Food Retail Market Structure and Produce Purchases in the U.S.

Xiaowei Cai
Cal Poly State University
cai@calpoly.edu

Richard Volpe
Cal Poly State University
rvolpe@calpoly.edu

Christiane Schroeter
Cal Poly State University
cschroet@calpoly.edu

Lisa Mancino
USDA Economic Research Service
lmancino@ers.usda.edu

Selected Paper prepared for presentation at the 2017 Agricultural & Applied Economics Association Annual Meeting, Chicago, Illinois, July 30-August 1

Food Retail Market Structure and Produce Purchases in the U.S.

Abstract

We are creating a novel dataset by merging IRI household-level purchase records with Nielsen TDLinx store-level data on store location to study the food retail environment and produce purchases. Treating zip codes as markets, we find a number of significant and robust relationships: increased market concentration is associated with decreased produce expenditures. In addition, the presence of most nontraditional store formats such as convenience stores and dollar stores is associated with decreased produce purchases. However, the opposite is true for club stores and natural/gourmet supermarkets. The estimated effects of market entry are small, supporting the literature on supermarket intervention studies.

Introduction

The majority of U.S. households' grocery purchasing patterns do not reflect adherence with the Dietary Guidelines for Americans (DGA). In particular, Americans purchase relatively too few fruits and vegetables to meet the DGA recommendations (Kimmons et al., 2009; Volpe and Okrent, 2012). The USDA Center for Nutrition Policy and Promotion (CNPP) showed large discrepancies between average and recommended expenditure shares by food group, based on the USDA Thrifty Food plan (Carlson et al., 2007). The disconnect between recommendations for food expenditure shares and actual consumer purchasing behavior is commonly cited as one of the important determinants of the growing rates of obesity, heart disease, type-2 diabetes, and other related ailments in the U.S. (Segal et al., 2016).

Research on the economic and health impacts of various store formats is motivated as the U.S. food retail industry continues to evolve and churn.[1]There are a number of reasons to expect that the number and variety of grocery stores available to households might influence consumers' food purchasing decisions. Store formats differ, in some cases substantially, according to their prices and product assortments, among other key attributes. For example, the ubiquity of convenience stores, particularly in areas with few to no large food retailers, has long been cited as factors driving dietary quality down in urban areas (Bustillos et al., 2008; Rummo et al., 2015). Other studies have identified the impacts of supercenters on BMI (Courtemanche and Carden, 2011) and diet quality (Volpe et al., 2013).

In the present paper, we study the potential influence of the food-at-home retail environment on fruit and vegetable purchases using 28,613 food markets throughout the U.S.,

[1] The Nielsen TDLinx data are ideal for measuring and illustrating the structural change in food retail, with respect to store formats. We calculated the share of total establishments, by count, for all formats in the TDLinx data across 28 U.S. metropolitan areas. From 2004 to 2014, the average share of club stores fell 8%, the average share of dollar stores increased 30%, and the average share of independent supermarkets increased 55%.

defined at the zip code level. The food markets exhibit a great deal of variation with respect to the concentration and distribution of food retailers as well as consumer demographics. Food retailers are classified according to 10 distinct formats in the data, and the share of total stores for each format ranges in the data from 0 to 100 percent.

We focus on fruit and vegetable expenditures for two reasons: One is that reviews of the epidemiological literature have established their singular importance towards dietary quality and health-protective effects (e.g. Van Duyn and Pivonka, 2000). The other is that by doing so, we obviate the need for the construction of overall dietary quality measures. Such measures are usually calculated based on consumption, rather than purchases, and we are only able to observe purchases. This may introduce a number of assumptions that may lead to measurement error.[2] A commonly used measure of dietary quality is the Healthy Eating Index (HEI), which requires food pattern equivalents for all items purchases in the scanner data, and this information is currently not available from IRI.

The econometric findings suggest a number of important associations between the food retail environment and produce purchases. The share of retailers that are supercenters, dollar stores, superettes, convenience stores, and limited assortment supermarkets are all negatively and significantly associated with the share of grocery expenditures attributed to fruits and vegetables. However, the association is positive and significant for club stores and natural/gourmet supermarkets. The estimated impact of store entry on the share of grocery expenditures allocated to fresh produce is generally very small in magnitude for all formats. We also find market concentration, a factor with ambiguous expectations based on theory and literature, to have a

[2] For this study in particular, given our interest in the role of different store formats, it is straightforward to imagine one potential problem. Certain formats, such a club stores, encourage consumers to purchase in bulk, and with storable foods it may take households a long time to consume certain purchases. Therefore, storable foods are more likely to be recommended for decreased consumption, according to the DGA, than are perishable foods.

small but negative and significant impact on fruit and vegetable expenditures. In addition, we categorize households by the extent to which they meet recommendations for produce expenditures, and our key findings hold up in this setting as well.

Overall, our findings suggest that the makeup of the local food retail environment is important in determining the household share of grocery expenditures spent on fruits and vegetables. For example, for every one-percent increase in the share of stores within zip codes that are convenience stores, the average produce expenditure share falls one percent. However, store entry impacts are small, at least in the short term, supporting the body of evidence that demonstrated limited impacts from direct supermarket interventions.

Background: Food Retailers and Food Choices

Multiple formats are competing for consumers' food dollars in retail landscape of the U.S. Table 1 reports definitions and summary statistics for the store formats that we analyze in this study.

Table 1 here.

Consumers choose among formats for multiple reasons and, accordingly, there are many reasons why purchasing decisions may vary across store formats. Stores vary dramatically across formats in terms of overall size, product assortment, prices, promotional strategy, customer service, and amenities offered. Food retailers, regardless of format, engage in complex and multifaceted competitive strategies that include all of the factors mentioned above and more. Glanz et al. (2012) showed that all of the major components of competitive strategies, including product placement on shelves, influence food choices and ultimately obesity rates.

