STRUCTURING THE NEW PRODUCT DEVELOPMENT PIPELINE

Min Ding and Jehoshua Eliashberg

Marketing Department
The Wharton School
University of Pennsylvania

April 12, 2000

The authors acknowledge various constructive comments from participants in presentations given at the Marketing Science Conference at Syracuse, the Wharton School, Emory University, University of Florida, Cornell University, Pennsylvania State University, MIT, University of Rochester, Roterdam School of Management, University of Pittsburgh. Support from ISBM at Pennsylvania State University is gratefully acknowledged. The authors also want to thank the pharmaceutical industry experts who have graciously participated in our survey as well as for providing insightful feedback.

ABSTRACT

In many new product development (NPD) situations, the development process is characterized by uncertainty, and no single development approach (e.g., a particular technological version) will necessarily lead to a successful product. In order to increase the likelihood of having at least one successful product at the end of the NPD process, managers may choose to fund simultaneously multiple approaches. This strategy becomes a lot more complicated when the number of stages (e.g., concept screening, prototype testing) characterizing the NPD process increases. The managerial challenge is thus to construct *ex-ante* an appropriate NPD pipeline by choosing the right (i.e., optimal) number of approaches to be funded simultaneously at each stage. The so-called pipeline problem is present in other contexts as well. These include advertising copy selection, national rollout of new products with test markets as well as situations such as recruiting for academic positions. In this paper, we present a normative model for structuring such pipelines -- using a decision theoretic framework. The model incorporates inter-disciplinary considerations such as R&D, marketing, and product development. The structure of the optimal pipeline is driven by three critical factors: the cost of a development approach, its probability of survival, and the expected profitability if a successful product is developed and launched. We illustrate the workability and implications of the model by applying it to a number of real-world scenarios in the pharmaceutical industry, and by comparing its normative pipelines recommendations against actual pipelines. We also present general qualitative insights with regard to the optimal pipeline structure under two scenarios: one-stage NPD and two-stage NPD. Our results suggest, in general, that the pharmaceutical firms we studied employ narrower pipelines for their new drugs development than they should, and thereby they underspend on R&D.

1. INTRODUCTION

In many situations, there is more than one way (approach) to develop a new product in order to satisfy some specific consumer needs and capture a business opportunity. In cases where no dominant approach can be identified *a priori*, managers must decide how many approaches should be supported in parallel. Consider the following problem as a case in point -- the development of a preventive AIDS vaccine.

Acquired Immunodeficiency Syndrome (AIDS) is caused by the human immunodeficiency virus (HIV) and "is now the leading cause of death among adults between the ages of 25 and 44 -- the age range of more than half the nation's 126 million workers." (Gerson, 1997). The cumulative (national) costs of treating all people with the human immunodeficiency virus (HIV) reached $10.3 billion in 1992 and has been increasing ever since (Hellinger, 1992). The severity of this disease is further underscored by its infectious nature. This presents a significant business opportunity to the pharmaceuticals industry and, at the same time, an even bigger concern for public policy makers. As a result, substantial effort has been made, both by pharmaceutical/biotechnology industries and the U.S. government, to develop a preventive vaccine for HIV. May18, 1998 was even designated the first HIV/AIDS vaccine awareness day. To increase the probability of success, many prototype vaccines have been developed based on different mechanisms, including subunit vaccine, recombinant vector vaccine, peptide vaccine, virus-like particle vaccine, anti-idiotype vaccine, plasmid DNA vaccine, whole-inactivated virus vaccine, and live-attenuated virus vaccine. A number of prototype AIDS vaccines are being tested now in Phase I and II human clinical trials, sponsored by various companies (e.g., Bristol-Meyers Squibb, British Biotech PLC, Chiron/BIOCINE, Genentech, and Pasteur Merieux Connaught), and organized by the National Institute for Allergy and Infectious Disease (NIAID, which has a branch specifically formed to organize AIDS vaccine clinical trials). By February 1998, NIAID has conducted 29 phases I or II clinical trials with 19 different vaccine candidates (see NIAID website).

While the goal is to obtain one successful preventive vaccine at the end, both companies and the public policy makers believe that more than one approach should be pursued concurrently (Henderson, 1996). They, however, differ in their opinions about what is the right number of approaches that should be pursued simultaneously. The evidence suggests that while most of the companies mentioned above have supported more than one prototype vaccines, they rarely pursue more than three simultaneously. They seem to believe this strategy is in their best interest. The public policy makers, on the other hand, seem to believe that even the combined number of known prototype vaccines (larger than 20) is not large enough. A government sponsored review indicates "the dilemma … is related to the paucity of promising new AIDS vaccine candidates." To address this problem, a new two-year innovation grants were awarded in FY1997 through NIAID to encourage new ideas of prototype AIDS vaccines (NIH website).

The AIDS vaccine example leads to the critical question faced by a pharmaceutical company: what is the optimal number of prototype AIDS vaccines that should be pursued simultaneously at each of the clinical trial phases? This is the essence of structuring an optimal pipeline. The general pipeline problem could be defined as: there exists a business opportunity (or payoff) that could be captured by launching an appropriate new product. Multiple development approaches may be chosen and funded to develop this new product, none of which guarantees a successful product at the end of new product development (NPD) process. The NPD process is composed of multiples stages and the managerial challenge is to determine whether single or multiple (if multiple, how many) approaches should be funded at each of these stages. This paper addresses this problem.

The pipeline problem is highly relevant in many other contexts. For example, the development of an advertising campaign also involves the structuring of an optimal pipeline. In order to develop a successful advertising campaign, the ad agency usually creates multiple copies for the campaign. From this pool of potential ads, a subset is selected for copy testing. The copy testing itself may be done in a multi-stage fashion. For instance, focus groups could be used to do the first round screening, followed by second round screening in test markets. After reviewing

the results, one final copy is selected for the campaign. Deciding on how many test markets to employ prior to national rollout of a new product represents another pipeline structuring business problem. The pipeline problem is critical in non-business situations as well. One example is academic recruitment. The first stage of screening involves reviewing application package (c.v., recommendation, etc.). The second stage usually takes place in a conference. The fortunate ones will be invited to campus for the third stage of the process. Finally, schools need to decide how many offers to make, given that not everybody will accept the offer.

The rest of this paper focuses our modeling and analyzing the pipeline structure problem in the context of multiple-stage NPD. We take an interdisciplinary perspective by incorporating R&D, marketing, and product development considerations. The paper is organized as follows. In section 2 we review the literature that is most relevant to the problem. We then present (in section 3) the model formulation and its analytical implications. In section 4 we move from theory to practice, demonstrating the workability and the implications of the model by implementing it in a number of real-world situations. Section 5 provides concluding remarks as well as a discussion and suggestions for further research.

2. RELEVANT LITERATURE

Two streams of literature have studied problems related to the one of concern in this paper – marketing and R&D. The marketing literature has examined issues related to pipeline structuring, mainly for one-stage processes as well as issues related to managerial fallacies in pulling the plug to stop new product development projects. The R&D literature has focused mainly on resource allocation and portfolio models, employing mainly static mathematical programming models.

Some simple heuristics for structuring pipelines for NPD, and their corresponding budgeting implications, can be found in marketing management (Kotler, 1994) and NPD (Urban and Hauser, 1993) textbooks. The guidelines given in these books, however, focus only on the pass ratios and they consider the process deterministically. Figure 1 illustrates this line of thinking for a firm whose objective is to launch one successful new product.

Insert Figure 1 Here

Although the pass ratios (also known as probability of survival) represent indeed a critical driver in structuring the NPD pipeline, they are not the only driver. Gross (1972) and Feinberg and Huber (1996), for instance, recognized it in their models of selecting advertising copies and the number of candidates to be invited for campus interviews in academic recruitment, respectively. Their models are, however, one-stage models. Srinivasan et al., (1997) focused on the concept selection stage of NPD and studied the question of "how many concepts should be carried forward?" This paper offers empirical support to the idea that more detailed design work should be performed on several concepts in parallel (before selecting the final concept) in some NPD situations. Similar to Gross (1972) and Feinberg and Huber (1996), this paper is framed as a one-stage problem. A recent working paper by Dahan (1998) examines a

related problem. He also treats the entire NPD process as a single-stage problem, and asks the question of how many such stages (repeated development) should be considered by the firm, and within each repeat, how many approaches should be funded simultaneously. Relatedly, Bhattacharya, Krishnan, and Mahajan (1998) found that the traditional practice, recommended in the literature, of reaching a sharp definition for the new product early in the NPD process (i.e., support one prototype), may not be optimal, desirable or even feasible in some dynamic situations. Boulding, Morgan, and Staelin (1997) demonstrate experimentally that the actual pipeline observed in practice may be sub-optimal due to managerial misjudgment and/or fallacies. The authors suggested that a predetermined budgeting rule will alleviate such problems.

Managers responsible for developing really new products often recognize that attempting to capture the business opportunity with multiple approaches is inherently better (but more costly) than relying only on a single approach (This was indicated by executives we interviewed, who are responsible for resource allocation). A recent article (WSJ, 1999) cited " Werner Schiebler, technology license director of Hoechst Marion Roussel, said … 'We need to … (be) doing things in parallel.' That means using more leads to develop a compound through phase I and II trials …". This practice of funding multiple alternatives concurrently has been observed in the development of "really new products" in other industries as well. During the development of the videotape recorder technology, for example, Sony had pursued 10 major approaches where each approach had two to three subsystems alternatives (Rosenbloom and Cusumano, 1987). AT&T and the major oil companies usually start several programs in parallel before finally selecting a technology for system-wide usage (Quinn, 1985). According to the SVP and CTO of Texas Instruments, TI had pursued several alternative approaches on the 16-megabit DRAM chip while collaborating with Hitachi at the same time (Dreyfuss et al, 1990). During the development of Celcor (a honeycomb structure used to hold catalyst in a catalytic converter) at Corning Incorporated, six R&D teams had worked concurrently on a same problem using different approaches (Morone, 1993). Pursuing multiple approaches (parallel new product development) is

also common from public policy standpoint. The Department of Defense of the U.S. government often support multiple approaches simultaneously.

Firms who understand the importance of multiple approaches, may run, however, into the risk of funding too many (if not all) proposed alternative approaches for a single business opportunity and thus they may be running into the problem of overspending. That is, managers may not realize that sometimes they should only fund a subset of approaches and invest the saved money elsewhere. Sometimes, a strictly sequential NPD process would be appropriate. A sequential approach develops, tests, and launches one approach at a time until one alternative becomes successful (Chun 1994). That is, it takes the same approach all the way through the process until the uncertainty surrounding its performance is completely resolved. By contrast, a parallel new product development procedure will pursue more than one approach at the same time. Since only one commercially successful product will be needed, there is potential waste of redundant new product development resources in the parallel approach. On the other hand, the parallel approach helps the company cope with uncertainties in development, motivates people through competition, and improves the amount and quality of information available for making final choices on scale-ups or introduction (Quinn, 1996). The decision to adopt either sequential or parallel approach depends on several factors (Abernathy and Rosenbloom, 1968, 1969): the probabilities of stage-wise success, the funding level for each research alternatives, the expected profit, and the constraint of new product development time. If the benefits of parallel approach outweigh the extra new product development investment, then parallel approach should be used. The sequential approach should be used if the opposite is true.

The various pipelines observed in practice, could thus be grouped into two categories (Figure 2). The first category is Funnel structure in which the number of alternatives that a firm is committed to at each stage gradually decreases as the development process moves towards completion. According to the second category, Tunnel, the firm makes a commitment to the same number of alternatives at each NPD stage. The two different pipelines (funnel vs. tunnel) have, of course, financial budgeting as well as organizational implications. A tunnel, for

instance, may reflect management commitment to a stable R&D personnel and to their emotional attachment to the project they have been assigned to. The managerial challenge of determining the optimal pipeline structure for a specific situation, however, has not been addressed adequately in the literature.

Insert Figure 2 Here

Another stream of literature that is somewhat related to the pipeline structuring problem can be founded in the R&D literature. There is a copious collection of resource allocation models for observing evolutionary new products. The early development of the literature has been reviewed by Cetron et al (1967) and Souder (1978). Reviews could be found in Jackson (1983), Souder and Mandakovic (1986), Steele (1988), Weber et al (1990), and Schmidt and Freeland (1992). According to Souder and Mandakovic (1986), the population of project selections models could be categorized as classical methods, portfolio models, project evaluation techniques, and organizational decision methods. Classical methods try to prioritize available projects and fund the projects that are on top of the list. Some of the most common classical methods are profiles, checklists, scoring models, and economic indexes. Classical models are simple to use whenever the projects can be prioritized. On the other hand, they fail to reflect the dynamic decision making process. Portfolio models are usually structured as an optimization problem, the goal of these models are usually to optimize an objective function under a given set of constraints (Schmidt and Freeland, 1992). The most fundamental mathematical programming tool employed is linear programming. Linear programming based models have several weaknesses. They do not handle the interdependencies between new product development projects and they are static. Project evaluation techniques are methods developed to evaluate individual new product development projects, including goal-contribution models, decision tree, utility theory, Monte Carlo simulation, and risk analysis models. To our knowledge, however, none of these methods has been used to address the pipeline structuring problem of concern here.

While some existing studies have addressed the risk issue associated with developing new products, to our knowledge, no study/model has been conducted to investigate the optimality of parallel/sequential resource allocation for new products in a dynamic multi-stage decision making framework and the extent to which companies over/under spend on the development of such new products. Under the (rather strong) assumption that every approach will eventually succeed, optimal parallel approach problem has been investigated allowing managers to make either one intermediate decision (Nelson ,1961) or multiple intermediate decisions (Marschak et al, 1967). However, these normative models could not be used for developing new products where probability of ultimate success (p) is less than 1. Other researchers have considered this scenario ($p<1$) but under fairly simplistic conditions. Abernathy and Rosenbloom (1968, 1969) formulated a model with two alternative approaches. Dean and Hauser (1967) formulated a model for the new product development planning of the Army Materiel Command with more than two alternative approaches. These studies, however, did not explicitly incorporate the multiple stage and the dynamic nature of decision making associated with the development of new products. Often, the process is considered exogenously as funnel, where the number of options pursued becomes smaller as the project progresses towards launch.

3. MODEL FOUMULATION AND ITS IMPLICATIONS

We begin by introducing the basic model that addresses the issues discussed earlier. Relaxation of the key assumptions which leads to a refined model are discussed in section 5. Relaxation of other (non-key) assumptions is discussed in this section.

