NBER WORKING PAPER SERIES

SCALE, SCOPE AND SPILLOVERS: THE DETERMINANTS OF RESEARCH PRODUCTIVITY IN THE PHARMACEUTICAL INDUSTRY

Rebecca Henderson
Iain Cockburn

Working Paper No. 4466

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 1993

This research was funded by four pharmaceutical companies and by the Sloan Foundation. Their support is gratefully acknowledged. We would also like to express our appreciation to all of those firms that contributed data to the study, to Bronwyn Hall, Richard Caves, and Zvi Griliches, and to seminar participants at the NBER, UBC, MIT, Stanford, Northwestern, Columbia, and Carnegie Mellon, for their helpful comments and suggestions. The conclusions and opinions expressed in this paper are our own, not those of the National Bureau of Economic Research, and any errors or omissions remain entirely our responsibility. This paper is part of NBER's research program in Productivity. .

NBER Working Paper #4466
September 1993

SCALE, SCOPE AND SPILLOVERS: THE DETERMINANTS OF RESEARCH PRODUCTIVITY IN THE PHARMACEUTICAL INDUSTRY

ABSTRACT

This paper presents the results of a study of the determinants of research productivity in the pharmaceutical industry. Using disaggregated, internal firm data at the research program level from ten major pharmaceutical companies, we find no evidence of increasing returns to scale at either the firm or the research program level. However our results suggest that there are three benefits to running research programs within the context of larger and more diversified R&D efforts: economies of scale arising from sharing fixed costs; economies of scope arising from the opportunity to exploit knowledge across program boundaries within the firm; and the enhanced ability to absorb internal and external spillovers. We also find that spillovers between firms may play a major role in increasing research productivity. The paper also speaks directly to the question of firm heterogeneity. A significant proportion of the "firm effect" identified in previous studies can be explained by the slowly changing composition of the research portfolio, as well as by less easily measured aspects of innovative capability.

Rebecca Henderson
MIT Sloan School of Management
E52-543
Cambridge, MA 02139
and NBER

Iain Cockburn
Faculty of Commerce
2053 Main Mall
University of British Columbia
Vancouver, BC
Canada V6T 1Z2
and NBER

Introduction

While there is considerable agreement that economies of scale and scope in performing R&D and the magnitude of R&D spillovers may have important implications for research productivity, research to date has been far from conclusive. The theoretical literature is divided, for example, as to the effects that relative firm size or the presence of spillovers should have on incentives to undertake research (Dasgupta and Stiglitz (1980); Spence, (1984)), while empirical work has generated surprisingly inconclusive and sometimes contradictory results. (Cohen and Levin, (1989).) Several observers have suggested that this is the result of a historical reliance upon aggregated data, combined with a failure to control adequately for firm or industry effects (Baldwin and Scott, (1987); Cohen and Levin, (1989)). In this paper we look inside the firm for evidence on the importance of scale, scope and spillovers for research productivity, using detailed, disaggregated, data at the level of the research program obtained from the internal records of ten major research-oriented pharmaceutical companies.

The enriched understanding of the production technology of innovation which this research provides may help to resolve several debates in the economics of industrial organization. There are also implications for the theory of the firm. For example several scholars, notably Chandler (1990), have suggested that the presence of very large firms in modern capitalist economies is driven by the opportunity to exploit internal economies of scale and scope. Moreover since the size of research programs, firm size and spillover regimes may be dramatically affected by regulatory policy, a better understanding of their effects upon research productivity can help to lay the ground work for informed public policy.

Data from the pharmaceutical industry are particularly well suited to studying these issues. The industry is extremely research intensive, and

successful research drives firm performance. As a result nearly every firm in the industry conducts multiple research programs that compete directly with each other, offering us a rich data set. Research effort is relatively homogeneous, so that a comparison of research programs across firms is not confounded by unobservable variation in technology. Moreover the nature of the technology is such that it is possible to construct reasonable measures of technological "distance" in order to evaluate the effects of spillovers. A study of research productivity in the pharmaceutical industry is also interesting in its own right, since the industry is one in which a better understanding of the importance of scope, scale and spillovers is likely to have immediate implications for the formulation of government policy and the conduct of individual firms.

We use a panel data set with almost five thousand observations, covering more than twenty years to explore the drivers of research productivity in the "discovery" phase of ethical drug research. Our results suggest that there are no returns to scale *per se* at either the level of the firm or the level of the research program. However we find that individual programs are more productive, all other things equal, in larger and in more diversified firms, and that research productivity is strongly correlated with the productivity of programs in related therapeutic areas within the firm, suggesting that there may be significant economies of scope at the level of the therapeutic class. Our results also suggest that research productivity in a given therapeutic class is positively associated with the patents applied for by competing firms. We interpret these results as evidence consistent with the presence of extensive spillovers within the industry, but we also note that they are consistent with correlated research efforts and outcomes due to industry-wide shocks to technological opportunity.

The paper begins with a short literature review. We then briefly

describe the nature of research in the pharmaceutical industry as background to the development of some hypotheses. Subsequent sections discuss some of the estimation issues and describe the data on which the study is based. A discussion of the empirical results follows, and the paper closes with a discussion of their implications and an outline of future research.

1. Hypothesis Development and Literature Review.

Scale, scope, spillovers and research productivity.

Schumpeter (1934,1950) was the first to argue systematically that there was an enduring relationship between firm size and research productivity. As several observers have noted, his argument can be decomposed into several elements. In the first place, he suggested that there might be a significant relationship between research productivity and monopoly power. *Ex ante* monopoly power might serve to mitigate problems in the financial markets, while *ex post* monopoly power would mitigate problems of appropriability. In the second place, Schumpeter suggested that there may be significant economies of scale in the research production process itself, and that larger firms may enjoy significant economies of scope. Thus he implied that not only will larger research programs be more productive than smaller ones, but also that a research program of any given size will be more productive if it is conducted within a larger firm.

Later work has fleshed out this idea to suggest that there may be three major advantages in performing R&D conferred by size (Fisher & Temin (1973), Cohen and Levin, (1989)). In the absence of fully functioning markets for innovation, larger firms may be able to spread the fixed costs of research over a larger sales base. Larger firms may also be able to exploit complementarities between other assets of the firm -- marketing or manufacturing expertise for example -- to increase the productivity of research.

They may also have advantages in the financial markets over smaller firms: to the degree that they are able to mitigate problems of adverse selection and moral hazard in raising capital, they may be better positioned to fund risky projects. Much qualitative, historical evidence is consistent with these ideas, although a paucity of data at the program, as opposed to the firm level, has made it difficult to distinguish between economies of scale and scope (Chandler, (1990); Mowery (1989)).

Schumpeter's core conclusions have been challenged on two fronts. On the one hand, organizational researchers have suggested that increasing organizational size imposes significant diseconomies of scale in the form of increasingly bureaucratic and hierarchical organizations that reduce technological creativity. Some industry level and case based research is consistent with this hypothesis (Burns and Stalker, (1966); Tushman and Anderson, (1986)).

On the other hand, systematic cross-sectional studies of Schumpeter's ideas have had inconclusive results. Much early research explored the relationship between firm size and research intensity, seeking to explore the degree to which research intensity increased with firm size (Baldwin and Scott (1987), and Cohen and Levin (1989), provide excellent summaries of research in this tradition.). The results of this work are largely contradictory, and, as Fisher and Temin (1973) have pointed out, cannot be taken as tests of Schumpeter's ideas. Unfortunately the results of later work exploring the more appropriate relationship between research output and firm size have been similarly unconvincing. Bound et al. (1984) found that there was evidence of linear returns to scale for R&D programs between $2 and $100m when patents were used as a measure of research output, but they found that their results were quite sensitive to specification assumptions. Acs and Audretsch (1988), using a measure of "major innovations" as a measure of output found that in highly concentrated industries with high barriers to entry large firms were

likely to be the source of the majority of innovations, while in less concentrated, less mature industries smaller firms were likely to be more innovative. A study by Pavitt et al. (1987), using a similar measure of output, suggested that both very small firms and very large firms were proportionately more innovative than more moderate sized firms. In general, as Cohen and Levin (1989) suggest, a failure to account for specific industry effects such as variations in demand conditions, in technological opportunity and in appropriability conditions, coupled with an inability to use project level as opposed to aggregate firm level data may be responsible for the inconclusive nature of existing results.

The role of spillovers in determining research productivity is similarly not well understood. Two significant practical problems make it difficult to measure their impact (Griliches, 1991). In the first place it is difficult to distinguish between the influence that spillovers from other industries have on research productivity through their unmeasured effect upon input costs, and the influence that they have directly on research productivity by increasing the total stock of knowledge available to researchers. In the second place, the accurate estimation of spillover effects is dependent on adequate measures of technological "distance" and of R&D capital, both constructs that are exceedingly difficult to measure. Jaffe's work (1986,1988) provides the best example of a careful attempt to account for these problems, and his work suggests that spillovers have important effects on research productivity, but there have been few attempts to duplicate his results using disaggregated data or more detailed measures of technological distance.

Research in the Pharmaceutical Industry

Before turning to hypothesis development it is important to describe the nature of research and development in the pharmaceutical industry since it has several quite distinct characteristics that shape the determinants of

research productivity (Cocks (1973,1975), Caglarcan (1978), Chein (1979), Gross (1983), Spilker (1989)).

Pharmaceutical drugs are chemical compounds that are introduced into humans to cure or alleviate illness.[1] Ethical drugs can be usefully grouped into "therapeutic areas" such as "cardiovascular diseases" or "disorders of the alimentary system," although some drugs, such as aspirin, have multiple effects across several therapeutic areas. Within therapeutic classes, drugs address particular conditions such as hypertension or peptic ulcers. To a first approximation these more narrowly defined areas correspond to distinct "markets" but since, in general, drugs that affect a particular organ system tend to have multiple effects, a drug may be used to treat several conditions within a particular therapeutic class. Some of the drugs used in the treatment of hypertension, for example, are also useful in the treatment of congestive heart failure.

Pharmaceutical research takes place in two stages: drug discovery and drug development. The goal of the drug discovery process is to find a chemical compound that has a desirable effect in a "screen" that mimics some aspect of a disease state in man. For example, firms might screen hundreds of soil samples in the hope of finding a chemical compound that kills a particular kind of bacteria, or they might test fewer carefully synthesized compounds to find out if they block a particular enzymatic pathway. Drug discovery has become an increasingly complex process as our knowledge of chemistry and human physiology has increased. Whereas thirty years ago a majority of drugs were discovered through random screening, modern drug discovery relies on the integration of knowledge from a very large number of rapidly changing scientific disciplines. Screening remains important since it is very difficult to

---

[1] Our focus here is upon "ethical" pharmaceuticals - drugs that can only be dispensed with a prescription.

predict how a particular compound will react inside a biological system, but it is now informed to a much greater extent by a detailed knowledge of scientific disciplines such as biochemistry, molecular biology and physiology.

