Socioeconomic Factors, Nutrition, and Food Choices: An Analysis of the Salad Dressing Market

Alan D. Mathios

Most studies utilize consumer surveys or experimental data to investigate the relationship between socioeconomic variables and food choices. The author examines actual purchase behavior (supermarket scanner data) in the natural shopping environment. The author focuses on whether the propensity to purchase high-fat unlabeled products within a product category varies across supermarkets with different demographic profiles. The results show differences in this propensity across demographic groups, even after accounting for price and sale promotions. The findings are consistent with consumer survey and experimental design research on the relationship between demographic characteristics and the processing of nutrition information.

Despite significant reductions in per capita consumption of fat, sodium, and cholesterol over the last 20 years, a significant number of Americans, especially certain demographic groups, still exceed dietary recommendations for these nutrients. Consequently, a large amount of research has focused on identifying and explaining the relationship between demographic characteristics and food choices. For example, researchers have utilized consumer surveys to examine both demographic differences in food choices and factors that influence food choices, such as nutrition knowledge, nutrition information processing skills, and label usage and comprehension (Adrian and Daniel 1976; Bender and Derby 1992; Gould and Lin 1994; Kim 1995; Lenehan et al. 1973, Putler and Frazio 1991). Although extremely useful, the use of consumer surveys to investigate these issues has some disadvantages. As Glanz, Hewitt, and Rudd (1992, p. 273) note, "Self-report measures may be compromised by non-response, high respondent burden, incomplete data, and social desirability. Whenever possible, studies should attempt to use simulations, observations, and/or environmental measures (e.g. sales data) on at least a sub-sample of respondents."

Experimental studies also have focused on the relationship between demographic characteristics and factors that influence food choices. For example, studies have examined demographic differences in the comprehension of nutrition labels, perceptions of health claims, and the effects of stimulus on utilization of nutrition information (Cole and Balasubramanian 1993; Mitra et al. 1995; Moorman 1990). Experimental studies of this type also have contributed greatly toward understanding the relationship between demographic characteristics and food choices.[1] However, these studies often abstract from time, price, and product selection constraints faced by consumers in the supermarket.

I take an alternative approach by examining actual food purchase behavior (supermarket scanner data) in the natural shopping environment. I examine whether, within a product category, the propensity to purchase high-fat products that do not contain nutrition labels varies across supermarkets with different demographic characteristics. Although the data cannot be used to disentangle whether the differences across demographic groups occur because of the nutrient content of the product or the absence or presence of a nutrition label, the results of this research can shed light on several issues related to the determinants of food choice. First, most other studies have linked demographic characteristics to nutrition information processing skills. This study can be used to evaluate whether these demographic characteristics are associated with different food choices, under circumstances in which consumers face similar shopping environment and sets of products from which to choose. Second, how consumers vary in their behavior with respect to products with undisclosed, undesirable attributes is of importance to policy issues in food and nutrition research, marketing research, and advertising regulation. Third, the results can serve as a baseline for further analyses of the Nutrition Labeling and Education Act (NLEA) (1990, 21 U.S.C. 301). Fourth, the study can help identify demographic groups who potentially have the most to gain from the NLEA, namely those consumers who have high propensities to purchase unlabeled high-fat products when choosing foods within a product category.

Demographic Characteristics and Food Choices

I utilize actual purchase data to examine the relationship between demographic factors and food choices. The majority

ALAN D. MATHIOS is an associate professor, Department of Consumer Economics and Housing, Cornell University. The author thanks Wegmans Supermarkets for generously providing the data for this project. He especially thanks Tom Dinardo, who patiently prepared the data reports and answered numerous questions about the data, and Ayda Yurekli, Margaret Coleman, and David Waigner for their research assistance. Pauline Ippolito and Debra Scammon provided helpful comments and the four anonymous reviewers helped to significantly improve the paper. A U.S. Department of Agriculture Hatch Grant provided financial support for this project.

[1]There are a host of nutrition information experiments conducted in supermarkets. See, for example, Russo and colleagues 1986 and Schucker and colleagues (1992).

of research in this area does not focus directly on the relationship between these variables but utilizes self-report and experimental data to examine the links between demographic characteristics and other measures of consumer behavior. For example, many studies focus on the association between demographic characteristics and nutrition knowledge, nutrition label usage, and nutrition information processing, all of which are likely to have impacts on food choice behavior. The literature review summarizes this research and focuses on how the results of these studies led to hypotheses regarding the association between demographic factors and the types of food choices examined here. The review is organized by the key demographic variables analyzed in the study.

Education

Education has been found, in a variety of settings, to be associated with information acquisition and healthy behaviors (Moorman and Matulich 1993). There is only a limited amount of research that has focused directly on the relationship between consumer characteristics and actual food shopping behavior (Glanz, Hewitt, and Rudd 1992). In their study of correlates of food shopping behavior, Fusillo and Beloian (1977) find that consumers who score poorly on a nutrition knowledge test are likely to be less educated, older, and male and to have lower incomes. Nutrition knowledge, in turn, was associated with more careful shopping behavior. Bassler and Newel (1982) also find that education is related to healthy food consumption.

Although there is relatively little research focused directly on food shopping behavior, there is a host of research that uses cross-sectional survey data to examine how nutrition information processing skills vary with education. Many of these studies focus on the relationship between education and the use and comprehension of nutrition labels. Bender and Derby (1992) utilize Food and Drug Administration Health and Diet Surveys to investigate the number of consumers who use ingredient and nutrition labels. They find that better-educated consumers are more likely to use nutrition labels. The Food Marketing Institute (1989), Klopp and MacDonald (1981), and Lenahan and colleagues (1973) all find education to be positively correlated with label usage. Moorman (1990) finds a positive correlation between education and information utilization, as well.

