Faculty Research Working Papers Series

Markets and Uncertainty in Pharmaceutical Development

F. M. Scherer
John F. Kennedy School of Government – Harvard University

September 2007

RWP07-039

MARKETS AND UNCERTAINTY IN PHARMACEUTICAL DEVELOPMENT*

F. M. Scherer
Harvard University
Revised September 2007

The origins of the modern pharmaceutical industry can be traced to Germany: to Bayer AG's development of aspirin in the late 1890s; to the work of Dr. Paul Ehrlich, among other things in discovering in 1909, at the Institut für experimentelle Therapie, the first medicine effective against syphilis; and at I.G. Farben with the discovery in the 1930s of sulpha drugs. Work leading to new pharmaceuticals spread to other nations, and during World War II, a quiescent U.S. ethical drug industry grew explosively after military authorities issued contracts to 20 companies to produce penicillin, discovered and then rediscovered by British researchers during the 1920s and 1930s. Since then, the development and production of new pharmaceutical entities has been mainly the province of industrial firms, often working on a more or less worldwide scale with, or following up efforts by, academic and medical institution researchers.

This paper addresses two main questions: how the search for new pharmaceutical entities by profit-seeking enterprises is guided by market forces in tandem with advances in medical science, and how organizations advancing the state of the art in pharmaceuticals cope with the uncertainties they face in performing their important work.

1. The Profit Lure

Private companies invest in research and development with the hope of creating new products that, protected by patents and other first-mover advantages, will have at least a temporary monopoly or differentiation advantage and hence will yield quasi-rents, defined as revenues in excess of variable production and distribution costs. The standard economic model is reflected with considerable simplification and without an attempt to approximate real-world values, in Figure 1. There is a demand curve (solid

*This paper is written for the conference on The Global Organization of Biomedical Innovation, Kiel, Germany, October 4-8, 2007.

line) for some new therapeutic contribution. Process development defines the dashed marginal cost function MC. With at least a partial monopoly, the responsible firm maximizes its profits by setting marginal cost equal to marginal revenue MR, producing an output of approximately 39 million pills per year and selling them at $8.40 per ten pills. Quasi-rents -- i.e., the net left over after payment of variable costs to meet fixed costs and reward R&D investments – are roughly $25 million per year, as shown by the horizontally shaded rectangle (i.e., $0.64 per pill times 39 million units).

A diversity of things can go wrong with this scenario. For one, most nations impose price controls on pharmaceutical products. Suppose price controllers hold the allowable wholesale price at $6.50, i.e., 23 percent lower than the profit-maximizing price. Then quasi-rents will be reduced, and the stimulus to R&D, if anticipated, will be weaker. In the case shown the shortfall is small -- annual quasi-rents (the vertically shaded rectangle) are reduced by only $1.6 million because of the large assumed increase in the quantity demanded. With a different demand configuration, e.g. concave downward, or more severe price restraints, the quasi-rent shortfall could be considerably greater.

A more troubling case is illustrated in Figure 2. It assumes that the drug is targeted for markets in which most of the consumers have much lower disposable income, and/or much weaker insurance coverage, than citizens in the rich industrialized nations. This "income effect" shifts the demand curve downward relative to Figure 1. The best the profit-maximizing drug producer can do in this market is a price of $2.95 per ten pills, eliciting demand of 14 million pills annually. Total quasi-rents are reduced from $25 million per year in Figure 1 to $1.33 million per year. It is well known that profit stimuli are insufficient to elicit much in the way of new drug development for diseases mainly found in low-income nations.[1] Standard market forces fail. R&D subsidies from governments or charitable agencies must fill the gap, if it is to be filled. Even for drugs combatting ubiquitous diseases, whose development is induced by the lure of profits in rich nations, the constricted purchasing power in low-income nations can severely limit the quantity

1 . See Medicins sans Frontieres Access to Essential Medicines Campaign and Drugs for Neglected Diseases Working Group, Fatal Imbalance: The Crisis in Research and Development for Neglected Diseases. Geneva, September 2001.

demanded and consumed. If differential pricing is practiced[2] and the medicine is made available at marginal cost, only 30 million pills are demanded annually in the market whose conditions are described by Figure 2. Production subsidies can help alleviate the market failure in this instance.

2. The Theory of R&D Inducement

Economists have conceived richly elaborated models of how the lure of profits in a market economy induces private enterprises to invest in research and the development of new products. As a framework for subsequent insights, it is useful to present here a bare-bones version. Figure 3 provides an introductory case. R&D is supported by private firms only if it is expected to yield discounted quasi-rents in excess of R&D costs. The quasi-rents, i.e., the horizontally shaded rectangle in Figure 1, will continue to flow in from the time of initial marketing until shortly after patents expire. We assume this period to be 12 years and the firm's time discount rate to be 10 percent. The discounted present value of a quasi-rent stream starting at year 0 and continuing through year 12 is approximately $250 million. This is the starting point of the solid quasi-rent curve shown in Figure 3. However, as population grows and incomes rise, all else equal, the quasi-rent potential for a given drug is likely to increase relative to its year-zero value. We assume here an average real (i.e., price level-adjusted) quasi-rent growth rate of 4.2 percent per year -- the value found in a study of U.S. pharmaceutical industry profitability over the years 1962 through 1996.[3] This growth causes the quasi-rent function in Figure 3 to have an upward slope.

On the other side of the inducement mechanism equation is the trend in research and development costs required to discover and test a successful new pharmaceutical entity. One would normally expect this cost to decline over time with advances in knowledge. There have been astonishing advances in the knowledge base useful for predicting which molecules will have therapeutic action in the human body and in narrowing the search to a relatively few promising

---

2 . See e.g. F. M. Scherer and Jayashree Watal, "Post-TRIPS Options for Access to Patented Medicines in Developing Nations," Journal of International Economic Law, vol. 5 (December 2002), pp. 913-940.

3 . F. M. Scherer, "The Link Between Gross Profitability and Pharmaceutical R&D Spending," Health Affairs, vol. 20 (September/October 2001), p. 217.

targets.[4] At some early moment in time, the science base is so poorly defined that development of the new drug may be impossible, i.e., infinitely expensive. Figure 3 assumes the less extreme case of a $300 million development cost at year 0 (assumed for counter-factual simplicity to be incurred instantaneously). As knowledge advances, the search can be more narrowly targeted, with fewer false starts and better laboratory equipment. The rate of advance, i.e., real R&D cost decline, is arbitrarily assumed in Figure 3 to be 3 percent per year.

Up to year 7.5, costs are greater than quasi-rents (payoffs), and so no investment in development will be induced. Breakeven -- R&D costs equal to discounted quasi-rents -- occurs at year 7.5. After that, costs fall relative to quasi-rents, so the development becomes increasingly profitable (and hence increasingly likely).[5] The investment-inducing crossing of the quasi-rent and R&D cost curves follows Alfred Marshall's metaphor of how supply and demand curves determine a product's price in the market, each component acting as if it were a scissors blade. Both supply (R&D costs) and demand (quasi-rent potential) must be propitious to induce investment in new technology by profit-seeking private enterprises.

This simple characterization is in several respects too simple and unrealistic. One immediate caveat is the assumption that R&D costs decline over time. Modern drug discovery is driven by advances in science, but to bring a drug to market, the entity must be clinically tested to the satisfaction of national or supra-national drug regulators. And there is compelling evidence that testing costs have been rising, not falling, over the past three decades -- perhaps by as much as 10 percent per annum.[6] If costs rise more rapidly than expected quasi-rents, the model breaks down and the implied market fails. Although there are abundant concerns that the pace of new drug development has been faltering during the past decade, it is clear that discovery and development do continue -- quite possibly because discovery

---

4 . Perspective on this point is obtained by comparing the article on "Pharmacology" in the eleventh edition of the Encyclopaedia Britannica (1911), vol. 21, pp. 347-353, with the Roche Diagnostics chart of human biochemical pathways, which will be presented as an unnumbered exhibit at the conference.

5 . See F. M. Scherer, "The Economics of Innovation and Technological Change," in the International Encyclopedia of the Social & Behavioral Sciences (2001), vol. 11, pp. 7350-6.

6 . From calculations in F. M. Scherer, "Pharmaceutical Innovation," forthcoming, drawing upon Tufts University survey data compiled by Joseph DiMasi, Ronald Hansen, and Henry Grabowski.

costs, once equal in magnitude relative to testing costs, have been falling over time relative to testing costs.

Alfred Marshall famously insisted (in the frontispiece to his Principles, 1890 et seq.) that *Natura non facit saltum* (nature does not make leaps). In this he was wrong, at least for the economics of research and development. Figures 4 and 5 lay out the two principal cases. In Figure 4, a scientific breakthrough around year 4 dramatically reduces the cost of developing the relevant new drug, so that development becomes profitable by year 5, in contrast to the year 7.5 breakeven with continuous demand and cost changes in Figure 3. Developments clearly induced on one side of the Marshallian scissors in this way are called "science-push" innovations, or in fields of technology more mechanical than pharmaceuticals, "technology-push" innovations. In Figure 5, the innovation-inducing change takes place on the demand side, e.g., as the emergence of a new disease such as HIV-AIDS or avian influenza greatly increases the demand for new pharmaceuticals. Such cases are called "demand-pull" inducements. Together the concepts focus our analysis on how investments in innovation that were at one time not profitable can with changes over time become financially attractive.[7]

An important implication of the model must now be clarified. As R&D costs decline, gradually or abruptly, and as the quasi-rent potential for an innovation rises over time, the diagrams postulate a unique point in time where "breakeven" occurs. At the breakeven date, R&D costs fall into equality with discounted expected quasi-rents, and for the first time the development is profitable. We assume provisionally -- with amendments to follow -- that sinking a large lump of resources into research and development is done by a single firm -- a monopolist, so to speak. But a firm with a true monopoly position would not want to innovate at the breakeven date! Instead, it would wait until costs fall and/or discounted quasi-rents rise, so that it gains a surplus of quasi-rents over R&D costs -- a monopoly profit. Indeed, it would want to wait until the discounted present value of that surplus is maximized -- a waiting period that under plausible numerical assumptions could be as long as ten years.[8] So two important questions arise: First, does innovation take place at or

---

7 . The "demand pull" and "technology push" schema was first crystallized by Jacob Schmookler, Invention and Economic Growth (Harvard University Press: 1967). Up to that time, economists tended to emphasize technology-push inducements.

8 . The initial version of this theory was advanced by Yoram Barzel, "Optimal Timing of Innovations," Review of Economics and Statistics, vol. 50 (August 1968), pp. 348-355. It was extended inter alia by me, e.g., in "Schumpeter and the Micro-Foundations of Endogenous

near the breakeven time, or is it substantially delayed relative to that time? And second, recognizing that innovation confers externalities upon members of society other than the innovator, i.e., that social benefits can exceed private quasi-rents by factors of two or more, what is the socially optimal time for innovation? Is it later or earlier than the breakeven time?

