I. Current Status

Yoram Wind

The author reviews the current status and recent advances in segmentation research, covering segmentation problem definition, research design considerations, data collection approaches, data analysis procedures, and data interpretation and implementation. Areas for future research are identified.

Issues and Advances in Segmentation Research

Market segmentation long has been considered one of the most fundamental concepts of modern marketing. In the 20 years since the pioneering article by Wendell Smith [97], segmentation has become a dominant concept in marketing literature and practice. Besides being one of the major ways of operationalizing the marketing concept, segmentation provides guidelines for a firm’s marketing strategy and resource allocation among markets and products. Faced with heterogeneous markets, a firm following a market segmentation strategy usually can increase the expected profitability (as suggested by the classic price discrimination model, which provides a major theoretical rationale for the segmentation concept [35]).

Realizing the potential benefits of market segmentation requires both management acceptance of the concept and an empirical segmentation study before implementation can begin. The practical importance of segmentation research is reflected in the marketing literature and the actual practice of firms which often undertake segmentation studies and rely, at least to some extent, on the findings in their development and evaluation of marketing strategies.

Examine the current state of the art of segmentation research reveals some discrepancy between academic developments and real-world practice. Academic segmentation research has been one of the most advanced areas of research in marketing. Many of the new analytical techniques proposed in marketing have been applied to and tested in the segmentation area. Real-world segmentation studies, in contrast, have followed one of two prototypical research patterns with little creativity in design or analysis:

1. An a priori segmentation design in which management decides on a basis for segmentation such as product purchase, loyalty, customer type, or other factor. The survey results show the segments’ estimated size and their demographic, socioeconomic, psychographic, and other relevant characteristics.

2. A clustering-based segmentation design in which segments are determined on the basis of a clustering of respondents on a set of “relevant” variables. Benefit, need, and attitude segmentation are examples of this type of approach. As in a priori segmentation studies, the size and other characteristics (demographic, socioeconomic, purchase, and the like) of the segments are estimated.

Though segmentation studies have been dominated by these two prototypical designs, several major conceptual and methodological developments have been proposed in the academic literature. The advancement of market segmentation research requires, therefore, narrowing the gap between the academically oriented research on segmentation and the real-world application of segmentation research. Segmentation

---

1 See, for example, the selected bibliography on market segmentation recently published by the American Marketing Association [74].

2 The classification of segmentation studies into two categories, the a priori and post hoc, was suggested by Green [40].
studies should be reevaluated and new designs and analytical approaches considered. Also, the academic work on segmentation should be examined to reflect management's information needs, and to incorporate better the various separate developments in marketing theory (such as the change in the focus of analysis from a single product to an assortment of products [110]) and research (such as the development in overlapping clusters [19] or new measures of attribute importance [48]).

Achievement of these objectives requires a critical examination of the current state of the art in segmentation research, assessment of the major issues involved in the design and implementation of segmentation studies, and a perspective on the direction of future work. The purpose of this article is to provide such a review, for both consumer and industrial markets.3

Most segmentation studies have been conducted for consumer goods. Yet the concept of segmentation and most of the segmentation research approaches are equally applicable to industrial market situations [80, 106]. Hence, the intention of this review is to discuss the problems and perspectives of segmentation research in both consumer and industrial markets. The review, as summarized in Table 1, is organized on the basis of the five major phases of segmentation research—problem definition, research design considerations, data collection approaches, data analysis procedures, and data interpretation and implementation.

**PROBLEM DEFINITION**

The problem definition stage of segmentation research involves three major considerations: (1) managerial requirements versus the requirements proposed by the normative theory of segmentation, (2) the considerations of conducting a large-scale baseline segmentation study versus a continued series of ongoing segmentation studies, and (3) the specification of the desired segmentation model.

**Managerial Requirements**

Management acceptance of the market segmentation concept has resulted in the use of segmentation research to answer a wide range of marketing questions about the response of market segments to the firm’s marketing strategies (price changes, new product offerings, product changes, advertising themes, promotional efforts, and the like) as well as the selection of target market segments for each of the firm’s planned marketing offerings. The typical management questions which trigger and guide many market segmentation studies can be illustrated by the following objectives of several recently conducted concept-testing studies:

Which new product concepts evoke the highest respondent interest, and how do the evaluations of these concepts differ by respondent groups—heavy versus light product users, and users versus nonusers of the company's brand?

In terms of target markets for a new product concept, how do interested and disinterested respondents differ by demographic and socioeconomic characteristics, attitudes, and product usage characteristics?

Can the market for new product concepts be segmented in terms of the respondents price sensitivity (or other benefits sought), and what are the concept evaluations, attitudes, product usage, demographic, and other background characteristics of the various price-sensitive segments?

These and similar questions, which have guided most of the real-world market segmentation studies, have little to do with the type of questions suggested by the normative theory of market segmentation. Normative theory focuses on how information about customer characteristics and their relation to marketing strategy

---

3The term “industrial market” refers to all organizational markets [106].
variables should be used in the development and evaluation of a marketing strategy. Several models for the allocation of promotional resources to segments in the presence of uncertainty (both about the response to promotion by customers with given characteristics and about the degree to which each advertising medium reaches these customers) have been proposed [35, p. 186-241] and conceptually can be extended to other marketing strategy variables. Yet the normative segmentation models have rarely, if ever, been implemented. This discrepancy between the prescriptions of the normative theory and the practice of segmentation research can be attributed, at least in part, to the difficulties in operationalizing the segmentation theory. It is this difficulty in determining the needed data on both the response elasticities and reachability by media and distribution outlets that calls not only for undertaking the recommendations for further work suggested by Frank et al. [35, p. 249], but also for a critical reexamination of the nature and value of normative segmentation theory. Understanding management’s use of and difficulties in operationalizing segmentation [381], in both profit and nonprofit organizations, is thus a necessary first step toward the reevaluation of the normative theory of market segmentation. The recent work by Tollefson and Lessig [102] and Mahajan and Jain [70], and the new componental segmentation conceptualization suggested by Green and his colleagues [47] therefore represent promising first steps in this direction.

Baseline Versus Ongoing Segmentation Studies

Many segmentation studies have been designed as large, single effort, and expensive “baseline” studies. Such studies are justified in a variety of situations in which management would like to understand the basic structure of the market. Yet this type of study rarely has had subsequent effect on the other studies and strategies of a firm. Not uncommonly, after conducting a baseline study, firms continue to undertake other marketing studies such as concept and product testing with no regard for the concept of segmentation or the findings of the earlier study. This peculiar situation is partly due to habit—the continuation of past practices of analyzing most marketing research studies at the total sample level and not the segment level—and partly due to management disappointment with the relevance and operationalization of the guidelines provided by many segmentation studies.

To benefit fully from segmentation, whether or not baseline studies are conducted, management must employ the concept of segmentation in all studies, i.e., analyze the data of all studies at the segment level. Furthermore, any monitoring of the firm’s product performance (on profitability, market share, or growth, for example) should be undertaken not only at the total market level but at the segment level as well [115].

The Segmentation Model

The segmentation model requires the selection of a basis for segmentation (the dependent variable) as well as descriptors (the independent variables) of the various segments. The variables used as bases for and descriptors of segments have included all variables suggested in the consumer behavior literature. These variables can be divided into two types—general customer characteristics, including demographic and socioeconomic characteristics, personality and lifestyle characteristics, and attitudes and behavior toward mass media and distribution outlets, and situation-specific customer characteristics such as product usage and purchase patterns, attitudes toward the product and its consumption, benefits sought in a product category, and any responses to specific marketing variables such as new product concepts, advertisements, and the like.

In building an organizational segmentation model, the variables to be included should be not only the characteristics of the relevant organizational decision-making units (DMUs) but also organizational characteristics such as size and SIC. Both sets of variables include “general” and “situation-specific” characteristics.

Selection of variables for the model. In the selection of variables for the segmentation model, the major considerations are (1) management’s specific needs and (2) the current state of the marketing and consumer behavior knowledge about the relevance of various variables as bases for, and descriptors of, market segments.

Management needs are an obvious but somewhat neglected consideration. Conceptually, the bases for segmentation could vary depending on the specific decisions facing management. If, for example, management is concerned with the likely impact of a price increase on its customers, the appropriate basis might be the current customers’ price sensitivity. If, however, management is concerned with the loss of customers, the basis for segmentation could be actual or likely switching behavior, which in turn could be coupled with a secondary basis such as the reasons for switching. In contrast to the theory of segmentation that implies that there is a single best way of segmenting a market, the range and variety of marketing decisions suggest that any attempt to use a single basis for segmentation (such as psychographic, brand preference, or product usage) for all marketing decisions may result in incorrect marketing decisions as well as a waste of resources.

