The Effect of Retailer Concentration and Store Format on the Healthfulness of Consumers' Food Purchasing Decisions

Richard Volpe[♣]
Abigail Okrent[♣]
Ephraim Leibtag[∃]

USDA-Economic Research Service

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association 2011
AAEA, & NAREA Joint Annual Meeting, Pittsburgh, Pennsylvania, July 24-26, 2011

[♣] Authors are Research Economists with the Food Economics Division of the USDA-Economic Research Service. The contact author is Richard Volpe, who can be reached at rvolpe@ers.usda.gov.
[∃] Author is the Deputy Director for Research, Food Economics Division, USDA-Economic Research Service.
The views expressed are those of the authors and may not be attributed to the USDA or the Economic Research Service.

Abstract

We examine the effect of the market share of supercenter stores on consumers' food-at-home purchasing habits in the United States. We measure healthfulness several different ways to ensure robustness, but all measurements place a greater value on fresh fruits and vegetables and whole grains than on processed foods high in sugar and sodium. The results show that consumers, on average, purchase less healthful foods at supercenters than they do at supermarkets. Moreover a one-percent increase in the local market share of supercenters results in a decrease in purchase healthfulness for groceries of 0.5 to two percent.

Keywords: food purchases; supercenters; obesity; consumer behavior; food retail

JEL codes: L66; P46; D12; L11

Introduction

One of the most visible and well-documented changes in food retail in the United States in the past two decades has been the rise of supercenters, specifically Wal-Mart Supercenters. Supercenters are defined as large stores that combine the typical offerings of department stores with full grocery stores.[1] Though the supercenter format has existed in the U.S. for many decades through examples such as Meijer and Fred Meyer, the mass merchandiser Wal-Mart opened its first Supercenter in 1988. At writing, there are over 2,900 Supercenters operating in the United States (Wal-Mart Inc, 2011). Wal-Mart is by far the largest food retailer in the U.S. (Supermarket News, 2011) and several other chain stores such as Target and Kmart are opening supercenter stores across the country in an effort to emulate the firm's success.

Concurrently, obesity and a range of other health-related issues are widespread and on the rise in the United States (Ogden et al., 2006). Obesity in particular is frequently cited as a major contributor to healthcare costs (Bhattacharya and Bundorf, 2009). The thrust of this paper is to examine if there is a connection between the market share of supercenters, which is classified as a market-structure factor, with the healthfulness of consumers' food purchases. The extent to which consumers' shopping baskets conform to USDA recommendations for healthful eating has direct implications on the prevalence of obesity and overall health in the United States. Hence any relationship between food retail structure and consumer purchases can offer insights into the causes of obesity and related issues.

The econometric results, controlling for market concentration and regional characteristics as well as household effects, indicate increases in local supercenter share result in decreased purchase healthfulness. Across space, increased supercenter share is associated with decreased

[1] Supercenters are also referred to as hypermarkets and superstores in the literature.

healthfulness. Over time, the effect persists but is less robust. We argue that, on average, supercenter shoppers purchase less healthful foods than their counterparts at conventional supermarkets, but that the relative quality of the food being purchased at supercenters may be increasing due to concerted efforts on the part of these stores to offer more healthful options.

Background

There has been much economic research on the economic effects of supercenters. In this section we provide a brief overview of the relevant literature for this market characteristic and justify the modeling of food basket healthfulness as a function of supercenter share. We also discuss our expectations for the marginal effect of supercenter share on healthfulness, showing that the theory and literature lead to ambiguous conclusions. The vast majority of supercenter studies focus on Wal-Mart Supercenters, but to prevent confusion we use the lowercased term "supercenter" when discussing the relevant literature to make clear that our economic reasoning pertains to all stores of this format.

A key component to understanding the proposed supercenter share-healthfulness relationship is the understanding of what differences, if any, exist between the food product menus at supercenters and supermarkets. Selected survey evidence (Market Force, 2011; Seiders et al., 2000) has suggested at consumers find overall quality inferior at supercenters, particularly for fresh meat and produce. However other survey results have found that consumers can discern no difference in food quality between supercenters and supermarkets (DSN Retailing Today, 2003). Additionally both academic researchers and authorities on the food supply chain have determined that the products sold at supercenters and supermarkets are essentially identical (Singh et al., 2006; Vallianatos et al., 2004; Western Farm Press, 2005). Therefore we do not

assume that the quality and healthfulness of supercenter product menus differ importantly from those of supermarkets.

Having set aside intrinsic differences in quality across store formats as a determinant, we turn our focus to price and competitive considerations. Supercenter prices are almost uniformly lower than supermarket prices, for comparable products and particularly for fruits and vegetables (Leibtag et al., 2010). Moreover, low-income consumers are more likely to shop at supercenters than any other store format (Basker, 2007a). Therefore we suggest the possibility that supercenters make healthful foods more affordable for consumers who otherwise may consider them to be prohibitively expensive (Blisard et al., 2004). Supercenter entry and share expansion has been connected to significantly lower prices at competing supermarkets (Hausman and Leibtag, 2007; Volpe and Lavoie, 2008; Woo et al., 2001), and thus as supercenter share increases we expect local food prices to decrease on average. The latter two studies highlighted this effect for categories such as fresh produce and dairy, which contain many foods recommended as a part of a healthy diet.

Despite the fact that competitors have been shown to lower their prices in competition with supercenters, the size and scope of the supercenter chains, particular Wal-Mart, ensures that conventional supermarkets are unable to consistently beat supercenters in price. The literature prescribes a number of alternative avenues for competition with supercenters and common among them is a focus on product quality and variety (Basker and Noel, 2009; Seiders et al., 2000; Singh et al., 2006). Therefore as the local supercenter share increases, there is reason to expect that incumbent supermarkets will increasingly focus on improving the quality and variety of their offerings and thus expanding on healthful options for consumers throughout the market. These arguments support a positive supercenter share-healthfulness relationship.

However the price effect of supercenters, as measured by both the prices at supercenters and their effect on competitors' prices, suggests a competing hypothesis. A body of research (Chou et al., 2004; Cutler et al., 2003; Philipson and Posner, 2003) has demonstrated that decreases in the real cost of food, in terms of both time and money, are associated with increases in obesity, typically as measured by the body mass index. Citing this relationship as motivation, Courtemanche and Carden (2011) studied the relationship between Wal-Mart Supercenters and obesity and ultimately attributed 10.5 percent of the total rise in U.S. obesity since the 1980s to the expansion of Supercenters. More specifically they identified an increase in the obesity rate of 2.3 percent for one additional Supercenter per 100,000 people. It is worth noting that in their study the authors also concluded that the theory on the potential link between Supercenters and obesity was ambiguous and depended on the relative price effects across healthful and unhealthful foods, among several other factors.

Data on Consumer Food Purchases

We use the 1998-2006 Nielsen Homescan data to measure consumer expenditures by food category for our analysis. The Homescan dataset consists of the self-scanned purchases of a sample of households, who are asked to record their complete purchases of each shopping trip by scanning UPCs. The sample includes households across most of the geographic United States and the Homescan data include detailed information on product names and characteristics as well as household demographics.

The first step in testing for the market structure determinants of food purchases is to organize the Homescan data, which include thousands of individual products, into larger food groups that are tractable for our analysis. The Quarterly Food-at-Home Price Database (QFAHPD) provides an excellent starting point by constructing 52 food groups based on the

2005 Dietary Guidelines (see Todd et al. (2010) for more complete discussion). These food groups form the basis of our analysis and enable the comparison of purchasing habits between store formats. The food groups include fruits and vegetables, whole and refined grains, prepared foods, meats, fats, sweeteners, and more, as shown in table 1.

Table 1 here.

We constructed shopping baskets of total expenditures by Homescan household and quarter, yielding a total of 656,212 shopping baskets for nearly 40,000 households. We then calculated expenditure shares for the 52 QFAHPD food groups. Expenditure shares for household *i* and food group *g* (g = 1-52) in quarter *q* (q = 1998Q1–2006Q4) is given by

$$share_{igq} = \frac{exp_{igq}}{\sum_{g=1}^{52} exp_{igq}}, \quad (1)$$

where *exp* is expenditure as reported in the QFAHPD. As an illustrative example, the row for food group 1 in table 1 provides the average expenditure shares of whole fresh or frozen fruit for quarterly shopping baskets. Hence households in the Homescan database spend an average of 2.25 percent of their food-at-home dollars on whole fresh or frozen fruit.

Methods for Measuring Basket Healthfulness

The major goal of this paper is to understand how market structure impacts consumers' food purchasing decisions, with a focus on the healthfulness of food purchased for at-home consumption. Without complete data on household demographics, food-away-from-home consumption, and at-home preparation and consumption of food, no measurement of basket healthfulness will be without limitations or assumptions. Accordingly, we measure basket healthfulness three different ways and apply our analysis to each.

Food Groups Recommended as Part of a Healthy Diet

The simplest and most direct method for assessing the nutritional quality of quarterly shopping baskets begins by sorting the QFAHPD categories into two groups: those comprised of foods recommended as a part of a healthy diet according to USDA guidelines and those not. The executive summary of the USDA *Dietary Guidelines for Americans, 2010* describes a healthy diet as being one that:

- Emphasizes fruits, vegetables, whole grains, and fat-free or low-fat milk and dairy products
- Includes lean meats, poultry, fish, beans, eggs, and nuts
- Is low in saturated fats, trans fats, cholesterol, salt (sodium), and added sugars

Most of the QFAHPD food groups contain foods for which it is readily apparent whether or not they meet these guidelines. Those food groups meeting the USDA guidelines are henceforth referred to as "healthful." Working in numerical order, groups 1 through 15 are all healthful. The only potentially questionable group in this list is fruit juice, number 3. Certain fruit juices contain significant amounts of added sugar per serving, which is not recommended for consumption by USDA. Our default classification of fruit juice as healthful owes to the fact that most fruit juices count as at least one serving of fruit. Food groups 16 through 18 are all healthful, as they contain various whole grain foods and cooking products. Food groups 19 through 21 are all not healthful, as USDA guidelines recommend against the consumption of refined and processed grains. Food groups 22 through 24, low-fat milk and dairy, are healthful. Alternatively, regular-fat milk and dairy, groups 25 through 27, are not healthful. Low-fat meat, poultry, fish, nuts, and eggs are all recommended and hence groups 28 and 31 through 37 are all healthful. Regular-fat meats, food groups 29 and 30, are not healthful.

The proper classification of fats is not as clear as most other food groups. Fats are important sources of calories and fatty acids, however their consumption is recommended only in strict moderation by USDA. While there are exceptions within food groups, oils are more likely to contain cholesterol-improving monounsaturated fatty acids while solid fats are more likely to contain saturated fats and trans fats, neither of which are recommended for consumption by USDA. Therefore we classify food group 38, oils, as healthful and food group 39, solid fats, as not healthful.[2]

The remaining categories lend to more straightforward classification. Raw sugars and sweeteners, food group 40, is not healthful. USDA guidelines dictate that Americans ought to obtain sufficient dietary sugars from fruits, vegetables, and dairy. All of the beverages found in the supermarket beverage aisle except for water are not healthful. Hence groups 41 and 42 are not healthful, group 43 is healthful. Finally, food groups 44 through 52 are not healthful. Processed and packaged foods, whether canned, boxed, or frozen, typically contain excessive added sugar and sodium in addition to a number of additives not recommended for consumption by USDA. Table 1 summarizes the classification of the 52 groups by healthfulness.

To assess the effects of market structure on the nutritional content of shopping baskets using this classification, we calculate the share of each household's total expenditure that is attributed only to those healthful food groups. Using the same subscripts as in (1), healthful expenditure shares are given by

$$healthexpshare_{iq} = \frac{\sum_g exp_{igq} \left| g \in healthful \right.}{\sum_{g-1}^{52} exp_{igq}}. \tag{2}$$

[2] Performing the empirical analysis by changing the classification of those food groups for which classification is most challenging, specifically fruit juice and oils, does not change the findings qualitatively. For the case of oils, the expenditure shares are near enough in magnitude between store formats that the results do not change perceptibly.