Yoram Barzel pioneered the analysis of the first question and suggested an approach to the second.[9] He assumed that the first firm to come up with a successful innovation obtained a monopoly on the early production and sale of the relevant product.[10] Monopoly ensued once the product was developed and marketing commenced. But before then, firms competed -- one might say, raced -- for that monopoly position. And as they competed, accelerating the completion date and hence increasing the cost of R&D and possibly also duplicating each others' R&D costs, they dissipated the surplus of quasi-rents over costs until no surplus was left, leaving zero net profits in equilibrium, analogous to the zero-profit equilibrium long assumed in models of price rather than new product competition. The Barzel innovation model was one case in a broader set of so-called rent-seeking models, except that in R&D races, consumers gain from earlier availability of the new product.[11] Thus, Barzel analyzed what might be called "virtuous rent-seeking," clearly distinguishable from the waste that characterized other contemporary models of rent-seeking competition for price-raising monopolies and the like.[12]

Growth," in Horst Hanusch and Andreas Pyka, eds., The Elgar Companion on Neo-Schumpeterian Innovation (Edward Elgar: 2007).

9 . Note 8 supra.

10 . In an analysis that probably induced Barzel's contribution, I argued that new markets were shared by possibly multiple innovators, with the share obtained by each rival proportional to the size of its head start into the market. I believe that assumption is more realistic in most, but not all, actual cases. See F. M. Scherer, "Research and Development Resource Allocation under Rivalry," Quarterly Journal of Economics, vol. 81 (August 1967), pp. 395-394.

11 . Both Barzel, supra note 8, and I analyze whether the zero-profit rent-seeking equilibrium to an R&D race might approach social optimality, after external benefits not captured by the innovator are taken into account. For plausible parameter constellations, rent-seeking can come close to social optimality, assuming no unwarranted duplication of R&D investments. On this, more later.

12 . See e.g. Gordon Tullock, "The Welfare Costs of Tariffs, Monopolies and Theft," Western Economic Journal, vol. 5 (June 1967), pp. 224-232; and Anne Krueger, "The Political Economy of the Rent-Seeking Society," American Economic Review, vol. 64 (June 1974), pp. 291-303.

3. The Profitability of Drug Innovation

We pause now for a reality check on our conceptual schema. Does anything remotely resembling a zero-net-profit equilibrium emerge as firms compete for leading positions in new pharmaceutical markets? And if it does, what are the dynamics by which the equilibrium is approached? And what are the implications for judging how well the pharmaceutical industry contributes in the net to human welfare?

My principal evidence on this point comes from the pharmaceutical and biopharmaceutical industries of the United States -- by far the largest single national market for modern pharmaceuticals, and one served inter alia by numerous companies whose home bases are outside the United States. Thus, although the full international picture will not be portrayed, an important and in many respects representative view can be had.

There is a long tradition, dating back to the investigations by Senator Estes Kefauver during the late 1950s, asserting that drug makers depart widely from the notion of a zero-profit competitive equilibrium. Year after year, pharmaceutical companies ranked at or near the top of the annual tabulations published by Fortune magazine analyzing such profit variables as the return on stockholders' equity. In the 39 years between 1968 and 2006, pharmaceuticals ranked either first or second 27 times among from 22 to 50 broad industry groups in its after-tax returns on stockholders' equity. Charges of profiteering surfaced anew in the early 1990s while Hillary Clinton led efforts to reform U.S. health care institutions and threatened to impose price controls upon the principal ethical drug producers. Those charges led to a thorough critical analysis of drug makers' defense that the standard measures of profitability were biased, overstating their true returns. That investigation, by the U.S. Congress' Office of Technology Assessment, concluded that there was in fact an upward bias in published profit ratios. Once appropriate accounting adjustments were made, pharmaceutical companies enjoyed returns on investment only two or three percentage points higher on average than their real (i.e., inflation-adjusted) cost of capital, approximating 10 percent.[13] And at least part of that differential, separate evidence suggests, is attributable to unusual non-

---

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

diversifiable risks.[14] The reason why true returns are much lower than those reported in formal accounting statements resides mainly in the fact that under standard accounting principles, R&D expenditures are written off annually as a current expense, even though in truth they are a long-term investment with payoff periods of from 10 to 20 years. They should ideally be capitalized and then depreciated over an appropriate life span. Current R&D expensing severely understates the denominator of profit/assets or profit/ stockholders' equity calculations while, at least given the gross profit rates and growth rates experienced in pharamaceuticals during recent decades, the profit numerator is understated less severely under expensing.[15]

That investment in pharmaceutical development is in fact a rent-seeking activity is suggested strongly by an analysis of U.S. data spanning the years 1962 through 2004.[16] Two variables were defined: R&D expenditures in the United States by members of the Pharmaceutical Research and Manufacturers of America (PhRMA) and gross margins realized by the entire U.S. pharmaceutical manufacturing industry, as defined by the Census Bureau. The gross margins are equivalent, aggregated up to the industry level, to the horizontally shaded quasi-rent area in Figure 1. Both variables were deflated to constant purchasing power using the gross domestic product deflator. Tightly-fitting exponential trends in both time series were estimated, and percentage deviations from those trends were then computed. The result is shown in Figure 6. There are distinct cycles in trend deviations for both variables. When gross margins (quasi-rents) rise relative to trend, so also do expenditures on pharmaceutical R&D; when the margins decline

---

14 . See Dietmar Harhoff and F. M. Scherer, "Technology Policy for a World of Skew-Distributed Outcomes," Research Policy, vol. 29 (April 2000), pp. 559-566.

15 . For a simple numerical example, see F. M. Scherer, "Pricing, Profits, and Technological Progress in the Pharmaceutical Industry," Journal of Economic Perspectives, Summer 1993, pp. 97-116. The definitive conceptual source is Thomas R. Stauffer, "The Measurement of Corporate Rates of Return," Bell Journal of Economics and Management Science, vol. 2 (Autumn 1971), pp. 434-469.

16 . See F. M. Scherer, "Pharmaceutical Innovation," paper written for the forthcoming Handbook on the Economics of Technical Change, updating Scherer, "The Link Between Gross Profitability and R&D Spending," supra note 3. See also Frank Lichtenberg, "Public Policy and Innovation in the U.S. Pharmaceutical Industry," in Douglas Holz-Eakin and Harvey Rosen, eds., Entrepreneurship and Public Policy (MIT Press: 2004); and Carmelo Giocotto et al., "Drug Prices and Research and Development Behavior in the Pharmaceutical Industry," Journal of Law & Economics, vol. 48 (April 2005), pp. 195-214.

relative to trend, so do the R&D outlays.[17] The principal break in the pattern occurs during the 1990s, at the time Mrs. Clinton's health care reform task force was threatening pharmaceutical price controls and the government extracted from drug-makers a commitment not to increase their prices at a more rapid rate than the general pace of inflation. Gross margins dipped relative to trend. A possible interpretation of the continuing high levels of R&D investment is that pharmaceutical companies were defending their pricing in public fora, claiming that high prices led to high R&D. To cut R&D at such a time would at least have presented a bad public image.

Gross margins in the pharmaceutical industry -- i.e. roughly, quasi-rents as a percentage of sales -- are extraordinarily high relative to the comparable ratio for all U.S. manufacturers. In 1987, for example, they amounted to 60 percent as compared to the all-manufacturing average of 30.5 percent. But the escalation in tandem of R&D costs (and also advertising and direct sales promotion costs) tends to dissipate those margins, leaving net profit returns only slightly above the rates associated with industries of similar risk. Thus, an all-industry equilibrium analogous to the 7.5 year breakeven point in Figure 3 is approximated.

Although no exactly similar study is known for the biotech cousins of pharmaceutical manufacturers, the dissipation of "rents" appears from exhaustive tabulations by Gary Pisano to be even more extreme.[18] Figure 7 reproduces his Figure 6-2, showing for all U.S. biotech companies with publicly traded securities (more than 200 cumulatively) their total sales and operating income before depreciation. Both sales and operating income are tiny during the industry's infancy in the late 1970s and 1980s. Sales then exploded, but operating profits remained in negative territory until 2003. And if one extremely profitable company -- Amgen -- is excluded from the analysis, the remainder of the industry as a whole remained unprofitable in 2003 and 2004. A few other firms, to be sure, had positive cash flows. Their returns were outweighed by the large losses, attributable mainly to high R&D outlays with little or no sales revenue, of the vast majority. Pisano reports that among the companies generating positive cash flows, a mere 15 accounted for

---

17 . The sharp upturn in the 1980s is probably attributable to the adoption of a research strategy innovation, "rational drug design," by many drug companies, and by rapidly increasing coverage of consumers' pharmaceutical outlays by health care insurers, which reduced elasticities of demand and, by a well-known theorem in monopoly pricing, led to increased margins.

18 . Gary P. Pisano, Science Business: The Promise, the Reality, and the Future of Biotech (Harvard Business School Press: 2006), especially p. 115.

93 percent of the above-zero returns.

In sum, the broad tendency for profit-seeking pharmaceutical companies appears to be the realization of returns on investment only slightly above normal, and for the biopharmaceutical industry as a whole, returns have tended on average to be negative. This raises several new questions. First, how does the dissipation process operate? What competitive instincts lead companies to invest so much, notably in R&D, that overall returns on investment are only modest? Second, what are the implications of such behavior for economic welfare? Is virtuous rent-seeking through R&D investment a good thing for consumers or a wasteful activity? And third, why do private enterprises continue to invest and indeed increase their investments despite the absence (at least on average) of bonanza profit returns?

I approach this formidable set of questions in three stages. First, we will investigate the role of uncertainty in pharmaceutical R&D and the strategies evolved in a market context to deal with uncertainty. A strong distinction will be made between supply-side, i.e., scientific and technological, and demand-side uncertainties. Second, a speculative hypothesis is offered for the tendency of investors, at least in the United States, to place substantial bets on highly uncertain research and development prospects. Third, we consider cross-cannibalization of revenues as firms compete for position within a particular therapeutic sub-market.

4. Uncertainty and Risk-Hedging

To suggest that investing R&D funds to develop new products is uncertain is to be trite. A more nuanced view is required. Some R&D activities are much riskier than others.[19] I would assert without solid support that pharmaceutical R&D (along with biopharmaceutical R&D) is among the riskiest innovative activities, along with investment in new airliners, in the domain of product research and development.

It is conventional to divide pharmaceutical R&D into three more or less sequential but sometimes overlapping time stages: discovery research, pre-clinical formulation and testing, and clinical trials (often paralleled by production process development efforts). (R&D activities that follow the introduction of an approved new product will be ignored.) The discovery stage can extend from as long as 20

---

19 . For pioneering research on industrial R&D risks, see Edwin Mansfield, Industrial Research and Technological Innovation (Norton: 1968), especially Chapter 3; and Mansfield et al., Research and Innovation in the Modern Corporation (Norton: 1971).

years to as little as a year. Pre-clinical testing, extending over one to a few years, takes molecules originating from the discovery stage and tests them in animals (ranging from mice to Guinea pigs and monkeys) for possible therapeutic efficacy and toxicity. Clinical trials, occurring over periods of from three to ten years, test candidate molecules for efficacy and safety in samples of human beings, with the number of subjects escalating from a few at the outset to as many as thousands or even tens of thousands as the trials progress. Each stage has unique risk profiles.