Key Assumptions:

Several assumptions, validated by interviews with executives in pharmaceutical industry, have been made in developing the basic model:

1. Multiple approaches may be taken to develop the new product and there is no dominant approach that guarantees success. Hence, initially we assume that the probabilities of success and the costs incurred within the various new product development stages are the same for all alternative approaches. They may vary, however, across stages.
2. The expected profit from the business opportunity can be captured if one successful product is launched. Profits generated by additional successful products are negligible.
3. The firm does not repeat any of the new product development stages, nor does it repeat the whole new product development process.

These three basic assumptions establish a useful framework. We observe that in practice Assumption 1 is employed. One company we surveyed makes even a more restrictive assumption than assumption 1 by not allowing for variations of probabilities of success and costs across stages. Assumption 2 is quite reasonable as judged by executives in the pharmaceutical industry we have interviewed. Assumption 3 may seem quite restrictive at first, but it is an accurate description of many really new product development scenarios including drugs. For instance, in many situations, a firm can capture a large market share if it launches its product first (pioneer advantage) and thus becomes the market leader (Bond and Lean 1977, Parry and Bass, 1990, Urban et al, 1986). Under this scenario, the potential profit of a late launch (due to repetition of certain new product development stages) is minuscule compared to launching the product first.

To focus on the key drivers of the pipeline structure, we assume that all monetary terms have been transformed into present value based on the cost of capital and time. In analyzing the NPD process below, we move backwards, that is from product launch to the early stages of the NPD process.

Stage 0 (just prior to launch):

The expected degree of market success of any new product depends on two factors. First, whether the product is likely to meet consumers' needs. Second, how many other products it is likely to compete with. For the sake of exposition, we invoke, as an example, the assumption of no obvious product differentiation in the market. That is, all successfully launched products will divide the market equally among them. For example, a firm will capture the whole business opportunity if no competitor has successfully developed a similar product, while it will capture 1/3 of the market if two of its competitors have launched simultaneously similar products. If there are **m** competitors in such market, each has probability **p** of developing at least one successful product, one way to express the expected profit of any firm, viewed just prior to launch is:

$$E[\pi_0(s_1)] = \begin{cases} 0 & \text{if } s_1 = 0 \\ \alpha R \sum_{i=0}^{m} \left[\frac{1}{1+i}\binom{m}{i} p^i (1-p)^{m-i}\right] & \text{if } s_1 \geq 1 \end{cases} \quad (1)$$

s_1: the number of projects successfully passed the completion stage.

$E[\pi_0(s_1)]$: the expected cumulative profit when viewed from stage 0.

R: the expected cumulative revenue for a business opportunity;

α: the average contribution rate (the pretax profit and development cost as a percentage of revenue);

i: the number of competitors who have developed at least one successful product;

The probability of success (p) in the binomial distribution in (1) represents the (equal) strength of each firm in capturing the business opportunity. Since the number of competing firms

(m) is usually quite small, it should be fairly easy to modify equation (1) and allow different probabilities of success for different firms.

Of course, many other approaches can be taken to model $E[\pi_0(s_1)]$. An alternative method, based on trial and repeat behavior, could be used to estimate the magnitude of business opportunity for frequently purchased products (e.g., drugs treating chronic diseases). This method is described in section 4. It has been applied in estimating the business opportunities that faced by firms for seven new drug development situations.

Stage 1 (last NPD stage):

The probability of having a certain number of successful projects at the end of stage 1 can be modeled as a binomial distribution:

$$Pr(s_1 | p_1, n_1) = \binom{n_1}{s_1} p_1^{s_1}(1 - p_1)^{n_1 - s_1} \tag{2}$$

n_i: the number of approaches initiated in stage i.

s_i: the number of approaches which have successfully passed stage i.

p_i: the probability of success per approach at stage i.

$Pr(s_i | p_i, n_i)$: the probability of having s_i successful approaches in the end of stage i given p_i and n_i, modeled as a binomial distribution as in stage 1.

The expected profit at this stage can be expressed as:

$$\begin{aligned} E[\pi_1(n_1)] &= \Pr(s_1 = 0 \mid p_1, n_1) E[\pi_0(s_1 = 0)] + \Pr(s_1 > 0 \mid p_1, n_1) E[\pi_0(s_1 > 0)] - n_1 c_1 \\ &= [1 - (1 - p_1)^{n_1}] E[\pi_0(s_1 > 0)] - n_1 c_1 \end{aligned} \tag{3}$$

c_i: the cost of funding one approach at stage i.

It is straightforward to establish expressions for the variance and the probability of obtaining at least one successful product at this stage. They are, respectively,

$$V[\pi_1(n_1)] = (1 - p_1)^{n_1}[1 - (1 - p_1)^{n_1}]\{E[\pi_0(s_1 > 0)]\}^2 \quad (4)$$

$$L_1(n_1) = 1 - (1 - p_1)^{n_1} \quad (5)$$

Following similar arguments we can also show that:

Stage k (k ³ 2):

The expected profit at this stage can be formulated as:

$$E[\pi_k(n_k)] = \sum_{s_k = n^*_{k-1}}^{n_k} \left[\Pr(s_k | p_k, n_k) E[\pi_{k-1}(n^*_{k-1})]\right] + \sum_{s_k=0}^{n^*_{k-1}-1} \left[\Pr(s_k | p_k, n_k) E[\pi_{k-1}(s_k)]\right] - n_k c_k \quad (6)$$

Parameters are defined as in stage 0 and 1.

The variance for stage k and probability of obtaining at least one successful product at the end of NPD pipeline could be calculated, respectively, by:

$$\begin{aligned} V[\pi_k(n_k)] = & \sum_{s_k = n^*_{k-1}}^{n_k} \left\{\Pr(s_k | p_k, n_k)\left[E\left[\pi_{k-1}(n^*_{k-1})\right]\right]^2\right\} + \sum_{s_k=0}^{n^*_{k-1}-1} \left\{\Pr(s_k | p_k, n_k)\left[E\left[\pi_{k-1}(s_k)\right]\right]^2\right\} \\ & - \left\{\sum_{s_k = n^*_{k-1}}^{n_k} \left[\Pr(s_k | p_k, n_k) E\left[\pi_{k-1}(n^*_{k-1})\right]\right] + \sum_{s_k=0}^{n^*_{k-1}-1} \left[\Pr(s_k | p_k, n_k) E\left[\pi_{k-1}(s_k)\right]\right]\right\}^2 \end{aligned} \quad (7)$$

$$L_k(n_k) = \sum_{s_k = n_{k-1}*}^{n_k} \left[\Pr(s_k \mid p_k, n_k) L_{k-1}(n_{k-1}*)\right] + \sum_{s_k=1}^{n_{k-1}*} \left[\Pr(s_k \mid p_k, n_k) L_{k-1}(s_k)\right] \quad (8)$$

Having set up the model, it is now possible to investigate its implications. We begin with equation (3).

Proposition 1:

$E[\pi_1(n_1)]$ is a strictly concave function with a unique global maximum at n_1* which equals to:

$$n_1* = \frac{\ln\left(\dfrac{-c_1}{E[\pi_0(s_1 > 0)]\ln(1 - p_1)}\right)}{\ln(1 - p_1)} \quad (9)$$

Proof: Strict concavity can be shown by:

$$\frac{\partial^2 E[\pi_1(n_1)]}{\partial n_1^2} = -E[\pi_0(s_1 > 0)][\ln(1 - p_1)]^2 (1 - p_1)^{n_1} < 0 \ . \qquad (10)$$

The global maximum could be obtained by solving:

$$\frac{\partial E[\pi_1(n_1)]}{\partial n_1} = 0 \qquad (11)$$

Q.E.D.

Corollary 1:

- n_1* in (9) increases as the ratio between cost per approach and expected cumulative profit ($c_1/E[\pi_0(s_1>0)]$) decreases.
- n_1* in (9) increases when p_1 increases from 0 to p_1*, peaks at p_1*, and decreases when p_1 decreases from p_1* to 1. p_1* is defined as: $p_1^* = 1 - e^{\frac{-c_1}{(1/e)*E[\pi_0(s_1>0)]}}$.

The proof is straightforward.

Investigating equation (6) for optimality becomes less tractable. However, it can be shown that:

Lemma 1:

$E[\pi_k(n_k)]$ is a strictly concave function with a unique global maximum at n_k*, where n_k* is implicitly defined by the following equation:

$$p_k \sum_{s_k=0}^{n_{k-1}^*-1} \{\Pr(s_k \mid p_k, n_k^*)[E[\pi_{k-1}(s_k+1)] - E[\pi_{k-1}(s_k)]]\} - c_k = 0 \qquad (12)$$

where k is a positive integer and k ³ 2.

Proof: see Appendix A.

The next proposition provides more insights into the nature of n_k*:

Proposition 2:

For stage k (k ³ 2):

n_k* in (12) increases when c_k decreases;

n_k* in (12) reaches maximum at an interior value of p_k(between 0 and 1);

An approximation (upper bound) for n_k* in (12) is given by:

$$n_k^* < \frac{\ln c_k - \ln(p_1 p_2 \cdots p_k E[\pi_0(s_1 > 0)])}{\ln(1 - p_1 p_2 \cdots p_k)}. \qquad (13)$$

Proof: See Appendix A.

Proposition 3:

If, for all k (k ³ 2),

$$c_k < p_k \sum_{s_k=0}^{n_{k-1}^*-1} \{\Pr(s_k \mid p_k, n_{k-1}^*)[E[\pi_{k-1}(p_{k-1}, s_k + 1, c_{k-1}, E\pi_{k-2})] - E[\pi_{k-1}(p_{k-1}, s_k, c_{k-1}, E\pi_{k-2})]]\} = \bar{c}_k \qquad (14)$$

then the NPD pipeline will take the shape of a **funnel** (n_k* > n_{k-1}*). Otherwise the pipeline will be a **tunnel** shape (n_k* = n_{k-1}*).

Proof: See Appendix A.

Based on Propositions 1-3, the decision rules for structuring a three-stage optimal NPD pipeline are captured by a decision tree (see Figure 3). This decision tree could be easily extrapolated to k stages. When supplied with the required inputs (parameters) for a given NPD project, the model can then produce a specific decision tree to be used to construct the optimal pipeline by the managers.

Insert Figure 3 Here

Discussion:

It is possible to represent geometrically the dependence of the optimal pipeline on the three key problem's drivers: expected profit, cost and probability of success for one stage scenario. See Figure 4 (Appendix B provides the formal analysis).

Insert Figure 4 Here

Figure 4 represents a one-stage pipeline with some given $E[\pi_0(s_1 > 0)]$. Changes in $E[\pi_0(s_1 > 0)]$ can be captured by increases in the horizontal line where $E[\pi_0(s_1 > 0)]$ is currently fixed and thereby expanding the diagram upwards. Any point in the rectangle captures a combination of (p_1, c_1). The line and the curve shown in Figure 4 represent boundaries. Point E, for Example, represents a pipeline situation characterized as (p_{1E}, c_{1E}) where the normative number of development approaches is equal to 3.

The numbers in the figure refer to the optimal numbers of approaches to be funded for various regions in the parameter space. Three key insights for the one-stage scenario are: (1) the managerial decision is reduced to a binary choice (fund a single approach or none) when the cost per approach (c_1) is larger than ¼ of the expected market potential ($E[\pi_0(s_1 > 0)]$). (2) for a fixed probability of success (e.g., p_{1E}), the optimal number (n_1*) increases when the cost (c_1) decreases. (3) for a fixed cost (e.g., c_{1E}), the optimal number (n_1*) first increases then decreases as p_1 increases from 0 to 1. The intuition behind the last insight is that the marginal benefit of an additional approach is small under either small p_1 (this additional approach is less likely to be successful) or large p_1 (a successful product is likely to be developed by other approaches).

The story becomes more complex in a two-stage scenario (see Figure 5 and Appendix B for formal analysis). There are essentially three types of normative pipelines that will emerge for the two-stage process, namely,

M-M: fund multiple approaches in both stages;

M-S: fund multiple approaches in the initial NPD stage (e.g., concept screening), and focus on one approach in the second NPD stage (e.g., prototype testing);

S-S: fund a single approach in both stages;

Insert Figure 5 Here

Given that $E[\pi_2(n_2)]$ is concave with respect to n_2 (Lemma 1), the corresponding conditions under which each of the three scenarios is optimal could thus be simplified as:

Scenario	First NPD Stage (e.g., concept screening)	Last NPD Stage (e.g., prototype development)
S-S	$given \quad n_1^* = 1$ $E[\pi_2(1)] > E[\pi_2(0)] \quad and \quad E[\pi_2(1)] > E[\pi_2(2)]$	$E[\pi_1(1)] > E[\pi_1(0)] \quad and \quad E[\pi_1(1)] > E[\pi_1(2)]$
M-S	$given \quad n_1^* = 1$ $E[\pi_2(2)] > E[\pi_2(1)]$	$E[\pi_1(1)] > E[\pi_1(0)] \quad and \quad E[\pi_1(1)] > E[\pi_1(2)]$
S-M [1] **(reduced to S-S)**	$given \quad n_1^* > 1$ $E[\pi_2(1)] > E[\pi_2(0)] \quad and \quad E[\pi_2(1)] > E[\pi_2(2)]$	$E[\pi_1(2)] > E[\pi_1(1)]$
M-M	$given \quad n_1^* > 1$ $E[\pi_2(2)] > E[\pi_2(1)]$	$E[\pi_1(2)] > E[\pi_1(1)]$

Thus, there are two conceptually different determinants that affect the structure of the two-stage pipeline. One is the overall profile of the NP (the relationship among c, p where

[1] Note it is possible to have a scenario where the optimal pipeline is a reverse funnel, i.e., support one approach in the initial stage and multiple approaches in the second stage (S-M). Under the logical constraint that an approach, which is developed internally, must pass all earlier stages in order to be available for later development, this scenario is reduced to S-S scenario in the analysis followed. This scenario is realistic, however, in pharmaceutical industry where pharmaceutical companies let external biotech firms do the initial development and then they acquire (or form alliance with the biotech firms) a new compound that survived the earlier stages at the biotech firms.

$c=c_1+c_2$ and $p=p_1*p_2$, and $E[\pi_0(s_1 > 0)]$). In Appendix B (B10) we provide precise definitions for low, moderate, and high overall costs. The other is the distribution of the overall cost and probability of survival between the two stages (c_2/c and p_2/p) which are represented by the axes in Figure 5. The boundaries for different pipelines are shown in Figure 5, where each rectangle represents the results for each overall cost region.

