

What drives nutritional disparities? Retail access and food purchases across the socioeconomic spectrum ^{*†}

Jessie Handbury

University of Pennsylvania and NBER

Ilya Rahkovsky

Economic Research Service, USDA

Molly Schnell

Princeton University

March 17, 2015

Abstract

The poor diets of many consumers are often attributed to limited access to healthy foods. In this paper, we use detailed data describing the healthfulness of household food purchases and the retail landscapes in which these consumers are making these decisions to study the role of access in explaining why some people in the United States eat more nutritious foods than others. We first confirm that households with lower income and education purchase less healthful foods. We then measure the spatial variation in the average nutritional quality of available food products across local markets, revealing that healthy foods are less likely to be available in low-income neighborhoods. Though significant, the spatial differences in access are small and explain only a fraction of the variation that we observe in the nutritional content of household purchases. Systematic socioeconomic disparities in household purchases persist after controlling for access: even in the same store, more educated households purchase more healthful foods. Our results indicate that policies aimed at improving access to healthy foods for underserved socioeconomic groups will leave most of the disparities in nutritional consumption intact.

*The materials contained in this report are strictly confidential and not cleared for release. This report should not be used or quoted in any fashion until the USDA releases it for publication.

†Prottoy Aman Akbar and Yue Cao provided us with outstanding research assistance. Jessie Handbury would like to thank the Wharton Social Impact Initiative, the Research Sponsors' Program of the Wharton Zell-Lurie Real Estate Center, and the Economic Research Service at the USDA for generous financial support. The views expressed are those of the authors and should not be attributed to the Economic Research Service, the USDA, Nielsen, Gladson, or IRI.

1 Introduction

It is well known that there are large nutritional disparities across different socioeconomic groups in the United States, but little is known about why such disparities exist. Poor diets are often attributed to three factors: food deserts restricting consumer access to healthy foods, preferences for unhealthy foods, and higher prices of healthy foods. Under the assumption that differential access is to blame for nutritional disparities, the Agricultural Act of 2014 introduced \$125 million to be spent annually in each of the next five years to promote access to healthy foods in underserved communities (Aussenberg, 2014). Many states have also introduced programs to improve access by providing loans, grants, and tax credits to stimulate supermarket development and to encourage retailers to offer healthy foods in food deserts (CDC, 2011).¹

Despite the growing popularity of such programs, little is known about their success in narrowing nutritional disparities. This paper measures the maximal impact of these policies by quantifying the role that access plays in generating socioeconomic disparities in nutritional consumption. We first employ novel data describing the nutritional quality of food products available to and purchased by U.S. households to characterize the degree of socioeconomic disparities in access to and consumption of nutritious foods. We then use a theoretical framework to demonstrate the main challenge that we face in identifying the causal role that access plays in generating disparities in purchases. Since households sort into neighborhoods and retailers cater to local tastes, consumption disparities across locations with differential access may reflect not only the role of access but also demand-side factors. Our theory suggests that we can separately identify these demand-side factors by looking at the purchases of households living in the same location. We therefore use the detailed residential and shopping location information in our household purchase data to measure an upper bound for the role that access plays in generating the existing disparities. We complement this cross-sectional analysis with a look at how households respond to observed changes in their retail environment. Together, our results indicate that improving access to retail outlets alone will do little to close the gap in the nutritional quality of diets across different socioeconomic groups. While equating access would help to reduce differences in nutritional consumption across different income groups, over 90% of the disparities across education groups would remain.

Using two novel measures of the healthfulness of household purchases, we first document significant differences in the nutritional quality of foods purchased by different socioeconomic groups across the U.S. This generalizes the results of previous studies which have documented disparities in nutritional consumption by focusing on purchases of a few products, such as fruits or vegetables, or in specific localities (see Bitler and Haider (2011) for a detailed survey of this work).² To obtain a more comprehensive picture of the nutritional consumption of household purchases, we combine consumption data from Nielsen with nutritional information from Gladson and IRI to construct a dataset describing the full nutritional profile of the grocery purchases made by over 60,000 households from 52 markets across the U.S. between 2006 and 2011. We calculate two complementary household-level indexes, an “expenditure score” and a “nutrient score,” that represent the healthfulness of the products purchased relative to USDA category-level expenditure recommendations and FDA recommendations for per calorie nutrient consumption, respectively.³ An examination of these household-level nutritional indexes reveals significant