In the early decades of pharmaceutical industry evolution, most discovery research was characterized as more or less random screening, also called the "try every bottle on the shelf" approach.[20] Large numbers of substances occurring in nature and, more often, those synthesized in academic and industry laboratories, were tested at relatively low cost per test for possible pharmacological activity in vitro and in such animals as earthworms and mice. Those that gave promise of favorable activity were carried into pre-clinical animal tests. David Schwartzman estimates that in 1970, 126,000 substances were tested for pharmacological action by U.S. pharmaceutical companies. The likelihood of a marketable drug coming from any particular analysis was apparently on the order of one in several thousand. In the late 1970s and early 1980s, medical science (much of it originating in academic institutions) had advanced to the point that many therapeutic processes could be associated with specific cells, enzymes, or proteins in the human body. Candidate molecules could be synthesized to interact with those specific targets. This was the advent of so-called rational drug design. It eliminated the need for much random screening, but understanding of human biological functions was insufficient to tell in advance which specific molecules would have the desired effects. In one outstanding example of rational drug design, a Cambridge MA biotech company explored 367 different molecules before finding one that bound to its target in the hoped-for way.[21] The advent of recombinant gene-splicing and other techniques of the biotech revolution taking root in the 1980s reduced the fraction of dead ends even more, especially for molecules such as artificial insulin or erythropoietin, where the task was mainly to find a way to synthesize substances known to have desirable effects in the human body, but whose absence caused disease and debility. In other biotech projects, however, considerable trial and error remained the norm.

---

20 . This paragraph summarizes a more detailed analysis in Scherer, "Pharmaceutical Innovation," supra note 16, which in turn assesses a huge parent literature.

21 . See Barry Werth, The Billion Dollar Molecule (Simon & Schuster: 1994), p. 251.

There is little systematic evidence on the success rates in finding through pre-clinical animal tests molecules with enough promise to be carried into costly human clinical testing. The only known quantitative evidence is from a study of biopharmaceuticals. and for it, there are appreciable interpretation difficulties. Mark Struck reports a 57 percent rate of transition from preclinical to early clinical trials, but his detailed Table 2 suggests a success rate nearer 18 to 19 percent.[22] He asserts that success transitions were somewhat lower in conventional (so-called small-molecule) pharmaceutical tests, but again, our ignorance here is considerable.

For small-molecule pharmaceuticals, the evidence is much more solid, at least for drugs tested in the United States. For drugs advanced into clinical testing during the 1970s, 23 percent emerged successfully with marketing approval; for molecules tested in the 1980s, the reported success rate was slightly lower at 21.5 percent.[23] Thus, only one out of four or five molecules put into expensive clinical testing succeeds in reaching the market. The out-of-pocket R&D cost per success under 1980s conditions has been estimated by DiMasi et al. as roughly $400 million (including the cost of failed drug candidates).

Clearly, at both the discovery stage and in clinical testing, success is much rarer than failure. And the costs are substantial. The strategies chosen to cope with these risks and uncertainties are therefore important.

When it is so difficult to identify in advance which of many possible molecules will be therapeutically interesting and, eventually, therapeutically successful, a rational strategy is to pursue multiple research and development paths, in parallel or in series or in some combination of the two.[24] Individual companies clearly do this in their pre-clinical research: they investigate hundreds or even thousands of target molecules in the quest for candidates with interesting potential therapeutic

---

22 . Mark M. Struck, "Biopharmaceutical R&D Success Rates and Development Times," Bio/Technology, vol. 12 (July 12, 1994), p. 676.

23 . See Joseph A. DiMasi et al., "Cost of Innovation in the Pharmaceutical Industry," Journal of Health Economics, vol. 10 (1991), pp. 107-142; and DiMasi et al., "The Price of Innovation: New Estimates of Drug Development Costs," Journal of Health Economics, vol. 22 (2003), pp. 151-185.

24 . See F. M. Scherer, "Parallel R&D Paths Revisited," working paper, Harvard University, July 2007.

effects. It is less common for them to bring into costly human testing more than one candidate drug at a time. However, the market does so: it is conventional for several firms more or less simultaneously to be pursuing different leads in clinical testing, propelled in the race by a new medical threat (e.g., the emergence of AIDS or the rising incidence of Alzheimer's disease) or scientific advances that suggest interesting new lines of therapy. In the most interesting analysis of this point, Joseph Di Masi and Cherie Paquette found that 72 first-in-class drugs approved in the United States between 1960 and 1998 were followed by at least 235 new drugs in the same narrow therapeutic categories by the year 2003.[25] And especially for later drug cohorts, the evidence pointed strongly toward parallel development and testing of drugs. Thus, in the 1990s, the average lag between the pioneer (i.e., the winner of the race) and the first follower was 2.25 years -- a period much too short for the follower to have initiated its R&D project only after observing the first-mover's success. The third mover followed the second during the 1990s by 2.5 years on average; the fourth mover followed the third by 1.4 years. It seems virtually certain that parallel clinical testing paths were being pursued -- perhaps but not necessarily inadvertently.

I am unable to say that such parallel paths behavior in the discovery and testing of new drugs is optimal. No one knows what is optimal in such a complex and ever-changing environment. But it is clear that such behavior at least tends in the direction of optimality. To assess the economic rationality of parallel paths strategy in research and development generally, various simulation exercises were conducted. Results of the simplest version are presented in Figure 8. Assuming vast uncertainty, each of many possible R&D paths was considered to have an identical ex ante probability of yielding a successful new product ranging from 0.2 (one success in five, as in clinical testing) to 0.01 (more like the probabilities in synthesizing therapeutically interesting molecules). Quasi-rents, varying over a wide range of values (and measured in thousands of dollars per year), were assumed to be realized at a constant annual rate from the time a success was achieved through year 25, with later payoffs discounted at an interest rate of 6 percent. Assuming the cost per experiment to be $1,000, profit-maximizing parallel

---

25 . Joseph A. DiMasi and Cherie Paquette, "The Economics of Follow-on drug Research and Development," Pharmacoeconomics, vol. 22 (2004 supplement), pp. 1-14. For earlier insights on drug R&D "racing," see Iain 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. Cockburn and Henderson find little evidence of conscious racing against known rivals, but suggest that the races are set off by scientific discoveries and that entrants respond in accord with their individual research capabilities.

paths strategies were computed. One sees in Figure 8 that, except with very low annual quasi-rent (benefit) realizations, some degree of parallelism is always optimal under conditions of uncertainty. Two additional important generalizations emerge: the deeper the benefits (quasi-rent) stream, the larger is the optimal number of parallel paths; and the smaller the probability of a single-path success, the more sensitive the optimum number of paths is to the depth of the benefit stream. For single-path success probabilities of 0.2 -- like those encountered in clinical testing -- the profit-maximizing number of paths ranged from 8 to 22. For 0.01 success probabilities, the optimal number ranged from 17 to as many as 177, depending upon the depth of the benefits stream.

This is to be sure a greatly simplified model. It cannot show that the number of paths pursued in real-world drug discovery is too small or too large. But clearly, substantial parallelism is called for, given the uncertainties. It is simply wrong to say that parallelism (often pejoratively called "duplication") is wasteful. Among other things, my analysis shows, parallelism advances the date at which a successful solution is available. Thus, in the context of Figure 3, it moves the R&D completion date to the left. My analysis also shows that over most payoff (quasi-rent) possibilities, parallel paths behavior yields positive profits to an individual firm seeking to come up with a successful new product under substantial uncertainty. In this respect the model cannot demonstrate the complete exhaustion of supra-normal profits, toward which quantitative analyses of pharmaceutical and biopharmaceutical profit records point. But if several firms competitively "race" toward a demand-satisfying solution, knowingly or inadvertently pursuing parallel paths, they will not only advance the time of completion toward the breakeven point in Figure 3, but they will increase total R&D expenditures and approach, even if not exactly achieve, breakeven.[26]

Whether such behavior is optimal in a broader sense depends upon the specific constellation of parameter values. When social benefits exceed the quasi-rents appropriated by private participants in an R&D marketplace, the social benefits curve lies at all points above the indicated quasi-rent curves, and the social benefits-maximizing time of project completion is earlier than the schedule that

---

26 . It should be noted that the "R&D Cost" curves in Figure 3 and its successors are severe abstractions from reality. They should ideally assume not a single R&D project, but a portfolio of parallel (and series) R&D projects that satisfies optimality criteria. And in this context, the more projects conducted in parallel, the higher R&D costs will be, but the sooner success will be achieved -- an alternative version of what Figure 3 says.

maximizes the net profits of a single (near monopolist) firm. Under plausible values of the relevant parameters -- e.g., when social benefits are 2.25 times private quasi-rents, as Mansfield found on average,[27] when the discount rate is 10 percent, when benefits are rising at 4.2 percent per year, and when R&D costs are falling at a rate of 3.05 percent per year -- breakeven, that is, zero-profit R&D equilibrium -- is socially optimal![28]

Thus, acceleration of pharmaceutical innovation through the pursuit of parallel R&D paths, both consciously within individual firms and collectively through inter-firm competition, is to be encouraged. The "invisible hand" surely does not achieve an optimal solution in inducing innovation, but it at least nudges in the proper direction. Chairman Mao's maxim, though wrongly implemented during the Cultural Revolution, can continue to guide pharmaceutical research: "Let 100 flowers bloom."

5. Valuation Uncertainties

Our focus thus far is on uncertainties in finding molecules that are interesting therapeutically, and in the end, those that can pass regulators' safety and efficacy hurdles. Another dimension of R&D uncertainty is less widely recognized: uncertainty about the ultimate market value of a new product, given some threshold level of technical success.

A clear indication of this valuation dimension comes from the pioneering research of Henry Grabowski, John Vernon, and Joseph DiMasi.[29] They compiled data on new drugs approved by the U.S. Food and Drug Administration after clinical trials in three different time intervals beginning in the 1960s. For each drug they estimated the discounted present value of inflation-adjusted quasi-rents, net of income taxes. They arrayed the individual drug payoff estimates in descending

---

27 . Edwin Mansfield et al., "Social and Private Rates of Return from Industrial Innovations," Quarterly Journal of Economics, vol. 91 (May 1977), pp. 221-240.

28 . This result is a slight variation from the proof in Scherer, "Schumpeter and the Micro-foundations of Endogenous Growth," supra note 8.

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

value deciles, with the result, for their first sample of drugs (introduced between 1970 and 1979) shown in Figure 9. The distribution of payoffs is highly skew. The top ten drugs in terms of discounted payoff value contributed 55 percent of the payoffs from all 98 drugs in the sample. Only drugs in the top three deciles achieved payoffs in excess of the average $81 million capitalized R&D cost per drug successfully achieving FDA approval, including the pro-rated costs of failed research and clinical trials. Thus, mere approval of a drug does not ensure financial success; that depends in addition in how the drug is received on the market and, among other things, how much competition it faces. Research on later approved-drug cohorts revealed a quite similar skew distribution of returns.