Research also has focused on the relationship between education and knowledge of diet-disease issues. Fullmer, Geiger, and Parent (1991) find that education is associated with greater comprehension of diet-disease messages. Ippolito and Mathios (1991) and Cotungna and colleagues (1992) find that people with higher levels of education had greater knowledge of the relationship between diet and disease.

Dietary recall data also have been used to address the relationship between demographic characteristics and dietary choices. Adrian and Daniel (1976), Ippolito and Mathios (1995), and Putler and Frazio (1991) all find that, after controlling for a variety of factors, well-educated consumers consume less fat than others.

Overall, the research concludes that education is correlated with superior information processing skills, greater use and comprehension of labels, and lower levels of fat intake. All these effects suggest that compared to less-educated consumers, well-educated consumers choosing foods within a product category are more likely to choose the low-fat products that contain nutrition labels. This suggests the following hypothesis:

> H_1: The propensity to purchase high-fat unlabeled products within a product category should be lower in supermarkets where shoppers have higher levels of formal education.

Income

Household income may indicate human capital beyond that given by formal education, and thus, may reflect greater efficiency in processing nutrition information. However, higher income may reflect higher opportunity cost of time, which may reduce time spent seeking out nutrition information. Household income also reflects the consumers' financial ability to purchase health. How this effects the decision to purchase healthier foods depends on the relative price of these foods.

There are several studies that link household income with health-information acquisition and healthy behaviors. The results are not as consistent as those found for education, though in most cases, studies find either no effect or a positive effect. In an exhaustive review, Moorman and Matulich (1993) examine the relationship between income and health behaviors. Of the relationships examined, 23 showed a positive or no effect of income on health behavior. Only 9 found a negative relationship.

In some studies, income has been found to be correlated with the use of a nutrition label (Fusillo and Beloian 1977; Lenahan et al. 1973). Cotugna and colleagues (1992) find that income is positively related to knowledge of the link between diet and disease. In a study of the food label reading habits of women with infants and children receiving food assistance (WIC clients), Michel and colleagues (1994) find that even low-income women who report concern about diet and nutrition have a low level of nutrition knowledge. Similarly, Gould and Lin (1994) find that household income has a positive impact on health knowledge.

Studies of dietary recall show mixed results with respect to the relationship between income and fat consumption. Adrian and Daniel (1976) find a positive but declining effect of income on fat intake. (It should be noted, however, that these results are based on data from over 20 years ago.) Gould and Lin (1994) find a positive impact of income on fat consumption, though they find that income elasticities with respect to fat consumption are low.

Overall, the research linking household income to label use, nutrition knowledge, and fat consumption show mixed results. However, taken together, we believe that the majority of empirical research supports the following hypothesis:

> H_2: The propensity to purchase high-fat unlabeled products within a product category should be lower in supermarkets where shoppers have higher incomes.

Age

There are several studies that focus on the relationships among age, nutrition information acquisition, and food choices. The results on age are somewhat mixed, but many studies find a negative correlation between age and nutrition information processing ability. Cole and Balasubramanian (1993) find that elderly consumers do not use nutritional information as accurately as do young consumers and point to age-related changes in information processing skills (Cole and Gaeth 1990) as a potential explanation. Similarly, Fusillo and Beloian (1977) find that elderly consumers have more difficulty comprehending nutrition labels. Bender and Derby (1992) examine the characteristics of those who pay attention to ingredient lists and nutrition labels. They find that consumers who use both types of label information are more likely to be young (25–34 years of age). In a study of nutrition knowledge of older and younger elderly in rural areas, Fischer and colleagues (1991) find that the younger elderly had higher levels of knowledge about fat and sodium and tended to make more healthful food selections. In a study of the influence of individual differences on the quantity and content of information recalled from product labels, Heroux, Laroche, and McGown (1988) find that the most important covariate related to the amount of information recalled from product labels is age of the respondent. They find that elderly subjects recall less information than do young subjects. Moorman (1990) finds a negative effect of age on utilization of nutrition information.

Other studies, however, find less conclusive effects of age. In their examination of health motivation and health ability, Moorman and Matulich (1993) find that age affects health maintenance behaviors positively and health-information acquisition negatively. In their review of the literature, they also find mixed results with respect to the relationship between age and health behaviors. Klopp and McDonald (1981) find no relationship between age and label use, and Gould and Lin (1994) find no significant correlation between age and health knowledge.

In summary, most of the research on information comprehension points toward a negative relationship between the ability to process nutrition information and age. The research on the relationship between age and health behavior, including fat consumption, is less conclusive. Taken together, we believe that consumer survey and experimental research in this area lead to the following hypothesis regarding the link between actual food purchase behavior and age:

H_3: The propensity to purchase high-fat unlabeled products within a product category should be greater in supermarkets with older shoppers.

Gender

There is some evidence to suggest that female shoppers read nutrition labels more often than do male shoppers (Bender and Derby 1992; Fusillo and Beloian 1977). In a survey by the Food Marketing Institute, 38 percent of women read nutrition labels, compared with 29% of males. Nonworking women were the most likely of any group to read ingredient lists or nutrition information.

Studies of dietary intake show that women consume significantly less fat than do males, though this result may stem from consumption of fewer calories (Ippolito and Mathios 1994). Consequently, it is difficult to assess how this differential fat consumption reflects nutrition information. Many studies of nutrition information acquisition do not focus on gender, though there are many studies that focus on gender differences with respect to other consumer behavior. Moorman (1990) and Moorman and Matulich (1993), in studies designed to evaluate the effect of consumer characteristics on information acquisition, do not evaluate the effect of gender. The research that has focused on the relationship between gender and nutrition information acquisition led to the following hypothesis:

H_4: The propensity to purchase high-fat unlabeled products within a product category should be lower in supermarkets with a greater proportion of female shoppers.

Data

I utilize data collected as part of a long-term project designed to evaluate the impact of mandatory disclosure requirements on food purchases. The data consist of two major components: Wegmans Supermarket data and information from nutrition labels.