¹Between 2004 and 2010, the Pennsylvania Fresh Food Financing Initiative provided \$73.2 million in loans and \$12.1 million in grants to stimulate supermarket development in food deserts in the state. In 2013, North Carolina House Bill 957 began granting tax credits to retailers who offer healthful foods in food deserts. In 2014, Maryland House Bill 451 provided \$1 million in assistance to food deserts through loans and grants, and the New Jersey Food Access Initiative started a private-public partnership to attract supermarkets to underserved areas.

²While there is a large literature in economics on the relationship between socioeconomic status and various health behaviors (e.g., Cutler and Lleras-Muney (2010); Jones (1997)), grocery purchases are one health behavior which has received surprisingly little attention.

³Our expenditure score is an extension of the measure used by Volpe et al. (2013). Given the nutritional information we have from Gladson and IRI, however, we can go further than looking at expenditures on food group categories. Our nutrient score directly measures the

disparities in the healthfulness of purchases across households with different income and education levels. The products purchased by households in the highest terciles for income and education are 40 percent closer to both USDA recommendations for product category expenditure shares and FDA recommendations for per calorie nutrient consumption than the products purchased by households in the lowest terciles of income and education.

Next, we provide the most comprehensive picture of the healthfulness of products available at retail locations across the U.S. to date and quantify the degree to which retail environments differ by socioeconomic status. Consistent with previous studies, we find that access to healthy foods is greater in wealthier and more educated neighborhoods (Beaulac et al. (2009); Ver Ploeg et al. (2009)). Using geo-coded data on the location of over 200,000 retailers across the U.S., we first document that there are large disparities in the concentration of stores across neighborhoods with different socioeconomic profiles. We then use weekly store-level sales data from Nielsen to identify the products that are available at over 30,000 participating retailers between 2006 and 2011. Analogous to the household-level analysis, we merge the Nielsen data with nutritional information from Gladson and IRI to calculate two complementary store-level healthfulness indexes. We find small, but statistically significant, correlations between observable market characteristics and the store-level healthfulness indexes, with stores in high-income and high-education neighborhoods offering more healthful products. Together, these results indicate that households in high-income and high-education neighborhoods have access to a significantly larger choice set of stores than households in other neighborhoods, with slightly more nutritious foods being offered in these stores.

While there is agreement among researchers that spatial and socioeconomic disparities in nutrition exist, the actual effects of access to healthy foods on food purchases is heavily contested (Bitler and Haider (2011)). Some studies find no relationship between store density and consumption (see, for example, Pearson et al. (2005) and Kyureghian et al. (2013)). Other studies that do find a positive relationship infer the role of food environments from a cross-sectional correlation between local store density and food purchases (Rose and Richards (2004); Morland et al. (2002); Bodor et al. (2008); Sharkey et al. (2010)). Determining the direction of causality in this relationship is crucial in assessing the potential impact of policies that encourage the entry of new stores into food deserts on the nutritional consumption of households in these areas. Up to this point, data limitations have led to measurement and identification issues which have hindered a clear understanding of the role that access plays in generating nutritional disparities.⁴

The detailed nature of our data allows us to go beyond existing work in examining the direction of causality in the relationship between nutritional availability and nutritional consumption. In two complementary analyses, we quantify the role that the spatial disparities we document using the store-level data play in generating the consumption disparities that we observe using the household-level data. As we expect disparities in consumption that are due to differential access to exist only between households living in different neighborhoods, we first look at whether consumption disparities persist when we control for the location of households. While the correlation between income and the healthfulness of food purchases is reduced by half when we control for the household's census tract, the relationship between education and healthfulness is only reduced by 10%. While informative, our "within-location" approach has its limitations. It is possible that households living in the same neighborhood still have differential access, either because they live in different locations within the neighborhood or because of differences in mobility (e.g., car ownership). To eliminate the effect of access entirely, we look at purchases made

healthfulness of the relative quantities of macro-nutrients in the products purchased.

