Metrics That Matter

Identifying the Importance of Consumer Wants and Needs

MARTIN R. LAUTMAN
Marketing Channels, Inc.
martin.lautman@yahoo.com

KOEN PAUWELS
Tuck School of Business
Dartmouth College
Koen.H.Pauwels@Tuck.Dartmouth.edu

How can the importance of consumer wants and needs be quantified? Using data sets from multiple consumer brand and advertising tracking studies, several standard traditional market research techniques are compared to vector autoregression (VAR) modeling. It is demonstrated that by utilizing VAR models and resolving causal ambiguity, key performance indicators can be identified that not only correlate with traditional market research summary metrics, such as overall ratings and purchase interest, but that also drive brand sales/share and thereby qualify as metrics that matter. The analytic philosophy underlying the VAR analytic approach also is shown to be consistent with (and complementary to) market mix modeling analysis. Presented is a procedure for the simultaneous assessment of the relative and absolute impact of multiple marketing initiatives on baseline and incremental sales—including advertising and promotion effects and traditional consumer awareness and attitudinal metrics—facilitating resource-allocation decisions and providing marketers within a single framework for return on marketing investment optimization.

INTRODUCTION

The efficient allocation of marketing and promotion resources frequently is dependent on the ability of an organization to answer the deceptively simple retrospective question, "Why did customers, noncustomers, or prospects act as they did?" or, prospectively, "Why did customers, noncustomers, or prospects indicate that they are intending to take some future action?"

Market research practice provides a well-accepted forum for addressing what might be termed the "what" question that underlies the "why" question; that is, within the relevant competitive set and target universe, "What is or was important to customers, noncustomers, and prospects at the point in time when the decision was made?" The logic of typical market research protocols is that by revealing "what" was important to the consumer then the "why" question can be readily answered. Given that analytic framework, it is not surprising that many companies expend significant resources to identify and monitor the "what's" in the form of appropriate attributes, benefits, and claims—concisely termed ABCs—for relevant brands and services (Lautman, 1993).

When ABCs are perceived to be reflective of consumer wants and needs related to their choice behavior, they often are referred to as key performance indicators (KPIs). Unlike the approach taken with market mix modeling, however, KPIs identified using traditional market research importance-assessment techniques tend to be treated as "drivers" of in-market sales and/or share without having first undergone the rigorous analytic steps necessary to substantiate a causal claim.

We will begin with a review of the strengths and weaknesses of the currently popular research

DOI: 10.2501/S002184990909045X

It is demonstrated that by utilizing VAR models and resolving causal ambiguity, key performance indicators (KPIs) can be identified that not only correlate with traditional market research summary metrics such as overall ratings and purchase interest, but that also drive brand sales/share and thereby qualify as metrics that matter.

methodologies for assessing "importance." This exposition will be followed by the description of a relatively new application of the well-known vector autoregression (VAR) econometric model. When applied to tracking studies and actual sales data within a comprehensive brand health framework, a VAR analysis will be shown to be a valuable research tool for marketing, able to avoid some of the technical challenges and interpretive pitfalls inherent in traditional quantitative market research importance-assessment methodologies. Examples from multiple brand health tracking studies will demonstrate the broad application of this brand steering technology.

TRADITIONAL RESEARCH PARADIGMS FOR ASSESSING "IMPORTANCE"

The review presented here of classical market research analytic techniques for identifying "what is important?" will be limited to verbal quantitative methodologies. We recognize that important recent research efforts have focused on developing and implementing such quasiverbal techniques as picture sorts and metaphor elicitation (Zaltman, 2003). Nonverbal procedures also have included the monitoring of cognitive and emotional responses using brain waves, heart rate, and galvanic skin responses and discretely observing and/or filming consumer behavior at the "moment of truth" at store shelves—either actual or simulated. [Editor's note: please see "Winning the Super 'Buzz' Bowl," p. 293.]

Our experience suggests that these qualitative techniques, while valuable, currently account for only a small proportion of research expenditures. Qualitative research tools, such as in-depth interviews, triads, focus groups, and ethnographic research, traditionally are considered prequantitative and hypothesis-generating techniques, with their findings generally recognized as not projectable to general populations without subsequent quantitative validation from a representative sample of the relevant universe.

There are four widely used quantitative research designs for empirically deriving answers to "why" from "what" questions. Each will be briefly reviewed, profiling their strengths and weaknesses. These four approaches for identifying what is important to consumers can be clustered under two general rubrics: direct and indirect (also termed modeled or derived) assessment methodologies.

DIRECT METHODOLOGIES

The most straightforward methodology for determining what is important to consumers is simply to ask the "why" question. This is the protocol typically utilized in qualitative research, where consumers are asked to explain and/or provide a rationale for the attitudes, opinions, or beliefs that they express or the behaviors that they manifest. A parallel approach exists in many quantitative surveys, where consumers are asked in a follow-up to a closed end question to explain "why" they responded as they did. Importance is determined on the basis of "intuitive expert content analysis" typically exercised by experienced market researchers who code and aggregate the data.

Conclusions are drawn based on the frequency, sequence, and/or pattern of responses. These efforts produce the well-known "open-ended" summary response table. It is generally accepted that the earlier (more "top-of-mind") and/or the more frequently a "reason why" is offered—either by a single individual or by a group—the greater is its importance in consumer decision making. Data can be analyzed at the individual or aggregate level, with the typical scenario being responses aggregated up to the segment or population level.

The primary criticism of this intuitively appealing approach is that this methodology—even when the data are collected anonymously—suffers from the potential inability (or lack of desire with socially sensitive issues) of some individuals to communicate openly and accurately what is important to them. This challenge can be of particular concern with self-administered surveys, such as those conducted on the internet, where respondents do not have live interviewers probing for specifics and clarity. In fact, consumers simply may not be able to introspectively intuit an answer to the question "why."

Market research practice provides a well-accepted forum for addressing what might be termed the "what" question that underlies the "why" question.

Another limitation of this direct assessment method is that the accurate portrayal of the results must rely on the skill of experts not only to faithfully and with great fidelity record consumer responses (if consumers do not write/type/dictate the responses themselves), but also to code and aggregate the data in an efficient and unbiased manner. Biases often are the byproduct of the human proclivity to construct patterns by processing and interpreting information in a manner consistent with their historical knowledge, predispositions at the time, and perceptual selectivity (mental maps). Thus, careful controls in code development and multirater scoring are prescribed to mitigate the risks of inconsistencies and inaccuracies in the summarization process.

The second direct-response methodology employs ratings, rankings, and/or check-offs: consumers are asked to identify the ABCs that are more important in determining their preferences or choices. Decision tasks can be either constrained or unconstrained in that consumers are either forced (with Q-sorts and constant sum scales, for instance) or not forced (with magnitude estimation and Likert scales, for instance), respectively, to prioritize ABCs. Although sometimes only one product or service is evaluated, this approach most often is utilized to gain an understanding of what is important to the consumer in choosing and/or developing preferences from among a competitive set of alternative products or services in a clearly defined universe.

Market researchers utilizing this second direct approach typically believe they already know the important reasons underlying consumer preferences or choices. Their purpose now is to understand the magnitude of their priority. Not surprisingly, like the first direct assessment method, this approach also can suffer from the potential biases of respondent unwillingness or inability to accurately report what is important to them.

INDIRECT METHODOLOGIES

Two popular indirect measurement and analytic methodologies are designed to reduce the potential social and self-reflective introspective response biases inherent in direct questioning approaches. In both methods, consumers are not asked directly to identify what or how much various ABCs associated with an item contribute to its preference level. Rather, importance is determined by utilizing statistical analyses. Since the importance of ABCs tends to be established across competitive space—not for any single product or service—what is determined to be important is assumed to represent the entire category of interest and all the items in the set.

In one approach, consumers are asked to participate in a relatively simple attribution task: identify to what extent a product or service possesses (or might possess) a particular attribute, deliver an identified benefit, or be associated with a specific claim. Typically, the analytic protocol followed consists of a methodology often attributed to the PERCEPTOR/ASSESSOR system (Hauser and Urban, 1977; Silk and Urban, 1978; Urban, 1975; Urban and Hauser, 1993). The popularity of this method in common research practice is such that it is rare to find a positioning or market segmentation study that does not employ a PERCEPTOR/ASSESSOR-type indirect analysis for importance assessment.

Multiple consumer ratings on relevant ABCs for a competitive set of products or services are aggregated across a sample of consumers and then statistically related to a positive dispositional criterion metric, such as overall favorability, overall preference, overall liking, or purchase intent. The importance of each ABC in "driving" behavior is determined based on the strength of its association with the overall criterion measure as determined through a simple cross-tabulation (Myers and Chay, undated) or, more commonly, through statistical analyses focused on quantifying systematic variance.

The absence of variance between alternative products or services on an ABC—by definition and design—can lead to the conclusion that the ABC is not a "driver" and, therefore, not important to consumers in establishing preferences or making choices between alternatives. By focusing on variance, the importance of an ABC is defined by differentiation between the rated items. Due to the need for sufficient variance, aggregate-level analyses primarily are employed. Quantifying how many people find a specific ABC to be important is not the focus of indirect assessments, as it can be in a direct analysis.

