



Faculty Research Working Papers Series

Pharmaceutical Innovation

F.M. Scherer

John F. Kennedy School of Government - Harvard University

February 2007

RWP07-004

The views expressed in the [KSG Faculty Research Working Paper Series](#) are those of the author(s) and do not necessarily reflect those of the John F. Kennedy School of Government or of Harvard University. Faculty Research Working Papers have not undergone formal review and approval. Such papers are included in this series to elicit feedback and to encourage debate on important public policy challenges. Copyright belongs to the author(s). Papers may be downloaded for personal use only.

PHARMACEUTICAL INNOVATION

F. M. Scherer
Harvard University

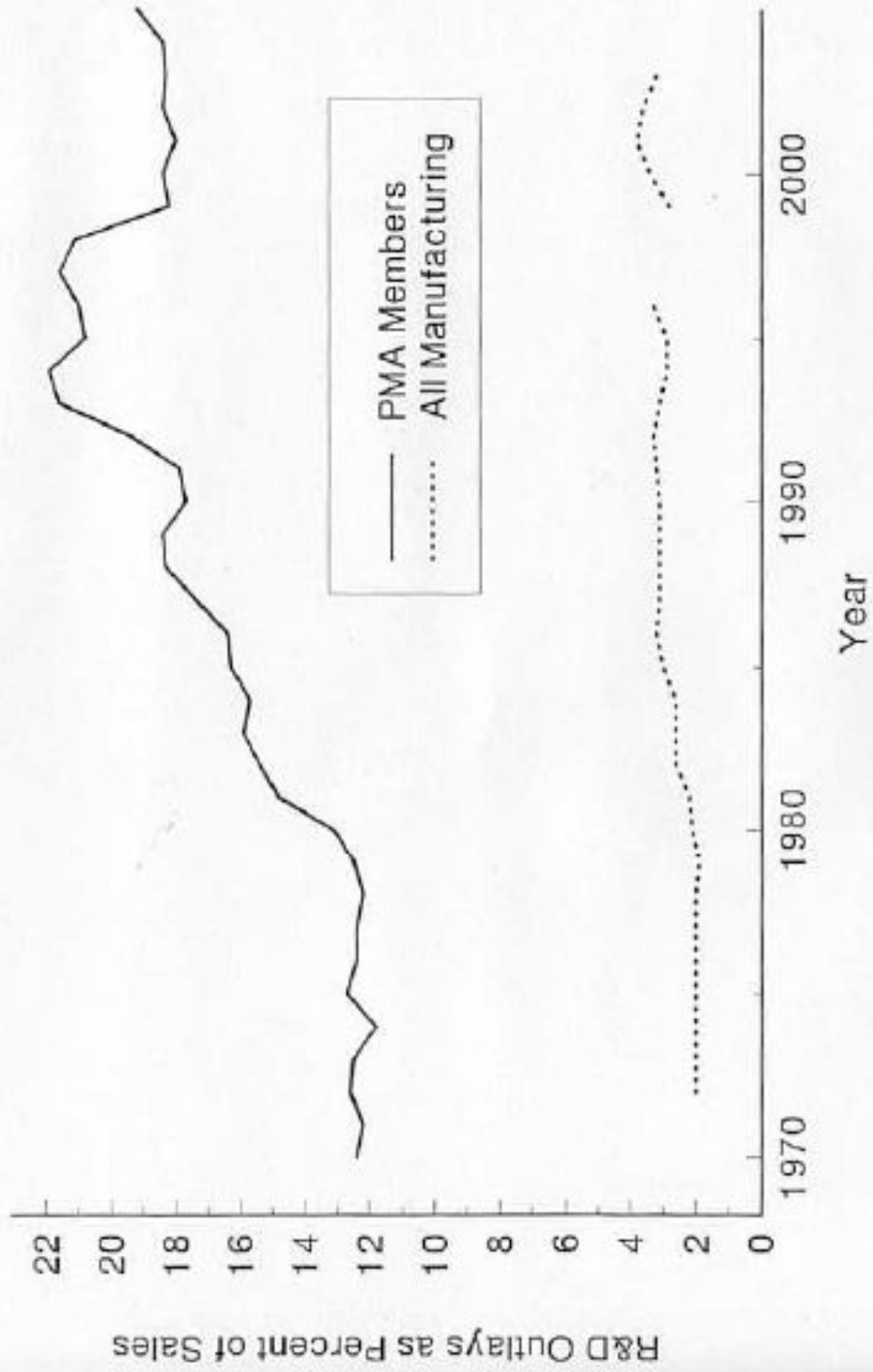
January 2007

Introduction

The discovery and development of new pharmaceutical substances are among the most interesting of innovation processes. Unusually large privately-financed expenditures on research and development (R&D) outlays are required to achieve a successful new product, and relative to the pharmaceutical industry's sales, R&D/sales ratios are extraordinarily high. The links to academic science and basic research performed in government laboratories are rich. The expectation of patent protection plays a more important role than in most other high-technology industries. New products must meet not only the test of market acceptance, but also survive rigorous scrutiny from government regulatory agencies. And the medical services market into which pharmaceuticals sell is itself unusually complex, with a significant fraction of consumers' purchases, at least in the wealthier nations, covered by insurance and hence subject to diverse moral hazard and adverse selection imperfections. Despite these problems, there is compelling evidence that the introduction of many new pharmaceutical products has yielded substantial net benefits in extending human lives and reducing the burden of disease. See e.g. Lichtenberg (2001, 2003) and Long et al. (2006).

Among these various characteristics, we focus preliminarily on one: the high ratio of pharmaceutical companies' R&D spending in relation to their sales. The clearest indicator of this trait comes from data systematically collected over the years by the U.S. industry trade association, previously called the Pharmaceutical Manufacturers Association (PMA) and more recently the Pharmaceutical Research and Manufacturers of America (PhRMA). The solid line in Figure 1 shows aggregate PhRMA member company-financed R&D expenditures within the United States as a percentage of U.S. sales for the years 1970-2005. From 12.4 percent in 1970 declining to 11.8 percent in recession year 1974, the spending ratio rose to a peak of 21.6 percent in 1996 before declining to its 2005 value of 19.2 percent.

Figure 1
R&D as Percentage of Sales, Pharmaceuticals and All Manufacturing

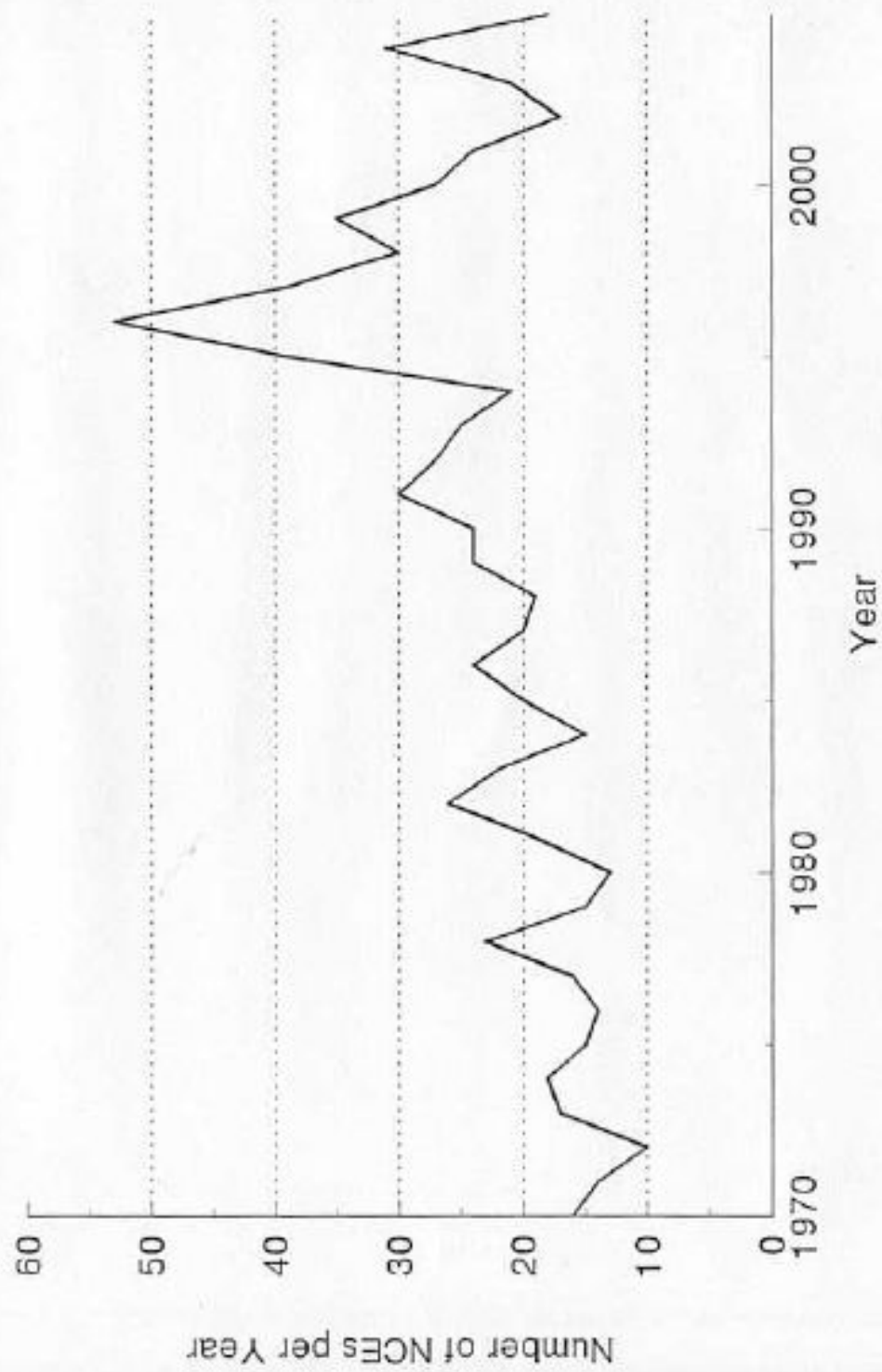


Included for comparison is a more fragmentary series (dotted line) of data on company-financed R&D outlays as a percent of sales for all manufacturing corporations reporting R&D expenditures in periodic National Science Foundation series. In recent years, pharmaceutical R&D/sales ratios have been nearly seven times as high as those of their all-manufacturing cohort.

The National Science Foundation data are often cited as an indicator of research intensity in pharmaceuticals. For the years 1999-2003, the average R&D/sales ratio for the NSF industry category "pharmaceuticals and medicines" was 9.2 percent, compared to 18.3 percent with the PhRMA series in Figure 1. But the NSF figures are biased downward because they are assembled using what is called the whole company method, under which all the R&D outlays and sales for a company are assigned to the industry in which the company has its largest volume of sales. Thus, they are aggregations of pharmaceutical companies' R&D activity in their home industry along with data from the much less research-intensive toiletries, cosmetics, first aid supplies, and insurance payment processing industries, among others. That the PhRMA data present a more accurate picture of what transpires in modern pharmaceutical manufacturing is revealed through data collected by the U.S. Federal Trade Commission (1985, p. 31) for the 1970s in narrowly-defined "lines of business." For "ethical drugs," the reported company-financed R&D/sales ratio was 10.2 percent, compared to 11.3 percent in the contemporaneous PMA report (which excluded many older, i.e. generic, drug sales). Among the 220 lines of business for which the FTC reported data in 1977 (and also in 1974 through 1976), ethical drugs had the highest R&D/sales ratio of any industry. For proprietary (e.g., over-the-counter) drugs, the comparable ratio in the 1977 FTC report was 2.9 percent, for toiletries and cosmetics, 2.5 percent, and for surgical and medical supplies, 3.8 percent.

From the R&D efforts of pharmaceutical companies has flowed a stream of new products, which is traced in Figure 2 through a count of the "new chemical entities" approved for marketing in the United States by the governing regulatory agency, the Food and Drug Administration. Excluded from the count are new formulations of existing products, combinations of previously approved entities, new uses of approved products, and most new biological entities and vaccines (in principle, the output of a

Figure 2
New Chemical Entities Approved for Marketing in the U.S., 1970-2005



somewhat different industry covered elsewhere in this volume).¹ The unusually robust upward fluctuations for 1995 and 1996 came from backlog reductions on drugs awaiting approval by the Food and Drug Administration. When those two years are excluded, the average number of new chemical entities approved per year between 1970 and 2005 was 21.4, with a statistically significant upward time trend.

Time Phases

It is customary to characterize new drug discovery and development in terms of time phases. The principal dichotomy is between the pre-clinical and clinical phases.

In the pre-clinical phase, research efforts are oriented at first toward isolating chemical or biological molecules that might have interesting therapeutic action in vitro and then testing for such action in diverse animals. The "animal model" tests often involve several different species, progressing over time from worms, mice, guinea pigs, and the like up the evolutionary scale to dogs and monkeys.

If those tests suggest that the drug might be efficacious and not cause toxic effects, the drug developer works to formulate the candidate in a form (e.g., pill) suitable for medical use and seeks permission to begin tests in human beings -- i.e., in the clinical trial phase. In the United States, formal approval for clinical testing occurs when the Food and Drug Administration issues a so-called IND -- i.e., investigation of new drug -- permit. The clinical phase is in turn conventionally divided into three main phases. In Phase I, the drug is administered to a small number of subjects, sometimes with the target disease and sometimes not, to test for the safety of various dosages and (when diseased subjects are included) for preliminary indications of therapeutic efficacy. More careful and extensive tests for therapeutic efficacy are administered in Phase II. If those tests are promising, the effort moves into Phase III, in which the drug is administered to at least two panels of patients who might be expected to

1. In collecting R&D statistics, the U.S. Census Bureau and National Science Foundation record biologically-oriented companies as chemicals manufacturers when they sell substantial amounts of product but as research and development service providers when they have not yet marketed products.