⁴There is also no consensus on the impact of a household's retail environment on obesity and other health problems. Anderson and Matsa (2011) find no effect of fast food entry on obesity, while Currie et al. (2010) find impacts on school children and pregnant women. Courtemanche and Carden (2011) find that Walmart entry increases local obesity rates, though non-causal results from Chen et al. (2010) and Volpe et al. (2013) suggest that the impact of store entry varies with neighborhood characteristics and the type of store entering.

within a given store. The results from the within-store analysis mirror those from the within-location analysis: the correlation between income and the healthfulness of food purchases is cut in half when we look at purchases made within the same store, whereas the correlation between education and nutritional quality is only reduced by 10%. In both the within-location and within-store analyses, the majority of the disparities that we observe between households persist when we control for access. We conclude that disparities in access play a minimal role in explaining observed disparities in consumption.

We present a simple model to formalize the intuition behind this empirical approach. The model nests two mechanisms, one driven by demand and one driven by supply, each which can independently explain the socioeconomic disparities in access to healthy foods that we observe. The demand-side explanation relies on within-group preference externalities. In a monopolistically competitive retail industry, firms will cater to the prevalent tastes in the local market. If high-socioeconomic households have stronger tastes for healthy foods than low-socioeconomic households, it follows that more healthful food products will be sold in high-socioeconomic neighborhoods. The supply-side explanation relies on two fairly general assumptions: (i) wholesale unit costs are increasing in product healthfulness, but do not vary across location, and (ii) location-specific marginal costs of retailing are increasing in the share of high-socioeconomic status residents in a neighborhood, but do not vary across products of different levels of healthfulness. These assumptions imply that firms in neighborhoods with a greater share of high-socioeconomic status residents have a comparative advantage in the distribution of nutritious products. As a result, they will sell more healthful food products than stores in low-socioeconomic neighborhoods, even if high- and low-socioeconomic status households have identical tastes. This model serves to demonstrate that the differences that we observe in the healthfulness of purchases made by high- and low-socioeconomic status households living in the same location act as a lower bound for the component of the overall socioeconomic disparities that we observe across households living in various locations that can be explained by factors other than the retail environment. We therefore conclude that the difference between the disparities we observe across locations and the disparities that we observe within locations is an upper-bound for the component of the existing disparities in purchases that can be explained by the retail environment alone.

Finally, we directly consider whether policies that improve access to healthy food products will have any impact on socioeconomic disparities in nutrition by looking at how households have responded to improvements in access in the past. Previous studies measuring the effects of changes in the retail landscape on food purchases are local in scope, looking at either the entry of a single supermarket or an intervention to increase the availability of nutritious food products in a single urban food desert, and find modest effects (Wrigley et al. (2003); Cummins et al. (2005); Weatherspoon et al. (2013); Song et al. (2009); Cummins et al. (2014)). We demonstrate that these results hold more generally by showing that the impact of store entry on the healthfulness of food purchases made by local households is limited in the 3,087 store entries that we observe. Improving the retail environment in low-income neighborhoods will only be effective in resolving socioeconomic disparities in nutritional consumption insofar as the nutritional quality of purchases made by low-income households improves in response to these changes. We find that the elasticity of the healthfulness of household food purchases with respect to the density and nutritional quality of retailers in their vicinity is positive, but close to zero. Improving the concentration and nutritional quality of the stores in the average low-income and low-education neighborhood to match those of the average high-income and high-education neighborhood would only close the gap in consumption by 1 to 3 percent. These results again suggest that policies aimed at improving access to healthful foods will do little to resolve the consumption disparities we document.