While clearly possessing some obvious advantages over direct methods, this second indirect assessment method suffers from two important limitations: First, when using traditional linear regression, the analytic approach is compensatory, meaning that the strengths of brands on some ABCs would be expected to offset weaknesses on others. Yet, without a potential product or service successfully achieving (if not exceeding) a noncompensatory threshold value ("ante") on a critical ABC, a consumer is unlikely to prefer that option. This performance hurdle is critical in product or service selection and can be termed the point of entry (POE) challenge.

A second concern with this indirect method is that, since establishing preference depends on the existence of variance, only the differentiation between options on ABCs—that is, uniqueness—has the potential to create "importance." Attribute levels achieved and benefits and claims believed and experienced that are valued equally or nearly equally by consumers for all alternatives would not be designated as important. Furthermore, an analysis attributing little or no importance to an ABC could occur regardless of whether the product or service options in the competitive set were all rated either as successfully delivering or as severely lacking acceptable performance on that ABC. In fact, in common research practice, insufficient differentiation often exists between products or services on intuitively critical attributes, such as emotionally-oriented claims, for them to emerge as "important." A related difficulty can be observed when an ABC is deemed important by a relatively few consumers (possibly, either in terms of existing as a market segment or as a function of early adoption) and does not emerge as important in the traditional aggregate analysis. All of these situations can reflect restriction of range issues and highlight what might be termed the point of differentiation (POD) challenge.

The second popular indirect importance analysis method includes tasks that ask a consumer to demonstrate their preferences by making choices between alternatives (Green and Rao, 1971; Green and Srinivasan, 1990). Unlike the first indirect method described above, no assessment of attribution, appropriateness, or belief of the ABCs for the choice options is required.

Operationally, this fourth method often takes the form of a research study design built around a conjoint or discrete choice task. The research rationale applied is that by requiring trade-offs, overall consumer preferences between systematically configured product or service options varying on factors based on ABCs can be decomposed and appropriately allocated among the underlying ABCs. This partitioning process enables the development of a hierarchy of importance (utilities) for determining the relative impact of each factor in accounting for choices between the options.

As with the initially described indirect methodology, importance is determined statistically with greater systematic variance defining greater importance. Consumer selections between the constructed alternative product or service offerings are analyzed to identify which factors make the greatest contributions in accounting for observed choices. When a factor is composed of levels, importance still is a function of variance: the more "differentiating" the levels—the larger the range of utilities between the levels—the more "important" the factor.

Several potential difficulties also are inherent in this second indirect importance assessment methodology: First, often these assessment tools need to be constrained by the number and complexity of the options that a consumer can reasonably consider and evaluate at any one point in time. Second, many products or services do not readily lend themselves to decomposition into relatively independent factors and ratio or interval levels. Third, maintaining the high fidelity of the product service within an elementary set of rational and emotional ABCs that are valid, believable, and accurate representations of perceived differences between choice alternatives can be a difficult task. Fourth, design factors—including how choices are shown and the number of factors and levels presented—all have been empirically demonstrated to affect results, thereby requiring very careful experimental and analytic controls in whose absence the robustness of the methodology and projectability of the results could be challenged. Finally, the reliance on variance as the arbiter of importance in these choice-based system burdens this approach with essentially the same limitations identified with the other indirect methodology.

ADDRESSING THE LIMITATIONS

Sometimes, to offset some of the weaknesses of each approach (and, in particular, those potentially raised by the POE and POD challenges), data from one of the two direct and one of the two indirect approaches are combined into a single analysis. This convergence of methods can augment one another, particularly when those ABCs rejected as unimportant when using one analytic method emerge as important in the other. ABCs identified as low on direct and high on indirect importance often are described as "unrecognized potential motivators"; and, those high on direct and low on indirect importance often are identified as POEs.

Not surprisingly, researchers working with each of the four techniques have developed study design and/or analytic enhancements and accommodations to address each assessment method's limitations. For example:

- The introduction of enhanced stimuli such as photographs and self-created montages has enabled consumers to represent visually what is important to them (Zaltman, 2003).
- Automated analysis of online customer reviews has been suggested as a tool to help identify relevant attributes and

levels that might otherwise have been missed (Lee and Bradlow, 2007).

- Since various independence assumptions inherent in regression analyses are relaxed and suspended in indirect assessment methods, a game theory analytic approach has been promoted (Shapley, 1953). A Shapley value bootstrapping analysis procedure normalizes the regression coefficients and relates the contribution of each ABC to the overall *R* squared (Conklin and Lipovetsky, 2005; Lipovetsky and Conklin, 2001), thereby attempting to mitigate some of the analytic challenges, including negative weights posed by highly correlated ("multicollinear") ABCs.
- Various segmentation and latent class models have been suggested to identify ABCs that are important to homogeneous subgroups of respondents.
- The choice method has seen the development of both "adaptive" and hybrid models for simplifying and accelerating the choice tasks (Johnson, 1991). Adaptive models have been incorporated both prior to the initiation of choices (by eliminating or simplifying options) and during the choice task itself (by applying Bayesian analyses to information inherent in early responses to limit the number of subsequent choices needed to obtain stable results). Hybrid approaches—combining both direct and indirect importance estimation within a common study design—have been utilized to reduce what might be recognized as an onerous number of consumer decisions.
- In conjoint trade-off designs, random effects hierarchical multinomial logit models have extended Luce's (1959) original approach to allow for individual part-worth estimates as well as aggregated group data (Allenby, Arora, and Ginter, 1998; Ding, Grewal, and Liechty, 2005).

THE CHALLENGE OF CAUSALITY

In all four assessment techniques cited above, consumer responses and ABCs are prioritized under the assumption that they can be related to in-market behavior in a well-defined and valid manner reflective of their importance. More specifically, by designating an ABC as important, it is implied that significant changes in consumer awareness, attribution, and/or belief of that item in a choice option would result in a behavioral (or, at least, an attitudinal) change. Marketing priorities then can be set consistent with the discovered pattern of results.

While all four of these traditional analytic methods have been accepted as aiding marketers in understanding what is important to the consumer, despite their respective design and analytic improvements, all continue to suffer from a critical shortcoming popularized in the book *Freakonomics* (Levitt and Dubner, 2005). These methodologies claim to identify and quantify what is important to the consumer from data typically collected at a single point in time. Since a correlation, by itself, explains nothing about the direction of a relationship, it can leave the analyst "fooled by association." At its essence, correlation is not equal to causation; it is only a requirement for it.

Even if the methodological shortcomings inherent in each of the techniques cited above were rectified, none of the four traditional approaches were designed to address the challenge of predictive validity. While from a philosophical perspective causality cannot be definitively proven, in practice, the only way to demonstrate likely causality is through experimental design or by collecting multiple measurements over time and applying time-based analytic techniques to that data. Stated in terms of the conceptual paradigm described here, responses to the "what" question can only answer the "why" question if they are predictive of behavior. In the final analysis, organizations want to create and implement marketing initiatives that they believe will affect future consumer behavior.

METRICS THAT MATTER

As a result of having emerged as "important" in multiple traditional assessment methodologies, common market research measures, such as top-of-mind brand recall, advertising awareness, and product/service quality regularly have been identified as critical to in-market success and termed KPIs. Many of these metrics have found their way into senior management reports in the form of corporate scorecards and dashboards, sometimes with employee compensation programs linked to performance benchmarks.

In the approach described here, a distinction will be drawn between KPIs that potentially *may be* important to a given brand's success at a specific point in time and those that *are* important—metrics that matter (MTMs). More specifically, MTMs are defined as KPIs that have been empirically demonstrated to drive sales and/or share over a defined time period within a real world competitive landscape. The identification of MTMs is a logical consequence of a systematic research progression for identifying what is important to consumers. This process can be visualized as a brand vitality funnel beginning with many ABCs hypothesized as potentially important to consumer attitudes, beliefs, and behavior and some eventually graduating to become MTMs that are proven to drive market sales/share (see Figure 1).

Traditional market mix modeling analyzes the impact of variations in pressure from specific advertising and promotion marketing initiatives to quantify relative and absolute effects on sales or share. In a complementary manner, VAR modeling

