ELSEVIER

Journal of Health Economics 20 (2001) 1033–1057

JOURNAL OF HEALTH ECONOMICS

www.elsevier.com/locate/econbase

Scale and scope in drug development: unpacking the advantages of size in pharmaceutical research

Iain M. Cockburn [a,*], Rebecca M. Henderson [b]

[a] *School of Management, Boston University and NBER, 595 Commonwealth Avenue, Boston, MA 02215, USA*
[b] *Massachusetts Institute of Technology and NBER, Boston, MA 02215, USA*

Received 1 October 1996; accepted 1 August 2001

Abstract

Drug development performance is examined using data on clinical research projects of 10 pharmaceutical companies. In contrast to previous work on the discovery phase of pharmaceutical R&D we find a strong correlation between the diversity of firms' development efforts and the success probability of individual projects, but no effect of scale per se. Large firms' superior performance in drug development appears to be driven by returns to scope rather than returns to scale. Scope is confounded with firm fixed effects, however, suggesting an important role for inter-firm differences in the organization and management of the development function.

JEL classification: O3; L2; L6

Keywords: Pharmaceuticals; R&D; Clinical; Economies of scale

1. Introduction

Economies of scale and scope in R&D are an important determinant of the economic performance of the pharmaceutical industry. The presence and magnitude of these effects have significant implications for returns to R&D as well as for the evolution of industry structure and its impact on welfare, but they are poorly understood. Though much of the literature has focused on economies of scale and scope in manufacturing and production, R&D activities have been identified as a particularly important source of advantages accruing to size and diversity. A range of theoretical arguments revolving around problems in the market for knowledge suggest that research or the production of new knowledge should be subject to considerable economies of scope (see Arrow, 1962; Teece, 1980). At

* Corresponding author. Tel.: +1-617-353-3775.
E-mail address: cockburn@nber.org (I.M. Cockburn).

0167-6296/01/$ – see front matter
PII: S0167-6296(01)00108-4

the same time qualitative work, such as Chandler (1990) and Freeman (1982), suggests that the organization of modern industrial R&D is such that the research process is likely to have significant economies of scale. Nonetheless, as Cohen and Levin (1989) and others have noted, despite the importance of these issues for the theory of the firm and for welfare analysis of mergers and market structure, the empirical work in this area has been largely inconclusive.

Arguably, the inconclusive or contradictory results of much of the previous research on this topic stems from the use of inappropriately aggregated firm level data. A large portion of observed variation in research productivity is likely to reflect differences in technological opportunity across research areas, but since most firms conduct R&D in a variety of areas it is very difficult to control for these effects at the level of the firm. Furthermore, while theory draws a clear distinction between economies of scope and economies of scale, firm level data cannot speak to the presence of economies of scope which derive from the diversity of activities *within* the firm, leading to confounding between these two effects. Using project level data derived from a detailed study of research in the pharmaceutical industry we showed in Henderson and Cockburn (1996) that, controlling for technological opportunity conditions, comparable research efforts located in larger firms were significantly more productive than rival programs located in smaller firms. In more recent years, however, larger firms seemed to benefit as much from economies of scope as from economies of scale per se.

This study left a number of important questions unaddressed. Most significantly, perhaps, the results addressed only the presence of economies of scale and scope in *research*, or "discovery", not in *development*, the translation of potentially fruitful new ideas into products that can be sold and marketed. This is an important issue since spending on development often greatly exceeds spending on research, and since the presence of economies of scale and scope in development activity are often advanced as a primary justification for both vertical and horizontal integration. In the case of pharmaceuticals, for example, economies of scale and scope in development are frequently cited in the business press as an important force driving the recent wave of industry mergers. Consolidation of the industry presents a challenge for antitrust and competition policy authorities: potential static welfare losses associated with increasing industry concentration must be weighed against dynamic welfare gains from increased research productivity. Since some firms are widening their research portfolios through mergers while others are deepening them, assessing these gains is difficult without a better understanding of the relative roles of scope and scale effects.

This paper uses disaggregated data at the project level to explore the degree to which the productivity of drug development is related to the scale or scope of a firm's development effort. We begin with a brief literature review and then turn to a qualitative discussion of the nature of pharmaceutical development as background for hypothesis development. Section 3 discusses the data and estimation techniques, Section 4 presents the results, and the paper closes with a brief discussion of their implications and of directions for further research. Although we find no evidence of returns to size per se, our results suggest that there are significant effects of scope on the likelihood of success of drug development projects. Intriguingly, while our earlier findings were robust to controlling for firm effects, the results reported here are not: productivity in the development phase of pharmaceutical research appears to be driven by other aspects of the organization and management of the firm.

2. Literature review and hypothesis development

2.1. Scale, scope, and research productivity

In principle size confers three major advantages in performing R&D. First, 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. Second, large firms may 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. Lastly, larger firms may be able to exploit economies of scale and scope in the conduct of research itself (see Panzar and Willig, 1981; Cohen and Levin, 1989; Schumpeter, 1950).

Early tests of these ideas attempted to establish a relationship between firm size and R&D intensity, and were largely inconclusive (see Baldwin and Scott, 1987; Panzar, 1989; Cohen and Levin, 1989). 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 constant returns to scale for R&D programs between $2 and $100 million when patents were used as a measure of research output, but their results were quite sensitive to specification assumptions. Acs and Audretsch (1988), using a count 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 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. Similarly, although the presence of economies of scope in production have been verified in many industries, including airlines, banking and advertising (see Caves et al., 1984; Glass and McKillop, 1992; Silk and Berndt, 1993, respectively), a lack of appropriately detailed data has meant that with some notable exceptions (such as Helfat, 1997), we have very little empirical evidence as to their importance in R&D.

Prior research on the pharmaceutical industry has faced many of the same problems. Use of aggregate data has not permitted researchers to make the important distinction between drug discovery and drug development, and in general the results reported in the literature have been inconclusive. Comanor (1965), Vernon and Gusen (1974), and Graves and Langowitz (1993) found evidence for decreasing returns to scale in R&D, while Schwartzman (1976) suggested that there were significant economies of scale in pharmaceutical research, and Jensen (1987) found that beyond a quite small threshold neither firm size nor the size of the research effort affected the marginal productivity of R&D.

2.2. Hypothesis formulation: the process of drug development

The process of searching for new drugs takes place in two distinct stages: (1) discovery and (2) development. The goal of the discovery phase is to find a novel compound that demonstrates some desirable effect in either an animal or chemical screen. For example, firms search for compounds that can make obese rats thinner or that can block the action of an enzyme that is known to regulate metabolic activity. In contrast, the goal of the development process is to explore the degree to which a particularly promising compound

is safe and effective in humans. Firms wishing to introduce drugs in United States must demonstrate both safety and efficacy to the FDA by providing data gathered from extensive clinical trials, and similar requirements are placed in many other countries.