Over the years almost all variables have been used

---

4For a review of the specific variables that have been used in consumer and organizational segmentation studies, see [35].
as bases for market segmentation. Although no systematic and exhaustive evaluation of this experience has been undertaken, a consensus seems to have emerged that some variables are “better” than others as a basis for segmentation. Some of the author’s “preferred” bases for segmentation for consumer and organizational marketing decisions follow.

For general understanding of a market:
—Benefits sought (in industrial markets, the criterion used is purchase decision)
—Product purchase and usage patterns
—Needs
—Brand loyalty and switching pattern
—A hybrid of the variables above

For positioning studies:
—Product usage
—Product preference
—Benefits sought
—A hybrid of the variables above

For new product concepts (and new product introduction):
—Reaction to new concepts (intention to buy, preference over current brand, etc.)
—Benefits sought

For pricing decisions:
—Price sensitivity
—Deal proneness
—Price sensitivity by purchase/usage patterns

For advertising decisions:
—Benefits sought
—Media usage
—Psychographic/life style
—A hybrid (of the variables above and/or purchase/usage patterns)

For distribution decisions:
—Store loyalty and patronage
—Benefits sought in store selection

Common to these variables (and in line with the theory of segmentation which centers on differing elasticities to marketing variables) is the focus on various consumers’ (and organizational buyers’) responses to marketing stimuli. This emphasis on “situation-specific variables” is consistent with the needs of most marketing managers and current findings [35, p. 66–89; 53, 73]. The heavy reliance on benefits sought [56] (and, in organizational segmentation studies, on criteria used in the purchase decision) and the product or service purchase pattern reflects not only the author’s personal bias but also the emphasis of a considerable amount of current segmentation research.

Surprisingly, despite the attention given over the years to hierarchy of effects models and more recently to consumers’ information processing [7, 104], few of the nonpurchase variables, with the exception of benefit, need, and attitude segmentation, have served as a basis for segmentation. Similarly, despite the recent interest in the entire consumption process (including usage, maintenance, and disposal [79]), few of the consumption (as opposed to purchase) variables have been used as a basis for segmentation. Yet, for certain purposes, and especially for public policy decisions, one may want to consider some of these neglected aspects of the consumer decision-making process as the dependent variables in a segmentation model [81].

Whereas the selection of a basis for segmentation drawn from management needs is straightforward, the selection of variables as descriptors of the segments is more complex. This complexity stems from three factors:

1. The enormous number of possible variables. Most of the variables covered in the consumer behavior literature can and have at some time been considered as segment descriptors.

2. The often questionable link between the selected basis for segmentation and the segment descriptors. Segments with varying elasticities to marketing variables may not be identifiable in terms of demographic and other segment descriptors. Conversely, segments defined in terms of demographic and other general customer characteristics tend to be identifiable but may not have varying elasticities to marketing variables. The latter situation restricts management’s ability to pursue a segmentation strategy aimed at the given segments (and calls for examination of other possible bases for segmentation). The former situation allows management to follow at least a “self-selection” strategy [35, p. 176].

3. The question of actionability, which relates to management’s ability to use information (on the discriminating descriptors) as inputs to the design of the firm’s marketing strategy (e.g., product design, copy execution, media scheduling, distribution coverage, etc.).

The second guide to the selection of descriptor variables for the segmentation model is current understanding of the link between the considered variables and actual consumer responses to marketing actions. Surprisingly, despite thousands of academic and commercial studies by marketing and consumer researchers, one can draw very few generalizations as to which variables would have what effect under what conditions. This frustrating state of affairs is primarily due to four factors:

1. Lack of a systematic effort (by both academicians and practitioners) to build a cumulative body of substantive findings about consumer behavior.

2. Lack of specific models which link behavior (and other bases for segmentation) to descriptor variables and thus predict which descriptor variables should be used.5

3. The nonrepresentative nature of most of the academic studies with respect to sample design (e.g.,

---

5For a notable exception, see Blattberg et al. [10], who built a model of the deal prone household, linked the parameters of the model to household demographic variables, and thus predicted which variables should be used as descriptors.
small, convenient samples), type of respondents (e.g., students), and tasks (e.g., non-marketing-related tasks). Even many of the real-world segmentation studies are based on relatively small and nonrepresentative samples.

4. Lack of comparable conceptual and operational definitions of variables across studies.

Because of these limitations all that can be expected from the behavioral sciences and the consumer and organizational buying behavior theories and findings is an initial set of hypotheses on the likely relationship between consumer (and organizational) concepts (variables) and some specific marketing response. Following this approach, one could establish a set of hypotheses such as:

Consumer likelihood to buy new products increases:
- the higher the education, income, and occupational status [90].
- the higher the respondent scores on venturesomeness, cosmopolitanism, social integration, and general self-confidence [82].
- the lower the perceived risk [90].

Consumer likelihood to switch brands increases:
- the larger the number of stores shopped.
- the less the perceived quality differentiation among the brands.

Consumer likelihood to choose a low priced brand increases:
- the lower the confidence in price as an indication of quality.
- the lower the perceived social significance of the brand choice.

Consumer likelihood to buy private label brands increases:
- the higher the education and income of the respondent.

The likelihood of an organization to buy new products (and have a fast rate of adoption) increases:
- the greater the prestige of the organization.
- the larger the organization.
- the more specialized the organization.
- the more financially stable the organization.
- the more cosmopolitan the decision makers' orientation.
- the greater the risk-taking propensity and the lower the perceived risk.
- the younger the decision makers.
- the higher the entrepreneurial orientation of the decision makers.
- the greater the exposure to supportive mass media communications.
- the less formalized the decision process.

These and similar hypotheses can be derived from the current literature on consumer and organizational buying behavior. Such hypotheses can serve as guidelines for the selection of variables in the segmentation model. Subsequent segmentation studies could provide the vehicle for testing these hypotheses.

A relatively neglected area of study has been the examination of the specification of the segmentation models. A notable exception has been regression segmentation models. Examination has suggested that if these models are misspecified by omitted variables or functional forms different from the "true" process, they may be useful descriptors of the true purchase rate (or the group means for similar individuals), although the estimated model coefficient will be biased, except in special circumstances [5]. Beckwith and Sasiemi [5] developed misspecification tests for the regression segmentation models. It is desirable to use such tests whenever appropriate and to develop similar ones for other (nonregression) segmentation models.

Traditional a priori and clustering-based designs versus newer flexible and componental segmentation models. Four major types of models can be used in an effort to segment a market. They include the traditional a priori and clustering-based segmentation models, and the new models of flexible [91] and componental segmentation [47], which are based on different sets of assumptions. The selection of a model depends on management's objectives in undertaking a segmentation study.

A priori segmentation models have had as the dependent variable (the basis for segmentation) either product-specific variables (e.g., product usage, loyalty) or general customer characteristics (e.g., demographic factors). The typical research design for an a priori segmentation model involves seven stages.

1. Selection of the (a priori) basis for segmentation.
2. Selection of a set of segment descriptors (including hypotheses on the possible link between these descriptors and the basis for segmentation).
3. Sample design—mostly stratified and occasionally a quota sample according to the various classes of the dependent variable.
4. Data collection.
5. Formation of the segments based on a sorting of respondents into categories.
6. Establishment of the (conditional) profile of the segments using multiple discriminant analysis, multiple regression analysis, AID, or some other appropriate analytical procedure.
7. Translation of the findings about the segments' estimated size and profile into specific marketing strategies, including the selection of target segments and the design or modification of specific marketing strategy.

Clustering-based segmentation models differ from a priori models only with respect to the way the basis for segmentation is selected, i.e., instead of an a priori selection of a dependent variable, in the clustering-based approach the number and type of segments are not known in advance and are determined from the clustering of respondents on their similarities on some

---

6These hypotheses are based on the Rogers and Shoemaker review of the diffusion of innovation literature [92].
selected set of variables. Most commonly the variables used in the clustering-based models are needs, attitudes, life style and other psychographic characteristics, or benefits sought. Frequently the clustering procedure is preceded by a factor analysis designed to reduce the original set of variables. Other than the fifth step in the prototypical \textit{a priori} model—the formation of segments—and occasionally the specific sample design, the rest of the steps in the segmentation research design are the same.