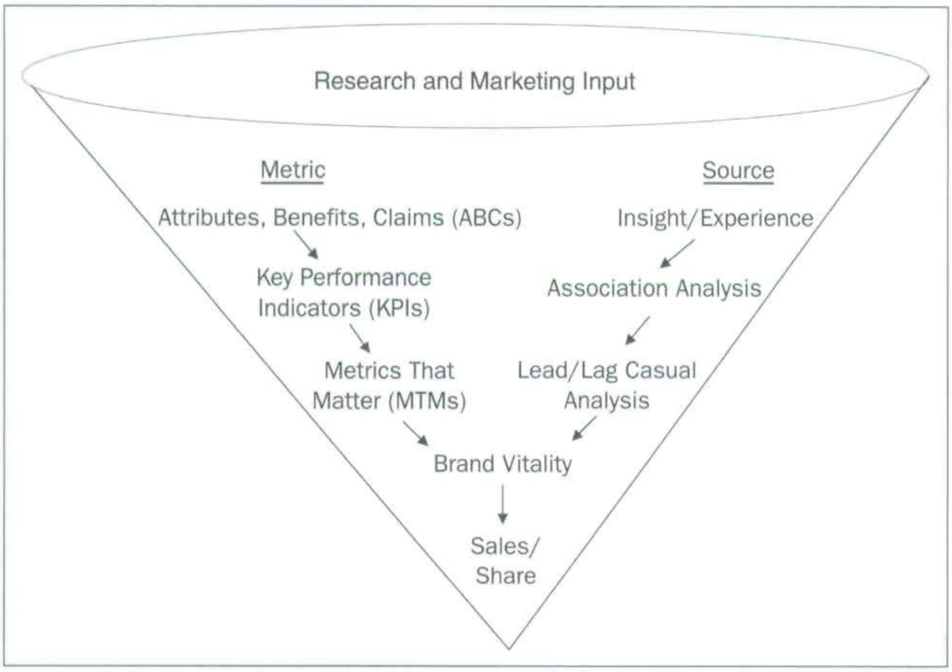

Figure 1 The Brand Vitality Sales Funnel

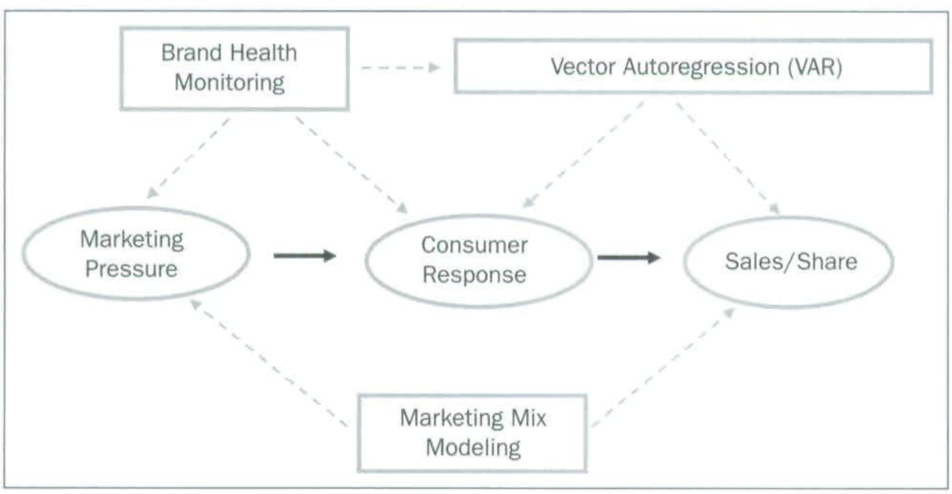

Figure 2 Integrated Marketing Mix and Vector Autoregression (VAR) Modeling

addresses the "intervening consumer effects" not explicitly measured by marketing mix modeling (see Figure 2). Thus, VAR technology provides a crucial link between specific consumer attitudes, beliefs, and opinions and their impact on market demand data (see Table 1).

BRAND HEALTH ASSESSMENT: THE DIAMOND MODEL

Providing a unifying framework for our VAR analysis to identify and represent the causal relationships among MTM is a model of brand health that, because of its integrated five-facet structure, has been termed the Diamond Model (see Figure 3). The Diamond Model provides a parsimonious, management-friendly representation of brand health as a dynamic concept. Unlike some other equity-based approaches that are more static (Aaker, 1991; Keller, 1993), this model differentiates in a rigorous fashion between the creators or causes (drivers) and consequences (effects) of brand health. When combined with a time-based data stream, the Diamond Model can formally represent the dynamic waxing and waning of a brand's health as reflected in consumer responsiveness to changing global and local market conditions, the pressure deriving from marketing initiatives and the availability of multiple touch points.

Each of the five facets in the Diamond Model is composed of multiple, broad-based ABCs that a detailed review of the market research literature across many categories had suggested were "important" in defining brand health. In each specific research program, client and industry research were used to supplement and configure the ABC list, thereby customizing the model for the category under study. All of these ABCs were characterized as potential KPIs.

The Diamond Model identifies:

- two drivers of brand health:
  - *Awareness*—presence-based metrics such as brand and advertising/promotion salience that create familiarity, and
  - *Attribution*—rational and emotional association metrics to reflect imagery and product experience that create perceptions of relevance and differentiation;
- two consequences of brand health:
  - *Attachment*—loyalty metrics such as affiliation and connection, and
  - *Accrual*—valuation metrics such as willingness to pay a price premium

TABLE 1
Stages of Importance Assessment

	Stage 1	Stage 2	Stage 3
Action	Say	Imply	Do
Metric	Direct importance	Indirect importance	Demand
Basis	Magnitude and frequency	Differentiation and uniqueness	Time-based, lead-lag covariance

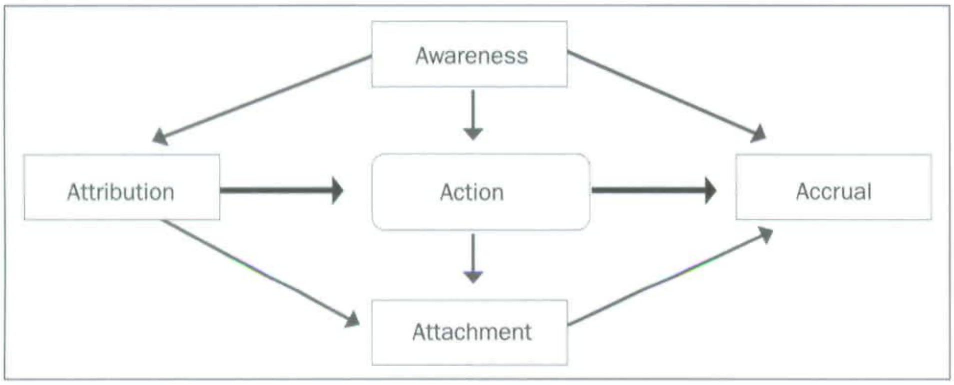

Figure 3 The Diamond Model of Brand Health

and purchase without a promotion or incentive that reflect the buildup or decrease of asset value;

- one intermediate consequence facet:
  - *Action*—activity metrics such as trial, rejection, and patterns of usage.

The arrows in the model represent the primary causal directions as brand health is created or diminished. While not represented in the basic model, feedback loops can be highly relevant between the model facets and their components. For example, consumer service/product experiences as reflected in the KPI related to Action are likely to affect Attribution.

The importance of the "paths" between the facets and components of the Diamond Model can be expected to vary with different categories, markets, brands/services, brand life stage, decision (brand) involvement, etc. Also, the consumer dynamics inherent in various categories would be reflected by appropriate ABCs within the facets. For example, the Accrual facet for a consumer packaged goods (CPG) brand would include outcome value metrics such as promotion sensitivity; in the financial space, it would include preference-related metrics such as share of wallet for banking and front of wallet for credit cards.

ANALYTIC APPROACH

The consumer data generation approach underlying the identification and quantification of MTMs is continuous surveys (tracking) in the relevant market space. The objective is to monitor the dynamics of a brand's "vital signs" on each of the five facets as contributors to and definers of its health. By integrating consumer survey feedback from tracking data with actual brand sales provided by IRI and/or Nielsen, we demonstrate how to incorporate VAR time series modeling (Cromwell, Hannan, Labys, and Terraza, 1994; Dekimpe and Hanssens, 2005; Enders, 2004) into a brand health monitoring program. This analysis will provide an understanding of the importance of each of the ABCs—that is, the identification and quantification of which ABCs (and, more specifically, KPIs) have been driving brand sales and, therefore, can be designated as MTMs.

Data sets from five consumer tracking programs involving a health and beauty aid (HBA) shaving product, three food products, and an over the counter (OTC) analgesic will be used to illustrate the VAR approach. The HBA data set is of particular note, as we will compare the results obtained with two traditional importance assessment approaches with VAR modeling.

The primary focus of our analysis will be contrasting the VAR results with brand ratings data as typically analyzed in the indirect importance PERCEPTOR/ASSESSOR-based system. This MIT-developed system is the one that, over the years, market researchers most typically have relied upon to identify importance "drivers" in many different types of research studies including awareness and usage, segmentation, and brand tracking. Glen L. Urban, in his 2002 autobiographical essay, "When I Stop Learning, I Will Leave," claimed in the pages of the *Journal of Marketing* that more than 3,000 new CPG products alone have been tested using this methodology. To provide a full assessment of this technique, three frequently used dependent variables—purchase intent, overall liking, and overall satisfaction—were utilized in our indirect analyses.

While a more detailed description of the analytic methodology is presented in the Appendix, a general overview follows: We conducted a unit root test (Enders, 2004) to assure stationarity of the stochastic process and to avoid the problems of spurious regressions and violations of the standard assumptions underlying hypothesis testing of the regression parameters. Granger tests served to establish causal relationships in the context of linear predictions as we sought to identify which ABCs demonstrate a causal relationship in driving sales (Granger, 1969). As each study included many ABCs, we have only shown those that have satisfied these standard tests.

After establishing causality, we used VAR models and OLS estimation to quantify the impact on sales for these causal variables. In an unrestricted VAR model, all variables were expressed as functions of lagged values (their own and the values of all others in the system). Based on the estimated VAR parameters, simulations using impulse response functions enabled the evaluation of each KPI independently, providing estimates of its effect on sales and share (Sola, undated). As is typical in VAR analyses, seasonal sales spikes were modeled using dummy variables corresponding to the data periods in which such spikes historically have been observed.