Research by Dietmar Harhoff and myself confirmed that such skew distributions are absolutely typical of the returns to investment in a wider array of new technologies, e.g., for the profits from individual patented inventions, the value of investments made by high-technology venture capital providers, and for the returns on common stock investments in high-technology companies that have floated an initial public stock offering.[30] Some of the distributions, e.g., on individual inventions, are more skew than those for the new drugs studied by Grabowski et al., since the latter had already passed the hurdle of FDA approval while patented inventions must still prove their mettle as technically feasible products along with market acceptance tests. But quite generally, the distribution of returns to high-technology investments is highly skew -- so skew in fact that the law of large numbers works at best poorly. Therefore, even though assembling feasible portfolios of projects can reduce the variability of pooled outcomes, substantial residual variability of returns remains.[31] Using the Grabowski-Vernon data plotted in Figure 9, we showed inter alia that the collectivity of all new-drug-developing companies active in the United States was subject to year-to-year profit fluctuations from their average value by as much as plus-or-minus 25 percent.

Nor are biopharmaceuticals an exception. Indeed, although systematic data are sparse, the degree of valuation uncertainty is almost surely higher than what Grabowski et al. observed for already-approved "small molecule" drugs. One indication of this is derived from the Pisano data graphed in Figure 7. Inclusion or exclusion of Amgen alone determines whether the whole biotech sample is seen to

---

30 . Harhoff and Scherer, "Technology Policy for a World of Skew-Distributed Outcomes," Research Policy, vol. 29 (April 2000), pp. 559-566.

31 . Ibid., pp. 563-566.

have positive or negative net profits. Further evidence comes as a by-product of research by Dietmar Harhoff and myself.[32] We tracked over time the values of common stock shares issued by 110 high-technology venture firms that had initial public stock offerings in the United States between 1983 and 1986. Of these, 52 survived until the end of 1995. Figure 10 plots the distribution of stock values for the 16 biotech survivors as of 1995, assuming that a $1,000 investment was made in each stock at the time of its initial public offering. The distribution is of course highly skew. One company, Amgen, towers over all the rest, with stock valued at $55,980 following a $1,000 investment in June of 1983. The top two -- Amgen and Chiron -- contribute 73 percent of the 16-company portfolio's value in December 1995. Further analysis reveals that the distribution of values became increasingly skew over time as valuation uncertainties were resolved by actual market events. Figure 11 tracks end-of-year value changes for a nine-firm subset of the 16 biotech companies. Most surviving companies, as Figure 10 implied, show little or no value gain relative to their initial investments.[33] Molecular Biosystems led the pack as the 1980s gave way to the 1990s, but then faded. Cohort leader Amgen had its own fluctuations reflecting important market events. Its blockbuster red cell-forming product, Epogen, was approved for marketing by the Food and Drug Administration in 1989. Competition from Amgen's licensee Ortho began in 1990 but intensified during 1993 and 1994 until suppressed under a license interpretation arbitration.[34] Amgen's fortunes also came under pressure as Mrs. Clinton's health care reform group focused attention on Epogen's high prices, and also by a Congressional investigation. But in September 1994 the Congressional furor ebbed and U.S. health care authorities decided to continue reimbursing for dialysis patients Epogen usage under the Medicare program at generous price levels.[35]

Investing in prospects with such skew-distributed outcomes is intrinsically risky. Averaging outcomes by assembling a sizeable portfolio is one way of reducing

---

32 . F. M. Scherer, Dietmar Harhoff, and Jörg Kukies, "Uncertainty and the Size Distribution of Rewards from Technological Innovation," Journal of Evolutionary Economics, vol. 10 (2000), pp. 175-200.

33 . Those that existed by merger or liquidation fared even less well.

34 . The author was a consultant to Amgen in the arbitration.

35 . Significant reductions in the reimbursement rate were announced in 2007.

the risk, although, as indicated above, one cannot diversify risk away fully.[36] Pharmaceutical companies hedge against risk by sustaining numerous development projects, many in different therapeutic classes. It is said that they also seek diversification through merger, although it is unclear whether the favorable large-numbers effect outweighs the unfavorable burden of greater bureaucracy. Venture capital firms also hedge by forming portfolios, although typically, they limit the size of their fund portfolios to 40 or 50 investment targets because close monitoring and the provision of hands-on advice to individual company leaders enhances the success of the overall portfolio investment.[37] In this respect, marginal statistical benefits of greater portfolio diversity are sacrificed for more effective managerial control, and venture capital investing continues to be a high-risk avocation.

The decision problem for the individuals who mortgage their homes and commit their life savings to launch a new high-technology company is more difficult. Most such entrepreneurs have limited wealth, and so they are unable to use portfolio strategies. They face the full panoply of technological risks and market valuation risks. Why? My speculative answer is rooted in a proposition advanced in 1948 by Milton Friedman and L. J. Savage.[38] It is hypothesized, among other things from the observation of gambling and insurance-buying behavior, that individuals have an ogive function defining the relationship between expected income and expected utility. At first, the marginal utility of income is diminishing, as Marshall among others postulated, but at income levels well above the individual's current situation, marginal utility is rising. In the case of high-technology entrepreneurs, a successful venture -- to be sure, at long odds -- is the best or only opportunity they have to achieve great wealth, a prospect to which a high utility value is attached. In addition, of course, there is the satisfaction that comes from hard creative work and being one's own boss. If the gamble fails, the high-technology entrepreneur typically has sufficiently attractive skills that he or she can find well-paying employment in someone else's company.

---

36 . See also the "dartboard" experiment reported in Scherer, "Schumpeter on the Micro-foundations of Endogenous Growth," supra note 8.

37 . See Ronald Gilson, "Engineerng a Venture Capital Market," Stanford Law Review, vol. 55 (2003), pp. 1067- .

38 . Milton Friedman and L. J. Savage, "The Utility Analysis of Choices Involving Risk," Journal of Political Economy, vol. 56 (August 1948), pp. 279-304. My own articulation of the hypothesis is found in "The Innovation Lottery," in Rochelle Dreyfuss et al., eds., Expanding the Boundaries of Intellectual Property ((Oxford University Press: 2001), pp. 3-21.

The United States has been extraordinarily successful in cultivating this climate of high-technology investment and entrepreneurship. I have often said that it is our secret economic weapon. Many other nations, including Germany, appear to have been less successful. One reason for the difference is the well-developed culture of venture portfolio investment and management in the United States.[39] But another differentiating factor may be the lower stigma associated with entrepreneurial failure in the United States, as compared, say, to Germany.[40] The propensity to fail at risky ventures is well-recognized in America, hardly anyone looks askance at failure, and unsuccessful high-tech entrepreneurs are often afforded additional chances to try their luck again. Whether the situation in Europe will change, with biotechnology as one of the beneficiaries, is something on which my knowledge is too limited to offer a prediction.

6. Cannibalization

One reason why the returns to R&D investment are highly skew is that the distribution of market opportunities is skew.[41] Some drugs are targeted toward diseases of high incidence among rich or well-insured consumers; others fall into the "orphan" class. This is to a considerable degree predictable in advance, but uncertainties remain. The first successful anti-AIDS drug, AZT, was classified as an orphan, but the disease proliferated so rapidly that supplying AZT became a bonanza. The profit returns may also be disappointing (or unexpectedly bountiful) because of the way competition to provide therapies for prevalent diseases fragments the market.[42] Usually, the first effective therapy has a first-mover advantage yielding both price and market share advantages and hence enhancing profitability. This is not always true, however. A second mover into the H2-antagonist anti-ulcer market, Zantac, was more successful than Tagamet, the

---

39 . See again Gilson, supra note 36.

40 . On the day I wrote this, the New York Times, August 11, 2007, p. B8, had an illustrative article about Heinz C. Schillelbusch's renaissance in the United States: "German Chief Felled by Scandal Returns to Market on a Smaller Scale."

41 . See Scherer, Harhoff, and Kukies, supra note 32.

42 . DiMasi and Paquette observe, supra note 25 at p. 4, that for orphan drugs, i.e., those with small potential markets, the multi-product entry characteristic of main-line drugs is "uncommon."

pioneer. A late mover into the cholesterol-reducing statin market, Prilosec, was almost discontinued in development because of its tardiness, but turned out to be the sales leader in its category and indeed the most profitable drug ever marketed.[43] But quite generally, to the extent that a therapeutic category attracts multiple entrants, each new entrant may cannibalize the sales and profits of other category members. This could lead to a less favorable interpretation of the evidence that competition for market positions in the pharmaceutical industry leads on average to a near-zero supra-normal profit equilibrium.

On the positive side, as we have seen, the pursuit of parallel R&D paths accelerates the availability of new therapies and thereby benefits consumers. Also unmentioned thus far but true, drugs in a narrow therapeutic category often have effects that differ among the patients to whom they are administered, and more choice means more opportunity that a better therapy can be prescribed for any given individual. And it has been shown that the more drugs there are in a particular therapeutic category, the lower prices tend to be, all else equal, at least in a largely unregulated market such as the United States.[44]

But on the negative side, when multiple entries fragment a therapeutic category, the quasi-rents that count as benefits in Figure 3 and the consumers' surpluses that attend them may be reduced as rivals cannibalize each others' surpluses. A relatively extreme example is shown in Figure 12. It assumes that at the outset, a product already exists with demand function D_1 and marginal cost function MC. Its producer maximizes profits by setting price P_1, realizing annual quasi-rents (i.e., producer's surplus) given by the rectangle P_1WXM, analogous to the horizontally shaded rectangle in Figure 1. Suppose now that another firm develops a substantially superior product that (exaggerating) captures all of the sales of the incumbent product and expands the overall demand in the therapeutic category. For geometric tractability, the demand shift is assumed to be linear and parallel to the original demand function, i.e., to D_2. We assume the same marginal cost MC. The newcomer maximizes its profits by setting the now-higher price P_2 and supplying expanded output OQ_2.

---

43 . John Simons, "The $10 Billion Pill," Fortune, January 20, 2003, pp. 58-68.

44 . See John Lu and William S. Comanor, "Strategic Pricing of New Pharmaceuticals," Review of Economics and Statistics, vol. 80 (1998), pp. 108-118; and DiMasi and Paquette, "The Economics of Follow-on Drug Research and Development," supra note 25, p. 12.

To analyze the welfare implications, we proceed in two stages. As a first approximation, ignoring competitive rent-seeking, producer's surplus (quasi-rent) 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 gain from the improvement, adding consumers' surplus increment GEP_2. Abstracting from cannibalized or transferred surplus, which is essentially a welfare wash-out, the net welfare gain is the lazy-L-shaped area GEBXWX (shaded with dots).

Now, however, suppose that the new product development is characterized by total rent-seeking dissipation. The stimulus to investment in the cannibalizing product is the new producer's surplus (quasi rent) rectangle P_2EBM. With complete rent dissipation, that surplus will be offset by the R&D expenditure of the firm(s) seeking to displace the original product. It can no longer be counted even in part as a welfare gain. Parallelogram GEKZ was before rent dissipation net new surplus. Because the two areas have equal heights and equal bases, the area of parallelogram GEKZ equals the area of rectangle P_2KP_1, all of which is dissipated by rent-seeking R&D costs. The remaining incremental welfare gain WKBX is also dissipated by R&D costs. After cannibalization plus rent dissipation, the net welfare effect is negative! Only after patent protection ceases do incremental welfare gains materialize.

