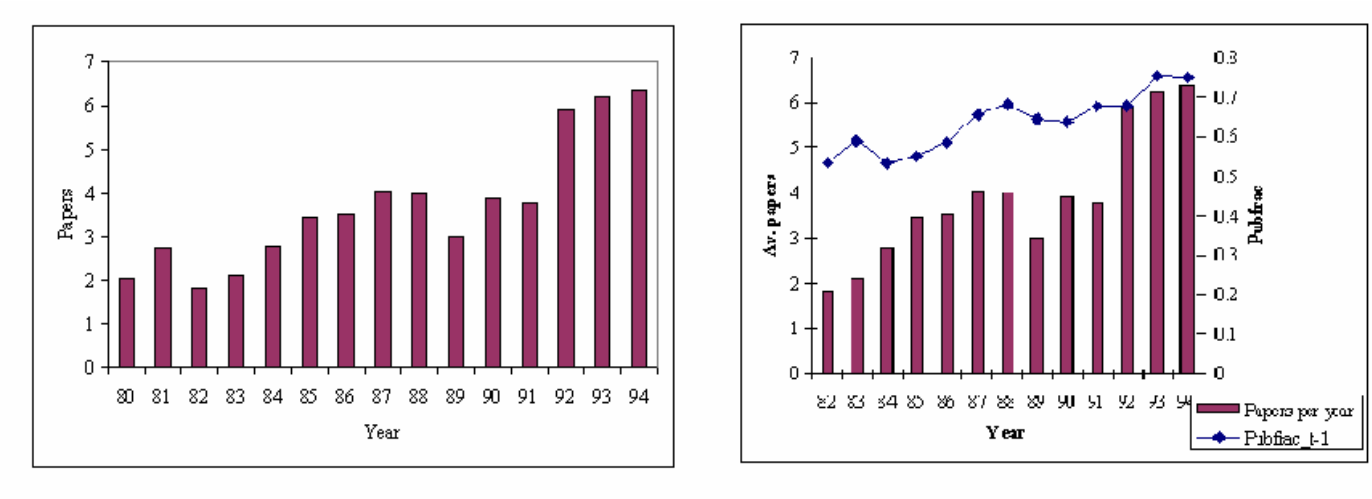
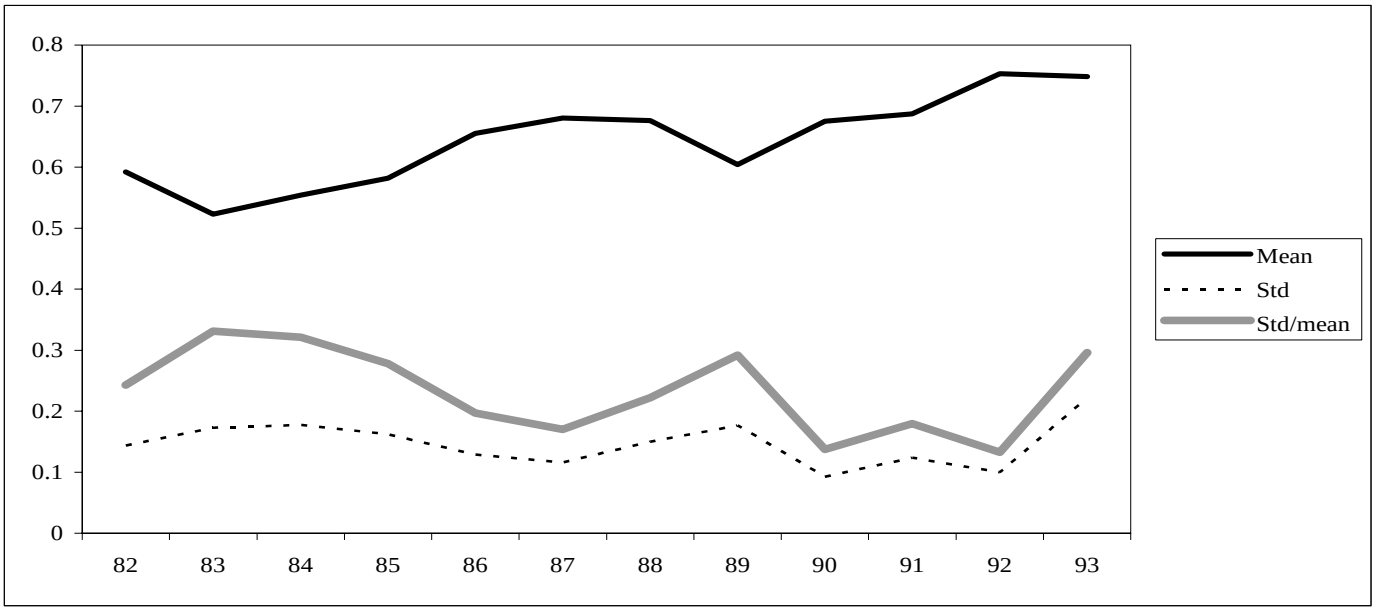