The FDA requires potential compounds to go through four phases: a "preclinical" phase, in which the compound is extensively tested for toxicity in animals; phase I in which the compound is tested in healthy human volunteers to ensure that it is safe; phase II in which it is tested in a small group of patients to establish efficacy; and phase III in which the compound is tested in very much larger numbers of more diverse patients to establish both safety and efficacy in a representative patient group. In general, these tests are very costly and take a long time to complete. Fig. 1 shows mean development costs for the 10 firms in our sample over time, both in constant 1998 dollars and as a percentage of total spending on R&D. Average spending in constant dollars on development increased from around $40 million in the middle sixties to well over $200 million in 1990. The average duration of projects (both successful and successful) was just under 5 years. These projects are also very risky: on average only one in five of the compounds that began substantial clinical testing in our data resulted in the filing of an application for new drug approval (NDA), and even fewer were granted an NDA and reached the marketplace.

The process of clinical development calls on a wide range of quite diverse skills, from clinical pharmacology to biostatistics and metabolic chemistry. These specialists are required to design trials that can yield statistically significant results, to identify critical "endpoints" in the progress of a disease that can be used to monitor the success of a new compound, and to administer and monitor clinical trials that often enlist thousands of patients in tens or hundreds of clinical centers located across the globe. Effective development also requires the participation of other scientists and technicians whose focus is on the manufacture and delivery of the new compound, including packaging experts, process chemists and operations engineers.

For diseases that are not well understood, the design and conduct of clinical trials may require significant investments in research about the progress and stages of the disease. For example, the design of clinical trials in the case of a compound designed to treat a relatively well understood condition such as hypertension is relatively straightforward, since techniques for monitoring blood pressure are well established and reductions in measured blood pressure are known to be linked directly to significant increases in patient well being. But in the case of a compound such as finasteride (Proscar), which was developed to treat benign prostatic hypertrophy, the company running the original clinical trials not only had to find methods for measuring progress in the treatment of the condition, but also had to demonstrate that these measures were linked to real improvements in patients' health status.

The nature of the drug development process thus suggests a number of reasons why it may be characterized by returns to scale and scope. Returns to scale are present when the costs of doing any single activity can be spread out over a larger activity base, or when undertaking an activity on a larger scale permits the adoption of more effective techniques. Since drug development is a complex, multifaceted activity requiring the coordination of a wide range of specialist expertise, both sources of economies of scale may well be present. On the one hand, a firm may be able to amortize the costs of "generic" specialized expertise that can be applied to a wide variety of diseases and large numbers of development candidates. Expertise in biostatistics, or in dealing effectively with regulatory authorities across the world may

Fig. 1. Trends in development spending: (—) mean development spending per firm; (- - -) mean share of development in total R&D.

have this property, for example. On the other hand, as the size of its development effort grows a firm may be able to afford to employ significantly more specialized resources. For example, a small firm may be forced to employ a small number of generalist clinicians to supervise all of its clinical trials while a larger firm can invest in employing specialized cardiologists, oncologists and so on to supervise trials of potential heart, cancer and other treatments. Similarly, smaller development programs may be forced to use general purpose software to monitor the progress of their trials, while clinical programs conducted on a larger scale may be able to invest in specialist software that is particularly tailored to the needs and experience of the firm. Thus we hypothesize the following.

H1. All else equal, larger development efforts are more productive.

Returns to scope arise when two or more activities are more efficient when conducted together within one organization instead of separately. For example, a firm may be able to apply knowledge generated within one research area to another project within the firm at a low or zero marginal cost. Here again, there are two important reasons for believing that development activities might demonstrate returns to scope. In the first place, larger firms may be able to efficiently transfer general knowledge about the efficient running of clinical trials across different projects within the firm. Large scale clinical trials are both costly and complex, requiring the tracking and integration of complex data across thousands of patients. Expertise gleaned from running large scale trials in, for example, depression, may reduce the costs of running large scale trials in hypertension or arthritis. In the second place, since few systems in the human body operate in complete isolation, particular medical knowledge gained through work in one area may be useful to another. For example, knowledge gained through work with anti-anxiety drugs may be of direct relevance to the conduct of trials of anti-depressants, while work in hypertension may be of immediate relevance to work in arrhythmia. Thus we hypothesize the following.

H2. All else equal, more diverse development efforts are more productive.

However, there are also a number of reasons for hypothesizing that pharmaceutical development may *not* be characterized by returns to scope or scale. Economies of scale arising from the sharing of a fixed cost, for example, arise only if the cost is truly fixed. While some specialist skills such as biostatistics can be applied to multiple disease conditions, the diversity of compounds and diseases is such that clinical trials may be highly idiosyncratic, and the design of each new trial may require a roughly equivalent input of expertise in biostatistics. We note also that the ability to realize efficiencies arising from the ability to employ specialized resources may only be related to the scale of the research effort when these resources cannot be traded in the open market. Some significant fraction of clinical trials are outsourced, and there is an extensive infrastructure of specialist firms whose skills can be accessed, in principle, by any firm making use of their services.

Equally, returns to scope in the management of clinical trials driven by knowledge externalities are likely to be present in firms only if this knowledge is "sticky", or not instantaneously transmitted across the boundaries of the firm. While this is a plausible assumption in the case of pharmaceutical discovery, it is much less so in the case of pharmaceutical

development where responsibility for the conduct of the actual trials themselves remains largely in the hands of physicians operating outside the firm. Trials are designed and paid for by the individual pharmaceutical firms but, historically, the vast majority are carried out at hospitals by physicians under contract. Thus knowledge about the conduct of clinical trials tends to be fairly rapidly diffused across the medical community. This is particularly likely to be true in the case of those diseases that are less well understood, since preliminary trials of compounds designed to address these conditions are often carried out under the supervision of a small set of leading physicians in the field who are well known to each other and who publish freely.

The presence or absence of scale and scope effects upon the productivity of drug development is thus an open question, and in the remainder of the paper we draw on project level data gathered from within 10 major pharmaceutical firms to test for them empirically.

3. Sources and construction of the dataset

The dataset used here comprises data on R&D inputs and outputs at the research project level obtained from the internal records of 10 pharmaceutical firms. Although for reasons of confidentiality we cannot describe these firms, we can say that they cover the range of major R&D-performing pharmaceutical manufacturers, that they include both American and European firms, and that we believe that they are not markedly unrepresentative of the industry in terms of size or technical or commercial performance. Between them they represent about 25% of the pharmaceutical research conducted world-wide. This section offers a brief description of the construction the dataset, the variables used in the econometric analysis, and presents descriptive statistics for the data. Appendix A gives fuller details of the construction of the data base.