Wegmans Supermarket Data

Wegmans is an upscale supermarket chain, located in New York State, that owns and operates approximately 50 stores, mostly in the upstate region. Wegmans provided two sources of data for this study: demographic data and product-movement data.

Demographic Information from Shopper Club Membership

Wegmans provided demographic information obtained from Wegmans Shopper Club (WSC) application forms[2] for 20 supermarkets. These forms include questions regarding the education, household income, age, family size, and sex of the shopper. The data are provided at the aggregate store level (i.e., the percentage of club members at a particular store that have a college education).

The demographic information for the 20 stores is presented in Table 1. The data indicate that the variation in education and income across the stores is large. For example, across the supermarkets, the percentage of shoppers that have a college degree ranges from 14% to 66%. Similar levels of variation are found for income.[3] There is a high correlation between income and education (.73). This may make it difficult to identify, within a regression model, the independent effects of each of these factors, and it raises

[2]The incentive to join the WSC is that most store price promotions only apply to WSC members.

[3]There are some disadvantages of using of store-level demographic data. First, not all shoppers are club members. Second, even for club members the card user is not necessarily the person who filled out the form. Third, these data are not continuously updated to match the weeks for which scanner purchase data are collected. However, the demographic profile of club membership is likely to be stable over one year's time (the time span used in this analysis). These problems are likely to make it difficult to identify a relationship between demographic characteristics and scanner purchase data. Yet, if strong links between demographic characteristics and purchases are found, it is likely that these data problems are not very severe.

Table 1. Demographic Characteristics by Store

STORE	% Graduate College	% Female	% Over 35 Years	% Income > $30,000
STORE 1	35	79	70	72
STORE 2	44	77	69	69
STORE 3	27	76	67	59
STORE 4	23	77	65	64
STORE 5	66	77	72	80
STORE 6	53	77	74	83
STORE 7	21	75	61	57
STORE 8	40	73	53	61
STORE 9	47	76	66	76
STORE 10	29	78	68	79
STORE 11	26	76	61	75
STORE 12	34	74	70	58
STORE 13	59	75	71	75
STORE 14	36	77	58	75
STORE 15	29	83	60	54
STORE 16	25	77	67	52
STORE 17	57	79	76	78
STORE 18	14	70	50	36
STORE 19	23	81	65	56
STORE 20	18	69	58	33

Notes: Percentages are based on Wegman's Shopper club card member application forms for each store.

concern about the robustness of the estimates (see the Results section).[4]

Scanner Purchase Data

From October 1992 to the present, Wegmans has provided product-movement data for a large number of products. Every four months, a product-movement report is made available that includes the number of units of each product sold during the previous week, the average retail price of each product during that week, the universal product code and size of the product, and whether there was a general price or a WSC promotion for that product.

Data were collected for a variety of packaged products, but I focused on salad dressings for several reasons. First, prior to the NLEA, only about half of the salad dressing products contained voluntary nutrition labels, making it an ideal category for looking at differences across groups.[5] Second, despite large declines in the per capita consumption of fat from meat in the United States between 1977 and 1985, there was a large increase in fat consumption from the category of salad dressings, sauces, and gravies.[6] Third, most salad dressings are sold in either 8-, 16-, or 24-ounce bottles, making it a relatively homogenous product to consider for analysis.[7] Fourth, for almost every type of salad dressing (e.g., French, Italian) and for most brands (e.g., Kraft), there is a labeled and unlabeled version of the dressing.

A weekly product-movement report was generated every fourth month beginning in the last week of October 1992. Product-movement data were collected for the last week in October 1992, February 1993, June 1993, October 1993, and so forth. I utilize only the October 1992, February 1993, and June 1993 data, because the focus is on pre-NLEA purchases.

Information from Nutrition Labels

The second major phase of the data collection effort consisted of collecting information on which products, prior to the NLEA, contained a nutrition label.[8] Information collected includes whether there was a nutrition label, and if so, the nutrient content of the food product. This data collection stage began October 1992 and ended March 1993. Data were collected from the local Ithaca Wegmans store.

Merging the Demographic, Scanner Purchase, and Label Data

Merging the scanner purchase data with the nutrition label data provides a data set in which purchase data is linked with nutrition label information. For each of the 20 stores, there are well over 100 salad dressings for which scanner data have been successfully linked with nutrition label information.[9] For standardization purposes, the analysis is limited to 8-, 16-, or 24-ounce bottles of dressing (these account for the majority of all sales). Although there are slight variations in which dressings are sold in which stores, the vast majority of salad dressings are available in every store during each of the time periods included in the analysis.[10] For analysis purposes, I restrict the data to those salad dressings that are in over 95% of the supermarkets.[11]

[4]Correlation between independent variables does not cause bias in the estimates. Rather, it typically results in high standard errors, making it difficult to obtain significant t-statistics. As is seen in the Results section, education and income are often both significant. This indicates that there is enough independent variation in these variables to identify these effects.

[5]The proportion of salad dressings with a voluntary nutrition label are consistent with the data reported in Caswell's (1992) study.

[6]See Ippolito and Mathios (1995) for a more detailed description of these results.

[7]A notable exception is Good Seasons Salad Dressings. These are powder dressings to which the consumer adds his or her own oil and vinegar. Almost all of these products did not contain a nutrition label prior to the NLEA. Although these products were not included in the results presented here, including them did not change the results in any important fashion.

[8] Label data were collected for virtually all packaged food products for sale in the local Wegmans. However, for this study, the products are, of course, limited to salad dressings.

[9]Our analysis is performed only on salad dressings for which scanner data and label data are both available. There are several salad dressings for which there are scanner data but no label data. This can occur for two reasons: First, Wegmans can sell a product that is not sold in the store in which the label data was collected. (Recall that the label data was collected only from the Ithaca Wegmans.) Second, if a new product was introduced after the data from labels were collected there will be scanner data but no label data. Overall, approximately 80% of total scanner data sales are accounted for by products for which label data are available.