Individual chemical compounds, and hence particular drugs, can be patented. Although patents can usually be effectively enforced and thus have great competitive importance, the ability to patent a particular compound does not guarantee full appropriability of the knowledge generated during its discovery. It is sometimes possible to "invent around" a particular patent. Rival firms can explore variants in chemical structure in the hope of finding an equally effective compound with fewer side effects, for example. Most importantly, the institutional structure of the industry, particularly the very high premium placed on publication and on open communication between scientists, means that the scientific knowledge generated in the course of a particular research program is often quickly and widely disseminated throughout the industry.[2]

Some of the compounds identified during the drug discovery process will enter drug development. The goal of the drug development process is to ensure that a particular compound is safe and effective in humans. Firms apply for a permission to test a compound in humans by filing an investigational new drug application, or "IND." If clinical trials seem successful, the company will then submit a new drug application, or "NDA" to the Federal Drug Administration, which will then rule as to whether the drug can be introduced commercially.[3] Both drug discovery and drug development are risky. Of

---

[2] Not all firms place a high premium on publication, and major biological discoveries are more likely to be published before major chemical discoveries. However the industry remains a strikingly "open" one in comparison to the majority of technologically driven industries.

[3] The process described here is in use in the United States: a comparable process is required by nearly every other developed nation.

approximately 1000 compounds tested in the laboratory and in animals, less than 10 will be introduced into humans, and that of these only about 20% will ultimately be approved (Sheck, (1984)).

Hypothesis Formulation

Although our database includes information about both drug discovery and drug development, since the two require quite different sets of skills we focus here upon the determinants of research productivity in ethical drug discovery.[4] Given the nature of R&D in the industry, all else equal we expect there to be significant economies of scale and scope in performing R&D, and we expect research productivity to be affected by both internal and external spillovers.

Returns to scale flow from two sources: from the ability to spread fixed costs out across a larger effort and from the ability to invest in specialized skills or more efficient techniques as the size of the effort increases. Conventional wisdom in the pharmaceutical industry suggests that beyond a minimum threshold size, under most circumstances there is little to gain from increasing the size of an individual discovery program, and our descriptive statistics (see below) certainly confirm this conclusion in outline. However since pharmaceutical research often requires investment in substantial fixed costs and since the complexity of the underlying science offers considerable scope for specialization, we expect significant economies of scale at the level of the firm. For example, to the degree that legal and regulatory

---

[4] Drug discovery is fundamentally a scientific process requiring the integration of complex scientific knowledge across a wide range of disciplines. Successful drug development, in contrast, requires the management of a program of clinical testing that is often initiated world wide using a diverse group of medical practitioners, and the integration of a far wider range of business functions, including toxicology, product formulation and process development (Spilker, 1989).

expertise are fixed costs, a larger research effort will gain economies of scale by spreading them over a larger base. A larger R&D effort might also be able to afford to invest in more highly specialized individuals or facilities. Whereas a smaller research and development effort might need to employ molecular biologists who can work across a relatively wide range of fields, for example, a larger one might be able to afford more narrowly focused specialists.

We also expect there to be significant economies of scope. Economies of scope exist when a tangible asset or a human resource can be used in more than one application at no additional cost. Consider, for example, the benefits of investing in a centralized laboratory devoted to peptide chemistry. Economies of *scale* exist if the costs of the laboratory are partially fixed, and if the lab can serve a larger and larger discovery effort for a less than proportionate increase in cost. They will also exist if the laboratory can become more efficient as it has more work to do, possibly through the specialization of its members. Economies of *scope* exist if the work of the peptide chemists is potentially relevant to a wide range of applications, and can be utilized in any one of them without diminishing its usefulness in the others.[5] Economies of scope may also arise if there are internal spillovers of knowledge, and the results of successful research in one field have implications for work in other fields.

Thus in general, while we expect economies of scale at the level of the firm to be simply related to the total size of the discovery effort, economies of scope should be related to the range and diversity of the research programs undertaken by the firm. Notice that it is difficult to distinguish between the benefits of economies of scope that arise from cost (or risk)

---

[5] The aggressive exploitation of a well established brand name by companies such as Coca-Cola and Johnson and Johnson is another well known example of the presence of economies of scope.

sharing and the benefits of internal spillovers. Conceptually, a clear distinction can be drawn between spillovers, in the sense that research input in one area may generate knowledge or research results applicable to other areas (research in neurochemistry directed towards diseases of the central nervous system may generate compounds which turn out to have clinical application in anti-spasmodics), and economies of scope arising from the public goods aspects of core bodies of knowledge applicable to many fields (peptide chemists). In practice, both effects are likely to be a function of the diversity of the firm's research activities and the linkages between them.

At both the level of the firm and the level of the research program, economies of scale and scope will flow from the firm's ability to invest in increased levels of specialization and to spread the fixed costs of scientific expertise across a wider base of research. The relative size of the two effects will depend upon the degree to which both the tangible resources and the less tangible knowledge upon which drug discovery is based are exclusively useful to a particular research program (or broader therapeutic area) or to the process of drug discovery in general. We thus expect economies of scale at the level of the research program to vary across fields, and economies of scope to vary by firm as firms pursue different mixes of research programs.

Research productivity may also be affected by the efforts of competing firms. On the one hand, firms may be "racing" with each other to reach a particular target, so that, all other things equal, research productivity will be negatively correlated with competitors' investments (Reinganum, (1989)). In the extreme, firms will keep entering the race until all of the potential (privately appropriable) benefits of the new drug will be dissipated by the costs of competing research (Dasgupta and Stiglitz, (1980)). On the other hand, firms may benefit from competitor's research since, all other things equal, if there are extensive spillovers between firms, the productivity of a research team may increase as others work in the same field. For

example, when Bristol-Myers Squibb announced that they had found an orally active ACE inhibitor, "Captopril," a potent hypertensive therapy, several competing firms were able to take advantage of their knowledge of Captopril's chemical structure to focus their own research efforts. If there are significant spillovers in the industry, a belief certainly consistent with the qualitative evidence, then competitive investment will increase the marginal productivity of investment in the industry.

In both theory and practice, of course, these two effects interact with each other in complex ways (Spence (1984), Dasgupta and Stiglitz (1980), Baldwin and Scott (1987), Reinganum (1989)). In our companion paper "Racing to Invest?: The Dynamics of Competition in Ethical Drug Discovery" we explore this issue in more detail. For the purposes of the analysis presented here, we hypothesize that, all other things equal, a *positive* correlation between research success in any given area across competing firms is consistent with the presence of significant spillovers or research complementarities between firms, while a *negative* correlation is consistent with a research environment that is primarily driven by racing behavior.[6]

Prior Research in Pharmaceuticals

Existing studies of the determinants of research productivity within the pharmaceutical industry provide some support for these hypotheses but have been hampered by a reliance upon aggregate, firm level data. A majority have

---

[6] This conclusion must be carefully qualified: our discussion has assumed that scientific opportunity - the base case relationship between effort and the odds of obtaining a drug - is constant across these regimes. If scientific opportunity changes in ways that we can not observe using our data, then a positive correlation of either investment in research or of research success across competing firms may reflect an increase in scientific opportunity, rather than the presence of spillovers. We return to this point in the interpretation of our results.

used new drug applications or new drug introductions as their primary measure of output, and firm level data that does not discriminate between research programs or between spending on discovery and development. These studies suggest that prior to 1962 there were significant diseconomies of scale but that since 1962 larger firms have enjoyed important economies of scale in research and development (Comanor (1965,1986), Baily (1972), Cocks (1973), Vernon and Gusen (1974), Grabowski et al. (1978), Schwartzman (1976), Wiggens (1979), Jensen (1987).) Most of these studies interpret this result as suggesting that the increased regulatory stringency that followed the 1962 amendments to the Food and Drug act gave an advantage to larger firms in allowing them to exploit economies of scope in dealing with the regulatory authorities. There has been little systematic exploration of the role of spillovers in the industry, with the notable exception of work by Dranove and Ward (1991) and Arora and Gambardella (1991). This work suggests that spillovers are important, but the limitations of their data allowed them only to explore their implications at the aggregate level.

2. Specification of the Econometric Model

The measurement of "true" research productivity is a project fraught with well known problems (Griliches (1984)). In the ideal case, we would like to measure the social as opposed to the private returns to research. Unfortunately the problems of measuring the social returns to investment in pharmaceutical research are likely to be particularly acute because prices for pharmaceuticals may not reflect the functioning of an efficient market.[7] Even private returns to R&D investment in this industry are difficult to estimate: the

---

[7] The attempted measurement of social return through hedonic analysis of pharmaceutical characteristics and pricing is underway in studies by Ernie Berndt at MIT and Valerie Suslow at Michigan.

economic returns realized by firms are the final result of a lengthy and uncertain process, driven by the uncertainties of clinical testing, the complexities introduced by marketing, competitive activity and the role of the forces that determine the demand for new therapies, as well as by scientific discovery at the laboratory bench. These problems are compounded by substantial difficulties in estimating costs of capital and in making appropriate adjustments for risk (Di Masi, (1991); Grabowski and Vernon, (1978)).

Our objectives in this paper are more modest. As a first step towards building a quantitative model of research productivity, we focus on the determinants of "technical success" in drug discovery, as measured by patent grants. Pharmaceutical companies patent prolifically, and patents are, of course, a rather noisy measure of research success, in part because the significance of individual patents varies widely. We control for this by counting only "important" patents, where an "important" patent is defined as one that was granted in two of the three major jurisdictions: Japan, Europe and the United States. We think of these patents as a useful measure of the generation of new knowledge, which is the "raw material" input to subsequent stages in drug development. Patents are clearly only one possible measure of success, and later work will explore the use of IND counts, NDA counts, sales and market share as alternative measures of research output. But our interest here is on determinants of technical success, defined in terms of producing new potentially important compounds, rather than on the ultimate commercial success or failure of new drugs, for which patent grants based on Patent Examiners' slowly changing objective criteria of novelty, non-obviousness, and potential industrial application, are one reasonable measure of research

output.[8]

Though the findings may not speak directly to the question of what determines economic or commercial productivity, we believe them to be an important contribution to further progress in understanding R&D in pharmaceuticals. Research results in programs which are not subsequently successful in the clinic may be valuable to other firms, in other fields, or in later efforts in the same program, and thus insight into their generation has broader implications. We believe that it would be misleading to base our analysis on the *ex post* success of a few blockbuster drugs.

We hypothesize that patent counts are generated by a production function $Y = f(X,\beta)$, where Y is patent counts, X is a vector of inputs to the drug discovery process, and β is a vector of parameters. We have no priors about the "true" functional form, so the model estimated should be thought of as a local approximation.

Some previous studies have looked at the dynamics of the R&D/patents relationship by estimating a lag structure on the input variables. Rather than make assumptions about distributed lags (and have to throw out much of our data in order to have 4 or 5 lags present in every program) we simply include "stocks" of the input variables as explanatory variables. The annual flows are reasonably smooth, so this is equivalent in many senses to imposing a geometric lag structure, where we have assumed a depreciation rate rather than estimated one. Notice that given a smooth series for the flow

---

[8] Since firms may differ in their patenting strategies, or in their "propensity to patent," care should be taken in interpreting our results. We include firm dummies in our models to control for differences between firms that persist over time, but in general our results are most robustly interpreted in terms of marginal effects upon patent output.

variable, it will be difficult as a practical matter to identify the estimated coefficient on the stock variable separately from the depreciation rate, making our assumption of a particular depreciation rate a second-order problem.

Since the dependent variable in this relationship only takes on non-negative integer values, some type of discrete dependent variable model is dictated. We assume that patent counts are generated by a Poisson process, which is appropriate if we are prepared to model research results as the outcome of an unknown (but large) number of Bernoulli trials with a small probability of success. This model certainly captures some aspects of drug discovery, such as screening. It may be less appropriate for mechanism-based research.