Our fundamental unit of observation is the "development project", defined as activity directed towards the testing and improvement of a candidate compound for human therapeutic use. This activity includes effort expended within the firm, as well as grants and other payments to hospitals and universities for conducting clinical trials. Assembling the data on these projects in a consistent and meaningful format was a considerable challenge. In nearly every case the process of data collection was an iterative one, involving close collaboration with key personnel from the companies participating in the study. The majority of the data were collected specifically for the purposes of this study. Each firm spent some months assembling its data, usually from primary documents, and the full data collection effort took nearly 2 years. As far as it was possible, definitions of research programs and of expense grouping were standardized across firms, data were collected at the same level of aggregation, and overhead expenses were treated in a consistent way, and steps were taken to ensure that, wherever possible, the data included world-wide research spending, not just US facilities. For the most part "projects" were clearly defined in the primary data sources in terms of a specific candidate compound, but in some cases, companies were only able to provide expenditure data grouped by therapeutic class, and expenditures were allocated to projects/compounds by pro-rating the total.

In total we observe 708 development projects, beginning as early as 1960 and ending (or being censored) in 1990. Unfortunately we were not able to comprehensively date the

movement of every project through the various phases of clinical development. While some firms were able to give us precise dates at which each compound entered each phase of clinical development, for most firms and for older projects we had to rely on regulatory filings which track the movement of the drug candidate through the development process. A variety of project outcomes are observed. In principle projects are either: (1) terminated before any regulatory filings; (2) generate an IND application (i.e. a request for permission to use the compound in human subjects); (3) receive IND approval; (4) generate an application for an NDA (permission to market the drug); or (5) receive NDA approval. Some projects were discarded from our working sample because of unresolvable difficulties in establishing dates. For example, in some cases all or some of the development effort took place outside US, while in other cases projects are licensed in or are conducted jointly with a partner firm, making their progress through the development process very hard to follow accurately. Along with the outcome of the project, we record its LENGTH, i.e. duration in years, total expenditure over the lifetime of the project (in constant dollars) and RATE, the average rate of expenditure per year over the lifetime of the project. For each project we also compute variables describing the nature of the firm's drug development program as a whole at the time the project began: SCALE which is the firm's total expenditure on development across all research areas, and SCOPE which is a count of the number of development programs in which the firm is spending more than $1 million 1998 dollars. Finally, as proxies for accumulated knowledge capital (or unobserved input quality) in the relevant therapeutic area, we compute a stock of previously obtained NDAs in the therapeutic class, depreciated at 20% per year.

Note that these projects do not constitute the entire development effort of the firm. In principle we observe the entire "population" of projects conducted by these firms, both successful and unsuccessful, but our dataset is not quite complete. Despite our best efforts and the cooperation of the firms participating in this study we were unable to successfully identify the compounds associated with some streams of expenditure, and thus could not accurately determine the outcome of the project. Along with a variety of other data problems this meant that some of the data was discarded. On average, therefore, the projects in our sample account for about 85% of the total development spending of these 10 firms.

Since our dataset ends before the outcome of some projects is determined, there is some censoring of the data. In most of our econometric analysis we confine our attention to the sub-sample of projects where we observe outcomes, inducing bias to the extent that we overweight shorter duration projects. However, results of exploratory calculations which adjust for censoring indicate that this is not a significant problem.

4. Results

4.1. Descriptive statistics

Tables 1–5 and Figs. 1–4 present summary statistics describing these data. The range and diversity of these firms' drug development efforts is quite striking. On average firms ran just under 16 identifiable projects simultaneously, started between and four and five new projects every year, and obtained 1.26 NDAs per year. The average firm had significant development activity (spending over $1 million 1998 dollars) in over 14 major therapeutic

Table 1
Descriptive statistics[a]

Variable	Minimum	Mean	Maximum	S.D.
Length of project	1	4.83	26	4.83
Average rate of development spending per project per year	0.01	2.57	>50	5.37
Total spending per project[b]	0.01	18.42	>450	47.26
Total firm development spending per year	4.67	82.57	>400	69.68
NDAs per firm per year	0	1.26	7	1.67
Number of active projects per year	1	15.92	42	10.14
Number of new projects started per year	1	4.43	12	4.08
Number of therapeutic classes with spending >$1 million	0	9.63	27	7.13

[a] $N = 708$ development projects. Spending in millions of constant 1998 dollars.
[b] Does not adjust for uncompleted projects, and does not include cost of capital.

areas per year, and was active in slightly more than 28 different therapeutic areas over the time period covered by our data. On average development projects lasted just under 5 years, spending $2.58 million 1998 dollars per year.

The intensity of development spending per project varies considerably, from $10,000 per year to over $50 million 1998 dollars per year. Fig. 2 gives the frequency distribution of the average rate of development expenditure per year over the lifetime of a project. The distribution is highly left-skewed: over 50% of the projects we identified spent less than $1 million 1998 dollars per year. At the other tail of the distribution, there are a small number of "blockbuster" projects spending at an average rate of over 40 million 1998 dollars per year. These averages also conceal significant additional variation in annual spending levels *within* projects. The same patterns are apparent in total expenditures over the lifetime of a project: as Fig. 3 shows, the distribution is highly skewed to the left, and covers a very wide range. Total project expenditures range from $10,000 to over $400 million 1998 dollars. It is important to note that distortions are introduced at the left tail of the distribution by censoring of projects at the end of our sample period, and these totals do not include any allowance for the opportunity cost of capital, so that it is difficult to infer anything from these data about the average cost of drug development.

Average rates of expenditure per project per year trended up slightly over the 30 years covered by our sample, and by the end of the 1980s they were roughly double their levels in the 1960s in real terms. Interestingly, as Table 4 illustrates and the ANOVA results in Table 5 confirm, there is relatively little variation in expenditure intensity or total expenditures across therapeutic classes. Much more variation is apparent across firms than across different drug

Table 2
Development project outcomes

Project outcome	Frequency	Percentage
NDA after IND	121	17.1
NDA, no IND	25	3.5
Terminated after IND	170	24.0
Terminated, no IND	269	37.9
Censored	123	17.5

Table 3
Summary statistics by year

Year project begins	Number of projects	Mean RATE of development spending per project per year ($1 million 1998 dollars)	Mean TOTAL development spending per project ($1 million 1998 dollars)	Mean LENGTH of projects (years)
61	15	0.52	1.05	1.93
62	19	0.53	4.17	5.21
63	34	0.76	2.67	3.26
64	14	0.77	10.30	3.57
65	19	1.45	9.94	4.89
66	7	0.66	3.50	2.43
67	18	1.27	10.98	4.22
68	8	3.64	30.49	4.75
69	10	1.29	8.03	3.10
70	14	1.57	9.01	3.29
71	11	0.47	3.71	6.64
72	25	1.39	17.59	8.96
73	13	1.50	14.29	7.69
74	17	5.27	53.16	9.53
75	17	2.15	21.39	8.18
76	13	2.72	27.55	5.46
77	23	0.71	4.72	4.09
78	24	4.27	42.87	8.50
79	17	0.91	6.77	4.71
80	15	1.78	9.32	5.13
81	37	3.71	32.46	4.39
82	73	2.43	17.48	5.03
83	37	2.87	17.44	4.65
84	37	2.59	15.63	4.43
85	47	5.42	32.48	3.24
86	32	2.81	14.73	3.55
87	42	3.77	14.18	3.00
88	28	7.09	18.81	2.14
89	15	3.32	9.05	2.23
90	15	3.07	6.04	1.08

classes. This stands in sharp contrast to our findings in previous work on drug discovery, where there were very substantial differences in research expenditures across different therapeutic classes as well as across firms.