For easier interpretation, we sum up the general insights with regard to the optimal pipeline structure under the two-stage scenario in Table 1.

Insert Table 1 Here

Table 1 suggests that a firm should always cast a wide net (fund multiple approaches) in the first stage and focus on one approach in the second stage if the screening (first stage) is effective (remove most of the uncertainty) and cheap. This insight is fairly robust with respect to the overall profile of the project (similar across the three regions). It should be pointed out that the exact definition of effective and cheap screening (as other similar terms used here) is relative and (the sizes of areas that fit this description) it may differ across the three groups of overall profiles. The optimal pipeline is also relatively straightforward for semi-effective and medium cost screening and again fairly robust with respect to the overall profile of the project. Under this situation, the firm should fund multiple approaches at both stages.

The pipeline structuring strategy becomes complicated when the screening is ineffective or expensive. Under this condition, M-M strategy should be used when the overall cost (c) is low; S-S strategy should be used when the overall cost is moderate; and the project should be abandoned (does not fund any approach) when the overall cost is high. For all other screening conditions, S-S strategy should be used except that a firm should adopt M-M when the overall cost is low.

There are some exceptions to the insights summarized in Table 1. First of all, even though we have stated that a cheap AND effective screening is required for the M-S strategy to be optimal, this requirement is relaxed to include expensive but very effective screening under the Low-Overall-Cost scenario and very cheap but ineffective screening under the Moderate-Overall-Cost scenario (see Figure 5). The second major exception is that the areas in Figure 5 with M-M as its optimal strategy may not exist under some situations (e.g., when the overall cost approaches the high end within each profile group).

So far we discussed optimal structures for one and two-stage processes, separately, some insights can also be obtained by examining and comparing the economic implications of multiple vis-à-vis single-stage development processes. The following simple example sheds some light into such comparison (see Figure 6). Note that in both cases the probability of ultimately success

---

Insert Figure 6 Here

---

is 0.36 and the total funding required is $10m. The best decision in the single-stage case is to GO if X>10/0.36 (see Decision Tree #1, Figure 7). The best decisions in the two-stage case is to GO

Insert Figure 7 Here

---

with stage 2 if X>6/0.6 **and** to GO with stage 1 if X>7.6/0.36 (see Decision Tree #2, Figure 7). This implies that the firm should fund ***both*** stage 1 and 2 if X>7.6/0.36. Since 7.6/0.36<10/0.36, this simple example illustrates that multiple (two)-stage development processes can lead to pursuing smaller business opportunities.

4. FROM THEORY TO PRACTICE

In this section, we will demonstrate the implementability of the model and its implications by studying first the motivating example discussed in section 1, the HIV vaccine. Next we analyze seven other new drug development cases. We will compare the models' (normative) recommendations to actual data. We will also demonstrate how the model can be used as a simulation tool to provide managers with a confidence region for its recommendations in the face of uncertainty. This is achieved by varying systematically the key parameter values.

The expected profit (equation 1) of an AIDS vaccine for any firm engaged in developing it can be calculated following the method used by Grabowski and Vernon (1990) with some modifications. The return to the firm from treating a person infected with AIDS is estimated to be $102,000 (Hellinger, 1992). The number of people infected with HIV every year is estimated to be at least 40,000 (Office of AIDS Research, NIH). Within one year of introducing a successful AIDS vaccine, the entire U.S. population (280 million) will be inoculated. The number of firms currently engaged in active development of preventive AIDS vaccines with NIAID is 12. Assuming each firm starts with 3 prototype vaccines, given the various phase-wise survival probabilities (see Table 2), the expected probability of success for each firm will then be $1 - (1 - 0.75 \times 0.48 \times 0.63)^3 = 0.5$. Thus the expected profit for any one firm, given that it succeeds in developing a vaccine, will be the expected market share (0.17) times the total business opportunity (equation 1). The cumulative cash flow (profit plus R&D costs) can be obtained using an average contribution rate of 40%, which is then adjusted for 36% tax rate and discounted using 10% cost of capital assuming 10 year development cycle prior to product launch. Finally, this domestic cumulative cash flow can be extrapolated to world-wide cumulative cash flow using a multiplier of 1.9 following Grabowski and Vernon (1990). The world-wide firm's expected profit will be

$$E[\pi_0(s_1 > 0)] = (\$102{,}000 \times \frac{40{,}000}{280{,}000{,}000}) \times 280{,}000{,}000 \times 0.17 \times 40\% \times 64\% \times 1.1^{-10} \times 1.9 = \$130\ m \qquad (15)$$

The expected benefit for the public policy makers, however, is quite different. In this analysis, we use the amount of national costs associated with treating AIDS over a long time

horizon (the resources that may be saved by using an AIDS vaccine) as the benefits for public policy makers. The cumulative (national) costs of treating all people with the human immunodeficiency virus (HIV) is estimated to be $10.3 billion in 1992 (Hellinger, 1992). Based on the average infection rate and the same cost of capital, we may calculate the present value of the benefits to public policy makers:

$$E[\pi_0(s_1 > 0)] = \sum_{i=0}^{\infty} \frac{(\$10.3b) \times (1 + \frac{40{,}000}{280{,}000{,}000})^i}{1.1^i} = \$113.5b \qquad (16)$$

The estimated cost for each prototype vaccine at any one of the three clinical trials should be same for both companies and public policy makers. In our initial analysis we will adopt the industrial averages from DiMasi et al (1991). Later we will vary the values of these parameters. Table 2 shows the cost and probability of success at each clinical trial stage, and the model's pipeline recommendations for a private firm and public policy makers. We have also included the currently known actual pipelines for developing AIDS vaccines by firms.

Insert Table 2 Here

From the table, it is clear that parallel approach is desirable for developing the AIDS vaccine, from both a for-profit firm's standpoint and public policy makers' standpoint. The number of optimal parallel approaches at each stage, according to our model, are quite different for these two parties. While our model recommends that a firm should support around 5 prototype projects in Phase 1, public policy makers would like to see up to 34 ***different*** prototype projects being supported in Phase 1. Similar differences in magnitude can be seen for the other two development stages as well. The actual pipeline of the firm is narrower than what the model recommends.

The probabilities of obtaining a return (or a range of return) within any given range, for either firm or the public policy makers can also be calculated for different NPD stages as shown in Table 3.

Insert Table 3 Here

To test the sensitivity of our analyses, the value of the parameters were varied one at a time. The results are shown in Table 4.

Insert Table 4 Here

Based on Table 4, it appears that, in general, our model's normative recommendations for structuring pipelines for developing AIDS vaccines are quite robust with respect to variations in the parameter estimates.

To further analyze the current practice in the area of new drugs development, we have analyzed seven additional new drug development categories. Given that there is only one paper (DiMasi, et. al., 1995) that has estimated therapeutical-category specific cost and probability of survival, we have selected seven chronic diseases for which these parameter values are available. These include three from cardiovascular class, namely, arrhythmia, hypertension, high cholesterol; three from neuropharmacological class, namely, depression, Alzheimer's disease, migraine; and one from NSAID, COX-2 drugs treating arthritis. Moreover, we know that different firms are engaged in developing drugs for each of these categories and they are at different stages in the development cycle. Our analyses below focus on the most advanced firm in each category.

The expected gross profit for such firms is calculated using a two steps procedure. First, the gross profit is estimated for a given competitive scenario. Second, the expected gross profit is obtained by weighing the gross profit for each scenario using the probability of occurrence for that scenario. Under each scenario, defined by a specific combination of the R&D outcomes for all firms involved (e.g., one scenario might be: Firm 1 launches its new drug in year 1, Firm 2 fails in its product development efforts, and Firm 3 launches its new drug 2 years after Firm 1, …), the revenue of the new drug at each period is calculated by summing the trial and repeat prescriptions for the pioneering firm. We assume that a patient has a given probability of trying a new generation of drugs during each office visit, and the physician does not discriminate among similar (me-too) drugs in deciding which drug to prescribe to the patient. Thus trial prescriptions at each period could be easily calculated if we know the market size and the probability of trial. We also assume that there is a given probability that a patient will respond well to the trial and will thus repeatedly use the same drug and will not switch to other me-too drugs. As a result, the repeat sales could also be easily obtained. Once the prescriptions at all periods have been obtained, the expected revenues and gross profit can then be calculated based on the following equations:

$$\mathbf{E}[\mathbf{p}] = \sum_{o=1}^{O} q_o \mathbf{p}_o$$
$$\textit{where} \quad \mathbf{p}_o = \sum_{t=1}^{T} \frac{s(t) \times (1 - \alpha) \times C}{(1+\beta)^t} \qquad (17)$$

where:

$E\pi$: is the expected gross profit;
π_o: is the gross profit under scenario o;
O: is the total number of possible competitive scenarios (vary from each other depends on which ones of the competing firms' NPD are successful.)
q_o: is the probability of having a particular competitive scenario o.
T: is product life (e.g. 12 years)
C: is the contribution rate (e.g., 40%)
α: is tax rate (e.g., 36%);
β: is cost of capital (e.g., 9%)
s(t): is revenues from the drug during period t.

Following the trial-repeat purchase structure employed in pretest market models (e.g., ASSESSOR, Silk and Urban, 1978), we have developed a formulation that captures the unique context in drug prescriptions. The revenues for a given drug during period t, s(t), can be obtained as following:

$$\begin{aligned} s(t) &= Trial\_Sale(t) + \text{Re}\, peat\_Sale(t) \\ &= MSize(t) \times [1 - CTC(t-1)] \times \frac{t_r}{n(t)} + MSize(t) \times CT(t-1) \times r_r \end{aligned} \tag{18}$$

where:

Trial_Sale(t):	is the revenue during period t generated by first time users;
Repeat_Sale(t):	is the revenue during period t generated by repeat users;
Msize(t):	is the market size ($) during period t;
CTC(t-1):	is the (cumulative) proportion of the market that has tried any new drugs up to period t-1;
CT(t-1):	is the (cumulative) proportion of the market that has tried the drug of interest up to period t-1;
t_r:	is the probability of trying the new drugs for the first time in one period;
r_r:	is the probability of getting a repeat prescription for the same drug after trial;
n(t):	is the number of new drugs available during period t.

For the seven cases studied here, we have estimated the most conservative gross profit, assuming that all competing firms will eventually succeed in their NPD effort, but their introductions of the new drugs will be sequential, based on their current development stages. The 1998 market size and growth rate for each disease have been obtained from "Pharmaceutical Therapeutic Categories Outlook" by SG Cowen (March 1999). The actual pipelines of all competing firms have also been obtained from the same source. The contribution rate, tax rate, and cost of capital have been obtained from literature (Grabowski and Vernon, 1994).

The trial (t_r) and repeat (r_r) probabilities for each new drug/compound have been obtained by surveying eight experts (two clinicians, two pharmacists who are also professors in pharmacy

schools, two marketing/forecasting managers in two major pharmaceutical companies, and two pharmaceutical marketing consultants). The cover letter of the survey informed the respondents that they will be asked to estimate two parameters for seven new drug/compounds based on their experience/intuition:

Percentage #1. What percentage of targeted patients is likely to be prescribed the new drug (get an least one Rx) within **two** years of the new drug launch?

Percentage #2. What percentage of the above patients is likely to be repeat users of the drug after using the drug for the first time?

The survey employed a list of relevant information for the seven new drug/compounds, namely, Indication Targeted (e.g., arthritis), the name of New Drug/Compound and the leading firm which is developing it (e.g., Celebrex by Monsanto), and the novel mechanism used by the new drug/compound (compared to existing therapies, e.g., selective NSAID, COX-2 only). The averages (across respondents) of the percentage values are used to estimate the trial (t_r) and repeat (r_r) probabilities in the following manner. For each drug/compound, the average of estimates for percentage #2 is used directly as the probability of repeat Rx (r_r). The trial rate is recovered from the average value of percentage #1 under the premise that there will be approximately eight office visits during the two-year period (an average Rx covers 30 days with two refills for another 60 days). Thus,

$$P_1 = 1 - (1 - t_r)^8 \tag{19}$$

where P_1 is the average value of percentage #1 for a drug (probability of trial within two years of the drug launch), and t_r is the first trial probability per office visit for the drug (the probability of receiving a Rx for the new drug per office visit, if a patient has not used the drug before).

The expected market returns and the normative/actual pipelines for the seven new drug development problems are presented in Table 5.

Insert Table 5 Here

Two interesting insights emerge from this analysis. First, the leading firm in each case seems to underspend on their corresponding new drug development throughout the clinical trials compared to the model's normative recommendations. These gaps, however, must be interpreted with caution. Managers may be under internal budget constraints, whereas the model has assumed the financial market is efficient. The budget constraint, if presented as a minimum Internal Rate of Return, could be easily incorporated into the model. Managers may also face creativity limitation. The observed underspending could be due to the lack of suitable new drug candidates. Different assessment of the market opportunity may also partially explain the gap. Another possible explanation is that the probabilities of survival of the alternative approaches/candidates are not independent of each other. As shown in the next section, the normative pipeline should indeed become narrower if there is correlation among alternative approaches. We also note from the analyses that **different** NPD pipelines are needed for **different** new drug development problems. In addition to different optimal numbers of approaches at each stage, the shapes of the pipelines are also quite distinctive for different cases. For instance, for all three cases in the neuropharmacological class, the optimal first two stages should have a tunnel structure (similar or same optimal numbers) and the firm should exhibit more focus (decrease the alternative approaches funded dramatically) only in the last stage. For the remaining cases (except arrhythmia), the optimal pipelines all exhibit a funnel structure (gradually decreased optimal numbers as the development progresses). In light of this observation, it is interesting to note that pharmaceutical firms, at least the ones studied here, adopt a one-size-fit-all funding strategy (either 1-1-1 or 2-2-2) for various new drug development cases.

5. CONCLUSIONS, DISCUSSION, AND FURTHER RESEARCH

In this paper we developed a parsimonious model that recommends optimal pipeline structures for mulitple-stage product development processes. When supplied with its key inputs: magnitude of the business opportunity, cost per development approach and survival probabilities, the model can shed insights into under(over) spending in new products development. Such results can force managers to engage in systematic thinking and examination of their product development pipelines and budgeting decisions. As decision support tool, the model developed here can also be used to simulate the uncertainty associated with really new product and provide a comprehensive understanding and internal analysis. In the real world, some mergers and acquisitions decisions are motivated by reviewing pipelines of new products for their appropriateness. "Most of the mergers we have seen have been made out of weakness (in their pipelines)", as declared by Pfizer chairman William Steere when Pfizer launched its hostile-takeover bid for Warner-Lambert, in an effort to pre-empt a merger between Warner-Lambert and American Home Products. However, "Some folks on Wall Street … argue that Pfizer's own bid could be no different from other drug mergers in its aim." (McGough and Deogun, 1999). Wall-Street analysts also rely on pipeline conditions in their valuation of firms' stocks.