Recently, differences across store formats have come under scrutiny from a policy perspective. The lack of access to larger-format stores (e.g. supermarkets and supercenters) has

been put forth as a potential determinant of poor diet quality and adverse health outcomes in the U.S. It is important to clarify that our study does not focus on or measure food access, but we are motived by research on this topic, as households only shop at stores and formats in their vicinity. Bustillos et al. (2008) was among the earliest studies to note that, in rural areas, healthful foods such as fruits and vegetables tend to be harder to find and more expensive than in urban areas. These differences are largely explained by the limited availability of large supermarkets in rural areas, and that rural households often shop at smaller convenience stores as a result. In a review of the literature, Story et al. (2008) found this to be generally true in dense urban areas as well, where the retail space for larger retailers is limited. A common theme in the studies reviewed by the authors is the importance of access to supermarkets, highlighting the finding that adults living more than four miles, depending on population density, from the nearest supermarket are significantly more likely to have low diet quality, controlling for demographic factors.

Not all studies on market structure and diet or health outcomes have uncovered significant impacts. Sturm and Datar (2005) did not find any significant relationship between the density of local food markets and restaurants with children's weight gain. Aggarwal et al. (2014) studied supermarket access and fruit and vegetable purchases specifically and found that proximity to supermarkets was not significant, controlling for covariates such as income, which account for vehicle access and mobility. Nevertheless, the authors found that supermarket choice, among store formats, was a significant driver of fruit and vegetable purchases for households.

Given the established importance of supermarkets, a line of research has been conducted on so-called "intervention" studies. Throughout this paper, the term *intervention* refers to the deliberate introduction of a new, full-sized supermarket, into a geographic market where

previously none operated. Interventions are commonly discussed as potential policy tools to address issues such as food insecurity, poor dietary quality, and adverse health outcomes in the U.S. Generally, ex-post analyses of intervention studies have found small or even no effects on food choices or diet quality (Kristal et al., 1997; Wrigley et al., 2003; Cummins et al., 2005; Escaron et al., 2015; Dubowitz et al., 2015). However, we are motivated to examine the potential impact of store entry, across formats, on fruit and vegetable purchases, given the advantages conferred via our rich dataset. In particular, we are able to observe households' comprehensive grocery purchases for extended times, while many previous studies relied on limited survey data.

A handful of studies have followed approaches similar to our own and examined relationships between the retail food environment and food choices or health outcomes. Much of the work in this vein is encapsulated by Cummins and Macintyre (2006), who noted that the estimated impacts of the food environment, measured directly, are likely to be small in magnitude. Similarly, Rahkovsky and Snyder (2015) found a weak, negative correlation between dietary quality and living in areas classified by USDA as low-income and low-access, across households.

We know that a large number of factors, beyond the physical food environment, affect food choices and health outcomes. For example, it is difficult to disentangle local market effects from the laws, regulations, and culture of the nation as a whole. Jilcott et al. (2011) found small but significant and negative impacts of the presence of both supermarkets and supercenters on local obesity rates. Chen et al. (2016) studied counties as markets and found that counties classified as USDA food deserts have significantly higher obesity rates than those not, both suggesting the importance of access to food retailers.

Finally, we measure and consider market concentration as a potential aspect of food retail structure that may influence fruit and vegetable expenditures. To our knowledge, no study has examined directly the relationship between fruit and vegetable choices and market structure. Theory and evidence leads to ambiguous expectations. On one hand, increased concentration may lead to fewer stores in a market. Applied research has long established a positive correlation between food prices and market concentration in food retail (e.g. Yu and Connor, 2002). One common interpretation of this finding is that fewer firms decrease the level of competition, while market power and prices rise. Given that retailers compete in many dimensions other than price (Richards and Hamilton, 2006), it stands to reason that competition in product quality or assortment may also lessen as concentration increases. As such, we expect to see a negative relationship between market concentration and fruit and vegetable purchases.

Alternatively, there are reasons to anticipate a positive relationship between concentration and fruit and vegetable expenditures. Ellickson (2013) has argued that the food retail industry is a natural oligopoly due largely to the extensive fixed costs associated with entry, though that even in markets with three to six unique firms, competition remains strong. Bonanno and Lopez (2009) demonstrate that this competition can take the form of firms differentiating themselves through quality, leading to a small number of firms offering high-quality foods and high level of services, at relatively high prices. Moreover, the effect of concentration may depend on the format of the largest firms in the market. Given that prices and products can vary across formats, the overall effect of concentration may not be captured in an econometric setting due to the heterogeneity of store formats and food retailers in general.

Data and Methodology

To measure fruit and vegetable expenditures, we use the IRI Household Panel, 2008-2012, henceforth 'the Panel.' Over this five-year period, the Panel consists of point-of-sale grocery purchase records for over 100,000 households, many of which participate for multiple years. Panel participants receive a small stipend in exchange for recording their itemized grocery shopping trips, using a handheld scanner provided IRI to record each food item purchased, the stores at which they were purchased, and the prices paid. Household purchase records can be paired with detailed demographics and information on store formats.

A major limitation of the Panel data is the lack of granularity for random weight, or non-UPC, purchase records. While the majority of all grocery items are easily scanned and recorded via UPC codes, a great deal of fresh produce products are not. Not accounting for this would likely underestimate their purchases. To account for this, Panel participants are asked to record total expenditures on fresh fruits and vegetables each shopping trip. Therefore while we do not observe the specific produce items purchased, we are able to combine the UPC-coded data with the total categorical expenditures reported by households to obtain a measure of total household fruit and vegetable expenditures for a given time period. We can then calculate the share of total grocery expenditures spent on fruits and vegetables. This serves as the foundation of our empirical approach and our key variable of interest.

To quantify the local retail environment, we use the Nielsen TDLinx data. The TDLinx data are store level and each record consists of the name, ownership structure, location, format, size, and entry date for a given store. Store-level grocery sales, or revenues, are recorded categorically in TDLinx based on annual revenues from grocery sales. Nielsen breaks annual grocery revenues into 19 categories, defined by upper and lower bounds. To assign numerical

dollar values to the stores in our dataset for measuring market concentration, we take the midpoint of each category and record that value for annual food revenues.[3]

Table 2 provides definitions and summary statistics for the variables used in our analysis. The average household's annual expenditures on fruits and vegetables accounts for 9.2% of all grocery expenditures. The CNPP recommends that households spend approximately 49% of their food expenditures on fruits and vegetables, indicating substantial room for improvement.[4] Compared to the U.S. population, the IRI Panel is disproportionately white, older and higher-income than the average U.S. household. These differences should be kept in mind when interpreting the results. Research has consistently shown (e.g. Larson et al., 2009) that households lacking access to supermarkets and healthy food are more likely to be low-income and minorities. Hence we argue that this indicates any estimated impacts of large-format access on produce purchasing are likely to be lower bounds of the effects on the U.S. population.