Despite a large policy literature on the topic, the relationship between access and nutritional consumption has been largely ignored by economists. Methodologically, our paper is closest to the literature in economics which uses the entry of fast food restaurants and large retailers, such as Walmart, to identify a causal relationship between the retail environment and obesity more generally (Currie et al. (2010); Anderson and Matsa (2011); Courtemanche and Carden (2011)). Our paper departs from these previous studies in two important dimensions. First, we are concerned not just with the relationship between access and nutritional consumption, but rather the interaction between access, nutritional consumption, and socioeconomic status. This is important from a policy perspective, as current policies aim to reduce disparities in consumption across socioeconomic groups. From a methodological perspective, our focus on disparities allows us to use both cross-sectional and time-series variation to consider the impact of the retail environment on disparities of health behaviors. Second, we look directly at the mechanism, food purchases, by which we expect changes in households' retail environments to impact obesity, rather than obesity itself. This is important because while access may have a causal impact on obesity, it need not work through the hypothesized mechanism, and the mechanism is of greater concern from a policy perspective.

If disparities in retail access do not generate the consumption disparities that we observe, then something else is to blame. In the context of our model, differences in demand are generated by differences in tastes. There are, however, a range of other explanations for consumption disparities, including differences in price sensitivities and budget constraints. For the purposes of this paper, we are agnostic as to the reasons why we observe systematic differences in the healthfulness of purchases between households either living in the same location or shopping in the same store. In future work, we aim to determine which factors are most important in explaining the large disparities that persist when we look at households in the same location.

The paper proceeds as follows. Section 2 describes the datasets that we use. Section 3.1 presents the indexes that we construct to measure the nutritional quality of households' consumption baskets and documents how these indexes vary across households with different levels of income and education. Section 3.2 shows how we measure access to nutritious foods and documents disparities in access across markets with different observable characteristics. Section 4 presents a model that nests two mechanisms that could each generate the observed disparities in both purchases and access and demonstrates how geo-coded household purchase data can be used to identify the role of access, separately from demand-side factors, in generating purchase disparities. Section 5 implements this procedure by looking at whether consumption disparities persist when we control for residential or retail location. In Section 6 we take an alternative, time-series approach and examine whether we observe the healthfulness of household purchases responding to changes in local access. Section 7 concludes.

2 Data

We use six different datasets that together describe the nutritional quality of food purchases that households make, the stores located in the neighborhoods where these households reside, the nutritional quality of the products offered in these stores, and the demographics of these neighborhoods. The first dataset is the Homescan data collected by the National Consumer Panel (NCP)⁵ and provided by Nielsen. The Homescan data contains transaction-level purchase information for a representative panel of 114,286 households across the U.S. Households in the panel use a scanner provided by NCP to record all of their purchases at a wide variety of stores where food is sold. After scanning the Universal Product Code (UPC) of each item purchased, the household records the date, store

⁵The National Consumer Panel is a joint venture between Nielsen and IRI.

name, quantity purchased, and price.⁶ Households participate in the NCP panel on average for two years and eight months, with the length of observed participation ranging from six months to the full period of analysis (2006 to 2011). In addition to household-level purchase activity, the Homescan data also provides us with information on the location and demographics of each household in the panel. For each year that a household is in the NCP panel, we observe the census tract in which the household resides and a range of demographic characteristics. We use the demographic data to measure two dimensions of socioeconomic status which are posited to impact a household's consumption decisions: income and education.^{7,8}

While the NCP Homescan data describes where Homescan panelists shop and what they buy, it only provides a limited picture of the retail environments in which households are making these consumption decisions. There are two problems with using the Homescan data to characterize retail environments: First, if no household in the Homescan sample shops at a given store, then we do not observe from the data that the store exists. Second, even if we do observe households shopping in a given store, we only observe the products that they actually purchase, not the full variety of products offered. Because of these limitations, we use two additional datasets, both maintained by Nielsen, to obtain a more detailed picture of the retail environments that households face. To see the full set of stores available to households, we use the Nielsen TDLinx data. The TDLinx data contains the names and geo-coded locations of nearly 200,000 food stores across the U.S. These stores fall into five categories: grocery, convenience, mass merchandise, and drug. To see the full set of food products available at a subset of these stores, we combine the TDLinx data with the Nielsen Scantrack (RMS) data provided by the Kilts-Nielsen Data Center at University of Chicago Booth School of Business.⁹ The RMS data contains UPC-level weekly sales values and quantities generated from point-of-sale systems in over 30,000 participating retailers across the U.S. We use this data to calculate indexes that summarize the nutritional quality of products offered by each store in the dataset.^{10,11}

The Nielsen datasets do not contain nutritional information for the products purchased by Homescan panelists or sold by Scantrack stores. We obtain this information from Gladson and IRI. The Gladson Nutrition Database provides nutritional information for over 200,000 unique UPCs. We supplement the Gladson data with nutritional information from the IRI database of over 700,000 UPCs. Each database contains information on the quantity of macro-nutrients and vitamins per serving, serving size in weight, and the number of servings per container.