We also demonstrated the time trajectory of lag effects—both short and long term—and their impact on both base and incremental sales, as determined through market mix modeling provided by an independent third-party research agency. Base volume is generally assumed to represent long-term effects, with incremental volume representing short-term effects attributable to marketing and promotion activity.

In the segmentation study, consumers directly rated a battery of ABCs on how important they were in the selection of a preferred brand. Seven ABCs—identified as KPIs by the client from multiple prior research studies conducted over many years—were included in both the segmentation and the tracking studies.

RESULTS

Two studies in the HBA personal hygiene segment (a segmentation study and a continuous brand health tracking study built on the Diamond Model platform) were analyzed. Both studies included random samples of a target male sample. The segmentation study included 1,000 respondents who qualified for the survey based on category use. In the tracking study, slightly more than two years of data were aggregated into 54 two-week periods, with each period based on approximately 90 category users.

In the segmentation study, consumers directly rated a battery of ABCs on how important they were in the selection of a preferred brand. Seven ABCs—identified as KPIs by the client from multiple prior research studies conducted over many years–were included in both the segmentation and the tracking studies. In the brand tracking study, ratings of brands on ABCs were analyzed using multiple methods. These analyses were VAR models for both base and incremental sales, as determined through market mix modeling and indirect importance analyses with three dependent variables: purchase intent, overall liking, and overall satisfaction. To address the issue of all KPIs not being able to be analyzed in a single VAR model due to the number of data periods available for analysis, those that passed the causality screen were grouped into facets and stepwise regression analyses were conducted.

The VAR results demonstrated that advertising/promotion awareness had nearly ten times the impact of any other significant KPI on base sales, clearly qualifying it as an MTM (see Table 2). The VAR model also demonstrated the KPIs of "allows shaving against the grain" and (the satisficing metric of) "good enough shave" to be the strongest product/brand MTM sales drivers. Similarly, being portrayed as "ordinary" (not a beauty model), but still "attractive to the opposite sex," and not "self-absorbed" were the strongest MTM user images and, ultimately, were selected by management to be part of the brand's future copy platform.

The MTMs driving incremental unit sales were found not to be as strong as those driving base sales. Since every MTM that had a significant impact on incremental sales also demonstrated a significant impact on base sales, movement on those metrics improved total unit sales from two perspectives.

None of the three traditional dependent variables—purchase intent, overall satisfaction, or overall liking—demonstrated a statistically significant effect in the VAR analysis. Moreover, comparing the VAR

TABLE 2
Health and Beauty Aids (HBA) Product

	VAR Unit Sales Base	VAR Unit Sales Incremental	Indirect Importance—Purchase Interest/Liking/Satisfaction	Ratings—Top-Box/Top-Two-Boxes[a] (%)	Ratings Mean[a]	Direct Importance—Top-Box/Top-Two-Boxes[a] (%)	Direct Importance—Mean[a]
I. Brand Awareness							
Top-of-mind awareness			0.12/0.16/0.17	5			
All unaided awareness			0.05/0.13/0.09	13			
Advertising/promotion awareness	2,866		n.s.[b]/0.08/n.s.	9			
II. Brand Attribution							
Features							
High-quality brand			0.30/0.53/0.54	47/80	4.23		
Growing more popular			0.37/0.28/NS	23/60	3.73		
Better shave than others	101	71	0.36/0.61/0.66	31/64	3.89		
Good enough shave	300	40	0.32/0.59/0.64	43/71	4.08		
Gives a close shave			0.30/0.57/0.57	37/75	4.07	80/97	4.75
Using fewer strokes			0.30/0.47/0.46	27/59	3.79	48/79	4.14
Prevents skin irritation			0.35/0.52/0.57	19/53	3.61	70/91	4.57
Shaving against grain	299		0.32/0.50/0.55	31/60	3.83	69/90	4.54
Technologically advanced	116		0.25/0.50/0.44	33/70	3.97	29/61	3.77
User imagery							
Handsome	42		NS/0.19/0.20	24			
Attractive to opposite sex	135		NS/NS/0.18	19			
Self-absorbed	–97		–0.14/NS/NS	7			
Ordinary	271	109	NS/NS/NS	17			
III. Brand Action							
Regularly/occasionally use			0.37/0.36/0.34	23	NA		
Purchase intent			NA/0.57/0.55	9/24	2.66		
IV. Brand Attachment							
Overall liking			0.57/NA/0.86	22/41	4.96		
Brand I trust			0.32/0.60/0.64	44/76	4.15		
Brand for people like me	201	120	0.44/0.59/0.57	29/58	3.73		
Overall satisfaction			0.55/0.86/NA	33/57	5.46		
V. Brand Accrual							
Good value for money			0.37/0.38/0.40	16/45	3.37	65/89	4.50
Worth paying more for	272	262	0.40/0.46/0.49	17/44	3.30	41/72	4.07

[a]All ratings are 5-point Likert scales except for overall liking and overall satisfaction, which are 7-point scales, and user imagery, awareness, and usage, which are dichotomous.
[b]n.s. = not significant.

findings with the indirect importance analyses revealed a different pattern of results. The strongest MTM (advertising/promotion awareness) did not emerge as statistically significant in the traditional indirect importance analysis with purchase intent or overall satisfaction as the dependent variables and had only an extremely weak effect with overall liking.

Although overall liking and satisfaction generally showed higher relationships with the KPIs, there was very little differentiation between the product/brand features, which would make setting marketing/communication priorities between them somewhat problematic. Only "growing more popular" dropped out as weaker than the other product/brand features,

The VAR model also demonstrated the KPIs of "allows shaving against the grain" and (the satisficing metric of) "good enough shave" to be the strongest product/brand MTM sales drivers. Similarly, being portrayed as "ordinary" (not a beauty model), but still "attractive to the opposite sex," and not "self-absorbed" were the strongest MTM user images and, ultimately, were selected by management to be part of the brand's future copy platform.

and only then with overall liking and overall satisfaction as the dependent variables. "Good enough shave" and "allows shaving against the grain" (both of which showed a strong impact on base sales in the VAR analysis) did not emerge as significantly stronger than the other product/brand features in the indirect importance analysis.

Particularly interesting were the user-imagery ratings that demonstrated low and/or not statistically significant results in indirect importance but qualified as MTMs, such as "attractive to the opposite sex" and "for ordinary people." Comparing the results on the direct and indirect ratings relative to the VAR model results, it can be seen that while "gives a close shave" was rated significantly higher than any other metrics in the direct importance rating, it was not among the highest scoring in our indirect importance analysis and was not statistically significant in either the base or incremental VAR models, suggesting that it may be a POE. Similarly, the claim "is technologically advanced"—a major foundation of the brand positioning in the client's brand architecture, and observed to be an MTM in the VAR analysis—was rated statistically significantly lower on direct importance than any of the six other client-designated KPIs.

Some consistency between the VAR results and the indirect and direct importance ratings was observed. This was particularly noticeable with respect to the relatively high scores for the Accrual claim "worth paying more for" and the Attribution benefit "allows shaving against the grain." Interestingly, on neither of these claims was the brand rated particularly strongly.

Finally, a review of the top-box, top-two-box, and mean ratings demonstrated that high performance on brand ratings or user imagery KPI did not necessarily lead to either high-indirect importance or strong performance in the VAR analyses, illustrating the competitive nature of the category.

A second application of VAR technology is shown in the form of the results of an analysis of two years of survey data for a food product. The results were aggregated into four-week intervals of 150 target market respondents. Shown are the VAR analysis, the derived importance analyses for purchase intent, overall liking, and overall satisfaction, and the top-box, top-two-box, and means for the KPIs (see Table 3).

Unlike the HBA analysis, a large number of metrics for the food product qualified in the VAR analysis as MTMs. In fact, every facet of the model incorporated as least one KPI qualifying as an MTM and all five Diamond Model facets represented significant unit sales volume. The Attachment facet included the two strongest KPIs (overall liking and favorite brand), and every brand personality image had a significant impact on sales.

Comparing the VAR results with the indirect importance measures demonstrated that many of the KPIs were recognized to be important in both analyses. Still, it was apparent there were a number of important differences between the two analyses. First, the advertising/promotion awareness—the core positioning for the brand and nearly all of the user images—demonstrated indirect importance scores much lower than the other metrics, even though the VAR analysis indicated that all of these metrics had a large impact on sales. In fact, it can be seen that advertising/promotion awareness was among the strongest MTMs.

Second, the brand personality imagery, while strongly associated with the brand, demonstrated much lower indirect importance scores than the product/brand attributes, possibly because of dichotomous ratings. In the VAR analysis, two brand personality images, "contemporary" and "engaging," generated more sales than any product/brand feature other than "a kitchen staple."