We model the single parameter of the Poisson distribution function, λ, as a function of some explanatory variables, X, and parameters β in the standard fashion:

$$E[Y_{it}] = \lambda_{it} = \exp(X_{it}\beta)$$

to guarantee non-negativity of λ, and estimate the parameters by maximum likelihood in the standard way. Note that the choice of whether to use explanatory variables in levels or logs has important implications in this model. If we use levels, the estimated elasticity of output with respect to each explanatory variable will vary with the magnitude of the variable *by assumption*. Conversely, if an explanatory variable enters in logs we impose the constraint that the elasticity is constant over its range of variation. In the case of R&D, we prefer to maintain the null hypothesis of constant elasticity of Y with respect to R&D expenditures[9] allowing easy comparison with

[9] In exploratory work we tested for non-linearities in this relationship by including separate quadratic terms in the log of R&D. Estimated coefficients were very small and insignificant.

previous work. Since we have a fair number of observations in which the R&D variables are zero, using the log of the R&D variables introduces the complication that we cannot take the log of zero. Following previous work, we deal with this by setting the log of R&D equal to zero in such cases, including an appropriately coded dummy variable to account for this in the regression. We have no strong priors about the appropriate way to include the other explanatory variables and we report results obtained by entering these variables in levels. Many of these variables also have substantial numbers of zeros, and this avoids numerical problems in the estimation caused by near collinearity of the VARIABLE=0 dummy variables with other regressors.[10]

A useful way to think about this specification is to divide the explanatory variables into two classes: the R&D variables, R, and other variables, Z, (which include spillovers, measures of scope and scale, plus the constant term and firm and therapeutic class dummies). The estimated function has a direct proportionate relationship between R&D expenditures and patent counts, mediated by a multiplicative "shift variable" which will vary according to the extent of spillovers etc.

$$E[Y_{it}] = \lambda_{it} = \exp(\beta \log(R_{it}) + \gamma Z_{it})$$
$$= R_{it}^{\beta} \cdot \exp(\gamma Z_{it})$$

Re-writing the equation in logs,

$$\log(\lambda_{it}) = \beta \log(R_{it}) + \gamma Z_{it}$$

thus we can interpret the coefficient on *log(R)* directly as the elasticity of Y with respect to R&D, while the elasticities of the Z additional variables are γZ.

The assumption that the dependent variable is distributed Poisson is

---

[10] Very similar results were obtained in exploratory work using all explanatory variables in logs.

quite strong: like most other data of this type, the mean=variance property of the Poisson distribution is violated here. In the presence of such overdispersion, though the parameters β will be consistently estimated, their standard errors will typically be under-estimated, leading to spuriously high levels of significance. Overdispersion is often interpreted as evidence that the statistical model is misspecified in the sense that there may be unobserved variables in the equation for λ,

$$E[Y_{it}] = \lambda_{it} = \exp(X_{it}\beta + \epsilon_{it})$$

As is well known (see Hausman, Hall, Griliches (1984), Hall, Griliches, Hausman (1986)) if ϵ is distributed gamma, then it can be integrated out giving Y distributed as a negative binomial variate. If ϵ is not truly gamma, however, then the maximum likelihood estimates of the coefficients of the model will be inconsistent. Gourieroux, Montfort, and Trognon (1984) suggest using a quasi-generalized pseudo-maximum likelihood estimator based on the first two moments of the distribution of Y, which gives consistent estimates for ϵ drawn from a wide variety of distributions. The GMT estimator is just weighted non-linear least squares estimates of the NLLS model

$$Y_{it} = \exp(X_{it}\beta) + \epsilon_{it}$$

with weights derived from the relation VAR[Y] = E[Y] $(1+\eta^2$ E[Y]) using initial consistent estimates of β. Below we present alternate estimates of some of our regression models using maximum-likelihood estimation of the Poisson and Negative Binomial models, non-linear least squares (with robust standard errors), and the GMT estimator.

3. Sources and Construction of the Data Set.

This paper uses a data set obtained as part of a larger study of

research productivity in the pharmaceutical industry. The larger study has both qualitative and quantitative components. The qualitative study draws upon the medical and scientific literature and upon a program of detailed field interviews, and is designed both to shape the choice of variables and hypotheses explored in the quantitative study and to explore the role of less easily measurable factors such as organizational structure and firm culture in driving research productivity.

The quantitative study draws upon data about spending and output at the *research program level* obtained from the internal records of ten pharmaceutical firms. Although for reasons of confidentiality we cannot describe specifics of the overall size or nature of the firms, we can say that they cover the range of major R&D-performing pharmaceutical manufacturers, that they include both American and European manufacturers, and that we believe that they are not markedly unrepresentative of the industry in terms of size, or technical and commercial performance. This section offers a brief description of the important variables used in the econometric analysis, and discusses descriptive statistics for the data set.

The data set used in this paper is an unbalanced panel of 4879 observations, indexed by firm, research program, and year. With a complete, rectangular, panel we would have 11,400 observations, made up of ten firms, 38 research programs, and up to 30 years of data. In practice not all of these observations are available: the average time period for which we have complete data is on average just under 20 years per firm, and not all firms are active in all research areas. Our working sample is drawn from a data base which currently has 5543 potentially useful observations. After deleting missing values, grossly problematic data and peripheral research areas we are left with 4879 observations. The number of observations per firm varies from over 1000 to less than 100, with a mean of 554.3. For each observation we

have data on both inputs and outputs to the research process. Our measures of input include person years and research spending in discovery and development, and our measures of output include patents, INDs, NDAs, new drug introductions, sales and market share.

Assembling the data in a consistent and meaningful format required considerable effort. In nearly every case the process of data collection was an iterative one, involving close collaboration between the researchers and key personnel from the participating companies. The majority of the data were collected specifically together for the purposes of this study. Each firm spent some months assembling the data, usually from primary documents, and the full data collection effort took nearly two years. We worked hard to ensure that, as far as was possible, definitions of research program and of expense grouping were standard across firms, data was collected at the same level of aggregation, and overhead expenses were treated in a consistent way.

Data was collected by research program rather than by therapeutic class or by project since analyzing the problem in this way best reflects the dynamics of discovery research. A grouping by therapeutic class is too general: "cardiovascular research," for example, includes research into widely different areas such as hypertension, cardiotonics, antiarrhythmics and hyperlipoproteinemia. However analysis of data by individual project is difficult and misleading. Not only is it difficult to assign effort to particular drug candidates with any accuracy, but the notion that research productivity is best measured at this level raises serious conceptual difficulties. A research program typically continues over many years. At the discovery stage, the firm invests in the program, rather than in particular candidates. The identification of a drug development candidate is an indication of the success of the program, and retrospectively assigning resources to its generation may introduce serious biases into the analysis.

Classification of our data into therapeutic areas is an important factor

in our analysis, since it drives both the fundamental organization of our data, and the notion of spillovers.[11] Note also a crucial distinction drawn between two tiers of aggregation: a detailed "research program" level, and a more aggregated "therapeutic class" level which groups related programs into therapeutic areas. For example, the therapeutic class "CARDIOVASCULAR" includes the research programs "ANTI-HYPERTENSIVES", "CARDIOTONICS", "ANTITHROMBOLYTICS", "DIURETICS" etc.

Full details of the construction of variables used in the study are given in the Appendix. Our primary variables are DISCOVERY, defined as "expenditure relating primarily to the production of new compounds", which excludes clinical development work and is measured in constant dollars; and PATENTS, a count of "important" patents. These measures of inputs and outputs are matched by year and research program. We count patents by their year of application, and define "importance" by the fact that the patent was granted in two of the three major markets: the USA, Japan, and the European Community. Applying this criterion screens out large numbers of patents: up to 60% of the number filed in the US in any given year are discarded. The first year in which we were able to obtain this data is 1961, and although we have data for 1989 and 1990, we believe that these are seriously undercounted. Patent grants may lag applications by as much as four years in the United States and six in Japan. Our data were obtained from files current in 1992, and since we are counting patents by year of application, many of the patents applied for in 1989 and 1990 may not yet have been granted, and our final

---

[11] We classified our data into therapeutic areas according to a scheme which closely follows the IMS Worldwide classes. A more detailed discussion of this issue is contained in Appendix One, and a complete listing of therapeutic classes is given in Appendix Two.

sample therefore includes only observations for the years 1961-1988.[12]

We tried our best to ensure that the DISCOVERY variable was measured consistently across research programs and across firms, despite serious differences in accounting conventions and reporting formats across firms and over time, and we took steps to ensure that, wherever possible, the data covered worldwide research spending, not just US facilities.

We also construct measures of the overall size of the firm's R&D effort and of the scope of its activities. SIZE is total discovery spending that year, which is intended to capture the effects of the overall scale of the firm's research effort.[13] SCOPE is a count of the number of narrowly defined research areas in which the firm is active, in that it spent more than $500,000, in constant 1986 dollars, in a single year. This variable is intended to capture the effects of the firm's diversification into different therapeutic areas. Note that SCOPE and SIZE are unique only to firm and year, not to firm, year, and research program.

We then construct variables intended to capture the effects of spillovers both within and between firms. For each observation on firm and

---

[12] Several industry experts have suggested that this is an unsatisfactory measure of "importance" in that many purely "defensive" patents are filed in at least Europe and the U.S. They suggest that a more appropriate gauge of importance would be filing in fifteen or twenty counties. There may also be important differences in the patenting policies of European vs US firms. However preliminary analysis using INDs as a measure of output suggests that our measure of patents is significantly correlated with INDs, confirming us in the belief that we are capturing an important dimension of performance.

[13] There are some grounds for believing that the relevant measure of size is the scale of the entire firm, captured by e.g. total sales, or total employees. We experimented with these types of measures in exploratory work, but obtained poor results. These are perhaps not too surprising given that size effect captured by a variable such as total firm sales is likely to be seriously confounded with e.g. demand.

research area, we start with the basic data on "own" annual flows of inputs and outputs. We then construct spillover variables at two levels. We capture spillovers internal to the firm by measuring the output of all of the other research programs within the relevant therapeutic class. We measure spillovers between firms using the output of 29 other firms, both in the observation's own narrow research area, and in the wider therapeutic class.[14] Finally, we construct "stocks" for all of these variables by accumulating the flows over time with a 20% depreciation rate, and also a "news" variable by subtracting 20% of the stock at the beginning of the year from the annual flow.

Descriptive Statistics

Tables (1) and (2) and Figures (1),(2) and (3) present summary statistics describing these data. The first five columns in Table 1 are for our full set of data with up to 5543 observations on each variable, and are "case-wise" calculations based on each variable's non-missing observations. Means for the sample of 4879 observations used in our regressions which is obtained by "list-wise" deletion and imputation of zeroes for some missing data, are given in the last column.

Averaging across all firms, research programs, and years, our firms spent on average $1.99m 1986 dollars on discovery per program per year for an average of 1.7 "important" patents. Each "important" patent, on average, thus cost about $1.2m in 1986 dollars. The average firm in our sample is

---

[14] We chose as our sample the ten firms that have given us data together with 19 other firms who have been consistently in the top 40 world wide pharmaceutical firms in terms of R&D dollars and sales. Note that these 19 firms are only a fraction of the population of other firms generating spillovers, and the estimated coefficient on our external spillover variables may therefore tend to overstate the magnitude of this effect. However we believe that these firms are a representative sample of the industry as a whole, and account for the majority of the industry's spillover pool.

highly diversified, investing substantially (more than half a million dollars) in just over ten programs a year. In addition firms invest more than ten thousand dollars a year in a further six programs.