⁶See Harding and Lovenheim (2014) for a detailed description of how this panel of households is recruited and encouraged to continue reporting purchases on a weekly basis.

⁷Households record whether their income falls into one of 16 categories, listed in appendix Table A.1. We limit our analysis to households that have at least one household head working over 30 hours a week and report annual earnings of over \$8,000. We assign households an income equal to the midpoint of their income category for each bounded category and an income of \$260,000 for the "\$200,000 and above" category. Where noted, we adjust the resulting household income for household size using the OECD equivalence scale. The first adult in the household receives a weight of 1 and all other adults receive weights of 0.5, while each child receives a weight of 0.3 (<http://www.oecd.org/eo/growth/OECD-Note-EquivalenceScales.pdf>).

⁸Households record the male and/or female household head's education in one of six categories: grade school, some high school, high school graduate, some college, college graduate, or post-college graduate. The distributions of household heads across these education categories by sex are recorded in appendix Tables A.2 and A.3. For our analysis, we exclude households in which either household head reports only a grade school education, as there are too few observations to obtain precise estimates. We assign each household head a number of years of education, assuming that some high school is equal to 10 years, some college is equal to 14 years, and post-graduate is equal to 18 years. For households with both a male and a female head, we take the mean years of education across household heads.

⁹Information on availability and access to this data is available at research.ChicagoBooth.edu/nielsen.

¹⁰We assume that every product available in a store is sold to at least one customer each month.

¹¹Despite this detailed information on prices and product offerings, the RMS data covers a more limited range of retail outlets than the TDLinx data and only provides us with the county, not the precise geo-coded location, of each store. Where possible, we obtain the geo-coded location of the stores in the RMS data by matching them to the TDLinx data as follows: If there is only one observation for a given combination of store name and county in both datasets, then we assume that this is the same store. If there are multiple observations for a given store name-county pair, we match the stores based on a comparison of the households that we observe shopping at both the TDLinx and the RMS store on the same day.

Gladson and IRI collect this information directly from product labels.^{12,13} We merge the Gladson and IRI data with the Nielsen Homescan and RMS data to obtain the nutritional profile of products we observe being purchased by households and sold in stores.¹⁴ In Sections 3.1 and 3.2, we describe how we use this information to measure the healthfulness of households' grocery purchases and the healthfulness of products available at the store-level, respectively.

The final dataset that we use contains tract-level demographics from the 2000 U.S. Census.¹⁵ We use this information to measure the distribution of income and education in the neighborhoods in which Nielsen households and stores are located.

3 Stylized Facts

3.1 Disparities in Nutritional Consumption

We begin by documenting the extent of the disparities in nutritional consumption across household types. We focus on the *quality* rather than the quantity of food a household purchases since the latter is affected by the extent to which a family eats at restaurants, and a propensity for eating out is likely related to household characteristics. We measure the quality of a household's purchases using two complementary indexes. We calculate these indexes at a monthly frequency for each household in the sample. The first index measures the extent to which a household's grocery purchases deviate from the USDA Center for Nutrition Policy and Promotion (CNPP)'s dietary guidelines for recommended expenditure shares by USDA food category. This index follows the measure used in Volpe et al. (2013). We will refer to this as the "expenditure score.". Given the nutritional information we have from Gladson and IRI, however, we can go further than looking at expenditures on food group categories. We therefore also calculate a "nutrient score" that directly measures the healthfulness of the relative quantities of macro-nutrients in the products purchased. The nutrient score measures the extent to which a household's purchases deviate from the FDA's recommendations for nutrients per calorie. Both indexes are based on inverse squared loss functions that penalize households with purchases above (below) the recommended amounts in unhealthful (healthful) food categories or nutrients.