We observe a similarly wide range of project durations, from 1 to 26 years. Again, as Fig. 4 shows, the frequency distribution is highly left-skewed, with 90% of all projects lasting 10 years or less. The censoring of project durations caused by the end of the sample period in 1990 is visible in the last column of Table 3 where mean project lengths fall substantially in the late 1980s. Fig. 5 presents the results of estimating a life table for the projects by the standard Kaplan–Meier method with an adjustment for censoring. Adjustment for censoring makes little difference except for lowering the estimated hazard rate at the left end of the scale. Note also that Fig. 5 makes no distinction between successful and unsuccessful termination modes for projects since we cannot determine the outcome of censored projects. Exploratory analysis using a restricted sample of projects started in the 1960s and 1970s

Table 4
Summary statistics by therapeutic class

Therapeutic class	Number of projects	Mean RATE of development spending per project per year ($1 million 1998 dollars)	Mean TOTAL development spending per project ($1 million 1998 dollars)	Mean LENGTH of project (years)
Alimentary tract and metabolism	95	1.99	12.53	4.54
Hematologicals	41	4.99	28.76	4.32
Cardiac and circulatory	135	3.01	22.16	4.52
Dermatologicals	17	1.53	11.34	5.12
Genito-urinary system	1	1.03	5.15	5.00
Systemic anti-infectives	121	3.78	27.24	5.97
Hormones	23	2.06	11.87	5.10
Cytostatics	36	1.27	13.01	5.83
Musculo-skeletal system	57	2.45	15.63	4.12
Central nervous system	100	1.86	17.87	5.21
Anti-parasiticals	8	1.73	11.49	7.00
Respiratory system	47	2.81	10.88	3.64
Sensory organs	8	3.45	23.90	3.88
Immunotherapeutics	5	0.71	6.17	3.40
Ethical therapeutics NEC	6	4.74	80.98	8.50

shows little difference in the estimated hazard functions for successful versus unsuccessful projects. Projects which fail to generate an NDA exit the sample a little earlier than those which are successful, but in both cases the hazard function shows a distinct peak at about 7 years duration. As for expenditure intensities, there are very substantial differences across firms in project durations, and relatively little across therapeutic classes. This is most easily seen in the ANOVA results in Table 5 (which do not control for censoring) but are also

Table 5
ANOVA results

Variable	Within mean square (d.f.)	Between mean square (d.f.)	$F\ (P > F)$
Firm effects			
RATE of development expenditure per project per year	9.17 (704)	146.35 (9)	15.95 (0.001)
TOTAL development expenditure per project	727.6 (708)	10335.4 (9)	14.21 (0.001)
LENGTH of projects	15.69 (708)	622.9 (9)	39.69 (0.001)
Therapeutic class effects			
RATE of development expenditure per project per year	11.31 (659)	17.05 (14)	1.51 (0.102)
TOTAL development expenditure per project	877.6 (663)	1330.9 (14)	1.52 (0.099)
LENGTH of projects	24.15 (663)	32.23 (14)	1.33 (0.181)

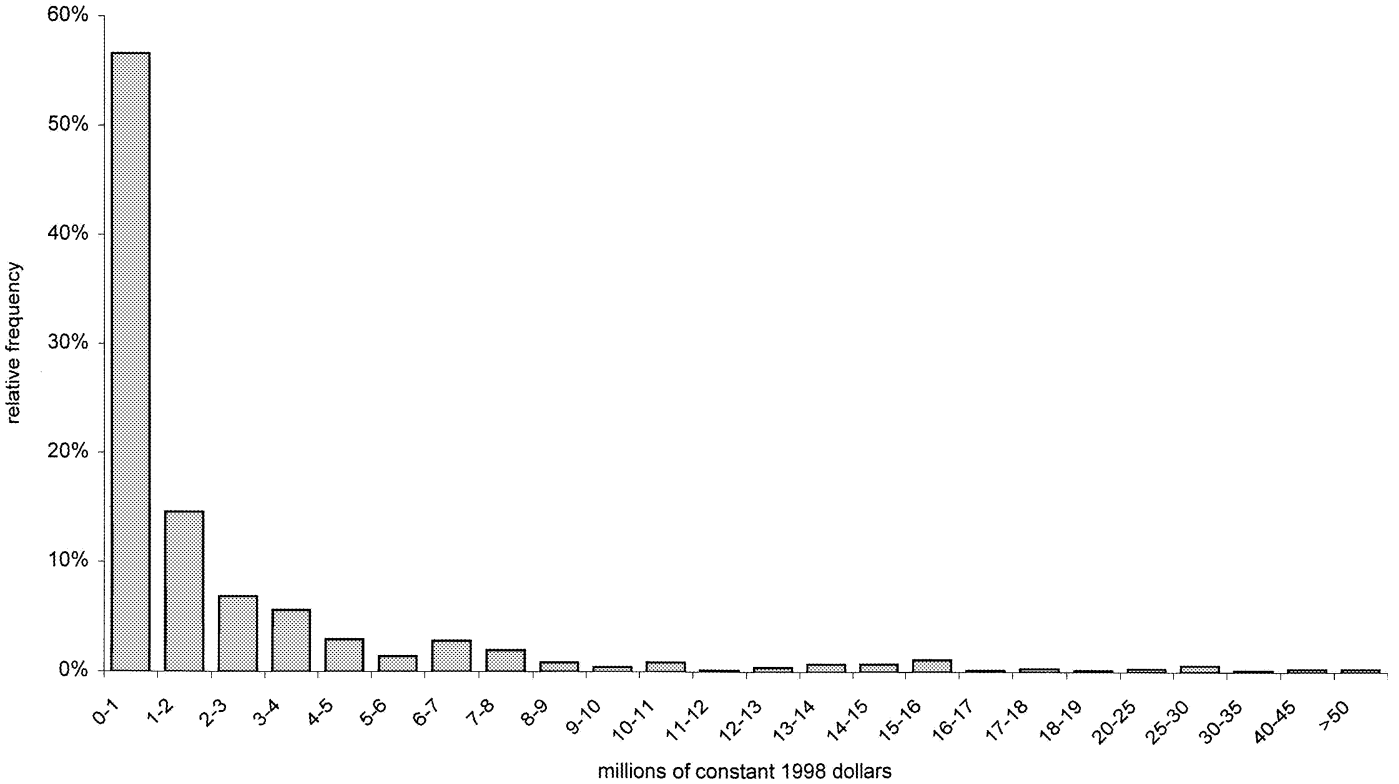

Fig. 2. Frequency distribution of annual rate of expenditure per project.