Table 2 here.

We measure household fruit and vegetable (including fresh, frozen, and canned) expenditures as shares, rather than using dollar amounts, to better measure adherence to the DGA. As expenditure shares on healthful foods, such as produce, increase, overall dietary quality is almost surely increasing. Increases in total produce expenditures, in dollars, may be marked by analogous or even greater increases in spending in other categories, resulting in no improvement in overall adherence. Alternatively, total expenditures can change over time as

[3] The highest category consists of those stores generating more than $100,000,000 per year in food revenues. We report the value $100,000,001 for these stores, effectively truncating sales volume in the data. Stores of this size constitute only 0.37 percent of the total stores in the dataset, and eliminating this category from the analysis leaves our results qualitatively unchanged. Ninety percent of all stores in our data range from $1 million to $16 million, annually.

[4] The recommended expenditure share for fruits and vegetables is calculated using the USDA Thrifty Food Plan, which provides recommended expenditures by food group. The numbers used for our calculation are available from Carlson et al. (2007).

household members grow into adults or pass away. Finally, households over time make errors or underreport their food purchases in these scanner-based surveys (Einav et al., 2008). Expenditures may fall and rise again due to inconsistencies in reporting, but expenditure shares are only subject to such impactful measurement error if produce purchases are underreported at a different rate than other foods. Given that households need to only report total expenditures on random weight foods, this seems unlikely as a systemic concern in the data.

In our statistical framework, one concern with using expenditure shares is the possibility that expenditure shares might increase because of increases in the relative price of produce. That is, households are not purchasing more produce but are paying more for it. While this may be possible for individual households due to selective shopping habits across store formats, it is only a systematic concern if produce prices increased more than retail food prices. We examined the Consumer Price Index (CPI) data for 2008-2012 to investigate this. During our time period, the food-at-home (grocery) CPI increased 8.2% while the fruit and vegetable CPI only increased 1.4%. This strengthens our identification strategy, as produce grew relatively cheaper for the average household during this time, and therefore increases in produce expenditure shares are more likely to reflect increases in purchase volumes.

The food retail environment, treating zip codes as markets, is measured using a series of variables.[5] First, we calculate market concentration using the Herfindahl-Hirschman Index (HHI), which is the sum of squared market shares of firms within markets. This variable is

[5] TDLinx data have been used in number of studies to measure market structure. Given the complexities involved in measuring geographic markets, there are no conventions for doing so. For example, Orhun (2013) measured markets using metropolitan statistical areas, Ver Ploeg et al. (2011) used Census blocks, and Lamichhane et al. (2014) used Census tracts.

We selected zip codes as they are the most granular geographic measure for which we have complete information on our intended model controls. Moreover, the geographic size of zip codes is inversely correlated with population, therefore serving as a proxy for population density. However, consumers very likely cross zip code boundaries in order to shop for groceries, meaning that these are imperfect measures of geographic markets.

henceforth the FoodHHI. It is vital to note that TDLinx only provides food revenues, as opposed to all merchandise revenues, for selected store formats. Hence, to avoid measurement error, the FoodHHI is calculated using only conventional supermarkets, limited assortment supermarkets, warehouse supermarkets, natural/gourmet supermarkets, supercenters, and superettes.[6] To account for the presence and number of additional formats, and to recognize the heterogeneity of the food retail environment, we also calculate the share of total stores, per zip code, accounted for by each format (henceforth FormatShares). Throughout all of our empirical work, conventional supermarkets serve as the reference case. Finally, to capture the effects of market entry, we create dummy variables equal to 1 for zip codes and years in which conventional supermarkets, supercenters, club stores, and natural/gourmet supermarkets opened new stores (henceforth FormatEntries).

Our baseline model to be estimated for household *i* in year *t* is

$$\begin{aligned} \text{FruitVegShare}_{it} = {} & \beta_1 + \beta_2 \text{FoodHHI}_{t-1} + \beta_3 \text{HHIncome}_{it} + \beta_4 \text{HHAge}_{it} + \beta_5 \text{Married}_{it} + \\ & \beta_R \text{Race}_i + \beta_{Ed} \text{Education}_{it} + \beta_{Emp} \text{Employment}_{it} + \beta_{Reg} \text{Region}_i + \beta_{FS} \text{FormatShares}_{it-1} + \\ & \beta_{FE} \text{FormatEntries}_{it-1} + \varepsilon_{it}. \end{aligned} \quad (1)$$

where FruitVegShare, the dependent variable, is the share of total annual grocery expenditures spent on fruits and vegetables. In our baseline estimation, we treat this variable as a simple percentage and estimate (1) using OLS. We also compare expenditures to recommendations. The CNPP provides expenditure recommendations for households based on age and gender composition and we categorize households depending on how their expenditure shares compare to recommendations. The categories are: households meeting 0 to 25% of recommendations, households meeting 25 to 50% of recommendations, and households exceeding 50% of recommendations. This enables us to estimate (1) in a limited dependent variable setting.

[6] The excluded formats, for which grocery revenues are not reported in TDLinx, include club stores, dollar stores, and convenience stores.

Many of the components of (1) are reported as vectors and require explanation. Race is a vector of dummies signifying the race each household reports for itself, with "other race" serving as the reference category. Education consists of vectors of educational attainment dummies for both the male and female household heads, where applicable. The reference case here is persons with high school graduation as their highest educational level. Region is a vector of IRI-defined regions of the country, with the West serving as the reference category.

To address endogeneity concerns, we lagged each market structure variable by one year. While we also experimented with contemporaneous impacts and longer lag lengths, this specification yielded the best goodness of fit (as measured by the adjusted R^2) and the most significant regression coefficients. It is also worth noting that in the contemporaneous case, given the timing framework by which the TDLinx data are recorded and the annual frequency of our data, some changes in purchasing patterns may be occurring after changes in market structure.

