The Geography of Poverty and Nutrition: Food Deserts and Food Choices Across the United States

Hunt Allcott, Rebecca Diamond, and Jean-Pierre Dubé*

July 20, 2015

Abstract

We study how and why healthful eating varies by income in the United States. First, we use the Nielsen Homescan and retail scanner data to document that the lowest-income households consume groceries that average 0.57 standard deviations less healthful than the highest-income households. Little of this difference can be explained by differences in local supply of healthful foods: in a reduced-form event study framework, entry of additional large grocery stores and supercenters reduces travel costs but does not affect average healthfulness, even for households in "food deserts." We then estimate a formal demand model and use estimated preferences to decompose differences in purchasing patterns between low- and high-income households. We find that 92% of the nutrition-income gradient is driven by differences in demand across products, while only 8% can be attributed to differences in supply. The demand impact can be broken down into 72% due to differences in UPC demand, 25% due to product group demand differences, and -5% due to nutrient preferences. The 8% supply effect is entirely driven by different product offerings, with prices playing essentially no role.

*Allcott: New York University and NBER. hunt.allcott@nyu.edu. Diamond: Stanford GSB. diamondr@stanford.edu. Dubé: Chicago Booth and NBER. Jean-Pierre.Dube@chicagobooth.edu. We thank Catherine Wright for exceptional research assistance. We thank seminar participants at the 2015 NBER Summer Institute for helpful comments. We are grateful for funding from the Chicago Booth Initiative on Global Markets.

I Introduction

Thirty-five percent of American adults are obese, up from 15 percent in the late 1970s (NCHS 2013, 2014).[1] Obesity is estimated to be responsible for 10-27 percent of US medical costs, amounting to several hundred billion dollars annually (Finkelstein *et al.* 2009, Cawley *et al.* 2015). At least for women, there are meaningful socioeconomic differences in obesity: low-income women are 45 percent more likely to be obese than high-income women, and women who have not completed college are about 70 percent more likely than those who have (Ogden *et al.* 2010). Differences in eating patterns are a leading potential cause of this health disparity.

Many factors could cause a "nutrition-income gradient." Low- vs. high-income people could have different tastes, perhaps as a result of different early-life eating patterns. Poverty is associated with lower education, which could drive differences in health knowledge and cognition (Cutler and Lleras-Muney 2010). Unhealthful high-calorie foods could cost less per calorie in money or preparation time.[2]

A large public health literature studies these issues, giving substantial attention to the idea that lack of availability of healthful foods in low-income neighborhoods causes the poor to consume less healthful food. This literature has documented that low-income areas have fewer supermarkets and generally lower supply of healthful food and that there is a cross-sectional correlation between living in an under-served "food desert" and eating less healthfully (Larson, Story, and Nelson 2009). One explanation for such a correlation is that supply is simply responding to demand in equilibrium. Another potential explanation is that higher costs of infrastructure, product distribution, and operation for supermarkets might reduce supply in low-income neighborhoods even if demand were the same; see Food Marketing Institute (1998) for evidence on such costs. Furthermore, in the presence of fixed costs, preference externalities could magnify the effects of local demand differences (Waldfogel 2003, 2008). Recent empirical evidence provides some indirect support for a food desert hypothesis: Ellickson (2006, 2007) highlights the role of fixed costs in grocery retail, and people who moved to higher-income neighborhoods in the Moving to Opportunity randomized experiment were less likely to be obese (Kling, Liebman, and Katz (2007), Ludwig *et al.* (2011)).

The idea that lack of supply causes lower demand for healthful foods has inspired substantial policy action to increase supply. The 2001 Food Poverty Eradication Bill required local and national governments in the United Kingdom to document and take actions to eliminate food deserts. In the U.S., the Healthy Food Financing Initiative has awarded over $140 million since 2011 through a suite of programs that finance and provide technical assistance to grocery stores, farmers markets, and other suppliers of healthful foods in under-served areas (TRF 2015a). Pennsylvania's Fresh Food

[1]Obesity is defined as having Body Mass Index (BMI) larger than 30 kg/m^2. BMI is weight in kilograms divided by the square of height in meters.

[2]Cutler, Glaeser, and Shapiro (2003), Lakdawalla and Philipson (2002), and Philipson and Posner (1999) highlight this issue and argue that cost decreases are largely responsible for the obesity increase of the past 40 years.

Financing Initiative provided $85 million grants and loans to retailers offering fresh foods in underserved low-income areas (TRF 2015b). Projects aimed at "eliminating food deserts" were eligible for the $100 million in Community Transformation Grants under the Affordable Care Act (HHS 2011). However, it is difficult to evaluate such supply side interventions without understanding the market failures that might justify them, the true underlying causes of a nutrition-income gradient, or at a minimum their causal impacts on healthful food consumption.[3]

This paper asks two sets of questions. The first is descriptive: *to what extent do low-income households purchase less healthful foods, and to what extent do stores in low-income neighborhoods offer less-healthful foods?* The second is causal: *to what extent do shifts in supply of healthful foods affect equilibrium consumption, and more broadly, what factors explain the "nutrition-income gradient"?* To answer these questions, we exploit Nielsen Homescan, a 60,000-household, nationally-representative panel survey of grocery consumption, and Nielsen RMS, which provides product-level sales data for 35,000 food retail stores nationwide.

We begin by laying out basic stylized facts of the nutrition-income gradient. First, Americans - even low-income households in urban areas - travel long distances for shopping. The average shopping trip is five miles one-way, and even the five percent of US households that do not own cars travel a mean of 2.5 miles one way. This will moderate the extent to which local neighborhood supply affects choice sets. Second, the RMS data show that stores in lower-income neighborhoods offer substantially less produce and overall less healthful grocery items than stores in higher-income areas. However, this is almost entirely explained by the sizes and types of stores that enter in different neighborhoods: low-income zip codes have 2-3 times more convenience stores and small grocery stores, and 40 percent fewer large grocery stores per capita. Third, low-income households get 78 percent of their groceries from what we call "supergrocers" - chain grocery stores, supercenters, and club stores, all of which tend to offer a wide variety of groceries, including healthful items and fresh produce - against 84 percent for the highest-income households. Fourth, there is a clear nutrition-income gradient on various metrics: low-income households consume less produce, fiber, and protein, and more sugar and saturated fat per 1000 calories. On our Health Index, a composite measure of healthfulness based on US government nutritional recommendations, the lowest-income households average 0.57 standard deviations worse than the highest-income households.

Taken together, these initial facts show that the nutrition-income gradient is correlated with supply, although the causality is yet to be established. Fifth, however, we show that the cross-sectional nutrition-income gradient is largely unaffected by controlling for measures of supply, and the magnitude attenuates by only 27 percent when measured within-Census tract, which isolates households with similar supply environments (up to within-tract variation). Sixth, we show that price is unlikely to explain the nutrition-income gradient: although the wealthy typically pay more

[3] In an influential review article, Bitler and Haider (2011) write that "it appears that much of the existing research implicitly assumes that supply-side factors cause any food deserts that exist." However, "we are unaware of any study that has systematically examined whether supply or demand factors explain the existence of food deserts."

for groceries (Broda, Leibtag, and Weinstein 2009), the *relative* price of more healthful groceries does not differ by income. Furthermore, healthful foods are not unambiguously cheaper or more expensive per calorie, which does not support the hypothesis that the poor eat less healthful foods because they are a cheaper source of calories.

We then turn to an event study framework to measure the causal effect of grocery supply on demand. We construct two datasets of "supergrocer" entry, one based on entry by several specific grocery and/or supercenter chains, and the other based on counts of retailers by size and channel type from Zip Code Business Patterns.[4] Using both datasets, we show that Homescan households significantly shift expenditures toward entrant supergrocers. However, more than half of the expenditure changes are simply diversion of sales from *other* supergrocers offering similar choice sets. Thus, entry has an economically limited impact on healthful grocery purchases. The qualitative results are similar when limiting to the subset of Homescan households in "food deserts," i.e. with no supergrocers in the zip code or at zip codes within three miles. For households in these food desert neighborhoods, the confidence intervals bound the effects of supply at a small share of the rich/poor gap in nutritional choices.

These reduced-form results leave open the question of what factors do cause the nutrition-income gradient. To answer this, we formalize and estimate a demand model based on Dubois, Griffith, and Nevo (2014). The model allows for Constant Elasticity of Substitution (CES) preferences over individual product UPCs, combined with Cobb-Douglas preferences for product groups (milk, breads, candy, vegetables, etc.) and aggregate preferences over specific macronutrients (saturated fat, sugar, salt, etc.). Like many continuous demand systems, the functional form assumptions are somewhat restrictive, but this approach has the benefit of giving a simple and transparent estimating equation.

These estimates show that preferences for healthful macro nutrients do not systematically vary across the income distribution. Low income households dislike unhealthful cholesterol and sodium more than higher income households. However, higher income households have a stronger demand for fruit and fiber.

In contrast, we find substantial preference heterogeneity across product group categories and individual UPCs. Higher income households prefer produce and prepared foods, while lower-income households prefer to consume more frozen foods. It appears other product characteristics, such as shelf life, lead lower income households to choose less healthful foods, as opposed to a direct preference to consume unhealthful nutrients.

Finally, we use our model of food demand to decompose the nutrition-income gradient into causes due to supply versus household preferences. We find that 92% of the nutrition-income gradient is driven by differences in demand across products, while only 8% can be attributed to differences in supply. The demand impact can be broken down into 72% due to differences in UPC

[4]We are not allowed to publicly identify specific retailers using the Nielsen data.

demand, 25% due to product group demand differences, and -5% due to nutrient preferences. The 8% supply effect is entirely driven by different product offerings, with prices playing essentially no role.

The remainder of this section discusses related literature. Sections II through IX, respectively, present data, stylized facts, the event study of retailer entry, the demand model, estimation, results, demand decomposition, and the conclusion.

I.A Related Literature

Our paper fits within a broad literature on the economics of nutrition and obesity; see Cawley (2015) for a recent review. Within this literature, Anderson and Matsa (2011), Currie, DellaVigna, Moretti, and Pathania (2010), Davis and Carpenter (2009), and Dunn (2010) study the effects of proximity to fast food restaurants on food consumption and obesity, complementing our analysis of supermarket entry. Their results are qualitatively consistent with ours in that they suggest that the causal impact of unhealthful food supply is small, either relative to the overall obesity rate or the nutrition-income gradient. Courtemanche and Carden (2011) study the effects of Walmart Supercenter entry on obesity, using the Holmes (2011) instrument for Walmart expansion (distance from the headquarters in Bentonville, Arkansas).[5] Also using the Holmes (2011) instrument, Volpe, Okrent, and Leibtag (2013) study the effects of supercenter expansion on the healthfulness of grocery purchases.[6]

In the public health literature, Larson, Story, and Nelson (2009) review 54 studies documenting differences in food access across neighborhoods. However, these studies are typically either detailed inventories of product availability at specific retailers in specific local areas (e.g. Sharkey, Horel, and Dean (2010)) or nationwide studies of store counts by neighborhood income without granular product data (e.g. Powell *et al.* (2007)). Even in establishing stylized facts, the Nielsen Homescan and RMS data are groundbreaking in that they provide both large nationwide samples *and* granular availability and purchase information. There are small handful of studies in the public health literature that examine food consumption before vs. after supermarket entry, e.g. Cummins *et al.* (2005), Cummins, Flint, and Matthews (2015), Elbel *et al.* (2015), and Wrigley, Warm, and Margetts (2003)." However, according to a major report by the USDA (2009, page v), "the findings are mixed," perhaps because the standard in this literature has been to study entry of *one* retail establishment, which limits statistical power and generalizability and makes it more difficult to

[5] Courtemanche and Carden (2011) find that an additional Supercenter increases BMI and obesity. Our results that grocery and/or supercenter entry does not significantly affect nutritional content per calorie consumed provide support for their interpretation that mechanism is reduced prices per calorie.

[6] Volpe, Okrent, and Leibtag (2013) find that supercenter expansion reduces the healthfulness of grocery purchases. Their results could differ from ours for a number of reasons; relative to their paper, we benefit from having precise dates and geocoded locations of store entry for several different grocery and/or supercenter chains, which allows crisp identification in a difference-in-differences estimator.

establish a credible counterfactual. By contrast, our study evaluates the effects of *thousands* of supermarkets and supercenters as they enter and exit across the U.S. during a nine-year period.

A very nice paper by Handbury, Rahkovsky, and Schnell (2015) is perhaps the most closely-related analysis. Like us, they present stylized facts and estimate the effects of store entry on healthful grocery consumption, and our stylized facts and difference-in-differences estimates in Sections III and IV are different but broadly consistent with theirs. They also present a theoretical model that illustrates the potential role of preference externalities and cost differences in causing healthful grocery supply to differ across low- vs. high-income neighborhoods. Our Sections V, VII, and VIII build on these reduced-form empirical results by specifying and estimating preferences and formally decomposing the sources of the nutrition-income gradient.

II Data

II.A Nielsen Homescan and Retail Scanner Data

We use the Nielsen Homescan Panel for 2004-2012 to measure grocery purchases. Homescan includes about 39,000 households each year for 2004-2006, and about 61,000 households each year for 2007-2012. Homescan households record UPCs of all consumer packaged goods they purchase from any outlet. We exclude all non-grocery purchases, considering only food and drink. See Einav, Leibtag, and Nevo (2010) for a recent validation study.

Homescan is less well-suited to study Americans' overall diets because it does not include data on food purchased in restaurants. However, the data are still well suited to measure how the grocery store supply environment affects grocery purchases.[7] One limitation of Homescan for our research question is that most households only record purchases of packaged items with UPCs, not non-packaged items such as bulk produce and grains. For 2004-2006, the data also include an 8,000-household "magnet" subsample that also recorded prices and weights of non-packaged items. We use the magnet data for robustness checks.[8] Appendix Figure A1 shows that about 60 percent of magnet households' produce calories are from packaged goods that are observed in the full Homescan sample, and this proportion does not vary statistically by income.[9]

[7] The National Health and Nutrition Examination Survey finds that 34 percent of calories are consumed away from home, including 25 percent that are consumed in restaurants (USDA 2014). For all income groups, the share of healthful and unhealthful macronutrients (protein, carbohydrates, saturated fat, etc.) consumed away from home is about the same as the share of calories consumed away from home, suggesting that the healthfulness of in-home food consumption might even be a reasonable proxy for overall diet healthfulness.

[8] The magnet data continue after 2006, but panelists now record only expenditures and not weights purchased. Because prices per unit weight can vary substantially across stores and neighborhoods, we do not use these data to construct food purchases.

[9] When limiting to fresh produce (excluding canned, frozen, and dried produce), magnet households purchase an average of 39 percent of calories from packaged items. Households with incomes less than about $20,000 buy about five percentage points less of their fresh produce calories from packaged items, but the proportion is constant at moderate and high incomes.

In addition to the standard Homescan data, we have additional questions from add-on surveys carried out by Bronnenberg, Dubé, and Gentzkow (2012) and Bronnenberg *et al.* (2013). As proxies for nutritional knowledge, we use indicator variables for whether a household member works in any health care occupation or in a subset of health occupations involving nutritional training.[10] We also use the score on a food knowledge quiz.[11] Panel A of Table 1 presents descriptive statistics for the Homescan data. Table 1 and all other Homescan results are weighted for national representativeness.

The Nielsen Retail Scanner Data (RMS) consists of weekly prices and sales volumes for each UPC sold at approximately 35,000 participating grocery, mass merchandiser, drug, convenience, and other stores at 94 retail chains for 2006-2012. RMS includes 53, 32, 55, and 2 percent of sales in the grocery, mass merchandiser, drug, and convenience store channels, respectively; we drop liquor stores. Analogous to Homescan, RMS does not include sales volumes of bulk produce and other non-UPD'd items, although it does include packaged produce.

II.B Grocery Retail Establishments

Studying the effects of grocery retailer entry requires reliable data on store open dates to avoid attenuation bias. Some datasets, such as InfoUSA and the National Establishment Time Series, might be reasonable for cross-sectional analyses, but they do not sufficiently precisely record the open dates of new establishments; see Bitler and Haider (2011, page 162) for further discussion. Furthermore, to measure true changes in availability experienced by consumers, we must use actual new establishments, not store locations that continue to operate but change management.

We measure entry with two datasets. First, we gathered the exact store open dates and addresses for several specific chains of large grocery stores and/or supercenters. This dataset includes 1732 new stores opened between January 2004 and December 2012.

[10]The health-nutrition subset includes dentists, dieticians, nutritionists, pharmacists, physicians, surgeons, physician assistants, registered nurses, and dental hygienists.

[11]Respondents were asked to identify the most common additive to table salt (correct answer: iodine), the scientific name for baking soda (correct answer: sodium bicarbonate), and the most common ingredient of granulated sugar (correct answer: sucrose). The variable "survey knowledge: food" is the share of these questions answered correctly.