All of the key variables show a substantial amount of variation. Discovery expenditures per class per year have grown substantially for the firms in our data set, from $730,000 1986 dollars per class per year in 1964 to $4.2m 1986 dollars per class per year in 1990 (Figure 1). At the same time, our data suggest that there has been a secular decline in the number of patents granted per program per year since the late 70s. (Figure (1).) This decline may be an artifact of our definition of "important" patents, since in many cases patents do not issue in Japan until quite late in the lifetime of a product, but it may also reflect the general decline in patenting rates that has characterized US and European firms over the last five years.[15]

Perhaps the most dramatic effect visible in the time series aggregates of our data is the huge increase in the costs of pharmaceutical research: the average firm in this sample's total investment in discovery research more than quadrupled in real terms from $16m constant 1986 dollars in 1964 to almost $67m constant 1986 dollars in 1990, reflecting increased intensity of highly skilled manpower in R&D and very substantial increases in the complexity and sophistication of research methods, equipment, materials etc. This aggregate figure masks some important movements in the program level data. While there was a very significant "deepening" of individual research programs (real discovery spending per program rises by almost 600% over the same period), this was accompanied by changes in the scope of the average firm's activities: the number of half million dollars plus programs in our average firm's

---

[15] This in turn may not reflect a real decline in inventive activity, but institutional factors such as resource constraints imposed on patent offices. See Griliches (1991).

portfolio rises by 31%, from 8.2 to 10.7, and concomitant movements in the numbers of smaller programs active in any given year. (See Figure (2).) It should be noted though that the variation in SCOPE of the whole sample is driven largely by differences between firms, rather than within firms. In fact SCOPE varies rather little within firms, with only the most diversified firm having a CV (std/mean) of more than 0.3, suggesting that it is likely to be difficult to distinguish between scope and firm effects in subsequent analysis.

The size distribution of discovery expenditures per research program per year, presented in Figure (3), is also interesting. For over 44% of the program-years in our sample, no expenditures on discovery were recorded.[16] Of the remainder, about one quarter of cases involved spending less than $0.2m 1986 dollars per program per year, and about one half spent less than $0.8m 1986 dollars per program per year. At the other tail of the distribution, just under 2% of cases involved expenditures of more than $10m 1986 dollars per program per year. There are also very substantial differences in the average level of expenditure across therapeutic classes: mean discovery expenditures per program range from $940,000 per year in alimentary tract and metabolic research to $4.0m per year in cytostatics.

On the output side, the distribution of annual counts of important patents per program (Figure (4)) is highly skewed to the left. About half of our observations show zero output per program-year, and almost 90% of the counts are less than 5 per year. Patent counts also vary significantly across firms, across therapeutic classes and across time. While the mean number of patents per program is 1.7 per year, this varies across firms from 3.1 to 0.09

---

[16] To the best of our knowledge these are "genuine" zeros and reflect intermittent expenditures over time, or programs which were "alive" in that they were spending money on development. Deleting these observations from the data set does not substantially change either the magnitude or the significance of our estimated coefficients.

per year, and across therapeutic classes from 0.7 per year for work in the genito-urinary system to 3.7 per year for cardiac and circulatory products.

4. The Empirical Results.

Analysis of aggregate data.

As a preliminary step in the analysis, Table (3) presents results obtained by aggregating our data to the firm level to explore the effects of scope and scale. This aggregated version of our data is a rather small sample, with only 181 observations on our 10 firms, and results should be therefore be treated with caution. Specifications (1) and (2) are Poisson regression estimates of "important" patents on flow and stock values of DISCOVERY, a quadratic time trend, and firm dummies. (Nonsensical results were obtained without the time trend, which is included in all our models.) The implied long run elasticity of "important" patent output with respect to research spending in model (2) is about 0.4, somewhat lower than, but not inconsistent with, the results obtained for much larger samples of firm level data by e.g. Bound et al. (1984), Hausman, Hall, and Griliches (1984) and Hall, Griliches and Hausman (1986).[17] This elasticity is well below 1, suggesting that there are not increasing returns to scale in drug discovery at the firm level. Note that these results suggest that firm effects are a very important determinant of innovative performance. Including firm dummies more than doubles the log-likelihood function, and failing to control for firm effects, as in model (1), gives a puzzlingly low coefficient on contemporaneous R&D, indicating that this version of the model may be badly miss-specified. In equations (3) and (4)

---

[17] There are two important differences between this work and previous studies which make these results not strictly comparable. Firstly, we have distinguished between discovery and development expenditures, and secondly, our patent counts are restricted to "important" patents.

we use our knowledge of the number of active research programs within each firm to investigate the presence of returns to scope. In equation (3) our measure of scope enters the regression linearly, with an implausible negative coefficient. Our preferred model allows for diminishing returns to increasing scope by including a quadratic term, giving us a more plausible inverted-U relationship between scope and research productivity. The final column of Table (3) presents results conditioning on past innovative success by including the stock of past patents. The coefficient on R&D stock falls somewhat, but the results on scope are largely unchanged.

Thus, in aggregate, our data appear to reflect the general cross sectional result: there is no evidence of returns to scale *per se* in research: increasing R&D intensity in and of itself does not produce a more than proportionate increase in innovative output. The results do however suggest that up to a point, size may confer substantial benefits: at the mean, an increasing the size of the total research effort by 10% (an amount equivalent to adding an additional program) is associated with a 3% increase in the patent output of existing programs.

If we did not have access to program level data we might stop here, noting that this result is consistent with the qualitative evidence that suggests that there are limited returns to increasing the size of research programs, and that the returns to scale in the industry suggested by the recent wave of mergers and by the results of scholars such as Baily (1972), Cocks (1973), Comanor (1965) and Jensen (1987) are driven by and by economies of scale in activities "downstream" from discovery research, notably in clinical development and in the regulatory arena.

Analysis of Program Level Data

The results of our analysis of program level data are given in Tables

(4), (5) and (5a).

Table (4) shows the results from re-estimating variants of the equations in Table (3) at the disaggregated research program level. Again, the dependent variable is "important" patent applications and the models were estimated using Poisson regression.

Two core findings are immediately apparent. Firstly, there is a statistically significant relationship between inputs and outputs in these disaggregated data. Although prior research has shown that there is a statistically significant relationship between inputs and outputs at the firm level, it is reassuring to find that at the research program level there is more in the quantitative data than simply noise.[18]

Secondly, there are very significant differences in research productivity across therapeutic classes and between firms. In these data we are able to control for an important source of heterogeneity in the data by including therapeutic class dummies (defined at a somewhat higher level of aggregation than our "program" unit of observation) as well as controlling for fixed firm effects. Likelihood ratio tests on excluding firm and therapeutic class dummies are highly significant, and interaction effects on discovery spending show large differences across firms and therapeutic classes.[19] Note that while firm dummies account for a substantial amount of the variance in the dependent variable, the R&D elasticities do not change much when they

---

[18] The very substantial increase in the log likelihood as we add variables to the model corresponds to an R^2 in excess of 0.5 which corresponds favorably to typical panel data results using micro data.

[19] With both firm fixed effects and therapeutic class fixed effects we are quite close to estimating a non-linear panel data model with fixed effects. While we believe that is important to control for these effects at these levels of aggregation, the large numbers of dummy variables (10 firms and 16 therapeutic classes) limits our ability to obtain stable estimates if we try to, for example, use time dummies instead of a non-linear trend.

are included in the model. By contrast the coefficients on the R&D variables fall by about a third in equation (8) when the class dummies are included. This suggests that an important part of the "firm effect" identified in the aggregate data and the results presented in Table (4) lies in the firm's choice of research programs. Just as in the firm level results, SCOPE has a significant and non linear impact on research productivity. In these data we are also able to test for scale effects derived from the total size of the research program: the SIZE variable enters with a positive and significant coefficient in equation (12), which increases somewhat when we control for SCOPE in equations (13) and (14). The final equation in Table (4) conditions on past success (or innovative capabilities) by including the stock of past patents. This variable is highly significant, and induces a correspondingly large drop in the likelihood function. The coefficient of 0.035 corresponds to an elasticity of about 0.25 at the mean. The SIZE and SCOPE effects are unchanged, but the coefficients on R&D fall sharply. We interpret this as being consistent with the hypothesis that patent applications are driven by the available stock of knowledge capital, and that the better measure of this stock is innovative output rather than innovative input. It may also however reflect two specification problems: the patent stock variable may be proxying for a variety of unobserved correlated effects, or there may be a problem with the exogeneity of the R&D variables with respect to patents. Conditioning on past success may simply be purging the R&D coefficients of the part which is the endogenous response to past success.

In Table (5) we introduce our measures of internal and external spillovers. (The first column of Table (5) simply duplicates the last column of Table (4) to allow easy comparison.) In exploratory work we included both flow and stock versions of these variables in the model, but concerns about the presence of measurement error (in many cases the coefficients were equal but opposite-signed suggesting that they were picking up the same underlying

factor) led us to use the "news" formulation presented here, in which news in X is given by $N_t = X_t - \delta K_{t-1}$ where K is the stock of X and δ is the depreciation rate. This construction reduces the measurement error problem and has an informative interpretation: own research productivity is higher when the output of spillover sources "spurts" beyond the level required simply to maintain their previous stock. Equation (15) includes a measure of internal spillovers: "news" in the output of related programs within the same firm, which enters with a positive and strongly significant coefficient. An alternative, and perhaps less satisfactory, measure of internal spillovers is effort in related classes (as measured by discovery spending) which we introduce to the basic model by itself in equation (16), and in conjunction with the output-based variable in equation (17). By itself, the effort variable is positive and marginally significant, but the "news in related patents" knocks it out completely in (17) and we drop it from further consideration.

In equation (18) the basic model is augmented with measures of external spillovers from outside the firm, both from within the narrowly defined research area, and from related therapeutic classes. Both enter with positive and significant coefficients, with elasticities of about 0.1 at the mean. Re-introducing the internal spillovers variable in (19) completes our "preferred" model, which captures the following effects: a rather low elasticity of research output with respect to R&D spending (about 0.1 in the long run); a positive effect on research productivity from increasing the scope of the firm's research activities (up to a point); a positive effect from increasing the overall scale of the research effort; a very significant effect of past research success on current performance, and positive effects of spillovers from both within the firm and from the research output of competitors.

Table (5a) addresses some of econometric issues raised previously by presenting alternative estimates of equation (19) using the various statistical

models discussed in section 2. (Again (19) is duplicated from the previous table to allow easy comparison of results.) Equation (20) gives Negative Binomial estimates, where the variance is modelled as an increasing function of the mean[20], equivalent to adding an unobserved gamma distributed random program effect to the model. Coefficients are broadly comparable to those obtained in the Poisson model, with slightly inflated standard errors, though the coefficients on stock of patents and current R&D rise quite substantially. Equation (21) is the non-linear least squares model, $Y = e^{X\beta} + \epsilon$, with robust Eicker-White standard errors. In this model, the R&D variables lose their significance altogether (and the coefficient on R&D stock falls by a factor of 4) though the other results carry through from the Poisson specification. Finally, (22) gives the results from the weighted NLLS GMT estimator which if the $\exp(X\beta)$ part of the model is correctly specified, will give consistent and efficient parameter estimates.[21] Results are broadly comparable with the Poisson estimates, with slightly larger standard errors, though the coefficients on the R&D variables are somewhat higher.