As demonstrated in the AIDS vaccine case, our model should also be of interest to public policy decision makers who are responsible for allocating tax money to biomedical research related to human diseases. There are always more fundable grant applications and more diseases than could be possibly supported. Furthermore, multiple approaches are often available to investigate the mechanism of a single disease. To cope with these problems, decision makers, in general, often try to divide the research budget among various diseases and support multiple (and different) labs for each disease. Unfortunately, instead of maximizing social welfare as public funding should do (which could be easily achieved by models such as ours once profit is replaced by a measure of social welfare), these decisions are sometimes influenced by other factors. The allocation of resources to different diseases is often influenced by political and social pressures (e.g., the case of breast cancer), and the allocation of resources for different projects related to the same disease is determined by scientific merit and budget constraint.

These decision rules may result in less than optimal welfare. The absolute magnitude of improvement, if systems such as our model are used, is signified by the sheer size of public funds in question. For instance, the National Cancer Institute, one of the 24 institutes and centers that is collectively known as the National Institute of Health, had a budget of $2.4 billion in fiscal year 1997 for the sole purpose of supporting research related to cancers.

This paper also contributes to the literature by filling the research gap regarding optimal resource allocation in parallel multi-stage new product development processes. As a marketing-R&D interface model, we have demonstrated how market inputs (size of the market, number and strength of competitors, and the proportion of the target market that can be addressed by each successful product) could be used to optimize resource allocation decisions during R&D and new product development.

The model proposed in this paper, while realistic for industries such as the pharmaceuticals, can, like any other model, be potentially expanded in several directions to incorporate additional considerations by relaxing its assumptions. Below, we discuss a number of possible extensions, some of these have already been undertaken.

Extension 1: Non-identical success probability and cost in each NPD stage

The basic model assumes that all available approaches have the same probability of success and cost at each stage (different across stages). This essentially captures the situation where there is no *a prior* advantages of any one development approach. In other situations, however, alternative approaches have different probability of success and cost and it is possible to rank the alternative approaches based on their probabilities of success, costs at each stage, or a composite measure of both probability and cost (e.g., probabilities of success per cost). The mathematical analysis is again based on the basic concept of comparing the marginal value of an additional project with the associated cost of supporting it. The actual analysis is straightforward but rather lengthy (each approach now has two parameters at each stage). We found that the

propositions still hold with regard to the existence of a unique maximum at each NPD stage, and the actual optimal number of approaches at each stage can also be identified. Furthermore, we found all the pipeline implications hold in spirit as well. Another promising avenue for further research is to represent the probabilities of success as a function of spending for that particular approach.

Extension 2: Modeling the Correlated Development Approaches

The basic model assumes explicitly that the alternative development approaches at each stage are probabilistically independent of each other. In other words, there is no commonality across different development approaches. This may not be true in all applications. It is possible that one or more common obstacles may exist across different development approaches, and that those need to be resolved before any approach could be successful. As a result, the outcomes of alternative development approaches will be positively correlated. One extreme example of such scenarios is the simultaneous support of multiple development teams using basically the same technology.

One way to address this aspect is to model the probability of success p_1 as p_1/p_c, where p_c is the probability of overcoming a common obstacle presented in all approaches (See Appendix A). We found that the insights from the basic model still hold in this more general environment. Furthermore, we have also examined the impact of the correlated approaches on the pipeline structure. The optimal number of approaches at Stage 1 will decrease as the (positive) correlation becomes stronger. The above result holds, at least when the correlation is high, for any stage k ($k \geq 2$). Of course, other possibilities exist to model the scenario where the development approaches are correlated.

Extension 3: Products are not identical (differentiation)

The basic model assumes products are identical and the business opportunity will be captured by a single successful product. In other situations, the final products may not be identical and the actual outcome may be more complicated. For example, more than one successful product may be launched by a firm to capture multiple niche markets. Alternatively, a firm may still launch a single product, but each successful product may have different market potential (different side effects of a drug, for example) which becomes clear only after last NPD stage and the one with the best market potential will be launched.

Both scenarios can be easily accommodated by revising the formulation of stage 0 (launch) model. Instead of assuming all products are equivalent, we could formulate probabilistically the payoff at stage 0 as an extreme value problem, with the payoff of a successful product following a specific probability distribution function (PDF). As a result, having more successful products is likely to generate more profit for a firm. This problem has the well known characteristic of concavity under most commonly used PDF. Hence, Proposition 1 remains valid.

The model could also accommodate the scenario where there are two levels of payoffs for a particular NPD project. We have essentially set the low payoff to be zero in our analysis, conceptually, we could easily accommodate the situation where the low payoff is nonzero. Under this scenario, the low payoff is guaranteed as long as one development approach is funded at all stages, while the high payoff requires the development of, for example, additional product attributes whose success is stochastic at each stage. The insights under this scenario, however, remain the same.

Extension 4: Certain NPD stages may be repeated.

The basic model assumes that delayed time to market due to repeating an NPD stage will render the business opportunity unprofitable and thus managers will drop the project altogether if no approach survives an NPD stage. This is true for most new drug development projects mainly

due to competition and, to certain extent, patent expiration if no new approaches (patents) are available. Under certain conditions, however, it might be worthwhile to repeat one or more new product development stages. This is true especially when there is no competitors or competitors are far behind, and the business opportunity is not expected to dwindle too much due to the time delay. In order to make the model applicable to these situations, we have examined the scenarios where delayed time to market will generate less profit and managers will consider repeating an NPD stage if no approach survives that stage. Instead of a linear three-stage new product development process used in our basic model, the customized process will more look like a tree structure where the process, when stalled in the original branch (failed all alternatives), is allowed to branch out to a previous stage.

The general approach in this situation is following:

- optimize the latest possible NPD repetition first;
- only downstream NPD stage repetition will affect current stage optimization;
- only the optimization of the stage where the repetition is allowed needs to be modified. It will, in general, involve simultaneous optimization of the total number needed and the optimal division among the original branch and the repeat branch.

The end of planning horizon (last profitable repeat) could be identified when the expected return allowing this repeat equals the expected return disallowing this repeat. Our analysis indicates that the optimal pipeline will become wider (larger optimal parallel approaches) prior to the branch, while the pipeline remains the same, compared to the no repetition scenario, once the development passes the branching point.

Finally, other possibilities for further research include: incorporating calendar time into the model and allowing managers to "crash" a development stage by increasing the amount of resources available for each approach; endogenize the number of stages for a particular NPD project, and incorporate learning into the process, allowing for information updating.

REFERENCES

Abernathy, W.J., R.S. Rosenbloom. 1968. Parallel and sequential R&D strategies: Application of a simple Model. IEEE Transactions on Engineering Management. EM-**15(1)** 2-10.

Abernathy W.J., R.S. Rosenbloom. 1969. Parallel strategies in development projects. Management Science. **15(10)** 486-505.

Bhattacharya, S., Krishnan, V., V. Mahajan. 1998. Managing new product definition in highly dynamic environments. Management Science. **44(11 (Part 2))** S50-S64.

Bond, R.S., D.F. Lean. 1977. Sales, Promotion, and Product Differentiation in Two Prescription Drug Markets. Washington, DC: US Federal Trade Commission.

Boulding, W., Morgan, R., R. Staelin. 1997. Pulling the plug to stop the new product drain. Journal of Marketing Research, **XXXIV** 164-176.

Cetron, M., Martino, J., L. Roepcke. 1967. The selection of R&D program content--survey of quantitative methods. IEEE Transactions on Engineering Management, EM-**14(1)** 4-13.

Chun, Y.H. 1994. Sequential decisions under uncertainty in the R&D project selection problem. IEEE transactions on Engineering Management, **41(4)** 404-413.

Dahan, E. 1998. Reducing technical uncertainty in product and process development through parallel design of prototypes. MIT Working paper.

Dean, B.V., L.E. Hauser.1967. Advanced Materiel Systems Planning. IEEE Transactions on Engineering Management, EM-**14(1)** 21-43.

DiMasi, J.A., Hansen, R.W., Grabowski, H.G., L. Lasagna. 1991. Cost of innovation in the pharmaceutical industry. Journal of Health Economics. **10** 107-142.

DiMasi, J.A., Hansen, R.W., Grabowski, H.G., L. Lasagna. 1995. Research and Development Costs for New Drugs by Therapeutic Category. PharmacoEconomics. **7(2)** 152-169.

Dreyfuss, J., Rapoport, C., Schlender, B., Calonius, E., Bylinsky, G., Shapiro, D., Alpert, M., Moore, A., Neumeier, S., S. Kirsch. 1990. Ideas for the 1990s: Today's Leaders Look to Tomorrow - Science. Fortune. **121(7)** 68-96.

Feinberg, F., J. Huber. 1996. A theory of cutoff formation under imperfect information. Management Science. **42(1)** 65-84.

Gerson, V. 1997. How business is dealing with the AIDS epidemic. Business & Health. **15(1)** 17-20.

Grabowski, H., J. Vernon. 1990. A new look at the returns and risks to pharmaceutical R&D. Management Science, **36(7)** 804-821.

Grabowski, H., J. Vernon. 1994. Returns to R&D on new drug introductions in the 1980s. Journal of Health Economics **13** 383-406.

Gross, A. 1972. The creative aspects of advertising. Sloan Management Review. **Fall** 83-109.

Hellinger, F.J. 1992. Forecasts of the Costs of Medical Care for Persons with HIV: 1992-1995. Inquiry. **29(3)** 356-365.

Henderson, C. 1996. Conference Coverage (NCVDG) Researchers Bullish as NIAID Reveals AIDS Vaccine Strategy. Blood Weekly. February 26.

Jackson, B. 1983. Decision methods for selecting a portfolio of R&D projects. Research Management, Sep/Oct 21-26.

Kotler, P. 1994. Marketing Management. Pretence Hall, Upper Saddle River, New Jersey 07458.

MaGough, R., N. Deogun. 1999. Mergers pose debatable cure for diseases of drug firms. Wall Street Journal. November 12.

Marschak, T., Glennan, T.K.J., R. Summer. 1967. Strategy for R&D: Studies in the Microeconomics of Development. Springer-Verlag New York Inc.

Morone, J. 1993. Winning in high-tech markets, the role of general management. Harvard Business School Press, Boston, Massachusetts.

Nelson, R.R. 1961. Uncertainty, Learning, and the Economics of Parallel Research and Development Efforts. The Review of Economics and Statistics, **43(4)** 351-364.

Parry, M., F.M. Bass. 1990. When to lead or follow? It depends. Marketing Letters. **1(November)** 187-198.

Quinn, J.B. 1985. Managing Innovation: Controlled Chaos. Harvard Business Review. May/June 73-84.

Quinn, J.B. 1996. Team Innovation. Executive Excellence. **13(7)** 13-14.

Rosenbloom, R., M.A. Cusumano. 1987. Technological Pioneering and Competitive Advantage: The birth of the VCR Industry. California Management Review. **XXIX(4)** 51-76.

Schmidt, R.L., J.R. Freeland. 1992. Recent Progress in Modeling Project-selection Processes. IEEE Transactions on Engineering Management. **39(2)** 189-200.

SG Cowen Securities Corporations. 1999. Pharmaceutical Therapeutic Categories Outlook. March.

Silk, A. J., G. L. Urban. 1978. Pre-test market evaluation of new packaged goods: a model and measurement methodology. Journal of Marketing Research, **15 (May)** 171-191.

Souder, W.E. 1978. A system for using R&D Project evaluation Methods. Research Management. September 29-37.

Souder, W. E., T. Mandakovic. 1986. Research Management. **29(4)** 36-42.

Srinivasan, V., Lovejoy, W., D. Beach. 1997. Integrated product design for marketability and manufacturing. Journal of Marketing Research. **34(1)** 154-163.

Steele, L.W. 1988. Selecting R&D programs and objectives. Research-Technology Management. **31(2)** 17-36.

Urban, G., L., Carter, TY., Gaskin, S., Z. Mucha. 1986. Market share rewards to pioneering brands, an empirical analysis and strategic implications. Management Science, **32(6)** 645-659.

Urban, G., J.R. Hauser. 1993. Design and Marketing of New Products. 2nd ed. Prentice Hall, Englewood Cliffs, New Jersey 07632.

The Wall Street Journal, April 21, 1999.

Weber, R. Werners, B., H.-J. Zimmermann. 1990. Planning Models for research and Development. European Journal of operational research, **48** 175-188.

APPENDIX A: Proof of Lemma 1, Propositions 2-3, and Impact of Correlation

Lemma 1: $E[\pi_k(n_k)]$ is a strictly concave function with a unique global maximum at n_k^*, where n_k^* is implicitly defined by the following equation:

$$p_k \sum_{s_k=0}^{n_{k-1}^*-1} \{\Pr(s_k \mid p_k, n_k^*)[E[\pi_{k-1}(s_k+1)] - E[\pi_{k-1}(s_k)]]\} - c_k = 0 \qquad \text{(A1) \& (12)}$$

k is a positive integer and $k \geq 2$.