There also can be label data on a dressing with no scanner data available. This can occur if the Wegmans in which the labels were collected carried a product that is not in any of the 20 other stores (the Ithaca store was not one of the 20). Moreover, it is possible that Wegmans could classify certain types of dressings in an alternative product code group, in which case it would not appear on the product-movement output. The impact of this type of data omission is likely to be small, because the vast majority of products for which label data are available also has scanner data.

[10]The maximum number of stores a salad dressing can be available is 60 (20 stores in three time periods).

[11]The results are not sensitive to this criterion.

The number of observations available for study equals the number of salad dressings for which both label data and scanner data are available, multiplied by the number of stores that sell that salad dressing, multiplied by the number of weekly product movements examined. This results in 4335 observations based on 20 stores, three time periods, and approximately 73 salad dressings.

Econometric Specification

Empirical Model at the Salad Dressing Level

Propensity to Purchase Labeled Products

One method to evaluate and assess whether consumers have different propensities to purchase high-fat products without nutrition labels when choosing foods from a product category is to specify an equation for the determinants of the number of purchases for each salad dressing. Equation 1 is designed to capture the important determinants of purchases. The estimation of Equation 1 uses data that vary by product (different salad dressings), store (20 different Wegmans stores) and time (three different weeks of data). Each variable contains subscripts indicating on which dimension it varies, and the normally distributed error term (e) varies by all subscripts.

$$(1)\quad NSOLD_{ijt} = a_1 + a_2\,EDUC_j + a_3\,FEMALE_j + a_4\,AGE_j + a_5\,INC_j + a_7\,HSIZE_j + a_8\,NOLABEL_i + a_9\,NOLABEL_i{*}EDUC_j + a_{10}\,NOLABEL_i{*}FEMALE_j + a_{11}\,NOLABEL_i{*}AGE_j + a_{12}\,NOLABEL_i{*}INC_j + a_{13}\,NOLABEL_i{*}HSIZE_j + a_{14}\,PRICEPER_{ijt} + a_{15}\,SALE1_{ijt} + a_{16}\,SALE2_{ijt} + a_{17}\,SMALL_i + a_{18}\,MEDIUM_i + a_{19}\,PFIEFFER_i + a_{20}\,WISHBONE_i + a_{21}\,KRAFT_i + a_{22}\,TOTSOLD_{jt} + e_{ijt},$$

where

$NSOLD_{ijt}$ = the number of units of salad dressing i sold in store j during week t;[12]

$EDUC_j$ = the percentage of the WSC members at store j who are college graduates;[13]

$FEMALE_j$ = the percentage of the WSC members at store j that are female;

AGE_j = the percentage of the WSC members at store j that are over the age of 35 years;

INC_j = the percentage of the WSC members at store j that have income above $30,000.

$HSIZE_j$ = the percentage of the WSC members at store j that have a household size greater than three persons;

$NOLABEL_i$ = 1 if salad dressing i is a high-fat unlabeled product and = 0 otherwise;[14]

$PRICEPER_{ijt}$ = the average retail price per ounce for salad dressing i in store j during week t;

$SALE1_{ijt}$ = 1 if salad dressing i in store j during week t is on sale to all shoppers and = 0 otherwise;

$SALE2_{ijt}$ = 1 if salad dressing i in store j during week t is on sale to Wegmans shoppers and = 0 otherwise;

$SMALL_i$ = 1 if salad dressing i is 8 ounces and = 0 otherwise;

$MEDIUM_i$ = 1 if salad dressing i is 16 ounces and = 0 otherwise;

$PFIEFFER_i$ = 1 if salad dressing i is the Pfieffer brand and = 0 otherwise;

$WISHBON_i$ = 1 if salad dressing i is the Wishbone brand and = 0 otherwise;

$KRAFT_i$ = 1 if salad dressing i is the Kraft brand and = 0 otherwise; and

$TOTSOLD_{jt}$ = the total number of salad dressings sold from store j during week t.

Equation 1 is estimated with ordinary least squares. The dependent variable is the number of units of salad dressing i sold in supermarket j at time t. This variable is explained by the price of that salad dressing at supermarket j at time t, whether a particular dressing is on sale at that store during that time period, the demographics of the supermarket, whether it lacked a nutrition label (which also indicates whether it is a high-fat product). This variable interacted with the demographic variables.

My central hypotheses can be evaluated by examining the coefficients on the variables that are interacted with the variable NOLABEL. The coefficient on the variable NOLABEL*EDUC, for example, provides an estimate of the degree to which high-fat products without a nutrition label differentially affects purchases across education levels, ceteris paribus. According to H_1, I expect the coefficient on NOLABEL*EDUC to be negative. That is, high-fat products without nutrition labels should be purchased less frequently in supermarkets with highly educated consumers. Similarly, according to H_2 and H_4, I expect the coefficients on NOLABEL*INC and NOLABEL*FEMALE to be negative. According to H_3, I expect the coefficients on NOLABEL*AGE to be positive. That is, elderly shoppers should be more likely to purchase high-fat products without labels than should young shoppers.

Equation 1 accounts for prices, the percentage of the shoppers at a supermarket that have a household size greater than three, whether the product had a special promotion, brand-specific effects, size of package effects, as well as a control for the total number of products sold at the particular supermarket during the relevant week. It is expected that SALE1 and SALE2 have positive coefficients and PRICEPER has a negative coefficient. The variable TOTSOLD is included to capture demand differences across the stores.

[12]Equation 1 is also modeled with the market share of each salad dressing at supermarket j at time t. The results, in most cases, are similar qualitatively. These results are presented subsequently along with those from Equation 1 as specified..

[13]The demographic data do not vary with the time period because these data were provided by Wegmans at the beginning of the first wave of data.

[14]Post-NLEA data indicate that almost every unlabeled product included in the study contained significant amounts of fat. Consequently, the variable NOLABEL is an indicator of a high-fat product, as well as one without a nutrition label.