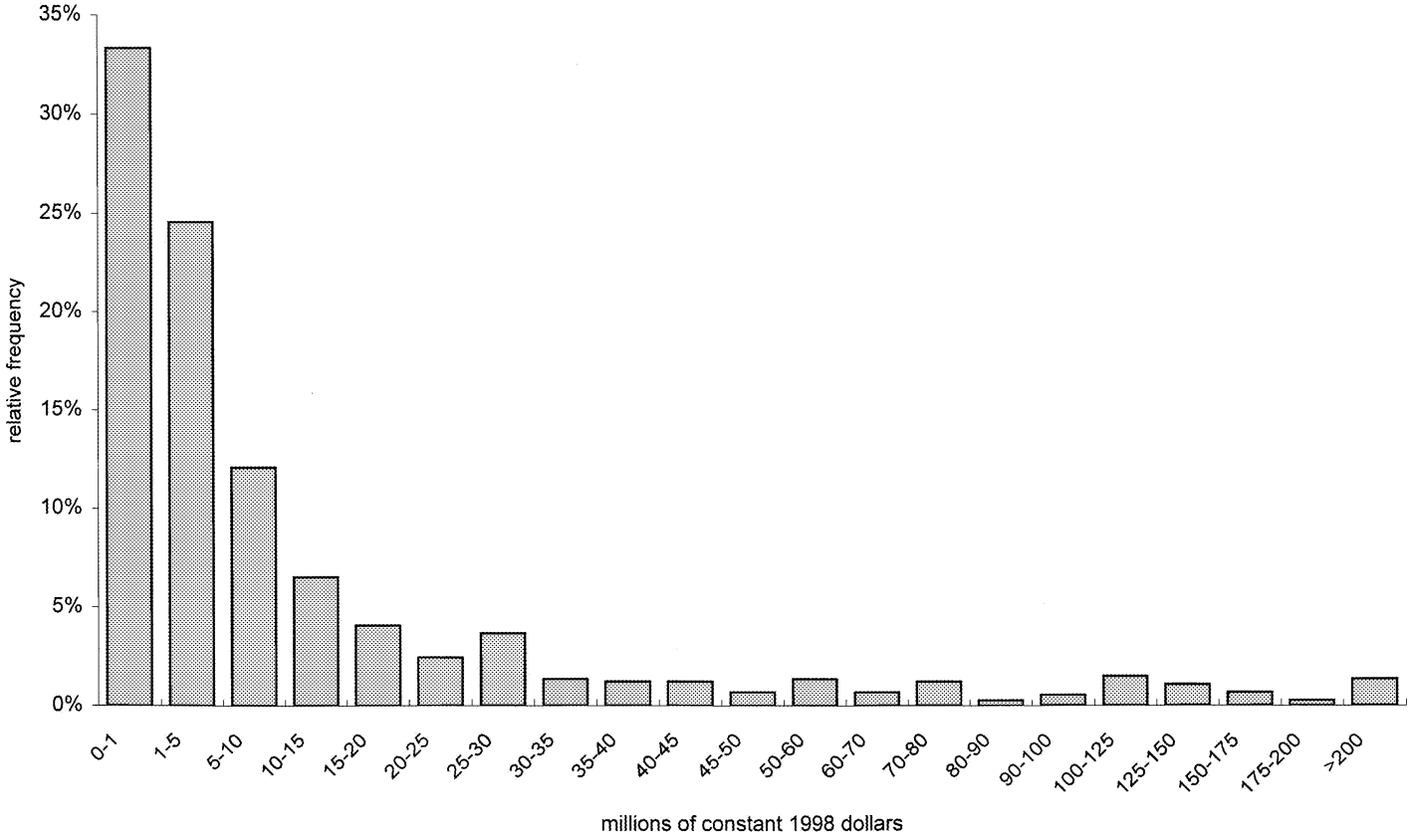

Fig. 3. Frequency of total expenditure per project.

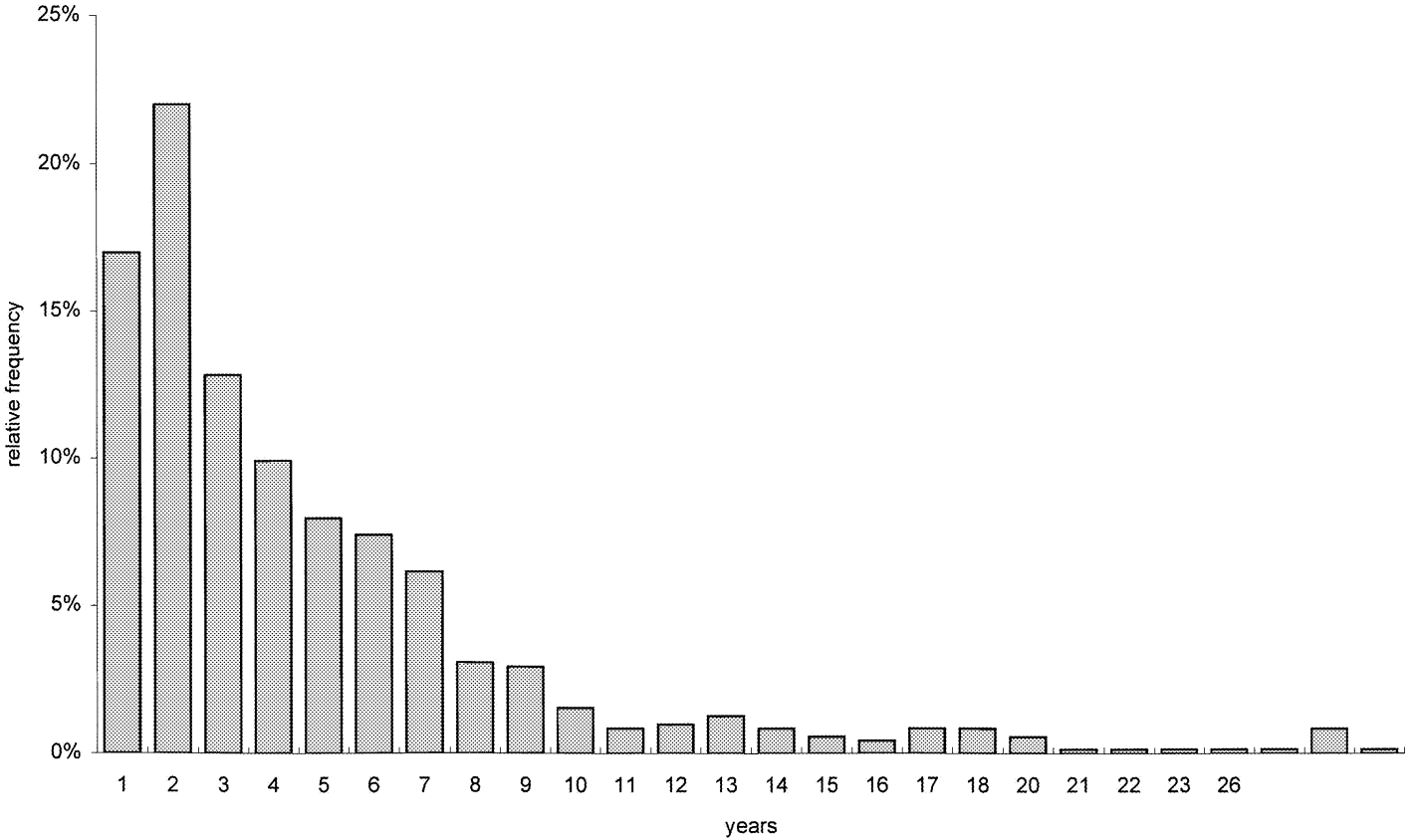

Fig. 4. Frequency distribution of project durations.