A drawback specific to this approach is that it does not account for dietary recommendations with respect to proportions. For example, a hypothetical household basket consisting entirely of eggs would be measured as being 100% healthful despite the fact that USDA guidelines recommend a balance of a variety of foods.[3]

Scoring Based on CNPP Expenditure Recommendations

The USDA Center for Nutrition Policy and Promotion (CNPP) calculates food plans to assist Americans in allocating their food budgets to meet USDA dietary guidelines. Assuming a close relationship between food purchase and consumption, these food plans provide a framework for measuring basket healthfulness as we can compare expenditure shares directly. The CNPP report provides separate food plans for consumers with food expenditures in the second quartile of the U.S. (Low-Cost), the third quartile (Moderate Cost), and upper quartile (Liberal). We are unable to accurately measure the total food expenditures for the households in our data, and we rely on the Liberal Food Plan as our basis for comparison.[4]

The CNPP recommended expenditure shares pertain to food categories that do not correspond perfectly with the QFAHPD food groups. We applied some degree of aggregation to both organizational schemes in order to facilitate direct compatibility (see table 2). In addition to combining food classifications within both systems, we used the USDA food plans to calculate recommended expenditure shares for households, rather than for individuals, in order to best facilitate measurements on Homescan household baskets. To do so, we began with the CNPP estimates of the weekly recommended dollar costs of feeding men, women, and children of all

[3] Additionally, we divide all household expenditures through by price in order to obtain quantities by food group. Thus we also calculate the quantity shares of the healthful food groups in order to account of the possibility that the relative prices of healthful products at one of the store formats may be driving observed differences in expenditure shares. The complete set of results based on this method, which are available from the author upon request, do not change our findings substantively.

[4] The relative expenditure shares for the various food categories compiled by CNPP are very similar across food plans. Using the Low-Cost or Liberal food plans does not change the findings substantively.

ages, by food category. From these values, it is possible to calculate total recommended food expenditure as well as optimal expenditure shares tailored for individual families, based on their demographic composition. To facilitate simple comparisons between store formats, we calculated total weekly food-at-home costs for a representative American family, as defined by the USDA Thrifty Food Plan (TFP). Such a family consists of one male aged 19-50, one female aged 19-50, and two children, aged 9-11 and 6-8, respectively. From total dollar costs, by food category, we calculated recommended expenditure shares for this representative household, as reported in the third column of table 2.

Table 2 here.

On average, the household shopping baskets feature significantly less expenditures on fruits, vegetables, fish, and whole grains than recommended by USDA, while expenditure shares are significantly higher than recommendations for refined grains, whole milk products, soft drinks, and other packaged and processed foods. For the case of sugars, sweets, and candies the deviations from USDA dietary recommendations are particularly dramatic. The USDA-recommended expenditure share for this food category is less than 0.5 percent for a typical family, yet the average expenditure share is 13.8 throughout the data.

The detailed demographic information available for each family is an asset of the Homescan database. Households report, with very few missing variables, the age and presence of male and female heads of household as well as the age and presence of children. With such information we are able to calculate USDA-recommended total food expenditure as well as expenditure shares individually for every family in the data.[5] Using the household-level recommended expenditure shares and the observed expenditure shares as calculated based on the

[5] Appendix A details the methodology for calculating recommended expenditures for Homescan families using the CNPP report.

QFAHPD food groups, we assign each household a score, by quarter, CNPP food category, and store format. This score, henceforth the USDA Score, is based on households' adherence to USDA recommendations. The USDA Score is calculated in three different ways, given by

$$(3)\quad USDAScore1_{icq} = \sum_c \left(\left(expshare_{icq} - USDAexpshare_{ic} \right)^2 \right)^{-1}$$

$$(4)\quad USDAScore2_{icq} = \sum_c \left(\left(expshare_{icq} - USDAexpshare_{ic} \right)^2 \middle| expshare_{icq} > 0 \right)^{-1}$$

$$(5)\quad \begin{aligned} USDAScore3_{icq} = & \sum_{c \in healthful} \left(\left(expshare_{icq} - USDAexpshare_{ic} \right)^2 \middle| expshare_{icq} < USDAshare_{ic} \right)^{-1} \\ & + \sum_{c \notin healthful} \left(\left(expshare_{icq} - USDAexpshare_{ic} \right)^2 \middle| expshare_{icq} > USDAshare_{ic} \right)^{-1} \end{aligned}$$

where all subscripts are familiar from (1) except for *c*, which denotes the CNPP food categories, as revised to enable compatibility with QFAHPD. For the purpose of calculating USDAScore3 we classified each CNPP category as being healthful or unhealthful based on the QFAHPD groups comprising each. CNPP does not incorporate bottled water into its food categories, and QFAHPD does not maintain prices for coffee and tea. The USDA scores do not incorporate any of these products.

The USDA scores are squared-error loss functions, designed to assign penalties for household expenditure shares that deviate from USDA recommendations.[6] *USDAScore1* is the simplest and operates on the assumption that the Homescan households report 100% of their food-at-home purchases to Nielsen. Therefore we input an expenditure share of zero for those food groups for which households report no purchases.

USDAScore2 makes a different assumption for completely empty food categories, allowing for the possibility that households simply have not recorded these purchases. Non-

[6] An alternative score design would rely on the absolute value of deviations from USDA recommendations. Squared-error loss functions assign greater weights to deviations of high magnitudes, while absolute value loss functions assume that each unit of deviation has the same effect on healthfulness.

recording of purchases is likely to be true for at least some cases in the data, because the sum total of recorded purchases for many household/quarter combinations is very low, relative to the caloric demands of adults. Therefore *USDAScore2* is calculated based only on those food categories with recorded purchases and expenditure shares greater than zero.

Finally, *USDAScore3* penalizes deviations from recommendations only if they detract from overall basket healthfulness. For this score we return to the assumption of calculating (3), that households record all of their food-at-home purchases. Under *USDAScore3*, households are penalized only if observed expenditure shares are below (above) those recommended for healthful (unhealthful) categories. We take the inverse of all three scores in order to lend ease of interpretation to the overall findings of the study, in that higher scores are always more indicative of healthfulness throughout.

Scoring Based on the Healthy Eating Index

The last method we use to score the healthiness of each household's food basket is the 2005 Health Eating Index (HEI). The 2005-HEI was constructed as a way to quantify and track the changes in the diet quality of Americans. The 2005-HEI is composed of 12 components where the maximum score for each component adds up to 100 (Table 3). For example, if an individual consumed less than 20% of his or her energy from solid fats, alcohol or added sugars then they would get 20 points. The standards for scoring each component are based on adherence to the 2005 Dietary Guidelines. An individual that completely follows the recommendations set forth in 2005 Dietary Guidelines would achieve a maximum score of 100 (Guenther et al., 2007). Because the HEI-2005 is not tied to individual requirements and is scored on a per 1000 kcal basis, it can be used to assess the overall quality of any mix of foods. Reedy, Kreb-Smith and Bosire (2010) showed how the 2005-HEI can be used to measure the diet quality of fast-food

restaurant menus and the U.S. food supply. We used the 2005-HEI to measure the diet quality of purchases made by households at different store outlets.

Table 3 here.

Unfortunately, the QFAHPD database does not contain the nutrient content of the foods purchased by households, and hence, HEI-2005 scores cannot be directly assigned to each household's basket of goods. To overcome this obstacle, we used the nutrient characteristics of foods consumed by sample respondents in the 2003-04 National Health and Nutrition Examination Survey (NHANES) to proxy for this missing nutrient information. The NHANES is a nationally-representative sample of civilian non-institutionalized individuals that collects two days of 24-hour dietary recall data from the sample participants and converts the dietary data into its nutrient composition (Center for Disease Control, 2006). We used the nutrient content of foods in the 2003-04 NHANES that closely match our QFAHPD food groups to indirectly assign an HEI-2005 score to household purchases in the QFAHPD.

First, we calculated the HEI for sample respondents in the 2003-04 NHANES based on one day of dietary intake. Only individuals with reliable dietary recall were scored and included in our analysis (i.e., DR1DRSTZ=1). Foods purchased at retail outlets other than grocery stores (e.g., restaurants, school and residential cafeterias, vending machines and vendors, community food programs and so on) were excluded from our HEI calculations (i.e., DR1FS = 1). We calculated the HEI scores using SAS program files and data sets published by the CNPP for the 2001-02 NHANES with adjustments for excluding food purchased at grocery stores (U.S. Department of Agriculture, 2010).

Second, the foods consumed by sample respondents in the NHANES were assigned to one of 30 food groups and we estimated the average effect of consumption of each food group on HEI-2005. We estimated the average effect of consumption of the 30 foods on the HEI-2005 as

$$HEI_j = \beta_0 + \sum_{n=1}^{30} \beta_n Q_{jn}^{NHANES} + \varepsilon_j, \quad (6)$$

where β_0 is a constant, β_n is the mean effect of consumption of food *n* on the HEI, Q_{in}^{NHANES} is the grams of food of food *n* consumed by individual *j* as reported in the NHANES, and ε_j is the error term The sample estimates were inflated to represent the entire U.S. civilian non-institutionalized population by using the inverse probability weights for the one day of dietary recall and the standard errors were adjusted to account for the stratified multistage probability of the sample. Table 4 contains the estimated mean parameters and standard errors using (6). Most of the regression coefficients are significant and have the expected signs.

Table 4 here.

Third, we used the estimated parameters in Table 4 and quantities of foods purchased in the QFAHPD to predict HEI for each QFAHPD household. The quantities consumed each quarter by a household for a food group were calculated by dividing total expenditure on the food group divided by household average price as reported in the QFAHPD.[7] Because the data in the NHANES represents daily consumption by individuals, we divided household quantities by household size to derive per capita quantities and then converted the per capita quarterly quantities into per capita daily quantities. We predicted HEI for each household as

$$\hat{HEI}_i = \hat{\beta}_0 + \sum_{n=1}^{30} \hat{\beta}_n Q_{in}^{QFAHPD} \quad (7)$$

[7] The price for a food group that consisted of more than one QFAHPD price was calculated as a weighted average of each QFAHPD price weighted by its expenditure share of the food group.

where $\hat{\beta}_0$ and $\hat{\beta}_n$ are the estimated parameters from Table 3 and Q_{hn}^{QFAHPD} is per capita daily consumption of food group *n* in household *i*.

Purchase Differences across Store Formats

The net effect of supercenter market share on consumers' food purchase healthfulness depends on a number of considerations, including the impact supercenters have on the pricing of their rivals and the extent to which consumers shop among competing stores to complete their shopping baskets. However a good place to start this empirical investigation and to justify the analysis is to compare the healthfulness of shopping baskets purchased at supermarkets with those purchased at supercenters. That is, to compare directly purchases between the two formats. If the foods purchased at supercenters are, on average, significantly different from those purchased at supermarkets in healthfulness, that would help to form our expectations on the effect of supercenter share and to understand why such an effect may be taking place. Responders to the Homescan survey report the format type of the stores at which they make their purchases, and thus we are able to measure these differences directly.

To accomplish this, we calculate the expenditure shares using the same method depicted in (1), only separately for purchases made at different store formats. That is, the expenditure share for household *i*, food group *g*, quarter *q*, and format *f* (f =supercenter, supermarket) is given by:

$$share_{igqf} = \frac{exp_{igqf}}{\sum_{g=1}^{52} exp_{igqf}} \tag{8}$$

Given (8), it is immediately possible to inspect how purchases differ by store format. Table 5 presents the average expenditure shares for Homescan households, by categories of QFAHPD food groups and store format. The QFAHPD groups were previously presented and categorized

according to healthfulness in table 1. The average quarterly shopping basket at supermarkets consists of more fruits, vegetables, whole grains, and healthful meats than does the average supercenter basket. However supercenter baskets also exhibit higher proportions of refined grains and unhealthful meats. The greatest difference between the two store formats pertains to packaged and processed foods, as the average expenditure share for these foods is 45 percent in supercenter baskets and 37 percent for supermarket baskets. Thus we have evidence that supercenter shoppers purchase considerably more foods that are not recommended for consumption by USDA.

Table 5 here.