Third, of the three dependent variables in the indirect analysis, only overall liking also demonstrated a strong contribution to sales in the VAR analysis. Analyzing

TABLE 3
Food Product

	VAR Total Unit Sales	Indirect Importance—Purchase Interest/Liking/Satisfaction	Ratings—Top-Box/Top-Two-Boxes[a] (%)	Ratings—Mean[a]
I. Brand Awareness				
Top-of-mind awareness	21,525	0.37/0.31/0.25	63	NA
All unaided awareness	20,903	0.46/0.40/0.34	73	NA
Advertising/promotion awareness	21,364	0.22/0.16/0.15	50	NA
II. Brand Attribution				
Core positioning "authentic"	12,823	0.22/0.19/0.13	49	NA
Features				
High-quality brand		0.41/0.66/0.60	50/66	5.93
Growing more popular		0.43/0.62/0.59	32/50	5.39
Has a fresh taste	14,317	0.54/0.65/0.71	47/63	5.69
Good for use anytime		0.55/0.66/0.68	50/66	5.82
High quality ingredients		0.46/0.71/0.64	47/61	5.79
Has a taste I love		0.50/0.77/0.72	42/59	5.55
Really different/unique	14,234	0.49/0.54/0.55	32/46	5.15
The best brand	12,364	0.62/0.70/0.70	46/60	5.56
A kitchen staple	21,333	0.64/0.61/0.60	47/58	5.26
Personality/imagery				
Bold	11,174	0.28/0.28/0.27	50	NA
Contemporary	15,794	0.18/0.16/0.15	31	NA
Life of the party	11,032	0.29/0.29/0.26	42	NA
Engaging	16,275	0.26/0.25/0.24	34	NA
Genuine	8,821	0.30/0.30/0.29	48	NA
Vibrant	6,911	0.28/0.28/0.28	45	NA
III. Brand Action				
Regular/most often used	19,373	0.54/0.49/0.43	60	NA
Purchase intent		NA/0.68/0.65	52/73	4.00
IV. Brand Attachment				
Overall liking	23,315	0.68/NA/0.81	52/69	5.81
Brand I trust		0.48/0.71/0.62	53/70	5.86
Favorite brand	24,470	0.52/0.48/0.44	52	NA
Relevant to you/your family		0.57/0.59/0.59	38/50	
Overall satisfaction		0.65/0.81/NA	57/73	5.98
V. Brand Accrual				
Good value for the money	18,780	0.50/0.57/0.56	38/55	5.61

[a] *All ratings are 7-point Likert scales except for purchase intent, which is 5-points, and brand personality imagery, awareness, and usage, which are dichotomous.*

the other KPI metrics in terms of their correlation with overall liking did not show a pattern of results different from that of the other two dependent variables. This lack of differentiation was not surprising, given the relatively high intercorrelation between the three dependent variables.

Our third VAR application involved a food (sauce) product. We compared our survey results for 25 four-week intervals (125 category users per period) with sales data partitioned using market mix modeling. Only two KPIs, advertising/promotion awareness and purchase intent among current users, accounted for significant base unit sales. The importance of advertising/promotion awareness to sales would not have been recognized from the indirect importance analysis (see Table 4).

Since the market mix model showed that incremental advertising/promotion awareness was being driven by television advertising, we conducted a separate VAR share model for television-generated sales. The model indicated that user imagery accounted for 86 percent of the incremental sales driven by television. The product/brand features, although particularly highly associated with the brand as shown in the ratings, were responsible for the remaining 14 percent of the incremental television-generated sales. Brand movement on the user ratings of "friendly" and "family pleasing" images had the greatest impact, accounting for nearly 60 percent of the television-driven incremental volume—results that likely would not have been recognized simply by directly comparing their indirect importance scores to those of the other user image KPIs.

Our fourth VAR analysis included three years of survey and sales data aggregated into four-week periods of 200 category users for a market-leading branded snack product. This analysis demonstrated the

TABLE 4
Food Product

	VAR Unit Sales		Indirect Importance—	Ratings—Top-Box/	
	Base	TV Incremental (%)	Purchase Interest/ Liking/Satisfaction	Top-Two-Boxes[a] (%)	Ratings—Mean[a]
I. Brand Awareness					
Advertising/promotion awareness	25,510		0.18/0.14/0.09	55	NA
II. Brand Attribution					
Features					
Meal whole family likes		7	0.49/0.58/0.66	53/72	4.17
Rich taste		6	0.42/0.56/0.60	46/70	4.12
High-quality ingredients		1	0.39/0.53/0.61	45/68	4.14
User imagery					
Friendly		33	0.24/0.26/0.26	38	NA
Family pleasing		25	0.32/0.36/0.39	49	NA
Genuine		18	0.27/0.29/0.31	43	NA
Italian		10	0.20/0.23/0.24	53	NA
III. Brand Action					
Purchase interest (users)	16,767		NA/0.64/0.64	26/56	3.41
IV. Brand Attachment					
Brand I trust			0.48/0.58/0.65	58/75	4.29
Favorite brand			0.50/0.51/0.51	50/72	4.17
Overall liking			0.66/NA/0.84	30/47	5.08
Overall satisfaction			0.64/0.84/NA	36/54	5.42

[a]*All ratings are 5-point Likert scales except for overall liking and overall satisfaction, which are 7-point scales, and user imagery and awareness, which are dichotomous.*

ability of our models to address concurrent and lagged effects as well as quantify the impact on actual product sales (by weight) for a one-time, one percentage point increase for each MTM (see Table 5).

Analysis of the Attribution facet demonstrated that short-term sales were driven by product/brand features. Total sales were driven more by occasions of use—not an unexpected finding in what is generally considered to be an impulse-driven category. Overall product satisfaction—a KPI widely considered to drive snack product consumption—was validated as an MTM, having both short- and long-term effects. In a demonstration of consistency with our VAR findings, overall satisfaction also had the directionally highest correlation with purchase interest and significantly highest with overall liking in the indirect importance analysis.

In our charts, "wear-in" refers to the number of periods it took to reach the respective MTM peak impact on sales and "wear-out" to the number of periods the MTM effect lasted beyond the peak period to complete decay (see Figure 4).

TABLE 5
Snack Product

	VAR Sales						
	Short-Term Sales (pounds)	Total Sales (pounds)	Wear-In	Wear-Out	Indirect Importance—Purchase Intent/Liking/Satisfaction	Ratings—Top-Box/Top-Two-Boxes[a] (%)	Ratings—Mean[a]
I. Brand Awareness							
Top-of-mind					0.17/0.16/0.13	10	
Advertising/promotion awareness					0.23/0.14/0.17	58	
II. Brand Attribution							
Occasions of use							
For on the go	0	88,326	1	0	0.27/0.30/0.23	39/52	3.39
Lift in the afternoon	0	79,753	1	0	0.31/0.33/0.30	42/57	3.56
Relaxing by yourself	0	76,633	1	0	0.37/0.39/0.32	46/61	3.71
Entertaining	0	70,322	1	0	0.35/0.38/0.39	61/79	4.27
Features							
Satisfying taste	44,488	44,488	0	0	0.36/0.47/0.49	59/80	4.34
Taste I love	23,890	23,890	0	0	0.38/0.54/0.55	56/76	4.26
III. Brand Action							
Regular/most often					0.26/0.20/0.17	14	NA
Purchase intent					NA/0.52/0.47	27/56	3.48
IV. Brand Attachment							
Brand I trust	54,962	54,962	0	0	0.30/0.39/0.46	77/85	4.51
Overall liking					0.52/NA/0.70	42/62	5.65
Favorite					0.29/0.26/0.21	15	NA
Overall satisfaction	69,485	288,673	2	0	0.47/0.70/NA	52/71	5.97
V. Brand Accrual							
Good value for money					0.31/0.30/0.37	41/60	3.84
BVS	139,031	260,761	0	1	NA	NA	NA

[a] *All ratings are 5-point Likert scales except for overall liking and overall satisfaction, which are 7-point scales, and awareness, favorite, and usage, which are dichotomous.*

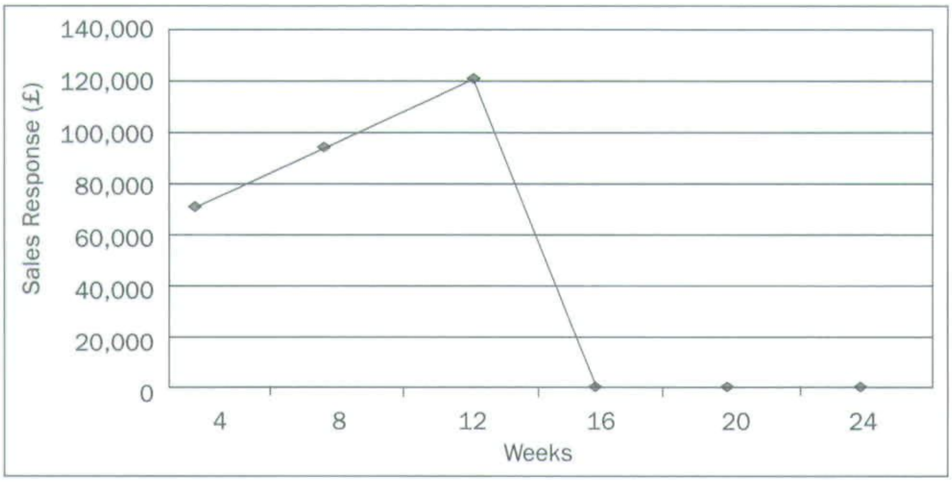

Figure 4 Sales Response to a One Percentage Point Increase in Overall Satisfaction