5. Conclusions

Previous research has suggested that large firms may have an advantage in R&D competition arising from their ability to exploit economies of scope and scale in conducting research. However little systematic work has

---

[20] Cameron and Trivedi's (1986) Negbin II model.

[21] The GMT estimates were calculated with weights derived from the Poisson estimates of β, rather than from unweighted NLLS. These give estimates of η^2 which are in line with the overdispersion parameter estimated in the Negative Binomial model. Using the NLLS βs gave an implausibly low estimate of η^2 of 0.11, and the resulting GMT weighted NLLS results were very difficult to interpret.

attempted to distinguish between the effects of scale and scope, and cross sectional studies using firm or line-of-business level data have found no simple relationship between size and research productivity. Our results confirm that moving within the firm through the use of detailed program level data may significantly improve our understanding of the relationship between scale, scope, spillovers and research productivity.

In line with previous studies of the pharmaceutical industry, we find no evidence of increasing returns to scale or scope at either the program or the firm level. We do however find evidence for complicated relationships between scale and scope and research productivity which are non linear and embedded in the structure of the research portfolio, and thus very difficult to pick up in aggregate firm level data.

"Up to a point," we find that research programs conducted within large firms will be significantly more productive those conducted within smaller firms, all other things equal. We identify three main advantages to running larger research efforts: economies of scale, economies of scope, and enhanced absorption of internal and external spillovers. The marginal effects suggested by our regression coefficients are difficult to interpret because SIZE, SCOPE and the program size are not independent. To get a sense of magnitudes, relocating the "average" program into a firm of twice the size would increase its productivity by 12%. Similarly, moving it from one of the least diversified firms into a firm running twice the number of programs would increase its productivity by about 20% through scope economies alone. This understates the return to diversification since the move to a more diversified firm will also increase the program's productivity through the absorption and generation of additional intra-firm spillovers.

Our results also suggest that inter-firm spillovers play an important role. "Own" research productivity is significantly positively associated with competititors' success in related research. A 10% increase in the flow of

patents in related areas at other firms generates an expected increase in patent count of around .7%. While this number may seem small, recall that these cross firm spillovers operate across the entire industry, so that the aggregate effects of on industry productivity may be quite high. If we knew that it was always equally difficult to get patents, our results would be consistent with the hypothesis that the effect of spillovers in the industry are such as to make the social rate of return to research significantly higher than the private rate of return. However we cannot conclude this from our results, since our coefficients estimate the net effects of scientific opportunity, investment behavior and the presence of spillovers. If investment behavior reflects a sudden change in the scientific opportunities in a particular field, for example, then our results are consistent with the hypothesis that when scientific opportunities in an area dramatically increase, both investments in discovery and the rate of patenting also increases. We explore this issue in more detail in our companion paper "Racing to Invest?: The Dynamics of Competition in Ethical Drug Discovery."

While the specifics of our results cannot be unilaterally extended to other industries, they do suggest that a very significant component of the firm effect evident in previous studies of research productivity is precisely the size and shape of the research portfolio. This is a finding that has fundamental implications for the larger question of the theory of the firm. It has long been known that in general minimum efficient scale in production cannot explain observed firm sizes (Panzar, 1989), and our results lend support to those that have argued that one of the most important determinants of firm size and scope in research intensive industries is the opportunity to take advantage of intra-firm spillovers of knowledge (Chandler, 1990; Teece, 1988).

The heterogeneity we observe across the firms in our sample in innovative performance reflects significant variation in their responses to sharply rising costs of doing research and the increasing role of core bodies

of knowledge about physiological mechanisms. While SIZE rose substantially for every firm in the sample, firms "grew" their research portfolios in very different ways. Though we cannot be specific, due to confidentiality issues, we note that while some firms held SCOPE (as captured by our rather crude measure) constant and simply increased their average program size, others held program size roughly constant and grew by adding new programs, others grew in both dimensions, and some chose to reduce SCOPE and greatly increase program size. At the same time, the firms in our sample employed quite different organizational structures in an attempt to ensure that the potential of both intra-firm economies of scope and inter-industry spillovers were fully realized. Pursuing these different strategies had, we believe, quite important consequences for their innovative and financial performance. We are actively exploring this issue in our research.

References

Acs, Zoltan J., and D.B. Audretsch (1988) "Innovation in Large and Small Firms: An Empirical Analysis." *American Economic Review*, September 1988, Vol. 78, No. 4, pp678-690.

Arora, Ashish, and Alfonso Gambardella (1991) "Evaluating Technological Information and Utilizing It: Scientific Knowledge, Technological Capability and External Linkages in Biotechnology." Paper presented at the NBER Summer Institute, July/Aug 1991.

Baily, Martin N. (1972) "Research and Development Costs and Returns: The U.S. Pharmaceutical Industry." *Journal of Political Economy*, January/February 1972.

Baldwin, W.L., and J.T. Scott (1987) *Market Structure and Technological Change*, Chur: Harwood Academic Publishers, 1987.

Bound, John, et al. (1984) "Who does Research and Development and who Patents?", in Zvi Griliches, (ed.) *R&D, Patents and Productivity*, Chicago: Chicago University Press, 1984.

Burns, T. and G. Stalker (1966) *The Management of Innovation*, London: Tavistock, 1966.

Caglarcan, E., et al. (Ed.) (1978) *The Pharmaceutical Industry: Economics, Performance and Government Regulation*, New York: John Wiley and Sons, Inc., 1978.

Cameron, A. Colin, and Pravin K. Trivedi (1986) "Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests", *Journal of Applied Econometrics*, Vol 1, pp29-53.

Chandler, A. (1990) *Scale and Scope*, Cambridge, MA: The MIT Press, 1990.

Chien, R.I. (ed.) (1979) *Issues in Pharmaceutical Economics*. Lexington MA: Lexington Books, 1979.

Cockburn, Iain, and Rebecca Henderson (1993) "Racing to Invest? The Dynamics of Competition in Ethical Drug Discovery", mimeo, MIT Sloan School of Management, April 1993.

Cocks, Douglas (1973) "The Impact of the 1962 Drug Amendments on R&D Productivity in the Ethical Pharmaceutical Industry." Unpublished Doctoral Dissertation, Oklahoma State University, 1973.

Cocks, Douglas, (1975) "Product Innovation and the Dynamic Elements of Competition in the Ethical Pharmaceutical Industry." in *Drug Development and Marketing*, ed. R. Helms, pp 255-254. Washington D.C.: American Enterprise Institute, 1975.

Cohen, W.M. and R. C. Levin (1989) "Empirical Studies of Innovation and Market Structure" in Schmalensee, R. and R. Willig (eds.), *Handbook of Industrial Organization, Volume 2*, Amsterdam: North Holland, 1989.

Comanor, William S. (1965) "Research and Technical Change in the Pharmaceutical Industry," *Review of Economics and Statistics*, May 1965, pp182-90.

Comanor, William S. (1986) "The Political Economy of the Pharmaceutical Industry," *Journal of Economic Literature*, September 1986, pp1178-1217.

Dasgupta, P., and J. Stiglitz (1980) "Industrial Structure and the Nature of Innovative Activity." *Economic Journal*, Vol 90, pp266-93.

DiMasi, J., et al. (1991) "The Cost of Innovation in the Pharmaceutical Industry." *Journal of Health Economics*, Vol. 10, pp107-142.

Dranove, David, and Michael Ward (1991) "The Vertical Chain of R&D in the Pharmaceutical Industry." Mimeo, Northwestern University, September 1991.

Fisher, R.M., and P.Temin (1973) "Returns to Scale in Research and Development: What Does the Schumpeterian Hypothesis Imply?" *Journal of Political Economy*, Vol 81., No. 1, pp 56-70.

Grabowski, H. G., Vernon, J. M., and Thomas, L. G. (1978) "Estimating the Effects of Regulation on Innovation: An International Comparative Analysis of the Pharmaceutical Industry," *Journal of Law and Economics*, 1978, pp133-163.

Griliches, Zvi (1984) *R&D, Patents and Productivity*, (ed.) Chicago: Chicago University Press, 1984.

Griliches, Zvi (1991) "The Search for R&D Spillovers." NBER Working Paper No. 3768, July 1991.

Gourieroux, C., A. Monfort, and A. Trognon (1984) "Pseudo-Maximum Likelihood Methods: Applications to Poisson Models", *Econometrica*, Vol. 52, No. 3, pp701-720

Gross, F. (1983) *Decision Making in Drug Research*, (ed.) New York: Raven Press, 1983.

Hall, Bronwyn H., Zvi Griliches, and Jerry Hausman (1986) "Patents and R&D: Is There a Lag?" *International Economic Review*, Vol. 27, No. 2, June 1986, pp265-283.

Hausman, Jerry, Bronwyn H. Hall, and Zvi Griliches (1984) "Econometric Models for Count Data With an Application to the Patents-R&D Relationship", *Econometrica*, Vol. 52, No. 4, pp909-938.

Jaffe, Adam B. (1986) "Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits and Market Value," *American Economic Review*, December 1986, pp984-1001.

Jaffe, Adam B. (1988) "Demand and Supply Influences in R&D Intensity and Productivity Growth." *Review of Economics and Statistics*, LXX (3), pp431-437.

Jensen, Elizabeth J. (1987) "Research Expenditures and the Discovery of New Drugs." *Journal of Industrial Economics*, September 1987. Vol XXXVI pp 83-95.

Mansfield, E. (1987) "Price Indices for R&D Inputs, 1969-1983. *Management Science*, Vol.33, No.1, 1987.

Mowery, David (1989) *Technology and the Pursuit of Economic Growth*, Cambridge, England: Cambridge University Press, 1989.

Panzar, John C. (1989) "Technological Determinants of Firm and Industry Structure," in Schmalensee, R. and R. Willig (eds.), *Handbook of Industrial Organization, Volume 2*, Amsterdam: North Holland, 1989.

Pavitt, Keith et al. (1987) "The Size Distribution of Innovative Firms in the UK.: 1945-1983," *Journal of Industrial Economics*, Vol. XXV, March 1987, pp297-319.

Reinganum, Jennifer F. (1989) "The Timing of Innovation: Research, Development and Diffusion", in Schmalensee, R. and R. Willig (eds.), *Handbook of Industrial Organization, Volume 1*, Amsterdam: North Holland, 1989.

Schwartzman, David (1976) *Innovation in the Pharmaceutical Industry*, Baltimore: Johns Hopkins University Press, 1976.

Schumpeter, Joseph A. (1934) *The Theory of Economic Development*, Cambridge, MA: Harvard University Press, 1934.

Schumpeter, Joseph A. (1950) *Capitalism, Socialism and Democracy*, Cambridge, MA: Harvard University Press, 1950.

Sheck, L. et al. (1984) "Success Rates in the US Drug Development System", *Clinical Pharmacology and Therapeutics*, Vol. 36. No. 5 pp574-583.

Spence, Michael (1984) "Cost Reduction, Competition and Industry Performance." *Econometrica*, Vol. 52., No. 1. January 1984, pp 101-121.

Spilker, Bert (1989) *Multinational Drug Companies: Issues in Drug Discovery and Development*. New York: Raven Press, 1989.