Next we take the step of applying our measurements of healthfulness to baskets comprised entirely of foods purchased at one store format or the other. We use (2), (3), (4), (5), and (7) and also calculate healthful quantity shares separately for the store formats to enable a robust comparison of supercenter and supermarket shopping baskets. Table 6 summarizes the results. Given that the various measurements require somewhat different interpretation, we calculate the percentage difference between the average scores of the two formats for all scores. Supermarket baskets significantly outperform supercenter baskets by all metrics. The difference is the most dramatic for the three USDA Scores, which show supermarket baskets to be 41 to 47 percent more healthful. The difference is the smallest for *HEIScore*, showing supermarket baskets to be 13.5 percent more healthful. These findings suggest that the differences between formats in terms of fruits, vegetables, and packaged foods are enough to induce a considerable margin in overall healthfulness.

Table 6 here.

Recognizing that much of these observed differences may be driven by demographics or regional characteristics correlated with supercenter shopping activity, we subject these format-specific healthfulness scores to a simple regression analysis. The regression equation is intended to identify the difference in purchase basket healthfulness between store formats while controlling for several potential confounding factors, as suggested by theory and the literature on consumer food choice. The model is given by:

$$\text{Healthfulness}_{iqfm} = f(\text{Supercenter}_{iqf}, \text{Concentration}_{iq}, \text{Income}_{iq}, \text{Education}_{iq}, \text{Free Time}_{iq}, \text{Profession}_{iq}, \text{Race}_{i}, \text{Regional Effects}_{i}, \text{Year Effects}_{q}, \text{Quarterly Effects}_{q}, \text{Supercenter-Year Interactions}_{qf}) + \text{Error}_{iqfm} \quad (9)$$

where all variables except for *Concentration* are drawn directly from the Homescan data and in turn the QFAHPD. *Supercenter*, the key variable of interest, is a dummy variable equal to one if the basket was purchased a supercenter-format store. All subscripts are familiar from (8) except for *m*, which denotes the healthfulness metric being used. Table 7 provides summary statistics for all of the variables used in the analysis, including the measurements of basket healthfulness. Additionally, all variable definitions as well as motivation for use can be found in appendix B.

Table 7 here.

Retailer concentration in particular is an important market structure consideration in many studies examining food retail. It is commonly used as a measurement of market power or competition among firms, as such factors are very difficult to measure directly. We have reason to expect concentration to impact consumers' purchasing decision through two avenues. For one, there is a well-established link between concentration and supermarket food prices (Cotterill, 1986; Lamm, 1981; Yu and Connor, 2002), which may drive consumers to adjust their shopping baskets or retail outlets. For another, supermarkets in concentrated markets tend to differentiate themselves from one another through food quality and variety (Bonanno and Lopez, 2009;

Ellickson, 2006), meaning we expect the product mix and in turn consumers' purchases to be more healthful in more concentrated markets. Data on market concentration was obtained from the 1998, 1999, 2000, 2001, and 2006 editions of the *Market Scope*, a Trade Dimensions publication.[8]

We estimated (9) using ordinary least squares (OLS) and the results are reported in tables 8 and 9. The results indicate that basket healthfulness is significantly lower at supercenters as compared to supermarkets. The average expenditure share representing healthful foods is 4.5 percentage points lower for supercenter baskets, relative to supermarket baskets. In terms of quantity shares, the difference is four percentage points. The average supercenter basket scores five points lower than supermarket baskets according to the HEI. The USDA scores do not lend themselves to easy interpretation, but the supercenter scores are significantly lower. To facilitate comparisons across all of the metrics, we calculated expected (predicted) scores for all six equations, by format, holding all other controls at zero. The average expected difference in healthfulness between baskets by format ranges from 11 percent for HEI scores to 26 percent for the USDA scores. Hence for all six measurements of food purchase healthfulness, supermarket baskets outperform supercenter baskets even while controlling for extensive demographics, regional and year effects, market characteristics, and seasonality.

Tables 8 and 9 here.

While the results strongly suggest that consumers purchase less healthful foods at supercenters, this difference between formats is not constant over time. For all six estimations of

[8] Market concentration data was calculated at the Nielsen Designated Market Area (DMA) level. There are 205 comprehensive and non-overlapping DMAs in the contiguous United States. We measured market share as the Hefindahl-Hirschman Index (HHI), which is given by the squared market share of each retailer operating in the DMA. The HHI was chosen over alternative measurements such as the four-firm or eight-firm concentration ratios because it exhibited the most variation across DMAs.

(2), the interactions between the supercenter term and the annual dummies generally show a closing of this gap throughout the years. For example, the average difference in healthful expenditure shares between formats was four percentage points wider in 1998 than it was in 2006. The average difference in HEI scores was nearly four points greater in 1999 than in 2006.

This trend may reflect developments by supercenters to position themselves as purveyors of healthful food. The supercenter format is dominated by Wal-Mart Supercenters which have received a lot of attention in the popular press in recent years for their forays into healthier food options. In 2006, Wal-Mart announced a major expansion in its offerings of organic foods, particularly produce (Warner, 2006). In 2007 Wal-Mart launched the Live Better Index, representing the company's commitment to "helping consumers save money and live better" (Wal-Mart Inc, 2007). In early 2011, the company announced a five-year program to improve the nutritional quality of its private label products, mostly by reducing sugar, sodium, and trans-fats (Torabi, 2011). Most of these developments have occurred at the tail end of our purchase data or after, but they reflect years of research and large financial investments towards the goal of successfully competing with conventional supermarkets in terms of healthful food options. Whatever the cause, this finding suggests it is important to focus on the time series aspect of the supercenter effect on food purchase healthfulness.

As a final exercise in studying differences between store formats, we restrict our attention only to those households that shop at both formats. We are unable to pinpoint the precise location of the stores at which Homescan responders shop and therefore we do not know whether households can realistically choose between supermarkets or supercenters when grocery shopping. In several major cities the nearest Wal-Mart Supercenter is at least 30 miles away, and in certain rural settings consumers' only realistic option is a supercenter. Therefore market-

structure characteristics may still be a major driver of the link between shopping format and healthfulness despite the controls in (9). We have established that, on average, households purchase less healthful foods at supercenters. However in this section we examine if this is still the case when consumers have the choice between the two.

To that end, we measure the percentage difference between supercenter and supermarket scores for each of the six healthfulness measures, by household and quarter. This difference is given by:

$$difference_{iqm} = \frac{(supermarketscore_{iqm} - supercenterscore_{iqm}}{supermarketscore_{iqm}} \quad (10)$$

Calculating percentage differences allows for comparisons not only across metrics but between analyses of the full sample versus the restricted sample. In this restricted setting, the results based on the HEI scoring are not meaningful and thus not reported.[9] We estimate a restricted version of (9) with these percentage differences as the dependent variables on the sample of only those cases of households shopping at both formats within a quarter. All variables pertaining to the supercenter dummy are removed, as each observation incorporates purchase data from both formats. The regression results are reported in table 10.

Table 10 here.

We measure the effect of shopping in the supercenter format by calculating expected percentage differences between formats using the estimated coefficients, i.e., the expectation of the dependent variable. All continuous variables are taken at their means. The regression results reflected a persistent phenomenon with respect to geographic regions that is not well explained by theory or evidence in prior empirical work. Specifically, for four of the healthfulness

[9] The HEI scores are based in large part upon variety in food consumption. HEI scores for shopping baskets showing insufficient variety are uniformly very low. In this restricted setting, too few baskets remaining in the sample exhibit sufficient variety to achieve meaningful regression results using HEI scores.

measures, the average difference in basket quality is considerably lower in the south than in the other three regions. Accordingly, table 10 includes the expected difference between formats, calculated individually for the four geographic regions.

Overall, examining only those consumers who regularly shop at both formats, food purchases are significantly more healthful at supermarkets than at supercenters. This is true based on expenditure shares and the USDA Scores, though the results are mixed for quantity shares of healthful foods. In the South, supermarket baskets score between five and eight percent higher than do supercenter baskets. For the West, Midwest, and Northeast, the results are very consistent and indicate that supermarket baskets are 21 to 22 percent more healthful. Certainly the difference between the south and the remaining regions calls for further investigation into product variety, and retail prices, among other factors. Nevertheless, the results based on the restricted sample support those of the full Homescan sample, indicating that consumers purchase significantly less healthful foods at supercenters than they do at supermarkets.

Measuring the Overall Effect of Supercenters

The previous section presented evidence that consumers purchase less healthful foods at supercenters, but did not account for consumer behavior across all shopping outlets. As such, it is possible that consumers partition their shopping between store formats, purchasing relatively healthful foods more frequently at supermarkets and less healthful foods more frequently at supercenters, resulting in no net effect to the healthfulness of household purchases. Such behavior may be explained by differences in relative prices between store formats. Additionally, the format-level analysis does not account for a potential indirect effect of supercenter presence, pertaining to consumers shopping at supermarkets. Recall that supercenter market share may

induce supermarkets to reduce their prices or enhance their product menus, either of which may have implications for the healthfulness of consumer purchases.

To understand the total effect of supercenter presence on food purchase healthfulness, we look at households total quarterly shopping baskets, including shopping trips to both supermarkets and supercenters. As before, we model food purchase healthfulness as a function of market structure variables, demographics, and temporal and spatial controls:

(11) $Healthfulness_{iqm}$ = f(Supercenter $Share_{iq}$, $Concentration_{iq}$, $Income_{iq}$, $Education_{iq}$, Free $Time_{iq}$, $Profession_{iq}$, $Race_i$, Regional $Effects_i$, Year $Effects_q$, Quarterly $Effects_q$, Household $Effects_i$) + Error

where all subscripts are familiar from (9) and once more all variable definitions are available in appendix B. Our approach in this section deviates from (9) in two key respects. One is that we now model supercenter share as a continuous variable, representing the market share of supercenter stores as reported by *Market Scope*, by DMAs. This allows for a testing of the change in consumers' purchase decisions as the local influence of supercenters grows. The other is that we model the healthfulness of consumers' shopping baskets as a function of unobservable characteristics that vary across households, specifically the attitude towards healthful foods. This approach suggests the use of panel estimation techniques, which are made possible by the nature of the Homescan data.

One potential issue endemic to this approach is the endogeneity of supercenter share, and in this instance we are referring primarily to Wal-Mart Supercenters. Although we assume that the differences in product menus between supercenters and supermarkets are negligible, this is not to say that there are not significant differences in the shopping preferences of consumers between the store formats. This gets to the heart of the classic endogeneity issue surrounding Wal-Mart's introduction and expansion in the U.S. Does Wal-Mart's success impact the

economic variable of interest, or is Wal-Mart's success due to the extant economic conditions? The issue is considerable enough to have spawned at least two literature reviews on the topic, by Basker (2007b) and Hicks (2008). As a result of this potential endogeneity and the related consequences on our estimate, we also estimate (8) while instrumenting for supercenter share. We employ two instruments, as selected from a variety of preliminary estimations. One is a classic Wal-Mart distance function, measuring the geographic distance of the DMA from Bentonville, AK, where Wal-Mart originated. The other is a county-level rurality score, as calculated by USDA-ERS. Wal-Mart stores have typically achieved greater success, in terms of market share, in rural areas as compared to urban ones (Franklin, 2001; Martens, 2008).[10]

Tables 11 and 12 present the results of estimating (11) for *Healthshare*, *USDAScore3*, and *HEIScore*. We estimate (11) two different ways-using OLS and fixed effects (FE) instrumental variable (IV) regression to exploit the panel nature of the data and to account for potential supercenter share endogeneity.[11] All regressions throughout are cleaned of probable outliers to eliminate recording errors in the Homescan data. The panel approach incorporates only those households that participated in the Homescan survey for at least four years, or 16 quarters. Hence these estimates are subject to smaller sample sizes.

Tables 11 and 12 here.

The OLS regressions for all three measures of healthfulness indicate that supercenter share has a negative and significant effect. Recall from the background section that we were unclear on the potential directions of this effect, though the format-level analysis provided cause to expect a negative relationship. Once we incorporate household FE through the panel nature of

[10] The county rurality codes are calculated and maintained by the Resource and Rural Economics Division of the Economic Research Service. The scores and the methodology behind their calculation are available at http://www.ers.usda.gov/briefing/rurality/ruralurbcon.