Table 1: **Data: Descriptive Statistics**

Panel A: Nielsen Homescan Households

Variable	Non-missing observations	Mean	Standard deviation
Education (years)	484,810	14.3	2.14
Age	484,810	52.3	14.4
1(Household includes children)	484,810	0.33	0.47
Income ($000s)	484,810	62.4	45.4
1(Health care occupation)	54,988	0.069	0.25
1(Health-nutrition occupation)	54,988	0.029	0.16
Survey knowledge: food	40,856	0.550	0.33

Panel B: Zip Code Establishment Counts

Variable	Mean	Standard deviation
Grocery	2.40	4.45
Grocery (<10 employees)	1.17	3.41
Grocery (10-49 employees)	0.58	0.98
Grocery (>50 employees)	0.66	1.19
Supercenters/club stores	0.15	0.44
Drug/convenience stores	6.06	7.62

Panel C: UPC Characteristics

Variable	Mean	Standard deviation
Package size (grams)	752	1656
Calories	1085	1425
Grams per 1000 calories		
Fat	31.1	34.5
Saturated fat	10.1	19.6
Cholesterol	0.094	0.58
Sodium	8.07	286
Carbohydrates	144	141
Fiber	10.7	22.9
Sugar	68.6	78.5
Protein	31.4	45.8
Health Index	-4.73	125

Notes: 1(Health care occupation), 1(Health-nutrition occupation), and Survey knowledge: food are from Homescan add-on surveys carried out by Bronnenberg, Dubé, and Gentzkow (2012) and Bronnenberg *et al.* (2013). "Health-nutrition" occupations include the subset of health occupations involving nutritional training: dentists, dieticians, nutritionists, pharmacists, physicians, surgeons, physician assistants, registered nurses, and dental hygienists. "Survey knowledge: food" is the share the following questions answered correctly: identify the most common additive to table salt (correct answer: iodine), the scientific name for baking soda (correct answer: sodium bicarbonate), and the most common ingredient of granulated sugar (correct answer: sucrose). Zip code establishment counts are from Zip Code Business Patterns data for 2004-2012. UPC characteristics are for all 1.54 million UPCs that ever appear in the Nielsen Homescan or RMS data.

Second, we use Zip Code Business Patterns (ZBP), which gives a count of establishments by NAICS code and employment size category for every zip code as of March 10th of each year. These data are drawn from tax records and other administrative data, an in particular the U.S. Census Company Organization Survey. Appendix Table A1 shows that the Zip Code Business Patterns data date openings of specific supercenters in the correct year 50 to 80 percent of the time, although they are sometimes recorded a year later and sometimes in a broader "general merchandise" NAICS code (452) instead of the specific "supercenter and club store" NAICS code (452910). Panel B of Table 1 presents descriptive statistics for the Zip Code Business Patterns data.

II.C Nutrition Facts and the Health Index

We purchased UPC-level nutrition facts from Gladson, and we gathered nutrition facts for non-packaged items from the USDA National Nutrient Database for Standard Reference (USDA 2013). In our analysis, we separately model demand for macronutrients (protein, fat, sugar, etc.), but we also wish to characterize goods and preferences using a one-dimensional index of healthfulness. The most natural option is the USDA's Healthy Eating Index HEI, which was designed to score entire diets on an easily-understandable range from 0-100. However, an item's contribution to the HEI depends on the other items consumed, which is less appropriate for datasets like Homescan where we do not observe consumers' entire diets.

We construct a modified version of the HEI that is based on the same U.S. government dietary recommendations but is linear and additively separable in macronutrients. The U.S. government Dietary Guidelines are clearly organized around "good" macronutrients to "increase" vs. "bad" macronutrients to "reduce". Our Health Index $H(\mathbf{x})$ is the sum of good minus bad nutrients, weighting each by its recommended daily intake (RDI): $H(\mathbf{x}) = \sum_k G_k \frac{g_k}{r_k} - (1 - G_k) \frac{g_k}{r_k}$, where g_k is the grams of macronutrient k, r_k is the RDI for a normal adult, and G_k takes value 1 for "good" macronutrients and 0 for "bad" macronutrients. (See Appendix A for additional details.) For example, fruit, vegetables, and candy have average Health Indices per 1000 calories of 6, 10, and -5, respectively. Across all Homescan households from 2004-2012, the mean Health Index per 1000 calories purchased is -2.60, the standard deviation is 0.70, and the interquartile range is [-2.20,-3.01]. Panel C of Table 1 presents nutritional summary statistics across all UPC codes in the Nielsen Homescan and RMS data.

III Stylized Facts: Supply and Demand for Healthful Food

III.A Consumers Travel a Mean of Five Miles for Shopping

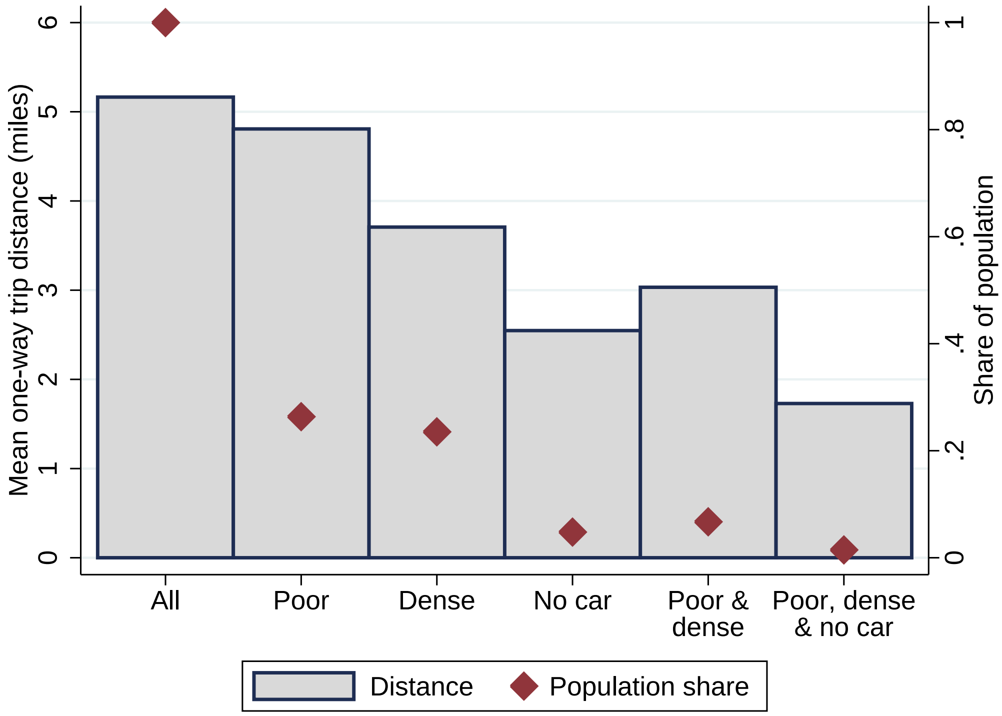
How far do people travel to shop? Living in a "food desert" would only affect individuals' choice sets if it is large enough to generate meaningful travel costs for shopping elsewhere. The 2009 National

Household Travel Survey (NHTS) is a nationally-representative survey that gathers demographics, vehicle ownership, and “trip diaries” for 150,000 households nationwide. Figure 1 shows average one-way distances for trips beginning or ending in “buying goods: groceries/clothing/hardware store.”

The data show that Americans travel a long way for shopping, typically in cars. The mean (median) trip is 5.2 (3.0) miles, and 90 percent of shopping trips are by car. Even low-income and urban subgroups typically travel some distance. “Poor” households (those with household income less than $30,000) travel a mean of 4.8 miles, with 82 percent of trips by auto. We define “Dense” census tracts as those with at least 4,000 people per square mile. These tracts contain 24 percent of US households; for comparison, Chicago has about 12,000 people per square mile, and Phoenix has 3200. “Poor” households in dense neighborhoods, which represent seven percent of US households, travel a mean of 3.0 miles, with 65 percent of trips by auto. Only 4.8 percent of households do not own a vehicle; they travel a mean of 2.5 miles, with 25 percent of trips by auto.[12] These facts help to understand results below showing that even households in zip codes without supergrocers make a substantial share of grocery expenditures at supergrocers.

[12]Appendix Figure A2 presents median travel distances and the share of trips by auto for the same subgroups. These results are broadly consistent with USDA (2009), which finds that 2.2 percent of US households live more than a mile from a supermarket and do not have access to a vehicle.

Figure 1: **Shopping Trip Distances by Household Income**



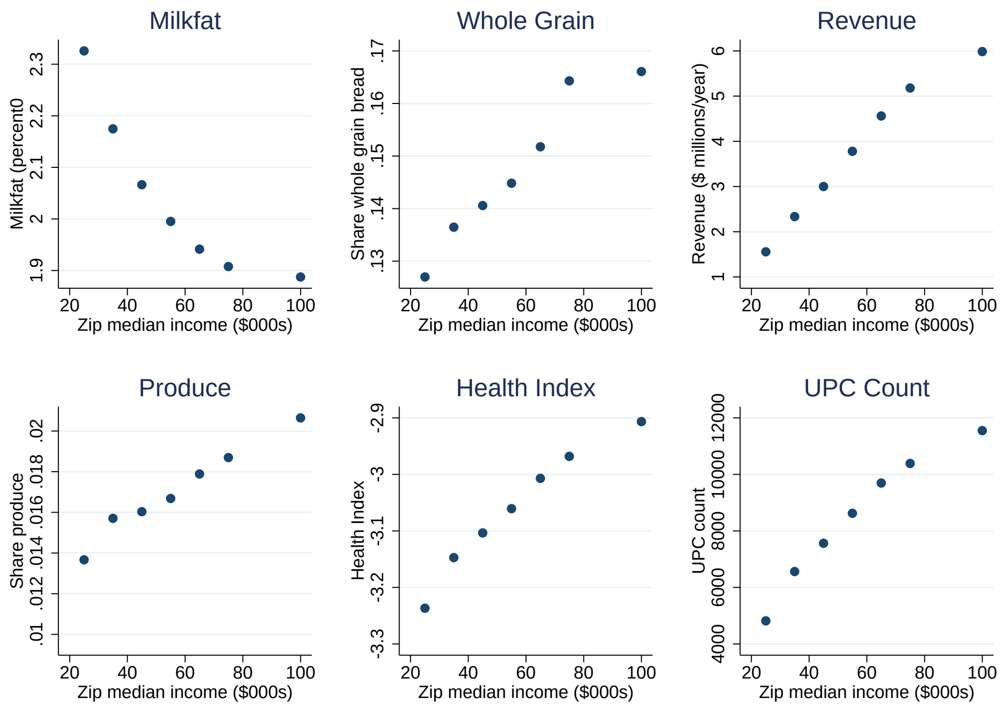
Notes: Data are from the 2009 National Household Travel Survey. Bars represent the mean one-way trip distance for trips beginning or ending in "buying goods: groceries/clothing/hardware store." "Poor" means household income less than $30,000, "Dense" means Census tract population density greater than 4,000 people per square mile, and "No car" means that the household does not own a car.

III.B Low-Income Zip Codes Have Less Healthful Supply

III.B.1 Stores in Low-Income Zip Codes Offer Less Healthful UPCs

Using RMS data, Figure 2 plots four different measures of the healthfulness of store product offerings: the average milkfat of milk UPCs, share of breads that are whole grain, share of UPCs that are produce, and mean Health Index across all UPCs offered. Because substantial amounts of produce are sold in bulk instead of packages with UPC codes, the produce figure is not a full account of the volume of produce available. All four panels show a consistent result: stores in wealthier zip codes offer much more healthful items. The two panels at the right of Figure 3 use RMS data to show that stores are significantly smaller in low-income neighborhoods: the mean store in zip codes with median household income below $30,000 offers about 5,000 UPCs, while the mean store in zip codes with median household income larger than $100,000 offers about 12,000 UPCs.

Figure 2: **Store Average Healthfulness and Size by Zip Code Median Income**



Notes: Using Nielsen RMS for year 2006, we constructed calorie-weighted mean milkfat of all milk UPCs, the calorie-weighted share of bread, buns, and rolls UPCs that are whole grain, the calorie-weighted share of UPCs that are produce, and the calorie-weighted mean health index across all UPCs offered, for each store. The left four panels of this figure presents the means of these variables across stores within categories of zip code median income. The right two panels present mean revenues and UPC counts by zip code income category, also using Nielsen RMS data.

Table 2 formalizes these correlations in store-level regressions using 2006 Nielsen RMS data. Columns 1 and 4 confirm that stores in higher-income zip codes offer substantially more produce UPCs and overall healthier items. Columns 2 and 5, however, show that conditioning on store size (as measured by revenues from grocery items with UPCs) explains almost all of the income-healthfulness relationship: large stores sell more healthful groceries and are much more likely to enter in higher-income zip codes. Columns 3 and 6 show that even when excluding the revenue variable, the store channel type also explains about 80 percent of the income-healthfulness relationship. Indeed, the R^2 in columns 3 and 6 show that channel types explain upwards of 95 percent of the variance in healthfulness across stores. This means that although the ZBP data have no information on the choice sets offered by different establishments, their data on count of establishments by channel type should still allow a very good prediction of these indices of healthful food availability for each zip code.

Columns 3 and 6 of Table 2 show that supercenters and grocery stores, in particular large groceries, have a wider variety of produce and much more healthful items than drug stores, convenience, stores, and other mass merchants, which includes discount stores like regular Walmart and K-Mart. There are no club stores such as Sam's Club or Costco in RMS, nor are there very small non-chain grocers. We refer to large (or chain) grocers, supercenters, and club stores as "supergrocers," using an intentionally awkward label to avoid confusion with other words. On average, supergrocers will offer more healthful items than non-supergrocers.

Table 2: **Healthful UPC Availability at Nielsen RMS Stores**

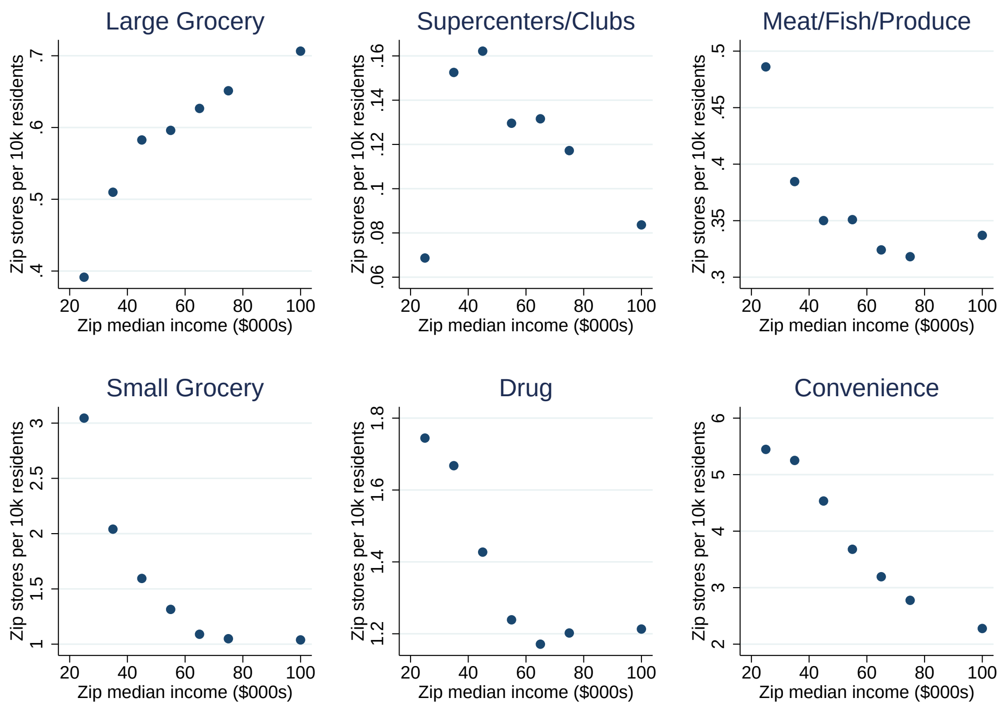
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	Count of Produce UPCs			Mean Health Index		
ln(Zip Median Income)	369.1	5.688	82.62	0.242	0.0148	0.0632
	(9.350)***	(6.359)	(3.290)***	(0.00596)***	(0.00392)***	(0.00328)***
ln(Annual revenue)		338.7			0.202	
		(1.636)***			(0.000954)***	
1(Large grocery)			629.1			-3.221
			(36.62)***			(0.0362)***
1(Small grocery)			261.9			-3.275
			(35.85)***			(0.0351)***
1(Supercenter)			217.1			-3.431
			(44.49)***			(0.0371)***
1(Drug store)			-832.0			-3.975
			(35.79)***			(0.0361)***
1(Convenience store)			-901.8			-4.087
			(36.36)***			(0.0358)***
1(Other mass merchant)			-810.8			-4.053
			(35.20)***			(0.0353)***
Constant	-3496.8	-4193.0		-5.685	-5.990	
	(100.5)***	(62.15)***		(0.0642)***	(0.0375)***	
Observations	31,539	31,539	31,539	31,539	31,539	31,539
R^2	0.0415	0.779	0.955	0.0493	0.805	0.997

Notes: Calendar year 2006 Nielsen RMS data at the store level. Omitted census division is Middle Atlantic, which consists of New York, Pennsylvania, and New Jersey. ln(Annual revenue) is revenue in 2006 from observed grocery items with UPCs. "Large" ("small") grocery stores are those with at least (less than) $5 million in observed revenue. Robust standard errors, clustered by zip code, in parentheses. *, **, ***: Statistically significant with 10, 5, and 1 percent confidence, respectively.