Figure (4): PUBFRAC over time



Tables

Table (1): Movements of scientists among the firms in the sample

	from																			
to	**1**	**2**	**3**	**4**	**5**	**7**	**8**	**9**	**10**	**12**	**13**	**15**	**16**	**26**	**33**	**35**	**36**	**40**	**41**	**Total**
1				1	1		2					1		1	1					**7**
2			1																1	**2**
3	1	1			2	1		1	1	3	1						1			**12**
4			1		1		1								3					**6**
5		1													1					**2**
7							1						1	1						**3**
8			2												2					**4**
11	1	1	2		1					2	1							1		**9**
12	2	1	1			1					1	1		1	1		1			**10**
13	2	5	3	1				1	1	1		1		3	1		1			**20**
15	1			1						1				1						**4**
16												2								**2**
26			1					1			1									**3**
33			3		1									1						**5**
35																		2		**2**
36				1											1					**2**
41																1				**1**
42																			2	**2**
Total	**7**	**9**	**14**	**4**	**6**	**2**	**4**	**3**	**2**	**7**	**4**	**5**	**1**	**8**	**10**	**1**	**3**	**3**	**3**	**96**

1=Abbott; 2= Burroughs Wellcome; 3= Merck; 4= Searle; 5= Hoffman La Roche; 7= Bristol-Myers; 8= Squibb; 9= Smithkline; 10= Beecham; 11= BMS; 12= SKB; 13= Glaxo; 15= Lilly; 16= Pfizer; 26= Upjohn; 33= Ciba-Geigy; 35= Sankyo; 36= Hoechst; 40= Fuijsawa; 41= Takeda; 42= Yamanouchi

Table (2): Variable descriptions

Variable name	Description	OBS	Mean	Stdev	Min	Max
Firm level						
PUBFRAC	see text	162	0.64	0.16	0.12	0.95
$PUBFRAC_0$	Initial value of PUBFRAC	162	0.52	0.16	0.20	0.77
TIME	Year (yy)-80	162	7.17	3.38	2	13
NEWSTAR_COMP	Scientists hired from competitors at t-1	162	0.40	0.81	0	4
NEWSTAR_OTH	Scientists hired from other at t-1	162	3.34	3.77	0	20
NEWSTAR	Total # of Scientists hired at t-1	162	3.74	4.00	0	20
NEWSTAR_2cum	Total # of Scientists hired at t-1 and t-2	162	7.59	7.53	0	36
LOGASS	log(total assets)	162	8.21	0.78	6.25	9.73
NEWSTAR/EMPL	10^5*Newstar(t-1)/employee at t-1	154	11.60	12.63	0	57.47
NetIncome	Net Income at t-2, 000 real $	162	430.63	308.64	-108.58	1557.78
Individual scientist level						
PUB	# of publications for each scientist in t	2639	4.41	3.90	0	32
COHORT1	1 if the year in which a scientist publishes his/her very first paper is <1974, 0 otherwise	2639	0.23	0.42	0	1
COHORT2	1 if the year in which a scientist publishes his/her very first paper is between 1974 and 1981	2639	0.51	0.50	0	1
TENURE	(Current year(yy)+1) – hiring date(yy)	2639	6.56	3.92	1	22
EXPER	(Current year(yy)+1) – year in which a scientist publishes his/her very first paper(yy)	2639	12.91	4.61	1	23
DUM89	1 IF year=1989, 0 otherwise	2639	0.10	0.30	0	1
INPUB	# of publications in very first year in SCI	2639	1.99	1.66	1	15
STARTOT (log)(t-1)	Total number of stars employed in the firm	2639	4.17	0.82	1.10	5.24

Table (3)

Correlations: Firm/year level data

		1	2	3	4	5	6	7	8	9
1	**PUBFRAC**	1.00								
2	**$PUBFRAC_0$**	0.37	1.00							
3	**TIME**	0.36	-0.10	1.00						
4	**LOGASS**	0.13	-0.02	0.45	1.00					
5	**NEWSTAR_COMP**	0.28	-0.13	0.44	0.18	1.00				
6	**NEWSTAR_OTH**	0.30	0.34	0.02	0.10	0.18	1.00			
7	**NEWSTAR**	0.34	0.30	0.10	0.14	0.38	0.98	1.00		
8	**NEWSTAR_2cum**	0.36	0.32	0.17	0.15	0.35	0.89	0.91	1.00	
9	**NetIncome**	0.34	0.16	0.60	0.62	0.52	0.29	0.38	0.45	1.00