The implications of this example are perplexing. Quasi-rents are dissipated not by competitive research that advances the time of product availability, which clearly increases welfare, but by substitution of one (superior) rent-yielding product for another. It is possible that in such product cannibalization cases, R&D can be targeted more narrowly and with greater certainty on displacing known products, R&D costs are saved, and the winning product's quasi-rent P_2EBM is less than fully dissipated. Or if on the contrary there is full dissipation, total R&D costs over time exceed quasi-rents P_2EBM because they expanded initially to offset the original product's quasi-rent P_1WXM and then to offset the new (cannibalizing) product's larger quasi-rent. Combined, they exceed the maximum pool of quasi-rents, and therefore, in the long run industry profits must be negative. At least for traditional pharmaceuticals, this is not what the data indicate, and so the example presented is too extreme. But it at least suggests caution in attributing unambiguous benefit to rent-seeking research and development investment processes.

Conclusion

The market-oriented creation and production of new pharmaceuticals is a

powerful but sometimes troubling phenomenon. It is responsive to changes in the science base and demand conditions. Competition among firms to capture the profits that come from success in developing and marketing a new product can accelerate the availability of new therapies and encourage product differentiation sensitive to the diversity of patient needs and (less clearly) facilitate price competition among contending drugs in a given therapeutic class. But it may be marred by excessive near-duplication of R&D investments, a proliferation of me-too drugs, and the dissipation of what otherwise would be producers' surplus.

The pursuit of numerous parallel paths at the discovery stage, which may superficially seem duplicative but in fact is an important vehicle for coping with uncertainty, is highly desirable. Competition among "Big Pharma" firms may be insufficient to ensure an optimal number and diversity of parallel paths, especially given the shrinkage in the number of independent centers of initiative attributable to a recent wave of pharmaceutical company mergers and consolidations. It is unclear to what extent this problem has been offset by the emergence of many small biopharmaceutical specialists, on whom "Big Pharma" companies rely significantly for product development candidates. Here too the discovery activities of not-for-profit university and hospital researchers play a crucial role. Without the substantial financial support provided by governments, a serious lacuna would exist.

Where profit-seeking markets fail most dramatically is in the provision of new medicines to prevent or cure the so-called "tropical" diseases -- those occurring mainly in low-income nations. In many if not most such cases, consumers' purchasing power and hence demand is insufficient to support the profits that would induce vigorous investment in research and development. To meet these human needs, government and philanthropic agency support is vital. It can operate on the supply side, targeting specific discovery research and product testing projects for subsidy, or on the demand side, e.g., as under new Advance Market Commitment programs, assuring the financial means to purchase new drugs in quantities and at effective prices sufficient to induce private sector research and development. These are themes whose consideration would prolong my paper unduly, but which, I assume, will be analyzed in depth by other conference papers.