Results and Discussion

We begin by estimating equation (1) using OLS in our baseline setting. To assess the robustness of the market structure impacts, we run three specifications of (1) on the share of total annual grocery expenditures spent on fruits and vegetables (FruitVegShare) as an expenditure percentage. The first specification features all market structure variables, the second one features only the store format shares, and the last features only the market entry variables. The results are reported in table 3. We note from the outset that all of the estimated impacts of our market structure variables are robust within our OLS specifications.

Table 3 here.

The results demonstrate a number of potentially important market structure impacts on fruit and vegetable purchases. HHIFood is negatively associated with produce purchases. The effect, which is consistent across all three specifications, is small but statistically significant. A marginal increase in market concentration is associated with a 0.2 percentage point decrease in FruitVegShare. Furthermore, significant changes to market concentration, such as mergers, acquisitions, or store closings, have the potential to impact local purchase patterns in an economically-significant manner. Therefore, our results in this respect suggest that market-level product assortment and quality lessen with retail market concentration.

The measures of store formats presence share significant associations with FruitVegShare, and in most cases the signs conform to expectations. Limited assortment share is negatively associated with FruitVegShare, and this may be due to the relatively small number of perishable items these stores carry. The association between the share of expenditures on produce and the share of supercenters in a market is small, but negative and significant, corroborating previous work conducted specifically on this format. A detailed review of the literature on larger store formats such as supercenters indicate that the nutritional impact of these stores on consumers can be positive or negative. Our findings suggest that the net effect of supercenters specifically on dietary quality remains likely negative. Club stores are associated with significant increases in FruitVegShare, which is notable, since this format is disproportionately patronized by higher-income households, though (1) controls for income. Natural/gourmet stores, which are differentiated in markets for their emphasis on perishables and locally grown or sustainably-sourced produce, meat, and seafood, demonstrate a positive and significant association with produce purchases.

Our work provides further evidence to support the importance of larger-format stores in shaping food choices and, in turn, diet quality. Convenience stores and superettes, the formats often discussed in the nutrition and epidemiology literature for their effects on highly rural or urban households, are both found to be negatively and significantly associated with FruitVegShare. Dollar stores are also negatively associated with produce purchases, and this effect is comparable in magnitude to that of limited assortment stores. Dollar stores carry very few perishable foods and therefore do not sell large quantities of fruits and vegetables.

The estimated impacts of entry, measured discretely as dummies, are mixed and very small in magnitude. Our findings in this regard are largely corroborative of the work to-date on intervention effects, in that discrete changes to the local food retail landscape appear unlikely to have significant or meaningful impacts on food choices, regardless of format, in the short term. The entry of a new club store is associated with a nearly 1-percentage point decrease in FruitVegShare, which is somewhat counterintuitive given our findings on format shares. It is worth emphasizing that the 28,613 markets we analyze range from urban and suburban zip codes with dozens of retailers, to sparse, rural settings with only one or two retailers per zip code. The effects of store entry may well vary in direction and magnitude depending on the size and characteristics of markets.

To capture the effects of store entry on household produce purchases in diverse markets, we divided the households into four groups by quartile based on the lagged one year number of retailers at the zip-code level. The first quartile contains markets with 1 to 7 retail stores, the second quartile includes markets with 8 to 14 stores, the third quartile has markets with 15 to 22 stores, and the fourth quartile encompasses all the markers that have 23 or more stores. The results are reported in table 4.

The quartile analysis provides potential policy implications with respect to store presence or entry and exit, depending on local market size. The estimated impacts for the respective format shares are substantially different when comparing the smallest markets to the largest. Limited assortment supermarket share is associated with a significant decrease in FruitVegShare for all markets with more than eight stores. Natural/Gourmet supermarket share, alternatively, is associated with significant positive impacts in all markets with eight or more stores. We found that the club store share is negatively associated with household produce purchases in markets with eight to 22 stores, i.e., the midrange zip codes. Dollar store share shows negative and significant impacts for all but the third quartile of markets, while supercenter share is negative and significant for the first and third quartiles. Convenience store share, as expected, is negative and significant across all quartiles. The largest magnitude of the coefficient is for the smallest markets, which are most likely to be dense urban or rural markets.

Table 4 here.

With respect to store entry, we continue to find limited impacts in the quartile regression setting. Club store entry shows small and significant negative impacts in the two smallest quartiles of markets, while natural/gourmet supermarket entry shows positive and significant impacts in the same two quartiles. The magnitude of the impact is much larger for the second quartile, those markets with between eight and 14 stores. There are no significant estimated impacts for larger markets, for any of the considered formats.

Given that we also measure our dependent variable categorically, we estimate the odds of household being in different categories in a multinomial logit setting in the following equation:

$$(2)\ \text{Log}\left(\text{Prob.}(\text{FruitVegCategory}_{itj})/\text{Prob.}(\text{FruitVegCategory}_{it0})\right) = \gamma_{1j} + \gamma_{2j}\text{FoodHHI}_{t-1} + \gamma_{3j}\text{HHIncome}_{it} + \gamma_{4j}\text{HHAge}_{it} + \gamma_{5j}\text{Married}_{it} + \gamma_{Rj}\mathbf{Race}_{i} + \gamma_{Edj}\mathbf{Education}_{it} + \gamma_{Empj}\mathbf{Employment}_{it} + \gamma_{Regj}\mathbf{Region}_{i} + \gamma_{FSj}\mathbf{FormatShares}_{it-1} + \gamma_{FEj}\mathbf{FormatEntries}_{it-1}.$$

In doing so, we estimate directly the impacts of the food retail market structure on adherence to the DGA with respect to fruit and vegetable intake. In our estimation, we use Group 0 as our reference category of households, those meeting between 0 and 25% of recommendations with their produce expenditures. Thus (2) is estimated with comparisons to group 1 (those households meeting 25 to 50% of recommendations) and group 2 (those households exceeding 50% of recommendations). The independent variables are the same as those described in equation (1). The results are reported in table 5.

