

STRATEGIC VALUE OF THE R & D INVESTMENT:

A SIMULATION MODEL APPROACH

Edmund B. Pyle, III, Wilfred J. Westlake

Bryce Douglas, and A. Douglas Bender

Smith Kline & French Laboratories

ABSTRACT

A Monte-Carlo Simulation Model was developed in an attempt to relate the size of the R&D budget to the future need for new pharmaceutical products in order to sustain the desired level of sales growth. This paper describes the simulation model in detail and presents an example of how it is used to evaluate and compare various R&D spending levels and strategies.

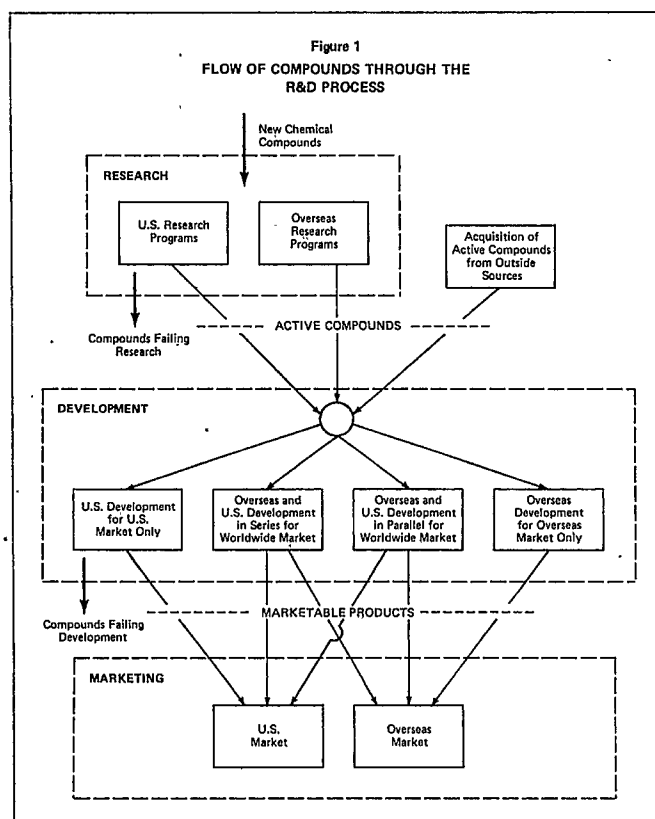
INTRODUCTION

The subject of the size of a Research and Development budget and the impact of this investment on the corporate growth have become of increasing concern to management as pressures to justify and control expenditures continue to increase. Traditionally, the size of the corporate R&D spend has been expressed as a percentage of sales. This figure ranges from a low of 1 to 2 percent in the capital intensive industries such as paper, heavy machinery and textiles, to 3 or 4 percent for the chemical industry, and up to 9 or 10 percent for the pharmaceutical industry. With the increasing attention now being given the R&D budget and the changing attitudes regarding the value of research and development, there have been attempts to express the level of R&D spend in other, and hopefully more meaningful terms such as its relationship to cash flow (1) and the impact of "capitalizing" rather than "expensing" R&D on rate-of-return on investment (2). Other studies have dealt with the performance or payoff from the investment in Research and Development (3-7), selection of projects and allocation of R&D resources (8-11), and the financial control of the R&D spend (12-13). None of these approaches, however, deals with the issue of the size of the R&D expenditure in relationship to the future goals and expectations of the corporation. In other words, the strategic value and implications of the R&D budget have been largely ignored.

The specific problem dealt with here basically concerns how much to spend annually on R&D to insure, within certain levels of confidence, future sales derived from new products which would meet desired group growth goals. In order to gain some insight into this issue, a model was designed which simulates, under certain specified conditions, the flow of chemical compounds from the research programs through the development process to the marketplace, and describes this flow in terms of costs and potential sales. This paper describes the mechanics of the simulation model and presents a hypothetical example problem in order to demonstrate how the model can be utilized to investigate the impact of various research spending levels and development/marketing strategies upon the future sales levels and development resources.