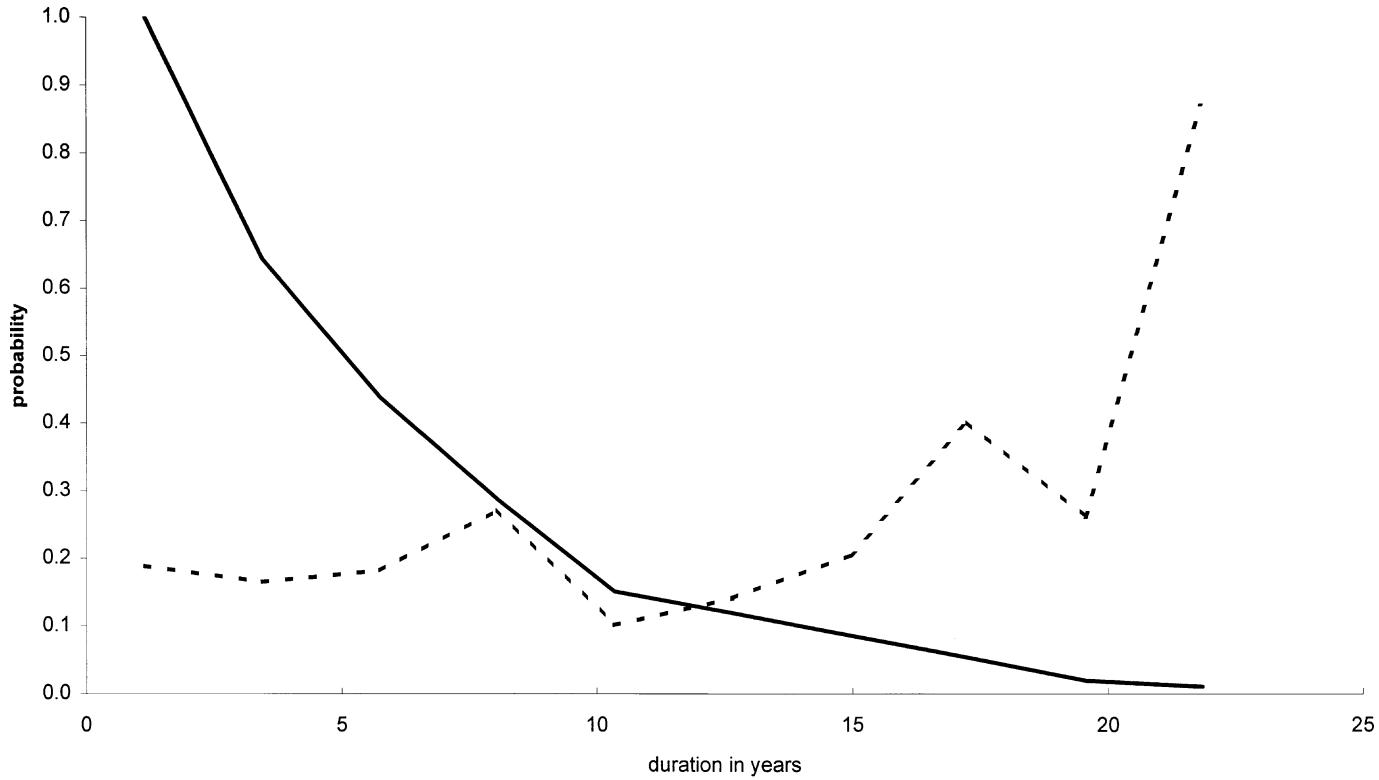

Fig. 5. Survival analysis (adjusted for censoring): (—) survival rate; (- - -) hazard rate.

apparent in tests for homogeneity in life tables stratified by firm and therapeutic class, where homogeneity across firms is strongly rejected by both log rank and generalized Wilcoxen tests at the 1% level, but only at 22% or worse for homogeneity across therapeutic classes.

4.2. Logit regression analysis

A closer examination of the hypotheses discussed above requires a statistical model of the drug development process. There are a number of dimensions upon which to measure the performance of a firm's development effort. The ultimate measure of success is, of course, the returns enjoyed by the stockholders of the firm and the benefits experienced by individuals who purchase its products. But long and variable delays in obtaining regulatory approval and a variety of other institutional peculiarities of the international marketplace make this a difficult exercise. Profitability of individual products is very difficult to calculate, as is the contribution of any particular product to the financial performance of large diversified firms. Equally, the literature on estimation of demand models for pharmaceutical products is at such a preliminary stage (and comprehensive demand data is so hard to obtain) that deriving estimates of consumer surplus is prohibitively difficult. We focus here therefore on the technological success of development projects, in the sense of obtaining regulatory approval.

A structural model of the drug development process might analyze it in terms of a series of sequential investment decisions by the firm, where in each period the firm learns more about the likely payoff to the project and decides whether to continue the project, and how much to spend. Capturing this learning process empirically is a daunting task, given the hidden information possessed by the firm and the potentially important role of strategic interaction with competitors and regulatory authorities.

A slightly less ambitious model would estimate success probabilities (and time spent) at each stage of the development process conditional upon success at the previous stage, and the impact upon them of factors such as the scope and scale of the firm's development effort. Unfortunately, data difficulties — particularly the fact that were unable to consistently measure spending at each stage — together with computational problems frustrated our efforts to estimate a decision tree by nested logit or a more appropriate competing risks duration model with multiple exit modes. "Exit modes" are not independent, and we have to address the issue of censoring of both durations and information on outcomes.

To analyze these data we therefore drastically simplify the structure of the problem by classifying projects as either "successes" or "failures", collapsing the time dimension, and deleting censored observations from the sample. Though the performance of a development project has two dimensions (success or failure in moving from one stage of the development process to the next versus the time spent at each stage) data problems prevent us from examining the time dimension of performance here. We look simply for evidence of scale and scope effects on the probability of a project successfully generating an NDA using a standard single equation, single level, logit framework. The results are thus presented as a descriptive analysis rather than estimates of a structural model, where our aim is to establish empirical regularities rather than to test behavioral assumptions about decision-making within the firm and at the FDA.