Table 5 here.

Our logit results readily confirm the OLS results, with respect to market structure impacts. HHIFood is negatively and significantly associated with the probability of being in groups 1 or 2, suggesting that as concentration increases, local households are less likely to meet recommendations in their produce expenditures. Supercenter store share is only significant in the comparison with group 2, though the effect is still negative and substantial. For convenience stores, the estimated impact is also only significant when groups 0 and 2 are compared. In addition, the natural/gourmet store share can increase the probability of local households meeting the produce expenditure recommendations. As the dollar store share increases in the local market, households are less likely to meet the recommendations. In term of store entries, the entry of a new club store is negatively associated with meeting the government recommendations. However, the entry of supercenters will increase the odds of households being in group 1 relative to group 0. Similarly, the entry of a natural/gourmet store will increase the probability of local households meeting more than 50% of expenditure recommendations by more than 25%, as we compare group 2 with group 0.

An additional note on our estimated findings with respect to supercenters is called for, given the attention these stores have received in research and in the popular press. Most of our estimates based on changes in supercenter store share, using continuous estimates of their presence, find negative associations with fruit and vegetable purchases. However we do find limited evidence that the discrete entry of supercenters may be associated with increased fruit and vegetable purchases and adherence to the DGA. More work is needed to flesh out these impacts, but our results are consistent with a story in which the introduction of supercenters in markets in which there were previously few or no large-format grocers may yield positive dietary impacts. However increased supercenter presence in markets that already feature larger stores may lead to decreased diet quality.

Conclusions

To investigate factors explaining consumers' choice of fruits and vegetables during 2008-2012, we use the IRI Household Panel and TDLinx data. Using the IRI Household Panel, we control for a large number of potential confounders, including physical characteristics, lifestyle choices, and geographical fixed effects. The rate and nature of the change in the food retail industry differs spatially across the U.S. Research has shown that consumers make different grocery purchases by store format (Bustillos et al., 2009; Volpe et al., 2013). The Nielsen TDLinx data provide store counts based on different store formats, the entrance and exit information of retailers over time, and other information such as total sales by store format. We merged the IRI panel data with TDLinx at the zip code level to test empirically whether structural changes in the local retail environment might play a critical role in determining produce purchases.

Using both the percentage and categorical measurements of produce purchase behavior, our results show that increased fruit and vegetable purchases and better meeting government produce expenditure recommendations are associated with higher local store counts of club stores and natural/gourmet stores. Conversely, average produce purchases and the likelihood of meeting the expenditure recommendations decrease with the count of supercenters, convenience stores, and dollar stores. The local entry of a new natural/gourmet store has a positive impact on produce purchase. However, the entry of a new club store may decrease consumer fruit and vegetable purchases, particularly in smaller or more concentrated markets. Our findings with respect to the heterogeneous store formats depend importantly on market size. Very few of our estimated impacts related to format-level store share are comparable between the first and fourth quartiles of markets, by total store count.

Our research supports a growing literature on the linkages between the retail food environment, food choices, and health outcomes. The estimated impacts of changes in the food retail market structure, in terms of both concentration and heterogeneity, are statistically significant, in particular for zip codes with fewer stores, which tend to be dense urban markets or rural areas. A fruitful avenue for future research might apply longitudinal data to examine long-term health impacts of market structure on consumers and households. Food prices also surely play a vital role in shoppers' choices, even within formats, and more work is needed to understand the impacts of prices in this context.

References

Aggarwal, A., Cook, A.J., Jiao, J., Seguin, R.A., Vernez Moudon, A., Hurvitz, P.M. and Drewnowski, A., 2014. "Access to supermarkets and fruit and vegetable consumption". *American journal of public health*, *104*(5), pp.917-923.

Bonanno, A. and Lopez, R.A., 2009. "Competition effects of supermarket services". *American Journal of Agricultural Economics*, *91*(3), pp.555-568.

Bustillos, B., Sharkey, J. R., Anding, J., & McIntosh, A. 2008. "Availability of more healthful food alternatives in traditional, convenience, and nontraditional types of food stores in two rural Texas counties." *Journal of the American Dietetic Association* 109(2008), pp.883-889.

Carlson, A., M. Lino, and T. Fungwe. 2007. "The Low-Cost, Moderate-Cost,and Liberal Food Plans". *U.S. Department of Agriculture, Center for Nutrition Policy and Promotion*.

Chen, D., Jaenicke, T., and Volpe, R.2017. "Relationships between home and neighborhood food environments and obesity: a national multilevel analysis using scanner data". *American Journal of Public Health* forthcoming.

Courtemanche, C. and A.Carden. 2011. "Supersizing supercenters? The impact of Walmart Supercenters on body mass index and obesity." *Journal of Urban Economics* 69(2), pp.165-181.

Cummins, S. and Macintyre, S., 2006. "Food environments and obesity—neighbourhood or nation?". *International journal of epidemiology*, *35*(1), pp.100-104.

Cummins, S., Pettigrew, M., Higgins, C., Findlay, A., and Sparks, L. 2005. "Large scale food retailing as an intervention for diet and health: quasi-experimental evaluation of a natural experiment". *Journal of Epidemiology and Community Health* 59, pp.1035-104.

Dubowitz, T., Ghosh-Dastidar, M., Cohen, D.A., Beckman, R., Steiner, E.D., Hunter, G.P., Flórez, K.R., Huang, C., Vaughan, C.A., Sloan, J.C. and Zenk, S.N. 2015. "Diet and perceptions change with supermarket introduction in a food desert, but not because of supermarket use." *Health Affairs*, *34*(11), pp.1858-1868.

Ellickson, P.B., 2013. "Supermarkets as a natural oligopoly". *Economic Inquiry*, *51*(2), pp.1142-1154.

Escaron, A.L., Martinez-Donate, A.P., Riggall, A.J., Meinen, A., Hall, B., Nieto, F.J. and Nitzke, S. 2015. "Developing and implementing 'Waupaca Eating Smart' a restaurant and supermarket intervention to promote healthy eating through changes in the food environment. *Health promotion practice*, pp.1524839915612742.