Teece, David (1988) "Technological Change and the Nature of the Firm." From Dosi, G. et al.: *Technical Change and Economic Theory*, (New York, NY: Columbia University Press, 1988)

Tushman, Michael, and Anderson, Philip (1986) "Technological Discontinuities and Organizational Environments." *Administrative Science Quarterly*, Vol. 31, pp439-465.

Vernon, J.M., and Gusen, P. (1974) "Technical Change and Firm Size: The Pharmaceutical Industry." *Review of Economics and Statistics*, August 1974, Vol xx, pp294-302.

Wiggins, Steven (1979) *Regulation and Innovation in the Pharmaceutical Industry*, Unpublished MIT PhD Dissertation, 1979.

Appendix One: Data Sources and Construction

The data set used in this study is based on detailed data on R&D inputs and outputs at the research program level for ten ethical pharmaceutical manufacturers.

Inputs

Our data on inputs to the drug discovery process are taken from the internal records of participating companies, and consist primarily of annual expenditures on exploratory research and discovery by research program. Several issues arise in dealing with these data.

(a) Discovery vs. Development

The distinction between discovery and development is important. We define resources devoted to discovery as all pre-clinical expenditures within a therapeutic class, and development as all expenses incurred after a compound has been identified as a development candidate. Where exploratory research was attributable to a particular research program, this is included in the discovery category. Non-program exploratory research was included in the overhead allocation for each research program. Clinical grants are included in the figures for development, and grants to external researchers for exploratory research are included in the total for discovery.

In some cases, the companies supplied us with data already broken down by discovery vs development by research program. In others, we had to classify budget line items for projects/programs into the appropriate category. This was done based on the description of each item in the original sources, and the location of items within the structure of the company's reporting procedure.

(b) Overhead

In order to maintain as much consistency in the data collection process as possible, we tried to ensure that these figures include appropriate overhead charges directly related to discovery activities, such as computing, R&D administration and finance etc., but exclude charges relating to allocation of central office overhead etc. The overhead also includes some expenditures on discipline-based exploratory research such as "molecular biology" which appeared not to be oriented towards specific therapies. Overhead was allocated across therapeutic classes according to their fraction of total spending.

(c) Licensing

We treat up-front, lump-sum payments in respect of in-licensing of compounds, or participation in joint programs with other pharmaceutical

companies, universities or research institutes, as expenditure on discovery. Royalty fees and contingent payments are excluded.

Outputs

In this paper we focus on patent grants as our measure of research output. We count patents by year of application. Our interest here is on determinants of technical success, defined in terms of producing new potentially important compounds, rather than on the ultimate commercial success or failure of new drugs. Since Patent Examiners award grants based on slowly changing objective criteria of novelty, non-obviousness, and potential industrial application, we believe that patent grants are an appropriate basis for measuring research output in this industry. Pharmaceutical companies patent prolifically, and patents are, of course, a rather noisy measure of research success, in part because the significance of individual patents varies widely. We partially control for this by counting only "important" patents, where we define "importance" by the fact that the patent was granted in two of the three major markets: the USA, Japan, and the European Community.

These data were provided by Derwent Publications Inc, who we asked to use their proprietary classification and search software to produce counts of "important" patents to us broken down by therapeutic class for 29 US, European, and Japanese pharmaceutical manufacturers for the 26 years preceding 1990. These firms were chosen to include the ten firms that have given us data together with 19 other firms chosen on the basis of their absolute R&D expenditures, R&D intensity, and national "home base" to try to get a representative, rather than exhaustive, assessment of world-wide patenting activity. These 19 firms have been consistently in the top 40 world wide pharmaceutical firms in terms of R&D dollars and sales.

Note that many of these patents will be "defensive" patents in that firms may patent compounds they do not intend to develop in the short term but that may have competitive value in the longer term. Alternative measures of "importance" include citation weighting and more detailed international filing data - "very" important patents are usually filed in nearly every major potential market. We hope to explore these alternative measures in later work.

Classification

Classification of inputs and outputs by therapeutic class is important because this drives our measure of spillovers. There are essentially two choices: to define programs by physiological mechanisms, e.g. "prostaglandin metabolism", or by "indications" or disease states, e.g. "arthritis". We have chosen to classify on the basis of indication, largely because this corresponds well to the internal divisions used by the companies in our sample (which is conceptually correct), but also because classification by mechanism is much more difficult (a practical concern.) In further work we intend to repeat the

analysis using a "cut" by mechanism. We classified both inputs and outputs according to a scheme which closely follows the IMS Worldwide classes. This scheme contains two tiers of aggregation: a detailed "research program" level, and a more aggregated "therapeutic class" level which groups related programs. For example, the therapeutic class "CARDIOVASCULAR" includes the research programs "ANTI- HYPERTENSIVES", "CARDIOTONICS", "ANTITHROMBOLYTICS", "DIURETICS" etc.

There are some problems with this procedure. Firstly, some projects and compounds are simply very difficult to classify. A particular drug may be indicated for several quite distinct therapies: consider serotonin, which has quite different physiological actions on either side of the blood-brain barrier. As a neurotransmitter it is believed to play important roles in mediating motor functions. As a systemic hormone it has a variety of effects on smooth muscle, for example it functions as a vasoconstrictor. Some companies report expenditures in areas which are very difficult to assign to particular therapeutic classes: a company doing research using rDNA technology might charge expenditure to an accounting category listed as "Gene Therapy/Molecular Biology" which is actually specific research performed on e.g. cystic fibrosis, but we have no idea about which diseases the research is directed towards treating, and are forced to include these expenditures in "overhead". Secondly, our two-tier classification scheme may not catch all important relationships between different therapeutic areas. We believe that we are undercounting, rather than overcounting in this respect, so that the importance of spillovers will be underestimated rather than overestimated. Thirdly, where firms supplied us with "pre-digested" data, they may have used substantively different conventions in classifying projects. One firm may subsume antiviral research under a wider class of anti-infectives, while another may report antivirals separately. Not surprisingly there are major changes within companies in internal divisional structures, reporting formats, and so forth, which may also introduce classification errors. After working very carefully with these data, we recognize the potential for serious miss-assignment of outputs to inputs, but we believe that such errors that remain are not serious. The use of patents as the output measure should reduce vulnerability to this problem, since we observe relatively large numbers, and a few miss-classifications are unlikely to seriously affect our results. When we move to INDs and NDAs as our output measures, the much more sparsely distributed data are likely increase our vulnerability.

Matching

Data series on inputs and outputs for each firm were matched at the research program level. This procedure appears to successfully match outputs and inputs unambiguously for the great majority of programs. In a very few cases, however, we ended up with research programs where patents, INDs or

NDAs were filed, but where there were no recorded expenditures. Of these the majority were obviously coding errors or reflected dilemmas previously encountered in the classification process, and appropriate corrections were made. In other cases, it was clear that these reflected "spillovers" -- research done ostensibly in, for example, hypertension, may generate knowledge about the autonomic nervous system which prompts patenting of compounds may be useful in treating secretory disorders (e.g. ulcers.) In such cases we set "own" inputs for the program equal to zero, and included these observations in the data base.

Deflation

Since our data sources span many years, it is important to base the analysis on constant dollar expenditures. We used the R&D price deflator constructed by Edwin Mansfield (1987) for his Oil and Chemicals industry grouping. This index is based on wage rates for R&D employees, and a price index for equipment and instrumentation purchases, and though its movement is quite different from the CPI or the GNP deflator, it varies much less across industries, leading us to believe that it may be a reasonable approximation to the "correct" index for pharmaceuticals. Mansfield's index exists only for 1969-1983, we extended it backwards to 1966 and forwards to 1990 using movement in the CPI. The periods 1966-1969 and 1983-1990 saw relatively little price inflation, so this approximation is unlikely to be serious problem. In a later paper we intend to exploit the information that some companies were able to give us on R&D inputs in units of labor hours to construct an index specifically for research costs in the pharmaceutical industry.

Construction of stock variables

Annual flows of discovery and expenditures were capitalized following the procedure described by Hall et al. (The R&D Masterfile: Documentation, NBER Technical WP #72). In brief, we first assume a depreciation rate for "knowledge capital" , δ, here equal to 20%. (This is consistent with previous studies, and as argued above is not going to be very important in terms of it's impact on the regression results since no matter what number we chose, if the flow series is reasonably smooth we would still find it difficult to identify δ separately from the estimated coefficient on the stock variable.) We then calculate a starting stock for each class within firm based on the first observation on the annual flow: assuming that real expenditures have been growing since minus infinity at a rate g, we divide the first observed year's flow by $\delta+g$. Each year, the end-of-year is set equal to the beginning-of-year stock net of depreciation, plus that year's flow. For the cases where the annual flow was missing "within" a series of observations, we set it equal to zero. In almost all instances, these missing values occur after

the expenditure flows have been declining towards zero: we are reasonably that these are "real" zeros and not missing data which should be interpolated. We used the same procedure to accumulate "stocks" of patents, based on the flow variables described above.

Definition of Variables Used in the Regression Analysis

Observations are identified by (encoded) FIRM, CLASS, and YEAR.

DISCOVERY	Expenditures by this firm in this research area, relating primarily to production of new compounds, by year, in millions of constant 1986 dollars.
own PATENTS	Important patents granted to this firm in this research area, by year, from the Derwent database. Note that throughout the analysis we count patents by year of *application*.
own PATENTS in related programs	Important patents granted to this firm in the related therapeutic class, net of the patents granted in this research area.
PATENTS in this program area by other firms:	Important patents granted this year in this research area to 29 other major multinational firms.
PATENTS in related programs by other firms:	Important patents granted in related therapeutic classes this year to 29 other major pharmaceutical firms.
SCOPE	The number of research areas in which this firm has spent at least $0.5m dollars on discovery this year.
SIZE	Total research expenditure by this firm in this year across all therapeutic classes.

Appendix Two: Therapeutic Class Definition

Class 10	Alimentary Tract and Metabolism
Class 20	Blood and Blood Forming Organs
Class 30	The Cardiovascular System
Class 40	Dermatologicals
Class 50	Genito-Urinary System and Sex Hormones
Class 60	General anti-infectives, systemic.
Class 90	Cytostatics
Class 100	Musculo-Skeletal System
Class 110	Central Nervous System
Class 120	Parasitology
Class 130	Respiratory System
Class 140	Sensory Organs
Class 150	Allergens and Immunomodulators
Class 160	All other therapeutics.