Tables 6–8 present estimation results. The unit of observation is the development project, and the explanatory variables are measured as of the year the project was initiated

Table 6
Logit regression results[a]

	Logit coefficients			Marginal effects on probability of success (×100)		
	1	2	3	1	2	3
Constant	−1.64** (0.16)	−2.64** (0.31)	−2.65** (0.29)			
SCALE: total firm development spending	0.09** (0.02)	0.01 (0.04)	0.01 (0.02)	1.61** (0.29)	0.13 (0.43)	0.16 (0.43)
SCOPE: number of therapeutic classes in which the firm is active		0.17** (0.04)	0.15** (0.04)		3.02** (0.68)	2.75** (0.69)
Stock of previously obtained NDAs			0.24** (0.10)			4.40** (1.79)
Likelihood of log	−280.6	−270.8	−267.8			

[a] Dependent variable: SUCCESS (1 if project resulted in an NDA) 515 observations. Likelihood of log for model with constant only (−296.3). Sample means: SCALE: 6.33 (tens of millions of constant 1998 dollars); SCOPE: 8.73 (development projects spending more than $1 million 1998 dollars per year); stock of past NDAs: 0.44.

** Significant at 1% level.

Table 7
Logit regression results: therapeutic class controls[a]

Model	Logit coefficients			Marginal effects on probability of success (×100)		
	1	2	3	1	2	3
Constant	−1.33** (0.26)	−2.32** (0.35)	−2.52 (0.37)			
SCALE: total firm development spending	0.09** (0.02)	0.002 (0.02)	0.01 (0.02)	1.66** (0.29)	0.03 (0.44)	0.17 (0.43)
SCOPE: number of therapeutic classes in which the firm is active		0.19** (0.04)	0.16** (0.04)		3.30** (0.16)	2.87** (0.70)
Stock of previously obtained NDAs			0.52** (0.15)			9.15** (2.68)
Twelve therapeutic class dummies	Insig[b]	Insig	Insig			
Likelihood of log	−275.6	−263.8	−257.5			

[a] Dependent variable: SUCCESS (1 if project resulted in an NDA) 515 observations. Likelihood of log for model with constant only (−296.3); likelihood of log for model with therapeutic class dummies only (−292.5). Sample means: SCALE: 6.33 (tens of millions of constant 1998 dollars); SCOPE: 8.73 (development projects spending more than $1 million 1998 dollars per year); stock of past NDAs: 0.44.

[b] Acceptance of the hypothesis that coefficients on these variables are jointly equal to zero at 5% level.

** Significant at 1% level.

Table 8
Logit regression results: therapeutic class controls and firm effects[a]

	Logit coefficients			Marginal effects on probability of success (×100)		
	1	2	3	1	2	3
Constant	−1.27 (0.62)	−0.95 (0.68)	−1.03 (0.68)			
SCALE: total firm development spending	0.02 (0.02)	0.04 (0.03)	0.04 (0.51)	0.22 (0.34)	0.56 (0.46)	0.54 (0.46)
SCOPE: number of therapeutic classes in which the firm is active		−0.06 (0.05)	−0.03 (0.06)		−0.92 (0.86)	−0.87 (0.85)
Stock of previously obtained NDAs			0.30* (0.18)			4.56* (2.87)
Nine firm dummies	Sig[b]	Sig	Sig			
Twelve therapeutic class dummies	Insig[b]	Insig	Insig			
Likelihood of log	−222.9	−222.3	−221.2			

[a] Dependent variable: SUCCESS (1 if project resulted in an NDA) 515 observations. Likelihood of log for model with constant only (−296.3); likelihood of log for model with therapeutic class dummies only (−292.5); likelihood of log for model with firm effects only (−229.2); likelihood of log for model with therapeutic class dummies and firm effects only (−223.1). Sample means: SCALE: 6.33 (tens of millions of constant 1998 dollars); SCOPE: 8.73 (development projects spending more than $1 million 1998 dollars per year); stock of past NDAs: 0.44.

[b] Acceptance or rejection of the hypothesis that coefficients on these variables are jointly equal to zero.

* Significant at 10% level.

(in these tables all the explanatory variables enter in levels, very similar results were obtained using logs). In Table 6 a quite strong result appears to hold: drug development projects are more likely to result in an NDA in firms which have significantly more diverse development efforts, rather than in those firms that simply spent more on development in total. Scale effects as captured by our measure of the overall size of the firm's development effort have a weak positive association with a project's success when entered alone, but this effect disappears when we control for scope. The SCOPE variable is positive and strongly significant and knocks out our scale variable. Past success in the therapeutic class, as measured by the depreciated stock of past successes is also positively associated with a successful outcome of the project. The marginal effects on success probabilities are also given these tables: in model (1), for example, the estimates imply that embedding the average development project in a firm which spends an additional $10 million in total on development would increase the probability of success by about 2%. Models (2) and (3) of Table 6 imply that relocating the average project to a firm active in one more area would increase the probability of success by about 0.03 (the unconditional success probability is 0.262 for this sample of projects).

In Tables 7 and 8 however, it becomes clear that these results may reflect confounding with more subtle effects. In Table 7 we reestimate the models of Table 6, but control for cross-sectional variation in technological opportunity, i.e. in differences in the costs of conducting clinical trials for different disease conditions, with 20 therapeutic class dummies. There is very little difference in the estimated coefficients. In Table 8 we also control for unmeasured systematic differences in research productivity across firms using firm dummies. The impact on the results is quite striking: while therapeutic class effects are very small and insignificant, the firm dummies are individually and jointly highly significant and have a marked effect on our estimated scale and scope coefficients. The strong scope effect found in Tables 6 and 7 goes away completely when firm dummies are introduced. In part this may reflect colinearity between our measure of scope and the firm dummies: much of the variation in the scope measure is between rather than within firms. But we believe that it may also reflect more interesting underlying phenomena such as persistent differences in organizational structure, incentives and decision-making procedures across firms.

5. Conclusions

One interpretation of the recent wave of mergers in the pharmaceutical industry is that firms are attempting to realize advantages conferred by size in the conduct of R&D. The source of these advantages is difficult to establish, however, without going within the firm. Here we use data disaggregated to the project level to explore the determinants of research performance in drug development. In Henderson and Cockburn (1996), looking within the firm revealed that large firms are at a significant advantage in drug *discovery*. Disentangling the effects of scope and scale suggested that, particularly in more recent years, these advantages appear to flow as much from economies of scope as from more conventional economies of scale, although before around 1975 both effects appear to have been

important. These results were quite robust to the inclusion of controls for firm and therapeutic firm effects.

Our analysis here of *development* performance confirms the value of using disaggregated data to explore the determinants of research productivity. In contrast to the discovery phase, we find no evidence for scale effects. The performance advantage of large firms appears to lie in economies of scope rather than of economies of scale: all else equal, a development program initiated within a more diverse development effort is significantly more likely to result in an NDA than one initiated within a more narrowly focused effort. These results are robust to the inclusion of controls for therapeutic class effects.

In marked contrast to our findings for drug discovery, however, these results are not robust to the inclusion of controls for firm effects. This result raises a number of intriguing issues for further research. It may be the case, for example, that success in drug development is purely a function of success in drug discovery: "good" firms discover promising compounds that they then move forward into development. Prior research suggests that there are very significant differences across firms in the productivity of discovery research, and it would be interesting to test the degree to which success in drug discovery correlates with development productivity.