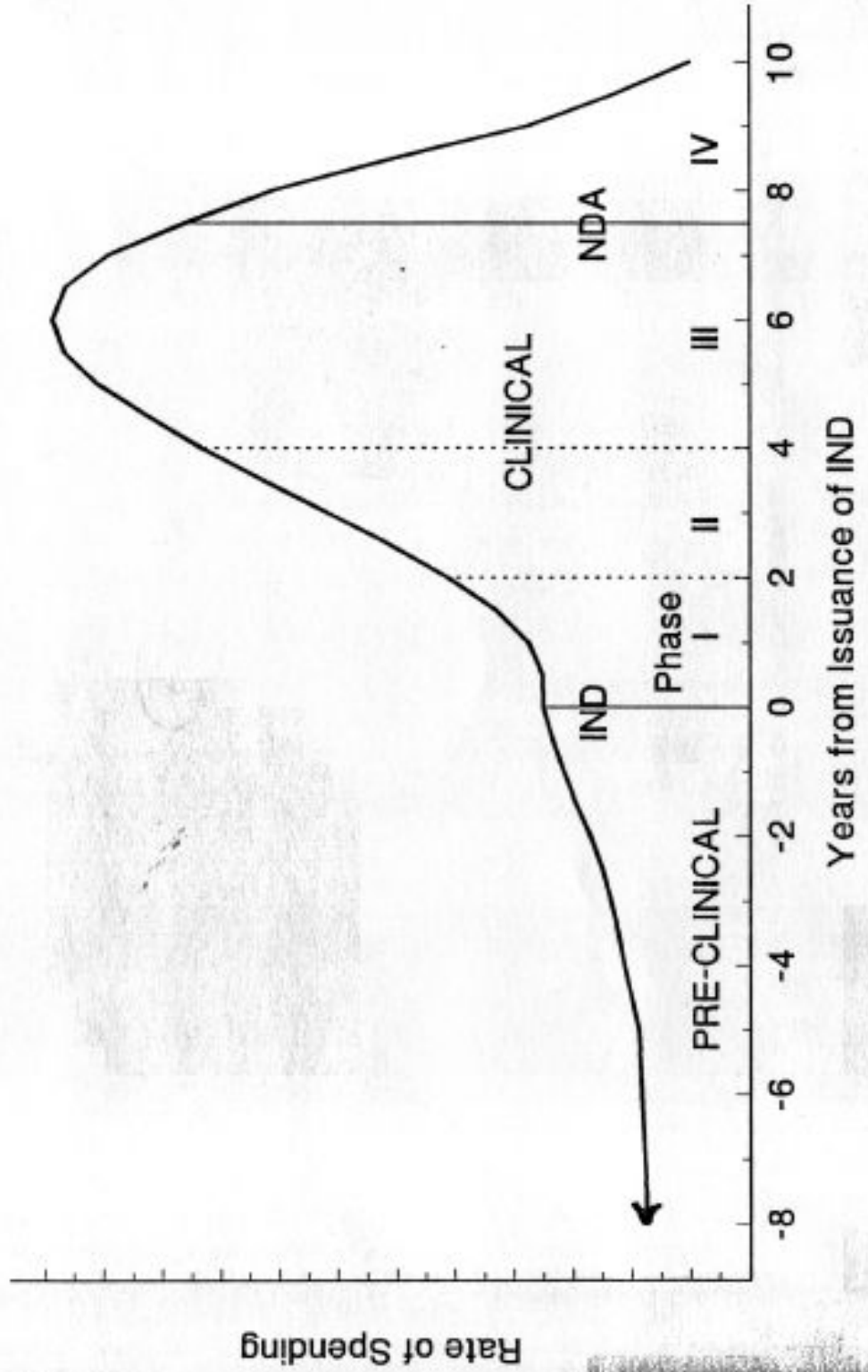
benefit from the therapy. The number of subjects can run to as high as several thousand, especially for drugs targeted toward already curable diseases or that will be used for long-term therapy. With the principal exception of drugs combatting diseases (such as AIDS) that could be lethal if untreated, the tests are double-blind, with half the subjects receiving the drug being tested and the other half a sugar pill or, less frequently, an alternative drug known to be effective against the disease. Under the double-blind approach, neither the recipients of the drug nor (to minimize subtle psychological influences) the administrators know whether a specific subject is receiving the drug being tested or a placebo. If the Phase III trials yield favorable results, the firm sponsoring the trials applies in the United States to the FDA (or in Europe to the European Medicines Approval Agency²) for a New Drug Approval (NDA), submitting voluminous trial data to support its application. Testing often continues during and after the approval interval through Phase IV trials, sometimes to answer unresolved questions posed by the regulatory agency and sometimes to provide additional evidence for the sponsoring company's planned marketing campaign.

Figure 3 presents a stylized characterization of the annual rate at which funds are spent in pharmaceutical discovery and development leading to a specific useful molecule and how the spending cycle is divided among the various phases. No overlap among phases is assumed, although some overlap can in practice occur. The spending rate is relatively modest in the early pre-clinical phases, rises sharply as clinical tests begin, peaks during Phase III human trials, and then tends to decline sharply with application for regulatory approval and movement into Phase IV tests, if any.

Detailed analyses of selected drug development histories by DiMasi et al. (1991, 2003) provide among other things estimates of the attrition rates marking transition into successive clinical testing phases and ultimate marketing approval. For self-originated drugs brought into clinical testing by multinational pharmaceutical companies between 1970 and 1982, some 23 percent of the molecules entering Phase I testing ultimately gained marketing approval from government regulators

2. See Healy and Kaitin (1999).

Figure 3
Rate of Pharmaceutical R&D Spending by Phase



following the completion of Phase III. For a later cohort initially tested in humans between 1983 and 1994, the estimated success rate was 21.5 percent.³ In the later study, the attrition rate between Phase I and Phase II was 29 percent; between Phases II and III it was 56 percent. Of the molecules brought into Phase III, 68.5 percent survived; thus, 31.5 percent failed during Phase III. In pharmaceuticals, it would appear that substantial risks of total failure persist later in the research and development cycle than in most other industries. More typically, as the rate of R&D spending rises, uncertainties abate, although uncertainties concerning market acceptance persist well past completion of the principal R&D tasks. See Branscomb and Auerswald (2001). Indeed, unless key technical uncertainties are resolved at relatively low spending levels, firms conducting the R&D will typically choose not to carry their efforts into higher levels of annual expenditure.

On average in recent U.S. experience, Phases I through III span six to seven years, with the regulatory approval process consuming an additional one to two years. On this, more later. There is considerable variation from case to case; drugs for rare and incurable diseases are often tested and approved on expedited timetables. Even more variation is found in the length of pre-clinical research, and indeed, as we shall see, relevant events in the discovery process can sometimes be traced out to decades before testing in human beings commences.

Changing Discovery Methods

Over time immemorial, humans found through trial and error that certain naturally-occurring substances had medicinal effects. Quinine, present in chinchona tree bark and first extracted by chemical methods in 1810, helped alleviate the symptoms of malaria. Vaccination, first with live smallpox toxin and then, thanks to Edward Jenner's experimentation in the 1790s, with a less dangerous vaccine based upon cowpox toxin, helped eradicate the scourge of smallpox. The bark of the white willow tree was known to provide relief against fevers and headaches. The active substance, salicylic acid, was extracted and identified in the 1830s.

3. For more recent success rate estimates using a different sampling approach, see Tufts Center (2006).

Salicylic acid, however, had unpleasant side effects -- ulcers and other gastric distress. Bayer A.G. of Germany had become one of the world's leading producers of synthetic dyestuffs, created through the manipulation and synthesis of organic chemical molecules. In 1896 Bayer established a laboratory to synthesize and test dyestuff formulations for medicinal effects in humans. One of its candidates, acetylsalicylic acid, proved to be as effective against fever and headaches as its parent molecule, salicylic acid, but with far milder side effects. The new formulation was named "aspirin," which was patented, trademarked, licensed, and sold profitably by Bayer throughout the world. See Mann and Plummer (1991). The beginnings of the modern pharmaceutical industry can be traced to Bayer's work on aspirin.⁴ German dye-makers (amalgamated under the umbrella of I.G. Farben) continued to test coal-tar-based dyestuff variants for therapeutic effects, and in 1935, they discovered a wholly new class of so-called sulfa drugs, the first of which was sulfathiazole. The sulfa drugs proved to be remarkably effective against a range of bacterial diseases such as spinal meningitis, various forms of pneumonia, and gonorrhoea. Variants were subsequently discovered to act as relatively safe diuretics, i.e., to reduce tendencies toward high blood pressure.

There was also progress on the theoretical front. During the 19th Century much was learned about the nature of cells, vaccines, and disease processes in the human body. In 1899 Paul Ehrlich became director of the Institute for Experimental Therapy in Frankfurt/Main, Germany. He conceived an "affinity" theory of small organic molecules' target-specific binding to particular sites in living organisms such as cells in the human body and postulated that if one could find the right "magic bullet," diseased and disease-causing organisms could be destroyed without otherwise harming their human carrier. Salvarsan, one of the more than 600 molecules Ehrlich and colleagues synthesized and tested for therapeutic effects against syphilis, a widespread and essentially incurable disease, proved in 1908 to be effective.

4. For various views on early pharmaceutical discovery approaches, see Schwartzman (1976, Chapter 2), Gambardella (1995), and Werth (1994).

The sulfa drugs provided a first line of defense against bacterial infections. A new line began to open up in 1928 when Alexander Fleming of London observed, but did not follow through on, the anti-bacterial action of a mold that had drifted onto and killed bacteria he was culturing in a Petri dish.⁵ His work was revitalized during the late 1930s by Howard Florey and Ernest Chain at Oxford University. Recognizing the anti-bacterial properties of Fleming's discovery, penicillium notatum, and with financial support from the Rockefeller Foundation, they struggled to develop methods of producing the substance in quantities sufficient first to conduct tests in live subjects and then to use in general medical practice. Their results were considered important to treating casualties from looming World War II, and so the British government approved their transmission to the United States government. American defense authorities set in motion a major effort to produce penicillin in large quantities, eventually, by deep-vat fermentation in corn steep liquor -- a process developed initially by a U.S. Department of Agriculture laboratory in Peoria, Illinois. Contracts to produce penicillin were let to 20 chemical companies, which expanded production for military hospitals and simultaneously gained expertise in the technology of antibiotics. See U.S. Federal Trade Commission (1958). This massive effort provided a major impetus to the emergence of a vibrant American pharmaceutical industry.

While Alexander Fleming's penicillin discovery lay dormant, Selman Waksman of Rutgers University began investigating whether naturally occurring spores might have antibiotic properties. With financial support from the Merck Company, Waksman and his students collected and tested against bacterial cultures approximately 10,000 soil samples. They made two discoveries -- a new antibiotic, streptomycin, effective among other things against tuberculosis, and more importantly, a method for discovering even more new pharmaceuticals -- the systematic screening of molds, fungi, and other substances occurring in nature.

With Waksman's success, the rapidly growing pharmaceutical industry had two main methods of identifying potential medicines

5. On the economic history of antibiotics production, see U.S. Federal Trade Commission (1958) and Kingston (1996).

-- the screening of naturally-occurring substances, plus the organic molecule synthesis approach pioneered half a century earlier by Bayer and Ehrlich. With each method, pharmaceutical action was ascertained empirically, that is, by testing the effects of a molecule on manifestations of disease. For antibiotic action, tests were initially conducted on cultures of bacteria growing in a Petri dish or a test tube -- i.e., in vitro. For anesthetic or tranquilizing action, initial tests might be conducted in earthworms. For blood pressure action and the like, the first target would be laboratory mice.

For the pharmaceutical industry, the antibiotic revolution also had unexpected negative effects. Penicillin technology was widely diffused to facilitate rapid expansion of wartime production. Waksman obtained patent protection on streptomycin, but licenses were made available widely. Price competition soon emerged and then intensified, driving the prices of penicillin and streptomycin down sharply, in some cases, below average production costs. Producing the new "wonder drugs" was found to unprofitable.

Salvation came with new discoveries. Building upon what had been learned with penicillin and streptomycin, the pharmaceutical companies began synthesizing or modifying naturally occurring antibiotic molecules to offer a new, more powerful line of antibiotics -- the so-called broad-spectrum antibiotics, starting with aureomycin in 1948 and then encompassing several molecular variants. These could be patented, and they were sold for many years at prices several multiples of their production costs. Developing and patenting new pharmaceutical entities was found to be a profitable endeavor. Efforts to discover new and different kinds of pharmaceuticals proliferated, precipitating a rapid rise in R&D outlays.

The search for new and effective drugs in the 1950s and 1960s was preponderantly empirical and intuitive, through the screening of plausible alternative molecules. Where similar molecules had already exhibited therapeutic activity, the screens were typically narrow. Investigators tried to identify comparable effects from "me too" molecular variants. When new molecules had no clear therapeutic antecedents, the screens were broad, i.e., covering a panoply of possible diseases. David Schwartzman (1976, p. 60) reports that in 1970, Pharmaceutical

Manufacturers Association members prepared, extracted, or isolated for medical research 126,060 substances and tested for pharmacological action some 703,900 substances (many, presumably, duplicative), among which only about a thousand showed enough promise to be advanced through higher animal tests into human trials. One company interviewed for an inter-agency government inquiry put 20,000 compounds through a narrow screen for antibacterial activity in 1966 and carried roughly 4,000 into further animal tests because of preliminary activity indications. The cost per individual screening test at that time was on the order of \$50, while tests on animals such as guinea pigs and monkeys had a reported average cost of \$10,000. Harbridge House (1967, pp. III-9 and IV-4). From such screening, sometimes called "random screening," companies developed extensive libraries of molecules with annotations on their effects, including serendipitous pharmacological activity. These were available for guidance when a search was begun for drugs that combatted a particular new medical problem of interest.