The coefficient on TOTSOLD is expected to be positive: The greater the total number of salad dressings sold in a supermarket, the higher the purchases of any product that is part of that total. There is little basis to hypothesize about the signs of the remaining coefficients.

Finally, I also consider an alternative version of Equation 1, identical to Equation 1 except that the dependent variable is NSOLD/TOTSOLD. In this case, I examine the effect of the independent variables on the market share of a particular salad dressing rather than on the number of units sold.

Empirical Model at the Store Level

The estimation of Equation 1 includes multiple observations from the same store, yet does not account for store specific effects. Because fixed effects are not possible in this model (the demographic variables of interest are at the aggregate store level and thus do not vary within a store) I estimate an alternative empirical model to evaluate the consistency of the results across model specification. In this model, the unit of observation is the supermarket. This model focuses on the percentage of salad dressing purchases at a supermarket that are accounted for by unlabeled high-fat dressings. Because the unit of observation is the supermarket, there are only 60 observations (20 stores for three time periods) as opposed to the thousands of observations used to estimate Equation 1.

$$(2) \quad RATIO_{jt} = a_1 + a_2\,EDUC_j + a_3\,FEMALE_j + a_4\,AGE_j + a_5\,HSIZE_j + a_7\,INC_{jt} + a_8\,RPRICE_{it} + a_8\,SALEL_{jt} + a_9\,SALENL_{jt} + e_{jt},$$

where

$RATIO_{jt}$ = the percentage of total salad dressing sales that are accounted for by high-fat unlabeled products in store j during week t;

$RPRICE_{it}$ = the market share weighted price of unlabeled products in store j during week t divided by the market share weighted price of labeled products in store j during week t;

$SALEL_{jt}$ = the percentage of labeled products at supermarket j during week t that are on sale; and

$SALENL_{jt}$ = the percentage of unlabeled products at supermarket j during week t that are on sale.

Because of H_1, we expect the coefficients on EDUC to be negative. In other words, the share of salad dressing sales accounted for by high-fat unlabeled products should be lower in highly educated supermarkets populations. Likewise, because of H_2 and H_4, the coefficients on INCOME and FEMALE are expected to be negative. According to H_3, the coefficient on AGE is expected to be positive. The coefficients on SALEL and RATIO are expected to be negative, because a sale promotion for a labeled product or a lower relative price for a labeled product should cause a decline in the proportion of sales accounted for by unlabeled products. Under analogous reasoning the coefficient on SALENL is expected to be positive.

Results

I first provide descriptive statistics, then regression results, and finally a discussion of the results.

Descriptive Statistics

In Table 2, I provide the means, standard deviations, minimums, and maximums for each variable used in the regression analyses.[15] The average of NSOLD equals 14.88, indicating that the typical salad dressing is purchased approximately 15 times per week. The average of TOTSOLD equals 1391, indicating that the typical supermarket in the sample sells close to 1400 salad dressings per week. Finally, 44% of salad dressings sold in these stores are high-fat products without labels. Table 3 includes data that help illustrate the relationship between sales of high-fat unlabeled salad dressings and one of the key socioeconomic factors: The rows of the table are arranged in order of educational attainment, with the supermarket with the lowest-educated population listed first. It is clear from Table 3 that there is a correlation between education and the share of sales accounted for by high-fat unlabeled salad dressing sales. In each of the 10 supermarkets with the lowest-educated populations, the share of total dressing, sales accounted for by high-fat unlabeled products is equal to or exceeds 50%. This is true in only 3 of the 10 supermarkets with the most-educated populations. Moreover, the share of total dressing sales accounted for by high-fat unlabeled products in the supermarket with the highest-educated population is 42%, compared with 66% for the supermarket with the lowest-educated population. Similar results are found for income. Of course, there are other factors that might account for these differences.

Regression Results at the Salad Dressing Level

In Table 4, I present the regression results for Equation 1. The first two columns apply to the equation with NSOLD as the dependent variable, and the second set of columns applies to the market share specification (NSOLD/TOTSOLD).

There are several factors that are statistically significant in explaining the market share of a salad dressing or the number of units sold. We first focus on coefficients of variables interacted with NOLABEL. The coefficient on NOLABEL*EDUC is negative and significant in both specifications, thus providing support for H_1. This indicates that, ceteris paribus, a high-fat salad dressing without a label is purchased more the lower the education of the supermarket is. For the NSOLD equation, the coefficient on NOLABEL*EDUC equals –.22 with a T-value of –5.28. A 10% increase in the percentage of shoppers that have a college degree is associated with a decrease of almost 3 units sold for each high-fat unlabeled salad dressing. This is a large effect, because the average number of units sold is just below 15 and there are dozens of high-fat unlabeled dressings in each store. Because of the high correlation between income and education, alternative specifications were estimated to examine the robustness of the education result. Instead of estimating one equation for the three time periods, I estimate each time period separately. In every case, the results supported H_1. Although the coefficient on NOLABEL*EDUC varied across the three equations, it was negative and statis-

[15] Table 2 should be used in conjunction with the variable definitions provided previously.