Proof:

This proof uses induction and is comprised of two steps:

Step 1: show the proposition holds when k=2;
Step 2: show the proposition holds for k=j+1, if we assume the proposition holds for k=j (j is a positive integer and $j \geq 2$);

In order to facilitate the proof, we need to first reformulate the expression of expected return at stage k. For convenience, we will use following abbreviation:

$$\Pr(s_k \mid n_k) = \Pr(s_k \mid p_k, n_k)$$

The outcome of m+1 parallel approaches at stage k could be viewed as a combination of outcomes from two groups. One group contains one approach, the other contains m approaches. Thus, the expected return at stage k for m+1 approaches could be reformulated as following:

$$\begin{aligned} E[\pi_k(m+1)] &= \sum_{s_k=n_{k-1}^*}^{m+1} \left[\Pr(s_k \mid m+1) E[\pi_{k-1}(n_{k-1}^*)]\right] + \sum_{s_k=0}^{n_{k-1}^*-1} \left[\Pr(s_k \mid m+1) E[\pi_{k-1}(s_k)]\right] - (m+1)c_k \\ &= p_k \left\{ \sum_{s_k=n_{k-1}^*-1}^{m} \left[\Pr(s_k \mid m) E[\pi_{k-1}(n_{k-1}^*)]\right] + \sum_{s_k=0}^{n_{k-1}^*-2} \left[\Pr(s_k \mid m) E[\pi_{k-1}(s_k+1)]\right] \right\} \\ &\quad + (1-p_k) \left\{ \sum_{s_k=n_{k-1}^*}^{m} \left[\Pr(s_k \mid m) E[\pi_{k-1}(n_{k-1}^*)]\right] + \sum_{s_k=0}^{n_{k-1}^*-1} \left[\Pr(s_k \mid m) E[\pi_{k-1}(s_k)]\right] \right\} - (m+1)c_k \end{aligned} \qquad \text{(A2)}$$

As a result, we could now simplify the incremental expected return (the discrete equivalent of 1st order derivative against m) offered by the $(m+1)^{th}$ approach:

$$
\begin{aligned}
E[\pi_k(m+1)] - E[\pi_k(m)] = {} & p_k \left\{ \sum_{s_k = n^*_{k-1}-1}^{m} \left[\Pr(s_k|m) E[\pi_{k-1}(n^*_{k-1})] \right] + \sum_{s_k=0}^{n^*_{k-1}-2} \left[\Pr(s_k|m) E[\pi_{k-1}(s_k+1)] \right] \right\} \\
& + (1-p_k) \left\{ \sum_{s_k = n^*_{k-1}}^{m} \left[\Pr(s_k|m) E[\pi_{k-1}(n^*_{k-1})] \right] + \sum_{s_k=0}^{n^*_{k-1}-1} \left[\Pr(s_k|m) E[\pi_{k-1}(s_k)] \right] \right\} - (m+1)c_k \\
& - \sum_{s_k = n^*_{k-1}}^{m} \left[\Pr(s_k|m) E[\pi_{k-1}(n^*_{k-1})] \right] - \sum_{s_k=0}^{n^*_{k-1}-1} \left[\Pr(s_k|m) E[\pi_{k-1}(s_k)] \right] + mc_k \\
= {} & p_k \left\{ \sum_{s_k = n^*_{k-1}-1}^{m} \left[\Pr(s_k|m) E[\pi_{k-1}(n^*_{k-1})] \right] - \sum_{s_k = n^*_{k-1}}^{m} \left[\Pr(s_k|m) E[\pi_{k-1}(n^*_{k-1})] \right] \right\} \\
& + p_k \left\{ \sum_{s_k=0}^{n^*_{k-1}-2} \left[\Pr(s_k|m) E[\pi_{k-1}(s_k+1)] \right] - \sum_{s_k=0}^{n^*_{k-1}-1} \left[\Pr(s_k|m) E[\pi_{k-1}(s_k)] \right] \right\} - c_k \\
= {} & p_k \Pr(n^*_{k-1}-1|m) E[\pi_{k-1}(n^*_{k-1})] - p_k \Pr(n^*_{k-1}-1|m) E[\pi_{k-1}(n^*_{k-1}-1)] \\
& + p_k \sum_{s_k=0}^{n^*_{k-1}-2} \left[\Pr(s_k|m) \left(E[\pi_{k-1}(s_k+1)] - E[\pi_{k-1}(s_k)] \right) \right] - c_k \\
= {} & p_k \sum_{s_k=0}^{n^*_{k-1}-1} \left[\Pr(s_k|m) \left(E[\pi_{k-1}(s_k+1)] - E[\pi_{k-1}(s_k)] \right) \right] - c_k
\end{aligned}
\tag{A3}
$$

The discrete equivalent of the 2nd order derivative against m could be obtained by calculating the difference between adjacent incremental expected return (1st order derivative):

$$
\left[E[\pi_k(m+2)] - E[\pi_k(m+1)] \right] - \left[E[\pi_k(m+1)] - E[\pi_k(m)] \right] \tag{A4}
$$

Using (A3) and similar partition technique, we could obtain:

$$
\begin{aligned}
\left[E[\pi_k(m+2)] - E[\pi_k(m+1)] \right] = {} & p_k \sum_{s_k=0}^{n^*_{k-1}-1} \left[\Pr(s_k|m+1) \left(E[\pi_{k-1}(s_k+1)] - E[\pi_{k-1}(s_k)] \right) \right] - c_k \\
= {} & p_k \left\{ p_k \sum_{s_k=0}^{n^*_{k-1}-2} \left[\Pr(s_k|m) \left(E[\pi_{k-1}(s_k+2)] - E[\pi_{k-1}(s_k+1)] \right) \right] + (1-p_k) \sum_{s_k=0}^{n^*_{k-1}-1} \left[\Pr(s_k|m) \left(E[\pi_{k-1}(s_k+1)] - E[\pi_{k-1}(s_k)] \right) \right] \right\} - c_k \\
= {} & p_k^2 \left\{ \sum_{s_k=0}^{n^*_{k-1}-2} \left[\Pr(s_k|m) \left(E[\pi_{k-1}(s_k+2)] - E[\pi_{k-1}(s_k+1)] \right) \right] - \sum_{s_k=0}^{n^*_{k-1}-1} \left[\Pr(s_k|m) \left(E[\pi_{k-1}(s_k+1)] - E[\pi_{k-1}(s_k)] \right) \right] \right\} \\
& + p_k \sum_{s_k=0}^{n^*_{k-1}-1} \left[\Pr(s_k|m) \left(E[\pi_{k-1}(s_k+1)] - E[\pi_{k-1}(s_k)] \right) \right] - c_k \\
= {} & p_k^2 \left\{ \sum_{s_k=0}^{n^*_{k-1}-2} \left[\Pr(s_k|m) \left(E[\pi_{k-1}(s_k+2)] - E[\pi_{k-1}(s_k+1)] \right) \right] - \sum_{s_k=0}^{n^*_{k-1}-1} \left[\Pr(s_k|m) \left(E[\pi_{k-1}(s_k+1)] - E[\pi_{k-1}(s_k)] \right) \right] \right\} \\
& + \left[E[\pi_k(m+1)] - E[\pi_k(m)] \right]
\end{aligned}
\tag{A5}
$$

Thus, we could obtain the equivalent 2nd derivative by rearranging (A5):

$$
\begin{aligned}
&\left[E[\pi_k(m+2)] - E[\pi_k(m+1)]\right] - \left[E[\pi_k(m+1)] - E[\pi_k(m)]\right] \\
&= p_k^{\,2}\left\{\sum_{s_k=0}^{n_{k-1}^*-2}\left[\Pr(s_k|m)\left(E[\pi_{k-1}(s_k+2)] - E[\pi_{k-1}(s_k+1)]\right)\right] - \sum_{s_k=0}^{n_{k-1}^*-1}\left[\Pr(s_k|m)\left(E[\pi_{k-1}(s_k+1)] - E[\pi_{k-1}(s_k)]\right)\right]\right\} \\
&= p_k^{\,2}\left\{\sum_{s_k=0}^{n_{k-1}^*-2}\left[\Pr(s_k|m)\left[\left(E[\pi_{k-1}(s_k+2)] - E[\pi_{k-1}(s_k+1)]\right) - \left(E[\pi_{k-1}(s_k+1)] - E[\pi_{k-1}(s_k)]\right)\right]\right] - \left[\Pr(n_{k-1}^*-1|m)\left(E[\pi_{k-1}(n_{k-1}^*)] - E[\pi_{k-1}(n_{k-1}^*-1)]\right)\right]\right\}
\end{aligned}
\tag{A6}
$$

Now we could prove (A1) holds for all k (k is a positive integer and $k \geq 2$) using induction.

Step 1: k=2.

(A6) could be simplified to:

$$
\begin{aligned}
&\left[E[\pi_2(m+2)] - E[\pi_2(m+1)]\right] - \left[E[\pi_2(m+1)] - E[\pi_2(m)]\right] \\
&= p_2^{\,2}\left\{\sum_{s_2=0}^{n_1^*-2}\left[\Pr(s_2|m)\left[\left(E[\pi_1(s_2+2)] - E[\pi_1(s_2+1)]\right) - \left(E[\pi_1(s_2+1)] - E[\pi_1(s_2)]\right)\right]\right\} \\
&- p_2^{\,2}\left[\Pr(n_1^*-1|m)\left(E[\pi_1(n_1^*)] - E[\pi_1(n_1^*-1)]\right)\right]
\end{aligned}
\tag{A7}
$$

Based results from stage 1, we could obtain:

$$
\begin{aligned}
E[\pi_1(s_2+1)] - E[\pi_1(s_2)] &= [1-(1-p_1)^{s_2+1}]E[\pi_0(s_1>0)] - (s_2+1)c_1 - [1-(1-p_1)^{s_2}]E[\pi_0(s_1>0)] + s_2 c_1 \\
&= p_1(1-p_1)^{s_2}E[\pi_0(s_1>0)] - c_1
\end{aligned}
\tag{A8}
$$

thus, assuming p_1 does not equal 0 nor 1,

$$
\begin{aligned}
&\left[E[\pi_1(s_2+2)] - E[\pi_1(s_2+1)]\right] - \left[E[\pi_1(s_2+1)] - E[\pi_1(s_2)]\right] \\
&= \left[p_1(1-p_1)^{s_2+1}E[\pi_0(s_1>0)] - c_1\right] - \left[p_1(1-p_1)^{s_2}E[\pi_0(s_1>0)] - c_1\right] \\
&= -p_1^{\,2}(1-p_1)^{s_2}E[\pi_0(s_1>0)] < 0
\end{aligned}
\tag{A9}
$$

Given (A9), the first term (summation) of (A7) must be smaller than 0. Given the definition of n_1^*, the second term $-p_2^{\,2}\left[\Pr(n_1^*-1|m)\left(E\pi_1(n_1^*) - E\pi_1(n_1^*-1)\right)\right]$ also must be equal or smaller than 0.

As a result, (A7) must be smaller than 0, i.e.,

$$
\left[E[\pi_2(m+2)] - E[\pi_2(m+1)]\right] - \left[E[\pi_2(m+1)] - E[\pi_2(m)]\right] < 0 \tag{A10}
$$

Thus, we have proved that the expected return at stage 2 is strictly concave. The value n_2^*, obtained by setting the 1st order derivative (A3) equals to 0, is the unique value that maximizes the expected return at stage 2.

Step 2: Assuming the proposition holds for k=j (j is a positive integer and $j \geq 2$), thus

$$
\left[E[\pi_j(m+2)] - E[\pi_j(m+1)]\right] - \left[E[\pi_j(m+1)] - E[\pi_j(m)]\right] < 0 \tag{A11}
$$

for stage j+1, the 2nd derivative could be represented by

$$\begin{aligned}
&\left[E[\pi_{j+1}(m+2)] - E[\pi_{j+1}(m+1)]\right] - \left[E[\pi_{j+1}(m+1)] - E[\pi_{j+1}(m)]\right] \\
&= p_{j+1}^2 \left\{ \sum_{s_{j+1}=0}^{n_j^*-2} \left[\Pr(s_{j=1}|m)\left[\left(E[\pi_j(s_{j+1}+2)] - E[\pi_j(s_{j+1}+1)]\right) - \left(E[\pi_j(s_{j+1}+1)] - E[\pi_j(s_{j+1})]\right)\right]\right\} \\
&- p_{j+1}^2 \left[\Pr(n_j^* - 1|m)\left(E[\pi_j(n_j^*)] - E[\pi_j(n_j^* - 1)]\right)\right]
\end{aligned} \qquad \text{(A12)}$$

Similar to step 1, (A12) is smaller than 0 because first term in (A12) is smaller than 0 because of (A11), and second term is smaller than or equal to 0 by definition of n_j*. Thus, the expected return function at stage j+1 is also strictly concave and thus proposition (A1) holds.

Combine the results from step 1 and step 2, we conclude the proposition holds for all k, where k is a positive integer and k ³ 2.

Q.E.D.

Proposition 2:

For stage k (k ³ 2):

n_k* increases when c_k decreases;

n_k* reaches maximum at an interior value of p_k(between 0 and 1);

An approximation (upper bound) for n_k* is given by

$$n_k^* < \frac{\ln c_k - \ln(p_1 p_2 \cdots p_k E[\pi_0(s_1 > 0)])}{\ln(1 - p_1 p_2 \cdots p_k)} \qquad \text{(A13) \& (13)}$$

Proof:

When the optimal number of approaches (n_k*) is funded, the marginal benefit of the n_k*th approach is equal to the marginal cost (c_k). If c_k decreases, the marginal benefit of the n_k*th approach will then be larger than its marginal cost. Given the concavity (from Lemma 1), then the new optimal number must be larger than the original optimal number (i.e., n_k* will increase).

The observation that n_k* reaches maximum at an interior value of p_k(between 0 and 1) could be made based on the fact that there is no marginal benefit when p_k equals either 0 or 1.