PHARMACEUTICAL RESEARCH AND DEVELOPMENT

The research and development process in the ethical pharmaceutical industry incorporates many varied activities. These may be grouped, however, into the following basic elements: acquisition and synthesis of novel new chemical entities, evaluation of these compounds in biological test systems designed to pick up desirable biological activities, assessing the safety of lead structures in animals, assessing the safety and efficacy in man of compounds judged to merit clinical trial, and finally securing permission to market the new drug. This process of discovery, development and marketing of a new drug is actually a flow of chemical compounds and is expressed schematically in Figure 1. New chemical compounds are passed through a number of research programs, both in the United States and Overseas. A small fraction of these survives primary and secondary biological testing, as well as all technical and economic evaluations, and are sent into the development process. Compounds also arrive in development from other sources, such as direct acquisition or joint ventures with other companies under licensing and/or royalty agreements. Once in development, a development/marketing strategy is selected



based on the predicted product profile, anticipated market, technical risk, and available resources. Of all the compounds entering the development process, a small fraction survive and become marketable products.

Despite the simplified system overview, the entire process is filled with uncertainties about the future, including when the next lead will be found for development, whether or not a particular compound will survive the development process and become a marketable product, and how the market size and competitive position of a potential product will vary between the time a research program is undertaken and the time a product is realized.

MODEL OF THE R&D PROCESS

Due to the complexity of the entire research and development process and the uncertainties associated with future events, it was apparent at the outset that a closed-form mathematical representation of the system was not possible. In addition, representing the system through a series of empirically derived relationships was not considered feasible due to the stochastic nature of many of the system elements. Since the behavior of many of the system elements was known in terms of historically derived probability distributions, and since the path through the R&D process for a particular chemical compound could be treated as a series of discrete events, a simulation model was considered to be the logical choice to represent the system. Conceptually, the model employs Monte-Carlo techniques to determine when each compound is ready to begin the development process, which development/marketing strategy will be employed for each,

whether or not each compound survives development, and the sales achieved for each compound that becomes a marketable product. A flow chart of the model is shown in Figure 2.

The queue of random compound arrivals is generated by considering three separate and independent sources: U.S. Research, Overseas Research, and Outside Acquisitions. Actually, a set of random arrivals is generated for each source independently and then merged to form the queue of compounds for input to the development process. This is accomplished by assuming the flow of compounds from each source to be completely independent and random events, i.e. a Poisson process with parameter "α" (the mean compound arrival rate). The time between successive compound arrivals thus has a negative exponential distribution. Pseudo random numbers from a negative exponential distribution can easily be generated from the observation that the cumulative distribution function of any continuous distribution is uniform. The cumulative distribution function for the negative exponential distribution is $[1-e^{-\alpha t}]$ and thus times between successive compound arrivals can be obtained by solving the equation:

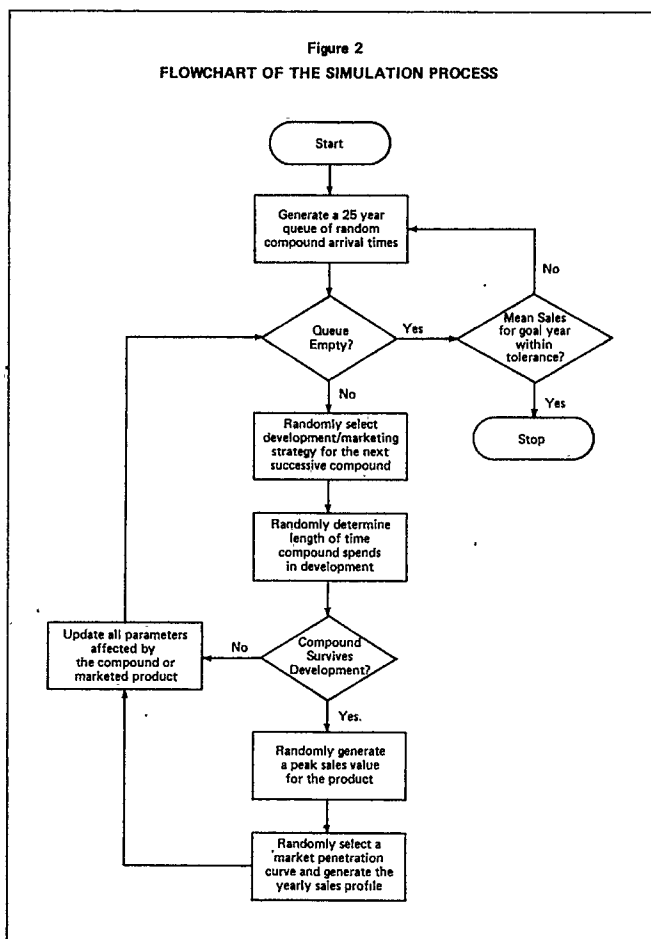
$$t = (1/\alpha) \log_e [1/(1-u)] \quad (1)$$