Figure 1

Profits under Unrestrained Monopoly and Price Controls

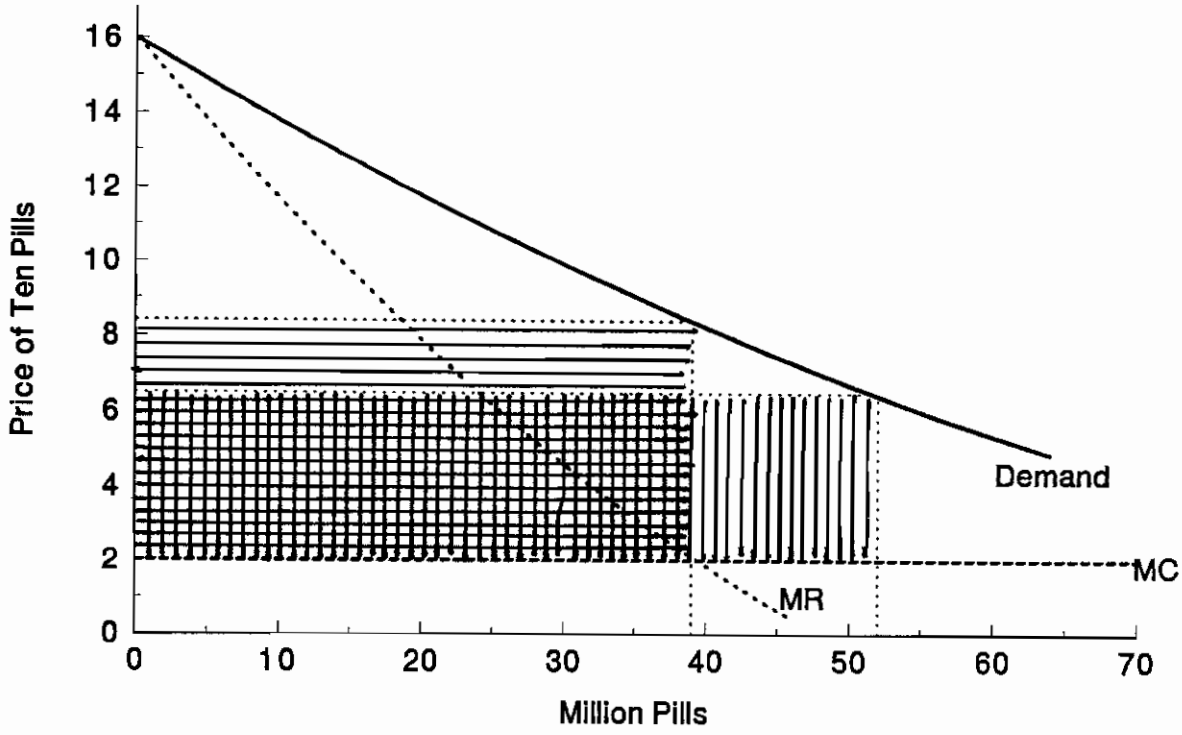


Figure 2

Reduced Demand in Low-Income Markets

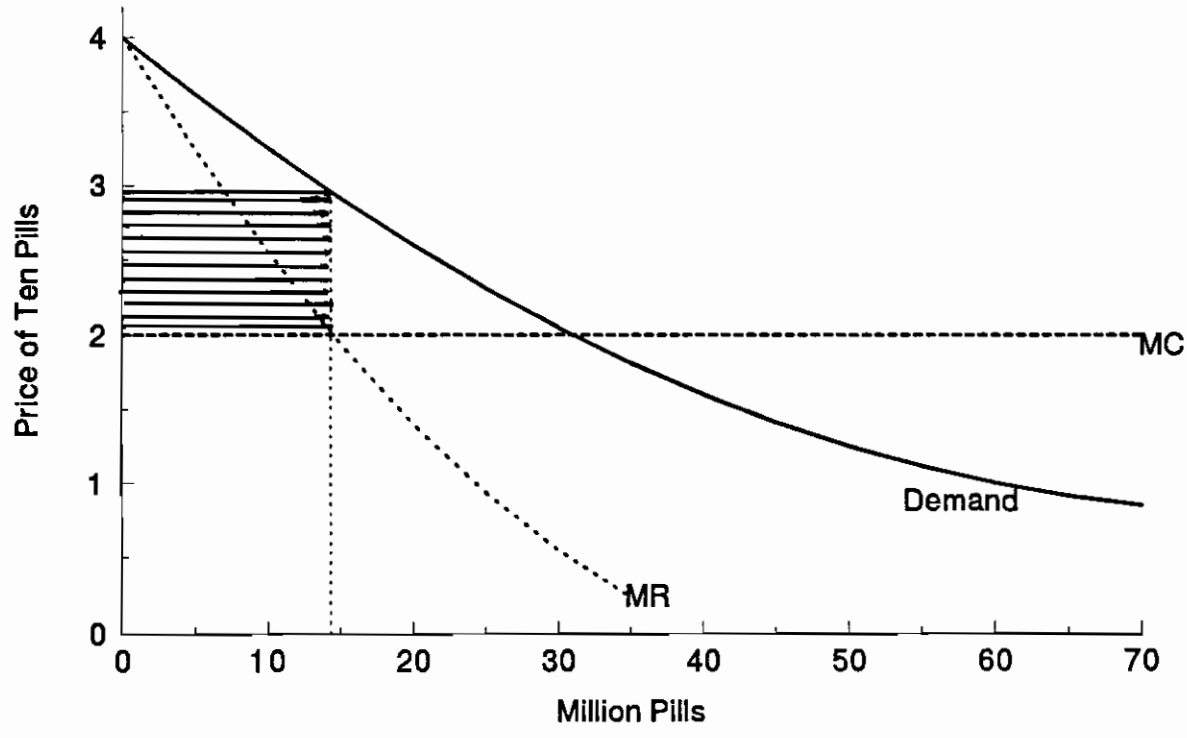


Figure 3

Demand-Pull and Science-Push with Continuous Change

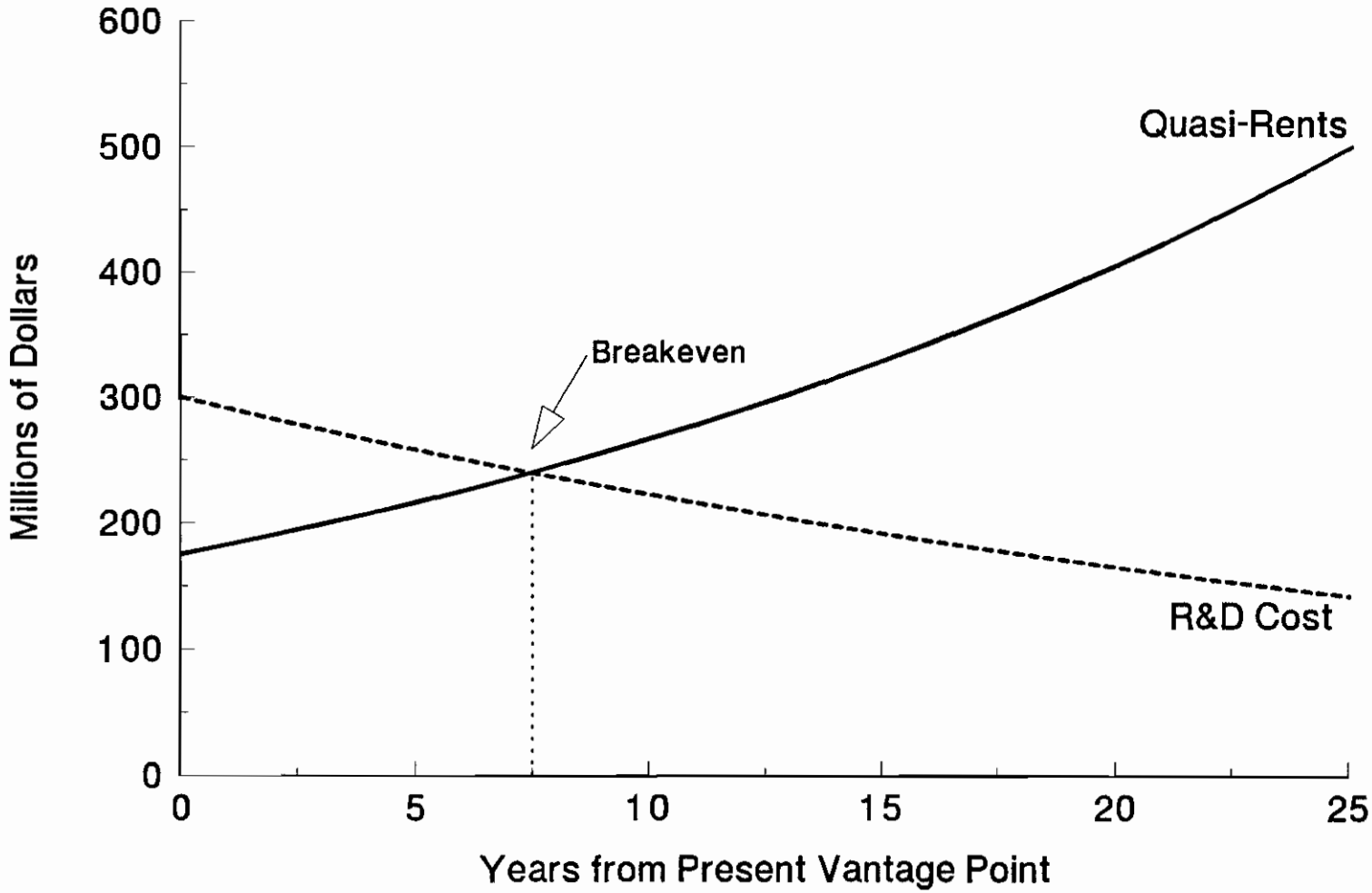


Figure 4

Results of Scientific Breakthrough

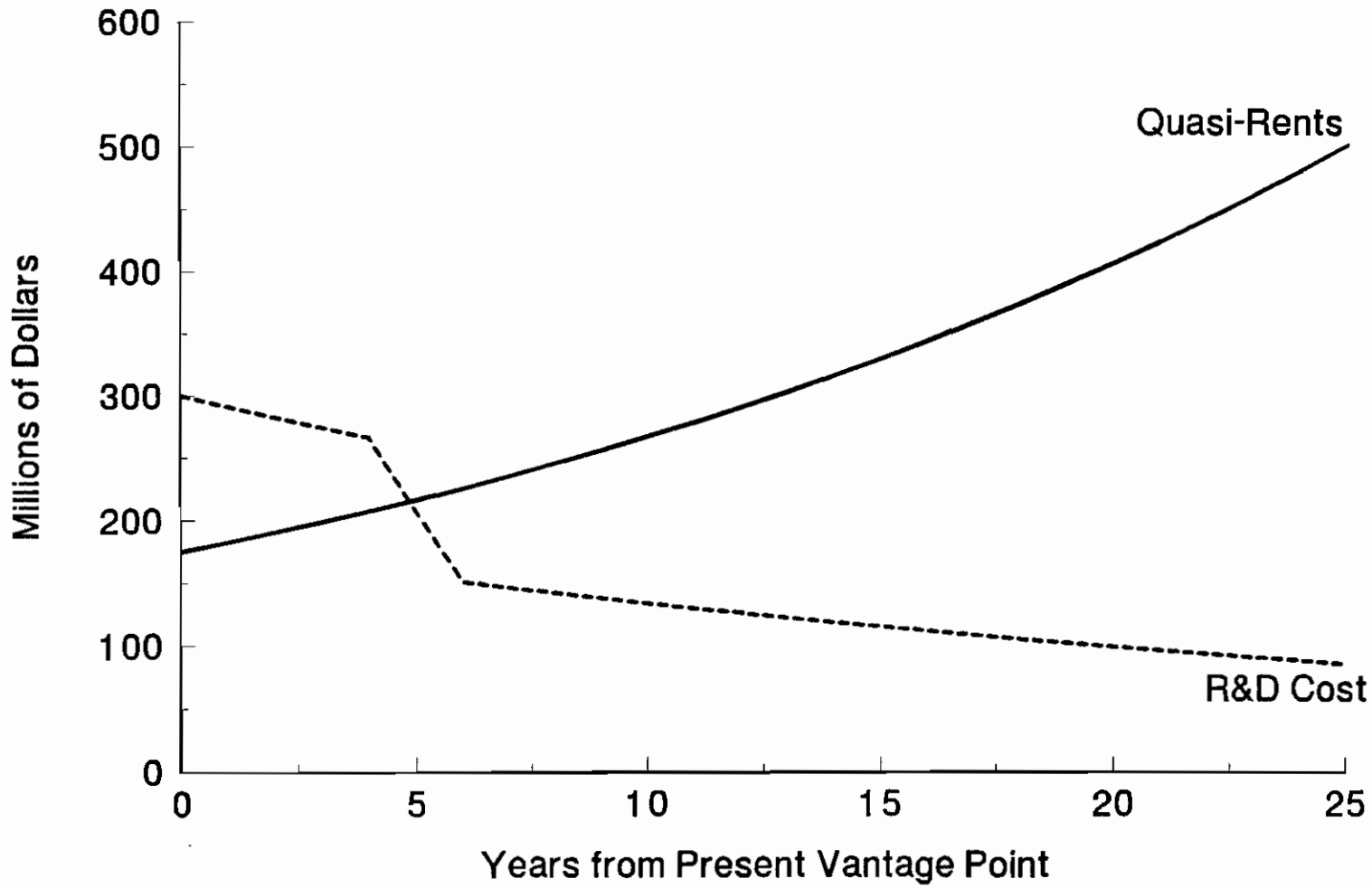


Figure 5

Results of Sudden Demand Increase