In order to prove the upper bound, we will first use induction to prove that

$$E\pi_k[(n_k+1)] - E[\pi_k(n_k)] < p_1p_2\cdots p_kE[\pi_0](1-p_1p_2\cdots p_k)^{n_k} - c_k \tag{A14}$$

For k=2:

$$\begin{aligned}
E[\pi_2(n_2+1)] - E[\pi_2(n_k)] &= p_2\sum_{s_2=0}^{n_1^*-1}\{\Pr(s_2\mid p_2,n_2)[E[\pi_1(s_2+1)] - E[\pi_1(s_2)]]\} - c_2\\
&= p_2\sum_{s_2=0}^{n_1^*-1}\{\Pr(s_2\mid p_2,n_2)[p_1(1-p_1)^{s_2}E[\pi_0] - c_1]\} - c_2\\
&= p_1p_2E[\pi_0]\sum_{s_2=0}^{n_1^*-1}\binom{n_2}{s_2}p_2^{s_2}(1-p_2)^{n_2-s_2}(1-p_1)^{s_2} - p_2c_1\sum_{s_2=0}^{n_1^*-1}\{\Pr(s_2\mid p_2,n_2)\} - c_2\\
&= p_1p_2E[\pi_0]\sum_{s_2=0}^{n_1^*-1}\binom{n_2}{s_2}\left(\frac{p_2(1-p_1)}{1-p_2}\right)^{s_2}(1-p_2)^{n_2} - p_2c_1\sum_{s_2=0}^{n_1^*-1}\{\Pr(s_2\mid p_2,n_2)\} - c_2\\
&= p_1p_2E[\pi_0]\frac{(1-p_2)^{n_2}}{\left(\frac{1-p_2}{1-p_1p_2}\right)^{n_2}}\sum_{s_2=0}^{n_1^*-1}\binom{n_2}{s_2}\left(\frac{p_2(1-p_1)}{1-p_2}\right)^{s_2}\left(\frac{1-p_2}{1-p_1p_2}\right)^{n_2} - p_2c_1\sum_{s_2=0}^{n_1^*-1}\{\Pr(s_2\mid p_2,n_2)\} - c_2\\
&= p_1p_2E[\pi_0](1-p_1p_2)^{n_2}\sum_{s_2=0}^{n_1^*-1}\Pr\left(s_2\mid\frac{p_2(1-p_1)}{1-p_1p_2},n_2\right) - p_2c_1\sum_{s_2=0}^{n_1^*-1}\{\Pr(s_2\mid p_2,n_2)\} - c_2\\
&< p_1p_2E[\pi_0](1-p_1p_2)^{n_2}\sum_{s_2=0}^{n_1^*-1}\Pr\left(s_2\mid\frac{p_2(1-p_1)}{1-p_1p_2},n_2\right) - c_2\\
&< p_1p_2E[\pi_0](1-p_1p_2)^{n_2} - c_2
\end{aligned} \tag{A15}$$

Assume the inequality for k=m-1 holds, i.e.,

$$E[\pi_{m-1}(n_{m-1}+1)] - E[\pi_{m-1}(n_{m-1})] < p_1p_2\cdots p_{m-1}E[\pi_0](1-p_1p_2\cdots p_{m-1})^{n_{m-1}} - c_{m-1} \tag{A16}$$

then for k=m,

$$E[\pi_m(n_m+1)] - E[\pi_m(n_m)] = p_m \sum_{s_m=0}^{n_{m-1}^*-1} \{\Pr(s_m \mid p_m, n_m)[E[\pi_{m-1}(s_m+1)] - E[\pi_{m-1}(s_m)]]\} - c_m$$

$$< p_m \sum_{s_m=0}^{n_{m-1}^*-1} \{\Pr(s_m \mid p_m, n_m)[p_1 p_2 \cdots p_{m-1} E[\pi_0](1 - p_1 p_2 \cdots p_{m-1})^{s_m} - c_{m-1}]\} - c_m$$

$$= p_1 p_2 \cdots p_m E[\pi_0] \sum_{s_m=0}^{n_{m-1}^*-1} \left[\binom{n_m}{s_m} p_m^{s_m} (1-p_m)^{n_m - s_m} (1 - p_1 p_2 \cdots p_{m-1})^{s_m}\right] - p_m c_{m-1} \sum_{s_m=0}^{n_{m-1}^*-1} \{\Pr(s_m \mid p_m, n_m)\} - c_m$$

$$= p_1 p_2 \cdots p_m E[\pi_0] \sum_{s_m=0}^{n_{m-1}^*-1} \left[\binom{n_m}{s_m} \left(\frac{p_m(1 - p_1 p_2 \cdots p_{m-1})}{1 - p_m}\right)^{s_m} (1-p_m)^{n_m}\right] - p_m c_{m-1} \sum_{s_m=0}^{n_{m-1}^*-1} \{\Pr(s_m \mid p_m, n_m)\} - c_m$$

$$= p_1 p_2 \cdots p_m E[\pi_0] \frac{(1-p_m)^{n_m}}{\left(\frac{1 - p_m}{1 - p_1 p_2 \cdots p_m}\right)^{n_m}} \sum_{s_m=0}^{n_{m-1}^*-1} \left[\binom{n_m}{s_m} \left(\frac{p_m(1 - p_1 p_2 \cdots p_{m-1})}{1 - p_m}\right)^{s_m} \left(\frac{1 - p_m}{1 - p_1 p_2 \cdots p_m}\right)^{n_m}\right]$$

$$- p_m c_{m-1} \sum_{s_m=0}^{n_{m-1}^*-1} \{\Pr(s_m \mid p_m, n_m)\} - c_m$$

$$= p_1 p_2 \cdots p_m E[\pi_0] \frac{(1-p_m)^{n_m}}{\left(\frac{1 - p_m}{1 - p_1 p_2 \cdots p_m}\right)^{n_m}} \sum_{s_m=0}^{n_{m-1}^*-1} \left[\Pr\left(s_2 \mid \frac{p_m(1 - p_1 p_2 \cdots p_{m-1})}{1 - p_1 p_2 \cdots p_m}, n_m\right)\right] - p_m c_{m-1} \sum_{s_m=0}^{n_{m-1}^*-1} \{\Pr(s_m \mid p_m, n_m)\} - c_m$$

$$< p_1 p_2 \cdots p_m E[\pi_0] (1 - p_1 p_2 \cdots p_m)^{n_m} \sum_{s_m=0}^{n_{m-1}^*-1} \left[\Pr\left(s_2 \mid \frac{p_m(1 - p_1 p_2 \cdots p_{m-1})}{1 - p_1 p_2 \cdots p_m}, n_m\right)\right] - c_m$$

$$< p_1 p_2 \cdots p_m E[\pi_0] (1 - p_1 p_2 \cdots p_m)^{n_m} - c_m$$

(A17)

Thus, the inequality (A14) holds for all k, where $k \geq 2$.

As a result, when $n_k = n_k^*$,

$$E[\pi_k(n_k^*+1)] - E[\pi_k(n_k^*)] = 0 < p_1 p_2 \cdots p_k E[\pi_0](1 - p_1 p_2 \cdots p_k)^{n_k^*} - c_k$$

$$\Rightarrow (1 - p_1 p_2 \cdots p_k)^{n_k^*} > \frac{c_k}{p_1 p_2 \cdots p_k E[\pi_0]}$$

$$\Rightarrow n_k^* \ln(1 - p_1 p_2 \cdots p_k) > \ln \frac{c_k}{p_1 p_2 \cdots p_k E[\pi_0]}$$

$$\Rightarrow n_k^* < \frac{\ln c_k - \ln(p_1 p_2 \cdots p_k E[\pi_0])}{\ln(1 - p_1 p_2 \cdots p_k)}$$

Q.E.D.

Proposition 3:

If, for all k (k ³ 2),

$$c_k < p_k \sum_{s_k=0}^{n_{k-1}^*-1} \{\Pr(s_k \mid p_k, n_{k-1}^*)[E[\pi_{k-1}(p_{k-1}, s_k+1, c_{k-1}, E\pi_{k-2})] - E[\pi_{k-1}(p_{k-1}, s_k, c_{k-1}, E\pi_{k-2})]]\} = \bar{c}_k$$

(A18) & (14)

then the NPD pipeline will take the shape of a **funnel** ($n_k^* > n_{k-1}^*$). Otherwise the pipeline will be a **tunnel** shape ($n_k^* = n_{k-1}^*$).

Proof:

From Lemma 1, we know that

$$p_k \sum_{s_k=0}^{n_{k-1}^*-1} \{\Pr(s_k \mid p_k, n_{k-1}^*)[E[\pi_{k-1}(p_{k-1}, s_k+1, c_{k-1}, E\pi_{k-2})] - E[\pi_{k-1}(p_{k-1}, s_k, c_{k-1}, E\pi_{k-2})]]\} = \bar{c}_k$$

represents the marginal benefit of supporting the n_{k-1}^*th approach in stage k, in other words, supporting the same number of approaches in both stage k-1 and k.

If (A18) holds, then the marginal benefit of supporting this approach is larger than the marginal cost (c_k), and thus this approach should be supported. Given the concavity (from Lemma 1), n_k^* must be larger than n_{k-1}^* and the pipeline will therefore have a funnel structure.

If (A18) does not hold, then the marginal benefit of supporting this approach is smaller than the marginal cost (c_k), and this approach should either not be supported or be the last one to be supported (if equality holds). Given the concavity (from Lemma 1), n_k^* must be the same or smaller than n_{k-1}^*. However, given the constraint that approaches have to pass earlier stages to move to later stages, the resulting pipeline will have a tunnel structure with its width equals to n_k^*.

Q.E.D.

Impact of Correlated Approaches:

The optimal number of approaches at Stage 1 will increase as the (positive) correlation becomes weaker.

Proof:

Let's assume that the success of an individual approach depends on two factors, one is a common obstacle presented in all approaches (with probability of success p_c), the other is the factor unique to each approach (with probability of success p_1/p_c). This reduces to the base model if $p_c=1$.

The pair-wise correlation could be calculated as:

$$corr = \frac{\mathrm{cov}(Y_1,Y_2)}{\sigma_1\sigma_2} = \frac{E(Y_1Y_2) - E(Y_1)E(Y_2)}{p_1(1 - p_1)} = \frac{p_c(p_1/p_c)^2 - p_1^2}{p_1(1 - p_1)} = \frac{p_1}{1 - p_1}\left(\frac{1}{p_c} - 1\right) \quad \text{(A19)}$$

Thus, smaller p_c (most of the uncertainty is due to the common obstacle) will result in higher correlation.

Under this scenario, the expected return at stage 1 is:

$$E[\pi_1(n_1)] = p_c[1 - (1 - \frac{p_1}{p_c})^{n_1}]E[\pi_0(s_1 > 0)] - n_1c_1 \quad \text{(A20)}$$

So, the marginal change in return is

$$E[\pi_1(n_1 + 1)] - E[\pi_1(n_1)] = p_c[\frac{p_1}{p_c}(1 - \frac{p_1}{p_c})^{n_1}]E[\pi_0(s_1 > 0)] - c_1 = p_1(1 - \frac{p_1}{p_c})^{n_1}E[\pi_0(s_1 > 0)] - c_1 \quad \text{(A21)}$$

the above quantity becomes 0 when $n_1 = n_1^*$. In which case, we could solve for n_1^*:

$$n_1^* = \frac{\ln c_1 - \ln\{p_1E[\pi_0(s_1 > 0)]\}}{\ln(1 - \frac{p_1}{p_c})} \quad \text{(A22)}$$

we could obtain the 1st derivative:

$$\frac{\partial n_1^*}{\partial p_c} = \frac{\ln c_1 - \ln(p_1E[\pi_0(s_1 > 0)])}{\left[\ln(1 - \frac{p_1}{p_c})\right]^2}\left(-\frac{\frac{p_1}{p_c^2}}{1 - \frac{p_1}{p_c}}\right) > 0 \quad \text{(A23)}$$

Thus, the optimal number increases as the probability of success for the common factor increases. Equivalently, the optimal number increases as the correlation decreases.

Q.E.D.

APPENDIX B: GEOMETRIC ANALYSIS OF ONE STAGE AND TWO STAGE SCENARIOS

The purpose of this analysis is to shed light to the optimal pipeline structure and be used as managerial guidelines when the specific parameter values are not available, or could not be accurately (or cost-effectively) obtained.

One-stage scenario:

First, we examine the condition under which at least one approach should be funded:

$$p_1 * E[\pi_0(s_1 > 0)] - c_1 > 0 \Leftrightarrow c_1 < p_1 E[\pi_0(s_1 > 0)] \quad \text{(B1)}$$

Given $E[\pi_1(n_1)]$ is concave over n_1 (Proposition 1), the condition under which two or more approaches should be funded simultaneously is following:

$$\begin{aligned} & E[\pi_1(2)] > E[\pi_1(1)] \\ & \Leftrightarrow \left[1 - (1 - p_1)^2\right] * E[\pi_0(s_1 > 0)] - 2c_1 > p_1 E[\pi_0(s_1 > 0)] - c_1 \\ & \Leftrightarrow c_1 < p_1(1 - p_1) E[\pi_0(s_1 > 0)] \end{aligned} \quad \text{(B2)}$$

In general, the condition under which funding n_1 approaches is better than funding n_1-1 approaches could be expressed as:

$$\begin{aligned} & E[\pi_1(n_1)] > E[\pi_1(n_1 - 1)] \\ & \Leftrightarrow \left[1 - (1 - p_1)^{n_1}\right] E[\pi_0(s_1 > 0)] - n_1 c_1 > \left[1 - (1 - p_1)^{n_1 - 1}\right] E[\pi_0(s_1 > 0)] - (n_1 - 1)c_1 \\ & \Leftrightarrow c_1 < p_1(1 - p_1)^{n_1 - 1} E[\pi_0(s_1 > 0)] \end{aligned} \quad \text{(B3)}$$

The relationship between the critical c_1 and the probability of success (p_1) is following:

$$\frac{\partial c_1}{\partial p_1} = (1 - p_1)^{n_1 - 2} E[\pi_0(s_1 > 0)](1 - p_1 n_1) \begin{cases} < 0 & p_1 > 1/n_1 \\ > 0 & p_1 < 1/n_1 \end{cases} \quad \text{(B4)}$$

$$\frac{\partial^2 c_1}{\partial p_1^2} = (1 - p_1)^{n_1 - 3} E[\pi_0(s_1 > 0)](n_1 - 1)(p_1 n_1 - 2) \begin{cases} < 0 & p_1 < 2/n_1 \\ > 0 & p_1 > 2/n_1 \end{cases} \quad \text{(B5)}$$

Based on the above analysis, the optimal strategy under various conditions could be represented in Figure 4.