where "u" is a pseudo random number from the uniform distribution, $0 \leq u \leq 1$. Equation (1) is used to generate the time between successive arrivals separately for each source because "α" is different for each source. In the case of outside acquisitions, "α" is an input variable, the value of which is determined from historical data. In the case of U.S. and Overseas Research, "α" is determined from the research spend (\$), manpower costs (\$/man) and productivity rate (compounds/man-year), all of which are input variables. The manpower costs and productivity rates for each source are determined by examining historical data. The research spend is the variable whose impact is to be assessed and its value directly affects the average compound arrival rates for both U.S. and Overseas programs. Therefore, as the research spend increases, the number of compounds in the research programs in the input queue also increases. The length of the input queue was arbitrarily fixed at 25 years with the first year of the span supplied as input.

Once the input queue of development compound arrivals is generated, the model begins selecting them for development, one at a time, until the queue is exhausted. This entire process assumes that a compound is started in development as soon as it is available, i.e. no waiting time involved. For each compound selected, the model randomly picks a development/marketing strategy. The selection of a particular strategy is made based on the occurrence fraction specified for each. The occurrence fractions are supplied as input and reflect the relative amount that each strategy is to be utilized during that particular simulation. They can have any value from 0 to 1 for each individual strategy but the sum of all occurrence fractions must be equal to 1. The particular development/marketing strategy is selected by generating a uniform random number and comparing it successively to the cumulative sum of the occurrence fractions for each option. We have identified seven basic development/marketing strategies which vary in the degree to which activities are planned in series and/or in parallel. Each is associated with a different number of years for development (N), yearly cost allocation over the development period and probability (P_s) that a new product emerges from the process. These values were obtained from historical data for each strategy and are fixed within the model.

After selecting a development/marketing strategy, the model determines how long the compound survives the development process. It does this by assuming that the first year of development is completed with certainty, and the probability of completing each successive year thereafter decreases linearly with

Figure 2
FLOWCHART OF THE SIMULATION PROCESS



time until "P_s" is reached for year "N". Mathematically, this can be expressed as:

$$P(k) = 1 - [(k - 1)(1 - P_s)/(N - 1)] \quad (k = 1, 2 \dots N) \quad (2)$$

A uniform random number (u) is generated and compared with successive values of P(k). When P(k) < u, "k" is assumed to be the year in which the compound fails and is dropped from the development process. The model adds the development costs for the number of years of survival into the appropriate yearly totals and proceeds to pick up the next compound in the queue.

If $u \leq P_s$, the compound is assumed to successfully complete development and become a marketable product. The model next randomly determines a peak year sales figure for sales in the U.S. by generating a Gaussian distributed random number whose distribution mean and standard deviation are determined from a combination of historical data and sales projections for existing research programs and are supplied as input. Also supplied as input are the upper and lower bounds for the peak sales. If the randomly generated sales figure lies outside the specified range, it is discarded and another value generated. Once the peak sales figure is obtained, the model randomly selects a market penetration rate. The model has three penetration rates: slow, normal and fast. These correspond to products which face stiff competition, average competition and little or no competition respectively. It is assumed that 25% of all new products will be slow, 50% normal and 25% fast. These values were obtained considering historical data. The value of a uniform random number is used to determine which penetration rate is used. Associated with each penetration rate is a curve of percent of peak sales as a function of time. Hence, once the penetration rate has been determined the yearly U.S. sales figures are obtained by applying the appropriate market penetration curve to the peak sales. The yearly overseas sales figures are obtained by applying a ratio of U.S. to overseas sales which is supplied as input. The model next adds the appropriate sales figures (depending upon the strategy) and the development costs into the appropriate yearly totals and proceeds to pick up the next compound in the queue.

