GLEN L. URBAN and GERALD M. KATZ*

The predictive accuracy of a widely used pre-test-market model (ASSESSOR) is analyzed. The standard deviation between pre-test-market and test-market shares is 1.99 share points before adjustments for achieved awareness, distribution, and sampling and 1.12 share points after adjustment. Sixty-three percent of those products tested passed the pre-test screen and 66% of these were subsequently successful in test market. A Bayesian decision analysis model is formulated and a "typical" case shows a positive value of information. Although some conditions are identified under which a test market may be bypassed, in the authors' opinion both pre-test and test-market procedures should be used in all but exceptional situations.

Pre-Test-Market Models: Validation and Managerial Implications

Laboratory pre-test-market models designed to forecast the sales and/or share of new frequently purchased consumer products have received substantial attention from academicians and practicing market researchers over the last decade. Eskin and Malec (1976) proposed a trial/repeat purchase forecasting procedure which later was extended by Kalwani and Silk (1980). In 1978 Silk and Urban documented the ASSESSOR model. Recently the COMP (Burger, Gundee, and Lavidge 1981), LTM (Yankelovich, Skelly, and White 1981), and NEWS (Pringle, Wilson, and Brody 1982) models have been documented in the published literature. Other pre-test-market models are documented by Urban and Hauser (1980).

These pre-test-market models have been implemented in many organizations. For example, between 1973 and 1980 ASSESSOR was used to evaluate more than 200 products in more than 50 organizations and is being used to evaluate about 75 new brands this year. NEWS has been applied on a pre-test basis to 57 brands (Pringle, Wilson, and Brody 1982, p. 26) and other models have been implemented numerous times.

Although pre-test-market models have been widely used in the packaged goods industry, there is a paucity of data available for determining their accuracy. Many advertising claims have been made, but few have been substantiated in professional journals.[1] COMP (Burger, Gundee, and Lavidge 1981) predictions recently were evaluated for eight brands for which predicted and actual test market shares were available. The reported mean absolute deviation is surprisingly small at .125 share point or 1.5% of the average share. This claim of exceptional accuracy is in contrast to Tauber's (1977) account of four validations of an unnamed simulated test-market procedure. One product was predicted to be a "winner" but failed, one was predicted to be a "loser" but sold over $40 million per year, and the other two were predicted to fail and did so. Recently more reasonable results were reported for the NEWS model (Pringle, Wilson, and Brody 1982). Based on a comparison of actual and predicted

*Glen L. Urban is Professor of Management, Alfred P. Sloan School of Management, MIT. Gerald M. Katz is Vice President, Management Decision Systems, Inc.

The authors extend special thanks to Steve Shugan who helped formulate the sequential decision theory model and John Hauser, Leigh McAlister, John Roberts, and Len Lodish who made valuable criticisms and suggestions. They also thank Management Decision Systems, Inc. and Novaction, S.A. for supplying the data for the study, Barbara Petroulis, Mary Lyons, and Robert Wallace who made major contributions to data collection, and the cooperating companies which responded to the questionnaire.

[1]For example, Elrick and Lavidge's COMP ad claims "correct to the nearest market share point more than 95% of the time" (*Marketing News*, Nov. 27, 1981, section 2, p. 13). NPD's ad for ESP claims "forecasts of year one volume within 9.9% of in-market results" (*Marketing News*, Nov. 27, 1981, section 2, p. 5). Yankelovich, Skelly, and White claim "95% record of accurate validations" for their LTM system (*Laboratory Test Market-Record of Validations*, New York, Yankelovich, Skelly, and White, undated).

Journal of Marketing Research
Vol. XX (August 1983), 221–34

test-market shares for 22 brands, a coefficient of variation of 25% is reported.

The only other journal record of comparisons between predicted and actual shares for a laboratory pre-test model is in the original Silk and Urban (1978) article. The mean absolute deviation across nine products was .9 share point or 13.8% of the average share. Since publication of this article, more than 200 new brands have been evaluated with the ASSESSOR model. We report results based on this experience.

The ASSESSOR model forecasts sales and/or market share for a new brand, provides a structure for evaluating alternative marketing strategies, and generates diagnostics that aid in improving the product. The forecast is based on the convergence of two models—a trial/repeat model and a preference model. In the trial/repeat model one uses laboratory measures of trial and survey measures of repeat purchase of the new product. In the preference model one uses existing brand preference and the last brand purchased to parameterize a logit model. Preference for the new product after use is substituted into the logit model to predict share. These data, along with the effects of management's introductory marketing plan (advertising, promotion, sampling, and couponing) on brand awareness and distribution, are used to derive a convergent forecast (see Silk and Urban 1978 for complete details).

The purpose of our article is to investigate this pre-test-market model's predictive accuracy and examine the managerial implications of such a predictive tool. We propose a Bayesian sequential decision model to help assess the value of the information from such a pre-test-market model. This decision model is useful in determining the effects of varying levels of accuracy and identifying any conditions under which a traditional test market might be eliminated.

VALIDATION

Difficulties

Assessing the accuracy of a pre-test-market model is fraught with conceptual and empirical difficulties. For instance, the initial forecast is developed before test market, but the actual results reflect any deviations from the original marketing plan. For example, the planned sampling might be late, the advertising might be increased, the promotion might be decreased, or the product formulation modified. The initial distribution plan may not be achieved or the estimated awareness from advertising may prove to have been optimistic. One could "adjust" for such errors by rederiving the forecast under the plan executed, but this procedure could bias the assessment of errors. It is too easy to find some explanation for the differences between actual and predicted results. If adjustments are to be made, they should be made for all cases—not just those in which deviations between actual and predicted results are large.

Another difficulty in validation is the long time lag between the pre-test-market forecast and the end of test market (usually 12 to 15 months). Many economic, political, and social events exogenous to the model may take place. During the time lag, the product may be subject to an organizational change such as a shift from the new products group to the brand group. In this case learning the actual share may be difficult because the brand manager is more concerned with confidentiality than model accuracy.

Comparison of the forecast with the actual test-market share is straightforward, but the test-market share is also subject to measurement error. In addition, competitive action in the test market may bias the share. Many test markets do not run long enough to obtain a steady-state reading of the actual share. Furthermore, the market share itself is subject to variance due to audit procedures and may not be calculated on the basis of the same category definition used in the pre-test-market analysis.

A final difficulty is sample size. Although many products are evaluated, only a fraction of them go to test and not all firms will disclose the market share for these products.

We report a systematic attempt to address these difficulties. The sampling plan is described, initial and adjusted shares are reported, nonresponse bias is examined, and an overall assessment of accuracy is made.

Sampling Design

All firms in which ASSESSOR was applied were surveyed and all product forecasts examined.[2] The 215 products included in the sample covered a variety of categories: 32% food, 23% household cleaning, 34% health and beauty aids, 9% over-the-counter pharmaceuticals, and 2% other. The studies reflected a wide spectrum of new product situations. Seventy-four percent were done in the United States and 26% were international (Europe, Japan, Australia, and Canada), 65% were major new product studies and 35% were line extension studies, 26% included multiple callbacks (sales waves), 15% were in categories having purchase frequency of less than three purchases per year, and 15% were in situations where the category was not well defined or did not exist.

For validation the pre-test-market share was drawn from the original report delivered to the sponsoring organization. The test-market share was obtained from a questionnaire sent to the sponsoring firm. In the survey the manager specified his or her interpretation of the managerial implication of the pre-test forecast ("definitely should go to test market, probably should go to test market, borderline case, probably should not go to test market, and definitely should not go to test market"), and the actual outcome ("went to test market, went national, was killed, went into redevelopment, or was already in

[2]Lists of firms were supplied by Management Decision Systems, Inc. and Novaction, Inc. The cooperation of these firms is gratefully acknowledged.

market").[3] The test-market result was specified by the observed market share and by the manager's qualitative assessment of the test-market result ("a big success, fairly successful, marginal, a slight failure, or a big failure"). Differences between the pre-test and test-market conditions were reported and the model rerun with the actual conditions to derive an "adjusted" share prediction. These adjustments were limited to three variables: awareness, distribution, and sampling.

Results

Two hundred and fifteen questionnaires were sent. After two followup mailings and phone calls, 81 questionnaires or 38% were returned. Within these 81, 29 products did not go to test market, 5 were in test market but it was too early to observe the final share, and 47 had completed test market. In three of the 47 test-market cases, market share could not be supplied because a competitor's product had been tested and the actual share was not known.