Table 2. Descriptive Statistics

Variable	Mean	Standard Deviation	Minimum	Maximum
Variables Used for Equation 1—Number of Observations = 4335				
NSOLD	14.89	19.74	1	265
RSOLD	.011	.011	.00042	.124
EDUC	35.43	14.35	14	66
FEMALE	76.4	3.14	69	83
AGE	65.11	6.67	50	76
HSIZE	52.88	5.14	45	63
INC	64.80	13.61	33	83
NOLABEL	.44	.50	0	1
NOLABEL*EDUC	15.52	19.98	0	66
NOLABEL*INC	28.40	33.39	0	83
NOLABEL*HSIZE	23.18	26.47	0	63
NOLABEL*FEMALE	33.48	37.96	0	83
NOLABEL*AGE	28.53	32.61	0	76
PRICEPER	.16	.03	.106	.236
SALE1	.023	.15	0	1
SALE2	.022	.15	0	1
MEDIUM	.38	.49	0	1
SMALL	.59	.49	0	1
PFIEFFER	.07	.25	0	1
WISHBON	.29	.45	0	1
KRAFT	.38	.49	0	1
TOTSOLD	1391	543.02	329	2404
Variables Used for Equation 2—Number of Observations = 60				
RATIO	.485	.109	.313	.76
EDUC	35.30	14.48	14	66
RPRICE	.850	.071	.711	.926
FEMALE	76.35	3.14	69	83
AGE	65.05	6.74	50	76
HSIZE	52.85	5.19	45	63
INC	64.60	13.89	33	83
SALEL	.039	.057	0	.132
SALENL	.053	.075	0	.185

Notes: Data are based on Wegman's supermarket scanner data and nutrition label data as described in text.

tically significant in each of the three specifications.[16]

The coefficient on INC*NOLABEL is not significant in the NSOLD equation but is negative and significant in the market share specification, thereby providing mixed results with respect to H_2. The results indicate that the share of high-fat unlabeled salad dressings is lower the higher the average income of the shoppers in the supermarket is, though this result is sensitive to the specification of the equation. As with the education result, these results are similar when estimating the equation separately for each week of data: The coefficient on INC*NOLABEL was significant in each of the three NSOLD/TOTSOLD specifications and was not significant in any of the NSOLD specifications.

The coefficient on FEMALE*NOLABEL is negative and significant in both specifications, thus providing support for H_4. Supermarkets with a greater proportion of female shoppers have fewer sales of high-fat unlabeled salad dressings, ceteris paribus. The coefficient on AGE*NOLABEL is positive and significant in both specifications, thus providing support for H_3. The results indicate that supermarkets with elderly consumers sell more high-fat, unlabeled salad dressings.

The coefficients on the other variables included in Equation 1 generally have the expected sign. The coefficients on the variables SALE1, SALE2, and TOTSOLD are positive and significant in the NSOLD equation. SALE1 and SALE2 are also positive and significant in the market share equa-

[16]In time period 1 (sales during the week of October 1992), the coefficient on NOLABEL*EDUC was equal to –.13 with a t-value of –3.22 in the NSOLD specification and equal to –.00007 with a t-value of 1.75 in the NSOLD/TOTSOLD equation. In time period 2 (sales during the week of February 1993), the coefficient on NOLABEL*EDUC was equal to –.33 with a t-value of –3.45 in the NSOLD specification and equal to –.00011 with a t-value of –2.64 in the NSOLD/TOTSOLD specification. In time period 3 (sales during the week of June 1993), the coefficient on NOLABEL*EDUC was equal to –.19 with a t-value of –3.46 in the NSOLD specification and equal to –.00007 with a t-value of –1.93 in the NSOLD/TOTSOLD specification. The numbers of observations for the three periods were 1447, 1445, and 1440.

tion. The effects of the WSC price promotion (SALE2) is enormous.

Regression Results at the Supermarket Level

I present the regression results for Equation 2 in Table 5, which are consistent with those from Table 4. The coefficients on EDUC and INC are both negative and significant, indicating that in supermarkets with highly educated or high-income shoppers (support for H_1 and H_2), the percentage of salad dressing sales accounted for by high-fat unlabeled products is significantly lower. The coefficient on AGE is positive and significant (support for H3), consistent with the results from Equation 1. The coefficient on FEMALE is negative, though the t-value is only 1.58. Overall, the results in Table 5 confirm the links between purchasing patterns and demographic characteristics, though there are only 60 observations in this sample.

The coefficient on SALENL is positive and significant, indicating that the greater the number of unlabeled products on sale, the larger the percentage of sales accounted for by these products. Finally, the sign on the variable, RPRICE, is not in the expected direction. However, if the sale variables are not included in the regression, the sign on RPRICE is negative and significant.

Discussion of Results

The regression results from equations 1 and 2 indicate that socioeconomic characteristics are strongly related to the propensity to purchase high-fat unlabeled products within a product category. These results are consistent with H_1–H_4, though the relationship between income and purchase patterns is sensitive to model specification. Consequently, results based on actual purchase behavior are consistent with the consumer survey and experimental design research in this area. Specifically, the relationship between socioeconomic factors and product selection in the salad dressing market are generally consistent with the results of consumer survey research on the use and comprehension of labels and experimental studies on information acquisition and health behaviors.

Unfortunately, the data used in this study cannot isolate the effects of labeling versus other effects. For example, it would be useful to know whether the results in the salad dressing market occur because of demographic differences in the reaction to the presence or absence of a nutrition label. Research suggests (Grossman 1981; Huber and McCann 1982; Johnson and Levin 1985; Meyer 1981; Yamagishi and Hill 1981; Yates, Jagacinksi, and Farber 1978) that, faced with undisclosed product attributes, consumers discount the missing attribute. If these effects vary with information processing skills, differences in the propensity to purchase products with undisclosed attributes might arise. After implementation of the NLEA, all products were required to disclose nutrition information, which may help identify whether this explanation is plausible. Examining pre- and post-NLEA changes in purchase decisions across the different supermarkets should help identify whether nondisclosure had differential effects across demographic characteristics. Even pre- and post-NLEA comparisons, however, must be interpreted carefully. The NLEA not only required mandatory labeling, but also format changes—as well as other changes in presentation.

It would be useful as well to gauge whether the purchase patterns identified here arise from differences in consumers' general health knowledge or valuation of health. In the salad dressing market, differences in the observed purchase patterns simply may reflect variation in the demand for fat, rather than variation in the response to the presence or absence of a nutrition label. Different propensities to purchase labeled products also could arise because of product lifecycle effects. The labeled salad dressings were likely to be newer innovative low-fat products. If high income, highly educated consumers are more likely to adopt new products, then they would have higher propensities to purchase these products.