III.B.2 Low-Income Zips Have Higher Counts of Unhealthful Store Types

Using the Zip Code Business Patterns data, Figure 3 plots the average count of stores by channel type for zip codes by median income category. Zip codes vary substantially in area and population, so this figure normalizes store counts per 10,000 residents; the mean zip code population is 12,000. Lower-income zip codes have more stores of all channel types, with two exceptions. First, the concentration of large grocery stores is sharply monotonically increasing in median income. Second, while supercenters and club stores are less likely to enter in wealthier zip codes, they are also highly unlikely to enter in the very lowest-income zip codes. Appendix Figure A4 presents raw counts per zip code, without normalizing by population. The qualitative trends are similar, and the wealthiest zip codes average about 1.2 more large grocery stores than the lowest-income zip codes.

Figure 3: **Store Counts by Zip Code Median Income**



Notes: Population-weighted mean store counts by zip code income category from Zip Code Business Patterns. Large (small) grocers are defined as those with 50 or more (fewer than 50) employees.

Putting the RMS choice set information together with the ZBP store location information, we predict whether lower-income neighborhoods have overall less-healthful supply. Projecting columns 3 and 6 or Table 2 onto the ZBP store counts, the predicted average Health Index of available UPCs increases monotonically from -3.17 to -2.90 between the lowest- and highest-income zip codes. The count of packaged produce UPCs increases less steeply, although zips with median income less than $30,000 are projected to have many fewer produce UPCs because they have so many fewer supercenters and large grocery stores. See Appendix Figure A5 for details.

III.C Low-Income Households Spend a Smaller Share at Healthful Store Types

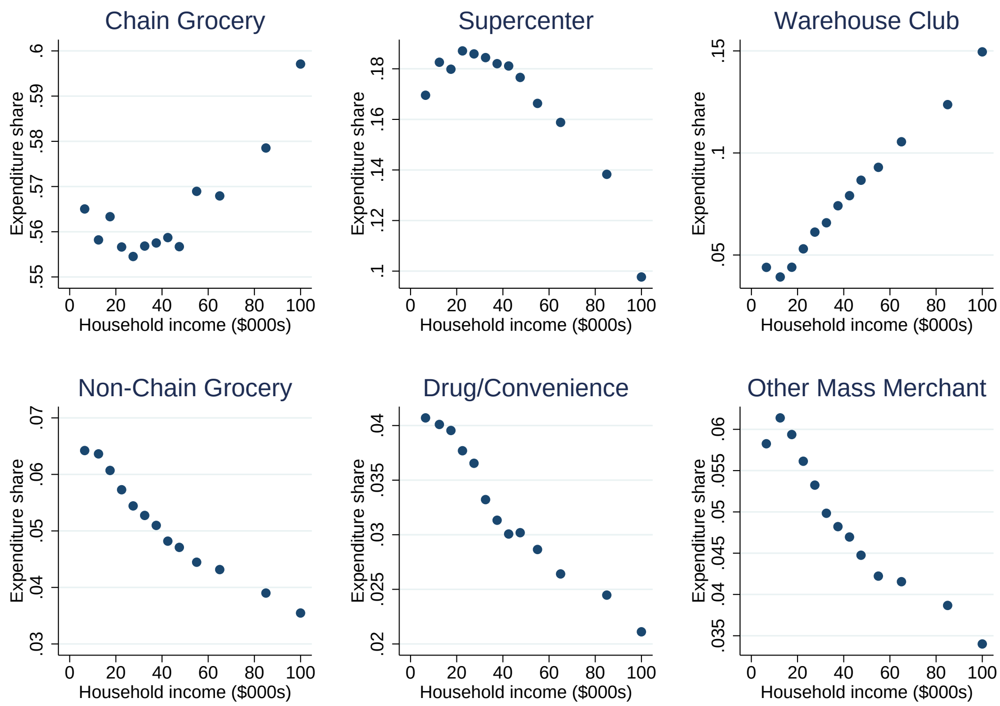
At what types of stores do people buy groceries?[13] According to the Homescan data, about 62 percent of groceries are bought at grocery stores, and this is fairly constant by income. However, Figure 4 shows other differences: low-income households spend more at non-chain groceries, more at supercenters such as Walmart Supercenter, SuperTarget, and Meijer, less at warehouse club stores such as Costco and Sam's Club, more at other mass merchants such as (regular) Walmart and Target, and more at drug and convenience stores.[14] Another 5-6 percent of expenditures are at channels not plotted in Figure 4, including bakeries, butchers, candy stores, liquor stores, fruit stands, and fish markets; this proportion is fairly constant by income.

The channel types offering more produce and more healthful items are plotted in the top row, while the channel types offering less produce and less healthful items are plotted in the bottom row. Adding across the "supergrocer" channels in the top row, there is an income gradient: 78 percent of low-income households' grocery expenditures are at chain groceries/supercenters/clubs, rising to 84 percent for the highest-income households.

[13]This paragraph extends the discussion on page 81 of Broda, Liebtag, and Weinstein (2009).

[14]We define a chain grocer as any retail chain that has no more than 75 percent of Homescan shoppers from any one county.

Figure 4: **Grocery Expenditure Shares by Household Income**

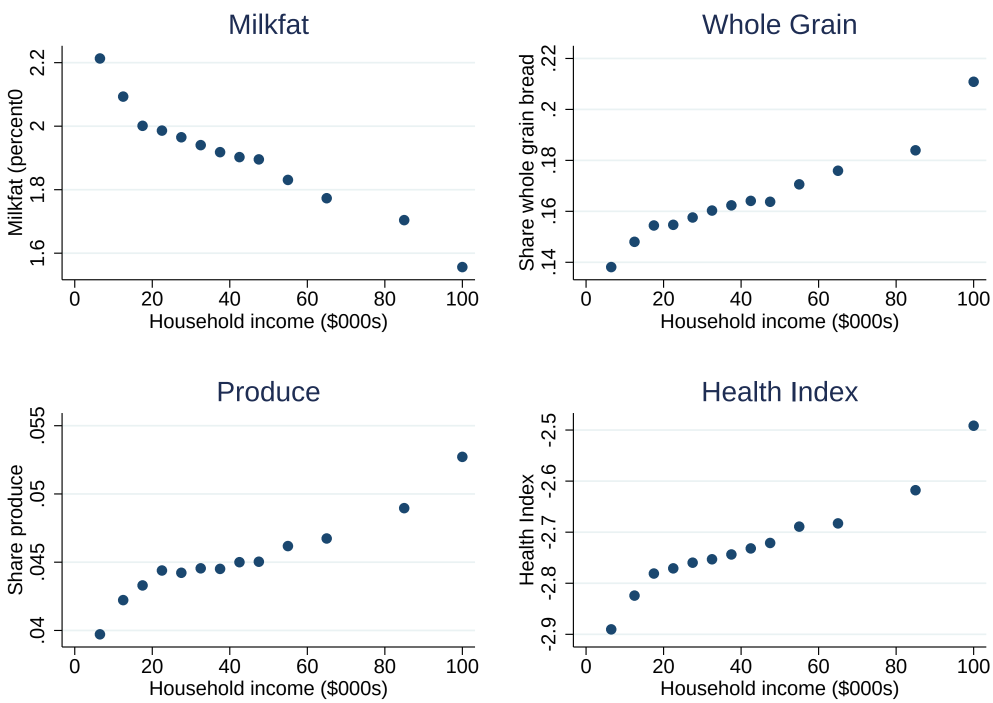


Notes: Nielsen Homescan data for 2004-2012. The x-axis presents nominal income bins; household incomes larger than $100,000 are coded as $100,000. Chain grocers are defined as any retail chain that has no more than 75 percent of Homescan shoppers from any one county.

III.D There is a Positive "Nutrition-Income Gradient," Largely Unexplained by Cross-Sectional Supply Differences

Figure 5 summarizes the healthfulness-income relationship for milkfat, whole grain, produce, and Health Index using Nielsen Homescan purchase data. Mirroring the availability presented in Figure 2, low-income consumers purchase higher-fat milk, less whole-grain breads, less produce, and lower-Health Index UPCs. The highest-income consumers get about 1.3 percentage points more calories from produce than the lowest-income (5.3 vs. 4 percent) and purchase items with 0.4 higher Health Index (-2.5 vs. -2.9). Given that the standard deviation of average Health Index across households is 0.70, this means that the lowest-income households' Health Index is 0.57 standard deviations lower.

Figure 5: **Healthful Purchases by Household Income**



Notes: Nielsen Homescan data for 2004-2012. The x-axis presents nominal income bins; household incomes larger than $100,000 are coded as $100,000. Milkfat is the calorie-weighted average milkfat of milk purchases, whole grain is the calorie-weighted average share of bread, buns, and rolls purchases that are whole grain, produce is the share of calories from fresh, canned, dried, and frozen fruits and vegetables, and Health Index is the average Health Index per 1000 calories.

The Health Index is comprised of individual macronutrients, and our demand model also allows individual macronutrients to enter utility separately. Appendix Figure A6 presents an analogue to Figure 5, but considering each individual macronutrient. Purchases of protein and fiber per 1000 calories clearly increase in income, while saturated fat and sugar per 1000 calories decreases in income, and sodium and cholesterol are not statistically associated with income. Thus, although the US Recommended Daily Allowances impose specific weights on macronutrients in our Health Index, our results would be qualitatively similar as long as one models protein and fiber as healthful, and saturated fat and sugar as unhealthful.

The produce and Health Index data in Figure 5 do not include bulk purchases. Appendix Figure A7 re-creates the figure using the magnet subsample for 2004-2006. The results are noisier due to the smaller sample, but qualitatively similar. For produce purchases, the positive slope in income is still positive, but slightly more attenuated until incomes exceed $80,000.

Table 3 formalizes the relationship between household income and Health Index per 1000 calories purchased, using the 2004-2012 Homescan data. In column 2, conditioning on household demographics (in particular, education) attenuates this relationship by about 20 percent. Columns 3 and 4 show that including observable cross-sectional measures of supply do not substantially attenuate the nutrition-income relationship. For column 3, we predict average health index across all UPCs offered and the count of produce UPCs available in each zip code by fitting RMS store characteristics from columns 3 and 6 of Table 2 to the store counts in each zip code from Zip Code Business Patterns. In column 4, we directly include the counts of stores by channel type in the household's zip code. Column 5 includes a vector of household census tract indicator variables, allowing us to compare households with the same supply environment but different incomes.

Columns 3-5 all suggest a limited role for the supply side in explaining the nutrition-income relationship. Certainly, the location of stores in columns 3 and 4 is endogenous to demand, which should bias the zip code characteristics away from zero and the ln(Household income) coefficient toward zero relative to the causal relationship. This only strengthens the suggestion that supply has a limited role. In Section IV, we take this a step further by testing how store entry affects healthful food purchases.

Table 3: **The Nutrition-Income Gradient**

	(1)	(2)	(3)	(4)	(5)
ln(Household income)	0.116	0.0900	0.0875	0.0872	0.0654
	(0.00380)***	(0.00396)***	(0.00398)***	(0.00396)***	(0.00402)***
Zip Health Index			0.125		
			(0.0198)***		
ln(Zip produce UPCs available)			0.0152		
			(0.00580)***		
Zip count large grocers				0.0205	
				(0.00210)***	
Zip count small grocers				0.00156	
				(0.000619)**	
Zip count supercenters/club stores				0.00635	
				(0.00419)	
Zip count drug/convenience stores				-0.00252	
				(0.000420)***	
Educ, age, children, race controls		Yes	Yes	Yes	Yes
Census tract indicators					Yes
Observations	484,696	484,696	481,185	483,015	449,752
R^2	0.0191	0.0648	0.0663	0.0665	0.516

Notes: Nielsen Homescan data at the household-by-year level. The dependent variable is mean Health Index per 1000 calories of groceries purchased; the mean is -2.60, and the interquartile range is [-2.20,-3.01]. Observations are weighted for national representativeness. Robust standard errors, clustered by household, in parentheses. *, **, ***: Statistically significant with 10, 5, and 1 percent confidence, respectively.

III.E Healthful Food Does Not Necessarily Cost More, and the Poor Do Not Pay Relatively More

The cost of healthful foods could also explain a nutrition-income gradient. Table 4 explores two different versions of this hypothesis. Columns 1 and 2 test whether healthful UPCs cost more. We take the sample of all transactions in 2012 and regress the natural log of price per 1000 calories on UPC attributes. Column 1 shows that it is cheaper to derive calories from carbohydrates and fats, and more expensive to derive calories from protein, fiber, sugar, and produce. "Good" and "bad" macronutrients (as classified by the US government for "increased" or "decreased" consumption) are sometimes cheaper and sometimes more expensive. Column 2 regresses price on the UPC's Health Index per 1000 calories, showing that in aggregate, more healthful UPCs are actually *less* expensive. We caveat that this result is contingent on the specific Health Index weights imposed by the US Recommended Daily Intakes, and the magnitude is also very small: a one-unit increase in the Health Index (more than one standard deviation) costs only 0.124 percent less.

Columns 3 and 4 test whether healthful UPCs cost *relatively* more in low-income neighborhoods. These regressions extend Table 1 of Broda, Leibtag, and Weinstein (2009), who show that conditional on UPC fixed effects, higher-income Homescan consumers pay slightly more for the same UPCs. Columns 3 and 4 present the coefficients on interactions between household income and the reported macronutrient, although with the main effect of household income. Column 3 shows that low-income consumers pay relatively less for UPCs containing more of both "good" and "bad" macronutrients, and Column 4 shows that low-income consumers do not pay relatively more or less for healthful UPCs. Estimates are very precise due to the large sample of transactions; standard errors in all columns are clustered by UPC.

Table 4: **The Price and Relative Price of Healthfulness**

	(1)	(2)	(3)	(4)
ln(Household income)			-0.0509	0.00241
			(0.0181)***	(0.000193)***
Fresh Fruit	0.168		0.00877	
	(0.108)		(0.00773)	
Fresh Vegetable	1.069		0.0135	
	(0.0842)***		(0.00877)	
Non-Fresh Fruit	0.325		0.00365	
	(0.0262)***		(0.00714)	
Non-Fresh Vegetable	-0.0258		-0.00200	
	(0.0265)		(0.00479)	
Fiber	0.0102		0.000116	
	(0.000584)***		(0.000119)	
Protein	0.00226		0.000213	
	(0.000313)***		(0.0000662)***	
Carbohydrates	-0.00772		0.000211	
	(0.000332)***		(0.0000806)***	
Fat	-0.0172		0.000281	
	(0.000723)***		(0.000184)	
Saturated fat	-0.00219		0.000199	
	(0.000549)***		(0.000124)	
Sugar	0.00223		0.00000865	
	(0.000153)***		(0.0000242)	
Sodium	0.0263		0.000628	
	(0.000813)***		(0.000232)***	
Cholesterol	-0.221		0.00401	
	(0.0123)***		(0.00160)**	
Health Index		-0.00124		0.0000137
		(0.000222)***		(0.0000102)
Observations	36,612,566	40,532,084	36,612,566	40,532,084
UPC fixed effects			Yes	Yes

Notes: 2012 Nielsen Homescan data at the transaction level. Dependent variable is natural log of price per 1000 calories. In columns 3 and 4, reported coefficients other than ln(Household income) are for the interaction of the stated variable with ln(Household income). Observations are weighted for national representativeness. Robust standard errors, clustered by UPC, in parentheses. *, **, ***: Statistically significant with 10, 5, and 1 percent confidence, respectively.

IV Effects of Retailer Entry

We use a difference-in-differences estimator to measure effects of grocery store entry on grocery purchasing patterns. Y_{it} is an outcome for household i in quarter t, μ_{rt} is a vector of Census region-by-quarter of sample indicators, and ϕ_i is a household fixed effect. P_{it}^d is an indicator taking value 1 if a known retailer from our known retailer dataset has previously opened within distance band d of household i. We use two distance bands, $d = 0 - 5$ miles and $d = 5 - 10$ miles. Using household-by-quarter data, the regression is:

$$Y_{it} = \sum_d \tau^d P_{it}^d + \mu_{rt} + \phi_i + \varepsilon_{it} \quad (1)$$

We drop observations from the quarter in which a store enters, as this is neither completely pre-entry nor post-entry.

We construct an analogous regression using the Zip Code Business Patterns data. Define S_{zt} and G_{zt}, respectively, as the count of supercenters/warehouse club stores and large (at least 50 employee) grocery stores in zip code z in year t. Using household-by-year data and now denoting μ_{rt} as Census region-by-year indicators, the regression is:

$$Y_{izt} = \tau^S S_{zt} + \tau^G G_{zt} + \mu_{rt} + \phi_i + \varepsilon_{izt} \quad (2)$$

In both regressions, standard errors are clustered by household. The τ coefficients measure the effect of entry and store counts under the identifying assumption that store entry and exit is exogenous to within-household preference changes over time. While retailers carefully plan entry and exit in response to local population growth and changes in local demographics, our identifying assumption seems plausible *within*-household.