Figure 1 shows the comparison of pre-test-market and test-market shares. The correlation is .95. The first column in Table 1 lists the overall means and deviations of these data. The average pre-test-market share forecast was 7.77 whereas the average test-market share was 7.16. Thus, a positive bias of .61 share point is present and is significant at the 10% level (t = 2.0). The average absolute deviation is 1.54 share points and the standard deviation is 1.99 share points.

As expected, the comparisons between adjusted and test-market shares show less error—mean deviation of −.01, average absolute deviation of 0.83, and standard deviation of 1.12. The correlation of the adjusted predictions with test-market shares is .98. Adjustments were made in 36 of the 44 cases. In most of these the adjustments improved the accuracy, but in six of the cases the deviation increased. The systematic overprediction for lower share values shown in Figure 1 was reduced substantially by the adjustments.

The validation sample consists of 13 health and beauty aid (HBA) products, 11 household cleaning products, and 20 food products. Table 1 reports the individual category results. The absolute differences are small and none of the paired comparisons of means or variances are significantly different at the 10% level. Similar levels of accuracy are observed across these product categories.

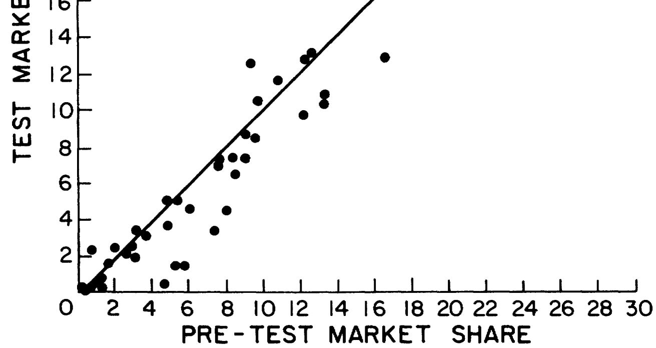
Nonresponse is a threat to the validity of the reported accuracy and must be examined. The pre-test share forecasts for all 215 studies have a mean of 7.13 and a standard deviation of 6.55. The 44 products in the validation sample have a pre-test share mean of 7.77 and a standard deviation of 5.72. They are not significantly different estimators of the true underlying population mean and variance (t = .46, $F(214, 43)$ = 1.31). Further analysis of nonresponse was done by comparing the first wave of 24 responses with the 20 later responses. The mean shares were 7.9 for the first 24 questionnaires and 7.7 for the last 20 questionnaire responses. These differences are not significant at the 10% level. The standard deviations be-

Table 1
COMPARISON OF PRE-TEST-MARKET, ADJUSTED, AND TEST-MARKET SHARES

	Overall (*n* = 44)	*HBA* (*n* = 13)	*Household* (*n* = 11)	*Food* (*n* = 20)
Average test market share	7.16	7.35	10.14	5.40
Pre-test versus test-market share				
Mean difference	0.61	0.43	0.61	0.73
Mean absolute difference	1.54	1.66	1.37	1.56
Standard deviation of differences	1.99	2.08	1.71	2.06
Adjusted versus test-market share				
Mean difference	−0.01	−0.29	−0.15	0.25
Mean absolute difference	0.83	0.88	1.04	0.68
Standard deviation of differences	1.12	1.09	1.23	1.02

[3]This response is allowed because some firms tested competitors' products that were already in test market or applied the model to their own product for validation purposes after it had gone to test market.

Figure 1
COMPARISON OF PRE-TEST-MARKET AND TEST-MARKET SHARES



tween pre-test and test-market shares were virtually identical (2.0 versus 2.0 for unadjusted and 1.0 versus 1.1 for adjusted comparisons). There is no apparent evidence of a nonresponse bias and the sample does not appear to be significantly different from the 215 studies.

Consistency with Sampling Theory

Knowing that the standard deviation between pre-test and test-market forecasts is 2.0 share points whereas the adjusted standard deviation is 1.1 share points, and that the average sample size for the studies in Table 1 is 489, we now ask whether the results are consistent with random sampling theory, i.e., is the reported accuracy "too good"?

We compare an estimate of the standard deviation due to sampling with the standard deviation of error in the trial/repeat portion of the pre-test-market model. If the standard deviation due to sampling is greater than the observed standard deviation, an inconsistency is implied or the convergence between the trial/repeat and preference models works to decrease error.

The sampling error can be calculated by recalling that one of the ASSESSOR share forecasts is the product of an estimate of ultimate cumulative trial (T) and the share of purchases among persons who have tried (S) (see Silk and Urban 1978 for complete details). If we assume the sampling errors for T and S are independently distributed random variables, we can calculate the variance of the product ($x = TS$) by (Goodman 1960; Shoemaker and Staelin 1976; Urban 1968):[4]

$$\sigma_x^2 = \sigma_T^2\sigma_S^2 + \mu_T^2\sigma_S^2 + \mu_S^2\sigma_T^2 \tag{1}$$

where:

σ_x = standard deviation of market share,
μ_T = mean ultimate trial proportion,
σ_T = standard deviation of ultimate trial,
μ_S = mean share of purchases among persons who have tried, and
σ_S = standard deviation of share of purchases among persons who have tried.

If we assume ultimate trial (T) and share among triers (S) are the result of sampling from a binomial population, we have:

$$\sigma_T = \sqrt{\frac{\mu_T(1-\mu_T)}{N_1}} \tag{2}$$

$$\sigma_S = \sqrt{\frac{\mu_S(1-\mu_S)}{N_2}} \tag{3}$$

where N_1 and N_2 are the sample sizes available for estimating T and S, respectively. If we approximate the true population means by the observed results $\mu_T = .21$ and $\mu_S = .34$, using the sample sizes from Table 1, $N_1 = 489$ and $N_2 = 152$, we now calculate $\sigma_x = 1.0$.[5]

This value of 1.0 is an estimate of the random sampling error and can be compared with the pre-test-market model standard deviation. In this comparison we use the adjusted standard deviation because it is our best estimate of the residual error and is a conservative value for this consistency check. We must realize that the reported standard deviation is the difference between pre-test and test-market forecasts and is not the standard deviation of the pre-test model alone because the test market is also subject to error. We know

$$\sigma_{x_1-x_2}^2 = \sigma_{\epsilon_1}^2 + \sigma_{\epsilon_2}^2 - 2\rho\sigma_{\epsilon_1}\sigma_{\epsilon_2} \tag{4}$$

where:

$\sigma_{x_1-x_2}^2$ = variance of difference between pre-test (x_1) and test-market share (x_2), i.e., variance of observed errors,
$\sigma_{\epsilon_1}^2$ = variance of pre-test errors (ϵ_1),
$\sigma_{\epsilon_2}^2$ = variance of test-market errors (ϵ_2), and
ρ = observed correlation of pre-test and test-market errors ϵ_1 and ϵ_2.

We observe $\sigma_{x_1-x_2}$ is equal to 1.12 share points from Table 1 and assume the pre-test and test errors are uncorrelated ($\rho = 0$).[6] If we further assume the test market is twice as accurate as the pre-test ($\sigma_{\epsilon_1} = 2\sigma_{\epsilon_2}$) we can calculate, by equation 4, $\sigma_{\epsilon_1} = 1.0$. This value can now be compared with an estimate of sampling error.

Using conservative assumptions we have found the reported error for ASSESSOR ($\sigma_{\epsilon_1} = 1.0$) to be equal approximately to the theoretical sampling error of the trial/repeat model. This finding indicates a possible inconsistency because there is no residual observed error to account for other than sampling. However, recall that the ASSESSOR model utilizes a convergent method to forecast share. A logit model based on changes in preferences after use of the new brand is also calculated. If we were to consider the use of two alternative methods as increasing the equivalent sample size, or if the errors in the two methods were negatively correlated, the estimate of random sampling error would decrease and the difference between the observed and sampling errors would increase. If this convergence argument is not accepted, we would conclude that the observed errors may be understated. In our subsequent evaluation of the value of information of the model we increase the observed errors to allow for this possibility.

[4]This formula was used by Shoemaker and Staelin (1976) to evaluate the random sampling errors in a test-market model. See their article for a complete discussion of the appropriateness of this approach. Note that the mean trial rate and mean repeat can be correlated across products. We assume only that the sampling errors are independent.