Gradually, as medical knowledge accumulated from research in hospitals, academic institutions, and industry, the search process narrowed to focus on molecules predicted on theoretical grounds to have desired therapeutic effects. Although earlier antecedents can be identified, this "rational drug design" approach is said to have flowered in the 1970s and early 1980s. Compare Gambardella (1995, Chapters 2 and 4); and Schwartzman (1996). A leading example was Tagamet, introduced by SmithKlineFrench in the late 1970s. Scientific research had shown that ulcers resulted from excess production of gastric acid in the stomach. Secretion of the acid was in turn instigated by histamine, an amine naturally present in the human body. Search for a therapy against ulcers more effective than traditional antacids such as sodium bicarbonate focused on finding agents that would block the acid-generating action of histamine. This narrowed the research agenda considerably, although trial-and-error research was still needed. SmithKline-French scientists synthesized and tested roughly 700 compounds over a period of ten years before seizing upon the highly successful H₂-antagonist Tagamet (chemical name, cimetidine). Its success in turn spurred others to explore molecular variants on Tagamet, which was soon surpassed by Glaxo's Zantac and Merck's Pepcid. These in turn were later overtaken by a different proton pump inhibitor approach embodied in Astra's

Prilosec. David Schwartzman (1996) argues that the "rational drug design" approach was not as revolutionary a break as some claimed it to be, because scientific knowledge provided imperfect guidance and much screening of alternative molecules, to be sure, targeted screening, was necessary before therapeutically successful molecules were obtained. His criticism is valid, but it is also true that scientific knowledge at least narrowed the searches and limited the use of "try every bottle on the shelf" approaches. Gambardella reports (p. 20) that some 5,000 drugs had to be synthesized for early screens to achieve one marketable product and observes (p. 40) that as of the early 1990s, "attrition rates ... do not seem to have diminished."

A further step forward toward rational drug design came with the perfection of methods such as X-ray crystallography and nuclear magnetic resonance imaging for ascertaining the exact structure of healthy molecules present in the human body that might serve as targets for hostile agents, and also the structure of the invading molecules.⁶ One can then try to design therapeutic molecules whose three-dimensional profile meshes exactly with receptor sites in the target molecule, often a protein, as a key meshes with a lock, and which, having bound to the target, block molecular functions or changes adverse to individuals' health. Here the search narrows to a particular structure for the therapeutic agent. Again, however, empiricism is not eliminated. Many alternative molecules, identified among other ways through computer modelling, might bind to a particular target but have no desirable therapeutic effect or be toxic. Thus, search must continue for a molecule with the right configuration and also the desired therapeutic interaction with its target -- a search much less precisely guided by received theory. For example, in its quest for a molecule that combatted the body's rejection of "foreign" kidney, liver, heart, and other tissue transplants with less severe side effects than the established inhibitor, cyclosporin, Vertex, a startup company pioneering the computer-aided design of therapeutic molecules, explored 367 different variants before finding one that bound to the receptor site appropriately and showed the hoped-for pharma-

6. X-ray crystallography images produced by Rosalind Franklin during the 1950s were crucial in the research by Francis Crick and James Watson ascertaining the structure of DNA.

cological effects.⁷ As a skeptic of the rational design school observed:⁸

A compound may be brilliantly deigned -- everything absolutely rational, but until that compound has been shown not only to do clinically what you want it to do, but to be safe, to be active orally, to stay around in the body, and not to give you nightmares, it's not a drug.

An even newer approach is to use modern genetic methods to identify and synthesize therapeutic molecules. Clinical studies are sometimes able to determine where and how the lack of a particular protein (such as insulin, human growth hormone, or erythropoietin) in the human body leads to ill health. Even more recently, using high-speed DNA-sequencing techniques, researchers can identify gene sequences which, when present in the human body, increase the probability of serious diseases, or alternatively, to isolate the sequences associated with individuals who might be expected from heredity or life style to acquire a disease but do not. Those DNA sequences usually express specific proteins which underlie the disease mechanism. When the absence of a protein is likely to render a person disease-prone, the protein can be synthesized by recombinant methods -- e.g., splicing the relevant strand of DNA into E.coli bacteria and growing the modified organisms in fermentation cultures.⁹ It can then be introduced, not always without difficulty, into the relevant organs of the human body. When a specific protein is found to increase the likelihood of disease, proteins or much smaller traditional organic molecules can be sought, as under rational drug design, to combat the action of those proteins or, as in the methods developed to combat HIV/AIDS, to interfere with the replication of harmful agents within them. These techniques probably provide paths to disease remediation with fewer blind alleys and detours than traditional screening approaches. However, the science is difficult, one cannot be sure whether a particular molecular modification will

7. See Werth, p. 251.

8. Ibid., pp. 215-216, quoting an unnamed Merck vice president.

9. The key discovery was made by Stanley Cohen of Stanford University and Herbert Boyer of the University of California, San Francisco, during the 1970s. Three patents on the Cohen-Boyer gene splicing techniques were then licensed to hundreds of other organizations.

work safely, and the manufacturing techniques tend to be more difficult and error-prone than tried-and-true "small molecule" processes. Uncertainty is by no means eliminated.

The advance of science and technology has also facilitated new, higher-powered methods of molecule screening.¹⁰ Three main methods are of interest. First, through "combinatorial chemistry," fragments of molecules can be treated chemically and combined in a host of different ways on a single multi-well micro-titre plate each well yielding a distinct isomer of some larger organic molecule. The process of generating new and possibly interesting molecules is thereby accelerated. Second, interesting proteins and other organic molecule targets can be arrayed on similar plates with numerous wells, to which are added diverse molecular variants of possible therapeutic interest. From these tests and the methods devised to interpret their results, one can quickly screen to see which of many therapeutically-interesting formulations bind to the targets. What would otherwise be a tedious process of testing for interactions one test tube or Petri dish at a time is accelerated in a kind of mass production. Third, as suggested in the previous paragraph, the process of DNA-sequencing has been accelerated greatly, bringing the costs of sequencing an individual's DNA down to such modest levels that sequencing the DNA or fragments thereof for large numbers of individuals has become feasible. The data gained in this way can be mined for disease proclivities and other characteristics of therapeutic interest.

This third technique has brought still another possibility onto the horizon. The typical drug does not work in all individuals with a given ailment, and the administration of a drug can have adverse side effects in some recipients but not in others. These variations in drug receptivity are undoubtedly related to difference in the subjects' DNA sequences. The new science of "pharmacogenomics" seeks to ascertain through large-scale DNA sequencing which individuals fit into which category.¹¹

10. For an excellent lay exposition of the principal advances, see Geoffrey Carr, "The Pharmaceutical Industry," special survey, The Economist, Feb. 21, 1998. See also "New Chips on the Block," The Economist, Dec. 2, 2006, pp. 24-26; and "Mining DNA for Biomarkers," Business Week, Sept. 5, 2005, pp. 82-85.

11. See Vernon and Hughen (NBER, 2005). According to the research director of a

In this way, individual drugs might be prescribed only for the individuals who will benefit from them and not be harmed by them, permitting better-targeted drug therapy. It is proposed too that pharmacogenomic screening might eliminate from clinical tests of new drugs the individuals likely to experience no effects or adverse side effects and therefore render drug testing processes more precise and less costly. The methods of pharmacogenomics are so new that few if any successful applications have been reported. But there are great hopes for them.

More generally, the enormous advances that have been made in computer-aided structurally-based drug design, low-cost molecular manipulation and screening, DNA screening, and recombinant genetics have inspired optimism about the possibility of a new "golden era" of pharmaceutical discovery. But the pot of gold has proved to be elusive. There have indeed been important breakthroughs, but they have been relatively scarce. As Figure 2 shows, after a burst of activity during the mid-1990s, the number of new pharmaceutical chemical entities introduced into U.S. medical practice during the late 1990s and early years of the 21st century has been stagnant or perhaps even declined. Three influences are evidently at work. On one hand, the development and introduction of new drugs depletes the inventory of long-established chemical possibilities and raises the hurdles a new drug must clear in order to displace already efficacious existing drugs. On the other hand, advances in medical knowledge, laboratory methods, and instrumentation open up new possibilities that should in time lead to new and superior drugs. But third, the latter dynamics work with substantial lags and a good deal of uncertainty, leading to substantial, more or less random, fluctuations in the rate of new drug introduction. Optimists believe a golden era of pharmaceutical discovery is coming. When it will materialize is one of the remaining uncertainties.

Industry - Academic Science Links

The evolution of pharmaceutical discovery away from unguided or at best intuitive random screening toward rational

biotechnology firm, the human genome is believed to contain roughly 5,000 pharmaceutically relevant genes, of which only 47 are targeted by the 200 best-selling drugs of 2003. "Fixing the Drugs Pipeline," The Economist, March 13, 2004, p. 37.

drug design and biological methods has led to increasingly rich linkages between the work of pharmaceutical companies on the one hand and academic science carried out in universities and governmentally-supported research institutes, both in the nations where the companies operate and across national boundaries. This has always been the case to some extent. The early work on sulfa drugs was conducted in German industrial laboratories by scientists trained at prominent German universities, which were at the time world leaders in chemical research and teaching. Penicillin moved quickly from the laboratories of Oxford University to numerous companies producing in quantity for the war effort. The first oral contraceptive was introduced by the G.D. Searle Company in 1960, a decade before the earliest date at which the trend toward rational drug design was said to begin. But a study by the IIT Research Institute (1968, pp. 58-72) for the U.S. National Science Foundation revealed an intricate "tree" of scientific discoveries extending back to 1849 that laid a foundation for the Searle contraceptive and later improvements. The more recent changes lie mainly in the richness and closeness of the science-industry linkages and the magnitude of the science base on which the industry could draw. In 2003, for example, against the \$27 billion of industry R&D expenditures within the United States reported by PhRMA members, the federally financed National Institutes of Health allocated a nearly equal amount to research intramurally and by outside grant recipients, much of it basic, and a considerable but unmeasurable portion of it on studies of direct or indirect interest in the discovery of new drugs. Among the knowledge "spillovers" traced by Adams and Clemmons (2006) through scientific journal article citations, drugs and biotechnology firms had five times the weighted citation volume from firm to university authors as the next most citation-intensive industry and nearly three times the volume of company scientist citations from firms to other firms.

Iain Cockburn and Rebecca Henderson (2000) studied the histories of the 21 new drugs introduced between 1965 and 1992 with the highest over-all therapeutic impact, as judged by industry experts. Among the 21, only five, or 24 percent, were developed with essentially no input from public sector research.¹² They contrast their results with an earlier analysis

12. Cockburn and Henderson leave the role of public science undecided for cyclosporine. But it

by Maxwell and Eckert (1990) concluding that 38 percent of an older sample of drugs were developed without public support. Cockburn and Eckert divide their sample of 21 into three categories -- drugs discovered through essentially random screening, drugs that might be said to fall under the rational drug design rubric, and drugs discovered through fundamental science. All but one of the drugs in the latter two groups built upon key enabling discoveries from public science, while four of the seven random screening drugs did not have clear public science antecedents. For 18 drugs in the Cockburn-Henderson sample on which requisite timing data were available, the lag from the key enabling discovery to synthesis of an effective drug was 17.3 years, with a median value of 12.5 years, a minimum of two years, and a maximum lag of 54 years. Although public sector research played a seminal role in facilitating high-impact drugs, 14 of the 18 drugs on which information was available were first synthesized or, for spore-based drugs, dug up, by private-sector firms. Plainly, a division of labor exists. Academic groups have comparative advantage in advancing the science underlying drug discovery and pharmaceutical companies excel at manipulating molecules into a form suitable for therapeutic use.¹³

Edwin Mansfield (1998) pursued a more aggregate approach toward identifying the importance of academic research to private-sector companies' innovations. He obtained from research and development laboratory heads in 76 to 77 companies, operating in seven broad fields of technology, estimates of the percentage of their new products introduced during two time periods, 1975-85 and 1986-94, that "could not have been developed (without substantial delay) in the absence of recent academic research." For all fields, the average research-dependence fraction was 10 percent for the earlier innovations and 11 percent for the later group. "Drugs and medical products" had by far the highest average research-dependence ratio -- 27 percent for the earlier period and 31 percent for the later period. Supplementing the "could not have been developed" cohort, 13 to 17 percent of the drug and medical

is clear from Werth (1994, pp. 48-50), that academic and hospital investigators played key roles in determining that, with the proper dosage, the substance could be used as an immunosuppressant.

13. See also Reichert and Milne (2002).

product innovations received "very substantial aid" from recent academic research -- a virtual tie with "information processing" for first place among the seven groups. With his "recent" framing question, Mansfield found that the average lag from key academic research results to commercialization in drugs and medicines was 8.5 to 8.8 years.

Academic science is transformed into pharmaceutical innovations through richly interconnected networks.¹⁴ Open science, to be sure, is available to pharmaceutical companies through journal articles and presentations at professional meetings. But in addition, there are tighter links. Pharmaceutical companies provide financial support for academic researchers, and their staffs sometimes perform joint research with academic researchers and co-author articles with them. They also enter into cooperative research and development agreements (CRADAs) with government laboratories such as the U.S. National Institutes of Health, permitting joint research, joint publication, and (under the Stevenson-Wydler Act of 1980) assignment of resulting patents to the companies. In recent years, some pharma companies have opened new laboratories in the vicinity of top academic research institutes in order to facilitate cooperation. Quick absorption of the newest scientific discoveries is facilitated when traditional pharmaceutical companies support their own active programs of basic research. In 1993, for example, drug companies reported that 13.6 percent of their total company-financed research and development budgets was devoted to basic research, as defined by the National Science Foundation -- 18 percent of the basic research spending of all industries covered by the National Science Foundation survey for that year. For the average research-performing company across all industries, basic research was 7.3 percent of total company-financed R&D spending.¹⁵

Even closer links between academia and industry are seen in the emergence of hundreds of small new biotech firms, which tend to locate near academic centers, have academic scientists as

14. See e.g. Gambardella (1995), Cockburn and Henderson (1998), Cockburn et al. (1999), Zucker et al. (1998), and Powell and Grodal (2005).

15. U.S. National Science Foundation, Research and Development in Industry: 1993 (NSF 96-304), p. 45.

their founding entrepreneurs, and count numerous distinguished academic researchers as members of their boards of directors and/or scientific advisory councils. Traditional "Big Pharma" companies in turn license molecules discovered in biotech startups for later-stage commercial development or, with increasing frequency, acquire the biotech companies outright, securing full ownership rights in their development "pipeline" molecules and adding staff associated with them to their own R&D staffs. See Kettler (2000). In this way they augment their inventories of interesting drug development candidates, among other things filling voids created when more traditional drug discovery approaches have yielded disappointing results.