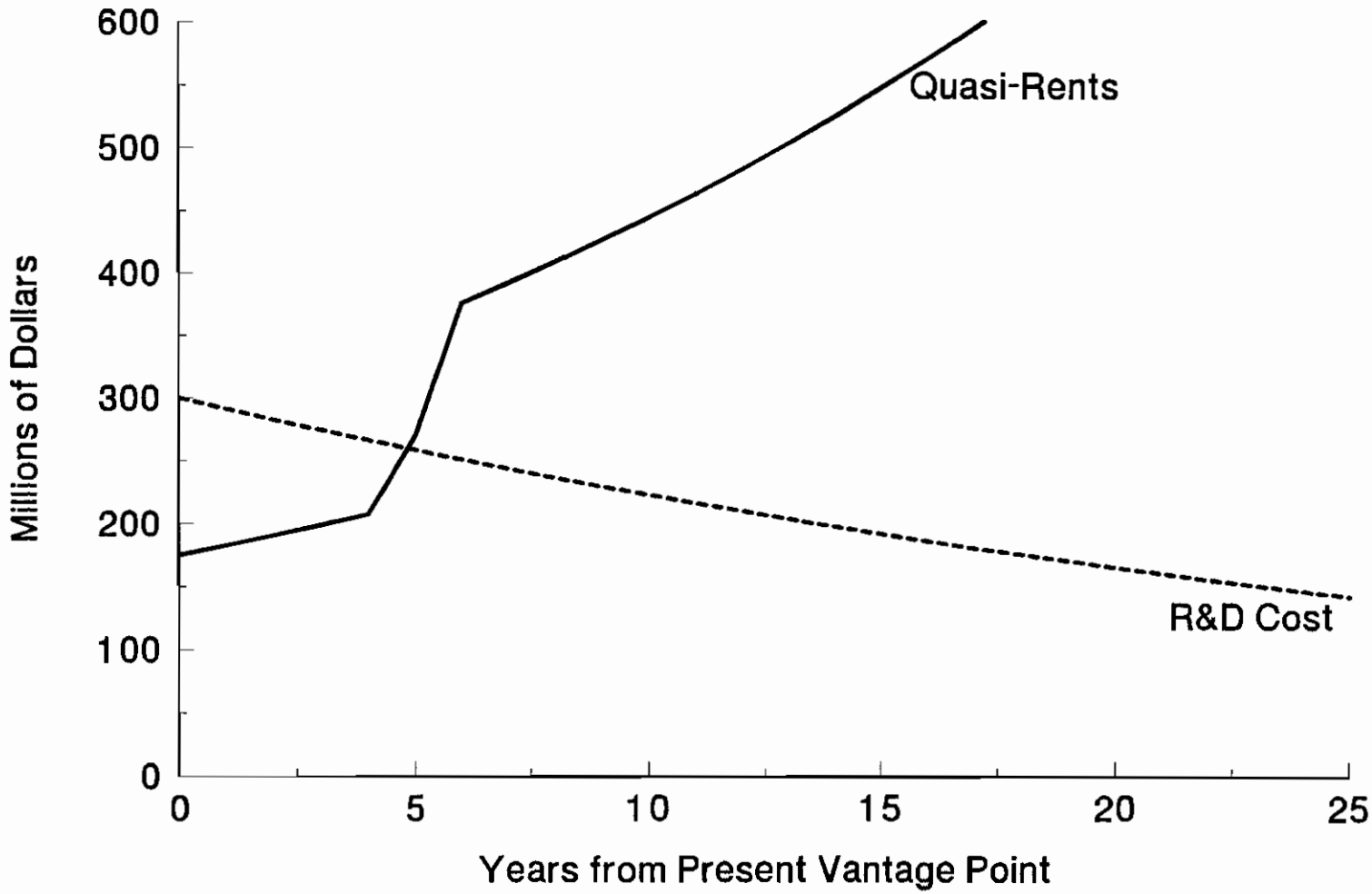


Figure 6

Trend-Adjusted Movements of Pharmaceutical Margins and R&D

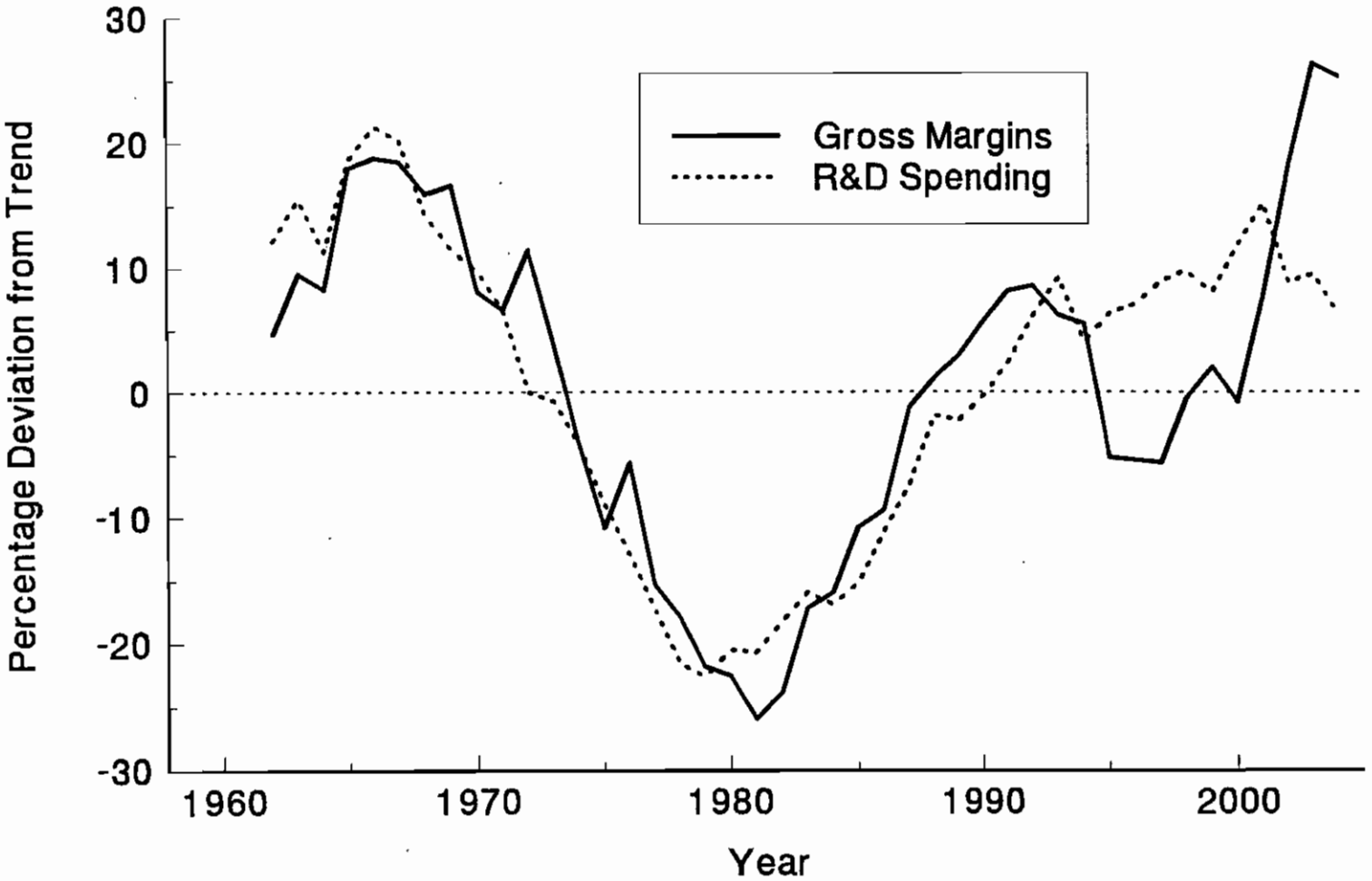
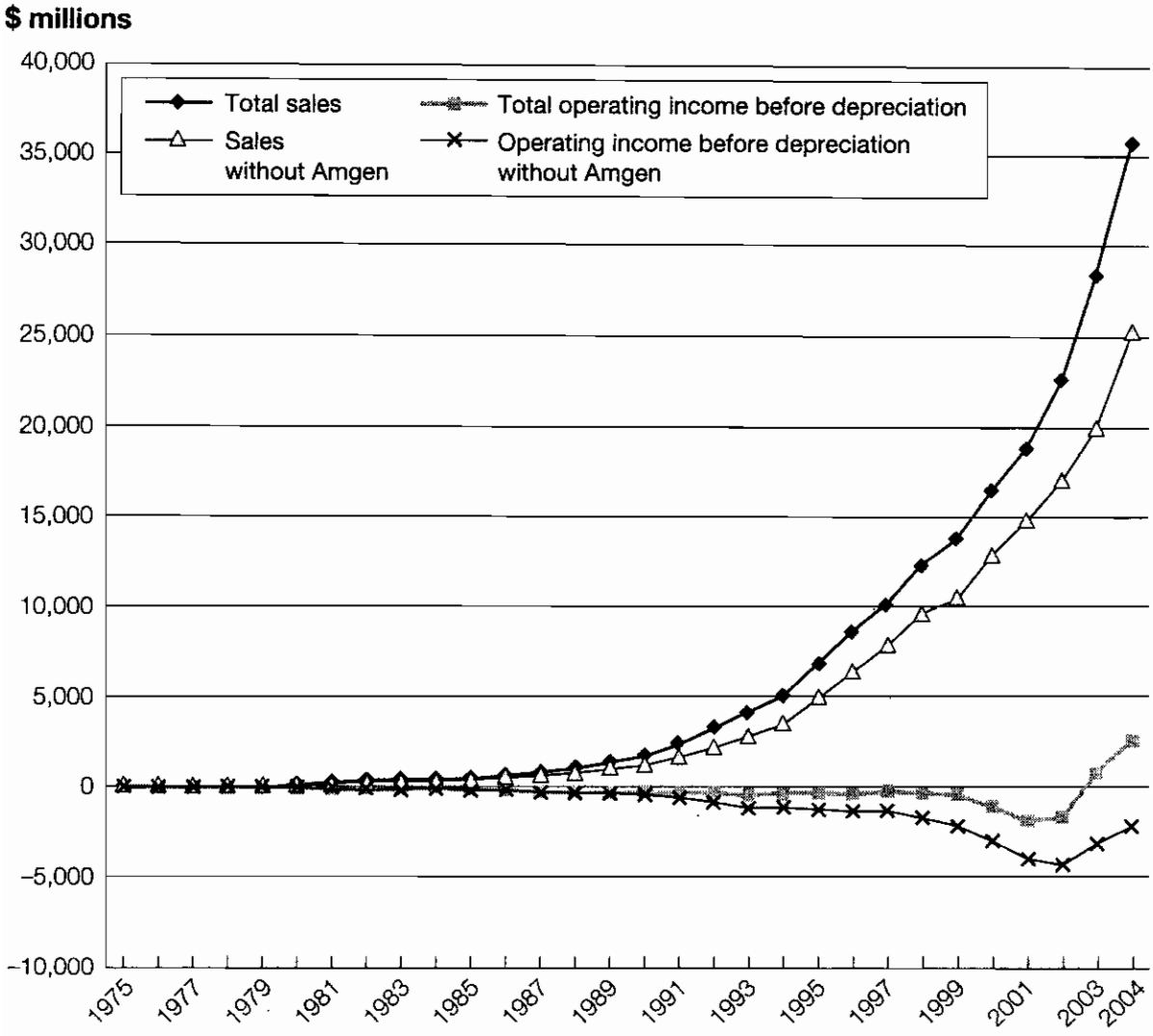


Figure 7

Biotech revenues and profitability with and without Amgen, 1975–2004*



*Values are inflation-adjusted.

Source: Gary P. Pisano, Science Business: The Promise, the Reality, and the Future of Biotech (Harvard Business School Press: 2006), p. 115.