Two-stage scenario:

The conditions for the last NPD stage could be obtained from the earlier analysis for one-stage process. Assuming $p = p_1 * p_2 \quad and \quad c = c_1 + c_2$, the conditions for the first NPD stage could all be expressed as inequalities between c_1 and a function of the p_1, p, c, and $E[\pi_0(s_1 > 0)]$:

Given $n_1^* = 1$,

$$E[\pi_2(1)] > E[\pi_2(0)]$$
$$\Leftrightarrow\ p_2(p_1E[\pi_0(s_1 > 0)] - c_1) - c_2 > 0 \ \Leftrightarrow\ c_1 < \frac{p_1}{p_1 - p}(c - pE[\pi_0(s_1 > 0)]) \tag{B6}$$

$$E[\pi_2(2)] > E[\pi_2(1)]$$
$$\Leftrightarrow\ \left[p_2^2 + 2p_2(1 - p_2)\right]\{p_1E[\pi_0(s_1 > 0)] - c_1\} - 2c_2 > p_2(p_1E[\pi_0(s_1 > 0)] - c_1) - c_2$$
$$\Leftrightarrow\ p_2(1 - p_2)(p_1E[\pi_0(s_1 > 0)] - c_1) - c_2 > 0 \tag{B7}$$
$$\Leftrightarrow\ c_1 > \frac{(c - pE[\pi_0(s_1 > 0)])\,p_1^2 + p^2E[\pi_0(s_1 > 0)]p_1}{p_1^2 - pp_1 + p^2}$$

Given $n_1^* > 1$,

$$E[\pi_2(1)] > E[\pi_2(0)]$$
$$\Leftrightarrow\ p_2(p_1E[\pi_0(s_1 > 0)] - c_1) - c_2 > 0 \ \Leftrightarrow\ c_1 < \frac{p_1}{p_1 - p}(c - pE[\pi_0(s_1 > 0)]) \tag{B8}$$

$$E[\pi_2(2)] > E[\pi_2(1)]$$
$$\Leftrightarrow\ p_2^2\left\{\left[1 - (1 - p_1)^2\right]E[\pi_0(s_1 > 0)] - 2c_1\right\} + 2p_2(1 - p_2)(p_1E[\pi_0(s_1 > 0)] - c_1) - 2c_2 > p_2(p_1E[\pi_0(s_1 > 0)] - c_1) - c_2$$
$$\Leftrightarrow\ c_1 > \frac{p_1}{p_1 - p}(c - pE[\pi_0(s_1 > 0)] + p^2E[\pi_0(s_1 > 0)]) \tag{B9}$$

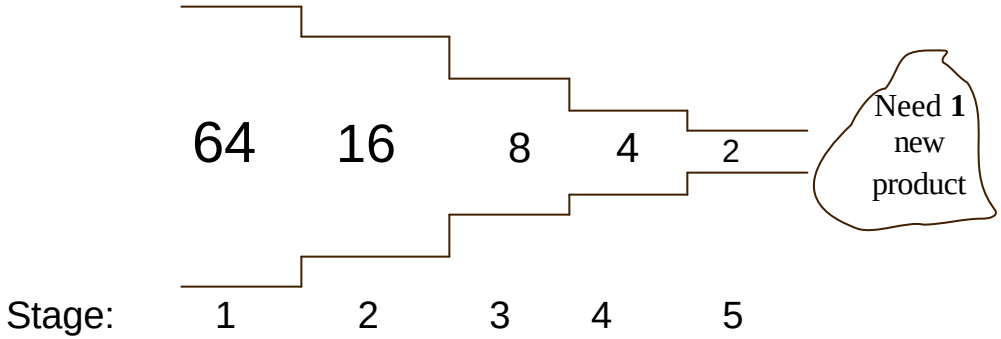
By analyzing the first and second derivatives of the above boundary conditions (similar to the one-stage problem), we found that the parameter space (p, c, $E[\pi_0(s_1 > 0)]$) could be divided into three regions:

$$\begin{array}{ll} \text{Region \#1(low overall cost)} & c < (p - p^2)E[\pi_0(s_1 > 0)] \\ \text{Region\#2(moderate overall cost)} & (p - p^2)E[\pi_0(s_1 > 0)] < c < pE[\pi_0(s_1 > 0)] \\ \text{Region\#3(high overall cost)} & c > pE[\pi_0(s_1 > 0)] \end{array} \quad \text{(B10)}$$

The behaviors of the boundary conditions differ dramatically across these regions (have different signs for first and second derivatives). The value of p_1 (and its interaction with p, c) only affects the boundaries (c_1) quantitatively, i.e., changes the relative positions of the boundaries but not the shape. The results are represented in Figure 5.

Figure 1: An Example of Structuring the NPD Pipeline

Stage	Pass Ratio	Cost per approach
1. Idea screening	1:4	$1,000
2. Concept test	1:2	$20,000
3. Product development	1:2	$200,000
4. Test Marketing	1:2	$500,000
5. National launch	1:2	$5,000,000



Total Pipeline's Budget Required: $13.984 m

Source: Kotler (1994), p.319, Table 13-1

Figure 2: Various Forms of Pipelines Structures Observed in Practice

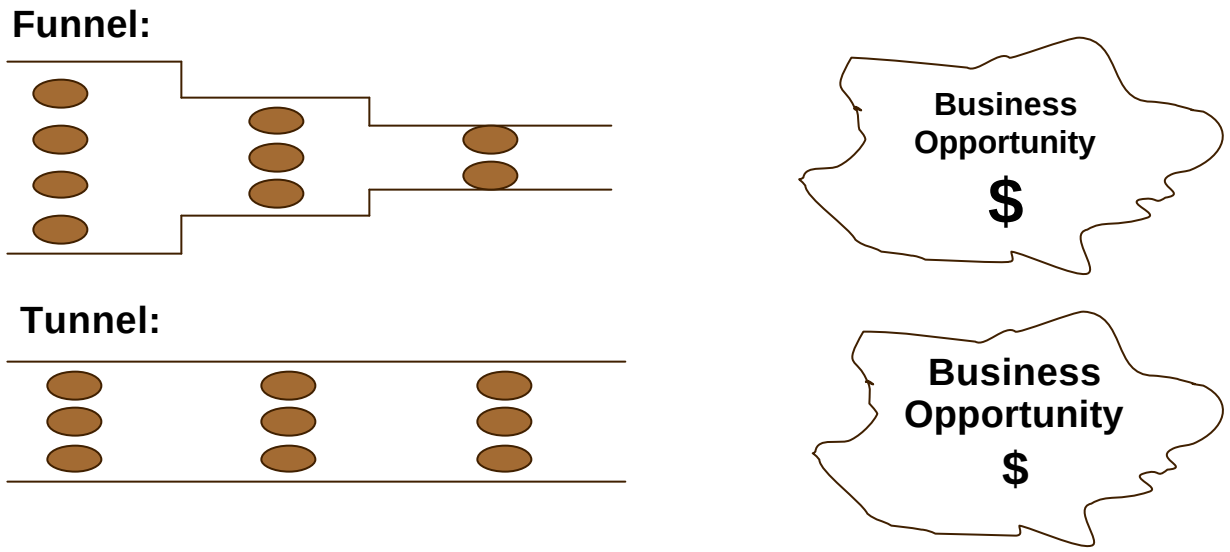


Figure 3: Decision Tree

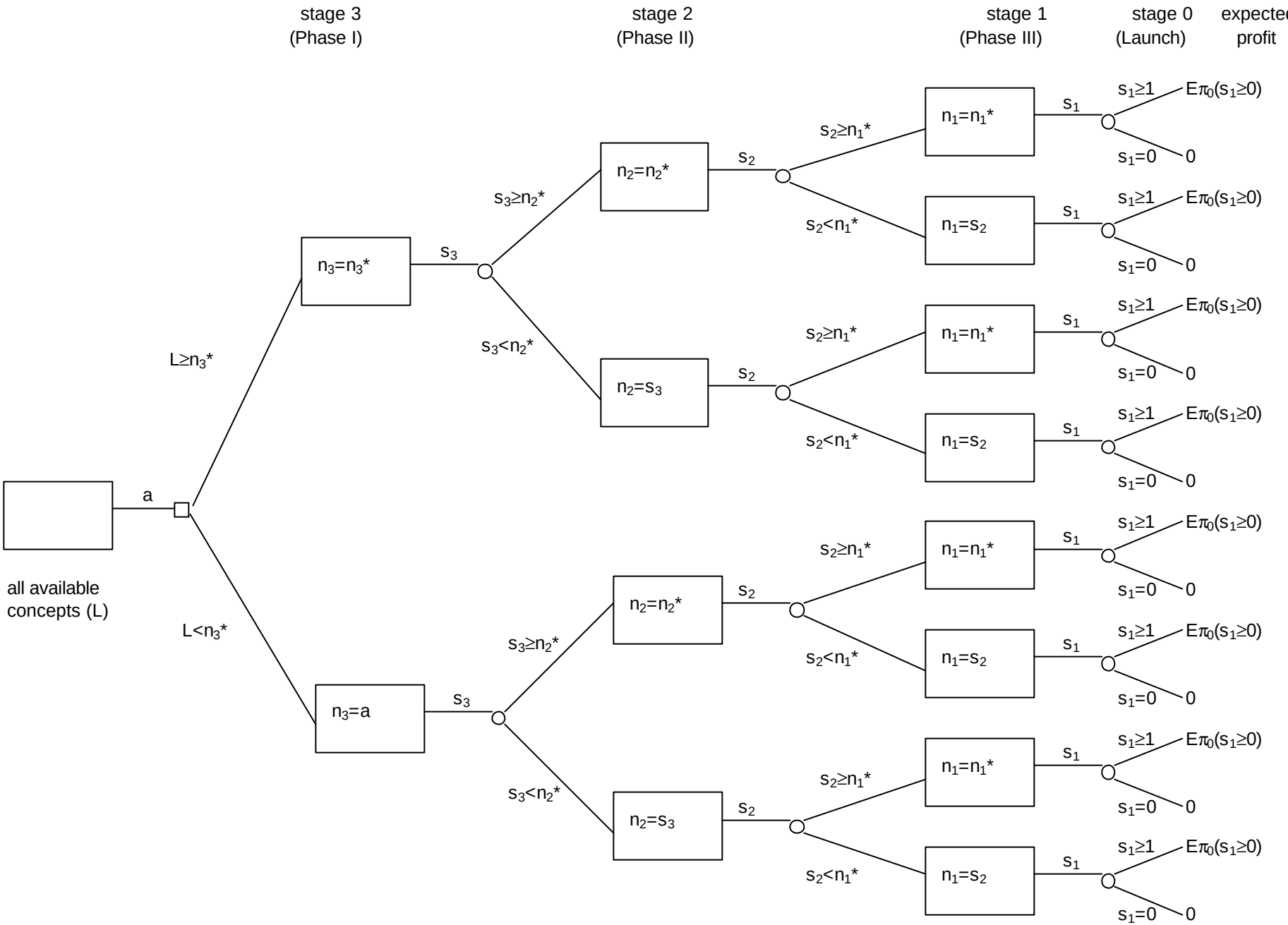


Figure 4: One Stage-Process Analysis

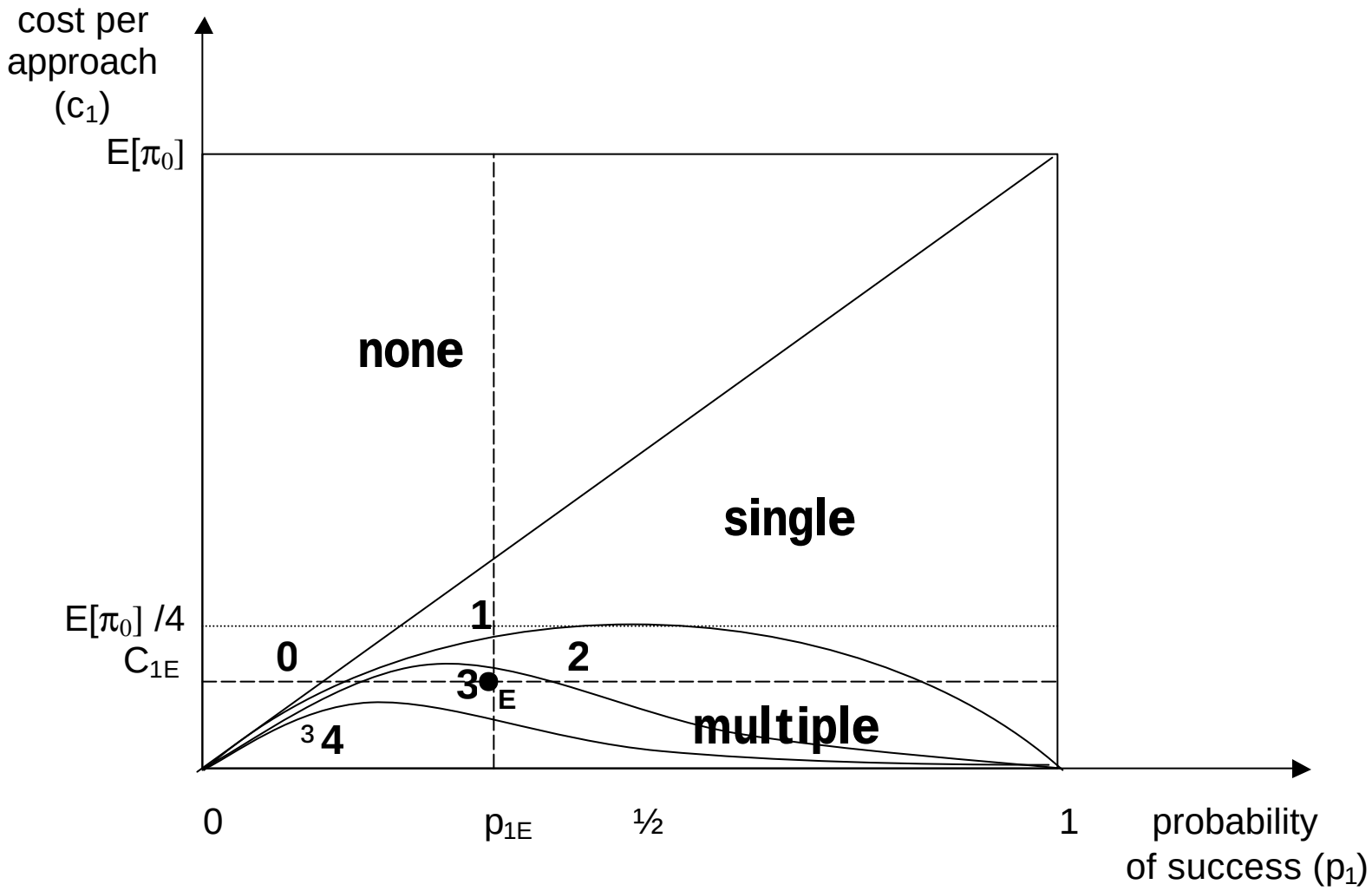
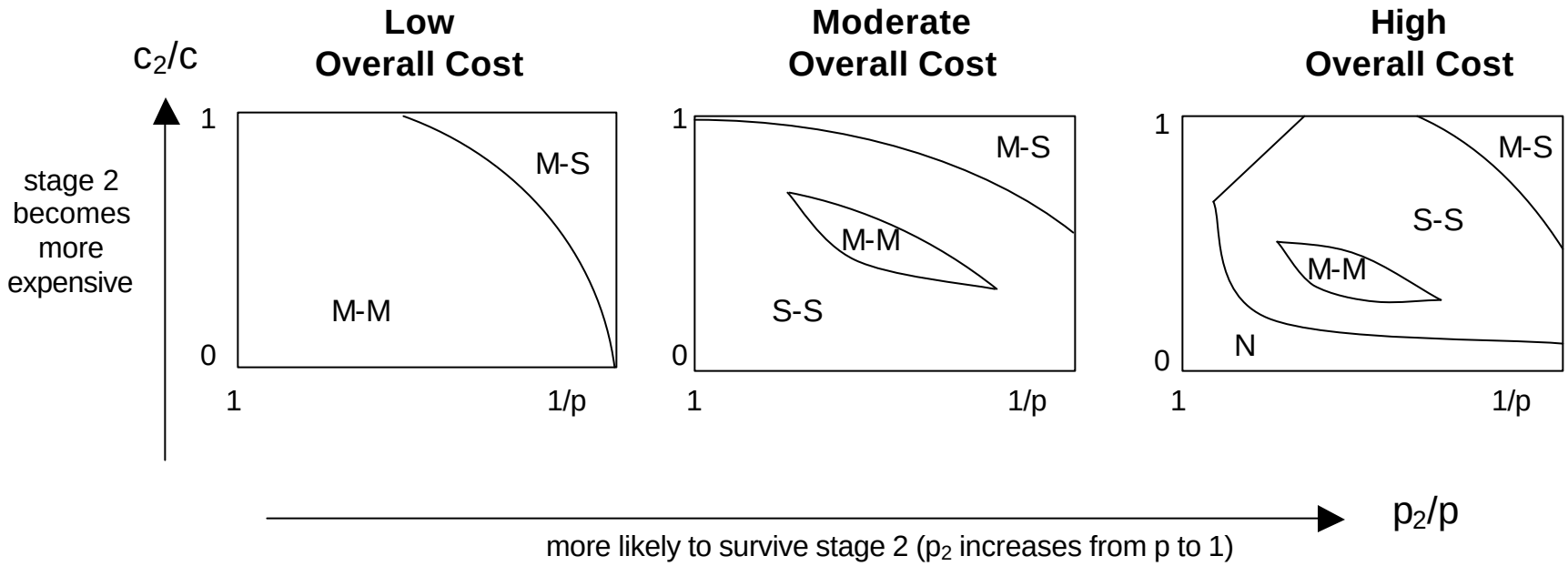