An indication of the extent to which firms introducing new drugs to the U.S. market depended upon others for early-stage discoveries is provided through a study undertaken by the author. For the five years 2001-05, the Food and Drug Administration's web site listing new medical entities approved for marketing during those years was searched. From information provided in the approval lists, the patents claimed by the drug developers as impediments to generic competition could be traced by searching the FDA's so-called "Orange Book." On the 85 new medical entities for which patent information was disclosed,¹⁶ 251 applicable patents were found, or an average of 2.95 patents per molecule. Altogether, 47 percent of the patents were assigned at the time of their issue to companies with names different from (abstracting from obvious name changes due to large-company mergers) the company authorized by the FDA to begin commercial marketing of the sample drugs. Patents issued in the earlier stages of development, i.e., prior to January 2000, were more likely (54 percent) to be assigned originally to firms other than FDA approval recipients than patents issued in later years (38.4 percent). The difference is statistically significant. Evidently, the companies carrying out final-stage development and testing relied disproportionately upon outsiders for early-stage discovery. Among the 251 patents, 10.4 percent went to essentially academic institutions, i.e., universities, hospitals, and independent research institutes. Seven percent went to universities, although a handful of the university assignments were joint with other institutions, including U.S. government laboratories. Seven of the 251 patents had multiple

16. Some drug categories are exempt from reporting their patent backgrounds.

organizational assignees and ten had only individual inventors as assignees. Many of the non-academic patent assignees were biotech companies, although an exact breakdown was not possible because information on companies that have not yet "gone public" is scarce. It cannot be ruled out that at least some of the assignees with names different from that of the company receiving FDA approval had common stock partially or wholly controlled by larger corporate parents, notably, the companies receiving FDA approvals.

That drug discovery has become more science-based and hence more efficient, as argued by Gambardella (1995), might be consistent with a remarkable result in the relevant literature that, at least at the time this was written, had no clear explanation. A research team at Tufts University has published two leading empirical analyses of new drug discovery and clinical testing costs. The first focused on 93 new chemical entities introduced into human testing between 1970 and 1982; the second on 68 NCEs first tested between 1983 and 1994. See DiMasi et al. (1991, 2003). Detailed data were obtained from ten to twelve pharmaceutical companies, and a consistent methodology was applied to allocate the costs of clinical trial failures to drugs that eventually succeeded in gaining approval and also to make the more difficult allocations of pre-clinical research costs to successful molecules. All entities in the sample were "self-originated," which means that the responsible companies did not license or buy discovered molecules from other companies, and hence presumably incurred the costs of discovery internally. The fraction of total constant-dollar pre-clinical outlays as a percentage of total pre-clinical plus clinical testing outlays, without adjustment for the cost of capital invested in the research, was as follows:

1970-82 cohort	57.7 percent
1983-94 cohort	30.0 percent

The decrease is remarkable.¹⁷ One possible explanation is that drug discovery became more sharply focused and hence consumed a much smaller fraction of total R&D outlays. Another possibility

17. An attempt by the author to elicit an explanation at the public forum in which the later results were first presented was unsuccessful, nor was an explanation forthcoming in published reports.

is that in the later period prototypes, even if not actual molecules, were licensed in from outside science-based laboratories, although presumably this should have been ruled out by sample design. A third possibility entails sampling variation or measurement error, although the differences seem too large to be explained in that manner alone. A fourth alternative is that for some reason clinical trial costs exploded during the later time period. We turn to that possibility now, although it must be admitted that an important mystery cannot be resolved here.

Clinical Testing Costs and Regulation

Since 1938, when the Pure Food and Drug Act of 1906 was amended after approximately one hundred persons were killed by sulfanilamide adulterated with poisonous diethylene glycol (used in antifreeze), the interstate sale of new drugs was prohibited in the United States unless the would-be drug provider obtained a safety certification (a New Drug Approval, or NDA) from the Food and Drug Administration. The FDA's powers were quite limited, and new drugs were often introduced into the market with claims of efficacy that were based on evidence that was more impressionistic than scientific. A demand for more stringent regulation emerged when thalidomide, a drug intended for use to combat morning sickness in pregnant women, caused severe birth defects in many infants born of women taking the drug. In Europe, where thalidomide had entered general use, some 8,000 malformed babies were victims; in the United States, there were only nine known cases. The U.S. Congress honored the FDA officer who had sidestepped regulations to keep thalidomide testing at low volumes in the United States, and in 1962, Congress passed the Kefauver-Harris Act to reform drug approval processes. It required the FDA to ensure that new drugs were not only safe, but also that they were efficacious, i.e., that they actually had the therapeutic effects their makers claimed. An earlier loophole allowing full-scale marketing if the FDA did not act within 180 days of an application's filing was eliminated. The FDA in turn issued new regulations requiring that pharmaceutical producers seeking approval for their new drugs follow a strict regimen of clinical testing that adhered to scientifically grounded sample design, experimental control, and statistical inference norms. Among other things, the three-phase approach to clinical testing was introduced, and in Phase III, double-blind testing against a placebo became "the gold

standard" for FDA oversight.

By the late 1960s, a hue and cry arose asserting that the new rules had drastically increased testing costs and that the number of new chemical entities receiving FDA approval had in turn been sharply reduced.¹⁸ Four reasonably well contrived clinical trial cost estimates, all attempting to pro-rate the cost of testing molecules dropped at diverse clinical test phases, provide an overview of what happened.¹⁹ All of the estimates are converted to year 2000 price levels using the gross domestic product implicit deflator:

Source	Test Period	Average Out-of-Pocket Cost per Approved New Chemical Entity ²⁰
Mansfield	Late 1950s	\$5.4 million
Clymer	Late 1960s	\$40.2 million
DiMasi I	1970-early 1980s	\$65.7 million
DiMasi II	1983-late 1990s	\$282 million

There are two striking increases. The first is from the pre-1962 to the post-1962 period, as the Kefauver-Harris Act took hold, with estimates based in both cases on data from a single pharmaceutical firm, SmithKlineFrench.²¹ The second, derived using a methodology consistent between periods, is for drugs entered into testing during the 1970s, as compared to those on which clinical testing began in 1983 or later.

18. The latter result, which stands out in statistical time series longer than the one in Figure 2 (e.g, Scherer 1996, p. 351), was intended, for the FDA wished to discourage the proliferation of "me too" molecules that offered at best trivial incremental therapeutic benefits.

19. The sources in order presented are Mansfield (1970), Clymer (1970), DiMasi et al. (1991), and DiMasi et al. (2003).

20. The Tufts group reports both out-of-pocket costs and costs capitalized at 9 to 11 percent interest rates to reflect the opportunity cost of capital tied up, sometimes for more than a decade, in testing. Capitalization roughly doubles the average NCE cost. It is the capitalized costs that are most frequently cited by pharmaceutical companies, although both alternatives have legitimacy. See Office of Technology Assessment (1994).

21. Mansfield does not disclose his source, but his close relationship with SKF at the time makes it likely that the estimates are comparable within the same firm.

Grabowski et al. (1978) took advantage of a natural experiment to explore the reasons for the first apparent cost increase. In the United States, proof of efficacy was required after 1962; in Great Britain, proof of efficacy was added to the approval standard only in 1971. Between 1960-61 and 1966-70, inflation-adjusted drug development costs in Great Britain increased by a factor of three while they increased in the United States by six times. This result suggests that the change in regulatory regimes in the United States was responsible for a twofold cost increase, with the rest attributable to other factors such as fear of tort liability (following substantial thalidomide damages paid by Continental European firms) and companies' desire to differentiate new drugs from the large number of entities already on the market.

What happened to cause a fourfold increase in price-level-adjusted clinical tests between the third and fourth samples remains, like the reason for the sharp relative fall in pre-clinical outlays, a mystery. The FDA did continue to strengthen its clinical evidence rules, in part as a reaction to further safety shortfalls. As the number of molecules competing to be prescribed against the typical disease increased, companies sometimes found it advantageous on marketing grounds to conduct more Phase III and Phase IV tests than the FDA required. It is also possible that the gross domestic product deflators used to render outlays comparable in terms of purchasing power are inappropriate for measuring inflation of R&D costs, both generally and in pharmaceuticals. Some R&D costs, e.g., for computer capabilities, rose much less than the rate of general inflation, and improvements in scientific apparatus sometimes made it possible to carry out experiments that would have been impossible, in effect, infinitely expensive, previously. However, other costs, and especially the costs of clinical testing, may have risen much more rapidly than the rate of general inflation. Clinical testing is a labor-intensive activity, and the wages of physicians, nurses, and related staff increased more rapidly than the GDP deflator. Much clinical testing is done in hospitals. The average cost per day of hospitalization rose at an average rate of 11 percent per year between 1970 and 1990 while the GDP deflator rose at 5.8 percent. It is also possible, although no research on the matter is known, that hospitals using clinical testing activities as a "profit center" dumped some of their soaring overhead costs into

the charges they levied on deep-pocketed pharmaceutical companies for clinical testing. And of course, sampling error cannot be ruled out. The DiMasi et al. samples are weighted in favor of large pharmaceutical companies targeting their new drugs toward long-term therapy markets able to bear heavy testing costs and requiring extensive proof of long-term safety. Drugs aimed at acute but rare diseases are probably undersampled.²²

Seeking to control their clinical testing costs, pharmaceutical companies have "outsourced" some of their clinical test tasks to independent contract research organizations (CROs) that specialize in test management, and more recently, they have sought to conduct tests in low-wage nations to the extent permitted by regulatory agency rules.²³ (The U.S. FDA requires that some clinical trials be conducted within the United States.) Azoulay (2003) calculated that on average, pharmaceutical companies covered by a large survey outsourced to CROs 23 percent of their clinical test activity on average in 1999. He found too that outsourcing is most frequent for Phase IV trials, and that in earlier-phase trials, use of CROs is less common for tests and disease indications that require complex protocols and quick feedback from unexpected events than for relatively routine testing. In a possible reversion to practices abandoned in the 1970s, pharmaceutical companies have also reconsidered using prisoners in late-stage clinical

22. For data on how testing costs vary over therapeutic categories within the Tufts University sample, see DiMasi et al. (1995). The highest average costs were for non-steroidal anti-inflammatories, which are often taken several times daily for extended periods. The lowest average costs were for anti-infectives, which are usually prescribed for only short periods. For so-called "orphan" drugs and drugs against lethal diseases such as HIV/AIDS, the Food and Drug Administration typically authorized smaller clinical trial samples and shorter trial periods. On orphan drugs, see Shulman and Manocchia (1997).

In recent years the FDA has distinguished between "priority" and "standard" approval candidates -- the former for molecules that offer significant improvements over existing approved therapies. Among the new medical entities approved between 2001 and 2005, 89 percent of the orphan drugs, i.e., those treating diseases affecting fewer than 20,000 U.S. residents, received priority designations, while 38 percent of the non-orphan drugs were designated "priority."

23. "Drug Companies Cut Costs with Foreign Clinical Trials," New York Times, Feb. 22, 2005; and "Outsourcing," Business Week, Jan. 30, 2006, p. 58.

trials.²⁴

The Decision-Theoretic Problem

Evaluating the results of clinical trials is a classic exercise in statistical decision theory. The Food and Drug Administration in the United States, and in the European Community since 1995, the European Agency for the Evaluation of Medicinal Products, attempt to keep unsafe products off the market and to approve only products that can demonstrate their therapeutic efficacy. Because drugs have differing activity in diverse humans, clinical test results are statistically noisy. Decision-makers for both the sponsoring company and the regulatory agency must try to sort out the true effects from the random variation. One must be wary of both Type I errors -- concluding that the drug is safe and/or efficacious when it is not, and Type II errors -- keeping a drug off the market when it actually will do good work. Tradeoffs are also required. Virtually every drug has some adverse side effects. Even century-old aspirin does; it can trigger ulcers, inhibit blood-clotting, and (more rarely) cause fatal Reye's syndrome. One must weigh the beneficial effects against the adverse effects, neither measured with certainty.

For the U.S. Food and Drug Administration, the "gold standard" again has been the double-blind test of a new drug against placebos, e.g., sugar pills. It is not always required, however. Denying an available potential remedy to patients with an otherwise fatal disease violates ethical canons, so the U.S. FDA either allows new drugs to be tested on all clinical trial subjects (e.g., in the early years of the HIV/AIDS crisis) or uses as a comparative benchmark a drug whose therapeutic efficacy has already been established. It has been argued, e.g., by Angell (2004), that, when they are available, established drugs should normally be used as the benchmark for comparison, because in a head-to-head competition, information may be generated that helps physicians choose among alternative therapies by comparing cost with benefits. This is not always a good idea, however. The largest trial of Merck's Cox-2 inhibitor pain reliever Vioxx was against an established over-the-counter drug, naproxen sodium (branded Naprosyn). The