Food Marketing Institute. 2015. "Supermarket facts." Available online at: http://www.fmi.org/research-resources/supermarket-facts

Glanz, K., Bader, M.D. and Iyer, S. 2012. “Retail grocery store marketing strategies and obesity: an integrative review”. *American journal of preventive medicine*, *42*(5), pp.503-512.

Jilcott, S.B., Keyserling, T., Crawford, T., McGuirt, J.T. and Ammerman, A.S. 2011. “Examining associations among obesity and per capita farmers' markets, grocery stores/supermarkets, and supercenters in US counties”. *Journal of the American Dietetic Association*, *111*(4), pp.567-572.

Kimmons, J., Gillespie, C., Seymour, J., Serdula, M., and Blanck, H. M. 2009. “Fruit and vegetable intake among adolescents and adults in the United States: percentage meeting individualized recommendations.” *The Medscape Journal of Medicine* 11, pp.26.

Kristal, A.R., Goldenhar, L., Muldoon, J. and Morton, R.F. 1997. “Evaluation of a supermarket intervention to increase consumption of fruits and vegetables”. *American Journal of Health Promotion*, *11*(6), pp.422-425.

Lamichhane, A.P., Warren, J.L., Peterson, M., Rummo, P., and Gordon-Larsen, P. 2014. “Spatial-temporal modeling of neighborhood sociodemographic characteristics and food stores.” *American journal of epidemiology*, p.kwu 250.

Larson, N.I., Story, M.T. and Nelson, M.C. 2009. “Neighborhood environments: disparities in access to healthy foods in the US”. *American journal of preventive medicine*, *36*(1), pp.74-81.

Orhun, A.Y. 2013. “Spatial differentiation in the supermarket industry: The role of common information.” *Quantitative Marketing and Economics*, *11*(1), pp.3-37.

Rahkovsky, I., and Snyder S. 2015. “Food Choices and Store Proximity.” *USDA Economic Research Service Economic Research Report* 195. September 2015.

Richards, T.J. and Hamilton, S.F. 2006. “Rivalry in price and variety among supermarket retailers”. *American Journal of Agricultural Economics*, *88*(3), pp.710-726.

Rummo, P.E., Meyer, K.A., Boone-Heinonen, J., Jacobs Jr, D.R., Kiefe, C.I., Lewis, C.E., Steffen, L.M. and Gordon-Larsen, P. 2015. “Neighborhood availability of convenience stores and diet quality: findings from 20 years of follow-up in the Coronary Artery Risk Development in Young Adults study”. *American Journal of Public Health*, *105*(5), pp.e65-e73.

Segal, L.M., Martin, A., and Rayburn, J. 2016. “The state of obesity: obesity data trends and policy analysis.” Retrieved from the WWW on December 14, 2016: http://stateofobesity.org/files/stateofobesity2016.pdf

Story, M., Kaphingst, K.M., Robinson-O'Brien, R. and Glanz, K. 2008.” Creating healthy food and eating environments: policy and environmental approaches”. *Annual Review of Public Health*, *29*, pp.253-272.

Sturm, R. and A. Datar. 2005. “Metropolitan area food prices and children’s weight gain.” *ERS-USDA Contractor and Cooperator Report No. 14*, December 2005.

Van Duyn, M. A. S., & Pivonka, E. 2000. “Overview of the health benefits of fruit and vegetable consumption for the dietetics professional: selected literature.” *Journal of the American Dietetic Association* 100, pp.1511-1521.

Ver Ploeg, M., Nulph, D. and Williams, R. 2011. “Mapping food deserts in the US.” *Amber Waves*, *9*(4), pp.46.

Volpe, R., & Okrent, A. 2012. “Assessing the healthfulness of consumers’ grocery purchases.” *USDA Economic Research Service Economic Information Bulletin* 102. November.

Volpe, R., Okrent, A. and Leibtag, E. 2013. “The effect of supercenter-format stores on the healthfulness of consumers’ grocery purchases”. *American Journal of Agricultural Economics*, 95(3), pp.568-589.

Wrigley, N., Warm, D. and Margetts, B. 2003. “Deprivation, diet, and food-retail access: Findings from the Leeds ‘food deserts' study”. *Environment and Planning A*, *35*(1), pp.151-188.

Yu, C.C. and Connor, J.M. 2002. “The price-concentration relationship in grocery retailing: Retesting Newmark”. *Agribusiness*, *18*(4), pp.413-426.

Table 1. Grocery Store Formats in the U.S.

	Description	Annual Revenues ($1,000)	Chain Size	Count (2012)
Conventional Supermarkets	Stores offering full lines of groceries with at least $2M in annual sales, and at least 85% of revenues in food (e.g. Kroger, Albertsons)[a,b]	14,322	211	26,738
Limited Assortment Supermarkets	Low-priced stores with fewer than 2,000 perishable items (e.g. Trader Joe's, Aldi)	5,304	290	2,988
Warehouse Supermarkets	Low-priced stores offering limited customer services but not requiring paid memberships (e.g. Food 4 Less, Smart & Final)	7,447	169	490
Supercenters	Hybrids of traditional supermarkets and mass merchandisers (general department stores) (e.g. Walmart Supercenters, Super Targets)	46,611	485	3,777
Club Stores	Paid membership, large retailers specializing in bulk-sized products. (e.g. CostCo, BJs)	76,917	451	1,236
Convenience Stores	Smaller retail stores carrying limited selections of basic items and necessities and are typically open long hours (e.g. 7-Eleven, AM/PM)	2,361	89	148,257
Military Commissaries	Supermarkets restricted to use by active and retired military personnel	18,876	445	7,906
Natural/Gourmet Supermarkets	Supermarkets emphasizing perishables, local and organic options, and ethnic foods (e.g. Whole Foods, The Fresh Market)	10,016	74	2,629

Superettes	Small supermarkets, often independent, typically comparable in size to convenience stores, featuring self-service features[c]	1,410	2	13,115
Dollar Stores	Small department stores using fixed price points, typically $1 per unit (e.g. Family Dollar, Dollar Tree)	1,361	484	22,663

Source: All descriptive statistics are calculated using the Nielsen TDLinx data.
a: The definitions for most store formats are edited from the Food Marketing Institute (2015). Nielsen also provides definitions for store formats, however these are proprietary.
b: The chains listed as examples for each format are not necessarily included in the TDLinx data.
c: The definition for convenience stores and superettes are drawn from the Mirriam Webster Dictionary.