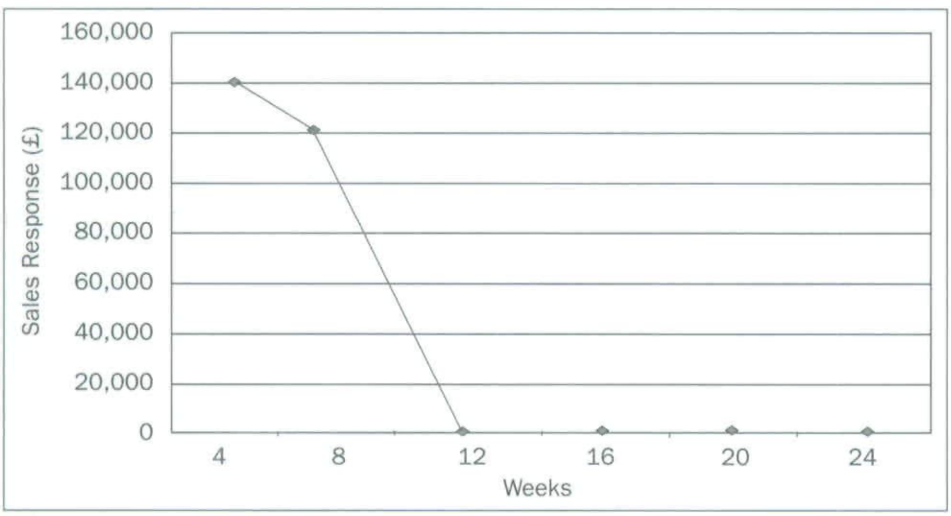

Figure 5 Sales Response to a One Percentage Point Increase in Brand Vitality Score (BVS)

Occasions of use were observed to have a one-period (week) wear-in period, while product/brand features had an immediate effect. An overview of the wear-in and decay function for overall product satisfaction showed that satisfaction peaked at week 12, dissipating by week 16.

In terms of relationships of the three dependent variables with the significant KPIs, overall liking and satisfaction were directionally higher for the product/brand features relative to occasions of use. This may have reflected the finding that the snack product performed significantly better (all $p < 0.01$) on the two product/brand feature KPIs relative to three out of the four occasions of use KPIs. In contrast to the indirect importance analyses, the VAR model showed that the product/brand variables did not have a larger effect than the occasions of use variables on total sales.

Also included in the VAR analysis was an overall measure of brand health, the Brand Vitality Score (BVS).[1] This measure differs from many other overall brand-health measures in that it provides a summary metric limited to consequences (effects). BVS is based on purchase intent, overall favorability, and overall liking; it is composed of consumers who claim to be familiar with a brand, thereby somewhat leveling the playing field for less well known and/or newer brands relative to better known (often larger) ones. BVS captures both short- and long-term effects with the impact of movement shown as having a relatively large impact on sales for up to 12 weeks (see Figure 5).

Given the importance of private label products in this snack category, we also conducted a VAR analysis of the impact of changes in the health of the branded product on private label brands. This analysis showed that significant positive movement on the branded product ratings on "need a lift in the afternoon" and "for on the go" led to a short-term sales lift for private label brands. Furthermore, a 1 percent enhancement in BVS for the branded product led to large increases in both short-term (67,279 pounds) and long-term total sales (190,452 pounds) for private-label products. Thus, the branded marketing/promotion activity that improved the brand equities of the branded product (the category leader) also grew the category.

[1] *BVS was developed by Dr. Orly Maravankin. It has been validated against market share and sales in multiple CPG categories.*

Our final VAR analysis was conducted for an analgesic product in the OTC drug category, four-week aggregations years were studied over approximately 2.5 years, yielding more than 32 periods of data. Approximately 160 target-market consumers were interviewed in the first 13 periods and 320 in each of the remaining periods. While facets other than Attribution were included in the study, only the results for KPIs related to symptom appropriateness are shown. It is generally accepted in the pharmaceutical industry that consumer selection of OTC brands is largely determined by perceptions related to efficacy in symptom alleviation and treatment.

Using the results from marketing mix models to parse out the effects of total unit sales and advertising driven incremental sales, the impact on average four-week unit sales of a one percent increase on ratings can be seen, both in terms of its impact on total unit sales (base plus incremental) and on incremental sales due to advertising. A clear MTM priority was apparent. Improvement in perceived performance on migraine headaches, muscle pains, and backaches was shown to have a greater impact on sales than improvement on general aches and pains, regular headaches, cold/flu, and especially general strains and sprains. In terms of sales directly associated with advertising, improvement was less differentiated, but related to higher ratings on appropriateness for "joint pain and stiffness," "general aches and pains," and "muscle pain" (see Table 6).

The three indirect importance analyses with purchase intent, liking, or satisfaction as dependent variables did not show statistically significant differences among the 10 symptoms, with all essentially equivalent. Thus, the indirect importance metrics did not provide a clear road map for addressing symptom priority in future marketing efforts.

TABLE 6
Over the Counter (OTC) Product

	Effect as a Percent of Total Unit Sales	Effect as a Percent of Advertising
Brand Attribution symptom appropriateness		
Muscle pain	1.8	0.9
Migraine headaches	1.8	0.0
Backache	1.7	0.0
Sinus headaches	1.4	0.8
Minor arthritis pain	1.4	0.2
Joint pain and stiffness	1.0	1.1
General aches and pains	0.8	1.0
Regular headaches	0.5	0.0
Cold and/or flu	0.5	0.6
General strains and sprains	−0.3	0.6

CONCLUSIONS

The acronyms KPIs and MTMs are often used interchangeably for identifying "important" areas where a company intends to focus its business efforts (e.g., ConAgra, 2004). The findings presented demonstrate that rigorously differentiating between KPIs and MTMs on the basis of driving market share and sales can provide significant insights and potentially a competitive marketing advantage.

Multiple examples from ongoing brand health tracking programs demonstrated how VAR models, when combined with sales data and applied to a brand-health model, could identify those metrics that have the greatest potential impact on in-market sales/share. More specifically, the VAR analytic protocol presented an approach for qualifying KPIs as MTMs ("brand health levers") and quantifying their impact on sales. Also, we illustrated how VAR models can take advantage of the partitioning of sales effects traditionally available through marketing mix modeling to identify which MTMs are driving different components of total sales.

The availability of VAR-derived information enables marketers to focus marketing and promotion initiatives on those metrics expected to have a meaningful and predictable sales impact. By allocating resources among MTMs in a manner reflective of their expected return, marketers can implement a formal return on marketing investment (ROMI) analysis to drive brand planning decisions. For example, assuming that both unaided brand awareness and high quality are MTMs, resource allocation decisions regarding the investment spending needed to apply leverage to yield additional performance on each of those two metrics could be weighed relative to their expected ROMI.

Since competitive brands also can be modeled, the VAR approach can address the challenges posed in identifying PODs and POEs by comparing actual ratings or rankings of the brands with the VAR results, identifying which ABCs meet MTM standards for each brand. Line extensions also can be independently evaluated with the goal of identifying MTMs that can facilitate brand differentiation and portfolio optimization.

VAR modeling efforts also can be directed toward private label and store brand products, either in aggregate or individually. One of the examples presented shows the unexpected sales lift enjoyed by private label products as they benefited from marketing initiatives for a market-leading, branded snack product.

Our VAR analyses clearly demonstrated the value of time series measurement in determining importance. Assessments of importance at a single point in time cannot address lead or lag effects. Furthermore, consumer metrics that do not demonstrate a current association with standard static direct and indirect assessment techniques would not be judged as important (and might even be rejected as having little merit), even though a dynamic analysis might show that they contribute to sales after accounting for significant time-based effects.

Traditional multiple assessments at different points of time (such as awareness and usage studies executed once or twice a year) or even reliability checks done during the same time period cannot adequately address this limitation. In essence, by focusing on trends rather than relying on data from a single point in time, market researchers will be less likely to be "fooled by randomness" (Taleb, 2005).

The indirect importance techniques that underlie many of the current analyses being used for identifying KPIs pose an additional potential problem in that they tend to be conducted at the category level. It is assumed that the same KPIs that drive a category also drive the sales of each of the brands in that category. We believe that it is risky to presume that KPIs observed to be category drivers function the same way for every brand in the category. In fact, our experience in multiple research projects with VAR models has identified significant differences in "drivers" for brands within a category. Moreover, the availability of an increasing proliferation of product options, multiple need states, and alternative distribution channels all have led to the definition of many categories becoming "fuzzy," complicating the creation of competitive sets.

It is reasonable to expect that the relative importance of MTMs for a given brand will vary as it matures in terms of its own life-stage and adoption cycle from being a new to an established product, as competitive brands emerge and change, as the market matures, and as marketing strategies and activity vary. Also, not all consumer metrics can be expected to demonstrate rapid change as a result of shifts in markets and marketing activity. Some metrics, particularly those directly related to a brand experience, are likely to generate slow moving signals, dependent on developing a critical mass of triers/users and purchase cycles. This would suggest the need to update a VAR analysis at least on a schedule reflective of a brand's evolution, if not continuously, as market conditions evolve.

Traditional importance-assessment systems have experienced challenges in appropriately measuring what has come to be recognized as the considerable influence of emotional and imagery related factors in the choice process. As demonstrated here, one advantage of the VAR approach is that brand- and user-oriented imagery and emotional ABCs not only can be readily included in these analyses, but also may be found to demonstrate a greater impact on brand sales than product/feature based metrics. This likely is due to the longer term nature of the data collection process and larger samples enabling sufficient sensitivity to capture the appropriate variance, even when the absolute levels of attribution of emotional user images relative to product/feature-based ABCs are relatively low and movement is slow.