When all the compounds in the queue have been exhausted, the model updates the mean yearly values of development costs and sales. In addition to these parameters, the model also updates mean yearly values for the number of compounds starting development, failing development and being marketed for each year in the 25 year span of interest. Following the updating process, the model reinitializes all variables and proceeds to generate a new queue of development compound arrivals for another run. This process is continued until some minimum number of runs have been completed, after which the model computes how many runs are necessary to achieve some degree of stability in the mean sales values. This is accomplished by utilizing a precision specification on the mean sales for some particular year of interest within the chosen 25 year span. This target year usually corresponds to some year for which a sales level goal has been established. Both the target year and the precision specification, with the precision specification being expressed as a fraction of the mean sales, are supplied as input.

The model computes the number of runs necessary by applying a statistical sampling technique. Suppose we let "x̄" be the mean and "s" be the standard deviation of the sales for a particular year as calculated after "n" runs have been made. These values are really estimators of an approximately normal distribution with a true mean "μ" and standard deviation "σ". Assuming "n" to be large enough so that the standard normal distribution can be used to estimate the true population of the calculated mean sales, then, with 95% probability, it can be shown that:

$$\left| \bar{x} - \mu \right| \leq 1.96 (s/\sqrt{n}) \quad (3)$$

Now we want the right side of equation (3) to be less than or

equal to some "a \bar{x} ", where "a" is the precision specification. Hence, solving for the required number of runs "n_r" yields:

$$n_r \geq [1.96 (s/a\bar{x})]^2 \quad (4)$$

After experimenting with many cases it was found that with a = 0.1 a good compromise between precision and run time is achieved. Since "n_r" varies inversely as the square of "a", decreasing "a" by a factor of 2 increases "n_r" by a factor of 4. It was also found that initially letting n = 25 yields good results. Therefore 25 runs are made before the model first uses equation (4) to predict the number of runs required. If n_r > 25, the additional runs are made and a new value of "n_r" is computed. This process is continued until n_r ≤ n, at which time the simulation is terminated. On the other hand, if n_r < 25 initially, 10 additional runs are made to insure that a low value of "s" was not achieved accidentally. After that, the simulation is again terminated when n_r ≤ n.

After the simulation is terminated, the model produces an output report consisting of two pages. The first page summarizes all input parameters and gives the yearly mean values for number of compounds starting development, number of compounds dropped from development, number of marketed products, development costs, and gross sales. The second page gives data pertaining to the detailed distribution of sales for the target year and the appropriate statistical parameters so that the probabilities of achieving various sales levels other than the mean may be ascertained. A sample output report is shown as part of the example case presented in the next section.

EXAMPLE OF MODEL USE

In order to aid in understanding the method in which the model is utilized, the details of operation and interpretation of results are demonstrated through the use of a hypothetical example case. The data presented in this example are for illustrative purposes only.

Assume a corporation had a projected gap between the desired sales (reflecting the planned growth level) and the forecasted sales from current products of \$220,000,000 by 1988. Since new products must ultimately come from R&D, the question of how much to invest today and in the future in order to find and develop a sufficient number of new drug products which would generate the required sales level can be addressed by using a simulation model such as described here. To look at this issue, management needs to know not only how much to spend on research, but how the development/marketing strategies will affect the sales of products emerging from current and anticipated research programs.

The input data assumed for this sample case was as follows:

Research Manpower Costs:

United States:	\$200,000/man-year
Overseas:	\$140,000/man-year

Productivity:

United States:	0.04 Compounds/man-year
Overseas:	0.02 Compounds/man-year
Outside Acquisitions:	1.00 Compounds/year

U.S. Product Size Distribution:

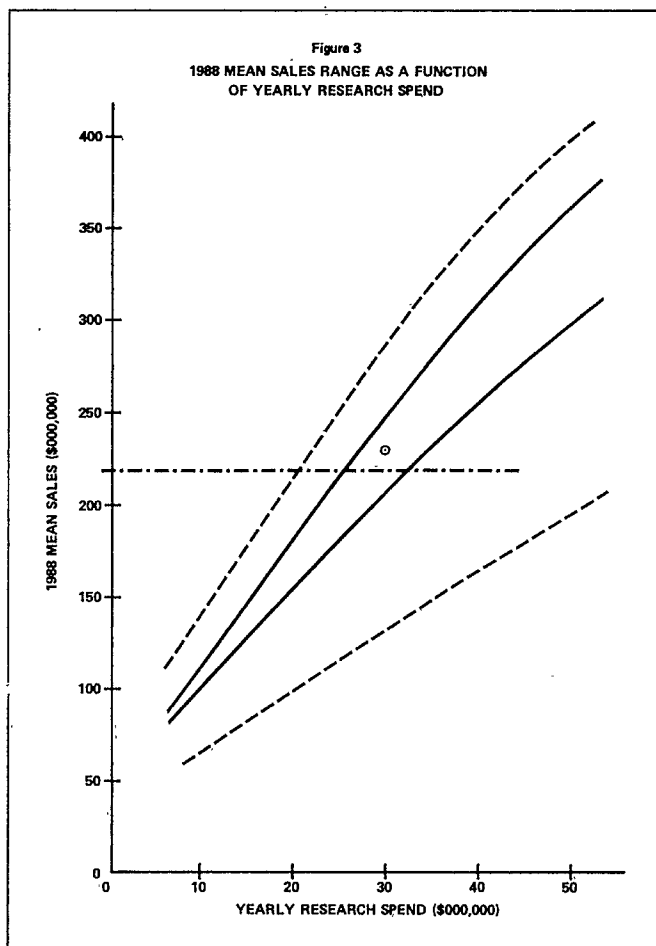
Mean:	\$ 24,000,000
Standard Deviation:	\$ 8,000,000
Lower Bound:	\$ 3,000,000
Upper Bound:	\$115,000,000

U.S./Overseas Sales Ratio: 2
 First year in 25 year span: 1973
 Target year: 1988
 Fractional Mean Sales Precision: 0.10

In addition, it was assumed that the development/marketing strategy was partially constrained by the fact that due to the nature of the current research programs, 10% of all compounds would be developed for an overseas market only and 20% of all compounds would be developed for a U.S. market only.

Using the simulation model, a parametric analysis was conducted by varying the research spend level and the development/marketing strategy mix. In order to simplify this analysis somewhat the U.S./overseas research spend ratio was held constant at 5:1. The results of the simulations are shown in Figure 3. The solid curves represent the ranges of possible 1988 mean sales values as a function of research spend for the specific case considered here. For each level of research spend the corresponding value for mean sales could lie anywhere between these two curves depending upon the strategy mix chosen. The upper curve corresponds to developing the remaining 70% of the compounds both overseas and in the U.S. in parallel for a worldwide market, and the lower curve corresponds to developing the remaining 70% of the compounds in the U.S. for a U.S. market only. This phenomenon is dependent upon the values of the input variables and the fixed parameters within the model and would probably be different for different conditions. The broken curves represent the range of values that would have been possible if the development/marketing strategy had not been partially constrained. They are included here for comparison purposes only. It can be seen from this figure that if the yearly research spend is not at least \$26,000,000 then the example case long-range sales goal of \$220,000,000 in 1988 probably will not be realized.

The specific point plotted in Figure 3 represents one possible solution to the example case. For this simulation, the research spend was taken as \$30,000,000 (\$25,000,000 in the U.S. and \$5,000,000 overseas) and the precision specification on 1988 mean sales was set at 5 percent. The development/marketing strategy was chosen to be: 10% developed overseas for an overseas market only, 20% developed in the U.S. for U.S. market only, 30% developed both overseas and in the U.S. in series for a worldwide market, and 40% developed both overseas and in the U.S. in parallel for a worldwide market. The model output for this particular simulation is shown in Figures 4 and 5. Figure 4 shows a recap of the input parameters plus the mean values of all important parameters for the 25 year span of interest. Figure 5



shows the entire range of sales which were simulated for 1988 as well as the important statistical parameters. In this case the mean sales are \$231,205,000 and the probability of achieving the desired goal of \$220,000,000 is approximately 53 percent. Utilizing the information in Figure 5, management can easily see the probabilities of achieving various other sales levels in the target year. It is also evident that increasing the research investment would result in increasing the probability of achieving the target year sales goal. Hence, management may assess the impact that the level of research spend has upon the probability of achieving the desired target year sales goal.