Clustering-based segmentation models occasionally are combined with some \textit{a priori} bases. For example, the sample first can be divided into brand users and nonusers (or heavy and light users, and the like) and then the respondents in each \textit{a priori} segment can be clustered according to some other basis for segmentation such as needs, benefits, or the like. This and similar hybrid approaches such as Peterson's \textit{sequential clustering} [86] (i.e., clustering on demographic characteristics followed by attitudinal clustering within each demographic segment), or Blattberg and Sen's multidimensional approach [11, 12] address the problem of intrasegment heterogeneity. Hybrid approaches are particularly useful in organizational segmentation problems in which the first level of segmentation is based on organizational demographic characteristics (e.g., SIC, size) followed by product usage or criteria used in the purchase decision [114]. Despite their conceptual attractiveness, hybrid approaches do require relatively large sample sizes.

In both the hybrid and “pure” clustering models the researcher is confronted with a series of decisions about the selection of a clustering algorithm and the determination of the number of segments. Some of the major considerations in this area are discussed in a subsequent section.

\textit{Flexible segmentation.} In contrast to \textit{a priori} segmentation in which the segments are determined at the outset of the study, and clustering-based segmentation in which the selected segments are based on the results of the clustering analysis, the flexible segmentation model offers a dynamic approach to the segmentation problem. By this approach one can develop and examine a large number of alternative segments, each composed of those consumers or organizations exhibiting a similar pattern of responses to new “test” products (defined as a specific product feature configuration). The flexible segmentation approach is based on the integration of the results of a conjoint analysis study [52, 120] and a computer simulation of consumer choice behavior [84, 117]. Conjoint analysis studies usually consist of three major parts: (1) preference ranking or rating of a set of hypothetical products, each described as different levels on two factors at a time [63, 64] or as different levels on the full set of factors [46, 53] (the data from this task constitute the input to one of the nonmetric or metric conjoint analysis algorithms [16, 50, 67, 68, 95, 100] which produce a vector of utilities— for the various factors and levels—for each respondent); (2) perceptual ranking or rating of the current brands on the same set of attributes used in the preference task; and (3) a set of demographic and other background characteristics.

The simulation uses these three data bases as inputs and requires the active participation of management in designing a set of “new product offerings” (each defined as a unique combination of product features—specific levels on each of the factors included in the conjoint analysis study). Management participation can be on a real time basis in which managers interact directly with the computer simulation. Alternatively, management can specify in advance a number of plausible new product concepts, or react to a number of “best” product combinations.

The consumer choice simulator is based on the assumption that consumers choose the offering (new product or existing brand) with the highest utility. The simulator is designed to establish (1) the consumer’s share of choices among the existing brands, which can be validated against current market share data if available, and (2) the consumer’s likely switching behavior upon the introduction of any new product. This phase provides a series of brand-switching matrices. Within each matrix management can select any cell or combination of cells as a possible market segment (e.g., those consumers remaining with brand $i$ versus those who switched to new brand $j$, etc.). Once the desired segments (cell or cell combination) have been selected, the demographic, life style, and other relevant segment characteristics can be determined by a series of multiple discriminant analyses which can be incorporated in the simulation.

The flexible segmentation approach departs from the traditional \textit{a priori} and clustering-based approaches by offering management the flexibility of “building up” segments (cell or cell combination within any specific brand-switching matrix) based on the consumers' response to alternative product offerings (under various competitive and environmental conditions). In addition, having selected a segment, management has easy (interactive mode) information on

---

7Simulators have been developed not only on the basis of such 0,1 choice rules but also on the basis of a probability of choice: i.e., the utility for a given item is calculated as a percentage of the total utility for the brands the respondent considers in his or her feasible set. More sophisticated choice simulators also have been developed and are suggested elsewhere [50]. Experience to date (with five simulators) suggests that the simple simulator described above can reproduce accurately the market share of an existing set of brands. Yet further experimental work is called for to refine the computer simulations to take into account new features such as consumers’ “learning” of factors over time and as a function of the marketing efforts of the firm and its competitors, the description of new item profiles not in terms of deterministic levels for each factor but as a probability distribution over factor levels, etc.
the estimated size of the segment and its discriminating characteristics.

Componential segmentation. The componential segmentation model proposed by Green et al. [47] shifts the emphasis of the segmentation model from the partitioning of a market to a prediction of which person type (described by a particular set of demographic and other psychographic attribute levels) will be most responsive to what type of product feature. The componential segmentation model is an ingenious extension of conjoint analysis and orthogonal arrays to cover not only product features but also person features. In componential segmentation the researcher is interested in developing, in addition to parameter values for the product stimuli, parameter values for various respondent characteristics (demographic, product usage, etc.).

In a typical conjoint analysis approach to market segmentation, a matrix of subjects by utilities is developed. This matrix can serve as the input to the determination of the profile of some a priori segments (e.g., product users versus nonusers) or alternatively as the input to a clustering program which would result in a number of benefit segments [54]. In componential segmentation, the same design principles which guide the selection of (product) stimuli are applied also to the selection of respondents. For example, in a study for a new health insurance product, four sets of respondent characteristics were identified on the basis of previous experience and management judgment: age (under 50, 50–65, and over 65), sex (male, female), marital status (married, single), and current insurance status (have some health insurance with the given company, have health insurance with another company, and do not have any supplementary health insurance). Given these factors and levels, if a full factorial design were used one would have 36 possible customer profiles (3 × 2 × 2 × 3). Employing an orthogonal array design, one can use only nine combinations. Such a design requires the screening of respondents to select those who meet the nine profile requirements.8 Each respondent is then interviewed and administered the conjoint analysis task for the evaluation of a set of hypothetical health insurance products (also selected following an orthogonal array design of 18 combinations of five product features each at three levels). Having completed the data collection phase, the researcher would have a 9 × 18 matrix of averaged profile evaluations of the 18 stimuli by the nine groups of respondents. This data matrix then is submitted to the COSEG (componential segmentation) model [38] which decomposes the matrix into separate parameter values (utilities) for each of the levels of the product feature factors (comprising the stimulus cards) and separate parameter values (saliences) for each of the levels of the four customer profile characteristics (describing the respondents) which indicate how much each profile characteristic contributes to variation in the evaluative responses.

Given these two sets of parameters, the researcher can make predictions about the relative evaluation of any of the 243 possible product features (3^5) by any of the 36 respondent types. The five product feature utilities and four respondent saliences and the standard deviation of the predictor errors (as estimated for the 9 × 18 matrix) are used with a Monte Carlo simulator to estimate (1) for each respondent segment the frequency of first choices for each of the considered new product combinations and (2) for each new product combination the frequency of first choices across segments.

The COSEG model offers a new conceptualization for market segmentation because it focuses on the building blocks of segments and offers simultaneously an analysis of the market segment for a given product offering and an evaluation of the most desirable product offering (or positioning). The concept and algorithm of componential segmentation can be extended to cover not only two data sets (product feature and respondent characteristics) but three or more data sets by adding, for example, the components of usage situations.

The model and measurement of componential segmentation apply to both consumer and industrial market situations. In fact, the first commercial application of componential segmentation was conducted by Green and his colleagues for a firm engaged in the marketing of credit cards. This study was concerned with the design of a new credit card and the respondents were retail establishments selected on the basis of four sets of establishment characteristics (type of establishment, size of establishment, the establishment’s current favorite card, and the number of credit cards honored by the establishment) [40].

RESEARCH DESIGN

Research designs for a segmentation study depend primarily on the segmentation model used and the researcher’s creativity and skills. Each of the four segmentation models (a priori, clustering-based, flexible, and componential segmentation) requires a unique research design. Because the specific research designs for the four approaches are discussed elsewhere,9 this section covers some of the more general considerations involved in the design of any segmentation research.

---

8Care should be given in the survey stage to count carefully the incidence of each profile in the population. This step is essential if one wants to use the componential segmentation model not only for the identification of segments but also for the estimation of the size of each segment.

9Research designs for a priori and clustering-based segmentation are described in [35], whereas the research designs for flexible segmentation and componential segmentation are described in [91] and [47], respectively.
project (and which to some extent are also relevant for the design of other marketing studies). These considerations include nine major questions.

1. What is the most appropriate unit of analysis (e.g., an individual, a household, a buying center)?
2. How should the variables be defined operationally, and what is the most appropriate stimulus execution?
3. What is the most appropriate sample design?
4. How reliable are the data and how should one treat unreliable data?
5. How stable are the segments over time?
6. How homogeneous are the segments?
7. How segmentable is the market?
8. How can the results of a segmentation study be validated?
9. What are the cost considerations in the design of a segmentation study?