Figure (1):
Patenting rates per program over time

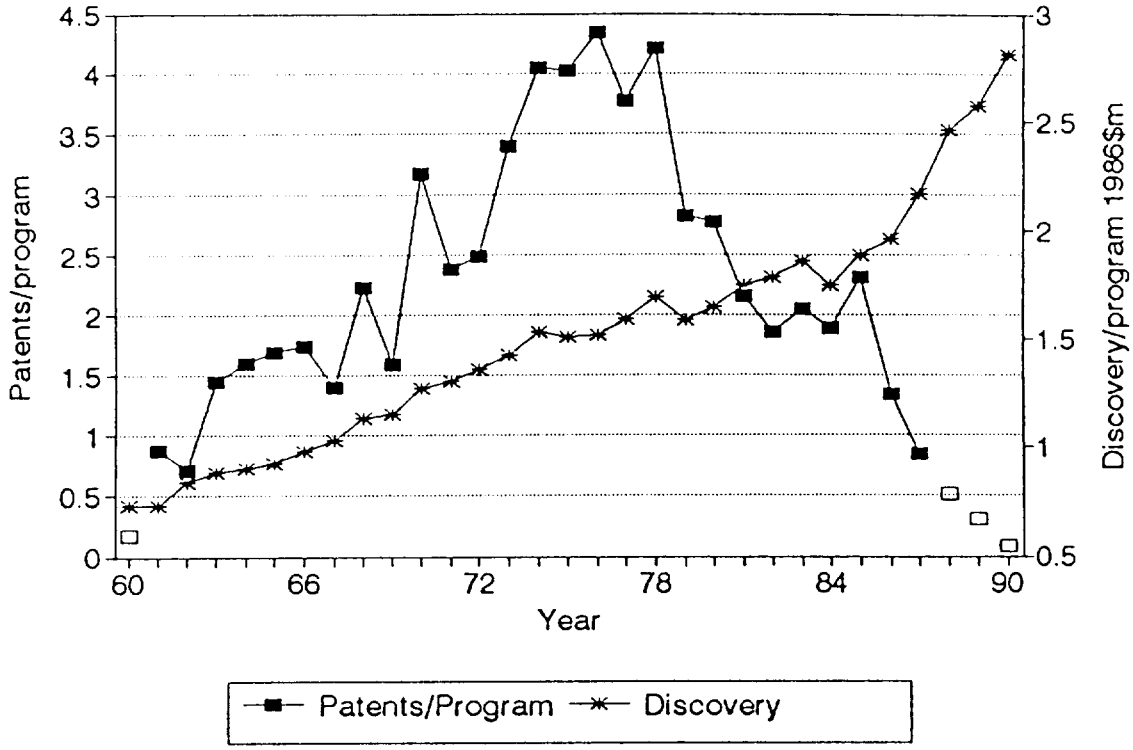

Figure (2):
The Evolution of Scope and Size

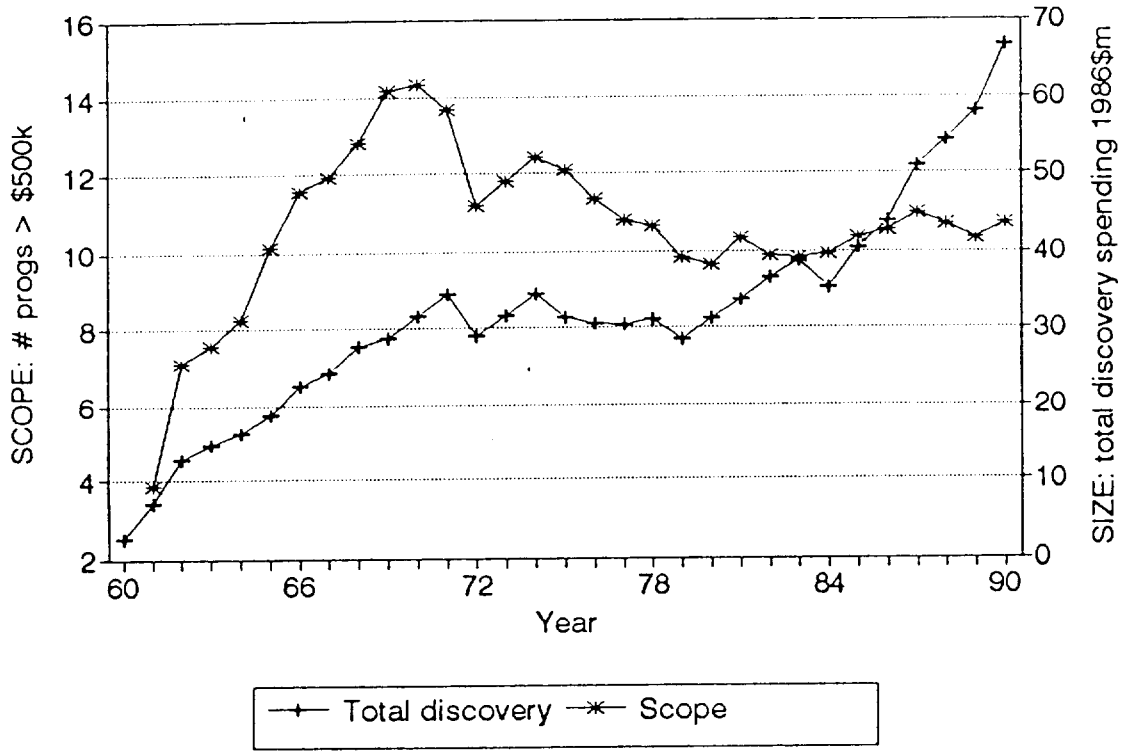

Figure (3)
Discovery/program, Frequency Distrib.

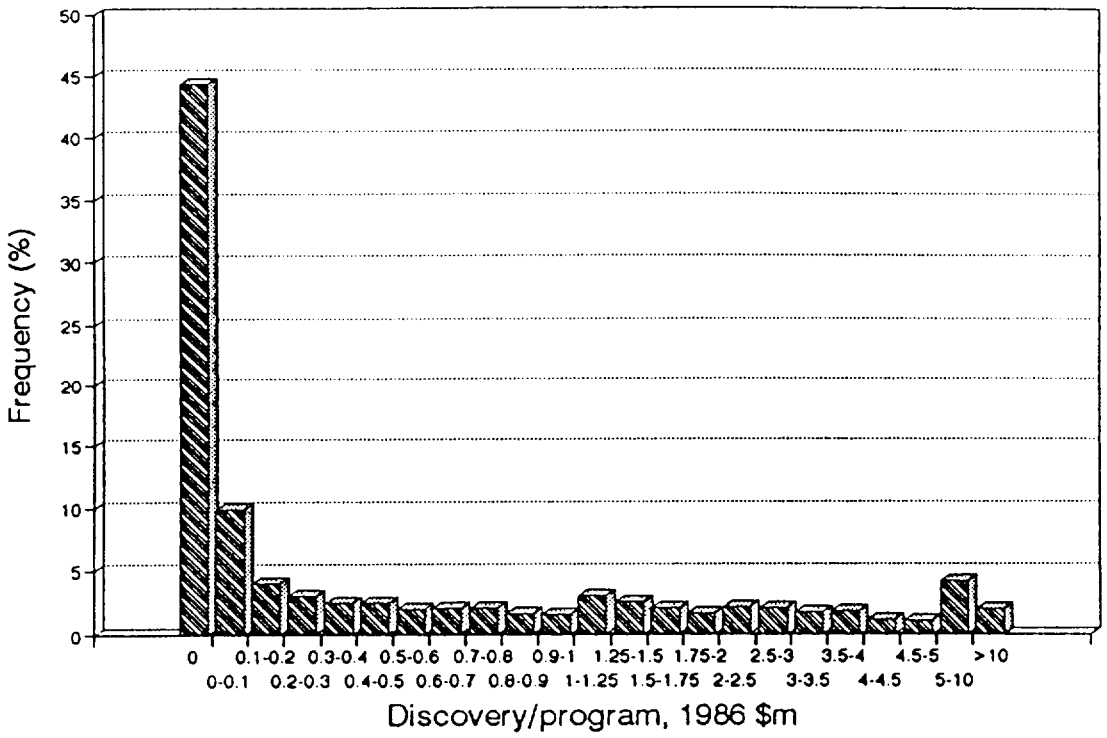

Figure (4)
Patents/program frequency distribution

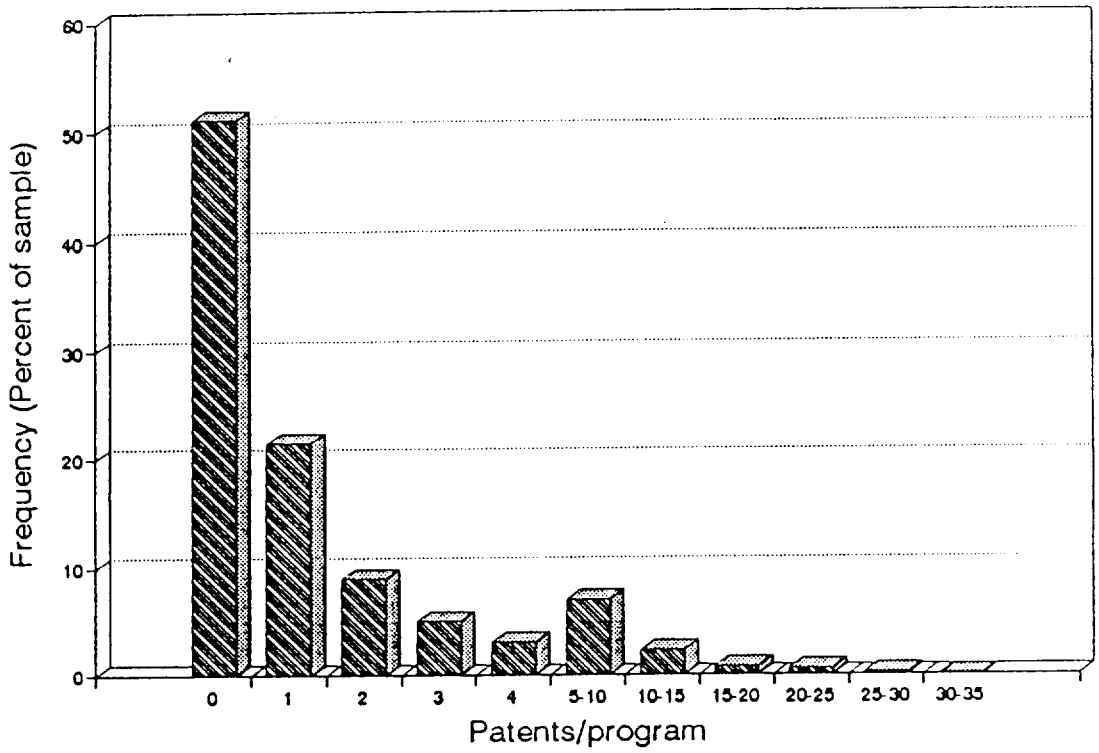

Table (1): Descriptive statistics: Selected variables at the research program level

Variable	Full Sample					Regression Sample N=4879
	N	Minimum	Maximum	Mean	Std Dev	Mean
Discovery, 1986$M	3089	0.00	>20	1.99	3.00	1.05
Stock of discovery	3089	0.00	64.11	5.33	8.15	2.92
Own discovery in related programs	3089	0.00	45.89	2.52	4.25	2.18
News in discovery in related classes	3089	-8.25	30.31	1.03	2.20	0.92
Own patents	5495	0.00	34.00	1.67	3.29	1.85
Stock of own patents	5495	0.00	96.18	6.12	10.70	6.54
Own patents in related programs	5495	0.00	60.00	3.30	6.03	3.57
News in own patents	5495	-13.60	33.97	0.76	3.66	0.96
Firm sales in program	5543	-2.03	889.59	19.72	74.66	19.77
SCOPE: Number of programs with disc>500K	5543	1.00	21.00	10.60	5.60	10.55
SIZE: Total discovery spending this year	5543	3.82	195.50	34.95	27.78	32.84
29 competitors' patents	4879	0.00	300.00	38.36	43.86	38.36
News in competitors' patents	4879	-91.84	128.26	10.93	19.71	10.93
Competitors patents in related programs	5486	0.00	353.00	107.99	81.55	116.80
News in competitors' patents in related classes	5486	-138.04	172.92	25.41	46.48	32.92