[5]N_2 is 489 times the trial in the laboratory given awareness and availability (.31).

[6]This is a conservative assumption in this consistency check. If ρ is greater than zero, σ_{ϵ_1} is increased and more likely to be greater than the random errors.

Table 2
FRACTION PASSING ASSESSOR

	ASSESSOR result					
Action	*Definitely GO*	*Probably GO*	*Borderline*	*Probably NO*	*Definitely NO*	*Total*
Went to test	13	6	1	2	1	23
Went national	5	6	0	0	0	11
Killed	0	2	0	4	9	15
Redevelopment	3	1	3	4	4	15
Already in market	0	3	3	2	1	9
Too soon to tell	1	0	0	0	0	1
Total	22	18	7	12	15	74

Fraction-Passing Tests

It is useful to consider the fraction of products that pass the pre-test-market screen and subsequently succeed in test market. The fraction of products that pass the pre-test-market screening can be calculated from the manager's qualitative assessment of the pre-test results along with the actions taken (Table 2). Overall, 40 of the 74 products (54%) passed the pre-test ("definitely or probably go").[7] If we split the "borderline" category proportionally to make a dichotomous classification, 60% pass [(22 + 18 + 7(40/67))/74 = .60]. Furthermore, some of the products in "redevelopment" will succeed. If we assume the same success rate (p) for them as for an average product, the overall proportion passing the pre-test is 63% [$(22 + 18 + 11p)/74 = p$, $p = .63$]. If we split the borderline class and consider redevelopment, 65% [$(22 + 18 + 11p + 3(40/67))/74$] is the estimate of successful completion of the test. Overall, we conclude that approximately 63% of the products pass the pre-test ASSESSOR.

Table 3 reports the manager's qualitative assessment of the pre-test result versus the actual in-market result. The fraction of products succeeding in test market given that they pass pre-test ("definitely or probably go") is 63% [(7 + 2 + 1 + 7)/(16 + 11)]. If the "marginal" class is split proportionally to define a dichotomous variable, the fraction is .69 [(7 + 2 + 1 + 7 + 3(18/35))/(16 + 11) = .69]. If the "borderline" class is also split proportionally, the proportion is 62% [(7 + 2 + 1 + 7 + 3(18/35) + 1(27/33) + 1(27/33)(18/35))/(16 + 11 + 6(27/33)) = .62]. Averaging these two estimates, we conclude that approximately two-thirds (66%) of the products passing the ASSESSOR pre-test-market screen subsequently succeed in test market.

A. C. Nielsen Co. has systematically studied the success rates for products in test market over the past two decades. "Success" was defined by whether or not the new product was subsequently launched nationally. The most recent study (1977) included 228 test-marketed items. Brands withdrawn from test markets or not launched nationally were considered failures. In this most recent study, Nielsen found a 35.5% success rate, a continuation of a trend toward lower success rates reported in their earlier studies (54.4% in 1961 and 46.6% in 1971). Above we estimate that among those products that passed ASSESSOR, 66% were judged by managers as successful in test market. If we assume these products will be or have been launched nationally, this 66% compares favorably with Nielsen's 35.5%. Other earlier studies by Buzzell and Nourse (1967) and General Foods (*Business Week* 1973) suggest higher success rates that are consistent with the Nielsen observations in 1961 and 1971. If the Nielsen trend is correct, introducing new products is becoming more risky than ever and pre-test-market analysis will become increasingly important in reducing the risk of test-market failure.

[7]On seven questionnaires, data on these questions were missing.

Table 3
FRACTION SUCCEEDING IN TEST MARKET

	ASSESSOR result					
In-market result	*Definitely GO*	*Probably GO*	*Borderline*	*Probably NO*	*Definitely NO*	*Total*
A big success	7	1	0	0	0	8
Fairly successful	2	7	1	0	0	10
Marginal	3	0	1	0	0	4
A slight failure	2	1	1	0	0	4
A big failure	2	2	3	3	3	13
Total	16	11	6	3	3	39

Table 3 also shows that six products were test marketed despite a negative pre-test evaluation. All of them subsequently were judged by management as "big failures" in test market. However, if the "borderline" and "marginal" classes are allocated as before, we can conclude that 3.8% would succeed after failing ASSESSOR [(1(6/33) + 1(6/33)(18/35))/(6(6/33) + 6) = .038]. In most cases, after a "NO" forecast at the pre-test stage, a product is dropped and it is not known whether it would have succeeded. In these conditions validation is not possible; however, in the six cases reported in Table 3 the products were either already in the test market or the model recommendations were overridden by other considerations. This sample is small, and although it indicates little chance of success after a product fails the pre-test-market evaluations, we must view this result as tentative.

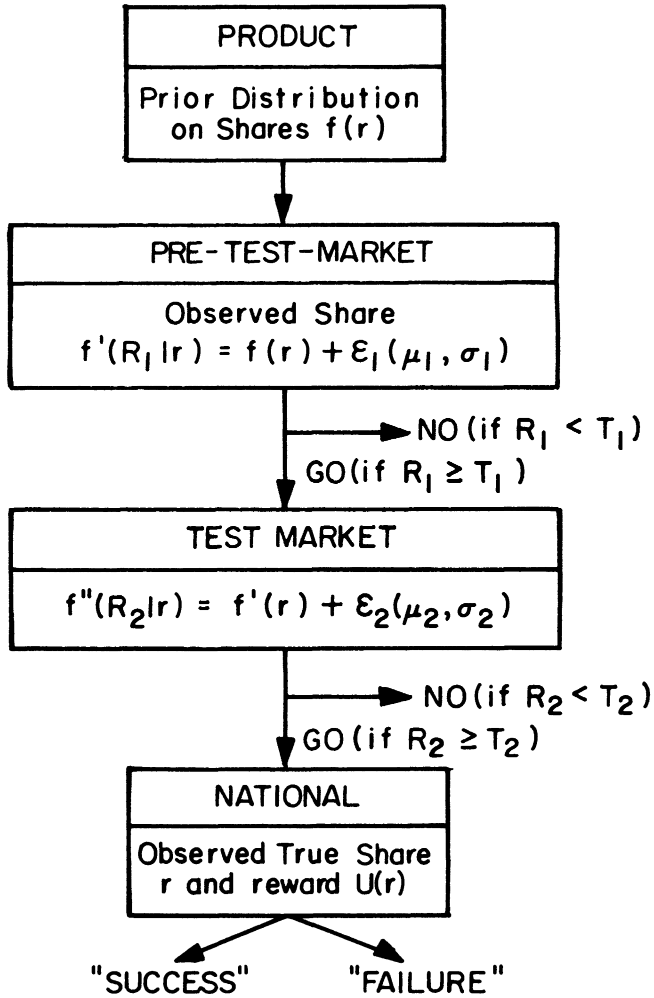
The preceding analysis also must be considered with some caution because the judgments of the ASSESSOR result were collected in our survey *after* the actual test market. There is a chance that the actual test-market result could affect the reported ASSESSOR result. Although we have no evidence to suggest this bias, the data could understate the model's misclassifications if it is present. Subject to this caveat we now summarize the validation results.

Summary of Validation Results

The comparison of pre-test-market and test-market shares shows a mean deviation of approximately one half of a share point. Adjustments for differences between pre-test and test-market conditions reduced the mean error to almost zero. The standard deviation also dropped from 2.0 to 1.1 share points with the adjustments. Approximately 63% of products tested pass ASSESSOR and two thirds of these succeed in test market. Estimates based on a very small sample of products launched in spite of a pre-test-market failure indicate that only 4% of products failing pre-test would succeed in test market.

Several questions naturally arise: Are these accuracy levels good enough? Is conducting a pre-test-market analysis worth the cost? What is the effect of delay due to testing? What is the value of increased accuracy? Are there conditions when test market can be bypassed? In the next sections we address these questions by formulating a sequential decision model and studying its behavior.