IV.A Entry by Known Retailers

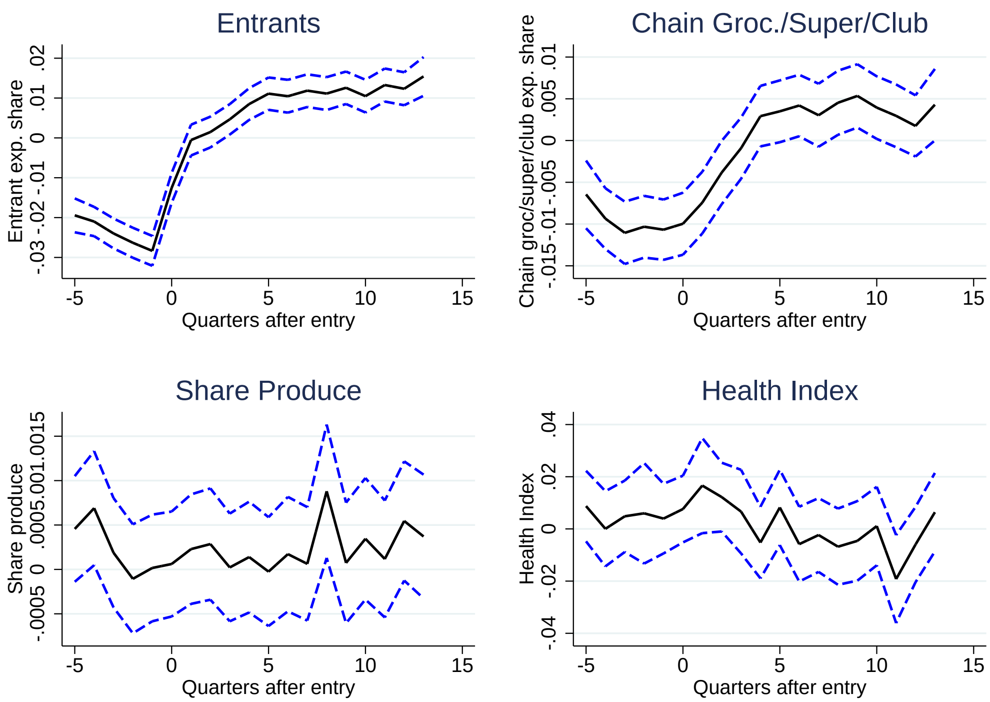
Figure 6 illustrates how store entry within five miles of household i affects purchases in event time, normalizing the quarter of entry to 0 on the x-axis. The top two panels present expenditure shares, showing that entry clearly affects purchasing patterns. The dependent variable in the top left panel is the combined expenditure share across the handful of retail chains we observe in our entry dataset. Expenditures at entering retailers increase by about 3.5 percentage points, and most of the adjustment occurs within the first four quarters after entry.[15] This is likely to understate the expenditure share increase at the specific entering store, as the entrant store will reduce expenditures at other stores within the same retailer or at the other retailers in our entry dataset. The top right figure shows the change in combined expenditures at chain grocers,

[15]This gradual adjustment of purchases is consistent with the results of Atkin, Faber, and Gonzalez-Navarro (2014), who study retail expansion in Mexico.

supercenters, and club stores (i.e. our "supergrocers"), which as we have shown above tend to offer the healthiest UPCs and widest variety of produce. To the extent that entry of large grocers diverts sales from other supergrocers, the actual changes in product availability (and thus the possible effects on healthful purchases) are more limited. Indeed, expenditure shares at supergrocers increase by only about 1.5 percentage points after entry (noting the difference in the y-axis scales between the top left and top right). Thus, a substantial effect of entry is to divert sales from other supergrocers.[16]

The bottom two panels show produce consumption and average Health Index per 1000 calories in event time. There is no statistically significant trend before or after entry.

Figure 6: **Known Retailer Entry in Event Time**



Notes: Presents effects of entry by several chains of large grocers and/or supercenters, using household-by-quarter Homescan data with household fixed effects and Census division-by-quarter of sample indicators. Top two panels present effects on expenditure shares, residual of controls. Bottom two panels present effects on healthful purchasing, residual of controls. Share Produce is the share of calories from fresh, canned, dried, and frozen fruits and vegetables; Health Index is the average Health Index per 1000 calories purchased.

Table 5 presents estimates of Equation (1). Panel A considers effects on expenditure shares at entrant retailers and all supergrocers. As a model with transport costs would predict, all effects are

[16]These results are consistent with those of Hwang and Park (2015), who look at a subset of Walmart Supercenters that opened between 2003 and 2006.

larger if a store enters within five miles of a household compared to if a store enters further away. Columns 1 and 2 consider Homescan households in all zip codes, while columns 3 and 4 consider the subset of households in "food deserts": zip codes with zero large (>50 employee) grocery stores, supercenters, or club stores in that zip code or at zip codes with centroids less than three miles away. These results formalize the graphical results in Figure 6, showing that about half of the expenditure share change is diverted sales from other supergrocers, while the other half is diverted from non-chain grocers, drug and convenience stores, other mass merchants, and specialty retailers that may offer a less healthful product selection. Columns 3 and 4 show that the expenditure share changes are larger in "food deserts," although a large proportion of sales are still diverted from other supergrocers. Appendix Table A3 shows that most of this diversion is from other mass merchants and non-chain grocers, and there is no statistically significant diversion from drug and convenience stores, although the point estimate is negative.

Panel B of Table 5 presents effects on the produce share of calories consumed and the Health Index per 1000 calories. Columns 1 and 2 show that for the average zip code, grocery entry has only a margin effect on healthful eating. For households within 5-10 miles of a new entrant, Health Index increases by a marginally significant point estimate of 0.01. (Recall that the standard deviation of Health Index across Homescan households is 0.7, and the difference between the lowest- and highest-income households in Figure 5 is around 0.4.) Column 3 shows that grocery entry may have a marginally significant impact on produce calorie share for households in food deserts that live within 5-10 miles of the entrant. However, Appendix Table A3 shows that this result is not robustly statistically significant under alternative plausible definitions of food deserts, although the point estimates for columns 3 and 4 tend to be positive.

The standard errors in column 3 bound effects on produce calorie share at around 0.006, or 6 calories per thousand. This is 12 percent of the mean, or about 1/2 the difference between the lowest- and highest-income households presented in Figure 5. The effects on Health Index in column 4 can be bounded at around 0.09. This represents just under 1/4 of the difference between the lowest- and highest-income households in Figure 5.

Table 5: **Effects of Retailer Entry: Known Retailers**

Panel A: Effects of Retailer Entry on Expenditure Shares

	(1)	(2)	(3)	(4)
Sample:	All Zip Codes		"Food Desert" Zip Codes	
Expenditure shares at store type:	Entrants	Chain Grocers/ Supers/Clubs	Entrants	Chain Grocers/ Supers/Clubs
Post entry: 5-10 miles	0.00662	0.00544	0.0416	0.0135
	(0.00184)***	(0.00183)***	(0.00643)***	(0.00597)**
Post entry: 0-5 miles	0.0366	0.0126	0.0687	0.0297
	(0.00238)***	(0.00203)***	(0.0103)***	(0.00752)***
Observations	1,826,027	1,826,027	345,031	345,031
Dependent var. mean	0.19	0.81	0.26	0.79

Panel B: Effects of Retailer Entry on Healthful Eating

	(1)	(2)	(3)	(4)
Sample:	All Zip Codes		"Food Desert" Zip Codes	
Dependent variable:	Share Produce	Health Index	Share Produce	Health Index
Post entry: 5-10 miles	0.000180	0.0101	0.00147	0.0295
	(0.000259)	(0.00560)*	(0.000846)*	(0.0187)
Post entry: 0-5 miles	0.0000949	-0.00251	0.00233	0.0404
	(0.000305)	(0.00719)	(0.00175)	(0.0300)
Observations	1,826,034	1,826,016	345,032	345,028
Dependent var. mean	0.050	-2.60	0.048	-2.67

Notes: Data are at the household-by-quarter level. Share Produce is the share of calories from fresh, canned, dried, and frozen fruits and vegetables; Health Index is the average Health Index per 1000 calories. Reported independent variables are indicators for whether a specific retailer has entered within 0-5 or 5-10 miles of the household; regressions also include Census division-by-quarter of sample indicators and household fixed effects. Robust standard errors, clustered by household, in parentheses. *, **, ***: Statistically significant with 10, 5, and 1 percent confidence, respectively.

IV.B Entry by All Retailers

The regressions above include only a limited set of retailers, which could reduce both power and generalizability. Table 6 presents estimates of Equation (2) using store counts of all retailers from Zip Code Business Patterns. The structure is similar to that of Table 5: Panel A presents effects on expenditure shares, while Panel B presents effects on healthful eating. Estimates for all zip codes and for food deserts are on the left and right, respectively. Columns 1 and 2 confirm that the ZBP data contain meaningful information. Column 1 shows that conditional on household fixed effects, a larger count of large grocery stores and/or a smaller count of supercenters and club stores in the zip code are both strongly positively associated with higher expenditure share at chain groceries. Column 2 shows the opposite: fewer grocery stores and more supercenters are strongly positively associated with higher expenditures at supercenters and club stores. As in Table 5, point estimates are larger in food deserts. Columns 3 and 6 show effects on combined expenditure shares for all supergrocers. As in Table 5, the coefficients are attenuated, meaning that entry by a large grocery retailer substantially diverts sales from other supergrocers. Consistent with the known retailer entry results, Appendix Table A4 shows that most of this diversion is from other mass merchants and non-chain grocers, and there is no statistically significant diversion from drug and convenience stores, although the point estimate is negative.

The bottom panel shows no statistically significant effect of the number of large grocers and supercenters/clubs on healthful eating. The standard errors are tighter than in Table 5. Even for the subset of households in "food desert" zip codes, we can bound the effect of entry at about 2 calories of produce per 1000 calories, or about 15 percent of the difference between the lowest- vs. highest-income households. Similarly, we can bound the effects on mean Health Index per 1000 calories at about 0.04, or about 10 percent of the difference between the lowest- vs. highest-income households. Appendix Table A4 shows that under one (but not a second) alternative definition of "food deserts," entry by supercenters and club stores does positively affect produce consumption and average Health Index, but the results are consistent in that the effect sizes can be bounded at only a fraction of the difference between the lowest- vs. highest-income households.

One reason to prefer the earlier regressions with specific known retailers is that we have high confidence that entry dates are correctly measured. In Appendix Table A1, we showed that ZBP data correctly date supercenter entry 50-80 percent of the time. We can also imagine using the true supercenter entry dates as an instrument for ZBP data, which are measured with error. Appendix Table A1 shows that the "first stages" of such a regression have coefficients around 0.9 and 0.66 for two different supercenter chains. If the average retailer in ZBP is measured with equal or perhaps somewhat more error than the less well-measured supercenter chain, this suggests that our bounds in the paragraph above should be increased by 50 to 100 percent due to measurement error. Even after this adjustment, however, our results in Tables 5 and 6 suggest that differences in supply, as measured by existence of a large grocery retailer, explain *at most* only a small share of the differences in nutritional decisions between low- and high-income households. In the next sections, we formalize a demand model to explore these issues further.

Table 6: **Effects of Retailer Entry: All Retailers**

Panel A: Effects of Retailer Entry on Expenditure Shares

	(1)	(2)	(3)	(4)	(5)	(6)
Sample:	All Zip Codes			"Food Desert" Zip Codes		
Expenditure shares at store type:	Chain Grocers	Supercenters/ Club Stores	Chain Grocers/ Supers/Clubs	Chain Grocers	Supercenters/ Club Stores	Chain Grocers/ Supers/Clubs
Large grocers	0.00583	-0.00545	0.000380	0.0102	-0.00852	0.00168
	(0.00104)***	(0.000817)***	(0.000846)	(0.00604)*	(0.00498)*	(0.00511)
Supercenters/clubs	-0.0142	0.0241	0.00996	-0.0544	0.0714	0.0170
	(0.00228)***	(0.00216)***	(0.00172)***	(0.0142)***	(0.0157)***	(0.00958)*
Observations	483,015	483,015	483,015	88,823	88,823	88,823
Dependent var. mean	0.56	0.24	0.81	0.49	0.30	0.79

Panel B: Effects of Retailer Entry on Healthful Eating

	(1)	(2)	(3)	(4)
Sample:	All Zip Codes		"Food Desert" Zip Codes	
Dependent variable:	Share Produce	Health Index	Share Produce	Health Index
Large grocers	-0.0000423	-0.00136	0.000705	0.00367
	(0.000105)	(0.00250)	(0.000703)	(0.0138)
Supercenters/clubs	-0.000342	-0.00152	-0.0000643	0.00657
	(0.000243)	(0.00541)	(0.00106)	(0.0227)
Observations	483,016	483,015	88,824	88,824
Dependent var. mean	0.050	-2.60	0.048	-2.67

Notes: Data are at the household-by-year level. Share Produce is the share of calories from fresh, canned, dried, and frozen fruits and vegetables; Health Index is the average Health Index per 1000 calories. Reported independent variables are the count of stores by channel type in the household's zip code; regressions also include Census division-by-year indicators and household fixed effects. Robust standard errors, clustered by household, in parentheses. *, **, ***: Statistically significant with 10, 5, and 1 percent confidence, respectively.

V Demand Model

We build on the model developed by Dubois et al (2013) to model households' demand for nutrients and groceries. Household i purchases groceries each week t by making shopping trips to local stores. Let S_{it} be the set of stores household i is aware of when making their shopping decisions at time t. We assume household i is aware of the prices of products sold within S_{it} in week t. The household elects which stores to shop at and purchases products in the stores to maximize utility. The utility function is:

$$\max_{s_{it} \epsilon S_{it}, x_{it}, y_{it}} U_{it}\left(x_{it}, z_{it}, y_{it}\right) - \alpha_{it} d_i\left(s_{it}\right)$$

$$s.t. \sum_{n=1}^{N} y_{int} p_{int} + p_{0t} x_{it} \leq I_{it}.$$

s_{it} is the set of stores within set S_{it} that the household elected to shop at in week t. $d\left(s_i\right)$ measures the total drive time needed to travel to all the stores. We will allow preferences to vary by household i and time t. $U_{it}\left(x_{it}, z_{it}, y_{it}\right)$ is household i's utility from products purchased in week t and α_{it} is household i's aversion to spending time driving in week t. Within the chosen stores, household i purchases groceries leading to an overall shopping bundle for the week. $y_{it} = \{y_{i1t}, y_{i2t}, ..., y_{iN_{it}t}\}$ is an N_{it}x1 vector of all the quantities (measured in calories) purchased by household i for every food product (UPC) stocked in stores s_{it}.[17] $p_{it} = \{p_{i1t}, p_{i2t}, ...p_{iN_{it}t}\}$ is an N_{it}x1 vector of prices (per calorie) for each product $n \epsilon N_{it}$ in week t. Each product n is characterized by C nutrient characteristics $\{a_{n1}, ..., a_{nC}\}$. Define the N_{it}xC matrix A=$\left\{\begin{array}{c} a_{11}, ..., a_{1C} \\ ... \\ a_{n1}, ..., a_{nC} \end{array}\right\}$. A measures the nutrient content (in kilograms) per calorie for each of the C nutrients in each of the N_{it} different products. Thus, the Cx1 be vector containing the total nutrient content of household i's bundle of goods, z_{it},can be written as: $z_{it} = A' y_{it}$.

Assuming quantities are continuous, the FOC for each product is:

$$\sum_{c=1}^{C} a_{nc} \frac{\partial U_{it}}{\partial z_{itc}} - \frac{\partial U_{it}}{\partial x_{it}} \frac{p_{int}}{p_{0t}} + \frac{\partial U_{it}}{\partial y_{int}} = 0. \tag{3}$$

We assume the utility function takes the form:

$$U_{it}\left(x_{it}, z_{it}, y_{it}\right) = \sum_{j=1}^{J} \mu_{ijt} log\left(\sum_{k=1}^{K_j} f_{ikjt}\left(y_{ikjt}\right)\right) + \sum_{c=1}^{C} h_{ic}\left(z_{ict}\right) + \gamma_i x_{it}.$$

[17] N_{it} is the number of products stocked in stores s_{it}.

Products have been grouped into J product categories, with K_j products within each group. An individual product is indexed by kj. $f_{ikjt}\left(y_{ikjt}\right)$ measures the sub-utility household i gets from product kj in week t. μ_{ijt} measures how much household i enjoyed product group j in week t for reasons other than nutrient content. $h_c\left(z_{ict}\right)$ measures the utility value household i gets from z_{ict} kilograms of nutrient c. γ_i measures the marginal utility household i gets from consuming the outside good (all other non-grocery consumption).

This setup allows household i to get value out of product jk through two mechanisms. First, household i gets direct value from consuming products in group j, as represented by $\mu_{ijt} log\left(\sum_{k=1}^{K_j} f_{ikjt}\left(y_{ikjt}\right)\right)$. This term represents that households value food consumption and variety in food consumption for reasons other than products' direct nutrient levels. Further, this terms allows for diminishing marginal utility from consuming large amounts of any one product or any one product group. Second, the household values the nutrient levels of the total bundle of groceries, as measured by $\sum_{c=1}^{C} h_{ic}\left(z_{ict}\right)$. This terms combines the total amount of nutrients across all the products consumed, allowing for direct preferences over nutrients, such as fat, sugar, or protein, to not depend on the exact product source of the nutrient. The combination of these two mechanisms allows households to diversity their food consumption across many types of products and product groups, but still adjust their consumption within and across product groups to account for the fact that the nutrient content of the food impacts the desirability of its consumption.