The Unit of Analysis

The marketing literature recognizes that most purchase and consumption behavior involves more than a single individual (the social context of, and influence on, purchase and consumption behavior). Yet most of the empirical market segmentation (as well as consumer behavior and marketing) studies ignore this premise and, with few exceptions, center on the individual as the sole unit of analysis.

This discrepancy between the “desired” and actual unit of analysis can be attributed primarily to the conceptual and methodological problems involved in moving from the individual to the multiperson situation. Such a move would require:

1. Identification of the relevant respondents—those who should be included in the unit of analysis. In a consumer segmentation study one should determine whether the unit of analysis includes the husband and wife, the mother and children, the husband, the wife, the entire family including both parents and children, etc. Similarly, in an organizational segmentation study one should determine which organizational members should be interviewed—the purchasing agent, the purchasing agent and intended product user, the controller, etc. [113]. The determination of the relevant respondents can either be made in advance or established empirically in the course of the data collection phase by snowball sampling procedures.

2. Determination of a multiperson dependent variable. One of the major problems in the analysis of multiperson data is the development of a dependent variable which reflects the decisions of two or more individuals. In this context two cases are trivial: whenever a single measure reflects the results of the preference and decision processes of two or more individuals (as in the case of the actual purchase of a single brand, excluding the case of multibrand purchases, by a household), and whenever the two or more individuals involved are congruent in their preference and choice behavior. There are no statistics on the actual incidence of these two cases. One can safely assume, however, that in many cases there is some degree of incongruency among the members of the buying center.

Analysis of this incongruency can follow two major approaches which differ in the level of analysis. The aggregate approach is based on analyzing separately the responses of the husbands and those of the wives and reporting any differences between the two groups (at the total sample or the segment level). This approach can be extended to cover children or any other members of the buying center. Although the easiest for analysis, this approach is conceptually undesirable because it does not solve the problem of intrahousehold (or intraorganizational) incongruency and its effect on the household (or organization) decision process and behavior.

The second approach is based on individual household analysis. In this case each household serves as the unit of analysis. For the simple husband-wife case, each household is described in terms of two vectors, one for the husband and one for the wife, and households are clustered on the basis of their similarities on the two sets of vectors. This approach was used in two recent studies. In one study among 120 couples in six metropolitan areas, life style attitudinal data were collected from both husbands and wives, and the 120 households were divided into six segments each representing a unique life style-attitudinal profile and different pattern of agreement between husband and wife with respect to the given attitudes. The small sample size does not justify delving into the substantive findings, but they suggest that the proposed approach is feasible. The second study on interest areas did not lead, however, to clearcut interpretable results.

The analysis of husband-wife pairs is relatively simple. Applying the same clustering approach to buying centers composed of three or more individuals is somewhat more complex, especially in the case of the household because households differ with respect to the number of children and adults present and the specific individuals involved in any given decision. In these cases a series of cluster analyses can be conducted. The total sample is divided into subsamples, each homogeneous with respect to the number of relevant decision makers (for example, a subsample of husbands and wives with no children, and another subsample of couples with one child, etc.). The data for each subsample are submitted to a separate cluster analysis to obtain a number of segments for each subsample. The resulting clusters are in turn clustered (with respect to some relevant criterion such as amount and type of household purchases) to obtain a set of higher order clusters which comprise segments from various subsamples.

Similar approaches also can be used in considering a segmentation of organizations. In these cases the first subsamples can be determined on the basis of the composition of the buying center, e.g., only purchasing agents, purchasing agent and user, etc.

3. Accounting for multiperson independent variables

Whether the dependent variable is a single or multiple measure, reflecting individual or multiperson choice

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
behavior, one can elect to use as explanatory variables the characteristics of a single individual, the characteristics of two or more individuals, or the characteristics of the household (or organization). Whenever the (demographic, psychographic, and other) characteristics of a single individual are included in the analysis, several multivariate statistical techniques (such as multiple regression analysis or AID) can be used to establish the nature and magnitude of the statistical association between the independent variables and the dependent variable.

If more than the characteristics of a single individual are included, one could treat these as another set of independent variables. Following this approach one could, for example, use the same analytical techniques as those used in the single respondent case (e.g., the brand choice of a household will be explained by using the husband’s psychographic characteristics and his wife’s psychographic characteristics). An alternative approach would be to rely on canonical correlation between the husband’s and wife’s response sets to reveal those items that are most contributory to husband-wife association. (Null cases could be set up by making random pairings of men and women.)

The third case involves the development of “group type” characteristics (such as group cohesiveness, autonomy, intimacy, polarization, stability, flexibility, etc.). This is the least developed area in both household and organizational studies despite some intriguing early developments in the measurement of group dimensions by social psychologists [60]. Once group type characteristics are developed, their analysis is straightforward and similar to the case in which one uses a single set of individual characteristics as the explanatory variable.

**Operational Definitions**

Having “passed” the managerial relevance and reasonableness considerations, the segmentation researcher should address the question of how to define operationally the dependent and independent variables. Developing operational definitions for selected variables is not a trivial task. Consider, for example, the situation of a clothing manufacturer who plans the introduction of a line with designs portraying musical instruments, notes, and personalities. A suggested basis for market segmentation is “music lovers.” But how could the variable be defined? Should it include persons who respond favorably to the statement, “I love music”? Those who go to classical concerts? Or perhaps those who go to rock concerts? Those who play a musical instrument? Or those who listen to an FM classical music station? Furthermore, how should these variables be measured? Should they be rated on some 2, 3, 4, 5, 6, 7, 8, 9, 10, or 11 point scale? Should they be ranked, assigned a constant sum value, or based on some free response data? Or should they be based on recall or on some diary-keeping procedure? It is evident that any of these or other possible definitions might result in different segment sizes and compositions. Such decisions also have a major impact on the nature of the required analytical procedures.

The importance of a careful operational definition of the dependent and independent variables of the segmentation model can hardly be debated. Yet in many segmentation studies little attention is given to an explicit evaluation of alternative operational definitions. Appropriate definitions and the testing of alternative operational definitions are especially crucial in psychographic research [107, 116] and in conjoint-analysis-based segmentation studies. The results of conjoint analysis are limited to the factors and levels included in the design. Omission of an important factor, exaggeration of the levels of any factor, or alternatively too small a difference among the levels of a given factor might lead to erroneous results.

Having defined the variables of interest and the respondent task, the researcher is still confronted with the question of stimuli execution. In benefit segmentation studies, for example, one can consider presenting the benefits (product features) as verbal descriptions, artist’s conceptions, color photographs of hypothetical products, or advertisements (or commercials) for the various benefits. A few studies have been conducted in this area, but little is known about the effect of alternative executions on consumer response and more work is needed.

**Sample Design**

The objective of any segmentation research study is not merely to explain differences among the specific respondents or to segment the sample, but rather to project the results of the study to the relevant universe. The projectability of results requires the use of an appropriate probability sample. Yet the great majority of segmentation studies rely on a quota sample. If quota samples or strict screening for specific respondent profiles (as required by the componential segmentation model) are used, the researcher must select the respondents for the screening interview on the basis of strict probability procedures and must keep accurate records of the screening data.

Another important consideration is the examination of any systematic differences between respondents and nonrespondents. Whereas nonresponse bias is widely recognized, most segmentation (and marketing) studies tend to ignore it. A growing body of evidence suggests that in many situations nonrespondents differ significantly from respondents. For example, a recent segmentation study for a bread product showed that the respondents to a personal interview had considerably higher bread consumption than those respondents who could not be reached in the first attempt and required 4–6 callbacks to recruit.

Because of the increasing cost of data collection
($30–60 per interview is not uncommon for a hour-long personal interview), the need for projectability, and the relatively small universe of many organizational buying firms, greater reliance should be placed, in both consumer and organizational buying segmentation studies, on rigorous telephone screening (which is less expensive) followed by interviews with a relatively small but select sample of respondents, and a small subsample of nonrespondents.

Data Reliability

Despite the importance of assuring that segmentation analysis is conducted on reliable data, little attention is paid to this issue in most segmentation (and marketing) studies, and the data analyzed are assumed to be reliable. Some variables (e.g., demographic characteristics) are more reliable than others (e.g., attitudes and psychographic characteristics). Whenever there is reason to suspect the reliability of some variables, certain safeguards such as test-retest reliability measures should be considered. With respect to attitude data, if test-retest data are collected, the degree of reliability can be considered explicitly in the factoring procedure.