Figure 4
SAMPLE OUTPUT REPORT—PAGE 1

PLANNING AND OPERATIONS
RESEARCH AND DEVELOPMENT SIMULATION MODEL

SIMULATION DATE - 6/16/73 IDENTIFICATION - STRATEGIC VALUE OF R+D * EXAMPLE CASE * NUMBER OF RUNS - 173

RESEARCH BUDGETS (\$M)

UNITED STATES = 25.0 (125 MEN AT 0.200 PER MAN)
OVERSEAS = 5.0 (36 MEN AT 0.140 PER MAN)
INFLATION RATE = 7.5 PERCENT PER YEAR

RESEARCH PRODUCTIVITY

UNITED STATES = 5.0 CPD/YR (0.0400 CPD/MAN-YR)
OVERSEAS = 0.7 CPD/YR (0.0200 CPD/MAN-YR)
OTHER SOURCES = 1.0 CPD/YR

U.S. PRODUCT SIZE SPECIFICATIONS (NORMAL DISTRIBUTION, \$M)

MEAN = 24.0 STANDARD DEVIATION = 8.0
LOWER BOUND = 3.0 UPPER BOUND = 115.0

PRODUCT SALES RATIO

UNITED STATES TO OVERSEAS = 2.0

DEVELOPMENT OPTIONS

OPTION 1 - 0.10 - OVERSEAS DEVELOPMENT FOR OVERSEAS MARKET
OPTION 2 - 0.20 - U.S. DEVELOPMENT FOR U.S. MARKET
OPTION 3 - 0.30 - SERIES DEVELOPMENT FOR WORLDWIDE MARKET
OPTION 4 - 0.00 - SERIES DEVELOPMENT FOR WORLDWIDE MARKET
OPTION 5 - 0.40 - PARALLEL DEVELOPMENT FOR WORLDWIDE MARKET
OPTION 6 - 0.00 - PARALLEL DEVELOPMENT FOR WORLDWIDE MARKET
OPTION 7 - 0.00 - PARALLEL DEVELOPMENT FOR WORLDWIDE MARKET

SIMULATION RESULTS (MEAN VALUES)

YEAR	CMPDS STARTED		CMPDS DROPPED		CMPDS MARKETED		DEVELOPMENT COSTS			GROSS SALES			INFLATION FACTOR
	OVS	USA	OVS	USA	OVS	USA	OVS	USA	TOT	OVS	USA	TOT	
1973	5.3	3.8	0.0	0.0	0.0	0.0	0.2	0.4	0.7	0.0	0.0	0.0	1.075
1974	5.3	3.8	0.0	0.0	0.0	0.0	0.9	1.9	2.8	0.0	0.0	0.0	1.155
1975	5.5	5.2	1.0	0.8	0.0	0.0	1.5	4.1	5.6	0.0	0.0	0.0	1.242
1976	5.4	5.2	1.9	1.9	0.0	0.0	2.0	6.1	8.2	0.0	0.0	0.0	1.335
1977	5.1	4.9	3.1	3.3	0.0	0.0	2.3	7.2	9.5	0.0	0.0	0.0	1.435
1978	5.8	5.1	4.1	4.5	1.2	0.3	2.3	7.4	9.8	1.9	1.3	3.2	1.543
1979	5.0	5.3	4.3	4.9	1.2	0.5	2.3	7.4	9.7	7.2	4.9	12.1	1.659
1980	5.4	5.1	4.2	5.0	1.2	0.6	2.3	7.5	9.8	15.0	11.3	26.3	1.783
1981	5.2	4.9	4.1	4.9	1.1	0.7	2.3	7.4	9.8	24.1	20.0	44.2	1.917
1982	5.3	5.2	4.4	4.9	1.0	0.6	2.3	7.3	9.6	34.4	31.4	65.9	2.061
1983	5.2	4.9	4.1	4.7	1.3	0.7	2.2	7.4	9.7	45.7	44.0	89.8	2.215
1984	5.3	5.0	4.1	4.9	1.0	0.5	2.3	7.3	9.6	58.1	57.4	115.5	2.381
1985	5.3	4.9	4.0	4.9	1.2	0.7	2.3	7.3	9.6	70.9	71.4	142.3	2.560
1986	5.4	5.1	4.1	4.6	1.2	0.6	2.2	7.3	9.6	84.6	87.0	171.7	2.752
1987	5.2	5.1	4.3	5.0	1.0	0.5	2.3	7.3	9.7	98.7	102.6	201.4	2.958
1988	5.1	5.1	4.0	4.7	1.1	0.6	2.3	7.4	9.7	112.8	118.3	231.2	3.180
1989	5.4	5.0	4.0	4.8	1.0	0.5	2.3	7.5	9.8	126.9	133.7	260.6	3.419
1990	5.5	5.1	4.0	4.9	1.1	0.6	2.3	7.5	9.8	140.6	148.7	289.4	3.675
1991	5.3	5.0	4.1	4.7	1.2	0.6	2.3	7.5	9.9	154.4	163.8	318.3	3.951
1992	5.2	5.2	4.3	5.0	1.1	0.6	2.3	7.4	9.7	167.8	178.8	346.7	4.247
1993	5.3	5.1	4.3	5.1	1.1	0.6	2.3	7.2	9.5	181.3	194.2	375.6	4.566
1994	5.4	5.1	4.2	5.0	1.1	0.7	2.3	7.3	9.6	195.3	209.8	405.2	4.908
1995	5.1	5.1	4.3	4.7	1.1	0.6	2.3	7.3	9.7	209.3	225.4	434.8	5.277
1996	5.6	5.0	4.1	4.7	1.0	0.6	2.3	7.3	9.6	223.0	241.3	464.4	5.672
1997	5.0	5.0	4.4	5.1	1.2	0.6	2.2	7.1	9.4	237.0	257.7	494.7	6.098