Conclusion and Further Research

The results indicate that the propensity to purchase high-fat products without nutrition labels within a product category varies greatly across key socioeconomic factors. Analysis of actual purchase behavior in the natural shopping environment yielded results that are consistent with consumer survey research and experimental design studies of the relationship between demographics and nutrition information processing skills. One limitation is the inability to clearly isolate the factors responsible for the differences in the propensity to purchase products with undisclosed undesir-

Table 3. Demographics by Percentage of Salad Dressings That Are Unlabeled (Stores Are Listed in Order of Education Level)

Store	% Income > 30,000	% College Graduate	Share of Sales by Unlabeled Dressings*
1	36	14	66
2	33	18	61
3	57	21	53
4	64	23	53
5	56	23	50
6	52	25	60
7	75	26	50
8	59	27	57
9	79	29	52
10	54	29	50
11	58	34	52
12	72	35	52
13	75	36	45
14	61	40	47
15	69	44	47
16	76	47	47
17	83	53	42
18	78	57	50
19	75	59	40
20	80	66	42

Notes: Data are from Wegman's scanner data and shopper club card application forms.
*The percentages reported in the table are based on the average of the three waves of data from each store.

Table 4. Determinant of Salad Dressing Sales

	Number of Units Sold		Market Share	
Coefficient	Estimate	T-Value	Estimate	T-Value
INTERCEPT	10.67	1.67	.014400	3.60**
EDUC	.089	3.21**	.0000401	2.43**
FEMALE	–.117	–1.39	–.0000450	–.90
AGE	–.050	–1.10	–.0000139	–.51
HSIZE	–.040	–.68	–.0000585	–1.67*
INC	.118	3.11**	.000114	5.00**
NOLABEL	–2.85	–.30	.008350	1.45
EDUC*NOLABEL	–.219	–5.27**	–.0000866	–3.49**
INC*NOLABEL	.00	.03	–.000126	–3.81**
HSIZE*NOLABEL	–.01	–.157	.000115	2.19**
FEMALE*NOLABEL	–.01	–.79	–.00016	–2.10**
AGE*NOLABEL	.24	3.54**	.00011	2.64**
PRICEPER	–11.03	–1.27	–.0110	–2.11**
SALE1	104.18	83.65**	.0609	81.68**
SALE2	11.08	10.15**	.0065	9.95**
MEDIUM	–2.84	–2.94**	–.0020	–3.44
SMALL	–.206	–.20	.00019	.314
PFIEFFER	3.17	3.25**	.0022	3.74**
WISHBON	3.20	7.32**	.0022	9.54
KRAFT	–.105	–.24	.00006	.22
TOTSOLD	.006	16.52**	–.000003	–14.14**

Notes:
*indicates significance at the 90 % level.
**indicates significance at the 95 % level of confidence.
N = 4,335.
Adjusted R^2 = .7405 for NSOLD equation and .7147 for market share equation.
Data are based on supermarket scanner data and nutrition label data set.

able attributes. These differences may occur for a variety of reasons, including health valuation, tastes, nutrition label usage and comprehension, adoption of new products, and general health knowledge. As was discussed previously, post-NLEA data can help isolate some of these effects, and this research is currently being undertaken.

I focus on a single market for a limited number of salad dressings. The research can be expanded in several ways. The consistency of these results across other product categories is of key concern. I focus on salad dressings because of several advantages that this market offered for analysis. Most other categories would be more difficult to analyze. Nevertheless, it is important to understand whether these results generalize to other product categories.

The consistency of these results with respect to markets other than food is also of interest. For example, the effects of nondisclosure of undesirable product attributes on purchase behavior is an important issue in the evaluation of the deceptiveness of advertising claims. The National Association of Attorney Generals has consistently pursued legal action against firms for advertising low prices yet omitting other attributes about the price, namely the restrictions associated with it.

Despite my narrow focus on the salad dressing market, the results raise several points that address broader issues in consumer health. In this study, as in many others, variables that are used to proxy information processing skills are associated with more healthy choices. This indicates that government policies that reduce information processing costs may lead to improved health choices. Policies such as mandatory labeling, standardized nutrition terms, and so on have the potential to lower the skills necessary to process nutrition information. Other policies, such as limitations on the type and number of health claims, are more controversial. Although limitation of health claims is supported by many, others

Table 5. Determinants of Percentage of Sales Accounted for by Unlabeled Products

Variable	Coefficient	T-Value
INTER	.542	2.36**
EDUC	–.002	–2.82**
INC	–.003	–2.58**
HSIZE	.003	1.70*
FEMALE	–.004	–1.58
AGE	.003	1.97*
RPRICE	.13	.73
SALEL	.01	.04
SALENL	1.25	4.98**

Number of observations = 60.
Adjusted R^2 = .8259.
*indicates significance at the 90% level.
**indicates significance at the 95% level.
Data are based on supermarket scanner data and nutrition label data as described in text.

maintain that these limitations will stifle the flow of truthful information to consumers, thereby raising nutrition information processing costs. (Ippolito and Mathios 1993). All these issues potentially can be addressed with supermarket scanner data and are important areas for further research.

References

Adrian, J. and R. Daniel (1976), "Impact of Socioeconomic Factors on Consumption of Selected Food Nutrients," *American Journal of Agricultural Economics*, 58 (1), 31–38.

Bassler, E. and G. K. Newell (1982), "Food Shopping Behavior and Food Use by Well-Educated Young Parents," *Journal of Nutrition Education*, 14 (December), 146–49.

Bender, M. and B. Derby (1992), "Prevalence of Reading Nutrition and Ingredient Information on Food Labels: 1982–1988," *Journal of Nutrition Education*, 24 (6), 292–97.

Caswell, J. A. (1992), "Current Information Levels on Food Labels," *American Journal of Agricultural Economics*, 74 (5), 1196–201.