Figure 2
SEQUENTIAL DECISION MAKING



r = true share
R_1 = observed pre-test forecasted share
R_2 = observed test-market share

IMPLICATIONS

Modeling the Decision Problem

Figure 2 depicts the new product decision process, from the point of availability of a "finished" product, which includes its price, packaging, advertising, and physical formulation. A concept test, product test, and/or advertising copy test has usually already been conducted by the time a product reaches this stage. The pre-test analysis is an additional screening device intended to eliminate product failures at a low cost (e.g., $50,000) rather than carrying them on to test market where they would fail at a high cost (e.g. $1.5 million). However, in eliminating failures the pre-test evaluation may also screen out a product that would have been a success. The manager's task, therefore, is to set a GO/NO cutoff value for the pre-test share prediction that balances these errors. If the share is greater than or equal to the cutoff (T_1), the product will go on to test market; if the share is below the cutoff, the product will be dropped or improved and retested. In test market a GO/NO cutoff must be applied again. If the share is greater than or equal to the cutoff (T_2), the product is launched nationally where its true potential is finally observed. The national share will be a key determinant of the profit the firm will earn and will lead management to declare the brand a success or failure.

The decision problem is to set T_1 and T_2 so that the firm maximizes its expected profit. This is a sequential decision problem which we model by use of decision theory (see DeGroot 1970; Raiffa and Schlaifer 1961; Rao and Winter 1981). The population of products available for pre-test-market analysis is depicted by a prior distribution of market shares ($f(r)$). All values of share are positive and the distribution is likely to be unimodal with many average brands and a few very large share brands. The distribution is also likely to be positively skewed (away from zero) because most very small share brands have been eliminated by previous testing. Figure 3 is a typical prior distribution.

We model the distribution of observed shares from pre-test-market ($f'(R_1|r)$) as the prior distribution ($f(r)$) plus an error term (ϵ_1).[8] This error distribution reflects the predictive ability of the test; the mean (μ_1) measures the bias and the standard deviation (σ_1) the inaccuracy due to random effects. This distribution is likely to be unimodal, but not necessarily normal. The cutoff is applied to the observed result (R_1), and if it is greater than or equal to T_1 the product goes to test market.

The distribution of market shares at this point is the prior distribution less those products eliminated at the pre-test stage. This distribution ($f'(r)$) plus a test-market error term ($\epsilon_2(\mu_2, \sigma_2)$) represents the observed distribution of observed test-market share. The test-market cutoff (T_2) is applied and the product goes national if $R_2 \geq T_2$. After national introduction the product produces profits or losses depending on its true share (r). A typical reward function for a major brand over its life cycle is illustrated in Figure 4. Introductory losses (advertising, promotion, plant and equipment) are depicted as \$55 million if share is zero; breakeven is at 5.75 share points; \$80 million is returned at 10% market share and \$190 million at 20% market share. The losses are relatively

[8]The additive error assumption is supported by the lack of significant difference between means and variances across categories where share ranged from 5.4 to 10.1 (see Table 1). A regression of the squared deviations between pre-test and test-market share versus test-market share was not significant at the 10% level (R^2 = .06).

Figure 3
PRIOR DISTRIBUTION OF MARKET SHARES

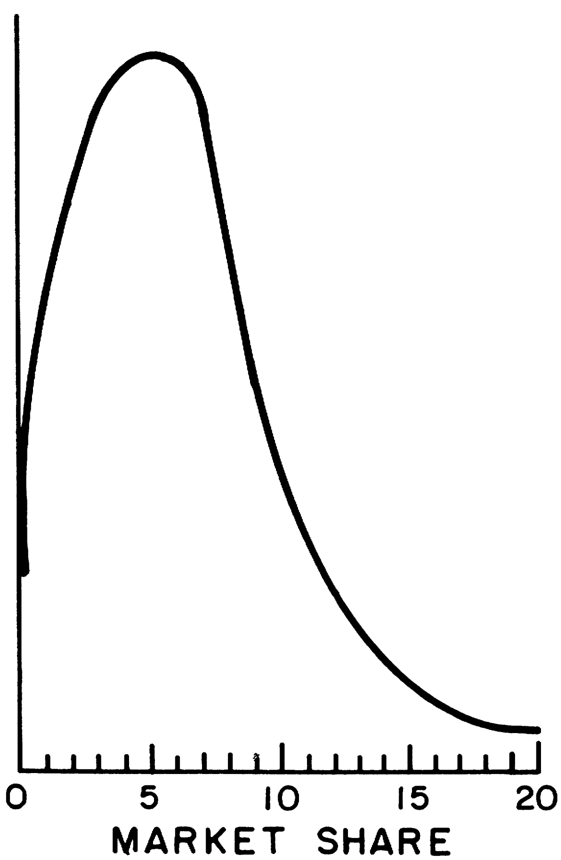
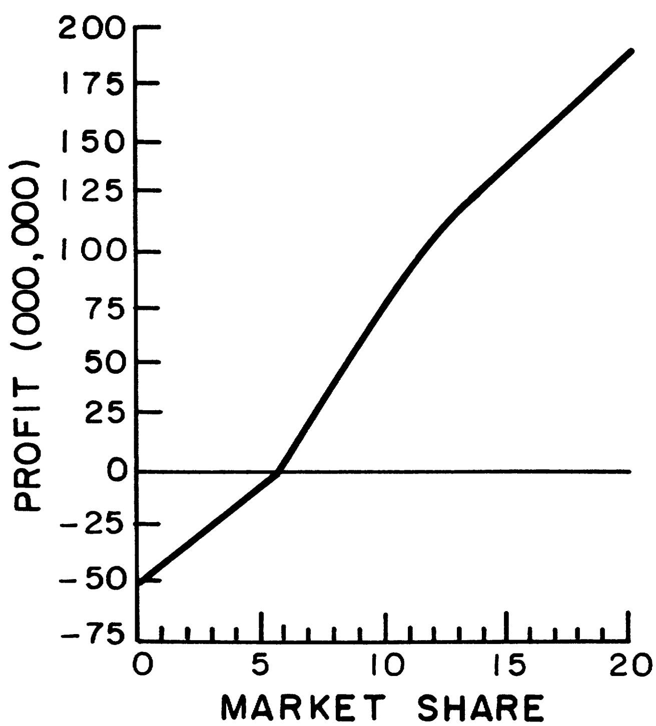


Figure 4
ILLUSTRATION OF TYPICAL REWARD FUNCTION U(r)



small in comparison with the gains because if the product fails it is usually removed from the market within the first year, but if it is a success the gains accrue over all the years in the product's life cycle.

We want to set T_1 and T_2 so that we eliminate products in the loss area without eliminating products that may generate the high rewards shown in Figure 4. Analytically, we want to maximize the product's expected profit (or more generally the expected utility) to the firm.[9] If we could assume the prior and conditional distributions to be of special conjugate forms that could be integrated to yield an expected utility function, and further that this function could be differentiated and solved for T_1 and T_2, an analytic optimal solution would be available. Unfortunately the natural conjugate distributions (Raiffa and Schlaifer 1961, p. 52–8) do not adequately represent the distributions that typify this problem (e.g. skewed prior and unimodal but non-normal error distribution). Even if they were available, the differentiation and solution to find the optimal T_1 and T_2 would be difficult because the utility function is nonlinear.

Numerical integration is used to find a best solution to the problem. We calculate the expected reward by:

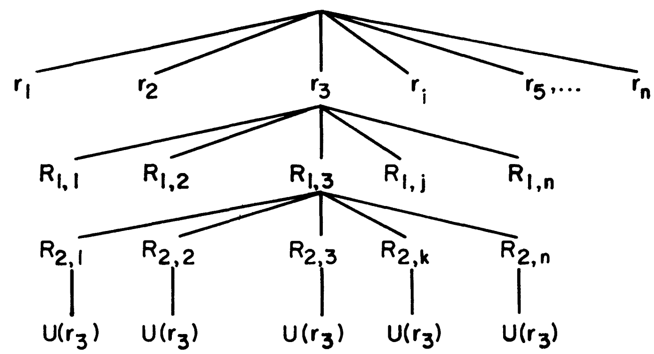
$$(5) \quad \mathrm{E}(\text{reward}|T_1, T_2) = \sum_{\substack{i,j,k \\ R_{1,j}\geq T_1 \\ R_{2,k}\geq T_2}} \mathrm{U}(r_i)\, \mathrm{P}(R_{2,k}|r_i)\, \mathrm{P}(R_{1,j}|r_i)\, \mathrm{P}(r_i) - C_t \sum_{\substack{i,j \\ R_{1,j}\geq T_1}} \mathrm{P}(R_{1,j}|r_i)\, \mathrm{P}(r_i) - C_p$$

where:

$\mathrm{U}(r_i)$ = utility associated with true share value r_i,
$\mathrm{P}(R_{1,j}|r_i)$ = probability of the j^{th} pre-test-market observation ($R_{1,j}$) given the true value of r_i,
$\mathrm{P}(R_{2,k}|r_i)$ = probability of the k^{th} test market observation ($R_{2,k}$) given the true value of r_i,
$\mathrm{P}(r_i)$ = prior probability of true share r_i,
C_t = cost of test market,
C_p = cost of pre-test-market evaluation,
T_2 = test-market cutoff, and
T_1 = pre-test-market cutoff.