Table 2. Summary Statistics for the Variables in the Data.

Variable	Description	Mean	St. Dev
FruitVegShare	Annual share of grocery expenditures on fruits and vegetables	0.092	0.058
Midwest	Dummy = 1 for households in the Midwest	0.173	0.378
South	Dummy = 1 for households in the South	0.364	0.481
Northeast	Dummy = 1 for households in the Northeast	0.270	0.444
HHIncome	Household income in $10000s	6.045	3.597
Fulltime_F	Dummy = 1 for households in which the female household head works at least 35 hours/week	0.401	0.490
Fulltime_M	Dummy = 1 for households in which the male household head works at least 35 hours/week	0.261	0.439
College_F	Dummy = 1 for households in which the female household head attended college	0.284	0.451
College_M	Dummy = 1 for households in which the male household head attended college	0.222	0.416
Gradschool_F	Dummy = 1 for households in which the female household head attended graduate school	0.112	0.315
Gradschool_M	Dummy = 1 for households in which the male household head attended graduate school	0.102	0.303
White	Dummy = 1 for households identifying as white	0.844	0.363
Asian	Dummy = 1 for households identifying as Asian	0.030	0.171
Married	Dummy = 1 for households in which the household heads are married	0.642	0.479
HHSize	Number of people living in the household	2.314	1.239
HHAge	Average age of adults in the household	51.577	13.604
HHIFood	The Herfindahl Index, calculated at the zip code level, for food retail firms	0.520	0.273
SupermarketShare	The zip code-level share of food retailers that are conventional supermarkets	0.146	0.128
LimAssortmentShare	The zip code-level share of food retailers that are limited assortment supermarkets	0.015	0.036
WarehouseShare	The zip code-level share of food retailers that are warehouse supermarkets	0.003	0.015
SupercenterShare	The zip code-level share of food retailers that are supercenters	0.019	0.046
ClubStoresShare	The zip code-level share of food retailers that are club stores	0.007	0.024
ConvenienceShare	The zip code-level share of food retailers that are convenience stores	0.722	0.184
NaturalGourmetShare	The zip code-level share of food retailers that are natural/gourmet supermarkets	0.014	0.042

SuperetteShare	The zip code-level share of food retailers that are superettes	0.052	0.119
DollarShare	The zip code-level share of food retailers that are dollar stores	0.015	0.050
SupermarketEntry	Dummy = 1 for zip codes and years in which a supermarket entered	0.760	0.427
SupercenterEntry	Dummy = 1 for zip codes and years in which a supercenter entered	0.744	0.436
ClubEntry	Dummy = 1 for zip codes and years in which a supermarket entered	0.741	0.438
NaturalGourmetEntry	Dummy = 1 for zip codes and years in which a supermarket entered	0.746	0.435

Source: Authors' calculations using Nielsen TDLinx data, 2004-2012.

Table 3. OLS Results for Estimating (1), the Determinants of Household Fruit and Vegetable Expenditure Share.

Variable	Full model	Store Share	Store Entry
LagHHIFood	-0.002**	-0.002***	-0.003***
	(0.001)	(0.001)	(0.001)
LagSupermarketShare(omitted)			
LagLimAssortmentShare	-0.018***	-0.018***	
	(0.005)	(0.005)	
LagWarehouseShare	-0.002	-0.002	
	(0.013)	(0.017)	
LagSupercenterShare	-0.008**	-0.008**	
	(0.004)	(0.004)	
LagClubStoresShare	0.035***	0.035***	
	(0.008)	(0.008)	
LagConvenienceShare	-0.010***	-0.010***	
	(0.002)	(0.001)	
LagNaturalGourmetShare	0.020***	0.022***	
	(0.005)	(0.005)	
LagSuperetteShare	-0.005***	-0.005**	
	(0.002)	(0.002)	
LagDollarShare	-0.016***	-0.015***	
	(0.004)	(0.004)	
LagSupermarketEntry	-0.001		-0.001
	(0.001)		(0.001)
LagSupercenterEntry	0.004		0.003
	(0.003)		(0.003)
LagClubEntry	-0.008**		-0.009***
	(0.003)		(0.003)
LagNaturalGourmetEntry	0.004**		0.006***
	(0.002)		(0.001)
Midwest	0.002**	0.001*	0.001*
	(0.001)	(0.001)	(0.001)
South	0.003***	0.002***	0.002***
	(0.001)	(0.001)	(0.001)
Northeast	0.001*	-0.001	-0.001
	(0.000)	(0.004)	(0.001)
HHIncome	0.002***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)
Fulltime_F	0.002***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)
Fulltime_M	-0.003***	-0.002***	-0.003***
	(0.001)	(0.000)	(0.000)
College_F	0.010***	0.010***	0.010***
	(0.000)	(0.000)	(0.000)
College_M	0.004***	0.005***	0.005***

	(0.000)	(0.000)	0.000)
Gradschool_F	0.018***	0.018***	0.018***
	(0.001)	(0.001)	(0.001)
Gradschool_M	0.009***	0.010***	0.010***
	(0.001)	(0.001)	(0.001)
White	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)
Asian	0.007****	0.007***	0.007***
	(0.001)	(0.001)	(0.001)
Married	-0.002	-0.001	-0.001
	(0.005)	(0.005)	(0.005)
HHSize	-0.004***	-0.004***	-0.004***
	(0.000)	(0.000)	(0.000)
HHAge	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)
Intercept	0.041***	0.034***	0.034***
	(0.002)	(0.001)	(0.001)

Standard errors in parentheses.
***: Coefficient is significant at the 0.01 level. **: At the 0.05 level. *: At the 0.10 level.

Table 4. OLS Market Structure Results for Estimating (1), Based on Four Quartiles of Retailer Counts, by Zip Code.