Green [40] proposes a data collection and analysis procedure in which individual item reliability is used in the factoring process itself. Each attitude item is replicated across a group of respondents and the product moment correlation between the test and retest scores constitutes the measure of each item’s reliability. The factoring procedure is a modified principal components method in which the entries of the correlation matrix (the usual starting point of the factor analysis) are modified to give greater weight to the more reliable items. In this way the first principal components—those that are usually retained in subsequent rotation—are the most errorfree, and the discarded components tend to be those that are more errorful. This approach leads to a more meaningful factor structure because the more reliable items have a greater influence in the delineation of that structure.

This and other procedures to handle unreliable data should be considered in the design of segmentation studies because the implicitly assumed reliability may not in fact exist. Of special interest in this context are the developments in the study of generalizability [24]. Traditional reliability measures assume that the design of a measuring procedure has been fixed with the possible exception of the number of items, and that one wants a numerical index of the precision of the device (i.e., reliability coefficient). A generalizability study, in contrast, covers (by appropriate analysis of variance design) all possible causes of discrepancies among observations. To date, the generalizability test has not been applied in segmentation (or other marketing) studies. The concept is intriguing and should be explored in the instrument-development stage of segmentation studies.

Segment Stability

An often neglected aspect of segmentation research is the question of segment stability over situations and time; i.e., given the assignment of individual $i$ to segment $j$, how likely is it that the individual will remain in the same segment over time and different situations? The answer to this question depends on three sets of factors.

1. The basis for segmentation. In general one might hypothesize that the more specific the basis for segmentation (e.g., price sensitivity for or purchase of a given brand) the less stable the segment. Similarly, the more general the basis for segmentation (benefits sought from the product category or needs) the more stable the segment.

2. The volatility of the marketplace. Changes in the competitive activities and other environmental (political, legal, cultural, economic, etc.) conditions are likely to disturb the stability of the segments, and increase the likelihood of switching among segments.

3. Consumer characteristics. All consumers go through basic life cycle changes; even in the short term (within a life cycle stage), consumers may differ with respect to their likelihood to change and the nature of the change.

The specific variables which operate in each of these three sets of conditions must be identified, and the nature and magnitude of their impact on changes in the stability of various segments (e.g., buyers versus nonbuyers, different benefit segments, etc.) should be assessed.

Most of the current segmentation research efforts are based on cross-sectional data. In these cases the stability question is critical in determining the “aging” of the data, i.e., the length of time—two years? one year? six months?—during which the findings of a segmentation study are likely to hold. To date there are only a few unpurchased industry cases in which the same market was segmented two or more times during a period of a few years. In these cases (ranging from frequently purchased food products to insurance policies), surprisingly few changes have been observed in the estimated size and composition of various segments over a relatively short period of time (1–3 years).

Longitudinal data are applicable to an analysis of the stability of market segments at the individual level. Yet most of the segmentation researchers using such data have tended to ignore the stability question and instead have tried to develop measures such as brand loyalty, deal proneness, and the like, which reflect the dynamics of consumer behavior over a period of time. Whereas the value of such efforts can hardly

---

10This stability has been observed at the aggregate and not individual level. It might, therefore, be similar to the case observed by Bass [3] of stable market shares despite considerable brand switching activities within the market.
be debated, it is desirable to use the available longitudinal data to address also the question of segment stability over time and situations. In the few cases in which segment stability was examined (such as [11, 12]), segments based on behavior over time did in fact show stability over three or four years—the length of time for which data were available. Some stability of benefit segments over a two-year period has been found [14]. Similar stability of psychographic segments from two independent samples over a two-year period also was found in an unpublished study by Wells.

Segment Homogeneity

Segmentation studies commonly involve the determination of the segments (based on either a priori judgment or a clustering-based approach) followed by the identification of the segment profile on the responders' other characteristics. The latter stage is usually undertaken by examining the possibility of significant differences between segment means on a set of background variables.

Finding that two or more segments are different in terms of their mean profiles does not provide any indication about the possible segments within each segment. Members of a "buyer" segment, for example, may buy a given brand for different reasons. They may be very heterogeneous in their needs, demographic characteristics, and information requirements. In principle, almost every segment may, in turn, be decomposable into subsegments. Hence, to achieve intrasegment homogeneity, a very specific multidimensional definition of the basis for segmentation is required.

To fully understand the structure of selected market segments, one must understand the degree of homogeneity of each segment [43] and, in particular, be able to identify the number of subsegments, their relative size, and composition. The most common approach to the identification of subsegments is a hybrid segmentation approach in which the researcher searches for further subsegments among the a priori segments by clustering the members of each segment, or alternatively by applying an a priori segmentation scheme to each of the segments established by a clustering of respondents. Similarly one can use the sequential segmentation approach suggested by Peterson [86].

More recently, Green and his colleagues [44] suggested the application of Lazarsfeld's latent class analysis [69] and orthogonal array designs to the problem of segment partitioning. Latent structure analysis is concerned with the discovery and measurement of the underlying structure of phenomena (e.g., attitudes), when these underlying variables (attitudes) are only probabilistically related to the responses. In the segmentation context one can envision an aggregate multiway contingency table which displays association among two or more categorical variables (such as high and low product usage and high and low brand loyalty) which is decomposed into two or more subtables (based on some descriptor variables such as age, education, etc.)—the latent classes. The entries of the new subtables do not show association across the original qualititative variables; yet when the cell entries of the subtables are added, the original association reappears.

The operationalization problems associated with latent structure analysis [78] are solved by use of Carroll's CANDECOMP (canonical decomposition) model and algorithm [18] to find the size of each latent class and the estimated probabilities of occurrence for each level of each variable for multiway tables involving up to seven variables. The CANDECOMP model and algorithm were applied to a large study on the characteristics of adopters of a new telecommunications service and revealed managerially meaningful subsegments of adopters [44].

The Segmentability of the Market

Segmentation studies are based on the premise that the given market is heterogeneous and can be segmented. Most empirical segmentation studies support this premise. It is not uncommon, however, to find markets in which no significant differences are found among various segments with respect to their demographic or other relevant consumer characteristics such as response elasticities to marketing variables. In these cases it might be of value to be able to assess the degree of segmentability of a market.

A measure of the segmentability of a market recently was proposed in the context of the componential segmentation model [40]. Before application of the COSEG model the researcher can take the respondents' ratings of the set of product stimuli and submit them to a conventional two-way ANOVA. Using Hays' omega square measure [59], the researcher can obtain an index of the importance of the respondents, the stimuli, and the interaction between stimuli and respondents. If the $\omega^2$ for the subjects and the interaction effects (between stimuli and subjects) are small and the $\omega^2$ for stimuli are large, the market does not seem to be amenable to segmentation given the specific set of respondent characteristics and stimuli (product features).

Validation

One of the major discrepancies between the academic and commercial studies of segmentation is with respect to the question of validity. Whereas many academic studies of segmentation do employ some form of validation (primarily cross-validation on a

---

11 It is important to note, however, that if segments do not have different characteristics (such as demographic features) but do show differing elasticities to marketing variables, management still can use a self-selection strategy [35].
holdout sample), most of the commercially based segmentation studies ignore the question of validity.

The validity of segmentation research is by far the most crucial question facing management. Do the segments discovered in a segmentation study exist in the population? Is the estimated segment size accurate? And how accurate are the estimated segment responses to the firm’s marketing actions? Despite the centrality of such questions, only a few validation reports have been published [33, 72] and relatively few firms have made any effort to validate the results of their segmentation studies. Most notable among these efforts is the AT&T validation study of the segments within the residential long distance market. In this and several other unpublished AT&T studies the segmentation results were found to be very predictable of actual market behavior.

In the absence of such external validation, certain validation procedures could be used. The procedure chosen would depend on the specific segmentation model employed. In a priori and clustering-based segmentation studies, validation can be undertaken by splitting the sample and predicting segment membership of a holdout sample. In flexible and componential segmentation designs, one can check both the internal validity (e.g., prediction of the evaluation of control stimuli which were ranked or rated by the respondents but were not used in the parameter estimation procedure) and external validity (by comparing the survey’s share of choices with the actual share data for each market segment). External validation requires purchase data (e.g., market share) by segment, which are not always available. Similarly, data availability is a problem when one wishes to validate the nonpurchase characteristics of the segments identified in segmentation studies. It is essential, however, for any firm using segmentation research to design and implement follow-up research—either in a test market or the national market—to validate the findings of the segmentation study and the subsequent segmentation strategy.