[11] The estimation results for all three USDA Scores are qualitatively similar and thus the results for scores 1 and 2 are not reported. The full set of results is available from the author upon request.

the data, we estimated magnitude of this effect falls for all three measures and becomes insignificant in the case of the USDA Score. For ease of interpretation, particularly across the different healthfulness measures, tables 11 and 12 include elasticities of purchase healthfulness with respect to supercenter share, calculated at the mean. We see that a one-percent increase in the market share of supercenter stores within a DMA is associated with an overall decrease in purchase healthfulness ranging from 0.5 to two percent. The elasticities are larger in magnitude for the OLS regressions, as the estimates for the FE IV regressions do not exceed 0.8 percent. The elasticity in the case of *USDAScore3* in the FE IV case is actually positive, though the coefficient on supercenter share is not significant.

The reason for the dampening of the magnitude may owe in part to the endogeneity of supercenter share, though this seems unlikely given that the Hausman test of endogeneity for supercenter share was rejected.[12] It is important to keep in mind when interpreting these findings that, when controlling for unobservable household characteristics through FE panel estimation, the estimated coefficients reflect the time series aspect of their relationships with purchase healthfulness. Therefore the smaller overall effect of supercenter share in a time series setting may owe to the efforts of such stores, particularly Wal-Mart, to offer and market increasingly healthful options (Stolberg, 2011). It is also possible that the regional dummies and Homescan demographics do not fully account for differences in preferences and shopping habits across markets. For example, Wal-Mart's food market share is the highest in the southern states and the Midwest, which is also where Americans eat the fewest fruits and vegetables and have the highest average BMIs (Baskin et al., 2005). Thus the OLS results may be capturing these persistent differences across space.

[12] However the first stage regressions indicate that the two instruments are good predictors of supercenter share.

It is worthwhile to discuss the supercenter findings in the context of Courtemanche and Carden (2011) (CC), which is closely related to our own work in motivation and implications. Recall that CC identified a causal relationship between Supercenters and obesity, calculating that an additional Supercenter per 100,000 residents increases the local obesity rate by 2.3 percent. We argue that our findings support and are in line with those of CC with respect to the effects of supercenter stores from the standpoint of health economics. We find that an increase in supercenter share, within a Nielsen DMA, significantly reduces the healthfulness of consumers' food purchases. Thus CC's findings may well be explained, in part, by this effect on the quality of food-at-home attributable to supercenters. Within our framework, it is not possible to estimate the potential effect of changes in consumer purchases on obesity. However it would be very worthwhile to decompose the total supercenter effect on health outcomes as a function of changes in food-at-home consumption, food-away-from home consumption, physical activity, and other relevant factors.

Conclusions

The prevalence and severity of obesity has been on the rise in the United States for many years across all regions of the country and nearly all demographics, especially children. The same is true for a number of other health concerns that can be connected, at least in part, to dietary quality. Diet and health have been the focus of research across many disciplines, not the least of which being economics, due to their substantial and varied implications. This study draws a connection between a key market structure characteristic in the industrial organization of today's food retail sector with the healthfulness of consumer purchases.

The results identify an important connection between the share of stores with the supercenter format and the healthfulness of consumers' food-at-home purchases. They also

demonstrate the importance of functional form and estimation technique when seeking to identify these economic relationships. In an OLS setting, we estimate that an increase in the market share of supercenters of one percent within a market area results in an overall decrease in purchase healthfulness of one to two percent. However if we exploit the panel nature of our data to isolate the time series effect, the magnitude of the effect is smaller, though still significant, between 0.5 and 0.8 percent.

In uncovering this relationship, we also showed that consumers typically purchase less healthful foods at supercenters as compared to supercenters. Quarterly shopping baskets purchased at supercenters, on average, contain significantly fewer fruits and vegetables and more packaged and processed foods by a wide margin. This holds true even when examining shoppers who regularly visit stores of both formats.

Finally, assuming a close correspondence between food purchase and consumption, this study provides further evidence that American households are not adhering to healthy eating guidelines, as defined by the USDA. Market-structure considerations aside, consumers purchase too few fruits, vegetables, and whole grains and purchase too many refined grains and processed foods high in sodium and added sugars. We demonstrate that this may be improving over time and in aggregate, but American consumers still have far to go before meeting USDA healthy eating guidelines.

The findings may have important policy implications for the future of food retail. Economists have long examined the impacts of market concentration and supercenters in order to inform policy on mergers, acquisitions, closures, expansions, and so forth. Proper analyses should account for as many key impacts as possible in order to get the complete picture. Given the prevalence of obesity and other concerns related to diet quality, the determinants of food-at-

home healthfulness should be among them. The continued expansion of supercenters, particularly Wal-Mart Supercenters, is planned by the firms (Wal-Mart, 2011) and has been predicted by researchers (Seiders et al., 2000). This may have important impacts on the food purchase decisions and dietary quality of Americans for many years to come.

This study poses several questions that suggest avenues for future research. We do not measure or comment on health outcomes resulting from changes in supercenter share. By linking our work with extensive data on calorie counts and nutritional attributes it may be possible to measure the effect on obesity or other health-related issues attributable to the effect of supercenter presence on food-at-home. Another important consideration is food-away-from home, which is substitutable with food-at-home, and it is not clear what effect, if at all, supercenters have on the quality or quantity of food-away-from-home purchases. Such information is important given that the latter category is typically less healthful for consumers (Todd et al., 2010). Finally, it is not at all clear why consumers might choose less healthful options at supercenters, as compared to supermarkets. This calls for an investigation into this phenomenon, likely starting with a study of the relative prices of healthful and unhealthful foods at both formats.

References

Basker, E. 2007a. "The Causes and Consequences of Wal-Mart's Growth." *The Journal of Economic Perspectives* 21(3):177-198.

Basker, E. 2007b. "When Good Instruments Go Bad: A Reply to Neumark, Zhang, and Ciccarella " *U. of Missouri Dept. of Economics Working Paper*.

Basker, E., and M. Noel. 2009. "The Evolving Food Chain: Competitive Effects of Wal-Mart's Entry into the Supermarket Industry." *Journal of Economics and Management Strategy* 18(4):977-1009.

Baskin, M.L.A., J., F. Franklin, and D.B. Allison. 2005. "The Prevalence of Obesity in the United States." *Obesity Reviews* 6(1):5-7.

Berges-Sennou, F., P. Bontems, and V. Requillart. 2004. "Economics of Private Labels: A Survey of Literature." *Journal of Agricultural & Food Industrial Organization* 2(3):1-23.

Bhattacharya, J. and M. K. Bundorf. 2009. "The Incidence of the Healthcare Costs of Obesity." *Journal of Health Economics* 28(3): 649-658.

Blisard, N., H. Stewart, and D. Jolliffe. 2004. "Low-Income Households' Expenditures on Fruits and Vegetables." *U.S. Department of Agriculture, Economic Research Service, Agirculture Information Bulletin No. 792-5*.

Block, J.P., R.A. Scribner, and K.B. DeSalvo. 2004. “Fast Food, Race/Ethnicity, and Income: A Geographic Analysis.” *American Journal of Preventative Medicine* 27(3):211-217.

Bonanno, A., and R.A. Lopez. 2009. "Competition Effects of Supermarket Services." *American Journal of Agricultural Economics* 91(3):555-568.

Center for Disease Control. "National Health and Nutrition Examination Survey, Examination Files, Dietary Interview - Individual Foods, First Day. 2003–04.".

Chou, S., M. Grossman, and H. Saffer. 2004. "An Economic Analysis of Obesity: Results from the Behavioral Risk Factor Surveillance System." *Journal of Health Economics* 23(3):565-587.

Cotterill, R.W. 1986. "Market Power in the Retail Food Industry: Evidence from Vermont." *The Review of Economics and Statistics* 68(3):379-386.

Courtemanche, C. and A. Carden. 2011. "Supersizing Supercenters? The Impact of Wal-Mart Supercenters on Body Mass Index and Obesity." *Journal of Urban Economics* 69(2):165-181.

Cutler, D.M., E.L. Glaeser, and J.M. Shapiro. 2003. "Why Have Americans Become More Obese?" *Journal of Economic Perspectives* 17(3):93-118.

Doyle, P., and C.B. Weinberg. 1973. "Effective New Product Decisions for Supermarkets." *Operational Research Quarterly* 24(1):45-54.

DSN Retailing Today (2003) Supermarket Survey Yields Ambiguous Results.

Eikenberry, N., and C. Smith. 2004. "Healthful Eating: Perceptions, Motivations, Barriers, and Promoters in Low-Income Minnesota Communities." *Journal of the American Dietary Association* 104:1158-1161.

Ellickson, P. 2006. "Quality Competition in Retailing: A Structural Analysis." *International Journal of Industrial Organization* 24(3):521-540.

Frank, L.D., M.A. Andresen, and T.L. Schmid. 2004. "Obesity Relationships with Community Design, Physical Activity, and Time Spent in Cars." *American Journal of Preventive Medicine* 27(2):87-96.

Franklin, A.W. 2001. "The Impact of Wal-Mart Supercenters on Supermarket Concentration in U.S. Metropolitan Areas." *Agribusiness* 17(1):105-114.

Galobardes, B., A. Morabia, and M.S. Bernstein. 2000. "The Differential Effect of Education and Occupation on Body Mass Index in a Sample of Working People of the General Population." *Annals of Epidemiology* 10(8):532-537.

Guenther, P.M., S.M. Krebs-Smith, and B.B. Reeve. 2007. "Development and Evaluation of the Healthy Eating Index-2005: Technical Report. ." *U.S. Department of Agriculture, Center for Nutrition Policy and Promotion*.

Hausman, J., and E. Leibtag. 2007. "Consumer Benefits from Increased Competition in Shopping Outlets: Measuring the Effect of Wal-Mart." *Journal of Applied Econometrics* 22(7):1157-1177.

Hicks, M.J. 2008. "Estimating Wal-Mart's Impacts in Maryland: A Test of Identification Strategies and Endogeneity Tests." *Eastern Economic Journal* 34:56-73.

Koski-Lahti, M, E. Vartiainen, S. Mannisto, and P. Pietinen. 2000. "Age, Education, and Occupation as Determinants in Body Mass Index in Finland from 1982 to 1997." *International Journal of Obesity* 24(12):1669-1676.

Lamm, R.M. 1981. "Prices and Concentration in the Food Retailing Industry." *The Journal of Industrial Economics* 30(1):67-78.

Leibtag, E., C. Barker, and P. Dutko. 2010. "How Much Lower are Prices at Discount Stores? ." *ERR-105, U.S. Department of Agriculture, Economic Research Service*.

Mancino, L., and C. Newman. 2007. "Who Has Time to Cook? How Family Resources Influence Food Preparation." *ERR-40, U.S. Department of Agriculture, Economic Research Service, May, 2007*.

Market Force (2011) Market Force Study Finds Consumers Perceive ALDI as Grocery Price Leader.

Martens, B. 2008. “The Effect of Wal-Mart Supercenters on Retail Grocery Concentration.” *Journal of Food Distribution Research* 39(3):14-28.

Mushi-Brunt, C., D. Haire-Joshu, and M. Elliott. 2007. "Food Spending Behaviors and Perceptions Are Associated With Fruit and Vegetable Intake Among Parents and Their Preadolescent Children." *Journal of Nutrition Education and Behavior* 39:26-30.

Philipson, T.J., and R.A. Posner. 2003. “The Long-Run Growth in Obesity as a Function of Technological Change.” *Journal of Perspectives in Biology and Medicine* 46(3):S87-S107.

Progressive Grover. 2011. “Smart Snacking.”

Reedy, J., S.M. Krebs-Smith, and C. Bosire. 2010. "Evaluating the Food Environment: Application of the Health Eating Index-2005." *American Journal of Preventative Medicine* 38(5):465-471.

Seiders, K., C. Simonides, and D.J. Tigert. 2000. "The Impact of Supercenters on Traditional Food Retailers in Four Markets." *International Journal of Retail and Distribution Management* 28(45):181-193.

Singh, V.P., K.T. Hansen, and R.C. Blattberg. 2006. "Market Entry and Consumer Behavior: An Investigation of a Wal-Mart Supercenter." *Marketing Science* 25(5):457-476.

Stolberg, S.G. (2011) Wal-Mart Shifts Strategy to Promote Healthy Foods. New York, NY, Bloomberg News.

Supermarket News (2011) 2011's Top 75 North American Food Retailers. http://supermarketnews.com/profiles/top75/.