Plugging in these functional forms to equation (3), multiplying both sides by y_{ijkt}, and normalizing the price of the outside good to one gives:

$$\mu_{ijt}\frac{f'_{kjt}\left(y_{ikjt}\right)y_{ikjt}}{\sum_{l=1}^{Lj} f_{ljt}\left(y_{ikjt}\right)}+\sum_{c=1}^{C} a_{kjc}y_{ijkt}h'_c\left(z_{ict}\right)=\gamma_i p_{kjt}y_{ikjt}. \tag{4}$$

We allow the within product group subutility functions to be flexible CES functions of the form:$f_{ikjt}\left(y_{ikjt}\right)=\lambda_{ikjt}y_{ijkt}^{\theta_{ijt}}$. θ_{ijt} determines the extent to which household i's utility from individual products diminishes with calories purchased for products within product group j in week t. λ_{ikjt} allows for household i have a different taste for each individual product k within product group j in week t. This allows for very flexible heterogeneity across individual products and product groups for reasons other than nutrient value.

The nutrient utility functions are parameterized as: $h_c\left(z_{ict}\right)=\beta_{ic}z_{ict}$. Plugging these into equation (4), and summing equation (4) across all products within product group j and summing over all weeks t in year T yields:

$$\sum_{t\epsilon T}\frac{\mu_{ijt}\theta_{ijt}}{\gamma_i}+\sum_{c=1}^{C}\frac{\beta_{ic}}{\gamma_i}\sum_{k=1}^{K_j}\sum_{t\epsilon T} a_{kjc}y_{ijkt}=\sum_{t\epsilon T}\sum_{k=1}^{K_j} p_{kjt}y_{ikjt}. \tag{5}$$

$\sum_{t\epsilon T}\sum_{k=1}^{K_j} p_{kjt}y_{ikjt}$ measures household i's annual expenditure on product group j. $\sum_{t\epsilon T}\frac{\mu_{ijt}\theta_{ijt}}{\gamma_i}$

captures how much household i would consume of product group j, absent the nutritional value of products in product group j. μ_{ijt} measures household i's taste for product group j in week t. household i's $\sum_{c=1}^{C} \frac{\beta_{ic}}{\gamma_i} \sum_{k=1}^{K_j} \sum_{t\epsilon T} a_{kjc} y_{ijkt}$ captures how much additional expenditure will be allocated to product group j based on the desirability of its nutrient contents.

It will be useful for estimation to rewrite equation 5 by solving for total calories purchased by household i in product group j in year T. Define total calories purchased by household i in product group j in year T as: $Y_{ijT} = \sum_{t\epsilon T} \sum_{k\epsilon J} y_{ikjt}$. equation 5 can now be written as:

$$\sum_{t\epsilon T} \frac{\mu_{ijt}\theta_{ijt}}{\gamma_i} + \sum_{c=1}^{C} \frac{\beta_{ic}}{\gamma_i} \tilde{a}_{iTjc} Y_{ijT} = \tilde{p}_{ijT} Y_{ijT}., \tag{6}$$

where $\tilde{p}_{ijT}$is the calorie weighted average price paid per calorie by household i in product group j in year T. Similarly, $\tilde{a}_{iTjc}$is the calorie weighted average amount of nutrient c per calorie in products purchased by household i in product group j in year T. Equation 6 can now be solved for total calories, Y_{ijT} :

$$Y_{ijT} = \frac{\sum_{t\epsilon T} \frac{\mu_{ijt}\theta_{ijt}}{\gamma_i}}{\tilde{p}_{ijT} - \sum_{c=1}^{C} \frac{\beta_{ic}}{\gamma_i} \tilde{a}_{iTjc}}. \tag{7}$$

Finally, taking logs of both sides yields:

$$log(Y_{ijT}) = log(\sum_{t\epsilon T} \frac{\mu_{ijt}\theta_{ijt}}{\gamma_i}) + log(\tilde{p}_{ijT} - \sum_{c=1}^{C} \frac{\beta_{ic}}{\gamma_i} \tilde{a}_{iTjc}). \tag{8}$$

A desirable property of this model is that it admits substantial flexibility and heterogeneity in preferences at the UPC level, as measured by $f_{ikjt}\left(y_{ikjt}\right) = \lambda_{ikjt} y_{ijkt}^{\theta_{ijt}}$. λ_{ikjt} allows each household to have heterogeneous preferences for each UPC in each week. The model also admits preference heterogeneity in product group preferences (including the outside good) over time and across households as measured by $\frac{\mu_{ijt}\theta_{ijt}}{\gamma_i}$, while at the same time capturing the utility value of nutrients, as measured by $\sum_{c=1}^{C} \frac{\beta_{ic}}{\gamma_i} \tilde{a}_{iTjc}$. One of the most desirable properties of the model is that is allows for estimation of nutrient and product group preferences from data has been aggregated to the annual product group level. This allows us to focus on estimation of preferences for nutrients without dealing with the parameters driving preferences at the UPC level or the dynamics in purchase behavior at the weekly level.

VI Estimation

Equation 8 will be the estimating equation to identify households' preferences for product groups and nutrients. All estimation will be done on data aggregated to annual product group consumption measures, as shown in equation 7.[18] We will estimate heterogeneous preferences across households of different income types. Household i's type is denoted $b\left(i\right)$. We now define the household i's preference for product group j in year T by the mean preference across years and households of type $b\left(i\right)$, and its deviation from this mean:

$$log(\sum_{t\epsilon T}\frac{\mu_{ijt}\theta_{ijt}}{\gamma_i})=\bar{\delta}_{b(i)j}+\widetilde{\delta_{ijT}}. \tag{9}$$

We also allow for one of the product nutrients to be unobserved to the econometrician that varies across household types. This is to capture unobserved overall quality differences of products bought by different types of households. Let c_u denote the unobserved product nutrient and C_o denote the set of observable product nutrients. Preferences for nutrients and the outside good will vary by household type $b\left(i\right)$ for estimation.

$$\sum_{c=1}^{C}\frac{\beta_{b(i)c}}{\gamma_{b(i)t}}\tilde{a}_{iTjc}=\sum_{c=1}^{C_o}\tilde{\beta}_{b(i)c}\tilde{a}_{iTjc}+\tilde{\alpha}_{b(i)c_u} \tag{10}$$

$$\tilde{\beta}_{b(i)c}=\frac{\beta_{b(i)c}}{\gamma_{b(i)}},$$

$$\tilde{\alpha}_{b(i)c_u}=\frac{\beta_{b(i)c_u}}{\gamma_{b(i)}}\tilde{a}_{c_u}. \tag{11}$$

Plugging equations (9) and 10 into equation 8 gives:lo

$$log(Y_{ijT})=\bar{\delta}_{b(i)j}+\widetilde{\delta_{ijT}}+log(\tilde{p}_{ijT}-\sum_{c=1}^{C_o}\tilde{\beta}_{b(i)c}\tilde{a}_{iTjc}+\tilde{\alpha}_{b(i)c_u}). \tag{12}$$

Equation (12) will be our main estimating equation. It relates total annual calories purchased in product group j by household i in year T (Y_{ijT}), to a fixed effect for each product group and household type $\left(\bar{\delta}_{b(i)j}\right)$, the average kilograms of each nutrient c per calorie in the bundle of products consumed by household i in product group j ($\tilde{a}_{iTjc}$), the value of unobserved nutrient $\left(\tilde{\alpha}_{b(i)c_u}\right)$, and a residual $\left(\widetilde{\delta_{ijT}}.\right)$[19] The product group by household type fixed effects capture the utility different types of household gain from each product group if all product groups offered zero nutritional content. The coefficients on the nutrient contents of the bundle of products $\left(\tilde{\beta}_{b(i)c}\right)$ capture how much households of type $b\left(i\right)$value the nutrient c. The residual represents household i's idio preference for product group j in year T.

To estimate households' preferences for nutrients and product groups we use non-linear least

[18] We will now refer to an annual time periods as T, weeks are still indexed by t.

[19] Note that we can not disentangle the preference for the unobserved nutrient from the total quantity of the unobserved nutrient. We estimate the combined effect of the value of the chosen quantity of the unobserved nutrient for each household type.

squares. The key identifying assumption is that individual households idiosyncratic preferences for product groups are uncorrelated with their preferences for nutrients. This identification assumption could potentially be violated, for example, if households purchased salty foods because they prefer product groups that have longer shelf lives. We will not be able to separate out how much of a preference for salty foods is due to a true nutrition value of sodium versus other impacts of sodium on the product. Essentially, we will be identifying the value of nutrients from a nutrition standpoint along with all the other impacts nutrients have on the desirability of food (taste, shelf life, ease of preparation.) It is hard to think that supply-side instrumental variables could solve this problem as it would be challenging to create variation only in the nutrition value of a product, without impacting it taste or shelf life.

Intuitively, the model will identify a strong preference for a given nutrient if variation in the intensity of that nutrient across product groups is strongly correlated with calories consumed within product groups across households or over time. Preference for nutrients are distinguished from a direct preference for the product group by whether the calories consumed in a given product group remains fixed, regardless of variation in nutrient content.

The main model estimates will explore preference heterogeneity based on household income. We bucket households into types based on whether their household income is below \$25,000, between \$25,000 and \$50,000, between \$50,000 and \$75,000, and above \$75,000.

VII Demand Estimation Results

We report the model estimates below. Table (7)reports the average product fixed effects for seven different categories of food. The coefficients have been normalized to represent the expenditure each type of household would choose for each category of food if all foods contained no nutrients. These preference represent the relative desirability of different foods for reasons other than nutrition.[20]

[20]Specifically, we take the estimated fixed effects for each product group within each income group $(\bar{\delta}_{b(i)j})$and sum them across product groups within the more aggregate product category listed in Table (7). We then divide this sum by the sum of all the product group fixed effects. For example, the reported expenditure share on Dry Grocery for Inc$\leq 25,000$: $\frac{\sum_{j\epsilon DryGrocery}\bar{\delta}_{Inc\leq 25K,j}}{\sum_{l=1}^{64}\bar{\delta}_{Inc\leq 25K,l}} = 0.5592$.

Table 7: **Average Preferences for Product Categories by Household Income**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Category:	Dry Grocery	Prepared Foods	Frozen	Meat	Dairy	Produce	Alcohol
Inc$\leq 25K$	0.5592	0.0943	0.0779	0.0438	0.1324	0.0360	0.0564
	(0.00002)	(0.000007)	(0.00001)	(0.000004)	(0.00001)	(0.000004)	(0.000008)
$25K < Inc \leq 50K$	0.5562	0.0964	0.0759	0.0445	0.1332	0.0396	0.0542
	(0.00001)	(0.000004)	(0.000006)	(0.000002)	(0.000008)	(0.000002)	(0.000004)
$50K < Inc \leq 75K$	0.5523	0.0979	0.0747	0.0439	0.1353	0.0426	0.0532
	(0.00005)	(0.000005)	(0.000006)	(0.000002)	(0.000009)	(0.000002)	(0.000004)
$75K < Inc$	0.5449	0.0985	0.0729	0.0400	0.1367	0.0485	0.0586
	(0.00001)	(0.000004)	(0.000004)	(0.000002)	(0.000007)	(0.000002)	(0.000003)
# of Product Groups	36	5	8	2	10	1	3

Notes: Parameter estimates are normalized to sum to one for each income group. The magnitudes represent households' expenditure shares on each product category if all products had no nutritional content. The model estimates a separate parameter for each calorie shares for each of the 66 product groups. This table aggregates these estimates into six broad categories. Standard errors cluster by household and use the delta method.

Table (7) shows that even without differences in nutritional content across product groups, there are differences in the types of products consumed across the income distribution. The lowest income group spends 6.9% more on frozen foods and 2.5% more on dry groceries.[21] The highest income group spends 35% more on produce, 3% more on dairy, and 3.4% more on prepared foods than the lowest income group. The higher income households appear to prefer products with shorter shelf life (more dairy, less frozen and dry grocery). Meat and alcohol budget shares do not vary much across the income distribution. To the extent that dry groceries and frozen goods are relatively less healthy than produce, dairy, and prepared foods, these preferences for products groups show that some of the nutritional differences between the low and high income households need not be due to a direct difference in preference for nutrients. Rather, these differences could be partially due to a preference for product groups characteristics, such as shelf life, which is correlated with nutrient content. Section VIII below will perform a formal calculation to measure what these product group preference differences drive nutritional outcomes.

Table 8 reports the estimated preferences for product nutrients across the four income groups. These have been normalized such that units represent how many kilograms of the nutrient offers the same utility value as a kilogram of fat.[22] This normalization removes differences in the marginal utility of a dollar ($\gamma_{b(i)}$) across the income groups.

Preferences for carbs, fiber, sodium, cholesterol, and fruit are linked with household income. However, the higher income groups do not always exhibit more desire for healthy nutrients. Higher income groups have a stronger desire for fiber and fruit, leading them to make more healthy choices. Carbs are twice as desirable to the highest income group than to the lowest (relative to fat). Fiber is undesirable to all households, however the higher income groups are willing to tolerate it more. Fruit 400% more desirable to the highest income group than the lowest.[23]

However, the higher income groups also have a stronger desire of cholesterol and sodium, leading them to make less healthy choices on this dimension. Cholesterol and sodium are both 23% more desirable to the highest income group than to the lowest. There does not appear to be a significant income link to households' preferences over the other nutrients that enter the health index (saturated fat, protein and vegetables). Overall, it does not seem that higher income households have more desire for healthy nutrients.

[21] These percentages are calculated by taking the difference in expenditure shares between the bottom income group and the top and then dividing this by the expenditure share of the top income group.

[22] Appendix Table A5 reports the raw point estimates for nutrient preferences measured in dollars. These estimates clearly show a declining marginal utility of a dollar across the income groups.

[23] These estimates for the value of fruit measures the nutrient desirability of fruit other than through our standard observed macronutrients. This reflects the value of the vitamins found in fruit.

Table 8: Preferences for Nutrients by Household Income

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Income	Carbs	Fat	Saturated Fat	Fiber	Protein	Sugar	Sodium	Cholesterol	Fruit	Vegetables
Inc$\leq 25K$	0.095	1.000	-0.200	-0.587	0.329	0.256	-0.710	-36.348	0.004	-0.066
	(0.0107)		(0.0158)	(0.0222)	(0.0131)	(0.0085)	(0.0271)	(1.746)	(0.0025)	(0.0027)
$25K < Inc \leq 50K$	0.104	1.000	-0.222	-0.578	0.283	0.242	-0.753	-29.272	0.007	-0.063
	(0.0074)		(0.0109)	(0.0155)	(0.0086)	(0.0058)	(0.0206)	(1.123)	(0.0018)	(0.0019)
$50K < Inc \leq 75K$	0.133	1.000	-0.240	-0.510	0.270	0.215	-0.712	-25.485	0.010	-0.064
	(0.0077)		(0.0120)	(0.0164)	(0.0097)	(0.0060)	(0.0217)	(1.175)	(0.0023)	(0.0022)
$75K < Inc$	0.190	1.000	-0.178	-0.431	0.308	0.157	-0.575	-27.895	0.012	-0.069
	(0.0041)		(0.0075)	(0.0101)	(0.0067)	(0.0031)	(0.0126)	(0.8425)	(0.0013)	(0.0015)

Notes: Standard errors are clustered by household. Magnitudes represent kilograms of the nutrient which offers the same utility as a kilogram of non-saturated fat. A given nutrient can enter through multiple preference parameters, such as saturated fat is both valued as fat and saturated fat. Fiber and sugar are also carbohydrates. Value of fruit and vegetables accounts for value over and beyond macronutient characteristics of the fruit and vegetables.