N=162

Correlations: scientist/year level data

		1	2	3	4	5	6	7	8	9	10
1	**PUB**	1.00									
2	**PUBFRAC**	0.17	1.00								
3	**STARTOT (ln)**	0.16	0.54	1.00							
4	**Tenure**	0.002	0.14	0.08	1.00						
5	**EXPER**	0.10	0.16	0.10	0.54	1.00					
6	**TIME**	0.26	0.40	0.32	0.37	0.43	1.00				
7	**dum89**	-0.11	0.01	0.007	-0.02	-0.02	-0.04	1.00			
8	**INPUB**	0.08	0.04	0.03	0.00	0.08	-0.01	0.00	1.00		
9	**COHORT1**	0.01	-0.04	-0.11	0.24	0.47	-0.14	-0.01	0.25	1.00	
10	**COHORT2**	-0.05	-0.07	-0.002	0.01	0.16	-0.07	0.02	-0.17	-0.56	1.00

N=2693

Table (4): Firm level results.
Dependent Variable: log[PUBFRAC/(1-PUBFRAC)]. Basic specifications

	1	2	3	4	5	6	7
$PUBFRAC_0$	3.47	3.16	3.16	3.30	3.20	3.32	2.83
	.50	*.52*	*.53*	*.61*	*.48*	*.54*	*.58*
$PUBFRAC_0$*t	-.22	-.22	-.22	-.22	-.21	-.20	-.21
	.08	*.07*	*.07*	*.09*	*.06*	*.07*	*.07*
NEWSTAR		.042					-.09
		.013					*.06*
NEWSTAR_2cum			.022				
			.007				
NEWSTAR/EMPL				.013			
				.004			
NEWSTAR_COMP					.16		
					.07		
NEWSTAR_OTH					.03		
					.012		
NEWSTAR*PUBFRAC(t-1)							.18
							.08
NEWSTAR_COMP/EMPL						.07	
						.02	
NEWSTAR_OTH/EMPL						.01	
						.004	
LOG(ASSETS)	-.023	-.04	-.04	.04	-.04	.05	-.03
	.09	*.013*	*.09*	*.10*	*.09*	*.10*	*.08*
TIME	.21	.21	.20	.20	.18	.17	.19
	.04	*.04*	*.04*	*.05*	*.03*	*.04*	*.04*
CONSTANT	-1.68	-1.46	-1.47	-2.25	-1.46	-2.27	-1.27
	.75	*.77*	*.77*	*.84*	*.75*	*.82*	*.67*
Method	OLS	OLS	OLS	OLS	OLS	OLS	OLS
F	21.68	31.31	31.41	25.54	29.10	25.13	27.01
R2	.31	.36	.36	.37	.37	.39	.38
Obs	162	162	162	154[+]	162	154[+]	162

Standard errors are in italics. Residuals clustered at the firm level.
F-tests for H_o: all coefficient = 0 reject H_o at .00% level in all models.
[+] *Reduced number of observations because data missing for EMPL in some cases*

Table (5): Firm level results.
Dependent Variable: log[PUBFRAC/(1-PUBFRAC)]. Further statistical tests

	1	2	3	4	5	6
$PUBFRAC_0$			2.84	2.76	2.95	3.09
			.82	*.78*	*.87*	*1.17*
$PUBFRAC_0$*t	-.20	-.16	-.22	-.22	-.22	-.29
	.09	*.11*	*.07*	*.07*	*.08*	*.10*
NEWSTAR	.043		.085	.096		.14
	.02		*.06*	*.036*		*.06*
NEWSTAR/EMPL		.01			.025	
		.06			*.017*	
LOG(ASSETS)	.06	.16	-.07	-.07	.07	-.08
	.26	*.28*	*.12*	*.11*	*.11*	*.10*
TIME	.19	.16	.20	.20	.20	.20
	.05	*.06*	*.03*	*.03*	*.05*	*.04*
CONSTANT	-.64	-1.36	-1.23	-1.17	-2.41	-1.51
	2.00	*2.17*	*1.11*	*1.03*	*.97*	*1.06*
Method	PANEL-FE (a)	PANEL-FE (b)	2SLS (c)	2SLS (d)	2SLS (e)	Selection model. Probit in first step, 2SLS in second step (f)
F	12.72	9.46	16.55	20.23	14.77	13.54
Obs	162	154[+]	162	162	154[+]	140

Standard errors in italics. Residuals are clustered at firm level in models 3, 4, 5, 6. F-tests for H_o: all coefficient = 0 reject H_o at .00% level in all models.