The expenditure score for the grocery purchases recorded by household h in month t is defined as

¹² The Gladson sample is based on the labels of products it receives from a variety of sources, including major retailers, food distributors, and food manufacturers. Gladson also maintains a "remote capture station" at a Safeway in California, where it records nutritional information for all of the products sold in the store. IRI purchases pre-market images from manufacturers and retailers and also conducts field-based imaging.

¹³ Product characteristics can change without a change in the product's UPC. When Gladson receives an updated version of the product that was already in the database, it updates the database, including a time stamp of when the product was added or updated. We use a version of the database that includes a snapshot of the market as of July 30th each year. We assume that these product characteristics are relevant for that calendar year.

¹⁴ These merges are not perfect. Only 45% of the UPCs in the Homescan data and 57% of the UPCs in the RMS data are in the Gladson or IRI nutrition data. We impute nutritional information for products not in the Gladson or IRI data using the average for UPCs in the same product module and product group, with the same values for all other relevant characteristics, such as brand, flavor, form, formula, style, and type.

¹⁵ The Nielsen data identifies household locations using 2000 tract definitions.

$$Expenditure\ Score_{ht} = \left[\sum_{c \in C_{healthful}} (sh_{cht} - sh_{ch}^{CNPP})^2 | sh_{cst} < sh_{ch}^{CNPP} + \sum_{c \in C_{unhealthful}} (sh_{cht} - sh_{ch}^{CNPP})^2 | sh_{cst} > sh_{ch}^{CNPP} \right]^{-1}$$

where c indexes CNPP food categories, sh_{cht} denotes the percent of household h 's observed grocery expenditures in month t on products in category c , and sh_{ch}^{CNPP} is the category c expenditure share, also in percent units, that the CNPP recommends for a household with the same gender-age profile as household h .¹⁶ We determine which CNPP food categories are healthful and unhealthful using the recommendations from the Quarterly Food-at-Home Price Database (QFAHPD) indicators for which of 52 food groups are healthful and unhealthful.¹⁷ The expenditure score penalizes households for spending less than their recommended expenditure share on healthful food categories ($c \in C_{healthful}$) and for spending more than their recommended expenditure share on unhealthful categories ($c \in C_{unhealthful}$). We follow Volpe et al. (2013) and take the inverse of the squared loss function so that higher scores are indicative of healthfulness.¹⁸

The nutrient score for the grocery purchases recorded by household h in month t is defined as

$$Nutrient\ Score_{ht} = \left[\sum_{j \in J_{healthful}} \left(\frac{nutr_{jht} - nutr_j^{FDA}}{nutr_j^{FDA}} \right)^2 | nutr_{jht} < nutr_j^{FDA} + \sum_{j \in J_{unhealthful}} \left(\frac{nutr_{jht} - nutr_j^{FDA}}{nutr_j^{FDA}} \right)^2 | nutr_{jht} > nutr_j^{FDA} \right]^{-1}$$

where j indexes a specific nutrient, $nutr_{jht}$ denotes the amount of nutrient j per calorie contained in household h 's observed purchases in month t , and $nutr_j^{FDA}$ is the amount of nutrient j that the FDA recommends an individual consume per calorie as part of a 2,000 calorie diet.¹⁹ The FDA indicates whether to consider its recommendation for a given nutrient as a lower bound or an upper bound. We assign the nutrients for which the FDA recommendation is an upper bound to the unhealthful category. These nutrients are fat, saturated fat, sodium, and cholesterol. The FDA recommendation is considered a lower bound for fiber, iron, calcium, and vitamins A and C, and we allocate these nutrients to the healthful category.²⁰ The resulting nutrient score penalizes households for purchasing less

¹⁶We use the recommended individual expenditure share from the "liberal food plan" to construct the household recommended expenditure share. We assign weights to each household member following the OECD equivalence scale and calculate the food expenditure weights as $w_{adult} = \frac{n_{adult}}{1+(n_{adult}-1) \times 0.5 + n_{children} \times 0.3}$ and $w_{child} = \frac{0.3}{1+(n_{adult}-1) \times 0.5 + n_{children} \times 0.3}$. The recommended category c expenditure share for household h is a weighted average of the recommended category c expenditure share of each household member, $sh_{ch}^{CNPP} = \sum_i w_i recshare_{ic}$, where i is a household member whose age and gender determine his/her weight (w_i). The recommended category c expenditure share, $recshare_{ic}$, is taken from Carlson et al. (2007).