Variable	Q1: 1-7 Stores	Q2: 8-14 Stores	Q3: 15-22 Stores	Q4: 23 or More Stores
LagHHIFood	-0.002	-0.002	-0.003*	0.0007
	(0.002)	(0.001)	(0.002)	(0.002)
LagSupermarketShare (omitted)				
LagLimAssortmentShare	-0.018	-0.022***	-0.022*	-0.028*
	(0.008)	(0.010)	(0.012)	(0.015)
LagWarehouseShare	0.014	-0.021	-0.015	0.004
	(0.019)	(0.026)	(0.028)	(0.04)
LagSupercenterShare	-0.010*	0.004	-0.031***	-0.004
	(0.006)	(0.008)	(0.011)	(0.04)
LagClubStoresShare	0.033	0.055***	0.049***	0.026
	(0.016)	(0.015)	(0.017)	(0.022)
LagConvenienceShare	-0.012***	-0.009***	-0.010**	-0.009*
	(0.003)	(0.004)	(0.004)	(0.005)
LagNaturalGourmetShare	-0.004	0.021***	0.043***	0.044***
	(0.007)	(0.009)	(0.011)	(0.014)
LagSuperetteShare	-0.008***	-0.002	0.002	-0.002
	(0.003)	(0.004)	(0.005)	(0.006)
LagDollarShare	-0.022***	-0.013*	-0.010	-0.017*
	(0.006)	(0.007)	(0.008)	(0.009)
LagSupermarketEntry	-0.0009	-0.0007	-0.0008	-0.0008
	(0.003)	(0.003)	(0.002)	(0.002)
LagSupercenterEntry	-0.0004	0.0057	0.0067	0.004
	(0.008)	(0.006)	(0.005)	(0.004)
LagClubEntry	-0.011*	-0.013*	-0.006	-0.005
	(0.009)	(0.007)	(0.006)	(0.005)
LagNaturalGourmetEntry	0.011**	0.06*	0.0007	0.0013
	(0.005)	(0.004)	(0.003)	(0.003)
Intercept	0.045***	0.043***	0.037***	0.04***
	(0.004)	(0.004)	(0.005)	(0.005)

Standard errors in parentheses.
***: Coefficient is significant at the 0.01 level. **: At the 0.05 level. *: At the 0.10 level.

Table 5. Multinomial Results for Estimating (1), the Determinants of Household Fruit and Vegetable Expenditure Share.

Variable	Group 1 Comparison	Group 2 Comparison
LagHHIFood	0.926**	0.879***
	(0.030)	(0.038)
LagSupermarketShare (omitted)		
LagLimAssortmentShare	0.530***	0.433***
	(0.118)	(0.127)
LagWarehouseShare	0.496	0.438
	(0.276)	(0.308)
LagSupercenterShare	0.999	0.663*
	(0.167)	(0.152)
LagClubStoresShare	2.899***	7.358***
	(1.028)	(3.250)
LagConvenienceShare	0.891	0.652***
	(0.067)	(0.063)
LagNaturalGourmetShare	1.421*	2.162***
	(0.291)	(0.531)
LagSuperetteShare	0.892	0.820*
	(0.076)	(0.091)
LagDollarShare	0.733**	0.411***
	(0.116)	(0.086)
LagSupermarketEntry	0.957	1.004
	(0.048)	(0.064)
LagSupercenterEntry	1.251**	1.114
	(0.143)	(0.168)
LagClubEntry	0.753**	0.692**
	(0.098)	(0.117)
LagNaturalGourmetEntry	1.123	1.268**
	(0.087)	(0.120)
Midwest	1.071**	1.212***
	(0.029)	(0.042)
South	1.142***	1.285***
	(0.027)	(0.040)
Northeast	1.103***	1.179***
	(0.028)	(0.039)
HHIncome	1.064***	1.110***
	(0.003)	(0.004)
Fulltime_F	1.099***	1.162***
	(0.020)	(0.028)
Fulltime_M	1.037*	1.020
	(0.022)	(0.028)
College_F	1.250***	1.556***
	(0.024)	(0.039)
College_M	1.086***	1.247***

	(0.023)	(0.035)
Gradschool_F	1.315***	2.062***
	(0.038)	(0.072)
Gradschool_M	1.141***	1.545***
	(0.034)	(0.057)
White	0.909***	0.909***
	(0.023)	(0.031)
Asian	1.078	1.510***
	(0.056)	(0.101)
Married	1.428***	1.274***
	(0.033)	(0.041)
HHSize	0.734***	0.525***
	(0.007)	(0.009)
HHAge	1.032***	1.063***
	(0.008)	(0.001)
Intercept	0.187***	0.023***
	(0.016)	(0.003)

Standard errors in parentheses.
***: Coefficient is significant at the 0.01 level. **: At the 0.05 level. *: At the 0.10 level.

Appendix A: Annual Sales Categories

TDLinx Category	Store-Level Annual Sales Range	Frequency (Percent of Total Stores)
1	$1 to $500,000	340 (0.05)
2	$500,001 to $1,000,000	14,039 (1.87)
3	$1,000,001 to $1,500,000	148,337 (19.76)
4	$1,500,001 to $2,000,000	141,189 (18.81)
5	$2,000,001 to $4,000,000	235,652 (31.39)
6	$4,000,001 to $6,000,000	66,517 (8.86)
7	$6,000,001 to $8,000,000	31,340 (4.19)
8	$8,000,001 to $12,000,000	22,309 (2.97)
9	$12,000,001 to $16,000,000	14,776 (1.97)
10	$16,000,001 to $20,000,000	14,334 (1.91)
11	$20,000,001 to $25,000,000	16,356 (2.18)
12	$25,000,001 to $30,000,000	11,412 (1.52)
13	$30,000,001 to $35,000,000	6,564 (0.87)
14	$35,000,001 to $40,000,000	6,671 (0.89)
15	$40,000,001 to $45,000,000	4,410 (0.59)
16	$45,000,001 to $50,000,000	3,499 (0.47)
17	$50,000,001 to $75,000,000	8,160 (1.09)
18	$75,000,001 to $100,000,000	1,986 (0.26)
19	$100,000,001 and up	2,782 (0.37)

Source: Nielsen TDLinx, 2004-2014.