[+] *Reduced number of observations due to the unavailability of some data on the number of employees.*

(a) Hausman Test for RE=FE: chi2=.60. F-test for all firm effects=0: F=2.72, prob>F = .0009

(b)Hausman Test for RE=FE: chi2= .63 F-test for all firm effects=0: F=2.67, prob>F = .0011

(c) NetIncome (t-2) as instrument for newstar. 1st stage regression: PUBFRAC(0) and netIncome(t-2) have positive and significant estimated coefficients (resp. at 6% and .0% level); R^2= .22 , Fstat for all coeff.=0: F=9.15, prob >F = .0000.

*(d) Instrument for Newstar is the predicted number of hires from a Poisson regression of newstar on the regressors in (4) and NetIncome(t-2). In the Poisson Regression, PUBFRAC(0) and netIncome(t-2) are positive and significant at .00% level, PUBFRAC(0)*t negative and significant at 1% level, LRchi(5)=150.46. In a correspondent neg. bin. regression: netIncome(t-2) significant at .00% level, PUBFRAC(0) significant at 2% level, LRchi(5)=40.80.*

Results do not significantly change if the predicted value of the Poisson regression is entered as a regressor (in place of newstar) in the second stage rather than used as instrument for newstar

(e) Country dummies as instr. for newstar/empl, US as reference. 1st stage regr.: PUBFRAC(0) has positive and 8% significant estimate. Germ. and Swiss dummies have significant (negative) estim. coefficients (resp at .0% and 1% level); R^2= .24, Fstat for all coeff.=0: F =6.56, prob>F = .000

(f) Second step regr. on selected sample for which NEWSTAR>0. NetIncome(t-2) and country dummies as IV and as added regressors in probit.

Inverse Mills Ratio defined from predicted values of first probit step is added to the 2sls regression, T-stat=2.05.

1st step Probit regression (dep var=1 if newstar>0, 0 otherwise): Germ Dummy has negative and 10% sig. estimate. NetIncome(t-2) has positive and 10% sig. estimated coefficient. Chi2(8) for all coeff.=0: 18.06, prob>chi2=.02; pseudoR^2=.12, log likelihood= -56.

1st stage of 2sls regression: R^2=.28, F test for all coeff. = 0: F = 5.98, prob>F = .000.

Table (6): Research productivity at the level of the individual scientist.
Dependent Variable: PUBS

	1	2	3	4	5	6
PUBFRAC(t-1)	1.47 *.11*	1.48 *.11*	1.47 *.19*	.47 *.13*	.45 *.21*	.26 *.13*
TENURE		.03 *.008*	.03 *.015*	-.008 *.01*	-.01 *.01*	-.03 *.01*
TENURE^2		-.002 *.0004*	-.002 *.0008*	-.0009 *.0005*	-.0005 *.0008*	.0004 *.0005*
LOG(startot)(t-1)				.57 *.07*	.67 *.11*	.58 *.07*
EXPER				.027 *.016*	.025 *.019*	.03 *.01*
EXPER^2				-.002 *.0005*	-.002 *.0007*	-.001 *.0004*
TIME				.06 *.012*	.06 *.01*	.04 *.01*
DUM89				-.37 *.04*	-.40 *.06*	-.40 *.04*
COHORT1				.39 *.13*	.43 *.11*	.26 *.10*
COHORT2				.16 *.09*	.20 *.07*	.10 *.06*
INPUB				.024 *.012*	.026 *.01*	.017 *.009*
PUB(t-1)						.05 *.002*
Firm Dummies	Yes	Yes	Yes	Yes	Yes	Yes
CONSTANT	.45 *.08*	.37 *.08*	.36 *.15*	-1.88 *.28*	-2.20 *.39*	-1.67 *.27*
Method	Poisson	Poisson	Neg. Bin. (Alpha=.5)	Poisson - RE	Neg. Bin. (Alpha=.45)	Poisson – RE
Chi	504.19	536.13	161.54	564.58	379.24	1047.05
Pseudo R2^	.031	.033	.012		.028	
Obs	2639	2639	2639	2639	2639	2693
LogLikelihood	-7888	-7872	-6651	-7014	-6542	-6836

Standard errors in italics.