Figure 5: Two-Stage-Process Analysis



The abbreviations represent the optimal NPD strategy under various conditions:
S-S: fund single (n*=1) approach at both stages;
M-S: fund multiple (n*>1) in stage 1, single (n*=1) in stage 2(last stage).
M-M: fund multiple (n*>1) approaches at both stages;
N: none (n*=0) should be funded.

Figure 6: An Illustrative Example

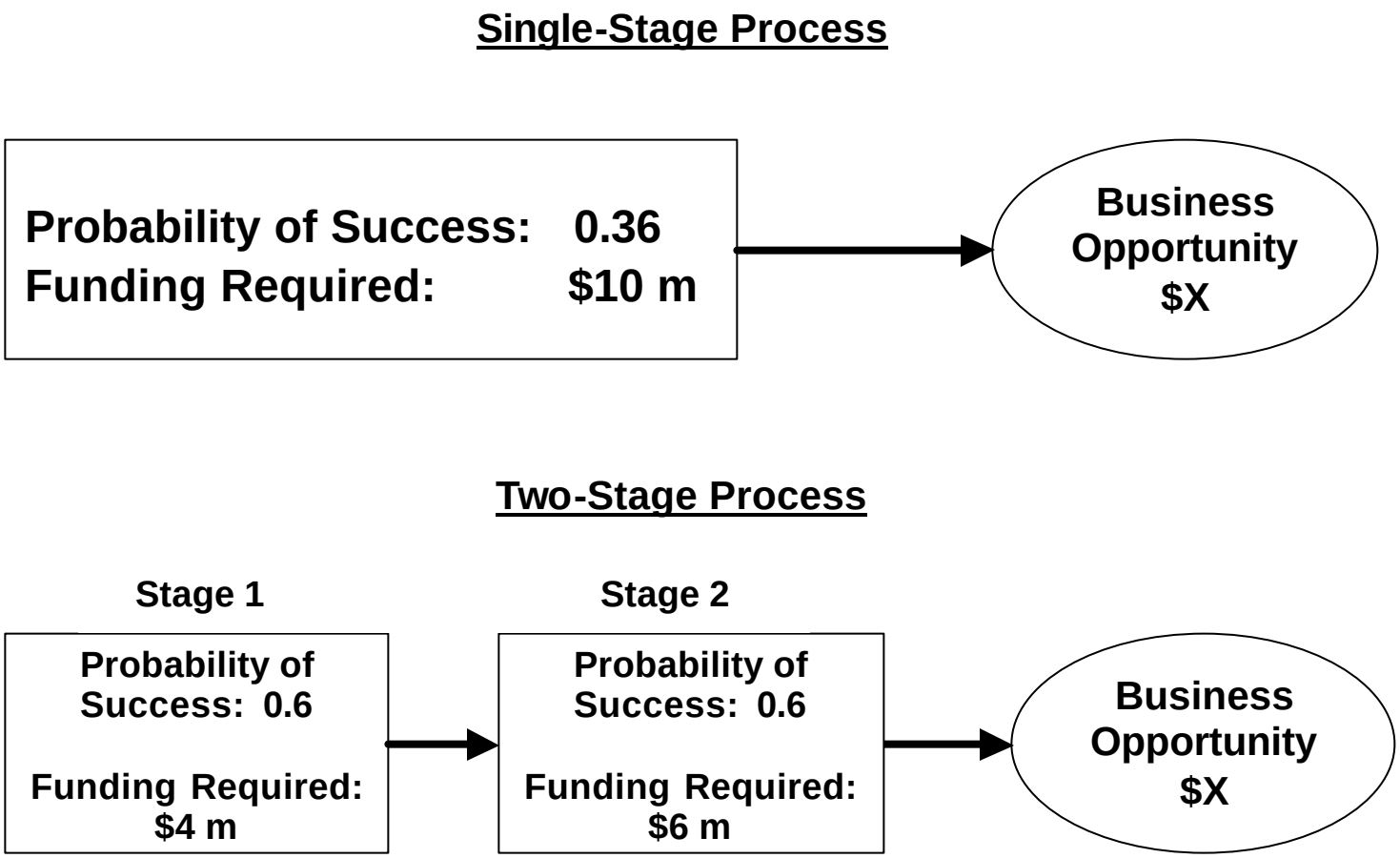
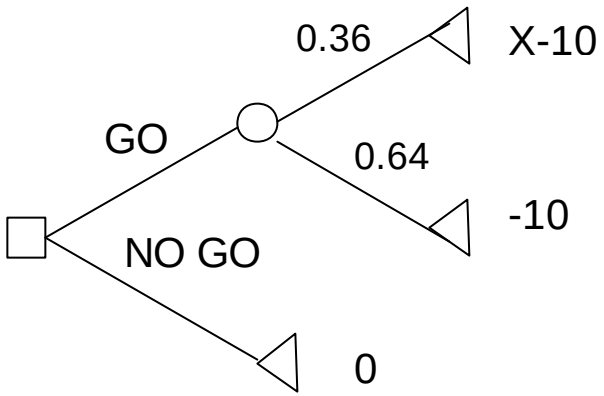


Figure 7: Decision Trees for the Illustrative Example

Decision Tree #1



Decision Tree #2

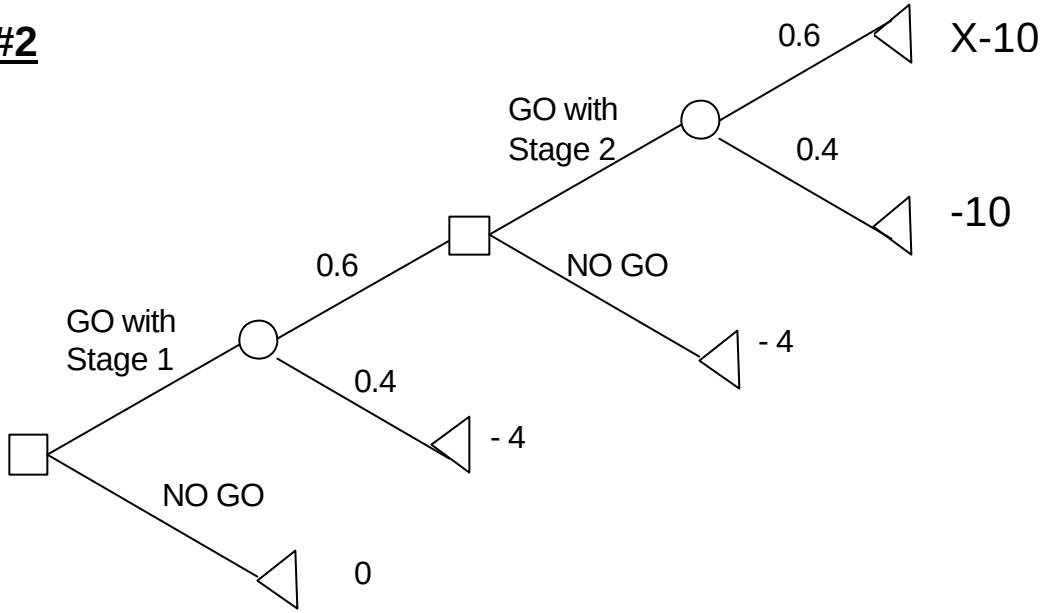


Table 1: Summary Results for Two-Stage NPD Pipeline

<table>
<tr><td></td><td colspan="4">First NPD Stage (screening stage)*</td></tr>
<tr><td></td><td>effective AND cheap[1]</td><td>semi-effective AND medium cost[2]</td><td>ineffective OR expensive[3]</td><td>In between[4]</td></tr>
<tr><td>Low Overall Cost</td><td rowspan="3">Multiple (>1) First
Single (1) Later</td><td colspan="3" rowspan="2"></td></tr>
<tr><td>Medium Overall Cost</td></tr>
<tr><td>High Overall Cost</td><td colspan="2">Both Multiple (>1)</td><td colspan="2">Both Single (1)
none (0)</td></tr>
</table>

Notes

* Relative to the second NPD stage's cost and probability of survival.

1. Large p_2/p AND Large c_2/c (Small p_1/p AND Small c_1/c)
2. Medium p_2/p AND medium c_2/c (Medium p_1/p AND Medium c_1/c)
3. Small p_2/p OR Small c_2/c (Large p_1/p OR Large c_1/c)
4. All other situations

Table 2: AIDS Vaccine Pipeline

Clinical Trial Stages	Cost per Prototype Vaccine ($m)	Probability of Success	Actual Pipeline for a Firm:	Optimal Pipeline of Prototype Vaccines Recommended by the Model for:	
				a firm ($130m)	public policy maker ($113.5b)
Phase 1	2	0.75	<3	**5**	**34**
Phase 2	4	0.48	<3	**5**	**25**
Phase 3	13	0.63	N/A	**2**	**9**

Table 3: Probabilities of Returns: AIDS Vaccine Pipeline

		probabilities	
		firm	public
Specified Range of Returns		$73-187 m	$113.5 b
condition under which the probabilities are assessed	Start with optimal number at Phase I	**39%**	**99.97%**
	Start with optimal number at Phase II	**44%**	**99.98%**
	Start with optimal number at Phase III	**48%**	**99.99%**
	Have at least one successful vaccine for Launch	**56%**	**100%**

Table 4: Sensitivity Analysis for AIDS Vaccine Pipelines

	optimal number of prototype vaccines recommended by our model					
	firm			public policy maker		
Clinical Trial Stage	Phase 1	Phase 2	Phase 3	Phase 1	Phase 2	Phase 3
Base Case	**5**	**5**	**2**	**34**	**25**	**9**
Vary the number of competitors (m)						
2	11	9	4	34	25	9
20	3	3	2	34	25	9
Vary the strength of competitors (probability of success p)						
0.78	3	3	2	34	25	9
0.23	8	7	3	34	25	9
Vary probability of success, Phase 3 (p_1)						
0.3	4	4	4	65	50	23
0.8	5	5	2	27	19	6
Vary probability of success, Phase 2 (p_2)						
0.2	6	6	2	77	60	9
0.7	4	3	2	23	16	9
Vary probability of success, Phase 1 (p_3)						
0.5	7	5	2	52	25	9
0.9	4	4	2	27	25	9
Vary cost, Phase 3 (c_1)						
8 m	6	5	3	34	25	10
20 m	5	4	2	33	25	9
Vary cost, Phase 2 (c_2)						
2 m	6	6	2	34	27	9
8 m	4	3	2	33	23	9
Vary cost, Phase 1 (c_3)						
1 m	6	5	2	35	25	9
4 m	4	4	2	32	25	9
Varying expected profit (firm)						
100 m	4	4	2	34	25	9
200 m	7	6	3	34	25	9
Varying expected benefit (public policy)						
100000 m	5	5	2	33	24	9
200000 m	5	5	2	36	26	10

Table 5: Seven NPD Challenges in Pharmaceutical Industry

Cardiovascular Class (3)			
Indication	**Expected Market Return (firm), E[p] ($m)**	**Actual Pipeline (leading firm)***	**Model Recommendation (leading firm)***
Arrhythmia	191	1 ® 1® N/A	2 ® 2 ® 2
High cholesterol	7,858	1 ® 1® N/A	16 ® 10 ® 4
Hypertension	10,334	1 ® 1 ® 1	17 ® 11 ® 5
	Phase 1	**Phase 2**	**Phase 3**
Cost (capitalized), c ($m)	8.47	13.48	33.38
Probability of Survival, p	0.639	0.566	0.724

Neuropharmacological Class (3)			
Indication	**Expected Market Return (firm), E[p] ($m)**	**Actual Pipeline (leading firm)***	**Model Recommendation (leading firm)***
Alzheimer's disease	8,021	1 ® 1 ® N/A	20 ® 19 ® 7
Migraine headache	1,099	1 ® 1 ® 1	11 ® 11 ® 4
Depression	5,238	2 ® 2 ® 1	18 ® 17 ® 7
	Phase 1	**Phase 2**	**Phase 3**
Cost (capitalized), c ($m)	4.31	8.05	33.94
Probability of Survival, p	0.898	0.442	0.511

NSAID Class (1)			
Indication	**Expected Market Return (firm), E[p] ($m)**	**Actual Pipeline (leading firm)***	**Model Recommendation (leading firm)***
Arthritis (COX-2)	3,059	2 ® 2 ® 2	12 ® 10® 3
	Phase 1	**Phase 2**	**Phase 3**
Cost (capitalized), c ($m)	11.53	18.15	68.34
Probability of Survival, p	0.750	0.417	0.709

*: The three numbers corresponding to phase 1, 2, and 3, respectively.
N/A corresponds to no data available.