A tree of all possible outcomes is generated for calculation of these probabilities (Figure 5). The prior distribution is described by discrete values (r_i) and the probability ($\mathrm{P}(r_i)$) for each of these prior true share values. Similarly, the error distributions are made discrete to produce $\epsilon_{1,j}$ and $\epsilon_{2,k}$ values and their associated probabilities $\mathrm{P}(\epsilon_{1,j})$ and $\mathrm{P}(\epsilon_{2,k})$. For each prior value ($r_i$) we calculate a set of pre-test-market observations $R_{1,j}$ by adding each possible error value ($\epsilon_{1,j}$) to the r_i value and calculating the conditional probability $\mathrm{P}(R_{1,j}|r_i)$. Test-market shares are produced similarly by incrementing each associated r_i by the possible error values ($\epsilon_{2,k}$) and calculating the conditional probability $\mathrm{P}(R_{2,k}|r_i)$. For each value of true share (r_i), the national reward is then calculated ($\mathrm{U}(r_i)$). The tree is usually large. For example, if the prior distribution has 40 points, the pre-test error distribution 20 points, and the test-market error distribution 10 points, 8000 end-points branches will be generated. Once the tree has been generated, the expected values in equation 5 can be calculated given T_1 and T_2 levels. Branches in which $R_{1,j} < T_1$ and $R_{2,k} < T_2$ are eliminated and the remaining probabilities used in equation 5.

Figure 5
DISCRETE TREE OF ALTERNATIVES



Note: r_i = true share value i
$R_{1,j} = r_3 + \varepsilon_{1,j}$ = conditional observed pre-test shares give true share r_3
$P(R_{1,j}|r_i) = P(\varepsilon_{1,j} = R_{1,j} - r_i)$
$R_{2,k} = r_3 + \varepsilon_{2,k}$ conditional observed test-market shares given true share r_3
$P(R_{2,j}|r_i) = P(\varepsilon_{2,k} = R_{2,k} - r_i)$
All branches are enumerated as shown for r_3

We also can use the tree to calculate the probability of (1) passing pre-test ($R_{1,j} \geq T_1$), (2) passing test market given that we passed pre-test ($R_{2,k} \geq T_2$ given $R_{1,j} \geq T_1$), and (3) the probability of national success ($r_i > \bar{r}$, where $\bar{r}$ is the breakdown value in $\mathrm{U}(r)$). Because we have enumerated all outcomes, we also can examine outcomes that were eliminated by the cutoff values. Specifically we can calculate the probability of a national success ($r_i > \bar{r}$) given that a product (1) failed pre-test ($R_{1,j} < T_1$), (2) failed test market after a successful pre-test ($R_{1,j} \geq T_1$, $R_{2,k} < T_2$), or (3) failed either test ($R_{1,j} < T_1$ or $R_{1,j} \geq T_1$, $R_{2,k} < T_2$).[10]

These are important managerial control variables and are meaningful measures of the two kinds of errors that can be made in the sequential decision system.

[9]Maximizing the expected reward is equivalent to minimizing the expected opportunity loss (Raiffa and Schlaiffer 1961, p. 83).

[10]Dynamic programming could be used to maximize more efficiently E(reward $|T_1$, T_2), equation 5, but it would not allow calculation of these probabilities.

Typical Product Case

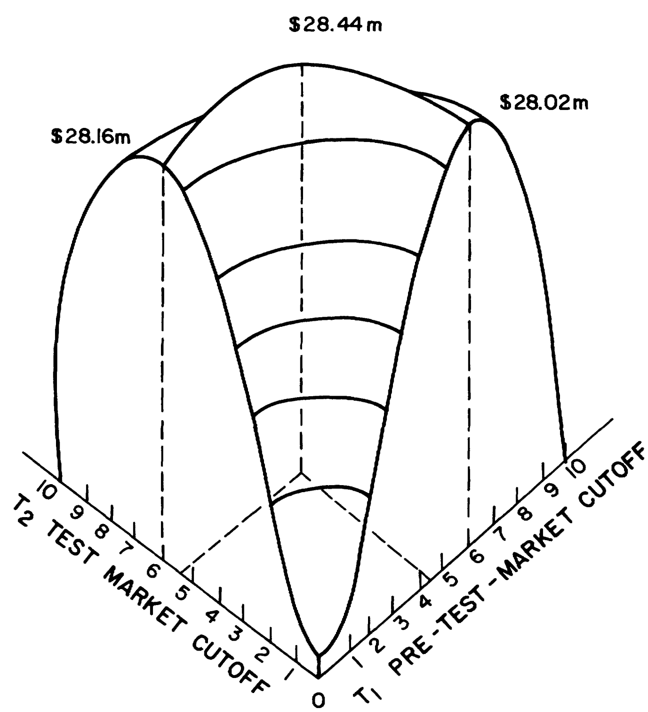
Now consider a typical launch of a major new product (e.g. toothpaste, deodorant, aerosol cleaner, or cereal). The purpose of this analysis is to determine whether a pre-test model contributes to the expected profit generated by the new product development process and whether there are any conditions under which a traditional test market may be bypassed. In this analysis we assume Figure 4 depicts the profit that would result with each level of national share accomplishment. The costs of test market are taken as \$1.5 million and the cost of pre-test as \$50,000. The distribution of pre-test results ($\epsilon_1(\mu_1, \sigma_1)$) is considered to be biased with $\mu = .6$ (see Table 1) and subject to a standard deviation of two share points ($\sigma_1 = 2.0$). This standard deviation is based on equation 4 with $\sigma_{x_1-x_2} = 2.0$. The pre-test error is calculated to be 1.8 and arbitrarily increased to 2.0 to allow for any possible underestimates of errors and to make this typical case a conservative evaluation of pre-testing.

Few published data are available to determine test-market accuracy, but in a comparison of first-year national share with test-market share for 50 new brands, Nielsen found 50% of the observations within 10% of the test-market shares (*The Nielsen Reporter* 1979). On the basis of the 7.16 average test-market share observed in our study (Table 1), this implies half the time share will be within .7 share point and, if we assume a normal distribution, the standard deviation of the error distribution is approximately one share point. Gold (1964) studied one source of error in test marketing—projecting test-area sales to national. He analyzed seven established products in six market areas commonly used for test marketing and projected sales by three common methods. His comparisons of store audit values in the test areas with national sales showed considerable error. In projections based on only one test city, the standard deviation was 32.8% of the mean. Based on two test cities the deviation was 21.7% and based on three cities the deviation was 15.6%. If we assume a mean share of 7.16 (see Table 1), the standard deviation would be 2.3 share points for one city, 1.6 for two cities, and 1.1 for three cities. The three-city value is consistent with the Nielsen data. However, Gold emphasizes that his study includes only part of the total error. He did not include effects such as competitive efforts to distort test markets, the extra attention the salesforce may lavish on the new product, or the interval of time between test market and national introduction. He believes these sources would increase the standard deviation by 50–75% (Gold 1964, p. 16). If this is true, the Nielsen error estimates would be optimistic. For our typical case, we chose the smaller test-market error estimates to prevent biasing the results toward pre-test-market models. We use a normal distribution with $\mu_2 = 0$ and $\sigma_2 = 1.0$ to approximate the test-market error distribution.

The final input for the model is the prior distribution of true share. This distribution can be derived from the results of all ASSESSOR pre-test evaluations. Over the 215 applications, the mean share is 7.13 and the standard deviation is 6.55 share points. The variance of the prior distribution can be calculated by deducting the variance of the pre-test error distribution under the assumption that they are independent random variables. The resulting prior distribution has a mean of 7.13 and a standard deviation of 6.23 share points and is skewed to the right. It is plotted in Figure 3.