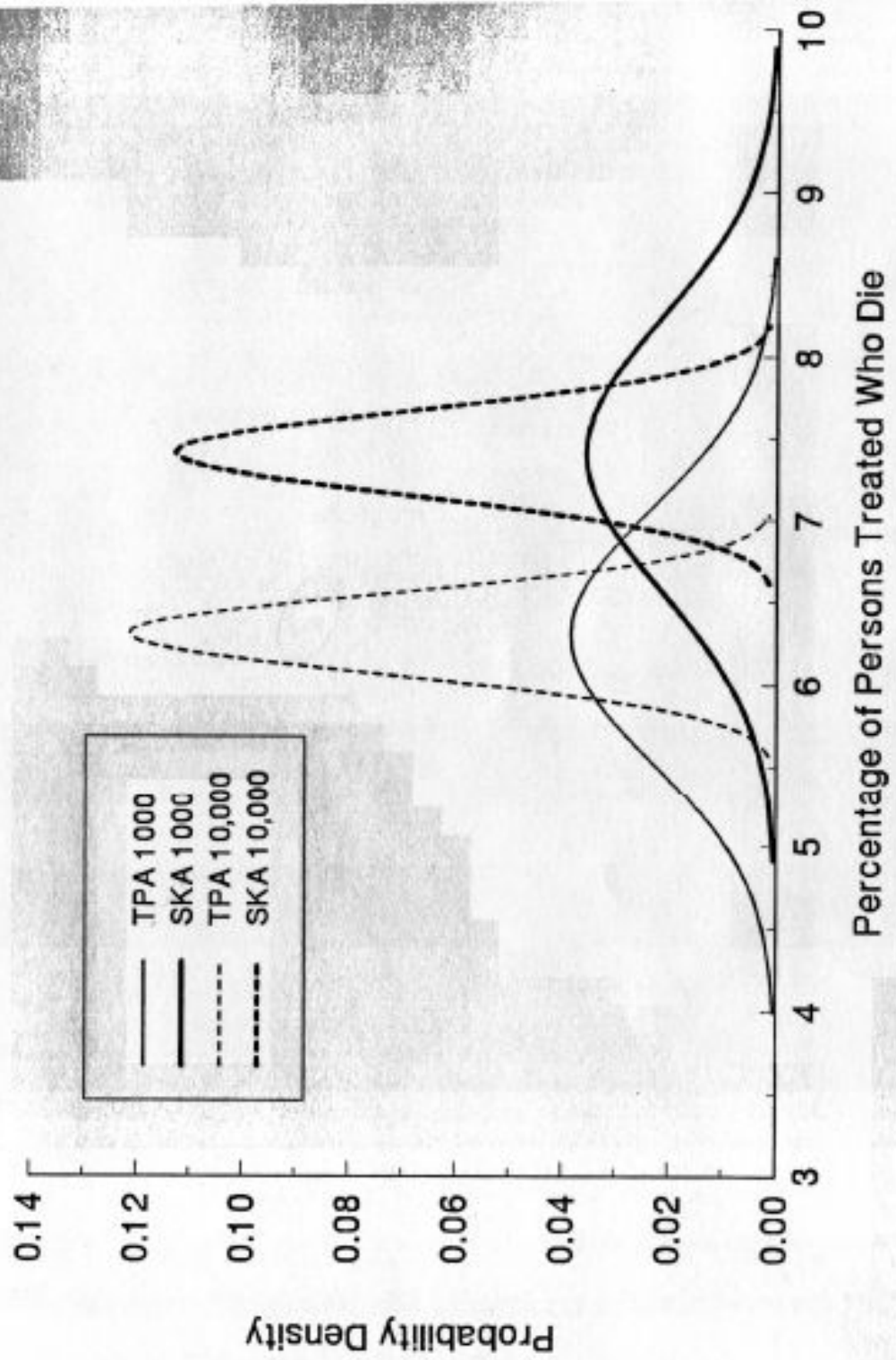
24. "Panel Suggests Using Inmates in Drug Trials," New York Times, August 13, 2006, p. 1.

latter was known to have blood-thinning properties which reduce the severity or likelihood of strokes and heart attacks. When clinical trial subjects were found to have a higher propensity toward adverse cardiovascular events with Vioxx than with naproxen sodium, the inference was drawn that naproxen sodium was having its well-known positive effect, and not that Vioxx was actually causing cardiac events. This was eventually found to be wrong, and when Vioxx was taken off the market in 2004, a torrent of tort litigation followed. An equally large trial against a placebo might have identified the adverse side effects from Vioxx more clearly, but left unclear Vioxx's superior record in avoiding ulcers as compared to naproxen sodium.

The decision-theoretic problems of clinical trial design are illustrated by the case of TPA (tissue plasminogen activator), a genetically engineered drug targeted against the blood clots that accompany heart attacks, versus an older, well-established drug, streptokinase (SKA).²⁵ The null hypothesis on efficacy would be that TPA was not more effective than a placebo or, under an alternative standard, than SKA. Carefully structured trials were conducted, and it was found that 6.3 percent of the subjects died after being injected with TPA plus heparin, while 7.4 percent died with SKA plus heparin. Assuming these values to reflect the true states of nature, Figure 4 shows how the statistical inference problem might evolve. The solid lines show the probability distribution of possible trial outcomes with samples of 1,000 on each alternative -- fairly typical of the Phase III sample sizes required during the early 1990s by the U.S. FDA. If the death rate with a placebo exceeded 8 percent, trials with 1,000 TPA injections would with high probability support a decision to allow marketing of TPA. The risk of a Type I error is quite small -- the area under the right-hand tail of the TPA distribution above 8 percent. But the test would be insufficient to tell whether TPA was more effective than SKA. If both drugs were matched against each other, there would be roughly one chance in three that SKA would be found to be superior when in fact it is inferior. Recognizing this, TPA's developer, the biotech firm Genentech, chose to sponsor head-to-head clinical trials with much larger

25. This example is drawn from Scherer (1996, pp. 354-355).

Figure 4
 Testing TPA Against Streptokinase



samples than those required by the FDA. With samples of 10,000, we see in Figure 4, there is only a small probability, measured by the proportional area under the overlap between the two dash-dash density functions, that one would err in concluding that TPA is not superior to SKA. With this evidence, TPA became the drug of choice for emergency treatment of heart attack and stroke patients, even with a TPA price ten times the price at which SKA was marketed.

Similar problems pervade testing for adverse side effects. Many side effects are rare events, occurring with probabilities less than 0.01. A 12-week-long trial of Vioxx, for example, yielded only about 1.2 heart attack cases per thousand subjects, suggesting a probability (that undoubtedly would have risen with longer therapy) of 0.0012.²⁶ With the sample sizes typically required by the Food and Drug Administration, determining whether Vioxx actually caused heart attacks or whether the relatively few cases observed would have happened in any event was intrinsically difficult.

Given these difficulties, policy-makers and company officials may wish to commence full-scale marketing of a new drug, expecting that rare side effects will reveal themselves when the population using the drug numbers in the hundreds of thousands or millions. There are, however, two main problems with this approach. When adverse effects are rare, physicians administering the drug to their typically small number of patients will seldom be able to discern that observed complications were caused by the drug rather than something that would have happened in any event. And, busy as they are, they are unlikely to take the trouble to report the event to a central office processing data on the entire population. Given this, extraordinary diligence is required on the part of agency responsible for overseeing drug safety, piecing together fragments of imperfect information from the field into a more coherent picture.

But here the wrong thing happened in the United States. The drug evaluation problem is sufficiently difficult that the U.S. FDA took a long time making its decisions, once companies

26. "Evidence in Vioxx Suits Shows Intervention by Merck Officials," New York Times, April 24, 2005, p. 1.

had deposited truckloads of clinical test information on its doorstep. During the 1980s, the time required for the FDA to make a definitive decision on companies' requests for a New Drug Approval averaged roughly 30 months. There were complaints of a "drug lag" relative to nations with less meticulous drug approval systems, and companies claimed that the regulation-induced delay of profitable sales impaired their incentives to sustain R&D efforts.

A solution to the problem was adopted by the U.S. Congress in the Prescription Drug User Fee Act (PDUFA) of 1992. It allowed the FDA to levy user fees upon pharmaceutical companies, in part with a fee per application, partly through a fee proportional to the number of plants licensed, and partly with fees rising with the number of new drug applications approved. The fees, eventually exceeding \$200 million per year, allowed the FDA to augment its new drug evaluation staff, in exchange for which it promised to reduce its decision-making lags to an average of 12 months.²⁷ The approval-linked component had potentially undesirable incentive effects, for the more drugs the FDA approved, the higher its revenues would be. Thus, it might be motivated to approve marginal drugs it would not favor if their approval did not yield additional revenue. The law was modified in 1997 to eliminate this incentive incompatibility by targeting an annual lump sum to be raised from drug approvals, with a high resultant fee per NDA, the fewer NDAs issued. Even so, another difficulty materialized. Congress made it clear that the FDA was not to divert the PDUFA revenues to activities other than new drug review, and when insufficient funds were separately appropriated to support all previously existing FDA functions, the FDA cut back on the resources it devoted to post-marketing surveillance. Its monitoring of post-marketing safety issues abated and responses to lethal side effects from already-approved drugs were delayed. Guided in part by a critical report from the Institute of Medicine (2006) of the National Academies of Science, corrective action was begun after the Vioxx crisis surfaced.

Uncertainty Revisited

27. An irony of the legislation is that at the time, the FDA had an effort underway, called Project 007, to reduce decision-making lags without additional budgetary resources. Even though it was beginning to succeed, the FDA favored having substantially more budgetary resources under the PDUFA system. See U.S. Food and Drug Administration (1993).

A recurrent theme in this essay has been the presence of uncertainty. As the research and testing process progresses, uncertainties are gradually mitigated. Several sources put the number of alternative molecules subjected to early screening at between 4,000 and 10,000 in order to have a single approved drug emerge at the end of the process.²⁸ According to PhRMA (2006, p. 4), the U.S. industry association, a single approved drug emerges on average from five compounds entering clinical testing, 250 molecules subjected to animal and other laboratory tests, and 5,000 to 10,000 molecules initially screened. As the number of drug candidates is winnowed, the costs of continued testing and hence the stakes in the game escalate.

Although no systematic analysis is known, there is reason to believe that the larger the anticipated benefits from curing previously untreatable diseases, and given a low probability that any single approach will succeed, the larger the number of projects pharmaceutical developers pursue. For example, in 2006, approximately 60 clinical tests for possible Alzheimer's disease medicines or preventatives were undergoing clinical trials.²⁹ Treating Alzheimer's patients was estimated to consume annual health care resources on the order of \$100 billion. Such "parallel paths" testing, rising with higher eventual benefits, at least approximates an optimal allocation of resources under uncertainty. See Scherer (1966).

Even when marketing approval is secured, risks do not vanish. As we have seen, severe safety hazards may become evident only when a drug has been accepted widely on the market. And approval is by no means synonymous with commercial success. Grabowski and Vernon (1990, 1994) have shown that the distribution of quasi-rents -- that is, the surplus of revenues over variable production costs for individual drugs -- is highly skew. Among any given 100 drugs introduced into the market, the top ten by number realize from 48 to 55 percent of their cohort's quasi-rents, while the least lucrative 80 out of 100 barely cover, or less than cover, their average capitalized research and testing costs. In its skewness, the profitability

28. See Gambardella (1995, pp. 20 and 40); Schwartzman (1996, p. 846); and PhRMA (2006, p. 4).

29. "Decoding Alzheimer's," Business Week, January 8, 2007, p. 54.

distribution for new drugs is similar to the distributions for most other new products, except that when one focuses only on approved drugs, one ignores the uncertainties that preceded approval and therefore obtains somewhat less skew outcome distributions than with samples that begin at earlier stages of the research and product innovation cycle. See Harhoff and Scherer (2000). Skewness of the distribution of rewards from innovation in turn makes it more difficult to hedge against risk by maintaining a portfolio of projects -- a standard feature of high-technology investment strategies. Scherer and Harhoff (2000) demonstrate, for example, in a simulation analysis of the Grabowski-Vernon data, that even averaging over all the products introduced into the U.S. market for a total of 21 years, skewness and the random appearance of a few extreme values lead to fluctuations in overall annual industry gross profitability of +/- 25 percent. In other words, the overall industry portfolio is insufficiently diverse to eliminate significant profit variations.

To be sure, pharmaceutical companies have some bases for predicting before marketing begins whether their new drug will enter a market with blockbuster potential or a niche in which quasi-rents will at best be modest. Among other things, first movers typically enjoy larger market shares than latecomers.³⁰ However, there are also surprises. The drug with the highest annual sales in pharmaceutical industry history, Lipitor (atorvastatin), was seen by its developers as at best a late entrant into the cholesterol-reducing statins drug market. With limited perceived prospects, the Lipitor project was on the verge of cancellation by its developer, Warner-Lambert, when a small clinical trial revealed, contrary to expectations, that it was more effective at given dosages than rival drugs.³¹ Confirmation of this result plus a marketing decision to set Lipitor's price at half the price of the leading rival propelled Lipitor's sales to record levels, ahead of several competing molecules. Similarly, Abbott Laboratories' Hytrin (terazosin) was synthesized through a minor manipulation -- the replacement of a two double-ring pentane bonds in a quadruple-ring molecule with single bonds -- of an anti-hypertensive drug marketed by Pfizer. Its performance as an anti-hypertensive was

30. See e.g. Bond and Lean (1977) and Robinson and Fornell (1985).

31. See John Simons, "The \$10 Billion Pill," Fortune, January 20, 2003.

unimpressive. But tests by academic researchers revealed serendipitously that Hytrin could ease the symptoms of benign prostate gland enlargement. It was retested for that use and approved, achieving annual sales in its category approaching a billion dollars per year -- a result far beyond the expectations of the team that created it. On the other hand, drugs with the most optimistic prospects sometimes prove to have unacceptable side effects, crashing and burning after hundreds of millions of dollars have been spent for development and testing.³²

The Unique Role of Patents

The expectation of patent protection on new products plays a particularly important role in pharmaceutical R&D decision-making. Levin et al. (1987) surveyed 650 corporate R&D managers, asking them inter alia to evaluate on a scale of 1 (not at all effective) to 7 (very effective) the effectiveness of patents as a means of protecting the competitive advantages from new products. From 17 pharmaceutical industry respondents, the average score was 6.53, compared to a response-weighted average of 4.33 for all 130 surveyed lines of business. Among the industries with more than one respondent, pharmaceuticals ranked second in its patent protection effectiveness score. This result is consistent with the findings of Edwin Mansfield et al. (1986), who asked the top R&D executives of 100 U.S. corporations what fraction of the inventions they commercialized between 1981 and 1983 would not have been developed in the absence of patent protection. For pharmaceuticals, the average response was 60 percent; for all industries, 14 percent.

The importance of patents to pharmaceutical R&D decision-makers stems not only from the large average investments in a typical new product and the many uncertainties lining the path to a new product approval. The differentiating factor is seen among other things through a comparison with another industry -- aircraft -- that taps a range of highly sophisticated technologies and spends billions of dollars developing the typical new product. For aircraft (both civilian and military), the average "effectiveness of product patents" score in the

32. See "Failure of Pfizer Cholesterol Drug Is Major Blow to Promising Approach," Washington Post, December 4, 2006.

Levin et al. survey was 3.79 -- in the lowest third among 130 industry categories.

The key difference lies in the relative ease of imitation, i.e., how difficult it would be, with vs. without patent protection, for new product imitators to launch their own competing products. Even without patents, the firm that would seek to imitate the Boeing 787 would have to build its own scale models, perform wind tunnel tests, compile detailed engineering drawings and specifications for all structural parts, work out electronic system interfaces, construct full-scale test models, test them for structural soundness and aerodynamic performance, and much else, spending very nearly as much as Boeing did to develop its 787. Presumably, it would have observed Boeing's design before undertaking the project, and by the time the imitator completed its developmental work, Boeing would be a decade ahead in sales and have progressed far down its learning curve, enjoying a substantial production cost advantage. But in pharmaceutical discovery and testing, much of the R&D is aimed at securing knowledge: knowledge of which molecules are therapeutically interesting, knowledge of which molecules work in animals, and most costly, knowledge as to whether a target drug is safe and efficacious in human beings. Once that knowledge is accumulated, absent patent protection, it is essentially there as a public good available to any interested party. Achieving it requires by recent U.S. standards an investment measured in the hundreds of millions of dollars. But for most new drugs, and especially small-molecule drugs,³³ a would-be generic imitator could spend a few million dollars on process engineering and enter the market with an exact knock-off copy. Generic entry in turn could quickly erode the quasi-rents anticipated by a pharmaceutical innovator to repay its R&D investment. Hence the importance attributed to patents by drug companies.