Brand awareness metrics for assessing the effectiveness of advertising and promotion activity often are not included by traditional importance assessment methods in their analyses of the drivers of choice and preference. When measures of these effects (including media source and touch points, if available) are included in a single brand health survey instrument, VAR analyses can assess their sales impact relative to other metrics and to the overall vitality of the brand.

The VAR results presented also demonstrated the importance of integrating copy testing results with on-going tracking and ultimately MTM development. Consumer recognition and recall of actual advertising, promotions, and even messaging can be included in a VAR analysis. This approach can facilitate the setting of message priority to reflect VAR modeling findings in much the same way as media spending strategy tends to be directed by the results from market mix modeling. Copy testing can confirm if the messages generating the strongest consumer registration and most playback are also having the greatest sales impact. Overlaying spending data on the effectiveness of "moving the dial" on these metrics can provide valuable input to a ROMI model.

Future research should compare the predictive validity of the metrics identified by the VAR approach relative to other research techniques. An analytic methodology commonly used in market research to address the "what" of importance is structural equation modeling (SEM). This approach is based on individual level data and has been proposed to address the causal shortcomings of traditional importance assessment methods. SEM consists of selecting an aggregate dependent variable, such as loyalty, and developing multiple regression models built around a path model.

While a number of applications of SEM have proven to be useful in understanding brand health (e.g., Morgan, 2000), multiple time periods are rarely included in what is typically conducted as a cross-sectional analysis. However, even if multiple time slices were nested in a panel design, this approach still would be restricted by the number of periods analyzed. Unless time periods are continuous or at least "plentiful and consistent" from a statistical perspective, a cross-sectional analysis will present only a data snapshot, with the statistical analyses potentially capitalizing on one-time chance relationships to give a best fit to the data. Expectations that observed results and conclusions will remain as market dynamics evolve over time are questionable.

Other technical issues inherent in an SEM analysis can also limit its applicability and utilization. These include:

- modeling at the category level (as opposed to the brand level) when manufacturers are most often interested in understanding which ABCs will improve the sales performance of their product/service, not the category in aggregate;
- difficulty in interpretation at the managerial level due to often complicated and sometime recursive "paths" between brand health components; and
- utilization of criterion measures such as loyalty, which may not directly relate to sales.

Still, the SEM approach can add value. The structure of the Diamond Model of brand health readily lends itself to an SEM analysis following a VAR analysis. SEM can expand on the understanding derived from a VAR analysis by developing a structural model based on prescribed theories of individual-level human decision-making behavior. Thus, an SEM analysis can help understand how to drive specific MTMs that marketing has targeted because of their expected significant impact on sales/share. Also, it is possible to extend the SEM approach to integrate aggregate level VAR time series analyses and, thereby, test a more complicated underlying model.

The results we present here also suggest a need for the development of data-fusion models that capitalize on the advantages and minimize the disadvantages of various importance assessment methodologies. In the world of market research, combining traditional direct and indirect methodologies largely has been limited to the identification of POEs and PODs. Multistage models reflecting hierarchical consumer decision making and competitive set formation would seem to have potential value in this regard.

Economists have approached a similar data-fusion analytic challenge by "stacking" data from different methodologies focusing on the same value assessment to improve statistical efficiency. They have used this approach to combine preferences based on hypothetical scenarios derived from discrete choice and other experiments (which they term "stated" or "direct" behavior) and actual behavior (which they term "revealed" preferences). A two-level nested (conditional) logit model (Hensher and Bradley, 1993) has been suggested to identify and calibrate differences in scale (variance) between data sets while estimating model parameters (Louviere, Hensher, and Swait, 2000).

Seeking cause-effect linkages is critical not only for marketing and brand health management, but also in nearly every social, political, and business research program. For example, economists have long recognized that when there is an interest among governments to foster economic growth, innovation is an area of great importance. Thus, at the global level, there has been an attempt to relate the consequences of local rules, regulations, and political structures to innovative scientific achievements that have attained economic success. Metrics such as intellectual property, education level, and R&D expenditure (as reflected by the number of trademarks, patents, and overall productivity) all can serve, in the language of our model, as ABCs to be tested for becoming KPIs and potentially graduating to designation as MTMs.

Causal ambiguity necessitates that strategy and execution will always involve risk (when probabilities are known) and uncertainty (when probabilities are not known). When sufficient data are available, VAR modeling can serve as a unifying protocol for addressing this ambiguity. It can supersede not only the "observe, correlate, and assume causality" market research paradigm, but also the traditional case history method that bases its insights for achieving business success on simplistic comparisons between practices in more and less successful organizations and then assuming causality. JAR

Martin R. Lautman is the managing director at Marketing Channels, Inc., a provider of marketing services, research, and venture capital to the information industry.

Koen Pauwels is an associate professor at the Tuck School of Business at Dartmouth College.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Vaman Kudpi, Dr. Frank Zinni, Dr. Keith Beauregard, Wayne Robertshaw, and Jane Orner for their assistance.

REFERENCES

AAKER, DAVID A. *Managing Brand Equity*. New York: The Free Press, 1991.

ALLENBY, GREG M., NEERAJ ARORA, and JAMES L. GINTER. "On the Heterogeneity of Demand." *Journal of Marketing Research* 35, 3 (1998): 384–89.

CONAGRA, *Shareholder Report*. Third Quarter, 2004.

CONKLIN, W. MICHAEL, and STAN LIPOVETSKY. "Marketing Decision Analysis by Turf and Shapley Value." *International Journal of Information Technology and Decision Making* 4, 1 (2005): 5–20.

CROMWELL, JEFF B., MICHAEL J. HANNAN, WALTER C. LABYS, and MICHEL TERRAZA. *Multivariate Tests for Time Series Models*. Thousand Oaks, CA: Sage Publications, 1994.

DEKIMPE, MARNIK G., and DOMINIQUE M. HANSSENS. "Sustained Spending and Persistent Response: A New Look at Long-Term Marketing Profitability." *Journal of Marketing Research* 36, 4 (1999): 397–412.

——, and ——. "Persistence Models and Marketing Strategy." *Tijdschrift voor Economie and Management* 50, 5 (2005): 1–29.

DING, MIN, RAJDEEP GREWAL, and JOHN LIECHTY. "Incentive Aligned Conjoint Analysis." *Journal of Marketing Research* 42, 1 (2005): 67–82.

ENDERS, WALTER. *Applied Econometric Time Series*, 2nd ed. New York: John Wiley and Sons, 2004.

EVANS, LEWIS, and GRAEME WELLS. "An Alternative Approach to Simulating VAR Models." *Economics Letters* 12, 1 (1983): 23–29.

GRANGER, CLIVE W. J. "Investigating Causal Relations by Econometric Methods and Cross-Spectral Methods." *Econometrica* 37, 3 (1969): 424–38.

GREEN, PAUL E., and VITHALA R. RAO. "Conjoint Measurement for Quantifying Judgmental Data." *Journal of Marketing Research* 8, 3 (1971): 355–63.

——, and V. SRINIVASAN. "Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice." *Journal of Marketing* 54, 4 (1990): 3–19.

HANSSENS, DOMINIQUE M. "Market Response, Competitive Behavior, and Time Series Analysis." *Journal of Marketing Research* 17, 4 (1980): 470–85.

——, LEONARD J. PARSONS, and RANDALL L. SCHULTZ. *Market Response Models: Econometric and Time Series Analysis*, 2nd ed. Boston, MA.: Kluwer Academic Publishers, 2001.

HAUSER, JOHN R., and GLEN L. URBAN. "A Normative Methodology for Modeling Consumer Response to Innovation." *Operations Research* 25, 4 (1977): 579–619.

HENSHER, DAVID A., and MARK BRADLEY. "Using Stated Choice Data to Enrich Revealed Preference Discrete Choice Models." *Marketing Letters* 4, 2 (1993): 139–52.

JOHNSON, RICHARD M. "Comment on Adaptive Conjoint Analysis: Some Caveats and Suggestions." *Journal of Marketing Research* 28, 2 (1991): 223–25.

KELLER, KEVIN LANE. "Conceptualizing, Measuring, and Managing Customer-Based Brand Equity." *Journal of Marketing* 57, 1 (1993): 1–22.

LAUTMAN, MARTIN R. "The ABC of Positioning." *Marketing Research* 5, 4 (1993): 12–18.

LEE, THOMAS Y., and ERIC T. BRADLOW. "Automatic Construction of Conjoint Attributes and Levels from Online Customer Reviews." Unpublished manuscript, 2007.

LEVITT, STEVEN D., and STEPHEN DUBNER. *Freakonomics: A Rogue Economist Explores the Hidden Side of Everything*. New York: Harper Collins, 2005.

LIPOVETSKY, STAN, and MICHAEL CONKLIN. "Analysis of Regression in Game Theory Approach." *Applied Stochastic Models in Business and Industry* 17, 4 (2001): 319–30.

LOUVIERE, JORDAN J., DAVID A. HENSHER, and JOFFRE D. SWAIT. *Stated Choice Methods: Analysis and Applications*. Cambridge, UK: Cambridge University Press, 2000.

LUCE, R. DUNCAN. *Individual Choice Behavior: A Theoretical Analysis*. New York: John Wiley and Sons, 1959.