Cole, C. A. and S. Balasubramanian (1993), "Age Differences in Consumers' Search for Information: Public Policy Implications," *Journal of Consumer Research*, 20 (June), 157–69.

——— and G. Gaeth (1990), "Cognitive and Age-Related Differences in the Ability to Use Nutrition Information in a Complex Environment," *Journal of Marketing Research*, 27 (May), 175–84.

Cotunga, N., A. F. Subar, J. Heimendinger, and L. Kahle (1992), "Nutrition and Cancer Prevention Knowledge, Beliefs, Attitudes, and Practices: The 1987 National Health Interview Survey," *Journal of the American Dietetic Association*, 92 (8), 963–68.

Fischer, C. A., S. J. Crockett, K. E. Heller, and L. H. Skauge (1991), "Nutrition Knowledge, Attitudes, and Practices of Older and Younger Elderly in Rural Areas," *Journal of the American Dietetic Association*, 91 (11), 1398–401.

Food Marketing Institute (1989), *Trends: Consumer Attitudes and the Supermarket, 1989*. Washington, DC: Food Marketing Institute.

Fullmer, S., C. J. Geiger, and C. R. Michael Parent (1991), "Consumers Knowledge, Understanding, and Attitudes Toward Health Claims on Food Labels," *Journal of the American Dietetic Association*, 91 (2), 166–71.

Fusillo, A. E. and A. M. Beloian (1977), "Consumer Nutrition Knowledge and Self-Reported Shopping Behavior," *American Journal of Public Health* 67 (9), 846–50.

Glanz, K., A. M Hewitt, and J. Rudd (1992), "Consumer Behavior and Nutrition Education: An Integrative Review," *Journal of Nutrition Education*, 24 (5), 267–77.

Gould, B., and H. C. Lin (1994), "Nutrition Information and Household Dietary Fat Intake," *Journal of Agricultural and Resource Economics*, 19 (2), 349–65.

Grossman, S. J. (1981), "The Informational Role of Warranties and Private Disclosure About Product Quality," *Journal of Law and Economics*, 24 (3), 461–89.

Heroux, L., M. Laroche, and K. L. McGown (1988), "Consumer Product Label Information Processing: An Experiment Involving Time Pressure and Distraction," *Journal of Economic Psychology*, 9 (2), 193–214.

Huber, J. and J. McCann (1982), "The Impact of Inferential Beliefs on Product Evaluations," *Journal of Marketing Research*, 19 (August), 324–33.

Ippolito, P. M. and A. D. Mathios (1991), "Health Claims in Food Marketing: Evidence on Knowledge and Behavior in the Cereal Market," *Journal of Public Policy & Marketing*, 10 (Spring), 15–32.

——— and ——— (1993), "New Food Regulations and the Flow of Nutrition Information to Consumers," *Journal of Public Policy & Marketing*, 12 (Fall), 188–205.

——— and ——— (1994), "Information, Policy, and the Sources of Fat and Cholesterol in the U.S. Diet," *Journal of Public Policy & Marketing*, 13 (2), 200–217.

——— and ——— (1995), "Information and Advertising: The Case of Fat Consumption in the United States," *American Economic Review*, 85 (2), 91–95.

Johnson, R. D. and I. P. Levin (1985), "More Than Meets the Eye: The Effect of Missing Information on Purchase Evaluations," *Journal of Consumer Research*, 12 (2), 169–77.

Kim, H. (1995), "The Effect of Consumers' Diet and Health Knowledge on Their Food Consumption," doctoral dissertation, Department of Consumer Economics and Housing, Cornell University.

Klopp, P. and M. MacDonald. (1981), "Nutrition Labels: An Exploratory Study of Consumer Reasons For Nonuse," *Journal of Consumer Affairs*, 15 (2), 301–16.

Lenahan, R. J., J. A. Thomas, D. A. Taylor, D. L. Call, and D. I. Padberg (1973), "Consumer Reaction to Nutrition Labels on Food Purchases," *Journal of Consumer Affairs*, 7 (1), 1–12.

Meyer, R. J. (1981), "A Model of Multiattribute Judgments Under Attribute Uncertainty and Informational Constraints," *Journal of Marketing Research*, 18 (November), 428–41.

Michel, P. M., M. K. Korslund, A. Finan, and J. Johnson (1994), "Food Label Reading Habits of WIC Clients," *Journal of Nutrition Education*, 26 (3), 146–48.

Mitra A., M. Hastak, G. T. Ford, and D. J. Ringold (1995) "Can the Educationally Disadvantaged Interpret Nutrition Information in the Presence of a Health Claim?" unpublished manuscript, Atkinson Graduate School of Management, Willamette University.

Moorman, C. (1990), "The Effects of Stimulus and Consumer Characteristics on the Utilization of Nutrition Information," *Journal of Consumer Research*, 17 (December), 362–74.

——— and E. Matulich (1993), "A Model of Consumers' Preventive Health Behaviors: The Role of Health Motivation and Health Ability," *Journal of Consumer Research*, 30 (September), 208–28.

Putler, D. and E. Frazio (1991), "Diet/Health Concerns About Fat Intake," *Food Review* 14, 16–20.

Russo, J. E., R. Staelin, C. Nolan, G. Russell, and B. Metcalf (1986), "Nutrition Information in the Supermarket," *Journal of Consumer Research*, 13 (June), 48–70.

Schucker, R., A. S. Levy, J. E. Tenney, and O. Matthews (1992), "Nutrition Shelf Labeling and Consumer Purchase Behavior," *Journal of Nutrition Education*, 24 (2), 75–81.

Yamagishi, T. and C. T. Hill (1981), "Adding Versus Averaging Models Revisited: A Test of a Path-Analytic Integration Model," *Journal of Personality and Social Psychology*, 41 (January), 13–25.

Yates, J., C. Frank, J. Jagacinski, and M. D. Farber (1978), "Evaluation of Partially Described Multi-Attribute Options," *Organizational Behavior and Human Performance*, 21 (April), 240–51.