Todd, J.E., L. Mancino, E. Leibtag, and C. Tripodo. 2010. "Methodology Behind the Quarterly Food-at-Home Price Database." *U.S. Department of Agriculture, Economic Research Service, Technical Bulletin No. 1926*.

Torabi, F. (2011) Shoppers Shifting to Private-Label Brands. Available at: http://moneywatch.bnet.com/saving-money/blog/so-money/shoppers-shifting-to-private-label-brands/2230/.

Trade Dimensions (2006) *Market Scope*.

U.S. Department of Agriculture. 2010. "Healthy Eating Index—2005 Development and Evaluation Technical Support Files, 2010." *Center for Nutrition Policy and Promotion*.

Vallianatos, M., A. Shaffer, M. Beery, R. Gottlieb, and A. Wheatley. 2004. "Thinking Outside the Big Box: Food Access, Labor, Land Use and the Wal-Mart Way." *UEPI Papers Report, 2004*.

Volpe, R.J., and N. Lavoie. 2008. "The Effect of Wal-Mart Supercenters on Grocery Prices in New England." *Review of Agricultural Economics* 30(1):4-26.

Wal-Mart Inc (2007) Wal-Mart Launches the Live Better Index with First Focus on the Environment.

Wal-Mart Inc (2011) Corporate Profile.

Warner, M. (2006) Wal-Mart Eyes Organic Foods. New York, New York, Available at: http://www.nytimes.com/2006/05/12/business/12organic.html.

Western Farm Press (2005) Wal-Mart Not a Bad Big Food Buyer After All.

Woo, B.-J., C.L. Huang, J.E. Epperson, and B. Cude. 2001. "Effect of a New Wal-Mart Supercenter on Local Retail Food Prices." *Journal of Food Distribution Research* 32:173-180.

Xie, B., F.D. Gilliland, Y. Ly, and H.R. Rockett. 2003. "Effects of Ethnicity, Family Income, and Education on Dietary Intake among Adolescents." *Preventive Medicine* 36(1):30-40.

Yu, C.-c., and J.M. Connor. 2002. "The Price-Concentration Relationship in Retailing: Retesting Newmark." *Agribusiness* 18(4):413-426.

Table 1: Average Expenditure Shares of QFAHPD Food Categories.

Food Group	Category	USDA Healthful	Mean Expenditure Share (N = 656,212)
	Fruits		
1	Whole fresh/frozen	Yes	2.25
2	Whole canned	Yes	1.01
3	Fruit juice	Yes	2.85
	Vegetables		
4	Dark green fresh/frozen	Yes	0.44
5	Dark green canned	Yes	0.06
6	Orange fresh/frozen	Yes	0.55
7	Orange canned	Yes	0.12
8	Starchy fresh/frozen	Yes	1.80
9	Starchy canned	Yes	0.37
10	Other-nutrient dense fresh/frozen	Yes	0.49
11	Other-nutrient dense canned	Yes	0.44
12	Other-mostly water fresh/frozen	Yes	1.52
13	Other-mostly water canned	Yes	0.52
14	Legumes fresh/frozen/dried	Yes	0.07
15	Legumes canned/processed	Yes	0.28
	Grains		
16	Whole grain packaged (bread, rolls, pita, tortilla, rice, pasta, cereal)	Yes	1.81
17	Whole grain flour and mixes	Yes	0.02
18	Whole grain frozen/ready to cook	Yes	0.01
19	Refined packaged (bread, rolls, pita, tortilla, rice, pasta, cereal)	No	8.45

20	Refined flour and mixes	No	0.42
21	Refined frozen/ready to cook	No	0.92
	Dairy		
22	Low fat milk	Yes	2.48
23	Low fat cheese	Yes	0.43
24	Low fat yogurt & other	Yes	1.45
25	Regular fat milk	No	3.34
26	Regular fat cheese	No	4.50
27	Regular fat yogurt & other	No	0.21
	Meats		
28	Low fat meat fresh/frozen	Yes	1.19
29	Regular meat fresh/frozen	No	5.30
30	Regular meat canned	No	0.22
31	Poultry fresh/frozen	Yes	0.66
32	Poultry canned	Yes	0.07
33	Fish fresh/frozen	Yes	0.70
34	Fish canned	Yes	0.83
35	Nuts and Seeds raw	Yes	1.34
36	Nuts and Seeds processed/nut butters	Yes	0.57
37	Eggs	Yes	1.33
	Fats and Oils		
38	Oils	Yes	0.86
39	Solids	No	2.31
	Sugar and sweeteners		
40	Raw	No	0.85

	Beverages		
41	Carbonated non alcoholic	No	3.88
42	Fruit drinks and other non-carbonated sugary beverages	No	2.45
43	Water	Yes	1.22
	Commercially prepared items		
44	Sweet frozen (ice cream, frozen desserts)	No	4.97
45	Sweet mixes (pancake, muffin and cake mixes)	No	1.18
46	Sweet packaged (cookies, candy bars, bars)	No	6.66
47	Sweet ready-to-eat (bakery items)	No	2.75
48	Not sweet frozen (pizzas,french fries, fish sticks and entrees)	No	9.21
49	Not sweet canned (soups, sauces, etc)	No	3.56
50	Not sweet packaged/Snacks	No	7.78
51	Not sweet packaged/Meals and sides	No	2.31
52	Not sweet ready-to-eat (hot and cold deli items)	No	1.02

Table 2: Average Expenditure Shares by QFAHPD-Compatible CNPP Food Categories.

CNPP Food Category	QFAHDB Food Groups	Household Shopping Basket Expenditure Shares	
Grains		USDA[a] Food Plan	Homescan households
All whole-grain products Whole grain breads, rice, pasta, pastries (incl. whole grain flours) (H) Whole grain cereals (incl. hot cereal mixes) (H) Popcorn, other whole grain snacks (H)	16, 17, 18	10.09	1.74
Non-whole grain breads, cereals, rice, pasta, pies pastries, snacks, and flours	19, 20, 21, 50, 51	6.10	16.91
Vegetables			
All potato products (H)	8, 9	1.77	1.87
Dark-green vegetables (H)	4, 5	5.59	0.56
Orange vegetables (H)	6, 7	2.61	0.59
Canned and dry beans, lentils, and peas (legumes) (H)	14, 15	8.32	0.28
Other vegetables (H)	10, 11, 12, 13	8.66	3.35
Fruits			
Whole fruits (H)	1, 2	16.49	4.82
Fruit juices (H)	3	1.86	2.62
Milk products			
Whole milk products Whole milk, yogurt, and cream Milk drinks and milk desserts	25, 27	0.86	2.98
Lower fat and skim milk and low-fat yogurt (H)	22, 24	8.77	3.30
All cheese (including cheese soup and sauce)	23, 26	0.60	4.67
Meat and beans			
Beef, pork, veal, lamb, and game	28, 29	5.31	9.03
Chicken, turkey, and game birds (H)	31, 32	2.69	2.09
Fish and fish products (H)	33, 34	11.92	2.06

Bacon, sausages, and luncheon meats (including spreads)	30, 52	0.91	2.06
	35, 36	3.16	2.22
Nuts, nut butters, and seeds (H)			
	37	0.12	1.06
Eggs and egg mixtures (H)			
Other foods			
Fats and condiments	38, 39	1.79	2.67
Table fats, oils, and salad dressings			
Gravies, sauces, condiments, and spices			
Coffee and tea (H)	N/A	0.02	N/A
Soft drinks, sodas, fruit drinks, and aids (including rice beverages)	41, 42	1.33	8.37
Sugars, sweets, and candies	40, 44, 45, 46, 47	0.41	13.80
Soups	49	0.51	3.26
Ready-to-serve and condensed soups			
Dry soups (dry)			
Frozen or refrigerated entrees (including pizza, fish sticks, and frozen meals)	48	0.18	8.40

a: The USDA Food Plan shares are based upon the recommended dollar costs of feeding a representative family consisting of one male and one female, aged 19-50, one child aged 9-11, and one child aged 6-8, according to the Liberal Food Plan. The authors gratefully acknowledge Mark Lino for providing the weekly dollar cost estimates, which are not available in the CNPP publications.

H denotes healthful food categories, as determined previously by the QFAHPD food groups.

Table 3: Components and Standards of the Health Eating Index-2005.

Component	*Max score*	*Standard*
Total fruit	5	≥0.8 cup eq/1000 kcal
Whole fruit	5	≥0.4 cup eq/1000 kcal
Total vegetables	5	≥1.1 cup eq/1000 kcal
Dark green and orange vegetables & legumes	5	≥0.4 cup eq/1000 kcal
Total grains	5	≥0.8 cup eq/1000 kcal
Whole grains	5	≥3.0 oz eq/1000 kcal
Milk	10	≥1.5 oz eq/1000 kcal
Meat and beans	10	≥1.3 cup eq/1000 kcal
Oils	10	≥12 g/1000 kcal
Saturated fat	10	≤7% of energy
Sodium	10	≤ 0.7 g/1000 kcal
Calories from fat, alcohol & added sugar	20	≤ 20% of energy

Source: Guenther et al. (2007).

Table 4: Mean Effects of Consumption of Different Foods on HEI-2005.

	Regression coefficient	Standard error	QFAHPD codes
Intercept	45.8553	0.7321	-
Whole fruit	0.0343	0.0027	1,2
Fruit juice	0.0100	0.0011	3
Dark green vegetables	0.0302	0.0082	4,5
Orange vegetables	0.0399	0.0062	6,7
Starchy vegetables	0.0059	0.0039	8,9
Other-nutrient dense vegetables	0.0132	0.0035	10,11
Other-mostly water vegetables	0.0203	0.0035	12,13
Legumes	0.0153	0.0048	14,15
Whole grains	0.0268	0.0023	16,17,18
Refined grains	0.0109	0.0031	19,20,21
Low fat dairy	0.0036	0.0010	22,23,24
Regular fat dairy	-0.0024	0.0006	25,26,27
Low fat red meat	0.0130	0.0049	28
Regular fat red meat	-0.0150	0.0019	29,30
Poultry	0.0030	0.0026	31,32
Fish	0.0112	0.0041	33,34
Nuts and seeds	0.1070	0.0125	35,36
Eggs	-0.0072	0.0040	37
Oils	0.0724	0.0823	38
Solid fat	-0.0253	0.0076	39
Sugar and sweeteners	-0.0092	0.0117	40
Carbonated beverages	-0.0050	0.0004	41
Noncarbonated beverages	-0.0052	0.0008	42
Water	-0.0007	0.0003	43
Frozen commercially prepared sweet items	-0.0150	0.0026	44
Other commercially prepared sweet items	-0.0086	0.0021	45,46,47
Frozen commercially prepared non-sweet items	-0.0033	0.0051	48
Canned commercially prepared non-sweet items	-0.0017	0.0012	49
Packaged commercially prepared snacks	0.0428	0.0071	50
Other commercially prepared non-sweet items	-0.0007	0.0007	51,52

Source: Authors' calculations using 2003-04 NHANES.

Table 5: Average Household Expenditure Shares by Selected Food Categories and Store Format.

Food Category	**QFAHPD Food Groups**	**Supermarket Mean Share** (%) **(N = 643,051)**	**Supercenter Mean Share** (%) **(N = 244,037)**
Fruits	1-3	6.63	4.77
Vegetables	4-15	7.27	5.02
Whole Grains	16-18	1.89	1.84
Refined Grains	19-21	10.04	9.13
Healthful Meats	28, 31-37	18.13	14.87
Unhealthful Meats	29, 30	5.65	5.17
Sugars and Sweeteners	40	0.84	0.88
Commercially Prepared Items	44-52	37.31	45.09

Source: Authors' calculations using Nielsen Homescan Data, 1998-2006.

Table 6: Average Scores for All Metrics of Basket Healthfulness, by Store Format.

	Supermarket Baskets	**Supercenter Baskets**	**Percent Difference**
Healthful Expenditure Share (%)	26.86 (13.14)	21.93 (18.59)	22.48
Healthful Quantity Share (%)	35.24 (18.96)	29.43 (24.69)	19.74
USDAScore1	6.29 (2.06)	4.26 (2.31)	47.65
USDAScore2	7.86 (2.65)	5.56 (3.18)	41.37
USDAScore3	6.59 (2.17)	4.60 (2.51)	43.26
HEI Score (0-100)	59.83 (15.72)	52.71 (11.42)	13.53

Standard deviations are in parentheses.
All percentage differences are statistically significant at the 0.01 level.