LÜTKEPOHL, HELMUT. *Introduction to Multiple Time Series Analysis*. Berlin: Springer-Verlag, 1993.

MORGAN, RORY P. "A Consumer-Oriented Framework of Brand Equity." *International Journal of Marketing Research* 42, 1 (2000): 65–78.

Myers, James H., and Richard F. Chay. "Direct vs. Derived Estimates of Ideal Product Attributes." Unpublished and undated manuscript.

Nijs, Vincent R., Shuba Srinivasan, and Koen H. Pauwels. "Retail-Price Drivers and Retailer Profits." *Marketing Science* 26, 4 (2007): 473–87.

Pauwels, Koen H., and Dominique M. Hanssens. "Performance Regimes and Marketing Policy Shifts." *Marketing Science* 26, 3 (2007): 293–311.

——, ——, and S. Siddarth. "The Long-Term Effects of Price Promotions on Category Incidence, Brand Choice, and Purchase Quantity." *Journal of Marketing Research* 39, 4 (2002): 421–39.

——, and Amit Joshi. "Counting What Will Count: Does Your Dashboard Predict?" Working Paper, Tuck School of Business at Dartmouth, Hanover, NH, 2008.

Pesaran, Hashem H., and Yongcheol Shin. "Generalized Impulse Response Analysis in Linear Multivariate Models." *Economics Letters* 58, 1 (1998): 17–29.

Shapley, L. S. "A Value for n-Person Games." In *Contributions to the Theory of Games*, Vol. II, H. W. Kuhn and A. W. Tucker, eds., *Annals of Mathematical Studies*. Princeton, NJ: Princeton University Press, 1953.

Silk, Alvin L., and Glen L. Urban. "Pretest Market Evaluation of New Packaged Goods: A Model and Measurement Methodology." *Journal of Marketing Research* 15, 2 (1978): 171–91.

Sola, Martin. "Lecture Notes." London: School of Economics, University of London, undated: [URL: http://www.ems.bbk.ac.uk/for_students/bsc_FinEcon/fin_economEMEC007U/VAR.pdf], accessed January 2007.

Taleb, Nassim N. *Fooled by Randomness*. New York: Random House, 2005.

Urban, Glen L. "PERCEPTOR: A Model for Product Positioning." *Management Science* 21, 8 (1975): 849–58.

——, and John R. Hauser. *Design and Marketing of New Products*, 2nd ed. Englewood Cliffs, NJ: Prentice Hall, 1993.

Zaltman, Gerald. *How Customers Think: Essential Insights into the Mind of the Markets*. Boston, MA: Harvard Business School Press, 2003.

APPENDIX

Our VAR analysis proceeds in five steps.

1. An augmented Dickey-Fuller test (Enders, 2004) unit root test verifies the univariate time-series properties (stationarity versus evolution) for each variable. This test addresses the question of whether a variable is mean-reverting (stationarity) or has changed permanently (the null hypothesis) in the data sample (evolution). If sales are mean reverting, no (marketing-induced) change has had a permanent sales effect.
2. We assess whether a metric is a leading indicator of sales performance by testing whether it Granger-causes sales (Granger, 1969; Hanssens, Parsons, and Schultz, 2001). The assumption underlying Granger causality (or, technically, noncausality) tests is that if an event y is the cause of another event x, then logically event y should precede x. A variable X is said to Granger-cause the performance variable Y if the mean-squared forecast error of Y using a bivariate model (i.e., explaining Y using past values for both X and Y) is smaller than that of using a univariate model (i.e., explaining Y using past values of Y only). Granger modeling involves assessing incremental forecasting power and is a statistical test of the joint significance of the other variable(s) in a regression, including dependent variable lags, enabling the testing of the impact on the dependent variable of multiple lag periods.

   An important choice is the number of time lags considered in these tests, as choosing a wrong number may lead to the erroneous conclusion that there is no Granger causality (Hanssens, 1980). Because we use these tests to eliminate variables (deleting variables that do not Granger-cause sales at any lag), we apply the tests for lags from one up to six months and consider a significant test result at any lag to be an indication that the variable is Granger-causing performance.
3. We estimate the dynamic interactions among sales and all leading performance indicators using VAR models. Instead of only treating performance as the dependent variable, VAR models regress a vector of endogenous variables (including performance, customer mindset metrics, and marketing actions) on the past of all endogenous variables, and thus capture complex dynamic interactions among these variables. In matrix notation, the VAR model is shown below:

$$Y_t = A_t + \sum_{i=1}^{p} \Phi_i Y_{t-i} + \Psi X_t + \Sigma_t,$$
$$t = 1,2,\ldots,T, \quad (1)$$

where Y_t is the vector of the endogenous variables, A_t the vector of intercepts, and X_t a vector of exogenous control variables, such as trend and seasonal patterns, and Σ_t the covariance matrix of the residuals. The number of lags p is selected by Schwartz' Bayesian information criterion, a consistent estimator of lag length (Lütkepohl, 1993).

4. We use the estimated VAR parameters to quantify the dynamic explanatory value of each endogenous variable on performance. Akin to a "dynamic R^2," generalized forecast error variance decompositions (GFEVD) provide a measure of the relative impact over time of shocks initiated by each of the individual endogenous variables in a VAR model, without the need to specify a causal ordering among these variables (Nijs, Srinivasan, and Pauwels, 2007; Pesaran and Shin, 1998). GFEVD estimates are derived using the following equation:

$$\theta_{ij}^{g}(n) = \frac{\sum_{l=0}^{n} (\psi_{ij}^{g}(l))^2}{\sum_{l=0}^{n}\sum_{j=0}^{m} (\psi_{ij}^{g}(l))^2},$$
$$i, j = 1,\ldots,m, \quad (2)$$

where $\psi_{ij}^{g}(l)$ is the value of a generalized impulse response function (GIRF, see fifth step below) following a one-unit shock to variable i on variable j at time t (Pesaran and Shin, 1998). We allowed the GFEVD sufficient lags (up to a year) to stabilize on the dynamic percentage of performance variation that is explained by a particular variable.

5. Finally, we quantify the magnitude and timing of the effects of each endogenous variable on performance by means of GIRFs. Based on all estimated reactions in the VAR model, the impulse response function estimates the net result of a one-unit "shock" to one variable on the performance variable relative to its baseline (its expected value in the absence of the marketing shock). GIRF uses the simultaneous-shocking approach (Dekimpe and Hanssens, 1999; Evans and Wells, 1983), in which the information in the residual variance-covariance matrix of Equation (1) is used to derive a vector of expected instantaneous shock values. The advantage of this approach is that it does not require selecting a temporal ordering among the variables

TABLE A.1

Unit Root Test, Granger Causality Test, and GFEVD for a Leading Snack Brand

	Unit Root Test	Granger Causality (*p*-Value)	Percent Dynamic Variation in Sales Explained
Sales	−5.92		
I. Brand Awareness			
Top-of-mind	−6.03	0.038	2.43
Advertising/promotion awareness	−8.16	0.048	1.95
II. Brand Attribution			
Occasions of use			
For on the go	−10.58	0.022	1.72
Lift in the afternoon	−11.19	0.020	2.31
Relaxing by yourself	−11.07	0.002	1.35
Entertaining	−10.98	0.032	1.01
Features			
Satisfying taste	−11.65	0.004	1.41
Taste I love	−6.45	0.012	1.02
III. Brand Action			
Regular/most often	−11.19	0.020	3.62
Purchase intent	−9.31	0.005	3.55
IV. Brand Attachment			
Brand I trust	−11.07	0.002	1.56
Overall liking	−10.14	0.001	4.13
Favorite brand	−11.29	0.043	1.63
Overall satisfaction	−9.71	0.004	1.35
V. Brand Accrual			
Good value for money	−10.18	0.031	1.21
BVS	−11.68	0.009	1.67

of interest. We derive the following three summary statistics from each GIRF: (a) the immediate performance impact on brand sales, (b) the permanent impact (that is, the value to which the GIRF converges), and (c) the total or cumulative impact, which combines the immediate effect with all significant effects. In the absence of permanent effects, this cumulative impact becomes the relevant metric to evaluate performance outcomes (Pauwels, Hanssens, and Siddarth, 2002). We do not report the statistical significance of this cumulative impact because we accumulate only the impulse response coefficients that are significantly different from zero; the absence of statistical significance is shown as zero cumulative impact. Finally, we obtain the wear-in time of each driver's effect on sales as the period with the highest (in absolute value) impulse response coefficient (Pauwels and Hanssens, 2007).

As an example of our approach, Table A.1 shows the results of the unit root tests, the Granger causality tests, and the GFEVD results for the snack product. For the unit root tests, we display the augmented Dickey-Fuller test statistic, for which values under −2.88 (−2.58) indicate a rejection of the null hypothesis (evolution) at the 95 percent (90 percent) significance level. For the Granger causality tests, we report the p-value of the F-statistic for those variables that are Granger causing sales at the five percent significance level. For GFEVD, we display the percentage dynamic variation in sales explained by the variable. The power of the VAR model to explain sales (R^2) was 0.89, representing a good model fit. In out-of-sample tests, the metrics produced by the VAR approach showed a better predictive ability than those obtained by reduced rank regression and by stepwise regression, two other frequently used analytical techniques (Pauwels and Joshi, 2008).