This asymmetry between pharmaceutical innovators and imitators was not nearly as glaring during the early 1980s. Because of FDA and Supreme Court rulings, generic drug providers had to invest nearly as much per molecule in clinical testing to

33. I.e., abstracting from biologicals, whose production tends to be more difficult and to entail more secret "black art." Check this against Ben Roberts' new results.

obtain marketing approval as the first-moving innovator.³⁴ Original developers also had problems. They typically sought patent protection just before beginning human tests, when probable "utility" could be documented, and at the completion of those tests, 30-month average decision-making lags at the FDA ate into the 17-year period over which their products were protected by patents. A grand compromise on these two points was achieved in the Hatch-Waxman Act of 1984.³⁵ It allowed patent holders to extend the lives of their patents, compensating for at least some of the period during which their new product introduction was delayed by regulatory oversight. It simultaneously reduced the clinical testing requirements for generic entrants once blocking patents had expired. Contrary to past precedents, the so-called Bolar Amendment also permitted generic drug developers to produce small quantities of the drug in question for their clinical trials before the drug's patents had expired so that they could complete their FDA paperwork and attempt entry as soon as patents expired. The generic entry provisions had a more dramatic impact, increasing the number of prescriptions filled generically in the United States from 19 percent in 1984 to 47 percent in the year 2000. The expectation of rapid generic entry following patent expiration in turn reinforced the incentive of pharmaceutical innovators to invigorate their R&D to compensate for impending profit losses.

Profitability and Research Investments

With strong patent protection and well-differentiated products, pharmaceutical producers enjoyed considerable discretion over the prices they set. Insurance coverage of drug outlays, expanded rapidly during the 1980s, reduced demand elasticities and conferred even more pricing power. One index for measuring pricing power is the price-cost margin (PCM), defined as:

$$PCM = \frac{\text{Sales} - \text{Material Costs} - \text{In-Plant Payroll Costs}}{\text{Sales}}$$

34. See Kitch (1973) and Bond and Lean (1977). A key Supreme Court ruling was *U.S. v. Generix Drug Corp. et al.*, 460 U.S. 453 (1983).

35. For a brief history, see Scherer (2007).

For 459 manufacturing industries on which data were published for the year 1987, pharmaceuticals had the sixth-highest margin, at 61.4 percent. For all manufacturing industries, the average PCM was 30.5 percent. Similarly, for decades pharmaceutical producers appeared at or near the top of Fortune magazine's annual list of broad industry groups, ranked in order of after-tax profit returns as a percentage of stockholders' equity. Beginning already in the late 1950s, the drug makers were accused in public fora of profiteering at the expense of consumers. They argued in return that high profits were a reward for superior innovation and a necessary spur to investment in risky R&D.

Another more subtle defense led, after considerable repetition, to a large-scale analytic investigation by the U.S. Office of Technology Assessment (1994). The basic argument was that, due to the R&D-intensity of pharmaceutical manufacturing and peculiarities in the way accepted accounting principles dealt with R&D outlays, reported profit returns on drug company assets and stockholders' equity were systematically overstated. Specifically, R&D outlays were recorded as a current year's expense when in fact they were investments yielding returns over decades following their incurrence. Ideally, they should be added to asset accounts and then depreciated only slowly. Ignoring their investment character understated drug company assets and hence, given the absolute R&D magnitudes and growth conditions experienced by drug companies, overstated profit ratios in which assets or stockholders' equity comprised the denominator.³⁶ A careful evaluation by the Office of Technology Assessment confirmed the validity of the underlying theory and concluded that, after appropriate accounting adjustments were made, pharmaceutical makers enjoyed returns on investment only two or three percentage points higher than the roughly ten percent real cost of their financial capital.³⁷ And at least part of that differential could be attributed to the riskiness of drug companies' investments.³⁸ In other words, drug companies did not appear to be realizing extraordinary supra-normal profit

36. For the relevant theory, see Stauffer (1971) (IMSEP p. 421).

37. See also Grabowski et al. (2002).

38. Not explicitly recognized in the analysis was the difficulty of risk-hedging through portfolio maintenance with highly skew payoff distributions, as shown by Harhoff and Scherer (2000).

returns.

This conclusion left unsettled the specific behavioral dynamics that reconciled unusually high price-cost margins, atypically high R&D/sales ratios, and bottom-line returns on investment only moderately above all-industry norms. Several studies of the links between profit potential and R&D investment have been published. I focus here on my own analysis, which has been brought as up-to-date as data availability permits.³⁹ The profit potential is measured from U.S. Census data for the "pharmaceutical preparations" industry as sales less materials purchases less in-plant payroll costs (including fringe benefits). Call this variable "gross margins." The data were adjusted for inflation to a 1992 = 100 price level base using the implicit GDP deflator. The coverage is for 1962 through 2004.⁴⁰ R&D data were spliced from various statistical reports of the Pharmaceutical Manufacturers of America and PhRMA. On the assumption that U.S. members allocated resources internationally to the most favorable locations and that overseas R&D was influenced by profit prospects in the firms' largest single market, the United States, the R&D data include "R&D abroad" as well as domestic R&D.⁴¹ Again, deflation was to 1992 price levels. For each deflated time series, an exponential growth trend was fitted using least-squares regression. For the gross margins variable, the average "real" rate of growth was 4.84 percent per year; for the R&D variable, 8.11 percent. Using the fitted trend, percentage deviations from the trend were computed. These are plotted in Figure 5, with margin deviations as a solid line and R&D deviations as a dotted line.

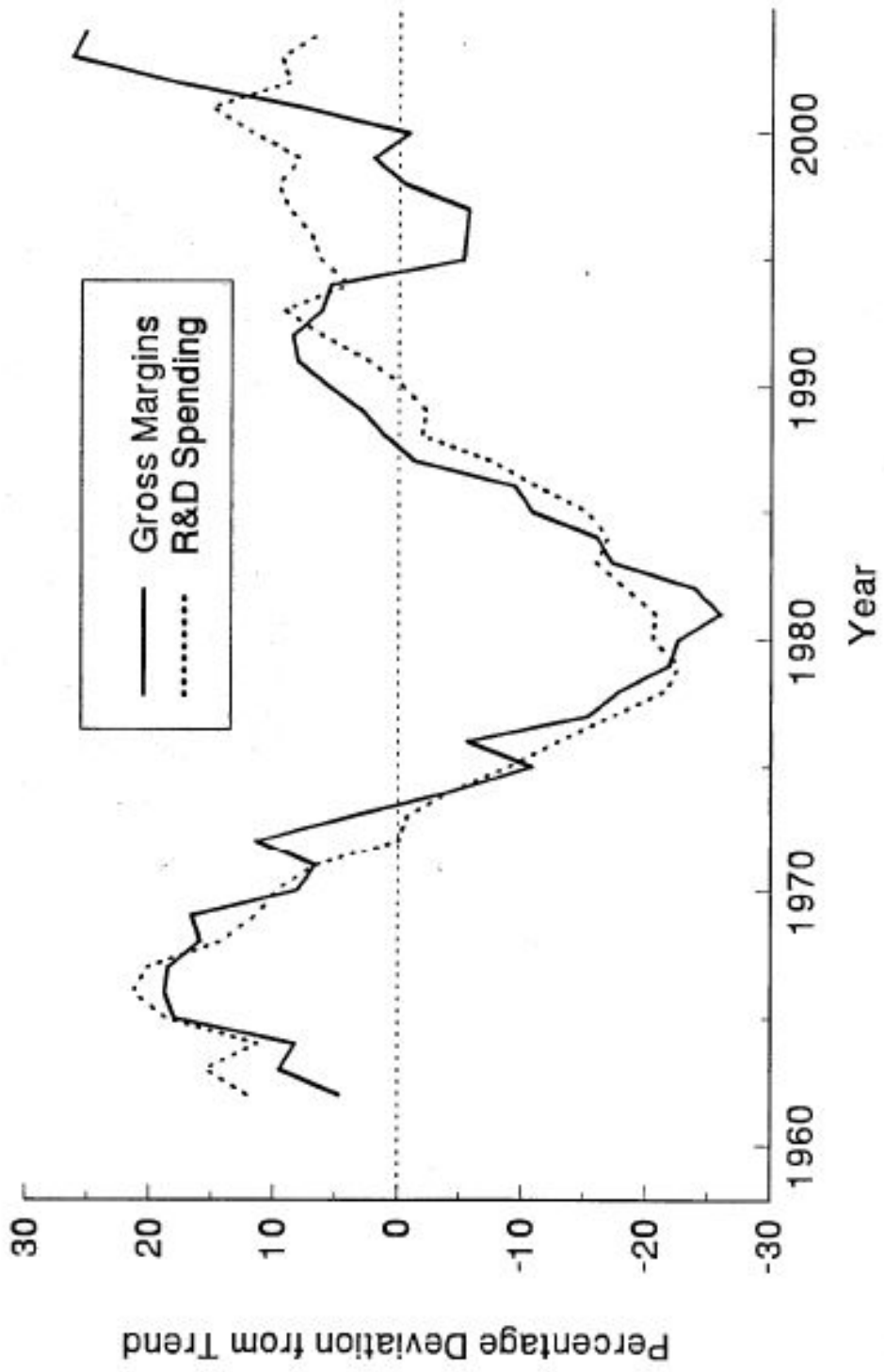
For at least the first three decades of the time series, the degree of coincidence is remarkable. When margins rise relative to trend, R&D rises in near tandem. The causation

39. Extended from Scherer (2001). See also Lichtenberg (2004), Scherer (1996, p. 388 note); and Grabowski et al. (2002).

40. A splice was necessary after 1996, when a new industrial classification was implemented and slight mismatches appeared between the new and old data.

41. R&D performed abroad by the foreign divisions of foreign-owned member companies was excluded.

Figure 5
Trend-Adjusted Movements of Pharmaceutical Margins and R&D



cannot plausibly run from R&D to margins, because R&D extends for a decade or so before marketing begins, and even then, it takes several years before sales and margins peak. The turning points are roughly coincident for the mid-1960s and the early 1990s, but R&D leads margins by three years for the early 1980s. Reconciliation might come from viewing margins as an imperfectly anticipated measure of profit expectations. Or companies might apply with some deviation a crude rule of thumb, raising R&D when margins increase and holding back its growth when their growth flags. The relationships appear to break down during the 1990s and early years of the new century. A possible explanation is that during the early 1990s, companies were under heavy pressure from the Clinton administration to curb their prices or face price controls. They argued against such policies, emphasizing the dependence of R&D investments on profits, and may have found it politic visibly to maintain R&D growth levels. Margin deviations rose sharply under a new and more conservative U.S. president, while a slight decline in R&D growth may have been due to disappointing new product approvals. Compare Figure 2.

Whatever the exact causal dynamics, two things appear clear: (1) Correctly accounting for R&D as a long-lived investment tends to reduce substantially, if not to eliminate altogether, the inference that pharmaceutical companies are on average achieving supra-normal profit returns. And (2), there are distinct links, both short-run and long-run, between gross margins and R&D investments. One possible theoretical explanation is that pharmaceutical companies are adhering to the Dorfman-Steiner (1954) theorem, which states that profit-maximizing investments in R&D (and also in drug promotion activity, generating nearly as much expenditure as R&D) are higher, the wider price-cost margins are. But under Dorfman-Steiner, one would not expect the nearly complete dissipation of profit margins for R&D and promotion.⁴² An explanation in better accord with the evidence and consistent with received theory⁴³ is that pharmaceutical companies engage in competitive rent-seeking behavior -- to be sure, of a virtuous character distinguishable

42. See also Stigler (1968). Compare Scherer (2004), which in a model with assumptions analogous to those of Dorfman-Steiner found equilibrium R&D to be approximately 29 percent of quasi-rents.

43. See especially Tullock (1967), Barzel (1968), and Krueger (1974).

from some of the early theories. That is, when rents (price-cost margins) are high, the companies compete vigorously to capture them by increasing their R&D (and promotional) outlays, and indeed, the companies compete so vigorously, there is little left over in the end for supra-normal profit. When rents decline, R&D outlays are also squeezed so that a competitive rate of return persists.

Exactly how this competitive rent-seeking evolves is left unclear. One possibility is that Firm A sees Firm B mounting an R&D project to develop new drug X, whereupon B initiates its own countervailing project to offer a variant of X and perhaps even to preempt B's innovation date. Cockburn and Henderson (1995) call such competition "racing" and, through interviews and an analysis of research focus data from pharmaceutical companies, find little support for it. Rather, they perceive investment decisions to be driven by the appearance of new technological opportunities and allocated among R&D laboratories on the basis of the firms' heterogeneous human capital capabilities. In this case, the more plausible chain of causation is that new science-based opportunities create profit potentials, and, recognizing them sooner or later, companies compete vigorously to exploit them, in the process dissipating most or all of the attainable rents.

Implications for Economic Welfare and Patent Policy

To the extent that these insights are anywhere near the mark, profound implications for welfare economics and the theory of patent policy follow.