Table 7: Summary Statistics for the Econometric Variables.

Variable	Mean	St. Deviation	Q1	Q3
Healthshare	32.01	13.31	22.49	40.37
USDAScore1	8.67	3.38	6.43	10.27
USDAScore2	10.24	4.21	7.54	12.09
USDAScore3	9.13	3.80	6.69	10.76
HEIScore	2.36	4.26	-0.10	4.09
HHI	0.21	0.09	0.15	0.23
Supercenter share (%)	5.48	8.43	0.40	5.80
HH Income	$47,300	$21,000	$32,500	$84,999
Education (Nielsen bracket)[a]	3.10	2.12	0	5
Hours Worked (Nielsen bracket)	3.37	3.15	0	3
Black (%)	13.58	34.25		
Asian (%)	2.78	16.44		
Other Race (%)	5.02	21.84		
Professional (%)	21.11	40.81		
Manager (%)	15.13	35.83		
Clerical (%)	6.38	24.43		
Sales (%)	5.13	22.05		
Craftsman (%)	9.89	29.86		
Operative (%)	6.27	24.24		
Service (%)	4.60	20.94		
Unemployed (%)	29.55	45.63		

a: Appendix B provides definitions of the explanatory variables and their categorization by Nielsen.

Table 8: Regression Results for Equation (9), the Determinants of Consumers' Healthful Purchasing Decisions, by USDA Recommended Food Groups and HEI Scores.

Variable	Expenditure HealthShare	Quantity HealthShare	HEI Score
Intercept	24.180***	30.910***	47.110***
	(0.144)	(0.218)	(0.555)
Supercenter	-4.498***	-3.795***	-5.090***
	(0.139)	(0.210)	(0.150)
Concentration	1.272***	3.009***	1.570***
	(0.214)	(0.323)	(0.234)
Income	0.211***	0.398***	0.150***
	(0.004)	(0.006)	(0.004)
Other Race	-0.014	-0.798***	-1.470***
	(0.072)	(0.110)	(0.079)
Black	0.020	-1.916***	-2.760***
	(0.052)	(0.078)	(0.056)
Asian	1.213***	1.777***	-1.700***
	(0.110)	(0.165)	(0.118)
Professional	1.785***	3.207***	1.480***
	(0.110)	(0.167)	(0.121)
Manager	0.550***	1.164***	0.731***
	(0.112)	(0.170)	(0.122)
Clerical	0.800***	1.736***	0.754***
	(0.122)	(0.186)	(0.134)
Sales	0.743***	1.585***	0.496***
	(0.123)	(0.187)	(0.135)
Craftsman	-0.805***	-1.127***	-0.329***
	(0.113)	(0.172)	(0.124)
Operative	-1.392***	-1.938***	-0.780***
	(0.118)	(0.179)	(0.129)
Service	-0.203*	0.070	-0.125
	(0.127)	(0.192)	(0.139)
Unemployed	4.545***	6.367***	3.630***
	(0.113)	(0.171)	(0.124)
Education	0.439***	0.491***	0.280***
	(0.12)	(0.019)	(0.014)
HoursWorked	-0.363***	-0.515***	-0.303***
	(0.009)	(0.014)	(0.010)
Married	2.295***	2.296***	2.470***
	(0.049)	(0.073)	(0.054)
HouseholdSize	-1.414***	-2.321***	-1.920***
	(0.014)	(0.021)	(0.015)
Northeast	1.170***	2.083***	1.970***
	(0.048)	(0.073)	(0.054)
Midwest	-0.214***	0.236***	1.110***
	(0.042)	(0.064)	(0.047)

West	-0.928***	-0.206***	-1.080***
	(0.045)	(0.068)	(0.049)
Q1	0.636***	1.256***	-5.500***
	(0.044)	(0.067)	(0.096)
Q2	0.205***	0.082**	0.395***
	(0.022)	(0.033)	(0.105)
Q3	-0.083***	-0.065***	0.465***
	(0.015)	(0.022)	(0.102)
1998	-0.220***	-3.728***	-0.075
	(0.088)	(0.132)	(0.099)
1999	-2.462***	-3.006***	-0.463***
	(0.094)	(0.143)	(0.096)
2000	-2.586***	-3.221***	-0.629***
	(0.091)	(0.139)	(0.096)
2001	-2.667***	-2.979***	0.304***
	(0.089)	(0.135)	(0.059)
2002	-2.412***	-2.939***	0.519***
	(0.086)	(0.131)	(0.060)
2003	-1.783***	-2.100***	0.327***
	(0.086)	(0.130)	(0.049)
2004	-1.323***	-1.705***	-0.812***
	(0.053)	(0.080)	(0.024)
2005	-1.155***	-1.129***	-0.687***
	(0.053)	(0.080)	(0.016)
Super1998	-4.014***	-3.488***	0.370*
	(0.236)	(0.355)	(0.216)
Super1999	-5.039***	-7.247***	-3.570***
	(0.216)	(0.326)	(0.227)
Super2000	-4.099***	-5.885***	-3.220***
	(0.205)	(0.309)	(0.216)
Super2001	-2.915***	-4.718***	-2.340***
	(0.187)	(0.282)	(0.198)
Super2002	-1.487***	-2.393***	-0.962***
	(0.184)	(0.278)	(0.196)
Super2003	-0.609***	-0.831***	-0.044***
	(0.181)	(0.272)	(0.193)
Super2004	-0.096	-0.037	-0.032***
	(0.101)	(0.153)	(0.109)
Super2005	0.323***	0.133	-0.186*
	(0.100)	(0.151)	(0.108)
N	722,813	735,712	682,421
Adj. R^2	0.094	0.072	0.1084
Expected % Difference between Formats[a]	18.602	12.278	10.804

Standard errors are in parentheses.

***: Coefficient is significant at the 0.01 level. **: At the 0.05 level. *: At the 0.10 level.

a: The expected difference between formats is calculated by holding all other controls at zero. Thus it is simply the percentage difference between the predicted score for supermarket baskets (the intercept), and the predicted score for supercenter baskets (the intercept plus the Supercenter coefficient).

Table 9: Regression Results for Equation (9), the Determinants of Consumers' Healthful Purchasing Decisions, by USDA Scores.

Variable	USDA Score 1	USDA Score 2	USDA Score 3
Intercept	5.492***	7.356***	5.895***
	(0.023)	(0.031)	(0.025
Supercenter	-1.431***	-1.801***	-1.406***
	(0.022)	(0.029)	(0.024)
Concentration	0.154***	0.179***	0.160***
	(0.035)	(0.046)	(0.037)
Income	0.019***	0.011***	0.018***
	(0.001)	(0.001)	(0.001)
Other Race	0.001	0.095***	0.011
	(0.012)	(0.015)	(0.012)
Black	-0.005	0.181***	0.050***
	(0.008)	(0.011)	(0.009)
Asian	-0.404***	-0.286***	-0.352***
	(0.018)	(0.023)	(0.019)
Professional	0.076***	0.058**	0.065***
	(0.018)	(0.024)	(0.019)
Manager	0.057***	0.050**	0.050**
	(0.018)	(0.024)	(0.019)
Clerical	-0.054***	-0.112***	-0.068***
	(0.020)	(0.026)	(0.021)
Sales	0.052***	0.056**	0.039*
	(0.020)	(0.026)	(0.021)
Craftsman	0.050***	0.074***	0.043**
	(0.018)	(0.024)	(0.020)
Operative	-0.023	0.008	-0.025
	(0.019)	(0.025)	(0.020)
Service	-0.041**	-0.050*	-0.045**
	(0.021)	(0.027)	(0.022)
Unemployed	0.132***	0.101***	0.180***
	(0.018)	(0.024)	(0.019)
Education	0.057***	0.063***	0.069***
	(0.002)	(0.003)	(0.002)
HoursWorked	-0.016***	-0.002	-0.014***
	(0.002)	(0.002)	(0.002)
Married	0.574***	0.503***	0.552***
	(0.008)	(0.010)	(0.008)
HouseholdSize	0.103***	0.061***	0.071***
	(0.002)	(0.003)	(0.002)
Northeast	0.027***	-0.083***	0.010
	(0.008)	(0.010)	(0.008)
Midwest	0.051***	-0.014	0.029***
	(0.007)	(0.009)	(0.007)

West	-0.298***	-0.326***	-0.332***
	(0.007)	(0.010)	(0.008)
Q1	0.015**	0.006	-0.006
	(0.007)	(0.009)	(0.008)
Q2	-0.005	-0.011**	-0.021***
	(0.004)	(0.005)	(0.004)
Q3	-0.039***	-0.036***	-0.053***
	(0.002)	(0.003)	(0.003)
1998	-0.583***	-0.247***	-0.651***
	(0.014)	(0.019)	(0.015)
1999	-0.229***	-0.422***	-0.291***
	(0.015)	(0.020)	(0.016)
2000	-0.246***	-0.401***	-0.307***
	(0.015)	(0.020)	(0.016)
2001	-0.303***	-0.409***	-0.356***
	(0.014)	(0.019)	(0.015)
2002	-0.333***	-0.441***	-0.386***
	(0.014)	(0.019)	(0.015)
2003	-0.296***	-0.380***	-0.332***
	(0.014)	(0.018)	(0.015)
2004	-0.144***	-0.193***	-0.160***
	(0.009)	(0.011)	(0.009)
2005	-0.125***	-0.200***	-0.147***
	(0.009)	(0.011)	(0.009)
Super1998	-1.073***	-1.728***	-1.069***
	(0.037)	(0.049)	(0.039)
Super1999	-1.266***	-1.567***	-1.326***
	(0.034)	(0.045)	(0.037)
Super2000	-1.049***	-1.321***	-1.077***
	(0.033)	(0.043)	(0.035)
Super2001	-0.759***	-0.970***	-0.770***
	(0.030)	(0.039)	(0.032)
Super2002	-0.431***	-0.535***	-0.433***
	(0.029)	(0.039)	(0.031)
Super2003	-0.189***	-0.265***	-0.209***
	(0.029)	(0.038)	(0.031)
Super2004	-0.049***	7.356***	-0.055***
	(0.016)	(0.031)	(0.017)
Super2005	0.036**	-1.801***	0.041**
	(0.016)	(0.029)	(0.017)
N	727,815	721,968	720,749
Adj. R^2	0.204	0.156	0.179
Expected % Difference between Formats[a]	26.056	24.483	23.851

Standard errors are in parentheses.

***: Coefficient is significant at the 0.01 level. **: At the 0.05 level. *: At the 0.10 level.

a: The expected difference between formats is calculated by holding all other controls at zero. Thus it is simply the percentage difference between the predicted score for supermarket baskets (the intercept), and the predicted score for supercenter baskets (the intercept plus the Supercenter coefficient).

Table 10: Regression Results for the Restricted Sample of Households that Shop at Both Formats.