U.S. PRODUCT SIZE STATISTICS

MEAN = 23.9 STD DEV = 7.9

OBSERVATIONS AND CONCLUSIONS

The model described here represents an initial step in permitting management to examine the relationships between investment and future demands for growth from new products. It also provides one approach to evaluating the impact of a number of R&D development/marketing strategies on the ability to achieve the desired level of growth.

Long range planning is an essential activity in any technologically-oriented industry and is especially important in the pharmaceutical industry, where the development process itself is relatively long (anywhere from 3 to 7 years) for any specific product. It is not sufficient to merely state long-range objectives and goals and assume that they will somehow be achieved through careful, although relatively short-range, operational planning. It is necessary to insure, within certain constraints and assumptions, that the level of investment commitment is consistent with the established goal. This model facilitates the generation of this type of information and thus is a valuable aid in the long-range, strategic planning process.

BIBLIOGRAPHY

1. Old, B.S., "Trends in Research and Development", Arthur D. Little, Inc., March 1972.
2. Friedman, J.J. and Friedman, M.N., "Relative Profitability and Monopoly Power", The Conference Board Record, 9:49, 1972.
3. Foster, R.N., "Estimating Research Payoff by the Internal Rate of Return Method", Res. Mgt., 14:27, 1971.
4. Galloway, E.C., "Evaluating R&D Performance - Keep it Simple", Res. Mgt., 14:51, 1971.
5. Faymour, M.E., "The Value of R&D in Relation to Net Sales", Res. Mgt., 15:47, 1972.
6. Newman, M.S., "Equating Return From R&D Expenditures", Financial Executive, 36:26, 1968.
7. Pearson, A.W. and Selen, J.M., "Assessing Research and Development", Science Journal, August 1969, p. 79.
8. Baker, N.R. and Pound, W.H., "R&D Project Selection - Where We Stand", IEEE Transactions on Engineering Management, EM-11:124, 1964.
9. Cetron, M.J., Martino, J. and Roepche, L., "The Selection of R&D Program Content - Survey of Quantitative Methods", IEEE Transactions on Engineering Management, EM-14:4, 1967.
10. Bobis, A.H., Cooke, T.F. and Paden, L.H., "A Funds Allocation Method to Improve the Odds for Research Success", Res. Mgt., 14:34, 1971.
11. Lindsay, E.M., "Financial Management of R&D: Planning and Budgeting, Project Authorization, Financial Reporting", Res. Mgt., 14:58, 1971.
12. Souder, W.E., "Cost/Progress - A Pattern for Operational Planning", Managerial Planning, Jan. - Feb., 1969, p. 1.
13. Jay, R.A., "Effective Use of Research and Development Resources", Adv. Management Journal, 32:34, 1967.