The prior distribution was divided into 40 discrete values, the pre-test error distribution into 20 discrete values, and the test-market error distribution into 10 discrete values. Simulations were conducted for 130 combinations of T_1 and T_2. In each case the expected reward was calculated by the foregoing procedure. The resulting expected profit surface is depicted in Figure 6. The best combination of cutoffs is $T_1 = 4.5$ and $T_2 = 5.5$ and the resulting profit is \$28.44 million. With no test or pre-test ($T_1 = 0$ and $T_2 = 0$), the reward is \$16.74 million, so the value of testing is \$11.7 million for each product which enters the full sequential decision system. If only a pre-test is done (i.e., $T_2 = 0$), $T_1 = 6.0$ is the optimal pre-test cutoff. The total reward is \$28.02 million in this case and the value of pre-testing is \$11.28 million. If only test market is done (i.e., $T_1 = 0$), $T_2 = 6.0$ is the optimal pre-test cutoff. The total reward is \$28.16 million in this case and the value of testing is \$11.42 million. As a benchmark, we calculated the value

Figure 6
EXPECTED REWARD SURFACE



of perfect information (a perfectly accurate test, i.e. no error) to be $13.58 million. We conclude that either test can contribute most of the value of testing. However, the incremental value of the test market given that a pre-test is done is $420,000 ($11.7 − $11.28 million) and the incremental value of the pre-test given that a test market is done is $280,000 ($11.7 − $11.42 million). Both pre-test and test market are worthwhile and are valuable components in this simulation of a new product development system.

Table 4 displays the implications of various cutoff policies and the associated probabilities of passing the tests, achieving national success, and eliminating brands which would have been successful. With only a test market, the probability of passing test market is 47%, the probability of national success (greater than breakeven) is 94% given the product passes the test-market screen, and the probability of eliminating a successful product is 10%. If both test and pre-test are done, the expected reward increases by $280,000, the probability of passing pre-test market is 69%, and the probability of passing test market increases to 72%. The probability of national success decreases 3 percentage points to 91%, but this is compensated by a 1 percentage point reduction in the chances of eliminating a success before national launch. Table 4 also shows the results of setting very high or very low cutoffs. In both cases the expected reward decreases substantially. High cutoffs virtually eliminate national failures but result in a 37% probability of eliminating successful products before introduction. Low cutoffs remove the risk of eliminating successful products, but result in a 46% chance of national failure for products taken to the market. The setting of the cutoffs demonstrates how the sequential testing system can be used to balance the risks of the two types of errors—carrying failures to the market and eliminating opportunities for success.

Effect of Delay Due to Testing

In reviewing the base case one notes the small incremental gain for test marketing given that a pre-test-market analysis has been done ($420,000). One could ask whether this value of information compensates for the 12 months or more of incremental delay in the introduction of the product and the opportunity competitors have to "read" the test-market results. Similarly one could ask why a pre-test-market analysis should be done if the incremental value is low given that a test market must be done and the pre-test will delay launch three months.

We can analyze this situation by reducing the reward function to reflect profit losses due to delay. Recall the no-delay reward occasioned by no testing is $16.7 million (Table 4, first column). We simulate a case of (1) 30% profit reduction from the gains in Figure 4 for a 12-month test-market delay to get a maximum expected reward of $19.2 million, (2) 10% reduction for a three-month delay to get a maximum $25.2 million reward, and (3) 40% reduction for a 15-month delay for both tests to get $16.6 million as a maximum expected reward. The expected value is highest for a three-month delay with pre-testing alone ($25.2 million) and has a value of information of $8.5 million (25.2 − 16.7). Although the test market causes a 12-month delay it also has a positive value of information of $2.5 million (19.2 − 16.7), but this is $6.0 million lower than the value of just the pre-test analysis. Under the delay penalties both pre-test and test market have negative incremental values if they are done in addition to the other testing procedure. This particular simulation indicates that if delay penalties are large, only one test should be done and the one that maximizes expected profit is a pre-test-market analysis.

Sensitivity to Changes in Reward Function

Now we return to the case of no delay penalties due to the time of testing and examine the effects of alternative reward functions on the expected reward, optimal cutoffs, probabilities, and value of information (Table 5).

First, consider the case in which the new products manager is making the GO/NO decision and applies his or her personal utility evaluations to the share outcomes.

Table 4
TYPICAL PRODUCT RESULTS

	No test	*Test market only*	*Test and pre-test*	*Pre-test only*	*High cutoff*	*Low cutoff*
Expected profit	$16.74m	$28.16m	$28.44m	$28.02m	$21.46m	$19.23m
T_1 (pre-test cutoff)	—	—	4.5	6.0	9.0	1.0
T_2 (test-market cutoff)	—	6.0	5.5	—	9.0	1.0
Probability of passing ASSESSOR (%)	—	—	69	54	28	94
Probability of passing test market (%)	—	47	72	—	70	96
Probability of national success if go national (%)	49	94	91	82	99.9	54
Probability of national success if fail ASSESSOR (%)	—	—	5	11	31	.01
Probability of national success if fail test market (%)	—	10	5	—	10	.01
Probability of national success if fail either ASSESSOR or test market	—	10	9	11	37	.01

Table 5
CHANGES IN REWARD FUNCTIONS

	Base: typical	*Case 1: personal utility*	*Case 2: smaller gain/loss*	*Case 3: low entry cost*	*Case 4: low gain*	*Case 5: shift*
Expected profit	$28.44m	$13.54m	$14.03m	$28.73m	$16.58m	$37.17
T_1 (pre-test cutoff)	4.5	6	6.5	6	5.5	3
T_2 (test-market cutoff)	5.5	6	—	—	5.5	4.5
Probability of passing ASSESSOR (%)	69	54	49	54	59	82
Probability of passing test market (%)	72	69	—	—	79	74
Probability of national success if go national (%)	91	95	86	82	93	92
Probability of national success if fail either ASSESSOR or test market (%)	9	13	14	11	11	11
Value of testing ($000)[a]	11,700	25,540	5,660	5,200	11,970	6,710
Value of pre-test-market only ($000's)	11,280	25,070	5,660	5,200	11,860	6,390
Value of pre-test-market if test market must be done ($000)[b]	280	570	380	300	380	140
Value of test market if pre-test must be done ($000)[c]	420	470	−320	−150	110	320

[a]Expected reward of best testing policy less reward of no testing.
[b]Expected reward of best testing policy with test market and pre-test less reward with only test market.
[c]Expected reward of best testing policy with test market and pre-test less reward with only pre-test.

The manager may be very averse to the risk of failure and feel that if the brand succeeds much of the recognition for its success will go to the established brand group over the course of the product's life cycle (see Keeney and Raiffa 1976 for a complete discussion of utility assessment procedures). We simulate this situation by doubling the monetary losses and halving the monetary gains of the reward function in Figure 4. Table 5 shows the results of this "personal utility" reward function. The expected reward is lowered by more than 50%, the cutoffs are raised to improve the probability of national success (95% versus 91% for the base case), and greater risk of eliminating a success is accepted (13% versus 9%). With the tighter screening criteria, fewer products pass pre-test and test market. The expected reward with no testing is −$12.0 million and the total value of testing is now $25.54 million (13.54 minus −12). The total value of information for pre-test-market is more than doubled in relation to the "typical" case ($25.07 million versus $11.28 million). The incremental value of information for pre-test given that a test market is done also doubles and the incremental value of test market given that a pre-test-market analysis is done increases by 12%.

Now consider another variation from the base case. If a market is small in total sales volume, the monetary gains and losses associated with each share will be lower. We simulate this situation by multiplying the entire reward function by one half (see case 2 in Table 5). The best sequential testing procedure now does not include test market because its added information is not worth the $1.5 million test-market cost. The best reward with test and pre-test is $13.71 million whereas the reward with pre-test only is $14.03 million. Even though the cutoff for pre-test is raised from its 4.5 base level to 6.5, more risk of error is accepted than in the "typical" case because the potential losses associated with failure are so much smaller (86% probability of national success instead of 91%).

A similar result is obtained when we simulate a 50% reduction in losses while leaving the gains at the base level (case 3). This situation could arise if the cost of entry (advertising, promotion, and production) were low. The incremental value of test market is negative (−$150,000) because the exposure to loss is low, and therefore the most efficient manner in which to manage such a brand is with a pre-test-market analysis only.