Table (2): Descriptive Statistics: Means of Selected Variables By Year

YEAR	N	DISCOVERY: Spending per program, 1986 $m	Own PATENTS per program.	SIZE: Total discovery spending by this firm, 1986 $m	SCOPE: Number of programs > $500K, 1986 $m	SALES Total pharma. sales, 1986 $m
61	83	0.42	0.99	7.08	3.9	1.99
62	115	0.61	0.90	12.81	7.1	4.81
63	115	0.69	1.30	14.60	7.6	5.41
64	115	0.73	1.39	16.24	8.2	6.36
65	120	0.77	1.44	18.72	10.1	7.52
66	124	0.87	1.47	22.61	11.6	8.74
67	126	0.96	1.28	24.36	12.0	10.27
68	123	1.14	1.74	27.69	12.8	11.80
69	128	1.18	1.38	28.72	14.1	12.73
70	120	1.40	2.26	31.65	14.3	14.77
71	117	1.45	1.82	34.57	13.7	16.49
72	163	1.55	1.88	29.14	11.2	18.14
73	173	1.67	2.39	31.60	11.8	19.35
74	175	1.86	2.75	34.49	12.5	20.24
75	196	1.82	2.73	31.38	12.1	20.19
76	196	1.83	2.91	30.58	11.4	20.69
77	198	1.97	2.59	30.57	10.8	20.91
78	200	2.15	2.84	31.14	10.7	22.17
79	232	1.96	2.06	28.42	9.9	22.04
80	224	2.07	2.04	31.20	9.6	23.02
81	222	2.25	1.70	33.81	10.3	24.39
82	241	2.32	1.53	36.50	9.9	23.39
83	251	2.45	1.64	38.78	9.8	22.62
84	284	2.25	1.55	35.32	9.9	21.76
85	278	2.50	1.79	40.39	10.4	22.57
86	275	2.63	1.25	43.84	10.5	24.64
87	262	3.01	0.97	50.93	11.0	23.94
88	251	3.53	0.51	54.42	10.7	25.45
89	195	3.73	0.30	58.16	10.3	29.70
90	193	4.16	0.08	66.72	10.7	31.36

Table (3): Determinants of patent output at the FIRM level.
Poisson Regression. Dependent variable = Total Firm PATENTS, 181 observations.

	(1)	(2)	(3)	(4)	(5)
Intercept	1.968** (0.101)	0.189 (0.132)	-0.283* (0.147)	-0.617 (0.157)	0.121** (0.159)
Ln(Total Firm Discovery)	0.084** (0.037)	0.154** (0.040)	0.330** (0.046)	0.288** (0.047)	0.273** (0.047)
Ln(Total Firm Stock of Discovery)	0.286** (0.033)	0.214** (0.049)	0.330** (0.053)	0.345** (0.053)	0.216** (0.054)
SCOPE: No. classes firm is active			-0.053** (0.007)	0.031** (0.015)	0.055** (0.016)
SCOPE * SCOPE				-0.004** (0.001)	-0.004** (0.001)
Total Firm Stock of own pats					0.0015** (0.0002)
Firm dummies	none	Sig.	Sig.	Sig.	Sig.
Time	0.155** (0.007)	0.153** (0.009)	0.156** (0.009)	0.162** (0.009)	0.118** (0.011)
Time * Time	-0.005** (0.001)	-0.006** (0.001)	-0.006** (0.001)	-0.006** (0.002)	-0.005** (0.001)
Log-likelihood	-2622.5	-1259.8	-1231.3	-1211.7	-1189.9

Standard errors in parentheses.
ln(variable) is set=0 when variable=0, and an appropriately coded dummy variable is included in the regression.

** Significant at the 1% level.
* Significant at the 5% level.

Table (4): Determinants of patent output at the research program level. Poisson Regression. Dependent variable = PATENTS, 4879 obs.

	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Intercept	-0.994* (0.053)	-1.681** (0.108)	-1.065** (0.066)	-2.623** (0.115)	-2.609** (0.115)	-3.191** (0.134)	-2.931** (0.128)	-3.765** (0.148)	-2.842** (0.149)
Ln(Discovery)	0.176** (0.015)	0.189** (0.011)	0.128** (0.010)	0.117** (0.010)	0.119** (0.010)	0.122** (0.010)	0.107** (0.010)	0.109** (0.010)	0.048** (0.010)
Ln(Stock of Discovery)	0.110 (0.009)	0.121** (0.009)	0.075** (0.009)	0.0845** (0.009)	0.082** (0.009)	0.085** (0.009)	0.089** (0.009)	0.093** (0.009)	0.035** (0.009)
SCOPE: No. classes firm is active					-0.013** (0.005)	0.114** (0.015)		0.074** (0.016)	0.112** (0.016)
SCOPE * SCOPE						-0.006** (0.001)		-0.005** (0.001)	-0.007** (0.001)
Ln(SIZE): Total disc. spending							0.175** (0.032)	0.369** (0.041)	0.295** (0.042)
Stock own pats in this class									0.035** (0.001)
Firm dummies	none	Sig.	none	Sig.	Sig.	Sig.	Sig.	Sig.	Sig.
Class dummies	none	none	Sig.	Sig.	Sig.	Sig.	Sig.	Sig.	Sig.
Time	0.163** (0.008)	0.153** (0.007)	0.165** (0.007)	0.161** (0.007)	0.169** (0.008)	0.176** (0.008)	0.148** (0.008)	0.169** (0.008)	0.105** (0.008)
Time * Time	-0.006 (0.000)	-0.005** (0.000)	-0.006** (0.000)	-0.005** (0.000)	-0.006** (0.000)	-0.006** (0.000)	-0.005** (0.000)	-0.006** (0.000)	-0.004** (0.000)
Log-likelihood	-12106.3	-10775.1	-10423.8	-9192.7	-9189.4	-9148.2	-9177.7	-9108.4	-8269.1

Standard errors in parentheses. ln(variable) is set=0 when variable=0, and an appropriately coded dummy variable is included in the regression. The omitted therapeutic class dummy is class60, systemic anti-infectives.

** Significant at the 1% level. * Significant at the 5% level.

Table (5): Determinants of patent output at the research program level.
Poisson Regression. Dependent variable = PATENTS, 4879 observations.

	(14)	(15)	(16)	(17)	(18)	(19)
Intercept	-2.842** (0.149)	-2.712** (0.149)	-2.805** (0.149)	-2.690** (0.149)	-2.951** (0.148)	-2.807** (0.149)
Ln(Discovery)	0.048** (0.010)	0.041** (0.010)	0.051** (0.010)	0.042** (0.010)	0.038** (0.009)	0.030** (0.010)
Ln(Stock of Discovery)	0.035** (0.009)	0.040** (0.009)	0.036** (0.009)	0.039** (0.009)	0.032** (0.009)	0.035** (0.009)
SCOPE: No. classes firm is active	0.112** (0.016)	0.106** (0.015)	0.111** (0.016)	0.106** (0.016)	0.110** (0.016)	0.105** (0.016)
SCOPE * SCOPE	-0.007** (0.001)	-0.006** (0.001)	-0.006** (0.001)	-0.006** (0.001)	-0.006** (0.001)	-0.006** (0.001)
Ln(SIZE): Total disc. spending by firm.	0.295** (0.042)	0.236** (0.042)	0.282** (0.041)	0.233** (0.042)	0.308** (0.042)	0.244** (0.042)
Stock own pats in this class	0.035** (0.001)	0.032** (0.001)	0.033** (0.001)	0.033** (0.001)	0.031** (0.001)	0.032** (0.001)
News in patents in related classes		0.032** (0.002)		0.032** (0.002)		0.033** (0.003)
News in discovery in related classes			0.025* (0.013)	0.003 (0.012)		
News in competitors' patents in this class					0.005** (0.001)	0.007** (0.001)
News in competitors' patents in related classes					0.003** (0.001)	0.002** (0.000)
Firm dummies	Sig.	Sig.	Sig.	Sig.	Sig.	Sig.
Class dummies	Sig.	Sig.	Sig.	Sig.	Sig.	Sig.
Time	0.105** (0.008)	0.105** (0.008)	0.105** (0.008)	0.106** (0.008)	0.066** (0.008)	0.069** (0.010)
Time * Time	-0.004** (0.000)	-0.004** (0.000)	-0.005** (0.000)	-0.004** (0.000)	-0.002** (0.000)	-0.003** (0.001)
Log-likelihood	-8269.1	-8192.0	-8259.1	-8183.9	-8105.1	-8026.9

Standard errors in parentheses. ln(variable) is set=0 when variable=0, and an appropriately coded dummy variable is included in the regression. The omitted therapeutic class dummy is class60, systemic anti-infectives.
** Significant at the 1% level.
* Significant at the 5% level.

Table (5a): Alternate Models
Dependent variable = PATENTS, 4879 observations.

	Poisson (19)	Negative Binomial (20)	Non-Linear Least Squares (21)	GMT (22)
Intercept	-2.807** (0.149)	-2.693** (0.202)	-2.267** (0.313)	-2.113** (0.234)
Ln(Discovery)	0.030** (0.010)	0.061** (0.015)	0.011 (0.021)	0.076** (0.019)
Ln(Stock of Discovery)	0.035** (0.009)	0.039** (0.014)	0.008 (0.022)	0.053* (0.018)
SCOPE: No. classes firm is active	0.105** (0.016)	0.104** (0.023)	0.085* (0.037)	0.049 (0.028)
SCOPE * SCOPE	-0.006** (0.001)	-0.006** (0.001)	-0.004** (0.002)	-0.004** (0.001)
Ln(SIZE): Total disc. spending by firm.	0.244** (0.042)	0.131* (0.069)	0.298** (0.085)	0.230** (0.095)
Stock own pats in this class	0.032** (0.001)	0.052** (0.002)	0.027** (0.002)	0.057** (0.026)
News in patents in related classes	0.033** (0.003)	0.041** (0.005)	0.022** (0.005)	0.069** (0.008)
News in competitors' patents in this class	0.007** (0.001)	0.012** (0.001)	0.003 (0.002)	0.011** (0.001)
News in competitors' patents in related classes	0.002** (0.000)	0.001 (0.001)	0.003** (0.001)	-0.002 (0.001)
Firm dummies	Sig.	Sig.	Sig.	Sig.
Class dummies	Sig.	Sig.	Sig.	Sig.
Time	0.069** (0.010)	0.084** (0.013)	0.020 (0.019)	0.071** (0.016)
Time * Time	-0.003** (0.001)	-0.003** (0.000)	-0.002** (0.001)	-0.002** (0.001)
overdispersion parameter	N/A	0.518 (0.003)	N/A	0.409
Log-likelihood		-7131.61	R^2=0.582 SER=2.240	SER=.905

Notes: see over.

Table 5a Notes: Poisson model as in column (19) of Table (5).
Negative Binomial variance modelled as $VAR(Y)=E[X](1+\alpha E[X])$
Non-linear Least Squares: $Y=\exp(X\beta)+\epsilon$
GMT: weighted non-linear least squares, with weights derived from Poisson estimates of β in column (19): $w_i=\exp(X\hat{\beta})+\hat{\eta}^2\exp(X\hat{\beta})$

Standard errors in parentheses. ln(variable) is set=0 when variable=0, and an appropriately coded dummy variable is included in the regression. The omitted therapeutic class dummy is class60, systemic anti-infectives.

** Significant at the 1% level.
* Significant at the 5% level.