Variable	Exp. Health Share	Qnt. Health Share	USDAScore 1	USDAScore 2	USDAScore 3
Intercept	-14.340**	-37.500***	-14.165***	-16.341***	-14.774***
	(5.814)	(11.638)	(3.320)	(3.753)	(3.403)
Concentration	11.436	20.153	17.450***	15.709***	16.043***
	(7.298)	(14.478)	(4.139)	(4.670)	(4.243)
Income	1.109***	1.545***	1.051***	1.052***	0.977***
	(0.149)	(0.295)	(0.084)	(0.095)	(0.087)
Other Race	-1.059	-0.350	-2.201	-2.048	-2.390
	(3.940)	(7.821)	(2.237)	(2.521)	(2.292)
Black	-4.775**	-9.210**	0.997	-0.196	1.009
	(2.182)	(4.328)	(1.237)	(1.398)	(1.270)
Asian	-3.704	-13.940	-14.762***	-9.190*	-18.126***
	(9.047)	(17.736)	(4.987)	(5.730)	(5.190)
Professional	-2.833	2.090	0.774	1.523	2.067
	(4.668)	(9.314)	(2.652)	(2.999)	(2.718)
Manager	-6.100	-4.655	-1.014	0.576	0.890
	(4.7410	(9.467)	(2.696)	(3.049)	(2.763)
Clerical	-2.362	-4.172	-1.196	-0.572	-1.346
	(5.267)	(10.513)	(2.996)	(3.388)	(3.070)
Sales	-2.614	-6.304	1.906	2.055	1.372
	(5.304)	(10.555)	(3.006)	(3.403)	(3.086)
Craftsman	-8.266*	-13.380	-5.786**	-6.798**	-4.647*
	(4.740)	(9.463)	(2.696)	(3.049)	(2.763)
Operative	-6.522	-8.042	-0.869	-1.721	-0.246
	(4.921)	(9.825)	(2.805)	(3.170)	(2.872)
Service	-11.441**	-8.795	-4.288	-3.168	-2.288
	(5.403)	(10.790)	(3.069)	(3.470)	(3.145)
Unemployed	1.527	11.760	6.477**	5.443*	6.210**
	(4.911)	(9.794)	(2.792)	(3.156)	(2.860)
Education	-0.165	0.035	-0.508	-0.750**	-0.764**
	(0.554)	(1.097)	(0.313)	(0.353)	(0.322)
Hours Worked	-0.575	-2.081**	-0.110	0.089	-0.117
	(0.431)	(0.854)	(0.244)	(0.275)	(0.250)
Married	1.972	8.868**	0.026	-0.235	1.150
	(2.288)	(4.532)	(1.286)	(1.451)	(1.324)
Household Size	0.007	-2.511**	0.775**	0.474	0.808**
	(0.619)	(1.235)	(0.354)	(0.400)	(0.363)
Northeast	15.227***	21.202***	17.385***	15.735***	16.364***
	(2.169)	(4.312)	(1.234)	(1.391)	(1.265)

Midwest	14.440***	7.834**	16.722***	16.159***	16.542***
	(1.798)	(3.577)	(1.022)	(1.154)	(1.050)
West	15.168***	19.871***	15.710***	17.297***	15.473***
	(2.628)	(5.224)	(1.492)	(1.683)	(1.531)
Q1	-0.085	0.284	0.721	0.550	1.149
	(1.874)	(3.722)	(1.067)	(1.204)	(1.094)
Q2	-0.942	-4.611**	-0.936*	-0.746	-0.902*
	(0.938)	(1.863)	(0.533)	(0.602)	(0.547)
Q3	-1.828***	-3.337***	-1.629***	-1.512***	-1.240***
	(0.624)	(1.241)	(0.355)	(0.401)	(0.364)
1998	16.860***	17.099**	28.067***	30.971***	25.903***
	(3.598)	(7.101)	(2.001)	(2.256)	(2.065)
1999	22.086***	28.032***	27.895***	28.458***	27.133***
	(3.055)	(6.075)	(1.744)	(1.968)	(1.789)
2000	16.460***	21.015***	22.736***	23.578***	21.283***
	(2.890)	(5.736)	(1.644)	(1.857)	(1.688)
2001	11.630***	18.135***	18.342***	19.142***	17.450***
	(2.677)	(5.314)	(1.522)	(1.717)	(1.561)
2002	7.248***	-0.303	12.772***	13.229***	11.806***
	(2.636)	(5.231)	(1.500)	(1.693)	(1.539)
2003	-0.355	-1.516	6.778***	7.315***	6.761***
	(2.618)	(5.191)	(1.490)	(1.695)	(1.529)
2004	-2.892	-9.295*	2.426*	2.918*	2.061
	(2.619)	(5.193)	(1.490)	(1.682)	(1.529)
2005	-6.513**	-7.767	0.439	1.650	0.267
	(2.607)	(5.168)	(1.482)	(1.676)	(1.522)
Exp. Difference, South[a]	6.775	-10.897*	7.734***	5.140*	4.567**
Exp. Difference, Northeast	22.002***	10.305*	25.119***	20.875***	20.931***
Exp. Difference, Midwest	21.215***	-30.628***	24.456***	21.299***	21.109***
Exp. Difference, West	21.943***	8.974	23.444***	22.437***	20.040***
N	31,017	31,981	33,194	32,824	32,605
Adj. R^2	0.012	0.005	0.040	0.0321	0.035

Standard errors are in parentheses.

***: Coefficient is significant at the 0.01 level. **: At the 0.05 level. *: At the 0.10 level.

a: For each expected value, we tested if the linear combination of coefficients producing the reported values is statistically different from zero. Hence (***) indicates that the value is different from zero at the 0.01 level, (**) at the 0.05 level, and (*) at the 0.10 level.

Table 11: Results of Estimating (11) for Healthshare and USDAScore2 using OLS and Fixed Effects IV.

	Expenditure Healthshare		USDAScore3	
Variable	OLS	Fixed Effects IV	OLS	Fixed Effects IV
Supercenter	-7.747***	-5.052**	-1.698***	0.305
share	(0.544)	(2.095)	(0.091)	(0.418)
HHI	-4.719***	2.197	-0.338***	-0.006***
	(0.393)	(1.429)	(0.066)	(0.284)
HH Income	0.205***	-0.101***	0.016***	-0.012***
	(0.006)	(0.009)	(0.001)	(0.002)
Black	-0.017	-0.416	0.088***	-0.006
	(0.095)	(0.358)	(0.016)	(0.070)
Asian	4.650***	-1.248***	-0.007	-0.185**
	(0.212)	(0.369)	(0.040)	(0.079)
Other Race	-0.808***	-0.172	0.097***	0.021
	(0.152)	(0.206)	(0.026)	(0.041)
Professional	2.557***	1.031***	-0.174***	-0.068*
	(0.228)	(0.221)	(0.038)	(0.042)
Manager	1.901***	3.008***	-0.474***	-0.103**
	(0.229)	(0.218)	(0.038)	(0.042)
Clerical	0.959***	0.545**	-0.127***	-0.070
	(0.250)	(0.243)	(0.041)	(0.046)
Sales	2.427***	1.032***	0.021	-0.068
	(0.254)	(0.254)	(0.042)	(0.049)
Craftsman	-2.736***	-2.215***	-0.378***	0.050
	(0.233)	(0.226)	(0.038)	(0.043)
Operative	-0.826***	0.839***	-0.069*	-0.031
	(0.250)	(0.246)	(0.041)	(0.046)
Service	-0.690***	0.309	-0.232***	0.009
	(0.263)	(0.269)	(0.043)	(0.085)
Unemployed	2.496***	-1.290***	0.452***	0.167***
	(0.239)	(0.233)	(0.039)	(0.044)
Education	0.948***	-0.196***	0.156***	0.161***
	(0.021)	(0.038)	(0.004)	(0.007)
Hours Worked	-1.751***	-1.089***	0.140***	0.097***
	(0.039)	(0.047)	(0.007)	(0.009)
Northeast	-0.657***	-0.441	-0.090***	0.116
	(0.089)	(0.429)	(0.015)	(0.085)
Midwest	-2.865***	-0.577	-0.552***	-0.120
	(0.073)	(0.522)	(0.012)	(0.104)
West	-0.233**	-1.506***	-0.118***	-0.124
	(0.093)	(0.506)	(0.016)	(0.103)
1998	0.017	-0.681***	0.243***	0.347***
	(0.152)	(0.232)	(0.026)	(0.047)

1999	-0.001	-0.198	0.170***	0.336***
	(0.144)	(0.224)	(0.024)	(0.045)
2000	-0.229**	-0.372	0.068***	0.198***
	(0.144)	(0.237)	(0.024)	(0.047)
2001	-2.700***	-1.105***	-0.256***	0.162***
	(0.132)	(0.207)	(0.022)	(0.042)
2002	0.831***	0.908***	-0.220***	0.128***
	(0.130)	(0.206)	(0.022)	(0.041)
2003	-3.131***	-2.393***	-0.518***	-0.155***
	(0.134)	(0.205)	(0.023)	(0.041)
2004	-0.351**	-0.498***	-0.053**	-0.070***
	(0.138)	(0.088)	(0.023)	(0.017)
2005	-0.205	-0.280***	-0.038	-0.035**
	(0.142)	(0.090)	(0.024)	(0.017)
Q1	-0.440***	-0.302***	0.122***	0.191***
	(0.070)	(0.044)	(0.011)	(0.008)
Q2	-0.468***	-0.390***	0.093***	0.135***
	(0.035)	(0.022)	(0.006)	(0.004)
Q3	0.553***	0.576***	0.015***	0.330***
	(0.024)	(0.015)	(0.004)	(0.003)
Intercept	28.658***	34.659***	7.026***	6.677***
	(0.297)	(0.466)	(0.049)	(0.093)
ε at Mean (%)	-1.325	-0.865	-1.019	0.183
N	177,758	177,758	153,029	153,029
Adj. R^2	0.199	0.045[a]	0.241	0.162

***: Coefficient is statistically significant at the 0.01 level. **: At the 0.05 level. *: At the 0.10 level.

Heteroskedasticity-corrected standard errors are in parentheses.

a: The R^2 value reported for the FE IV regressions are the overall values, rather than the within or between values.

Table 12: Results of Estimating (11) for HEIScore using OLS and Fixed Effects IV.

Variable	OLS	Fixed Effects IV
Supercenter share	-0.888***	-0.233**
	(0.134)	(0.107)
HHI	0.201**	0.107
	(0.097)	(0.69)
HH Income	0.013***	-0.022***
	(0.002)	(0.004)
Black	-0.492***	0.469***
	(0.023)	(0.179)
Asian	-0.028	0.159
	(0.051)	(0.182)
Other Race	-0.641***	0.271***
	(0.037)	(0.103)
Professional	0.204***	0.137
	(0.056)	(0.111)
Manager	0.109**	0.542***
	(0.056)	(0.110)
Clerical	0.252***	0.043
	(0.061)	(0.112)
Sales	0.224***	0.238*
	(0.062)	(1.87)
Craftsman	-0.206***	-0.429***
	(0.057)	(0.113)
Operative	0.063	0.242**
	(0.061)	(0.123)
Service	-0.148**	-0.142
	(0.064)	(0.135)
Unemployed	0.943***	-0.348***
	(0.059)	(0.117)
Education	0.172***	-0.028
	(0.005)	(0.019)
Hours Worked	-0.181***	-0.266***
	(0.010)	(0.024)
Northeast	-0.035*	-0.415*
	(0.022)	(0.213)
Midwest	-0.308***	-0.035
	(0.018)	(0.259)
West	-0.173***	-0.649**
	(0.023)	(0.255)
1998	-0.006	-1.217***
	(0.037)	(0.119)
1999	0.350***	-0.704***
	(0.035)	(0.115)
2000	0.340***	-0.759***

	(0.035)	(0.121)
2001	-0.032	-0.729***
	(0.032)	(0.106)
2002	0.406***	-0.109
	(0.032)	(0.106)
2003	-0.165***	-0.495***
	(0.033)	(0.106)
2004	0.134***	-0.156***
	(0.034)	(0.045)
2005	0.178***	0.069
	(0.035)	(0.046)
Q1	0.204***	0.620***
	(0.009)	(0.022)
Q2	-0.230***	0.946***
	(0.006)	(0.012)
Q3	-0.153***	0.676***
	(0.006)	(0.008)
Intercept	1.310***	2.423***
	(0.073)	(0.230)
ε at Mean (%)	-2.043	-0.534
N	172,647	172,647
Adj. R^2	0.084	0.059

***: Coefficient is statistically significant at the 0.01 level. **: At the 0.05 level. *: At the 0.10 level.
Heteroskedasticity-corrected standard errors are in parentheses.

Appendix A: Methodology for Calculating Household-Specific Expenditures, as Recommended by USDA

In order to calculate household-specific recommended expenditure shares, we used the average weekly dollar costs, by food category and age-gender group (see table A.1). For illustrative purposes, table A.2 includes the calculation of the expenditure shares for the TFP representative family. Recall that the TFP family consists of a male and female aged 19-50 and two children, aged 9-11 and 6-8 respectively. Hence the total weekly food expenditure for the TFP family is $273.38. Using this value as a denominator, we then calculate the USDA-recommended expenditure shares for each food category.