Next we simulate the effects of lowering the gains by 40% while leaving the losses at the base level (case 4). Both pre-test and test market are recommended with pre-test cutoff higher than in the "typical" case and a test cutoff equal to the "typical" case. The probability of passing pre-test is lower but the probability of test success and national success is higher with the tighter screens. There is also more chance of eliminating a success.

The final simulation reflects the case in which margins are very high and therefore the reward is higher at all share levels (case 5). We simulate this situation by shifting the reward function to the left by one share point (i.e., new reward value corresponds to old share minus one). Because the reward is very high, lower cutoffs may be used to achieve given probabilities of national success. The probability of passing pre-test is higher than in the "typical" case because so many profitable brands are present in the population (see prior distribution in Figure 3, with a breakeven share of 4.75 instead of 5.75). The chances of early elimination of a success are more likely, but can be tolerated.

The six cases show that the best testing procedure is sensitive to the reward function. If losses are relatively high, tighter testing criteria are recommended. If losses are relatively low, pre-test-market analysis may suffice. In most cases, pre-test-market and test-market work are used together to balance the risk of national failure and the chance of eliminating a potentially successful product before introduction.

Sensitivity of Testing Accuracy

The preceding analyses were based on the standard deviation of pre-test-market error of two share points. To examine the sensitivity of the testing system to alternate levels of pre-test accuracy, we repeated the evaluation using a small standard error of one share point. This would be possible if a large sample size ($n = 600$) were used and the firm were sure it would execute its marketing plan to achieve the estimated levels of awareness and distribution. In this case of improved pre-test accuracy, the maximum expected reward increases \$1.0 million and test market is not required. A higher cutoff (6.0) is used, so fewer products move past the pre-test stage, but a 92% chance of national success rate is observed. The probability of eliminating a success is 10%. This simulation shows that the pre-test alone does well in cases where sample size is large and adjustments are not required. This situation may be relatively rare, but it implies that the more accurate the pre-test, the more valuable it is and the more it can be relied upon as a screening procedure.

The "typical product" base case reflected an assumption of high test-market accuracy ($\mu_2 = 0$, $\sigma_2 = 1.0$). Work by Gold (1964) suggests that larger errors may be expected. We simulated cases with a mean error (μ_2) of .5 share point and increasing standard deviations for the distribution of differences between test-market and true share ($\sigma_2 = 1.25$, 1.5, and 2.0). As the test-market error increases, the total expected reward decreases, the incremental value of the pre-test-market analysis increases, the incremental value of the test market decreases, and the risk of national failure increases. At a standard deviation of 2.0, the test market no longer adds to the expected reward. The value of reducing the standard deviation can be calculated by the differences in the expected rewards. If the test-market standard deviation were 2.0, reducing it to 1.0 would be worth \$330,000.

Managerial Implications

The preceding simulations indicate that in one "typical" case, the expected value of adding a pre-test-market model analysis is positive. However, the output of the sequential decision model cannot be generalized because specific outcomes will depend on the individual firm's cost function, margins, investment policy, knowledge of the awareness it can generate, ability to achieve the targeted distribution, test market design, specific category prior distribution, and competitive practices. The model is a tool for managers. If they can estimate the appropriate prior distribution (e.g., Figure 3), reward function (e.g., Figure 4), pre-test accuracy, and test-market accuracy for their particular product and category, they can use the sequential decision model to set the best cutoffs and indicate which testing procedures should be used. The prior distribution reported here is based on 200 products. When this distribution is supplemented by a company's own new product experience, the required prior input can be generated. The reward function can be calculated on the basis of introductory cost and profit margin data. The test-market and pre-test-market accuracy levels we report, updated by a company's experience, complete the required input. After simulations, managers can examine the probabilities of passing each test, national success, and elimination of potentially successful products, as well as the expected reward, to evaluate the desirability of various test policies. Allowing the manager to change the probabilities by modifying the cutoffs is an alternative to direct assessment of the manager's utility function and the maximization of his or her expected utility (see Hauser and Urban 1979 for a discussion of some of the difficulties of utility assessment). The model gives decision makers a capability to manage the risks of new product development.

Decision makers must be careful in managing the enthusiasm for particular new products. Brand managers personally may have much to gain and little to lose by pushing their product despite results that are below the cutoff values. The usual persuasion methods are to propose changes that will be made to improve the product, identify unrealistic conditions in the test market (e.g., competitor's high sampling and promotion), or argue against the validity of the forecasting models. Though these responses sometimes may be legitimate, care must be exercised not to drop the screening criteria without careful reanalysis. For example, if the test market cutoffs in the base case were relaxed by one market share point, the probability of national success would drop from 91% to 83%. These errors would be compounded if the test market were biased and if this bias were not known or reflected in the cutoff rates. For example, with a one-half share point bias and moderate accuracy in test market ($\sigma = 1.5$), a relaxation of the test market cutoff by 1.5 share points will lower the probability of national success from 91% to 79%. This rate of approximately 80% national success strikes some managers as much closer to their experience. The implication may be that they have not enforced the correct screening criteria with enough tenacity.

Although the model must be reapplied in each firm, the simulations indicate there are some conditions under which test market can be eliminated by a pre-test-market model-based evaluation. A test market may not be required in the best testing system if (1) there is a large penalty for reaching the market 12 months late, (2) entry costs are low and losses at lower share levels are small, (3) a small market is to be entered and potential gains and losses are both small, (4) pre-test-market sample size is large, accurate awareness and distribution estimates can be made, and the firm's marketing plan will be faithfully executed.[11] Condition 3 is commonly observed in international markets. For example, in Europe, markets tend to be small and the associated rewards are small in relation to the fixed cost of test marketing. This is compounded by the difficulty of finding a representative test market city and controlling the marketing execution. We observe much less test marketing in Europe than in the U.S.A.

Our analysis indicates that conditions can exist under which bypassing test market could be considered, but a number of caveats must be emphasized. We have considered only the forecasting function of test market. Test markets serve other valuable functions. They assure the firm that it can produce and distribute a product in substantial volumes, provide opportunities to improve the products, and develop evidence to convince retailers and distributors to carry the product in the event of a national launch. If the firm is not sure it can make the product in large volumes at its specified level of quality, the estimated reward obviously will not occur. The firm must be sure it can produce the product within the design tolerances; if not, a test market may be required in any case. The test market can affect the reward function by improvement of the product and its marketing based on in-market experimentation and model analysis (see Urban and Hauser 1980). In this case, test marketing may be advisable because it increases the expected reward. An example might be found in case 2 of Table 5, where if the test market improved the brand's profit by more than $320,000, a test market should be included in the sequential testing system. The advent of pre-test-market models implies that test market designs should allocate increasing emphasis to finding the best marketing mix for a product rather than to sales forecasting alone. We believe in all but exceptional cases pre-test models and test markets should be used together to reduce the risk and improve the profitability of new product development. In this new system, test marketing should be aimed at maximizing the market mix elements (price, advertising, distribution, sampling, special displays) as well as forecasting sales.

[11]In addition, when the system is operating, a particular product may achieve a very high pre-test result. This would indicate a substantial probability that this observed share (R_1) was generated by a high true share (r). This posterior distribution ($f(r|R_1)$) can be calculated by Bayesian procedures and a decision analysis with the proposed model based on the remaining testing step may indicate that the best action is to bypass test market. We have analyzed a fixed screening system where T_1 and T_2 are set in advance. The more general case is to set T_2 after observing the particular pre-test result. The fixed screening system reflects a lower limit on the value of pre-testing. If the test-market cutoff is set optimally after each pre-test result, the value of information from pre-test will be higher than shown in this article. If the pre-test provides diagnostics useful in improving the product, this would also increase the value of the procedure.

CONCLUSION

Our validation study of the ASSESSOR pre-test-market system indicates good accuracy in predicting test market shares. This accuracy improves when adjustments are made for differences between pre-test and test-market conditions. Such a model is very useful in new product development. In a typical case application with the appropriate cutoff rules, it generated a substantial positive expected reward and reduced the risks of failure in test market. Although any outcome will depend on a particular firm's costs and utility function, it is our opinion that in most cases a pre-test-market analysis will have a positive value of information. The observed accuracies and the sequential decision theory model can be valuable tools in managing the dual risks of introducing a new product that fails and of not introducing a product that would have been a success.