A simple version of the situation is illustrated in Figure 6. Through research and development, a new product is created. Its existence gives rise to a new demand curve D^* that did not exist before, and through process R&D (which we have touched only lightly in previous sections) a (constant) marginal production cost function $M-MC^*$ appears. With a patent monopoly limited in time, the responsible firm maximizes its profits by setting marginal revenue MR equal to marginal cost, setting price OP^* and offering output OQ^* . During the on-patent period the firm realizes quasi-rents, also called producer's surplus,

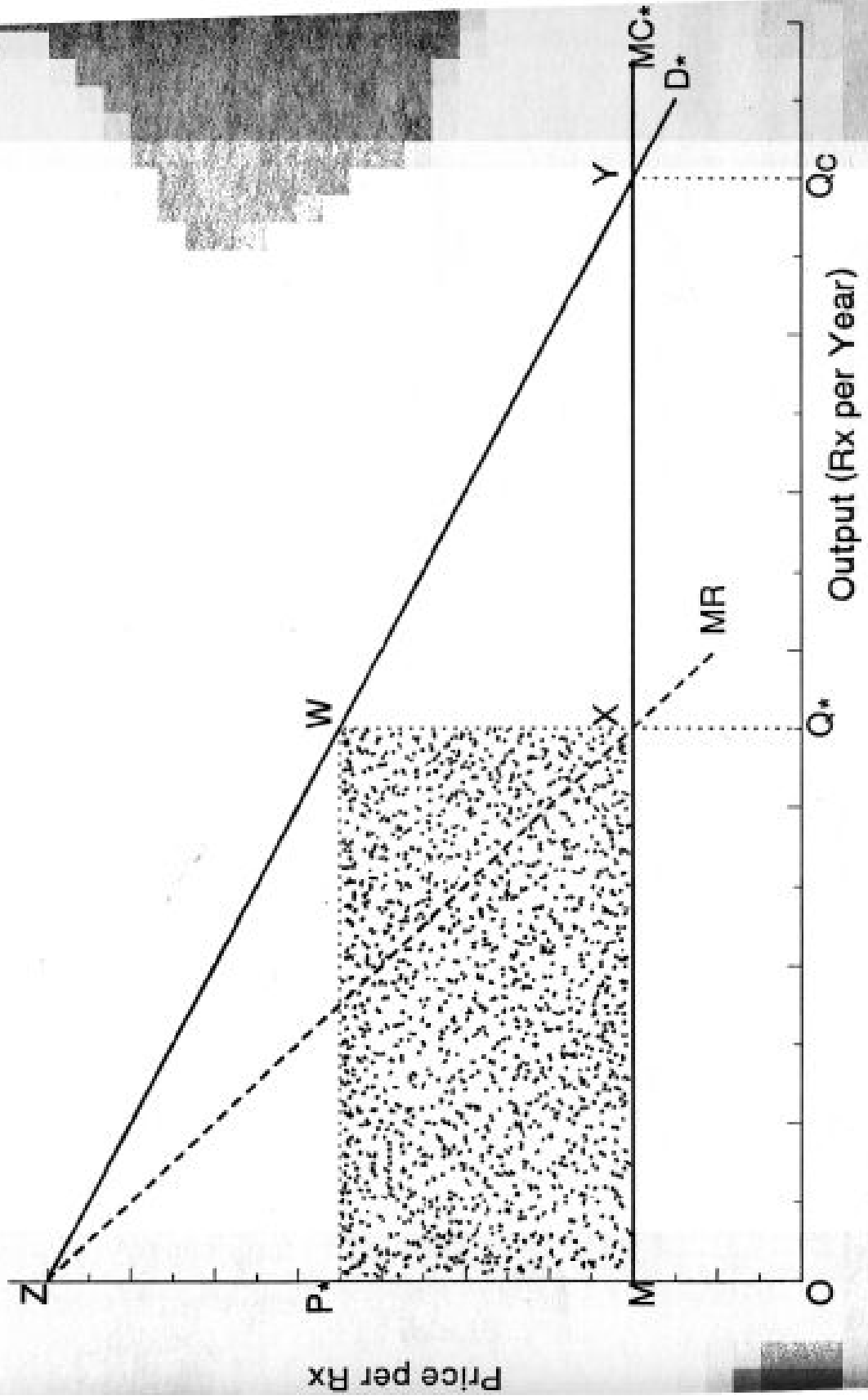
measured by the rectangular area P^*WXM . This producer's surplus represents a welfare gain, and its expectation motivates investment in forward-looking R&D. In addition, the availability of the new product yields consumers' surplus measured by the triangular area P^*WZ . When the patent expires, generic competition begins (before 1984, slowly), prices are driven to marginal cost OM , and output expands to OQ_c , giving rise to an additional surplus measured by triangle WYX and converting producers' surplus P^*WXM into consumers' surplus. Public policy limits patent lives in part because, if producer's surplus is sufficient to induce R&D, policy-makers do not want to delay indefinitely the realization of the consumers' surplus increment WYX . See Nordhaus (1969, pp. 70-90).

Now if competitive rent-seeking raises R&D costs to dissipate producers' surpluses totally, P^*WXM can no longer be counted as a welfare gain during the period of patent monopoly. Rather, the gain is offset by R&D cost (dotted area). After patents expire, rectangular surplus P^*WXM is not captured by product innovators, so it does not stimulate additional R&D. It becomes a pure surplus not offset by costs. Thus, the welfare gain pre-patent-expiry is only triangle ZWP^* rather than the larger trapezoid $ZWXM$, and after patent expiry, the incremental welfare gain (beyond ZWP^*) is trapezoid P^*WYM . The net welfare gain in a competitive rent-seeking context is smaller while the patent is in force and larger incrementally after the patent expires. This, as simpler models have shown in the past, leads to substantially shorter optimal patent lives -- e.g., in those models, as short as one year, compared to the much longer lives found in models without competitive rent dissipation.⁴⁴

Nor is this the end of the story. For what may be an extreme example, consider Figure 7, in which it is assumed that an efficacious patented product already exists as the scene unfolds. Its demand curve is D_1 and its marginal cost function $M-MC$. Profits are maximized at price OP_1 with output OQ_1 . Welfare surpluses are computed as in the previous example. Now let another firm develop a substantially superior product that (exaggerating) captures all the sales of the incumbent product

44. See especially McFetridge and Raffiquzzaman (1986) for an analysis that focuses on cost-reducing (process) innovations rather than product innovations. For a more complex variant covering the product innovation case, see Scherer (2004).

Figure 6
New Product Equilibrium with Rent-Seeking



and expands the overall demand in the relevant therapeutic category. For geometric simplicity, the demand shift is assumed to be in parallel to the original demand curve, i.e., to D_2 . Assume the same marginal cost MC . The newcomer monopolist maximizes its profits by setting the now-higher price OP_2 and offering expanded output OQ_2 .

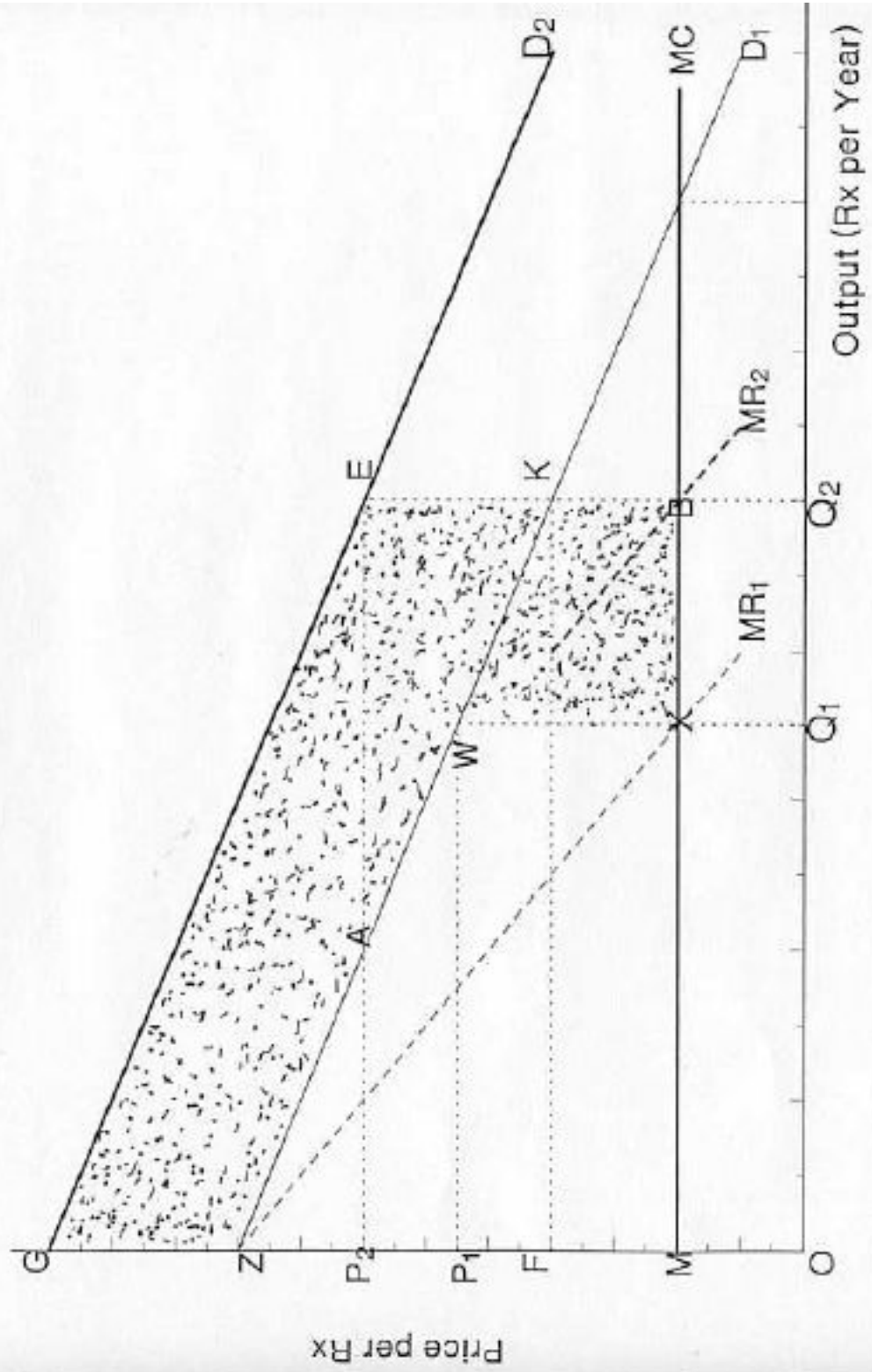
To analyze the welfare implications, we proceed in two stages. As a first approximation, ignoring competitive rent-seeking, producer's surplus is expanded from P_1WXM to P_2EBM , cannibalizing all of the original firm's producer's surplus and transforming P_2AWP_1 from consumers' surplus into producer's surplus. Consumers do gain from the improvement, however, adding consumers' surplus increment GEP_2 . Abstracting from cannibalized or transferred surplus, the net welfare gain is the lazy-L-shaped area $GEBXWZ$ (shaded with dots).

Now, however, we introduce competitive rent-seeking. The stimulus to investment during the new patent's life is the new producers's surplus rectangle P_2EBM . With perfect rent-seeking, that surplus will be dissipated by the R&D expenditure of firms seeking to displace Firm 1's product. It can no longer be counted as a welfare gain. Parallelogram $GEKZ$ was before rent-seeking net new surplus. Because of equal heights and equal bases, the area of parallelogram $GEKZ$ equals the area of rectangle P_2EKP_1 , all of which is dissipated through R&D costs. The remaining incremental welfare gain $WKBX$ is also dissipated through rent-seeking R&D costs. After cannibalization plus rent-seeking, the net welfare effect is negative! Only when patent protection ceases do incremental welfare gains appear. This again implies, given that rent-seeking investors are compensated competitively for their investments, that patent lives should not be protracted.

The implications of this analysis are perplexing. Taken at face value, they suggest that, to the extent that pharmaceutical R&D is indeed a virtuous competitive rent-seeking phenomenon, the welfare-enhancing virtue of the process is tarnished. And the case for exceptionally strong patent protection is weakened. The traditional wisdom might be redeemed through nonlinearities in the demand functions⁴⁵ or by solid evidence that rent-seeking

45. See Spence (1976).

Figure 7
Equilibrium with Cannibalization and Rent-Seeking



is less than fully competitive. At minimum, one must conclude that new theoretical and empirical perspectives are needed.

Developing New Drugs and Vaccines for Third World Diseases

Reverting to the simpler and less controversial assumption that pharmaceutical innovation is motivated by the lure of profits, a further dilemma presents itself. Rich consumers are able and willing to pay, either directly or through taxes and transfers, for an ample array of drugs to combat the diseases and debility afflicting them. For the consumers in nations with very low per-capita incomes, who tend to be concentrated in tropical areas harboring diseases such as malaria, sleeping sickness, and leishmaniasis seldom prevalent in the industrialized world, demand may be insufficient to yield quasi-rents inducing substantial investments in disease-alleviating R&D. A study by Medicins sans Frontieres (2001) found that among 1,393 new drug chemical entities introduced into world markets between 1975 and 1999, only 13 (or 15 counting tuberculosis drugs) were indicated for so-called "tropical" diseases. Clearly, the invisible hand falters in guiding research toward the needs of low-income populations.

There are several possible solutions. Prior to the Uruguay Round of international trade negotiations, concluded in 1994, many third-world nations (and some rich nations) did not offer patent protection on new pharmaceutical products. The resulting treaty required inter alia the provision of such patent rights in all World Trade Organization member nations by the year 2005 (later extended for the least-developed nations to 2016). One rationale was that this policy change would stimulate the development of medicines for tropical diseases, either by multinational pharmaceutical companies or enterprises based in low-income nations. (India, for example, was home to several of the world's leading generic drug suppliers.) Whether this strategy will work remains to be seen, but there are grounds for skepticism. See Lanjouw (1997, 2002). If under the logic of Figure 6, demand curves for drugs in low-income nations lie too close to marginal production cost functions, the pool of attainable quasi-rents will be too small to stimulate much development of tropical disease drugs by profit-seeking firms.

If private markets fail, a humanitarian case for governmental or philanthropic intervention exists. Governments and philanthropic agencies might intervene on either the supply

side or the demand side.

On the supply side, research and development on tropical drugs might be conducted in government or government-supported laboratories, or grants could be issued to private corporations to subsidize the development of tropical disease therapies. The U.S. Army's Walter Reed Hospital was once a leader in developing drugs to combat malaria and other tropical diseases. But as the desire to station American troops in tropical nations faded after the Vietnam War, so also did interest in developing such medicines. Thus, contracts and grants for altruistic motives remained the main supply-side recourse. Splendid work by the Gates Foundation, among others, has been done, but those activities, oriented thus far mainly toward basic research and therapeutic molecule discovery, are of too recent vintage to assess success. The alternative, especially when high-cost drug development and clinical testing stages are reached, is for governments to issue contracts to private enterprises -- presumably, the various pharmaceutical companies. Here the well-known agency-theoretic problems associated with national defense research and development contracting are encountered. Government agencies are not always adept at picking winning technological approaches, and indeed, given the uncertainties of drug discovery, one must be tolerant -- although legislatures seldom are -- of frequent failure. The choice problem is aggravated by the tendency of contract-seekers to exaggerate their chances of success at the early proposal stage and to underestimate the costs. Special contractual arrangements, such as cost-plus-fixed-fee contracts, may be necessary to transfer what would otherwise be unacceptable technological risks from private firms to the government sponsor. These often provide inadequate incentives for efficient operation and sustain other moral hazards. See Peck and Scherer (1962).