An alternative possibility is that the dominant role of firm effects in our econometric analysis reflects real, and important, economic phenomena (Henderson and Cockburn, 1994). The impact of the scope and scale of firms' drug development efforts upon their success rate appears to be quite small relative to enduring idiosyncratic differences across firms in the organization and management of the drug development process. Interviews conducted with senior managers of our sample firms suggest that differences in development *strategy* — in the pace and timing of development spending, and in the formulation of the research strategy that guides clinical development — are important determinants of development productivity. Answers to the deeper questions of how these strategies are formulated and implemented, and why the returns that accrue to them are not rapidly competed away by imitation, await more and better data. By capturing these organizational and managerial effects empirically, and evaluating them in the context of the modern economics of organization and incentives, we may be able to develop a richer understanding of the economic performance of the pharmaceutical industry, and its impacts on health and welfare.

Acknowledgements

This research was funded by the UBC Entrepreneurship Research Alliance (SSHRC Grant #412-93-0005), four pharmaceutical companies and 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, and to the Tufts Center for the Study of Drug Development for assisting us with data on project outcomes. We thank Paul Gertler for helpful comments on an earlier draft. Any errors or omissions remain entirely our responsibility.

Appendix A. Data sources and construction

The dataset 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.

A.1. Inputs

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

A.1.1. Research versus development

We distinguish between resources devoted to research (or "discovery", in the terminology of the industry) and "development". "Research" is defined as preclinical expenditures within a therapeutic class, whereas "development" is expenses incurred after a compound has been identified as a development candidate. We attributed exploratory research to a particular program wherever possible, but exploratory research that could not be so assigned was included in overhead. Clinical grants are included in the amounts for development, and grants to external researchers for exploratory research are included in the total for research. In some cases, the companies supplied us with data already broken down by research versus 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.

A.1.2. Overhead

In order to maintain as much consistency in the data collection process as possible, we tried to include appropriate overhead charges directly related to research activities, such as computing, R&D administration and finance, etc. but to 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.

A.1.3. 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 research. Royalty fees and contingent payments are excluded. Though increasing over time, realized expenditures on licensing are a vanishingly small fraction of research spending in this sample.

A.2. Outputs

We measure output of drug discovery by grants of "important patents", where importance is defined as filing in at least two out of three major jurisdictions: Europe, USA, and Japan.

The output of drug development programs is measured by movement through US regulatory process: applications and grants of INDs and NDAs.

A.3. 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). We classified both inputs and outputs according to a scheme which closely follows the IMS world-wide 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, for example, cystic fibrosis, but we were 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 spillovers in this respect. 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 significant miss-assignment of outputs to inputs, but we believe that such errors that remain are not serious.

A.4. 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 which 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.

A.5. Deflation

Since our data sources span many years, it is important to measure expenditures in constant dollar terms. We used the biomedical R&D price index constructed by James Schuttinga at the National Institutes of Health (NIH). The index is calculated using weights that reflect the pattern of NIH expenditures on inputs for biomedical research, and thus in large measure reflects changes in the costs of conducting research at academic institutions. However, since the firms in our sample compete directly with academic research laboratories for scientific talent we believe that this index is likely to the most appropriate publicly available index, and our results proved to be very robust to the use of alternate indices. 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.

A.6. Construction of stock variables

Annual flows of research and expenditures were capitalized following the procedure described by Hall et al. (the R&D master file: 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 in general is not going to be very important in terms of its 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 stock 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.

References

Acs, Z.A., Audretsch, D.B., 1988. Innovation in large and small firms: an empirical analysis. American Economic Review 78, 678–690.

Arrow, K., 1962. Economic welfare and the allocation of resources for invention. In: Nelson, R.E. (Ed.), The Rate and Direction of Inventive Activity. Princeton University Press, Princeton, pp. 609–619.

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

Bound, J., Cummins, C., Griliches, Z., Hall, B.H., Jaffe, A.B., 1984. Who does R&D and who patents? In: Griliches, Z. (Ed.), R&D, Patents and Productivity. Chicago University Press, Chicago, pp. 21–54.

Caves, D.W., Christensen, L., Tretheway, M., 1984. Economies of density versus economies of scale: why trunk and local service airline costs differ. RAND Journal of Economics 15, 471–489.

Chandler, A., 1990. Scale and Scope. MIT Press, Cambridge, MA.

Cohen, W.M., Levin, R.C., 1989. Empirical studies of innovation and market structure. In: Schmalensee, R., Willig, R. (Eds.), Handbook of Industrial Organization, Vol. 2. North Holland, Amsterdam, pp. 1059–1107.

Comanor, W.S., 1965. Research and technical change in the pharmaceutical industry. Review of Economics and Statistics 47, 182–190.

Freeman, C., 1982. The Economics of Industrial Innovation. MIT Press, Cambridge, MA.

Glass, J.C., McKillop, D.G., 1992. An empirical analysis of scale and scope economies and technical change in an Irish multiproduct banking firm. Journal of Banking and Finance 16, 423–437.

Graves, S.B., Langowitz, N., 1993. Innovative productivity and returns to scale in the pharmaceutical industry. Strategic Management Journal 14, 593–605.

Helfat, C., 1997. Know-how complementarities and knowledge transfer within firms: the case of R&D. Strategic Management Journal 18, 339–360.

Henderson, R.M., Cockburn, I.M., 1994. Measuring competence? Exploring firm effects in pharmaceutical research. Strategic Management Journal 15, 63–84.

Henderson, R.M., Cockburn, I.M., 1996. Scale, scope and spillovers: the determinants of research productivity in drug discovery. RAND Journal of Economics 27, 32–59.

Jensen, E.J., 1987. Research expenditures and the discovery of new drugs. Journal of Industrial Economics 36, 83–95.

Panzar, J., 1989. Technological determinants of firm and industry structure. In: Schmalensee, R., Willig, R. (Eds.), Handbook of Industrial Organization, Vol. 2. North Holland, Amsterdam.

Panzar, J., Willig, R., 1981. Economies of scope. American Economic Review 71, 268–272.

Pavitt, K., Robson, M., Townsend, J., 1987. The size distribution of innovative firms in the UK: 1945–1983. Journal of Industrial Economics 35, 297–319.

Schumpeter, J.A., 1950. Capitalism, Socialism and Democracy. Harvard University Press, Cambridge, MA.

Schwartzman, D., 1976. Innovation in the Pharmaceutical Industry. Johns Hopkins University Press, Baltimore.

Silk, A.J., Berndt, E.R., 1993. Scale and scope effects on advertising agency costs. Marketing Science 12, 53–72.

Teece, D., 1980. Economies of scope and the scope of the enterprise. Journal of Economic Behavior and Organization 1, 223–247.

Vernon, J.M., Gusen, P., 1974. Technical change and firm size: the pharmaceutical industry. Review of Economics and Statistics 56, 294–302.