Cost Considerations

The selection of a segmentation design cannot be done in isolation from cost considerations. Lip service often is given to the concept of cost and value of information. Yet most industry-based segmentation studies are launched with no explicit effort at determining the cost and value of the expected information. Regardless of the politics of setting marketing research budgets, the researcher has a major role in the cost decision because most of the research design considerations and the data collection and analysis procedures have cost implications. One can, for example, assure a higher reliability but it may involve somewhat higher costs. Explicit cost tradeoffs therefore should be made which lead to a selection of research designs based on the intended purpose, the expected information, and the costs involved.

Many design and analysis considerations have only minimal cost implications. In fact, whenever a researcher gives cost considerations as the reason for not using multivariate statistical procedures or for not undertaking some simple cross-validation procedure, one should be suspicious because many of these procedures do not involve higher costs. Furthermore, if a researcher is familiar with and has working knowledge of multivariate statistical techniques, their use in segmentation research can be considerably less expensive both in terms of computer costs and analysts’ time.

In most segmentation studies the cost of research design and analysis varies within very narrow ranges. The major variable cost component is data collection. Therefore, major cost savings can be expected from the data collection stage and not from cutting corners at the design and analysis stages.

Compared with the benefits believed to be associated with a segmentation strategy, the cost of segmentation research is often considered trivial. But few efforts have been directed at the assessment of (1) the expected value, in terms of alternative marketing segmentation strategies (versus the base case of no explicit segmentation effort) or (2) the expected cost of implementing different segmentation strategies. More work in this area is required.

DATA COLLECTION

The design and evaluation of alternative data collection procedures have received little attention in academically oriented market segmentation literature. Commercial researchers, however, have made significant contributions in this area, although relatively few innovative data collection approaches have been used in segmentation studies.

Primary Versus Secondary Data

Most of the commercially based segmentation studies have used a primary data collection effort and, in particular, cross-sectional surveys. In contrast, the technique-oriented academic segmentation studies tend to use available secondary sources and, in particular, panel data such as the Chicago Tribune and MRCA.

The development and offering of a psychographic data base for the MRCA panel has increased the commercial use of secondary sources in segmentation analysis. Increased usage of this and other secondary data sources (such as TGI or BRI) in segmentation analysis, and their integration with survey data, is desirable because it provides the longitudinal data required for dynamic segmentation analysis.

Other advantages of the greater reliance on secondary data sources are the better samples offered by most national panels, the greater attention (in comparison with the “typical” survey) to quality control of the data collection procedure, and the economic advantages of secondary sources (especially in light of
the recent increases in the cost of survey data collection). Of particular interest is the greater accuracy of panel purchase and usage data [118].

Another type of secondary data source of great potential value in segmentation analysis, for many products and services, includes all those data bases which are in some aggregate form rather than based on the individual (or household). For example, census demographic zip code data can offer a rich data base for segmentation analysis of zip codes. They are especially valuable for firms in the direct mail business. Two types of zip code segmentation efforts have been conducted on the basis of selected variables from the REZIDE tapes of the 35,592 U.S. zip codes [22].

1. “Tailor-made” segmentation efforts, on selected variables hypothesized to be related to the response to the specific firm’s offering, have been conducted by several firms, resulting in the selection of target zip code areas. Marketing efforts aimed at these segments have led to considerable improvement in response rates.

2. General segmentation of the U.S. based on 71 demographic and socioeconomic census variables was conducted by Computer Cartography, Inc. [23] and offered commercially. These data were factor analyzed and then clustered, resulting in 41 clusters which, in turn, were grouped in 12 major zip clusters. Each residential zip is within one of the 41 clusters (such as “upper class mobile professionals and college students in new towns”) and also in one of the 12 major clusters (such as “educated elite suburban white family areas” or “lower-middle to middle blue collar nonmetro areas”).

Similarly, the more recent developments of UPC data based on automated checkout data could offer a rich data base for “store area” segmentation analysis, especially when linked to electronic fund transfer payment data.

Conventional and Newer Procedures

Most segmentation studies are based on data collected in personal interviews or mail questionnaires. More recently, with the development of computer-based telephone interviewing, the telephone has become a feasible instrument for segmentation-type data collection. Such telephone interviewing can be very efficient because of response-dependent sequencing of questions, ease of reaching the sample units, and the ability to analyze the results as they come in.

In selecting a data collection method, therefore, one should consider a combination of approaches. The Human Population Laboratory [61] undertook a comparison of three data collection strategies containing personal interviews, telephone interviews, and mail questionnaires in different combinations (order of usage). Rate of return and rate of completeness were high in all three strategies. Similarly, substantive findings were virtually interchangeable and there was little difference in validity. The only difference was the cost per interview, which varied considerably by strategy. Accepting these findings suggests that substantial cost saving could be achieved in segmentation studies without sacrificing quality by designing a cost effective and task efficient mix of data collection procedures. However, other studies have shown differences in the response to, and validity of, various data collection procedures [15, 30, 31]. These differences can be accounted for by several factors. Self-administered questionnaires, for example, were found to be better when the information was available in records or possessed by other members of the respondent family. They were also better with highly educated and high income respondents [104].

The major new developments in data collection procedures include computer-terminal-based data collection procedures, two-way cable TV, and home computers. In addition to computer-based telephone interviews, computer terminals have been used for self-administered personal interviews in central interviewing facilities [80, 119] (such as shopping malls). With the advent of portable terminals, such interviews can be conducted at consumers’ homes or industrial buyers’ offices. Two-way cable TV has been presented for at least a decade as the panacea for data collection problems. Yet its development and dissemination are still more of a promise than a reality. Home computers are a newer development. The current pricing of microcomputers at less than $600 could have a major impact on their distribution and should not be ignored in development of new data collection procedures.

Other less frequently used data collection procedures for segmentation are (1) self-recorded protocol data (which, when content analyzed, can be used to identify the respondent’s decision process and serve as a basis for segmentation), (2) unobtrusive measures of respondents’ shopping, purchase, and usage behavior [105], and (3) projective techniques and other open-ended interviews. The major problem with using these data bases has been the difficulty in the content analysis of the protocols, open-ended responses, and some of the unobtrusive measures. If content analysis is required, care must be given to the development of a dictionary and other procedures commonly used in content analysis [37] and especially the development of interjudge reliability measures [66].

In selecting data collection procedure(s), one should give attention to the respondents’ ability to perform the task reliably, the cost in terms of money and time, and the scale properties required for the intended data analysis.

DATA ANALYSIS

The market segmentation field has been one of the primary and most prolific testing grounds for data analysis techniques. The diffusion of new data analysis techniques often has started with the publication of an academic article on the applicability of some analytical model and algorithm (usually borrowed from
mathematical statistics or mathematical psychology, and less frequently from mathematical economics and sociology). The procedure then is applied to a number of academically oriented marketing studies and finally is picked up by some of the more innovative marketing research firms. The diffusion of MDS and conjoint analysis provides good examples of this pattern.

To know the current state of applications of various data analysis techniques to market segmentation studies, therefore, one must continuously keep up to date with three areas:

1. The analytical techniques currently used in segmentation analysis (with a focus on possible problems associated with their use).
2. The analytical developments suggested in the recent marketing literature.
3. Analytical procedures which have not yet been introduced to the marketing literature and which offer some potential in the segmentation area.

The brief review of these areas is organized according to the three major types of analytical techniques used in segmentation research, namely, techniques for (1) classification of respondents into segments, (2) discrimination among the segments, i.e., the determination of segment profiles, and (3) simultaneous classification and discrimination.

Classification Procedures

Classification procedures for determining membership in market segments vary markedly according to the specific segmentation model used. The procedures most commonly employed in a priori approaches are sorting and cross-tabulation. Recently, Blattberg and Sen developed a Bayesian classification procedure [13] which they used successfully. When clustering-based segmentation is used, the most common procedures are clustering, MDS, and AID. Componential and flexible segmentation involve primarily conjoint analysis and computer simulation.

These and similar analytical techniques used for the assignment of consumers to segments are widely used and need no further explanation. The focus of this section, therefore, is on some of the possible pitfalls and unanswered questions involved in the use of these techniques, and some of the recent advances in classification procedures.

Sorting and cross-tabulation. The dividing line among a priori segments generally is arbitrary. Consider, for example, the commonly used segments of heavy and light brand users. How should the sample be divided? Should it be based on the top and bottom 50% of users or only the top 33% and the bottom 33%? Furthermore, should it be based on a number of items bought, frequency of purchase, total dollar

purchased, proportion of purchases devoted to the given brand, etc.? In many segmentation studies, too little attention is given to the conceptual implications of the various operational definitions and the sensitivity of the resulting segments to the definitions used.