VIII Decomposing Nutrition Choices

Using the model estimates above, we decompose how much of the observed nutrition differences observed across the income groups are due to supply forces (prices and product nutrient supply) versus demand forces (product group, nutrient, and UPC preferences). For a given set of prices for each product group ($\tilde{p}_j$), nutrients for each product group j $(\tilde{a}_{jc_1}, ..., \tilde{a}_{jC_o}, \tilde{\alpha}_{c_u})$, product group preferences $(\bar{\delta}_1, .., \bar{\delta}_J)$, and nutrient preferences $(\tilde{\beta}_1, ..., \tilde{\beta}_{C_o})$, we can calculate how many calories would be consumed within each product group $(\hat{Y}_j)$:

$$\hat{Y}_j = \frac{\bar{\delta}_j}{\tilde{p}_j - \sum_{c=1}^{C_o} \tilde{\beta}_c \tilde{a}_{jc} - \tilde{\alpha}_{c_u}}. \tag{13}$$

To evaluate the healthfulness of this bundle of goods, we calculate the health index, as discussed in Section II.C, where w_c are the health index weights on each nutrient:

$$\hat{H} = \sum_j \sum_c w_c \tilde{a}_{jc} \hat{Y}_j. \tag{14}$$

First, we explore the role of preferences at the UPC level. We restrict households' choices within product groups to a representative product for each product group for each income category of household. We take all the stores visited by households within income group b and use RMS to take an equally weighted average across all products stocked within this set of stores within product group j:

$$\hat{a}_{bjc} = \frac{1}{N_b} \sum_{s \epsilon S_b} \sum_{k \epsilon s_j} a_{kjc}, \tag{15}$$

where S_b is the set of all stores visited by households in income category b and s_j is the set of all products stocked within product group j in store s. Thus $\hat{a}_{bjc}$ equals the average nutrient content of nutrient c from product group j if households bought an equal amount of all products stocked from product group j from the set of stores shopped at by income category b.[24] Using the nutrient contents of these representative products, we then use our estimated to demand model to calculate what household would have purchased if they were only able to optimize across product groups, restricting them to buy this equally weighted average of all products supplied, within product groups. Each household's demand is:

$$\hat{Y}_{b(i)j} = \frac{\bar{\delta}_{b(i)j}}{\hat{p}_{b(i)j} - \sum_{c=1}^{C_o} \tilde{\beta}_{b(i)c} \hat{a}_{bjc} - \tilde{\alpha}_{b(i)c_u}}. \tag{16}$$

We then then calculate the health index for these new bundles of purchases.

Figure 7 shows that preventing households from targeting their purchases within products

[24] We weight the stores based on the number of trips households made to them.

groups at specific UPCs drastically closes the nutrition-income gradient. Higher income groups' purchases become significantly less healthful, while lower income groups become slightly more healthful simply by buying different products stocked within the stores these households are already shopping in.

Next, we directly investigate the role of product supply. We consider what households would buy if they faced the choice set we observe in the highest income group's set of stores. We use our representative products calculated above, but allow all households to choose from the set of products observed in the highest income group's (income group 4) stores:

$$\hat{Y}_{b(i)j} = \frac{\bar{\delta}_{b(i)j}}{\hat{p}_{b(i)j} - \sum_{c=1}^{C_o} \tilde{\beta}_{b(i)c} \hat{a}_{4jc} - \tilde{\alpha}_{b(i)c_u}}. \tag{17}$$

Figure 7 shows that this further closes the nutrition-income gradient, but by a much more modest amount.

We now add on the prices faced by households in income group 4 to assess whether prices play a role in the income gradient:

$$\hat{Y}_{b(i)j} = \frac{\bar{\delta}_{b(i)j}}{\hat{p}_{4j} - \sum_{c=1}^{C_o} \tilde{\beta}_{b(i)c} \hat{a}_{4jc} - \tilde{\alpha}_{b(i)c_u}}. \tag{18}$$

Figure 7 shows that price play essentially no role in the nutrition-income gradient.

Returning to demand, we now set product group preferences to those of the highest income, along with the prices and product nutrients observed in the top income group:

$$\hat{Y}_{b(i)j} = \frac{\bar{\delta}_{4j}}{\hat{p}_{4j} - \sum_{c=1}^{C_o} \tilde{\beta}_{b(i)c} \hat{a}_{4jc} - \tilde{\alpha}_{b(i)c_u}}. \tag{19}$$

Figure 7 shows that this more than 100% closes the nutrition-income gap completely. The bundles of food chosen by the lower income groups even are slightly healthier than those chosen the top income group.

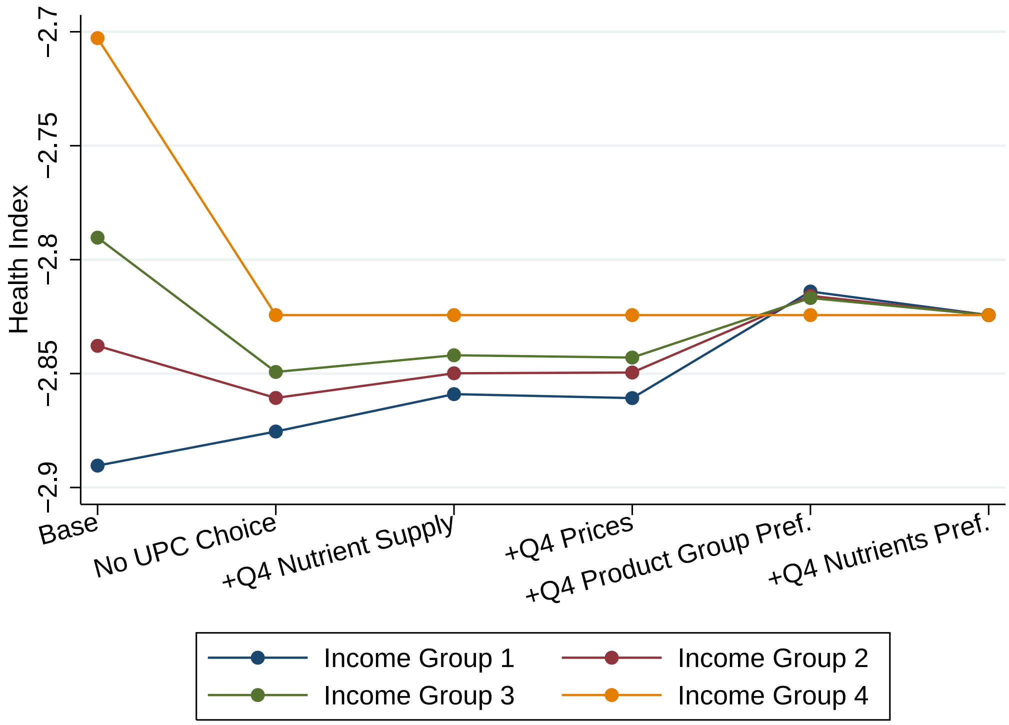
Finally, we now set the nutrient preferences to those of the highest income group, along with prices, product group preferences, and product nutrient levels:

$$\hat{Y}_{b(i)j} = \frac{\bar{\delta}_{4j}}{\hat{p}_{4j} - \sum_{c=1}^{C_o} \tilde{\beta}_{4c} \hat{a}_{4jc} - \tilde{\alpha}_{4c_u}}. \tag{20}$$

This mechanically closes the nutrition income gradient, as all households make identical purchases. The additional effect of setting the lower income groups' nutrient preferences to those of the highest income group actually make these lower income groups consume slightly less healthful food.

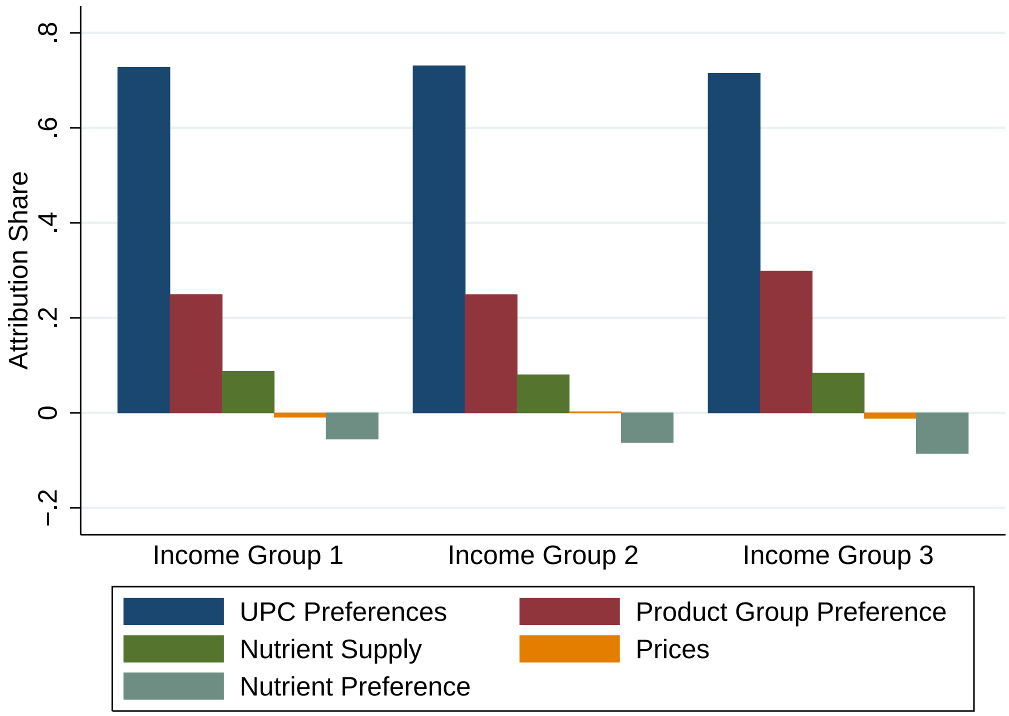
Figure VIII summarizes these effects into how how much of the healthy eating gap can be attributed to each of these supply and demand forces. Overall, individual UPC preferences explain 72%, product group preferences explain 25%, nutrient preferences explain -5% and product nutrient supply differences explain 8%. Demand differences are the overwhelming driver of the nutrition-income gradient.

Figure 7: **Predicted Health Index for Each Income Group**



Notes:Each category on the x-axis represented a separate counterfactual calculation. The base catergory measures the health index for each income group when each group retains their true preferences and face their own local supply conditions. The second category sets all prices to those observed in income group 4. The third category sets all prices and product nutrient characteristics to those in group 4. The fourth and fifth categories add on the product group and nutrient preferences, respectively.

Figure 8: **Percent of Health Index Differences From Supply and Demand Factors**



Notes: Units measure the percent of the health index difference between each income group and the richest income group that is caused be each factor: product group prices, product group nutrients, product group preferences, and nutrient preferences.

IX Conclusion

In this paper, we studied how and why healthful eating varies by income in the United States. First, we use the Nielsen Homescan and retail scanner data to document that the lowest-income households consume groceries that average 0.57 standard deviations less healthful than the highest-income households. Little of this difference can be explained by differences in local supply of healthful foods: in a reduced-form event study framework, entry of additional large grocery stores and supercenters reduces travel costs but does not affect average healthfulness, even for households in "food deserts." We then estimate a formal demand model and use estimated preferences to decompose differences in purchasing patterns between low- and high-income households. We find that 92% of the nutrition-income gradient is driven by differences in demand across products, while only 8% can be attributed to differences in supply. The demand impact can be broken down into 72% due to differences in UPC demand, 25% due to product group demand differences, and -5% due to nutrient preferences. The 8% supply effect is entirely driven by different product offerings, with prices playing essentially no role.

References

[1] Anderson, Michael, and David Matsa. 2011. "Are Restaurants Really Supersizing America?" *American Economic Journal: Applied Economics*, 3 (1): 152-188.

[2] Bitler, Marianne, and Steven Haider (2011). "An Economic View of Food Deserts in the United States." *Journal of Policy Analysis and Management*, 30 (1): 153-176.

[3] Broda, Christian, Ephraim Leibtag, and David Weinstein. 2009. "The Role of Prices in Measuring the Poor's Living Standards." *Journal of Economic Perspectives*, 23 (2): 77-97.

[4] Bronnenberg, Bart, Jean-Pierre Dubé, Matthew Gentzkow, and Jesse Shapiro (2015). "Do Pharmacists Buy Bayer? Sophisticated Shoppers and the Brand Premium." *Quarterly Journal of Economics*, forthcoming.

[5] Bronnenberg, Bart, Jean-Pierre Dubé, and Matthew Gentzkow. 2012. "The Evolution of Brand Preferences: Evidence from Consumer Migration." *American Economic Review*, 102 (6): 2472-2508.

[6] Cawley, John, Chad Meyerhoefer, Adam Biener, Mette Hammer, and Neil Wintfeld. 2015. "Savings in Medical Expenditures Associated with Reductions in Body Mass Index Among US Adults with Obesity, by Diabetes Status." *PharmacoEconomics*, November.

[7] Courtemanche, Charles, and Art Carden. 2011. "Supersizing Supercenters? The Impact of Walmart Supercenters on Body Mass Index and Obesity." Journal of Urban Economics, 69 (2): 165-181.

[8] Cummins, Steven, Anne Findlay, Mark Petticrew, and Leigh Sparks. 2005. "Healthy Cities: The Impact of Food Retail-led Regeneration on Food Access, Choice, and Retail Structure." *Built Environment*, 31 (4): 288-301.

[9] Cummins, Steven, Ellen Flint, and Stephen Matthews. 2015. "New Neighborhood Grocery Store Increased Awareness of Food Access But Did Not Alter Dietary Habits or Obesity." *Health Affairs* 33, (2): 283-291.

[10] Currie, Janet, Stefano DellaVigna, Enrico Moretti, and Vikram Pathania. 2010. "The Effect of Fast Food Restaurants on Obesity and Weight Gain." *American Economic Journal: Economic Policy*, 2 (3): 32-63.

[11] Cutler, David, and Adriana Lleras-Muney. 2010. "Understanding Differences in Health Behaviors by Education." *Journal of Health Economics*, 29 (1): 1-28.

[12] Davis, Brennan, and Christopher Carpenter. 2009. "Proximity of Fast-Food Restaurants to Schools and Adolescent Obesity." *American Journal of Public Health*, 99 (3): 505–510.

[13] Dubois, Pierre, Rachel Griffith, and Aviv Nevo. 2014. "Do Prices and Attributes Explain International Differences in Food Purchases?" *American Economic Review*, 104 (3): 832-867.

[14] Dunn, Richard A. 2010. "The Effect of Fast-Food Availability on Obesity: An Analysis by Gender, Race, and Residential Location." *American Journal of Agricultural Economics*, 92 (4): 1149-1164.

[15] Einav, Liran, Ephraim Leibtag, and Aviv Nevo. 2010. "Recording Discrepancies in Nielsen Homescan Data: Are They Present and Do They Matter?" *Quantitative Marketing and Economics*, 8 (2): 207–239.

[16] Elbel, Brian, Alyssa Moran, L. Beth Dixon, Kamila Kiszko, Jonathan Cantor, Courtney Abrams, and Tod Mijanovich. 2015. "Assessment of a Government-Subsidized Supermarket in a High-Need Area on Household Food Availability and Children's Dietary Intakes." *Public Health Nutrition*, February 26: 1-10.

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

[18] Ellickson, Paul B. 2007. "Does Sutton Apply to Supermarkets?" *RAND Journal of Economics*, 38 (1): 43-59.

[19] Finkelstein, Eric A., Justin G. Trogdon, Joel W. Cohen, and William Dietz. 2009. "Annual Medical Spending Attributable to Obesity: Payer-And-Service-Specific Estimates." *Health Affairs*, 28 (5): w822-w831.

[20] Food Marketing Institute. 1998. Urban Supermarkets. Washington, DC: Food Marketing Institute.

[21] Handbury, Jessie, Ilya Rahkovsky, and Molly Schnell. 2015. "What Drives Nutritional Disparities? Retail Access and Food Purchases across the Socioeconomic Spectrum." National Bureau of Economic Research Working Paper 21126.

[22] HHS (U.S. Department of Health and Human Services). 2011. "$100 million in Affordable Care Act grants to help create healthier U.S. communities." http://www.epa.gov/care/documents/AffordableCareActGrants.pdf

[23] Hwang, Minha, and Sungho Park. 2015. "The Impact of Walmart Supercenter Conversion on Consumer Shopping Behavior." *Management Science*, forthcoming.

[24] Holmes, Thomas J. 2011. "The Diffusion of Wal-Mart and Economies of Density." *Econometrica*, 79 (1): 253-302.

[25] Kling, Jeffrey R. Jeffrey B. Liebman, and Lawrence F. Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica*, 75 (1): 83-119.

[26] Lakdawalla, Darius N., and Tomas J. Philipson. 2002. "Technological Change and the Growth of Obesity." National Bureau of Economic Research Working Paper 8946.

[27] Larson, Nicole, Mary Story, and Melissa Nelson. 2009. "Neighborhood Environments: Disparities in Access to Healthy Foods in the U.S." *American Journal of Preventive Medicine*, 36 (1): 74-81.

[28] Ludwig, Jens, Lisa Sanbonmatsu, Lisa Gennetian, Emma Adam, Greg J. Duncan, Lawrence F. Katz, Ronald C. Kessler, Jeffrey R. Kling, Stacy Tessler Lindau, Robert C. Whitaker, and Thomas W. McDade. 2011. "Neighborhoods, Obesity, and Diabetes - A Randomized Social Experiment." *New England Journal of Medicine*, 365: 1509-1519.

[29] NCHS (National Center for Health Statistics). 2013. "Health, United States, 2013. Healthy weight, overweight, and obesity among adults aged 20 and over, by selected characteristics: United States, selected years 1960–1962 through 2009–2012" http://www.cdc.gov/nchs/data/hus/hus13.pdf

[30] NCHS (National Center for Health Statistics). 2014. "Health, United States, 2014. Table 59 (page 1 of 2). Selected health conditions and risk factors, by age: United States, selected years 1988–1994 through 2011–2012." http://www.cdc.gov/nchs/data/hus/hus14.pdf

[31] Ogden, Cynthia L., Molly M. Lamb, Margaret D. Carroll, and Katherine M. Flegal. 2010. "Obesity and Socioeconomic Status in Adults: United States, 2005-2008. NCHS Data Brief No. 50. http://www.cdc.gov/nchs/data/databriefs/db50.pdf

[32] Philipson, Tomas J., and Richard A. Posner. 1999. "The Long-Run Growth in Obesity as a Function of Technological Change." National Bureau of Economic Research Working Paper 7423.