Figure 8

Optimal Choice of Parallel Paths Strategies

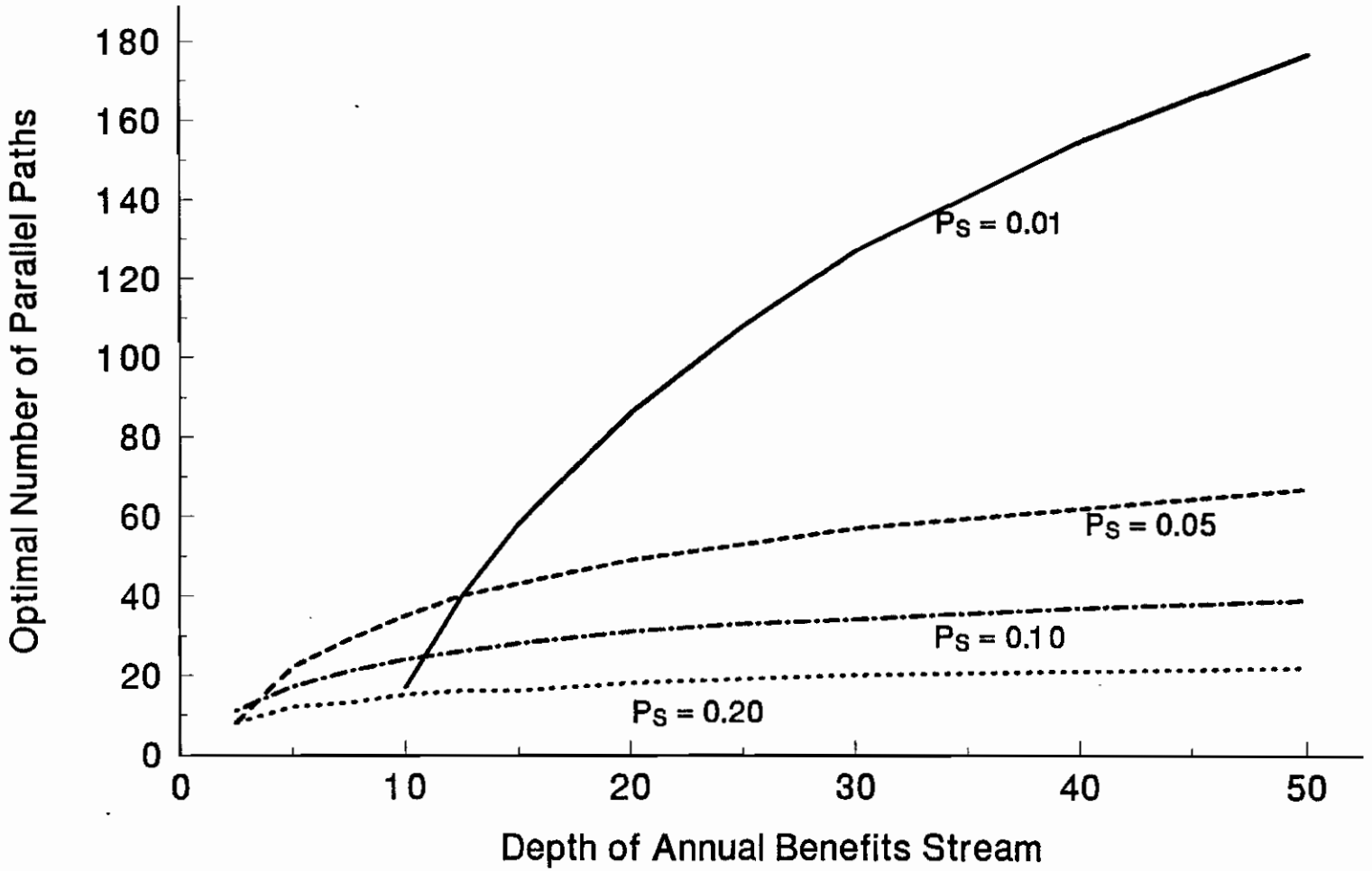


Figure 9

Discounted Present Value of 1970-79 New Drug Quasi-Rents

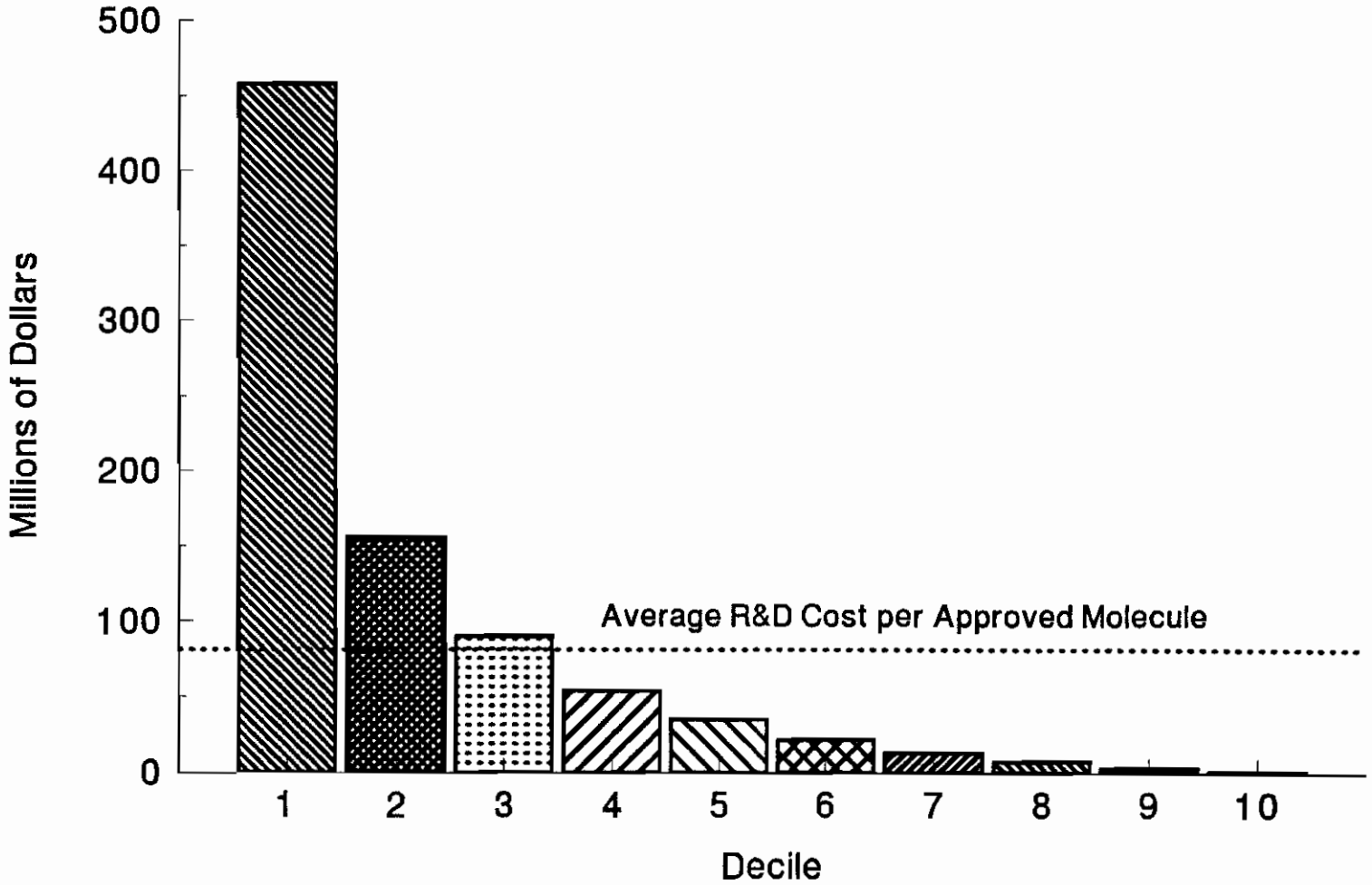


Figure 10

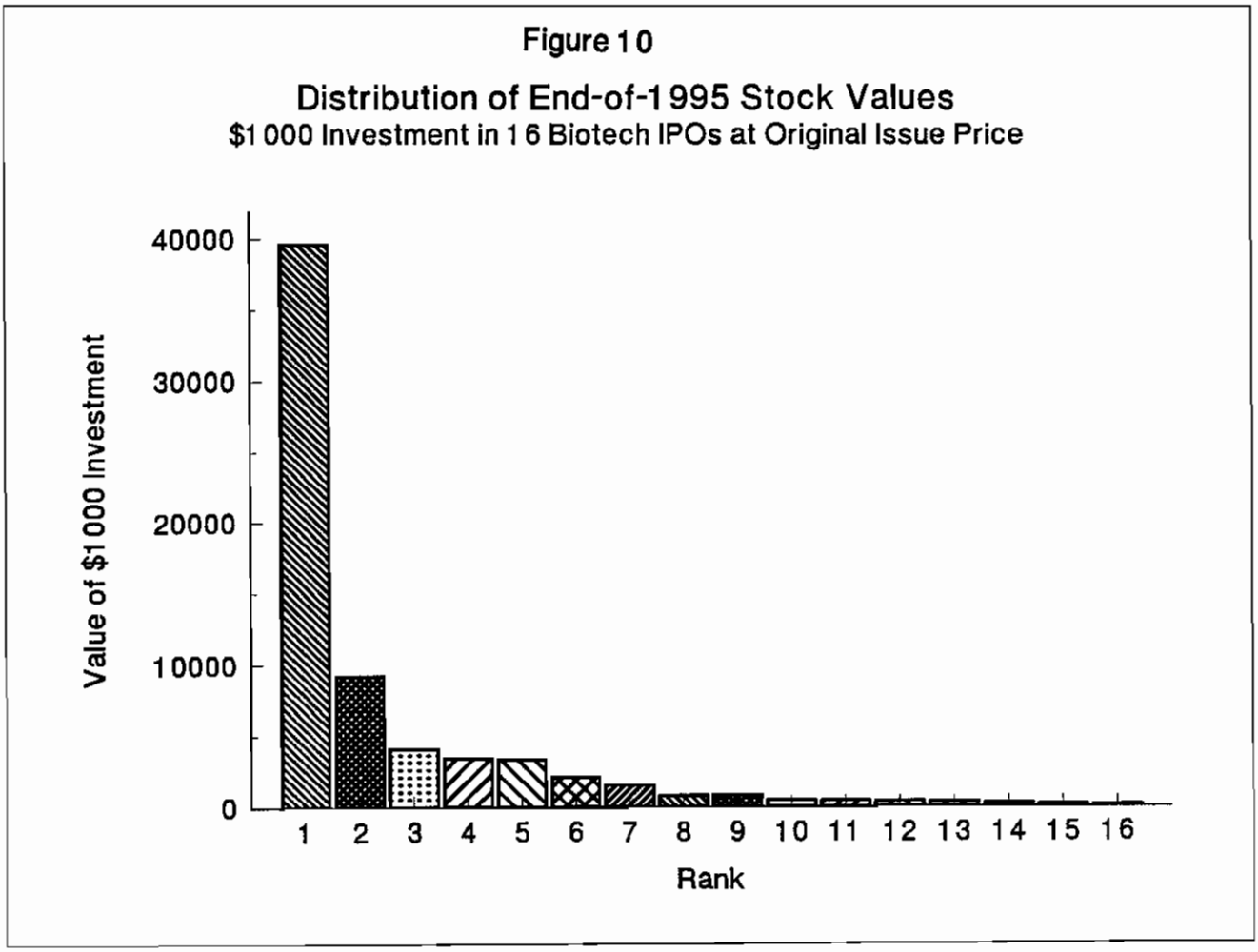

Distribution of End-of-1995 Stock Values
$1 000 Investment in 16 Biotech IPOs at Original Issue Price
Value of $1 000 Investment
40000
30000
20000
10000
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Rank

Figure 11

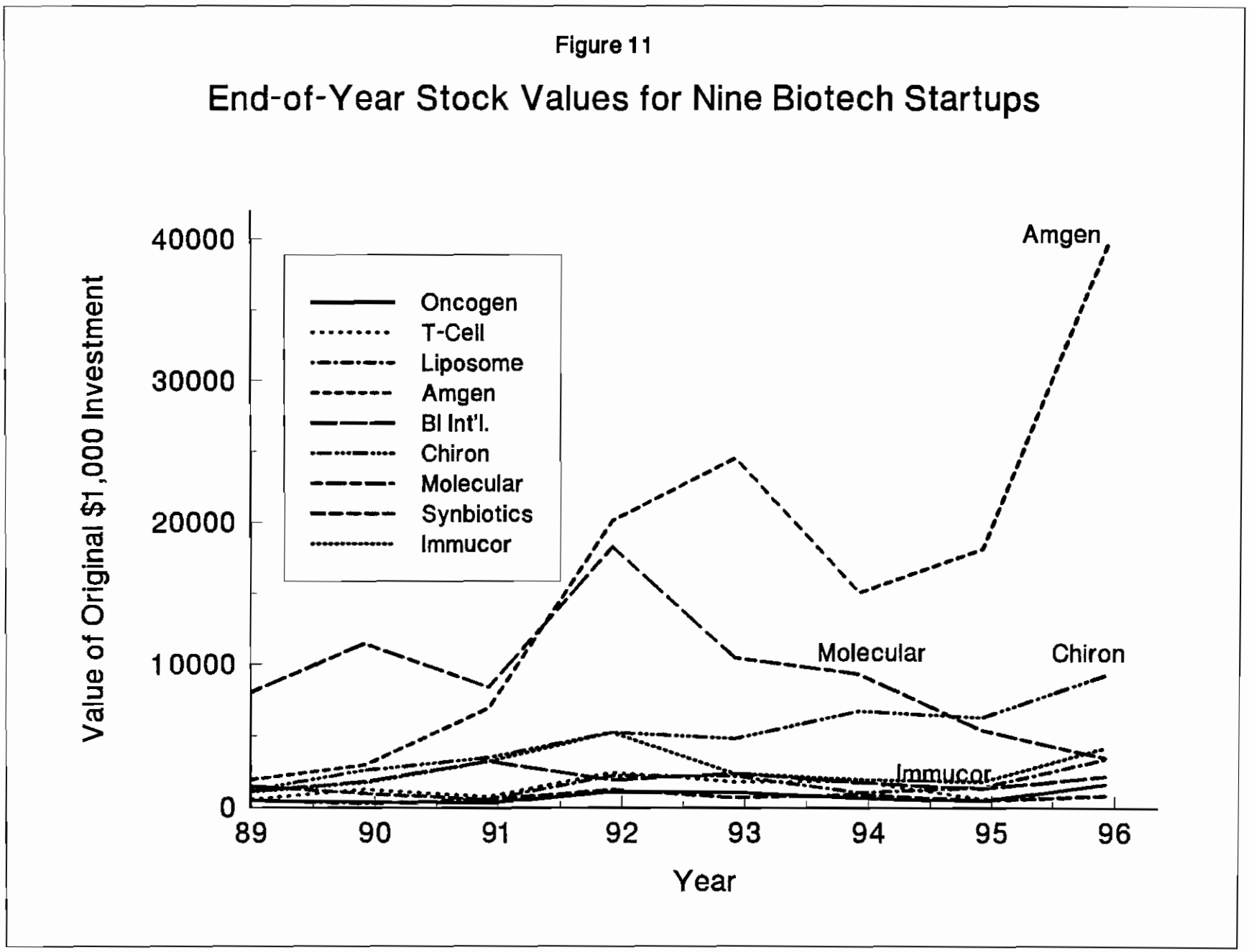

End-of-Year Stock Values for Nine Biotech Startups
Value of Original $1,000 Investment
40000
30000
20000
10000
0
89
90
91
92
93
94
95
96
Year
Oncogen
T-Cell
Liposome
Amgen
BI Int'l.
Chiron
Molecular
Synbiotics
Immucor
Amgen
Molecular
Chiron
Immucor

Figure 12

Equilibrium with Cannibalization and Rent-Seeking