A priori sorting of respondents can provide insight as demonstrated in the purchase segments of Blattberg and Sen [12]. It is judgmental and requires careful operational definitions of segment boundaries and the use of interjudge reliability measures among the judges who assign subjects to segments. Another neglected sorting procedure is the complete enumeration of all possible segments by following a set of assignment criteria. This procedure has been used in two cases.

1. The assignment of consumers to product assortment segments based on whether each of a set of products was used or not; i.e., if a product category has, for example, eight products, consumers are assigned to one of 256 (2^8) segments. In most product categories in which the product assortment patterns were analyzed, about 5% of the patterns accounted for 90% or so of the total purchases.

2. Benefit segmentation involving a relatively small number of benefits. If there are, for example, only five benefits and each individual can have either a high or low score on each benefit, there are 32 possible segments. Again, in the few cases in which such procedures have been used, a relatively small number of segments accounted for most of the respondents.

Of special interest in this context are the newer procedures for discrete multivariate analysis [8] which are mostly based on n-way cross-classification analysis. These procedures allow for nonlinearities and in many cases can be superior to the more conventional multivariate statistical techniques.

Clustering. A large number of clustering models and algorithms are currently available for the grouping of subjects. These techniques (including Q-type factor analysis and its modifications such as the linear typal analysis [26, 83]) differ in their objectives, type of clustering used, and the specific operationalization of the clustering procedures. The objective of clustering techniques as applied to market segmentation should not be limited to finding a typology of the market under study; one can use clustering also for data exploration, data reduction, and hypothesis generation.13

Clustering techniques can and have been categorized in a variety of ways. Hartigan [58], for example, suggested a classification based on mode of search into sorting, switching, joining, splitting, adding, and searching. For market segmentation purposes it might

---

12 For a basic review of these data analytical techniques, see [51].
13 Ball [1] has suggested three additional objectives—model fitting, prediction, and hypothesis testing. These objectives have rarely been used in a segmentation context.
14 For other classification systems, see for example [29].
be useful to distinguish between two major types of clustering techniques—those which build up clusters (a bottom-up approach) and those which break down a market into clusters (a top-down approach). The building up clustering techniques include several of the commonly used joining algorithms, such as the Johnson hierarchical clustering [65], Sneath single linkage algorithm [58], and various tree structures. Breaking down clustering procedures begin by partitioning subjects into two or more clusters, followed by subsequent partitioning of each of these clusters into subclusters. These techniques include a variety of splitting algorithms which, after an initial partitioning, obtain a new partition by switching objects from one cluster to another. The K means and Howard Harris algorithms are examples of this type of breaking down clustering procedure.

In view of the large number of available clustering [58] and pattern recognition [21] techniques, the critical question facing the researcher is which algorithm (and measures of similarity or distance) and type of measures (standardized, weighted, etc.) are most appropriate under what conditions and for what purposes. The current clustering literature does not provide satisfactory answers to these questions, and researchers tend to select an algorithm on the basis of such considerations as familiarity, availability, and cost of operations rather than appropriateness for a specific set of objectives and constraints.

Of special interest in this area are some of the recent developments in overlapping clustering. Clustering procedures used in traditional market segmentation studies were based on the premise of exclusivity; i.e., each individual could have been classified in one and only one cluster (segment). Conceptually, there are a number of situations in which a consumer can belong to more than a single segment (especially if one considers multiple brand usage, different usage occasions, multiple benefits sought, etc.). Several models and algorithms are now available for overlapping clustering [19] and can be used whenever it is conceptually desirable.

Multidimensional scaling. Most multidimensional scaling applications in marketing have been for product positioning [112]. Yet several MDS models and algorithms are suitable for the classification of respondents. Most notable among them is INDSCAL [17, 20] which develops for each respondent a weight on each of the relevant group dimensions. Respondents with similar salience on the perceptual dimensions can thus constitute a segment.

Segments also can be identified on the basis of the commonality of preferences either in a common (total market) perceptual (brand or product feature) space or in spaces unique to each perceptual segment. This grouping of respondents can be undertaken with several alternative models such as MDPref or PREFMAP [41, 50].

AID and stepwise regression. Although both of these techniques commonly are used to establish the profile of a segment (given some dependent variable such as purchase), occasionally the resulting segment descriptors (selected tree structures in AID, or the significant variables in the regression) can be used as a way of classifying the respondents into segments which, in turn, can serve as the basis for segmentation in subsequent analysis. Of particular value in this context is AID [98] which has been used successfully in a large number of commercial segmentation studies.

Newer approaches. More recently, multidimensional contingency table analysis [39] has been used to segment a market on the basis of the probability of being a “loyal” customer. This approach is utilized by Ogilvy and Mather under the name of logistic response analysis [25].

Discrimination Procedures

The analytical techniques used to profile the segments (which were selected a priori or identified in the classification stage) do not vary by type of segmentation model employed. To establish the segments’ profiles, market segmentation studies still occasionally include cross-tabulations and some univariate statistical technique. More commonly, however, the procedures applied are multiple regression, multiple discriminant analysis, AID, MCA (multiple classification analysis), or some combination of these techniques.

More recently new procedures have been suggested and applied to the problem of identifying segments’ profiles. Among these procedures are latent class analysis, logit and multinomial logit, multivariate probit, and nonparametric techniques.

Latent class analysis (as discussed in the section on segment homogeneity) utilizes the CANDECOMP model and algorithm. This approach was applied to the segmentation of the market for a new telecommunication service [44].

The multinomial logit model [42] has been applied to the same segmentation problem. It was preceded by an AID analysis (to determine which predictors, categories within predictors, and interactions should be considered). The logit model was applied to obtain parameter estimates for each predictor (or predictor combination) to test alternative models, and to develop probability estimates for the dichotomous criterion variable [42]. The AID/logit approach to segmentation can be extended to cover polytomous variables (and not only dichotomous data) by applying instead of AID the THAID algorithm [75] followed by a multinomial generalization of the logit model.

The multivariate probit model has been used recently in a pilot application to the segmentation area using the probability response coefficients (to each of a set of product features) as a basis for segmentation [88].

These and the more conventional discrimination
procedures, and in particular AID, THAID, and the AID-like algorithms such as SIMS (survey implement-
ed market segmentation), which is based on a simple cost-profit formulation [71], require relatively large data bases. Whereas this requirement does not create any special difficulty for most consumer market seg-
mentation studies, it does present a major stumbling block for most of the efforts to segment industrial
markets. Many industrial market situations involve a relatively small number of customers and require nonparametric statistical procedures and greater reliance on efficient experimental designs. More work on both areas is crucial to progress in industrial market
segmentation.

An interesting omission from the segmentation liter-
ature is the use of structural equation models (or path analysis) [9, 27]. Conceptually, at least, segmentation models can be constructed as causal diagrams or equivalent sets of linear equations that represent the assumed set of direct and indirect influences among a set of variables. The values of the unknown path coefficients can be estimated to indicate the relative strength of the determinants of the specified dependent variables (in effect a sequence of dependent variables is the recursive model). An excellent example of a segmentation-related problem in which this approach was used is a sociological study on the socioeconomic characteristics of occupational achievement [28].

**Simultaneous Classification and Discrimination**

Whereas most segmentation studies follow a two-
step procedure of first classifying respondents into
segments and then establishing their key discriminating
characteristics, attempts have been made to develop
and apply analytical techniques for the *simultaneous*
analysis of classification and discrimination. Canonical
correlation analysis has been applied in this context,
and more recently the flexible and componential seg-
mentation models.

**Canonical correlation** has been suggested as a meth-

od for segmentation by Frank and Strain [36] who
grouped respondents with regard to their cross-classi-
ified score categories on each significant canonical
variate in the predictor set. However, because canonical
correlation measures the shared variation between
two sets of variables (by transforming both configu-
rations to some best compromise position), it can be
used for segmentation as long as one does not want
to relate the predictor set to the criterion set. An
alternative procedure therefore can be the transforma-
tion of the space of predictor variables to best agree
with a fixed position occupied by the points in the
space of the criterion variables. It can be undertaken
by using any number of oblique or orthogonal Pro-
crustes transformations [6, 94].

If the researcher is not concerned with maintaining
the distinction of a criterion variable set, canonical
correlation or redundancy analysis [103] (a component
method which maximizes Stewart and Love's redu-
dancy index [100]) can be considered as a procedure
for simultaneously classifying the respondents and
determining the key discriminating profiles of the
various segments.