[33] Sharkey, Joseph R., Scott Horel, and Wesley R. Dean. 2010. "Neighborhood deprivation, vehicle ownership, and potential spatial access to a variety of fruits and vegetables in a large rural area in Texas." *International Journal of Health Geographics*, 9 (26): 1–27.

[34] TRF (The Reinvestment Fund). 2015a. "The Healthy Food Financing Initiative." http://healthyfoodaccess.org/sites/default/files/HFFI%20Fact%20Sheet_four%20pager-051315.pdf

[35] TRF (The Reinvestment Fund). 2015b. "Pennsylvania Fresh Food Financing Initiative." http://www.trfund.com/pennsylvania-fresh-food-financing-initiative/

[36] USDA (U.S. Department of Agriculture), Economic Research Service. 2009. "Access to Affordable and Nutritious Food: Measuring and Understanding Food Deserts and Their Consequences." http://www.ers.usda.gov/media/242675/ap036_1_.pdf

[37] USDA (U.S. Department of Agriculture), Agricultural Research Service. 2014. "Away from Home: Percentages of Selected Nutrients Contributed by Food and Beverages Consumed Away from Home, by Family Income (in Dollars) and Age." *What We Eat in America*, NHANES 2011-2012. http://www.ars.usda.gov/Services/docs.htm?docid=18349

[38] USDA (U.S. Department of Agriculture), Agricultural Research Service. 2014. "USDA National Nutrient Database for Standard Reference, Release 26." Nutrient Data Laboratory Home Page, http://www.ars.usda.gov/ba/bhnrc/ndl

[39] Volpe, Richard, Abigail Okrent, and Ephraim Leibtag. 2013. "The Effect of Supercenter-Format Stores on the Healthfulness of Consumers' Grocery Purchases." *American Journal of Agricultural Economics*, 95 (3): 568-589.

[40] Waldfogel, Joel. 2003. "Preference Externalities: An Empirical Study of Who Benefits Whom in Differentiated-Product Markets." *RAND Journal of Economics*, 34 (3): 557-568.

[41] Waldfogel, Joel. 2008. "The Median Voter and the Median Consumer: Local *Private* Goods and Population Composition." *Journal of Urban Economics*, 63 (2): 567-582.

[42] Wrigley, Neil, Daniel Warm and Barrie Margetts. 2003. "Deprivation, Diet, and Food Retail Access: Findings From the Leeds 'Food Deserts' Study." *Environment and Planning A*, 35: 151-188.

Online Appendix: Not for Publication

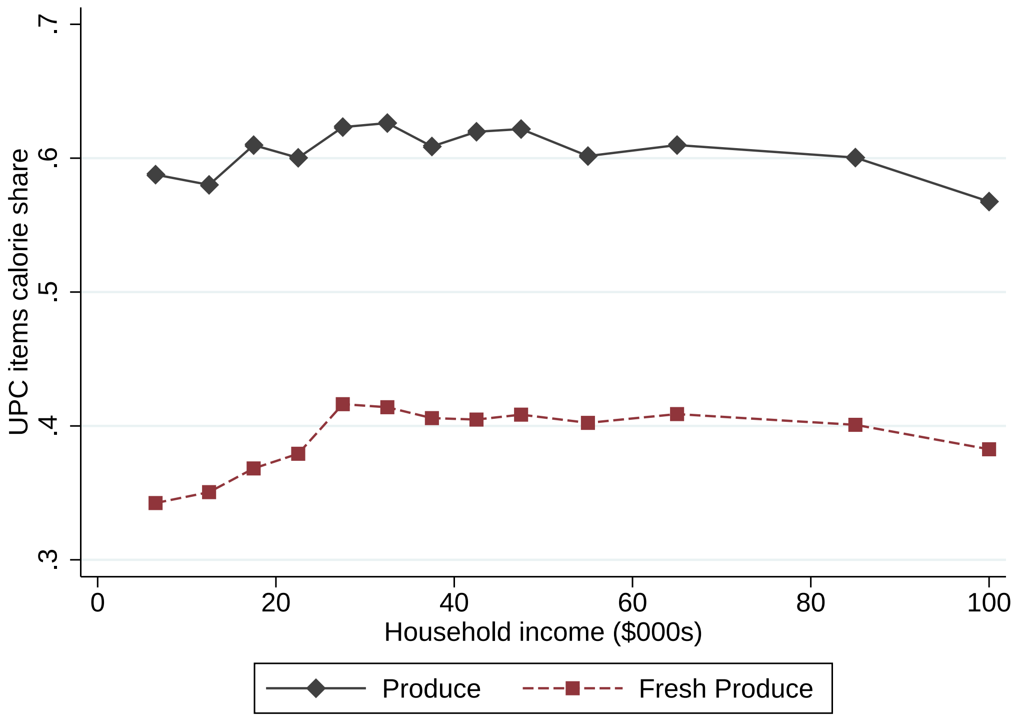
The Geography of Poverty and Nutrition: Food Deserts and Food Choices Across the United States

Hunt Allcott, Rebecca Diamond, and Jean-Pierre Dubé

A Appendix to Data Section

A.A Magnet Calorie Shares

Figure A1: **Calorie Shares from Produce Items with UPC Codes**



Notes: Uses Nielsen Homescan "Magnet" subsample for 2004-2006. "Produce" includes fresh, dried, canned, and frozen produce. Shows the share of produce and fresh produce calories coming from items with UPCs, which are the items we observe outside the Magnet subsample.

A.B Zip Code Business Patterns Accuracy Check

Table A1: **Zip Code Business Patterns Accuracy Check with Known Entry Dates**

	(1)	(2)
Dependent variable is count of channel type:	Supercenter	General Merchandise
Difference Estimator		
Supercenter Chain 1:		
2nd Lead	0.00478	0.00629
	(0.00622)	(0.0166)
1st Lead	0.0375	0.0577
	(0.00978)***	(0.0164)***
Entry Year	0.562	0.821
	(0.0199)***	(0.0210)***
1st Lag	0.208	0.0821
	(0.0169)***	(0.0191)***
2nd Lag	0.0777	0.0133
	(0.0127)***	(0.0145)
Supercenter Chain 2:		
2nd Lead	0.0158	-0.0172
	(0.0298)	(0.0421)
1st Lead	0.0133	-0.0451
	(0.0222)	(0.0462)
Entry Year	0.0621	0.480
	(0.0327)*	(0.0623)***
1st Lag	0.172	0.0701
	(0.0413)***	(0.0594)
2nd Lag	0.133	0.0918
	(0.0514)***	(0.0599)
Observations	264,734	264,734
Fixed Effects Estimator		
Post Entry: Chain 1	0.902	0.932
	(0.0138)***	(0.0227)***
Post Entry: Chain 2	0.667	0.665
	(0.0365)***	(0.0659)***
Observations	297,966	297,966

Notes: Data are at the zip code-by-year level. All regressions include year indicators; fixed effects regressions have zip code fixed effects. Robust standard errors, clustered by zip code, in parentheses. *, **, ***: Statistically significant with 10, 5, and 1 percent confidence, respectively.

A.C Health Index

Appendix Table A2 presents the DRIs r_k. For example, $H(\mathbf{x})$ would take value 1 for a UPC (say three cups of vegetables) that exactly satisfied the RDI of one "good" macronutrient, or -1 for a UPC that contained the maximum RDI of one "bad" macronutrient.

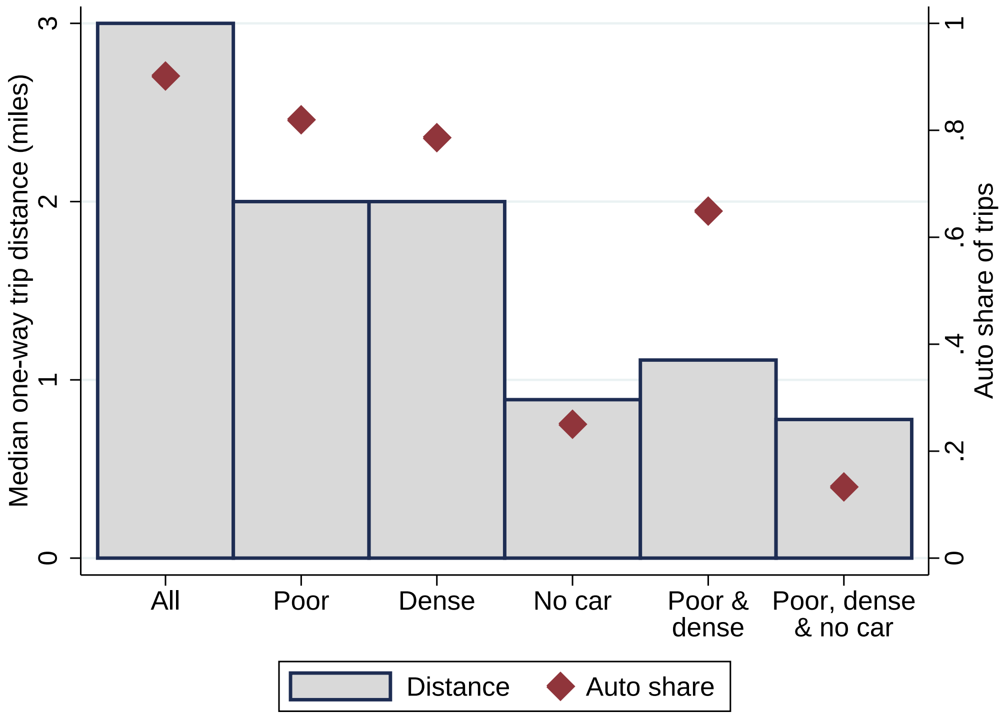
Table A2: **Health Index Function**

Attribute	Recommendation	Recommended Daily Intake (grams)	Explanation
		Fruits and vegetables	
Fruit	Increase	320	Two cups/day (Food Patterns); 160 g/cup
Vegetables	Increase	390	Three cups/day (Food Patterns); 130 g/cup
		All other items	
Protein	Increase	51	51 grams/day (DRI)
Fiber	Increase	29.5	29.5 grams/day (DRI)
Sugar	Reduce	32.8	45% of 282 calories/day from sugar+sat. fat (Food Patterns)
Saturated fat	Reduce	17.2	55% of 282 calories/day from sugar+sat. fat (Food Patterns)
Sodium	Reduce	2.3	2300 mg/day (Dietary Guidelines)
Cholesterol	Reduce	0.3	300 mg/day (Dietary Guidelines)

Notes: This table presents the Recommended Daily Intakes (RDI) for each attribute. The Health Index $H(\mathbf{x}) = \sum_k G_k \frac{g_k}{r_k} - (1 - G_k) \frac{g_k}{r_k}$, where g_k is the grams of macronutrient k, r_k is the RDI for a normal adult, and G_k takes value 1 for macronutrients recommended to "Increase" and 0 for macronutrients recommended to "Reduce."

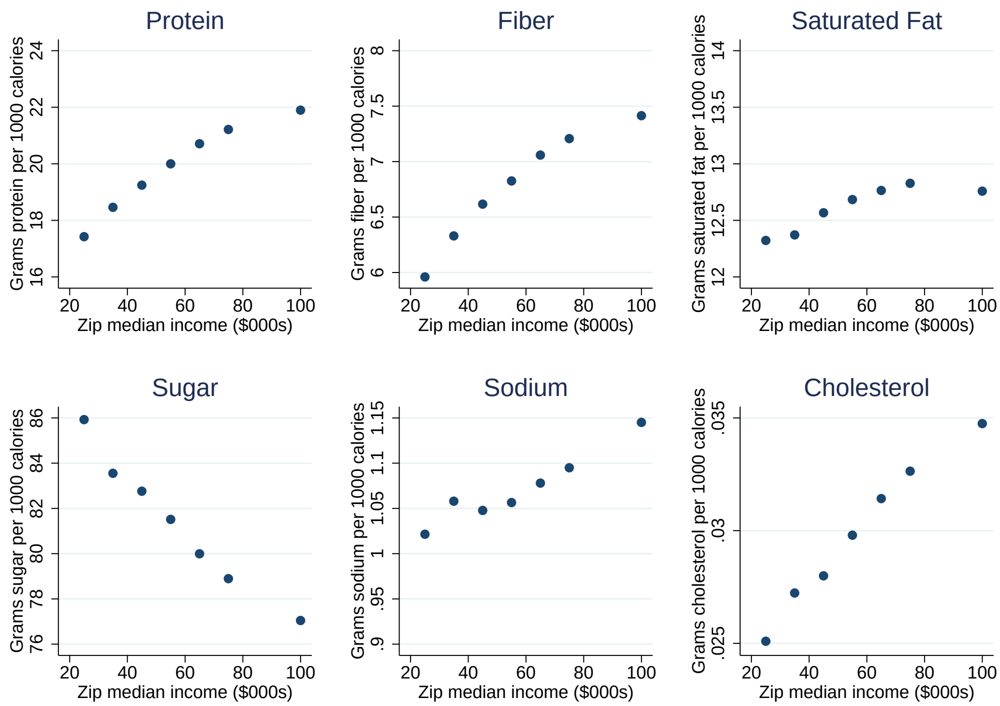
B Appendix to Stylized Facts Section

Figure A2: **Median Shopping Trip Distances by Household Income**



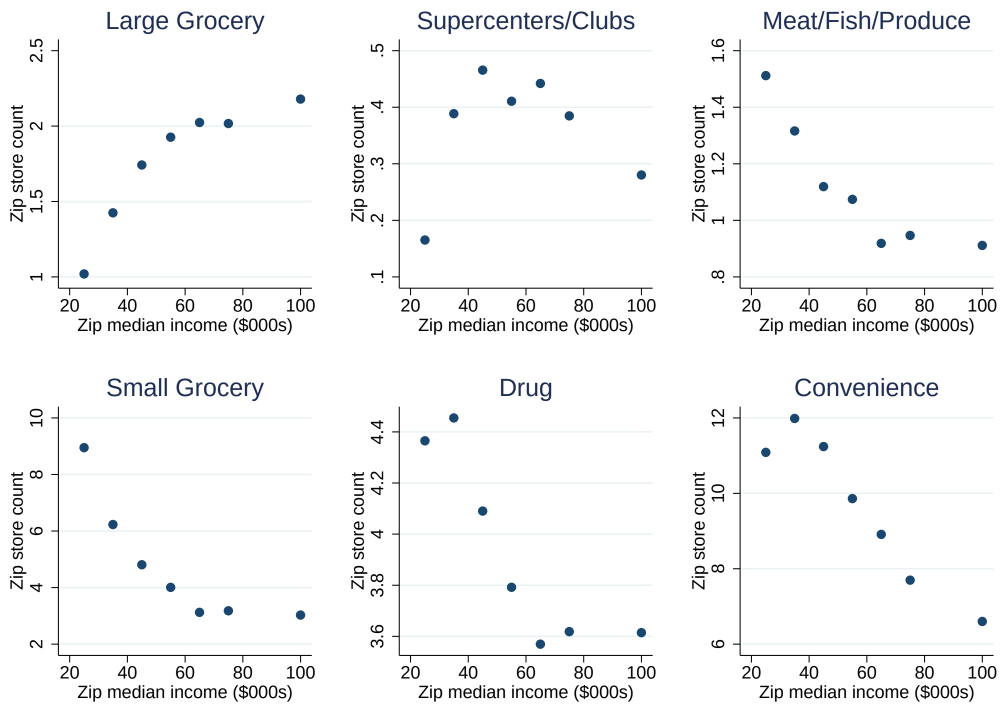
Notes: Data are from the 2009 National Household Travel Survey. Bars represent the mean one-way trip distance for trips beginning or ending in "buying goods: groceries/clothing/hardware store." "Poor" means household income less than $30,000, "Dense" means Census tract population density greater than 4,000 people per square mile, and "No car" means that the household does not own a car.