An alternative to intervention on the technology-push side is for government agencies to create special demand-side incentives for research and development. An interesting and attractive approach was the advance purchase approach to inducing new vaccine development endorsed by the G-8 nations in 2005 and 2006, but not yet funded at the time this essay was written. See Levine et al. (2005) and Berndt and Hurvitz (2005). Emphasizing the development of vaccines rather than traditional pharmaceuticals was attractive because vaccines can prevent disease through one or very few inoculations, whereas

treatment once a disease has taken hold often requires repeated and perhaps even life-long medical interventions that overstrain the healthcare delivery capabilities of low-income nations. One disadvantage of the vaccine approach is the particularly extensive and lengthy clinical testing required, since one cannot ethically tell in advance who would otherwise incur the target disease. The advantage of vaccines from the perspective of administration in low-income nations is a disadvantage for pharmaceutical companies, since each patient requires only one or a very few doses, which leaves much less demand than the demand for medicines that will be administered once a day for many days or even years. Recognizing these problems, the G-8 proposal identified three target diseases -- HIV/AIDS, malaria, and tuberculosis. For each, a generalized agreement to purchase 200 million doses at a prespecified subsidy of \$15 per dose would be announced, i.e., embodying a total commitment of \$3 billion per disease, paid only if the goal of successful new vaccine development were achieved. The purchase would be conditional upon the development of effective vaccines, with efficacy judged against standards articulated by a coordinating committee and by the national health authorities of nations administering the vaccine (which would add their own more modest subsidies to the purchase price). Quantities above 200 million would be procured at prices to be negotiated through a process that remained unclear at the time the proposal was approved. Because G-8 governments had not agreed to the required advance financial commitments at the time this essay was written, a judgment on the promise and success of the advance purchase commitment approach would be premature. What is clear is that an important market failure persists with respect to incentives for the discovery and development of therapies effective against diseases threatening one to two billion inhabitants of low-income nations, depending upon the disease.

Conclusion

The pharmaceutical industry provides a fascinating laboratory for studying what we know and what don't know about the economics of innovation. The industry has an extraordinary innovation record; it faces major risks and uncertainties in its efforts to solve new therapeutic problems; its links to academic science bases are unusually rich and deep; and the industry's responsiveness or lack thereof to economic stimuli is of considerable interest. That said, it must be admitted that

there is much we still do not understand about the pharmaceutical innovation process. As always, more work remains to be done.

References

James D. Adams and J. Roger Clemmons, "Industrial Scientific Discovery," working paper, Rensselaer Polytechnic Institute, October 2006.

Marcia Angell, The Truth About the Drug Companies (Random House: 2004).

Pierre Azoulay, "Acquiring Knowledge within and across Firm Boundaries: Evidence from Clinical Development," National Bureau of Economic Working Paper 10083 (November 2003).

Yoram Barzel, "Optimal Timing of Innovations," Review of Economics and Statistics, vol. 50 (August 1968), pp. 348-355.

Ronald S. Bond and David Lean, Sales, Promotion, and Product Differentiation in Two Prescription Drug Markets (Federal Trade Commission staff report: February 1977).

Lewis Branscomb and Philip Auerswald, Taking Technical Risks (MIT Press: 2001).

Geoffrey Carr, "The Pharmaceutical Industry," special survey, The Economist, February 21, 1998, pp. 1-18.

Harold M. Clymer, "The Changing Costs of Pharmaceutical Innovation," in Joseph Cooper, ed., The Economics of Drug Innovation (American University Center for the Study of Private Enterprise: 1970), pp. 109-124.

Iain M. Cockburn and Rebecca Henderson, "Racing To Invest? The Dynamics of Competition in Ethical Drug Discovery," Journal of Economics & Management Strategy, vol. 3 (Fall 1994), pp. 481-519.

Iain M. Cockburn and Rebecca Henderson, "Absorptive Capacity, Coauthoring Behavior, and the Organization of Research in Drug Discovery," Journal of Industrial Economics, vol. 46 (1998), pp. 157-182.

Iain M. Cockburn and Rebecca Henderson, "Publicly Funded Science and the Productivity of the Pharmaceutical Industry," Innovation Policy and the Economy, vol. 1 (2000), pp. 1-34.

Iain Cockburn, Rebecca Henderson, and Scott Stern, "The Diffusion of Science Driven Drug Discovery," National Bureau of Economic Research working paper 7359 (September 1999).

Joseph A. DiMasi, Ronald Hansen, Henry Grabowski, and Louis Lasagna, "Cost of Innovation in the Pharmaceutical Industry," Journal of Health Economics, vol. 10 (1991), pp. 107-142.

Joseph A. DiMasi, Ronald Hansen, Henry Grabowski, and Louis Lasagna, "Research and Development Costs for New Drugs by Therapeutic Category," PharmacoEconomics, vol. 7 (February 1995), pp. 152-169.

Joseph A. DiMasi, Ronald Hansen, and Henry Grabowski, "The Price of Innovation: New Estimates of Drug Development Costs," Journal of Health Economics, vol. 22 (2003), pp. 151-185.

Robert Dorfman and Peter O. Steiner, "Optimal Advertising and Optimal Quality," American Economic Review, vol. 44 (December 1954), pp. 826-836.

Alfonso Gambardella, Science and Innovation: The US Pharmaceutical Industry during the 1980s (Cambridge University Press: 1995).

Henry Grabowski and John M. Vernon, "A New Look at the Returns and Risks to Pharmaceutical R&D," Management Science, vol. 36 (July 1990), pp. 804-821.

Henry Grabowski, John M. Vernon, and Joseph DiMasi, "Returns on Research and Development for 1990s New Drug Introductions," PharmacoEconomics, vol. 20 Supplement 3 (2002), pp. 16-27.

Henry Grabowski, John M. Vernon, and Lacy Glenn Thomas, "Estimating the Effects of Regulation on Innovation: An International Comparative Analysis," Journal of Law & Economics, vol. 21 (April 1978), pp. 133-163.

Harbridge House Inc., Patent Policy Study: Task IV, Drug Study (Boston: May 1967).

Dietmar Harhoff and F. M. Scherer, "Technology Policy for a

World of Skew-distributed Outcomes," Research Policy, vol. 29 (April 2000), pp. 559-566.

Elaine M. Healy and Kenneth Kaitin, "The European Agency for the Evaluation of Medicinal Products' Centralized Procedure for Product Approval: Current Status," Drug Information Journal, vol. 33 (1999), pp. 969-978.

IIT Research Institute, Technology in Retrospect and Critical Events in Science (National Science Foundation: December 1968).

Institute of Medicine, National Academies of Science, The Future of Drug Safety (Washington: 2006).

Kenneth I. Kaitin and Joseph DiMasi, "Measuring the Pace of New Drug Development in the User Fee Era," Drug Information Journal, vol. 34 (2000), pp. 673-680.

David Kessler and Karyn Feiden, "Faster Evaluation of Vital Drugs," Scientific American, March 1995, pp. 48-54.

Hannah E. Kettler, "Pharmaceutical and Biotechnology Industries Coalescing," presentation at the American Enterprise Institute (October 2000).

William Kingston, "Antibiotics, Invention, and Innovation," Research Policy, vol. 29 (2000), pp. 679-710.

Edmund W. Kitch, "The Patent System and the New Drug Application," in R. L. Landau, ed., Regulating New Drugs (University of Chicago Center for Policy Study: 1973), pp. 81-108.

Anne O. Krueger, "The Political Economy of the Rent-Seeking Society," American Economic Review, vol. 64 (June 1974), pp. 291-303.

Jean O. Lanjouw, "The Introduction of Pharmaceutical Product Patents in India," working paper, NBER, 1997 [check number].

Jean O. Lanjouw, "Intellectual Property and the Availability of Pharmaceuticals in Poor Countries," Center for Global Development working paper no. 5 (April 2002).

Richard C. Levin, Alvin Klevorick, Richard R. Nelson, and Sidney Winter, "Appropriating the Returns from Industrial Research and Development," Brookings Papers on Economic Activity (1987, no. 3), pp. 783-820.

Ruth Levine, Michael Kremer, and Alice Albright, Making Markets for Vaccines: Ideas to Action (Center for Global Development: 2005).

Frank R. Lichtenberg, "The Benefits and Costs of Newer Drugs: Evidence from the 1996 Medical Expenditure Panel Survey," National Bureau of Economic Research working paper 8147 (March 2001).

Frank R. Lichtenberg, "The Impact of New Drug Launches on Longevity," National Bureau of Economic Research working paper 9754 (June 2003).

Frank R. Lichtenberg, "Public Policy and Innovation in the U.S. Pharmaceutical Industry," in Douglas Holtz-Eakin and Harvey Rosen, ed., Entrepreneurship and Public Policy (MIT Press: 2004).

Genia Long and six others, "The Impact of Hypertension Drugs on the Number and Risk of Deaths, Strokes, and Myocardial Infarctions," National Bureau of Economic working paper 12096 (March 2006).

Charles C. Mann and Mark L. Plummer, The Aspirin Wars (Harvard Business School Press: 1991).

Edwin Mansfield, "Discussion," in Joseph Cooper, ed., The Economics of Drug Innovation (American University: 1970), pp. 149-154.

Edwin Mansfield, "Patents and Innovation: An Empirical Study," Management Science, vol. 32 (February 1986), pp. 173-181.

Edwin Mansfield, "Academic Research and Industrial Innovation: An Update of Empirical Findings," Research Policy, vol. 26 (1998), pp. 773-776.

R. A. Maxwell and S. B. Eckhardt, Drug Discovery: A Case Book and Analysis (Humana Press: 1990).

D. G. McFetridge and M. Rafiquzzaman, "The Scope and Duration of the Patent Right and the Nature of Research Rivalry," Research in Law and Economics, vol. 8 (1986), pp. 91-129.

Medicins Sans Frontieres Access to Essential Medicines Campaign, Fatal Imbalance: The Crisis in Research and Development for Neglected Diseases (Geneva, Switzerland: September 2001).

William D. Nordhaus, Innovation, Growth, and Welfare (MIT Press: 1969).

Merton J. Peck and F. M. Scherer, The Weapons Acquisition Process: An Economic Analysis (Harvard Business School Division of Research: 1962).

Pharmaceutical Research and Manufacturers of America, Pharmaceutical Industry Profile: 2006 (Washington: 2006).

Walter N. Powell and Stine Grodal, "Networks and Innovators," Oxford Handbook of Innovation (Oxford University Press, 2005), pp. 56-85.

Janice M. Reichert and Christopher-Paul Milne, "Public and Private Sector Contributions to the Discovery and Development of 'Impact' Drugs," Tufts University Center for the Study of Drug Development White Paper (May 2002).

William T. Robinson and Claes Fornell, "Sources of Market Pioneer Advantages in Consumer Goods Industries," Journal of Marketing Research, vol. 22 (August 1985), pp. 305-317.

F. M. Scherer, "Time-Cost Tradeoffs in Uncertain Empirical Research Projects," Naval Research Logistics Quarterly, vol. 12 (March 1966), pp. 71-82.

F. M. Scherer, Industry Structure, Strategy, and Public Policy (HarperCollins: 1996).

F. M. Scherer, "The Link Between Gross Profitability and

Pharmaceutical R&D Spending," Health Affairs, vol. 20 (September 2001), pp. 216-220.

F. M. Scherer, "A Note on Global Welfare in Pharmaceutical Patenting," The World Economy, vol. 27 (July 2004), pp. 1127-1142.

F. M. Scherer, "The Political Economy of Patent Policy Reform in the United States," working paper, 2006.

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

David Schwartzman, "Has Pharmaceutical Research Become More Scientific?" Review of Industrial Organization, vol. 11 (December 1996), pp. 841-851.

Sheila Shulman and Michael Mannocchia, "The US Orphan Drug Programme: 1983-1995," PharmacoEconomics, September 1997, pp. 312-326.

John Simons, "The \$10 Billion Pill," Fortune, January 20, 2003, pp. -----.

A. Michael Spence, "Product Selection, Fixed Costs, and Monopolistic Competition," Review of Economic Studies, vol. 43 (June 1976), pp. 217-235.

Thomas R. Stauffer, "The Measurement of Corporate Rates of Return: A Generalized Formulation," Bell Journal of Economics and Management Science, vol. 2 (Autumn 1971), pp. 434-469.

George J. Stigler, "Price and Non-Price Competition," Journal of Political Economy, vol. 72 (January-February 1968), pp. 149-154.

Tufts Center for the Study of Drug Development, Impact Report, vol. 8 (May/June 2006).

Gordon Tullock, "The Welfare Costs of Tariffs, Monopolies and Theft," Western Economic Journal, vol. 5 (June 1967), pp. 224-232.

U.S. Federal Trade Commission, Economic Report on

Antibiotics Manufacture (Washington: 1958).

U.S. Federal Trade Commission, Statistical Report: Annual Line of Business Report: 1977 (Washington: April 1985).

U.S. Food and Drug Administration, Pilot Drug Evaluation Staff, "The 007 Experiment - An Approach to Change," unpublished internal assessment, February 25, 1993.

U.S. Office of Technology Assessment, Pharmaceutical R&D: Costs, Risks, and Rewards (Washington: 1993).

John A. Vernon and W. K. HUGHEN, "The Future of Drug Development: The Economics of Pharmagenomics," National Bureau of Economic Research working paper 11875 (December 2005).

Barry Werth, The Billion Dollar Molecule (Simon & Schuster: 1994).

Lynne Zucker, Michael Darby, and J. Armstrong, "Geographically Localized Knowledge: Spillovers or Markets?" Economic Inquiry, vol. 36 (1998), pp. 65-86.

15,775 words