**Selection of a Plan of Analysis**

The choice of data analysis techniques reflects the
researcher's familiarity with, and preference for,
varying techniques. Several questions should be ad-
dressed in selection of a specific analytical plan. They
are not unlike the ones facing researchers using any
multivariate data analysis technique and include: How
should one measure the relative importance of the
variables—-is $R^2$, for example, an appropriate mea-
sure [48, 76, 108]? Should discrimination among segments be based on individual or group level analysis [2,
4]? How should one identify and measure the effect
of any response errors [34]?

Unique to segmentation studies is the need to apply
a variety of analytical procedures in tandem. Most
segmentation studies involve “complex” designs [35]
revolving around several hybrid bases for segmenta-
tion [57, 101]. However, because one cannot know
in advance which basis for segmentation will lead to
the identification of meaningful segments, segmenta-
tion studies should be flexible, allowing diverse analy-

des aimed at the identification of relevant segments.
This need creates special demands for researchers
with knowledge of a large number of analytical proce-
dures, good conceptual understanding of alternative
segmentation models, and a high level of research
creativity.

**DATA INTERPRETATION AND
IMPLEMENTATION OF RESULTS**

Regardless of how sophisticated the segmentation
study, the key to a successful project is the re-
searcher's and user's (management) ability to interpret
the results and use them as guidelines for the design,
external, and evaluation of appropriate marketing
strategy. The data interpretation stage should be per-
formed *jointly* by the researcher and user, reflecting
the researcher's statistical judgment (which dif-
ferences among segments are statistically significant,
etc.) and the manager's product/market knowledge.
In this context, two major issues are (1) how to
determine the number of segments and select the target
segment(s) and (2) how to translate the segmentation
findings into marketing strategy.

**Determining the Number of Segments and Selection
of Target Segments**

One of the most complex questions facing re-
searchers who use a cluster-based approach to seg-
mentation is the determination of the number of
segments. The two relevant criteria are the statistical
measure of cluster (segment) stability and homogene-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
ity, and the managerial consideration of cost of segmentation. The question of segment stability can be assessed by comparing the results of alternative clustering procedures and computing measures of similarities between the different cluster solutions. Estimating the cost of segmentation involves management’s subjective judgment.

Whatever segmentation approach is used, it is management that selects the desired target segments. The selection of target segments is a complex “art” type of process which should take into account such factors as the segments’ expected response to marketing variables, the segments’ reachability, the nature of competitive activity within the segment, and management’s resources and ability to implement a segmented strategy for the selected segments.

The cost of segmentation, the problems inherent in any effort to reach effectively a large number of segments, and the complexity of managing a large number of segments all encourage the selection of relatively few segments. However, greater segment homogeneity requires a larger number of segments [32]. To balance these two forces, more rigorous analytical models should be developed. Furthermore, the segment selection decision should not be limited to a single product but should encompass the more realistic and complex case of a product line.

A more fundamental question in the selection of target segments faces public policy decision makers. Whereas a profit-oriented firm can decide to concentrate its efforts on a specific target segment(s), the public policy tradition has been to develop a single policy aimed at all consumers. As public policy decision makers move toward greater reliance on consumer (and marketing) research as input to their decisions, they face an increasing conflict between the empirical findings (on the heterogeneous nature of the market) and the traditional concept of the “average” or “typical” consumers. If policy formulators could implement a segmented strategy, undertaking public-policy-oriented segmentation research projects would present no difficulty. If public policy decision makers were to insist on a single strategy aimed at an “average” consumer, the concept of segmentation might have to be modified.

**Translating Segmentation Findings into Strategy**

The most difficult aspect of any segmentation project is the translation of the study results into marketing strategy. No rules can be offered to assure a successful translation and, in fact, little is known (in the published literature) on how this translation occurs.

Informal discussions with and observation of “successful” and “unsuccessful” translations suggest a few generalizable conclusions aside from the obvious ones such as:

1. Involving all the relevant users (e.g., product managers, new product developers, advertising agency personnel, etc.) in the problem definition, research design, and data interpretation stages.
2. Viewing segmentation data as one input to a total marketing information system and combining them with sales and other relevant data.
3. Using the segmentation data on a continuous basis. The reported study results should be viewed only as the beginning of a utilization program.

Difficulties with, and the nature of, the “translation” of segmentation findings into strategy depend on whether the segmentation study is used as input to:

1. Idea and strategy generation or strategy evaluation.
2. Product related decisions (i.e., product positioning, design, and price) or communication and distribution decisions.
3. Decisions about existing products (i.e., no change, product modification, repositioning, or deletion decisions) or new products.

The translation of segmentation findings into new ideas (and strategies) is usually limited only by the creativity of the users. Most segmentation studies, and especially those based on consumers’ needs, benefits, life styles, or other psychographic characteristics, offer a rich profile of potential target segments which, in turn, can lead to the generation of a large number of diverse ideas and strategies. Furthermore, if one is concerned, for example, with design of a product or a communication campaign, each idea can be executed in a variety of ways, the success of which depends more on the creativity of the designer than on the segmentation findings.

The translation of segmentation findings is more complex when they are used to evaluate (rather than generate) some marketing strategy. In this context, two situations are distinct—consumer reactions to a new strategy (e.g., new concept or new commercial) and consumer satisfaction with the firm’s current products and services. In both cases a meaningful evaluation should be done at the market segment level, and in both cases there is a strong tendency to define segments in terms favorable to the corporate decision makers’ objectives.

A major difficulty in the translation of segmentation findings into actionable strategy is management’s perceived ability to implement the strategy. In industrial marketing, one often hears the argument that salesmen cannot handle simultaneously a number of strategies aimed at a number of segments. Furthermore, organizations, whether in consumer or industrial markets, are on the average reluctant to undertake high-risk strategies. Strategies which depart from current strategy or which require new ways of reaching target segments (e.g., direct mail vs. TV, new distribution outlets, etc.) are viewed as high risks. In providing rigorous input to decisions perceived as involving high risks, segmentation findings can have a major impact.
CONCLUDING REMARKS

Market segmentation has served as the focal point for many of the major marketing research developments and the marketing activities of most firms. Yet too many segmentation researchers have settled on a fixed way of conducting segmentation studies. This tendency for standardization of procedures is premature and undesirable. Innovative approaches to segmentation have been offered in the past few years, and further work on the new conceptual and methodological aspects of segmentation should be undertaken. Of particular importance seems to be research on the following areas:

1. New conceptualization of the segmentation problem.
2. Reevaluation and operationalization of the normative theory of segmentation, with emphasis on the question of how to allocate resources among markets and products over time.
3. The discovery and implementation of new variables for use as bases for segmentation (i.e., new attitudinal and behavioral constructs such as consumption-based personality inventories and variables which focus on likely change in attitude and behavioral responses to the marketing variables) of the markets for products, services, and concepts.
4. The development of new research designs and parallel data collection and analysis techniques which place fewer demands on the respondents (i.e., data collection which is simpler for the respondent and data analysis procedures capable of handling missing data and incomplete block designs).
5. The development of simple and flexible analytical approaches to data analysis capable of handling discrete and continuous variables, and selected interaction at a point in time and over time.
6. Evaluation of the conditions under which various data analytical techniques are most appropriate.
7. The accumulation of knowledge on successful bases for segmentation across studies (products, situations, and markets).
8. Undertaking external validation studies to determine the performance of segmentation strategies which were based on findings of segmentation studies.
9. Designing and implementing multivariate, multimeter approaches to segmentation research aimed at both the generation of more generalizable (reliable) and valid data.
10. Integration of segmentation research with the marketing information system of the firm.
11. Exploring alternative approaches to the translation of segmentation findings into marketing strategies.
12. Studies of the organizational design of firms which were successful and unsuccessful in implementing segmentation strategies.

Review of some of the issues and current advances in segmentation research indicates that, despite the great advances in the management of and research practice of segmentation, numerous frontiers still require creative and systematic study. This review and the other articles in this special issue of JMR are presented with the hope of stimulating further advances in this important and challenging area.

REFERENCES

17. ---, "Individual Differences and Multidimensional Scaling," in R. N. Shepard, ed., Multidimensional...
55. ———, ———, and ———. "Analyzing Free Response Data in Marketing Research," Journal of Mar-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


92. Rogers, Everett M. and F. Floyd Shoemaker. *Commu-
107. Wells, William D. "Psychographics: A Critical Re-

view," Journal of Marketing Research, 12 (May 1975), 196-211.