Figure A3: **Store Average Healthfulness by Zip Code Median Income**



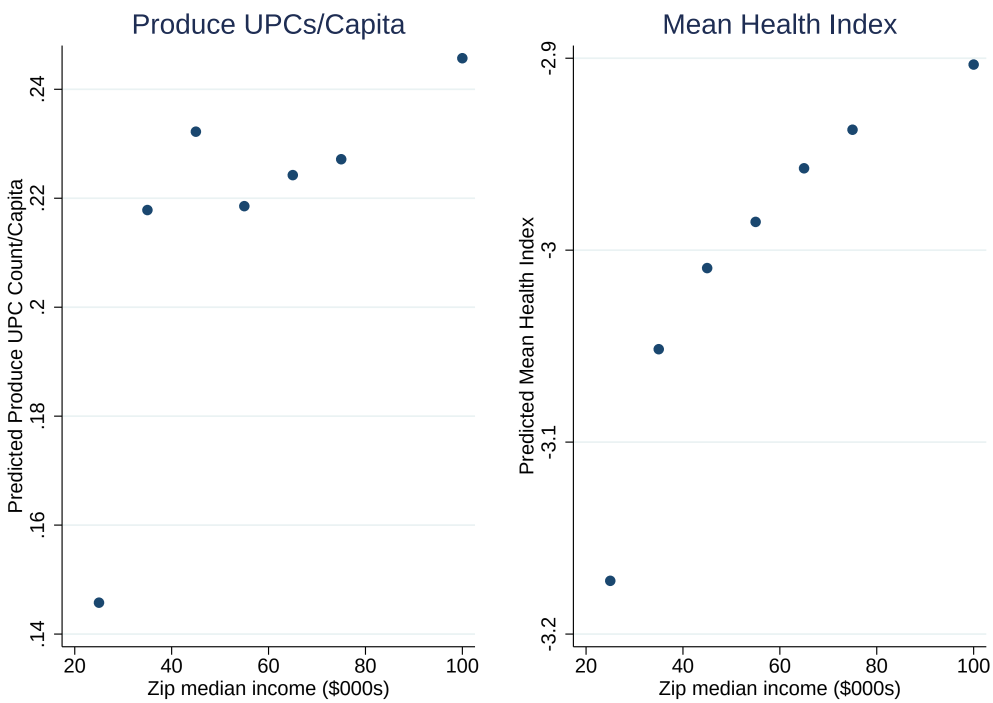
Notes: Presents unweighted means of store-level means of macronutrients across UPCs sold, for all stores in a given category of zip code median income. Data are from Nielsen RMS for year 2006. Constructed parallel to Figure 2 in the text.

Figure A4: **Store Counts by Zip Code Median Income**



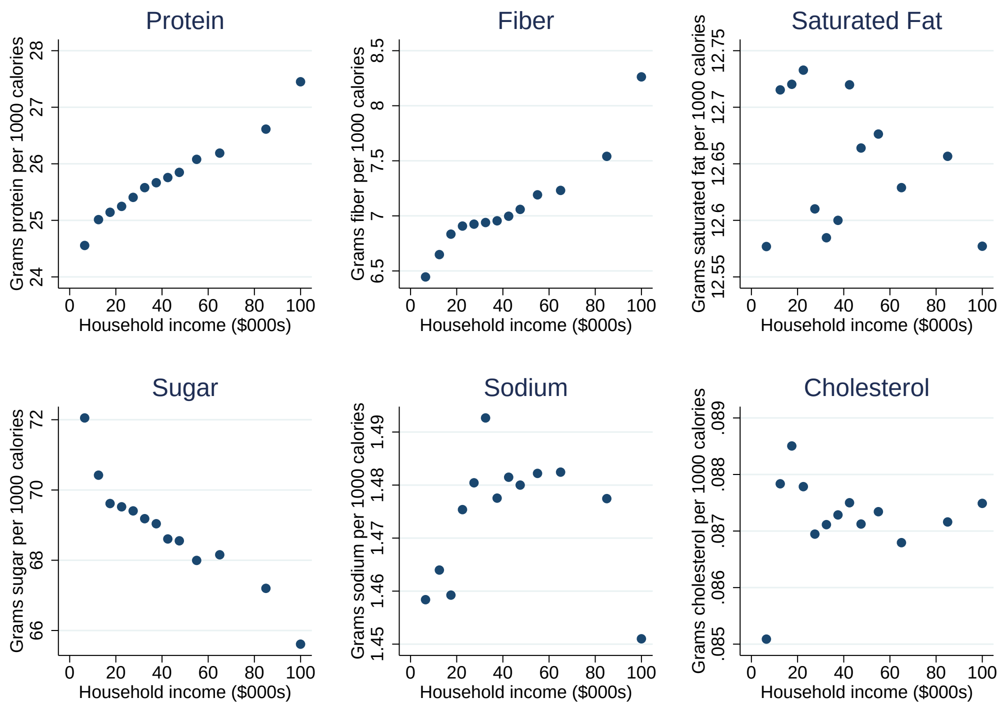
Notes: Notes: Population-weighted mean store counts by zip code income category from Zip Code Business Patterns. Large (small) grocers are defined as those with 50 or more (fewer than 50) employees. Parallels Figure 3 in the text, except does not normalize counts by zip code population.

Figure A5: **Store Counts by Zip Code Median Income**



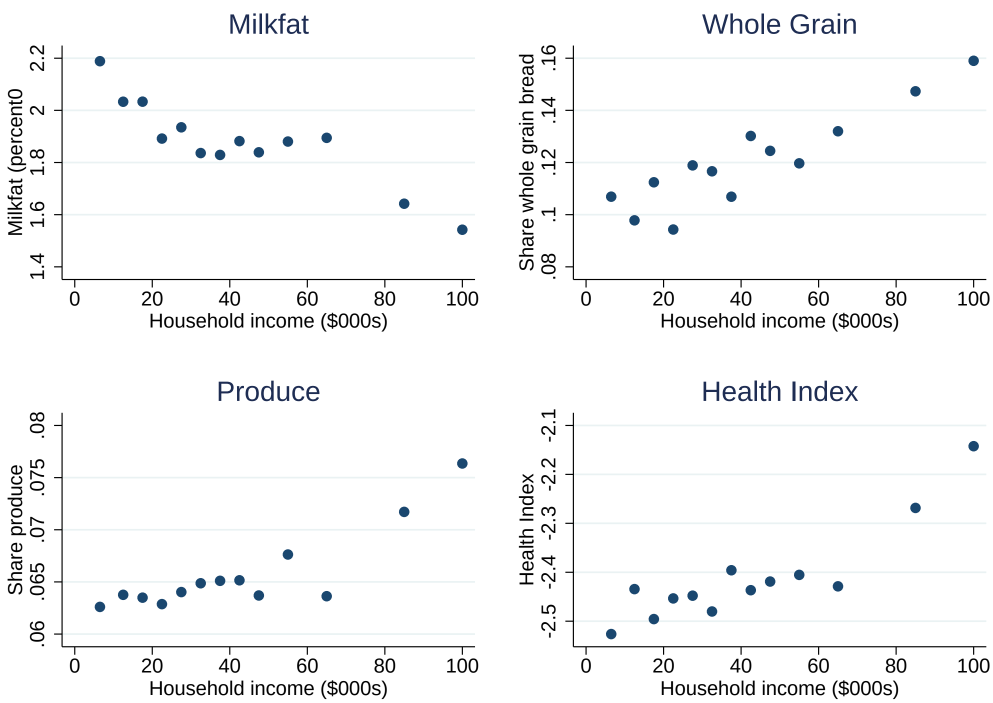
Notes: Population-weighted means across zip codes within categories of zip code median income of predicted count of produce UPCs per capita and mean Health Index across all UPCs offered. Predictions are based on projecting estimates from columns 3 and 6 of Table 2 onto Zip Code Business Patterns store counts.

Figure A6: **Macronutrient Purchases by Household Income**



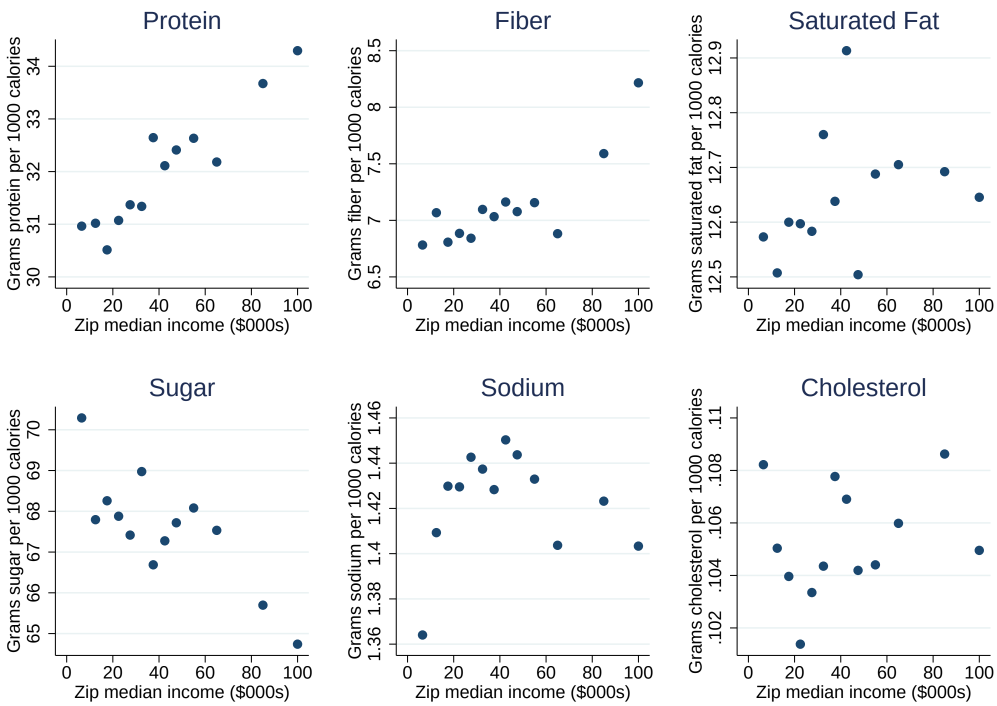
Notes: Presents calorie-weighted average macronutrient contents of purchases using Nielsen Homescan data for 2004-2012. The x-axis presents nominal income bins; household incomes larger than $100,000 are coded as $100,000.

Figure A7: **Magnet Subsample: Healthful Purchases by Household Income**



Notes: Nielsen Homescan data, magnet subsample, for 2004-2006. The x-axis presents nominal income bins; household incomes larger than $100,000 are coded as $100,000. This parallels Figure 5, except using the magnet subsample which also records purchases of non-UPC items such as bulk produce. Milkfat is the calorie-weighted average milkfat of milk purchases, whole grain is the calorie-weighted average share of bread, buns, and rolls purchases that are whole grain, produce is the share of calories from fresh, canned, dried, and frozen fruits and vegetables, and Health Index is the average Health Index per 1000 calories.

Figure A8: **Magnet Subsample: Macronutrient Purchases by Household Income**



Notes: Presents calorie-weighted average macronutrient contents of purchases using Nielsen Homescan data, magnet subsample, for 2004-2006. The x-axis presents nominal income bins; household incomes larger than $100,000 are coded as $100,000. This parallels Figure A6, except using the magnet subsample which also records purchases of non-UPC items such as bulk produce.

C Appendix to Retailer Entry Section

Table A3: **Effects of Retailer Entry: Known Retailers**

Panel A: Effects of Retailer Entry on Expenditure Shares

	(1)	(2)	(3)	(4)	(5)	(6)
Sample:	All Zip Codes			"Food Desert" Zip Codes		
Expenditure shares at store type:	Non-Chain Grocers	Convenience/ Drug Stores	Other Mass Merchants	Non-Chain Grocers	Convenience/ Drug Stores	Other Mass Merchants
Post entry: 5-10 miles	-0.00340	0.000609	-0.00140	-0.000655	-0.00224	-0.0122
	(0.00127)***	(0.000660)	(0.000923)	(0.00494)	(0.00207)	(0.00297)***
Post entry: 0-5 miles	-0.00418	-0.000746	-0.00763	-0.0118	-0.00125	-0.0149
	(0.00136)***	(0.000743)	(0.00108)***	(0.00430)***	(0.00172)	(0.00401)***
Observations	1,826,027	1,826,027	1,826,027	345,031	345,031	345,031
Dependent var. mean	0.052	0.034	0.052	0.071	0.026	0.056

Panel B: Effects of Retailer Entry on Healthful Eating

	(1)	(2)	(3)	(4)
Sample:	<1000 Produce UPCs		No Supergrocers in Zip	
Dependent variable:	Share Produce	Health Index	Share Produce	Health Index
Post entry: 5-10 miles	0.000853	-0.00186	0.000783	0.00936
	(0.000951)	(0.0203)	(0.000616)	(0.0137)
Post entry: 0-5 miles	0.00160	0.0381	0.000830	0.00144
	(0.00208)	(0.0357)	(0.00106)	(0.0199)
Observations	216,287	216,285	448,404	448,400
Dependent var. mean	0.048	-2.65	0.049	-2.65

Notes: Parallels Table 1, except Panel A presents effects on expenditure shares at alternative channel types, and Panel B uses alternative definitions of a "food desert." Columns 1 and 2 limit to zip codes with fewer than 1000 predicted produce UPCs, where predictions are based on projecting estimates from columns 3 and 6 of Table 2 onto Zip Code Business Patterns store counts. Columns 3 and 4 limit to zip codes with no large grocers, supercenters, or club stores. Data are at the household-by-quarter level. Share Produce is the share of calories from fresh, canned, dried, and frozen fruits and vegetables; Health Index is the average Health Index per 1000 calories. Reported independent variables are indicators for whether a specific retailer has entered within 0-5 or 5-10 miles of the household; regressions also include Census division-by-quarter of sample indicators and household fixed effects. Robust standard errors, clustered by household, in parentheses. *, **, ***: Statistically significant with 10, 5, and 1 percent confidence, respectively.

Table A4: **Alternative Estimates of Effects of Retailer Entry: All Retailers**

Panel A: Effects of Retailer Entry on Expenditure Shares

	(1)	(2)	(3)	(4)	(5)	(6)
Sample:	All Zip Codes			"Food Desert" Zip Codes		
Expenditure shares at store type:	Non-Chain Grocers	Convenience/ Drug Stores	Other Mass Merchants	Non-Chain Grocers	Convenience/ Drug Stores	Other Mass Merchants
Large grocers	-0.000566	-0.000109	0.000735	-0.000413	-0.000983	0.000836
	(0.000539)	(0.000283)	(0.000452)	(0.00385)	(0.00123)	(0.00238)
Supercenters/clubs	-0.00357	-0.000583	-0.00529	-0.0101	-0.00169	-0.00825
	(0.00103)***	(0.000557)	(0.000905)***	(0.00527)*	(0.00274)	(0.00559)
Observations	483,015	483,015	483,015	88,823	88,823	88,823
Dependent var. mean	0.051	0.034	0.051	0.069	0.026	0.055

Panel B: Effects of Retailer Entry on Healthful Eating

	(1)	(2)	(3)	(4)
Sample:	<1000 Produce UPCs		No Supergrocers in Zip	
Dependent variable:	Share Produce	Health Index	Share Produce	Health Index
Large grocers	0.00112	-0.0174	0.000886	0.00901
	(0.00138)	(0.0271)	(0.000564)	(0.0121)
Supercenters/clubs	0.00207	0.0510	0.00000378	0.00415
	(0.00111)*	(0.0261)*	(0.000821)	(0.0200)
Observations	55,746	55,746	116,510	116,510
Dependent var. mean	0.048	-2.65	0.048	-2.65

Notes: Parallels Table 6, except Panel A presents effects on expenditure shares at alternative channel types, and Panel B uses alternative definitions of a "food desert." Columns 1 and 2 limit to zip codes with fewer than 1000 predicted produce UPCs, where predictions are based on projecting estimates from columns 3 and 6 of Table 2 onto Zip Code Business Patterns store counts. Columns 3 and 4 limit to zip codes with no large grocers, supercenters, or club stores. Data are at the household-by-year level. Share Produce is the share of calories from fresh, canned, dried, and frozen fruits and vegetables; Health Index is the average Health Index per 1000 calories. Reported independent variables are the count of stores by channel type in the household's zip code; regressions also include Census division-by-year indicators and household fixed effects. Robust standard errors, clustered by household, in parentheses. *, **, ***: Statistically significant with 10, 5, and 1 percent confidence, respectively.

D Appendix to Model Estimation

Table A5: Preferences for Nutrients by Household Income

Income	Carbs	Fat	Saturated Fat	Fiber	Protein	Sugar	Sodium	Cholesterol	Fruit	Vegetables
Inc$\leq 25K$	1.07	11.28	-2.62	-6.62	3.71	2.89	-8.01	-410.00	0.04	-0.74
	[0.092]	[0.318]	[0.160]	[0.217]	[0.190]	[0.126]	[0.218]	[16.227]	[0.028]	[0.024]
$25K < Inc \leq 50K$	1.17	11.22	-2.49	-6.49	3.18	2.72	-8.45	-328.43	0.08	-0.71
	[0.104]	[0.220]	[0.111]	[0.130]	[0.126]	[0.033]	[0.167]	[10.746]	[0.021]	[0.017]
$50K < Inc \leq 75K$	1.72	12.89	-3.09	-6.58	3.48	2.77	-9.18	-328.46	0.13	-0.83
	[0.133]	[0.278]	[0.139]	[0.171]	[0.161]	[0.044]	[0.207]	[13.214]	[0.030]	[0.023]
$75K < Inc$	3.42	17.96	-3.20	-7.74	5.54	2.82	-10.32	-501.49	0.21	-1.24
	[0.117]	[0.249]	[0.128]	[0.157]	[0.151]	[0.039]	[0.184]	[13.519]	[0.024]	[0.023]

Notes: Standard errors are clustered by household. Magnitudes represent willingness to pay in dollars for 1 kilogram of nutrient. A given nutrient can enter through multiple preference parameters, such as saturated fat is both valued as fat and satuarted fat. Fiber and sugar are also carbohydrates. To get total willingnges to pay for these nutrients, the point estimates need to summed across the relavant nutrients. Value of fruit and vegetables accounts for value over and beyond macronutient characteristics of the fruit and vegetables.