In the future, more validation data will accrue and this analysis will be repeated for various categories of products. Future research will be directed at modeling the complete sequential decision process including concept, product, and advertising copy testing. Such a model would allow the screening criteria to be set to maximize expected rewards and allow analysis of the value of spending more funds to increase the accuracy at each testing phase. A final research direction is to apply the underlying concepts of laboratory measurement and modeling to the forecasting of consumer durables where risks are higher and test marketing is usually not done. For durable goods the cost of setting up the production line to produce in test-market quantities is great (e.g., millions of dollars), and after this investment is made the incremental risk in national launch is low. In terms of the preceding simulations, the incremental loss is low and the gain is high (case 3, Table 5), so the value of information from a test market is usually not worth its cost. However, any method of accurately predicting sales before the investment in production facilities would be extremely valuable. Pre-market models and measurement methods may provide the potential to improve the efficiency and effectivensss of new product development in other than the frequently purchased consumer product field.

REFERENCES

Burger, P., C. H. Gundee, and R. Lavidge (1981), "COMP: A Comprehensive System for the Evaluation of New Products," in *New Product Forecasting*, Y. Wind, V. Mahajan, and R. N. Cardozo, eds. Lexington, MA: Lexington Books, 269–84.

Business Week (1973), "The Rebuilding Job at General Foods," No. 2294 (August 25), 48–55.

Buzzell, R. D. and R. E. M. Nourse (1967), *Product Innovation in Food Processing 1954–1964*. Boston: Division of Research, Graduate School of Business Administration, Harvard University.

De Groot, M. H. (1970), *Optimal Statistical Decisions*. New York: McGraw-Hill Book Company.

Eskin, G. J. and J. Malec (1976), "A Model for Estimating Sales Potential Prior to the Test Market," *Proceedings 1976 Fall Educators' Conference*. Chicago: American Marketing Association, 230–2.

Gold, J. A. (1964), "Testing Test Market Predictions," *Journal of Marketing Research*, 1 (August), 8–16.

Goodman, L. A. (1960), "On the Exact Variance of Products," *American Statistical Association Journal*, 55 (December), 708–13.

Hauser, J. R. and G. L. Urban (1979), "Assessment of Attribute Importance and Consumer Utility Functions: von Newman-Morgenstern Theory Applied to Consumer Behavior," *Journal of Consumer Research*, 25 (March), 251–62.

——— and ——— (1982) "Prelaunch Forecasting of New Consumer Durables: Ideas on a Consumer Value–Priority Model," working paper, Alfred P. Sloan School of Management, MIT.

Kalwani, M. U. and A. J. Silk (1980), "Structure of Repeat Buying for New Packaged Goods," *Journal of Marketing Research*, 17 (August), 316–22.

Keeney, R. L. and H. Raiffa (1976), *Decision Analysis with Multiple Objectives*. New York: John Wiley & Sons, Inc.

——— (1979), "New Product Success Ratios 1977," No. 1.

Pringle, L. G., R. D. Wilson, and E. I. Brody (1982), "NEWS: A Decision Oriented Model for New Product Analysis and Forecasting," *Marketing Science*, 1 (Winter), 1–30.

Raiffa, H. and R. Schlaifer (1961), *Applied Statistical Decision Theory*. Boston: Division of Research, Graduate School of Business Administration, Harvard University.

Rao, V. R. and F. W. Winter (1981), "A Bayesian Approach to Test Market Selection," *Management Science*, 12 (December), 1351–68.

Shoemaker, R. and R. Staelin (1976), "The Effects of Sampling Variation on Sales Forecasts for New Consumer Products," *Journal of Marketing Research*, 13 (May), 138–43.

Silk, A. J. and G. L. Urban (1978), "Pre-Test-Market Evaluation of New Packaged Goods: A Model and Measurement Methodology," *Journal of Marketing Research*, 15 (May), 171–91.

Tauber, E. M. (1977), "Forecasting Sales Prior to Test Market," *Journal of Marketing*, 41 (January), 80–4.

The Nielsen Reporter (1971), "New Product Success Ratios," No. 5.

Urban, G. L. (1968), "A New Product Analysis and Decision Model," *Management Science*, 14 (April), 501.

——— and J. R. Hauser (1980), *Decision and Marketing of New Products*. Englewood Cliffs, NJ: Prentice-Hall Inc., 386–417.

Yankelovich, Skelly, and White, Inc. (1981), "LTM Estimating Procedures," in *New Product Forecasting*, Y. Wind, V. Mahajan, and R. N. Cardozo, eds. Lexington, MA: Lexington Books, 249–68.

LINKED CITATIONS

[Footnotes]

[4] **The Effects of Sampling Variation on Sales Forecasts for New Consumer Products**
Robert Shoemaker; Richard Staelin
Journal of Marketing Research, Vol. 13, No. 2. (May, 1976), pp. 138-143.
Stable URL:
http://links.jstor.org/sici?sici=0022-2437%28197605%2913%3A2%3C138%3ATEOSVO%3E2.0.CO%3B2-Q

References

Testing Test Market Predictions
Jack A. Gold
Journal of Marketing Research, Vol. 1, No. 3. (Aug., 1964), pp. 8-16.
Stable URL:
http://links.jstor.org/sici?sici=0022-2437%28196408%291%3A3%3C8%3ATTMP%3E2.0.CO%3B2-E

Assessment of Attribute Importances and Consumer Utility Functions: Von Neumann-Morgenstern Theory Applied to Consumer Behavior
John R. Hauser; Glen L. Urban
The Journal of Consumer Research, Vol. 5, No. 4. (Mar., 1979), pp. 251-262.
Stable URL:
http://links.jstor.org/sici?sici=0093-5301%28197903%295%3A4%3C251%3AAOAIAC%3E2.0.CO%3B2-G

LINKED CITATIONS

Structure of Repeat Buying for New Packaged Goods
Manohar U. Kalwani; Alvin J. Silk
Journal of Marketing Research, Vol. 17, No. 3. (Aug., 1980), pp. 316-322.
Stable URL:
http://links.jstor.org/sici?sici=0022-2437%28198008%2917%3A3%3C316%3ASORBFN%3E2.0.CO%3B2-%23

News: A Decision-Oriented Model for New Product Analysis and Forecasting
Lewis G. Pringle; R. Dale Wilson; Edward I. Brody
Marketing Science, Vol. 1, No. 1. (Winter, 1982), pp. 1-29.
Stable URL:
http://links.jstor.org/sici?sici=0732-2399%28198224%291%3A1%3C1%3ANADMFN%3E2.0.CO%3B2-W

A Bayesian Approach to Test Market Selection
Vithala R. Rao; Frederick W. Winter
Management Science, Vol. 27, No. 12. (Dec., 1981), pp. 1351-1369.
Stable URL:
http://links.jstor.org/sici?sici=0025-1909%28198112%2927%3A12%3C1351%3AABATTM%3E2.0.CO%3B2-%23

The Effects of Sampling Variation on Sales Forecasts for New Consumer Products
Robert Shoemaker; Richard Staelin
Journal of Marketing Research, Vol. 13, No. 2. (May, 1976), pp. 138-143.
Stable URL:
http://links.jstor.org/sici?sici=0022-2437%28197605%2913%3A2%3C138%3ATEOSVO%3E2.0.CO%3B2-Q

Pre-Test-Market Evaluation of New Packaged Goods: A Model and Measurement Methodology
Alvin J. Silk; Glen L. Urban
Journal of Marketing Research, Vol. 15, No. 2. (May, 1978), pp. 171-191.
Stable URL:
http://links.jstor.org/sici?sici=0022-2437%28197805%2915%3A2%3C171%3APEONPG%3E2.0.CO%3B2-P

Forecasting Sales Prior to Test Market
Edward M. Tauber
Journal of Marketing, Vol. 41, No. 1. (Jan., 1977), pp. 80-84.
Stable URL:
http://links.jstor.org/sici?sici=0022-2429%28197701%2941%3A1%3C80%3AFSPTTM%3E2.0.CO%3B2-2

NOTE: *The reference numbering from the original has been maintained in this citation list.*

LINKED CITATIONS

A New Product Analysis and Decision Model
Glen L. Urban; Benjamin Lipstein
Management Science, Vol. 14, No. 8, Application Series. (Apr., 1968), pp. B490-B519.
Stable URL:
http://links.jstor.org/sici?sici=0025-1909%28196804%2914%3A8%3CB490%3AANPAAD%3E2.0.CO%3B2-5

NOTE: *The reference numbering from the original has been maintained in this citation list.*