Recognizing that the TFP family may not be representative of those households in the Homescan database, we use the demographic information provided by the Homescan households to construct recommended shopping baskets based on the CNPP weekly dollar cost estimates. The three categorical Homescan variables that enabled this construction are AgeM, the age of the male head of household, AgeF, the age of the female head of household, and AgeC, the age and presence of children. Every household-specific weekly shopping basket consists of a dollar costs aggregated based upon these three variables (see table A.2). The CNPP age-gender groups do not correspond perfectly with the Homescan categorical variables, particularly because there are more of the former. Therefore it was necessary to use approximate matches or in certain cases average values to sync the two datasets together.

Table A.1: Weekly Dollar Costs for the Liberal Food Plan, by Food Category and Demographic

Age-Gender Group	Total Cost	Whole grain breads	Whole grain cereals	Popcorn & other snacks	Non-whl. grain breads	Potatoes	Dark green veg.	Orange veg.
Child 1	37.87	0.00	1.36	0.00	1.81	0.92	2.62	0.17
Child 2-3	41.15	0.02	0.46	2.44	3.05	0.95	1.12	0.75
Child 4-5	44.00	2.40	0.45	0.88	2.61	0.48	1.02	1.54
Child 6-8	57.81	1.49	0.27	2.22	3.86	0.49	1.20	0.87
Child 9-11	66.83	2.41	0.23	2.00	4.68	1.30	5.24	3.45
Male 12-13	74.02	0.74	0.71	7.55	7.02	1.70	2.26	1.09
Male 14-18	74.92	4.51	1.73	2.36	5.45	1.91	2.16	1.51
Male 19-50	78.40	2.82	6.38	1.69	4.58	1.60	2.95	1.65
Male 51-70	72.46	4.14	0.62	2.63	4.05	1.55	4.44	2.07
Male 71+	73.12	3.48	0.66	0.67	2.69	1.76	9.35	1.22
Female 12-13	64.12	5.77	0.81	1.91	5.72	1.49	2.93	1.11
Female 14-18	65.01	8.09	1.34	0.15	3.40	1.32	3.45	1.59
Female 19-50	70.34	3.38	4.57	0.12	3.56	1.44	5.88	1.16
Female 51-70	64.70	4.03	0.36	3.59	1.06	1.83	2.27	1.33
Female 71+	64.67	4.33	0.55	0.12	2.93	0.72	10.50	1.45
Family (TFP)	273.38	10.09	11.46	6.04	16.67	4.83	15.27	7.14
FC Shares		0.04	0.04	0.02	0.06	0.02	0.06	0.03

Age-Gender Group	Canned & dry beans	Other veg.	Whole fruits	Fruit juices	Whole milk, yog., etc.	Low-fat milk & yog.	Cheese	Milk drinks & desserts
Child 1	1.63	7.38	7.98	1.59	5.18	0.00	0.05	0.16
Child 2-3	2.89	4.94	3.92	1.17	0.22	6.64	0.17	0.08
Child 4-5	2.90	2.32	10.88	1.33	0.12	4.48	0.17	0.20

Child 6-8	4.22	4.10	12.53	1.32	0.11	4.50	0.12	0.23
Child 9-11	2.60	6.11	12.16	1.41	0.15	5.66	0.21	0.19
Male 12-13	4.15	7.30	9.99	2.99	0.24	6.69	0.38	0.22
Male 14-18	3.63	8.65	12.63	1.13	0.36	5.69	0.33	0.27
Male 19-50	10.21	7.30	8.96	1.18	0.36	5.93	0.91	0.75
Male 51-70	3.72	6.76	10.96	0.93	0.32	6.51	0.38	1.74
Male 71+	11.35	7.07	8.98	1.28	0.18	6.01	0.21	0.82
Female 12-13	3.67	4.36	9.82	1.28	0.19	5.71	0.36	0.22
Female 14-18	10.13	5.95	8.61	1.08	0.22	5.71	1.36	0.17
Female 19-50	5.71	6.17	11.43	1.17	0.24	7.88	0.42	0.33
Female 51-70	2.92	5.76	13.38	0.35	0.23	7.81	0.04	0.14
Female 71+	3.46	5.22	10.73	0.36	0.17	8.61	0.02	0.11
Family (TFP)	22.74	23.67	45.08	5.08	0.86	23.97	1.65	1.50
FC Shares	0.08	0.09	0.16	0.02	0.00	0.09	0.01	0.01

Age-Gender Group	**Red meat**	**Poultry**	**Fish**	**Bacon, saus., & lunch meat**	**Nuts & nut butters**	**Eggs**	**Fats & oils**	**Gravies & condmts.**
Child 1	1.34	0.31	2.74	0.04	0.99	0.01	0.34	0.55
Child 2-3	0.49	0.24	8.26	0.35	0.65	0.17	0.89	0.63
Child 4-5	1.32	0.66	7.88	0.12	0.86	0.03	0.32	0.51
Child 6-8	2.22	1.27	13.22	0.34	1.37	0.01	0.45	0.54
Child 9-11	2.79	0.97	10.48	0.93	1.93	0.04	0.44	0.29
Male 12-13	4.62	0.63	9.49	0.22	2.37	0.32	0.94	0.90
Male 14-18	10.28	0.51	2.64	0.47	3.29	0.11	0.69	1.22
Male 19-50	5.91	3.61	4.32	0.83	1.73	0.18	0.94	0.83
Male 51-70	5.26	3.37	7.11	0.31	2.02	0.17	0.77	0.30

Male 71+	3.83	5.75	3.45	0.22	0.86	0.08	0.89	0.46
Female 12-13	5.46	2.64	6.83	0.52	0.76	0.04	0.82	0.45
Female 14-18	3.90	0.63	2.39	0.57	1.93	0.07	0.69	0.72
Female 19-50	3.60	1.50	4.58	0.40	3.62	0.08	0.69	0.69
Female 51-70	4.03	1.67	9.55	0.02	2.83	0.12	0.36	0.27
Female 71+	3.94	1.85	3.71	0.47	3.66	0.05	0.32	0.11
Family (TFP)	14.52	7.36	32.60	2.50	8.65	0.32	2.52	2.36
FC Shares	0.05	0.03	0.12	0.01	0.03	0.00	0.01	0.01

Age-Gender Group	**Coffee & tea**	**Soft drinks**	**Sugars & sweets**	**Soups (nondry)**	**Soups (dry)**	**Froz./ refrg. entrees**
Child 1	0.00	0.16	0.45	0.02	0.06	0.01
Child 2-3	0.00	0.27	0.30	0.01	0.02	0.07
Child 4-5	0.00	0.21	0.09	0.08	0.07	0.08
Child 6-8	0.00	0.34	0.18	0.23	0.03	0.07
Child 9-11	0.00	0.66	0.28	0.10	0.04	0.07
Male 12-13	0.00	0.61	0.40	0.27	0.13	0.08
Male 14-18	0.01	1.19	1.43	0.67	0.07	0.02
Male 19-50	0.03	1.83	0.38	0.39	0.10	0.04
Male 51-70	0.10	1.39	0.36	0.35	0.06	0.07
Male 71+	0.04	0.25	0.18	1.31	0.06	0.00
Female 12-13	0.01	0.66	0.28	0.18	0.10	0.03
Female 14-18	0.07	0.52	0.41	0.46	0.02	0.05
Female 19-50	0.03	0.80	0.27	0.46	0.04	0.14
Female 51-70	0.03	0.13	0.12	0.34	0.10	0.00
Female 71+	0.01	0.06	0.01	1.12	0.07	0.00
Family	0.06	3.63	1.12	1.18	0.21	0.32

(TFP)						
FC Shares	0.00	0.01	0.00	0.00	0.00	0.00

Source: Dollar cost estimates are provided by Dr. Mark Lino of the USDA Center for Nutritional Policy and Promotion. Dr. Lino is not responsible for any of the numbers provided in the table. Any errors are the authors' alone.

Table A.2: The Use of Homescan Demographic Variables for Constructed Household-Specific Recommended Expenditure Shares

Homescan Variable Name	Variable Code	Description	CNPP Age-Gender Group(s)
AgeM	1	Under 25 years	Male 19-50
	2	25-29 years	Male 19-50
	3	30-34 years	Male 19-50
	4	35-39 years	Male 19-50
	5	40-44 years	Male 19-50
	6	45-49 years	Male 19-50
	7	50-54 years	Male 51-70
	8	55-64 years	Male 51-70
	9	65+ years	Male 71+
	0	No male head	None
AgeF	1	Under 25 years	Female 19-50
	2	25-29 years	Female 19-50
	3	30-34 years	Female 19-50
	4	35-39 years	Female 19-50
	5	40-44 years	Female 19-50
	6	45-49 years	Female 19-50
	7	50-54 years	Female 51-70
	8	55-64 years	Female 51-70
	9	65+ years	Female 71+
	0	No female head	None
AC	1	Under 6 only	Average (Child 1, Child 2-3, Child 4-5) (1)
	2	6-12 only	Average (Child 6-8, Child 9-11) (2)
	3	13-17 only	Average (Male 14-18, Female 14-18) (3)
	4	Under 6 and 6-12	(1) + (2)
	5	Under 6 and 13-17	(1) + (3)
	6	6-12 and 13-17	(2) + (3)
	7	Under 6, 6-12, and 13-17	(1) + (2) + (3)
	9	No children under 18	None

Appendix B: Explanatory Variable Definitions and Motivation for Inclusion in Equation (9)

Table B.1: Explanatory Variable Definitions and Motivation for Inclusion in (9), as Drawn from the Literature

Variable	Definition	Background
HH Income	Annual household income brackets, converted to US dollars.[a]	There is a persistent perception among lower-income Americans that healthful foods are more expensive (Eikenberry and Smith, 2004, Mushi-Brunt et al., 2007).
Education	Highest education level of the heads of household: 1) Grade School, 2) Some High School, 3) Graduated High School, 4) Some College, 5) Graduated College, 6) Post College.[b]	Researchers in health and medicine have identified links between educational attainment and obesity for individuals (Frank et al., 2004) and for children at home (Xie et al., 2003).
Hours Worked	Number of hours worked per week by the head of household: 0) None, 1) Less than 30 hours, 2) 30-34 hours, 3) 35+ hours.[c]	There is a link between healthful food and preparation time. In general, meals requiring little to no preparation time consist of packaged and processed foods that are not recommended as healthful by USDA guidelines. Low-income consumers, particularly single parents, may have difficulty meeting the time requirements for preparing menus at home that satisfy theUSDA recommendations for meeting dietary guidelines (Mancino and Newman, 2007).
Race/Ethnicity	Dummy variables identifying households as being black, Asian, white (the reference category), or belonging to another race.	Intended to account for intrinsic differences in food consumption preferences across races or ethnicities as found by researchers, e.g. Block et al. (2004).
Occupation/ Employment Status	Dummy variables identifying the occupational category of the head of household, physical laborer is the reference category.[d]	Occupation has been found to be a determinant of body mass index and obesity (Lahti-Koski et al., 2000; Galobardes et al., 2000).
Geographic Region	Dummy variables for regions of the U.S., as defined by Nielsen. East is the reference category.	Intended to capture differences in preferences across regions, should they exist.
1998-2005	Annual dummy variables for the	Intended to capture changing

	years in the data set, 2006 is the reference category.	purchasing habits in the population over time, owing to nutritional research, technological change, or any other factors.
Q1-Q3	Quarterly dummies representing three-month blocks of time. October through December is the reference.	Intended to capture seasonal shifts in consumers' purchasing decisions.

Source: Nielsen Homescan Survey Data, 1998-2006.

a: The Nielsen Data Dictionary lists the 19 income brackets into which Homescan households are categorized. For ease of interpretation, we converted the brackets to dollar amounts by taking the midpoint of each range. For example, bracket 4 is the range of $5,000 to $7,999. In our data, these households are assigned an income of $6,499.

b: For educational attainment, Nielsen records this value separately for male and female heads of household. For our purposes, we use the highest level attained in the household when both are reported. We have also used the average educational attainment in the case of two heads of household, without observing meaningful changes.

c: For hours worked we use the minimum number of hours worked in the case of two heads of household. We assume that the head working the fewest hours spends the most amount of time on domestic tasks such as food preparation. We also estimate using the average hours worked without observing meaningful changes.

d:In the case of two heads of household, we report the occupation of the head working fewer hours, under the assumption that the household head working fewer hours is more likely to be responsible for the purchasing and preparing of food at home.