¹⁷We aggregate the 52 QFAHPD food groups to the 24 CNPP food categories using the correspondence created by Volpe and Okrent (2013). In doing so, we find that two CNPP food categories (cheese and meat) contain both healthful and unhealthful food groups. Since the vast majority of cheese and meat purchases are of UPCs that fall into the unhealthful QFAHPD food groups, we assume that the aggregate CNPP cheese and meat categories are unhealthful. All of our results are robust to assuming that these food groups are healthful.

¹⁸We drop expenditure scores that are more than twice the distance between the 90th and 50th percentiles (nearly 5% of household-month scores).

¹⁹These recommendations come from the FDA's instructions for how to make use of nutritional labels (<http://www.fda.gov/Food/IngredientsPackagingLabeling/LabelingNutrition/ucm274593.htm>, last accessed on Dec. 4, 2014).

²⁰Some of these nutrients are identified as "nutrients of concern" in the USDA's Nutritional Guidelines for Americans, but others are not. We use all of the available recommended nutrients, whether they are nutrients of concern or not, as our goal is to assess the overall healthfulness

(more) than the recommended amount of healthful (unhealthful) nutrients per calorie. We normalize the deviation of a households' nutrient purchases from the FDA's recommendations to account for differences in the units in which nutrients are measured.²¹

The two scores consider the healthfulness of consumer purchases from two complementary perspectives. Each has their benefits and weaknesses. The expenditure score is closely related to consumer demand, since consumers choose foods rather than nutrients. Furthermore, the purchases of food groups, such as fruits and vegetables, are used by many other studies, and thus the expenditure score is more comparable to previous research. Finally, the expenditure score takes into account other micronutrients, such as zinc and potassium, which are not displayed on the nutritional facts panel and are therefore excluded from the nutrient score. The nutrient index, on the other hand, distinguishes between products inside a food category, *e.g.* regular versus low-fat cheese, that will be missed by the expenditure score. The nutrient score is also not sensitive to systematic variations in the price of foods purchased by different socioeconomic groups. If, for example, poor and rich consumers purchase identical quantities of cheese, but rich consumers purchase more expensive varieties, then for equal expenditures rich and poor consumers will have different expenditure scores. The nutrient score, on the other hand, will reflect that both groups have similar diets.

Table 1 shows the correlations of the expenditure and nutrient scores with log household income and years of education, conditional on other demographics.²² We see that wealthier and more educated households purchase more healthful foods, measured using either the expenditure or the nutrient score. Although both effects are statistically significant, the standardized coefficients reported in columns (4) and (8) reveal that education explains more of the variation in the quality of household purchases than income. Nutritional disparities across households with different levels of education but the same level of income are approximately 50% larger than the disparities across income levels controlling for education. One can see this graphically in Figure 1, which depicts the average log expenditure and nutrient scores by income and education tercile. For both measure, the average score varies more across education groups than across income groups.

of individual diets rather than larger public health concerns. The nutrient index highlights the choices that consumers made relative to the information and recommendations available to them at the time of purchase. It is likely that the included nutrients, such as vitamins A and C (both listed as "nutrients of concern" in 2005 but dropped in 2010 in response to increased consumption), are correlated with "nutrients of concern" for which we do not have information, such as potassium.

²¹As with the expenditure scores, we drop nutrient scores that are more than twice the distance between the 90th and 50th percentiles (nearly 5% of household-month scores).

²²All regressions include household size dummies, average head of household age, a dummy for marital status of household heads, dummies for households with either a female or male household head, a dummy for the presence of children, and dummies for whether the household reports being white, black, Asian, or Hispanic. See Table A.4 in the appendix for the full regression results.

Table 1: Consumer Characteristics and Nutritional Quality of Purchases

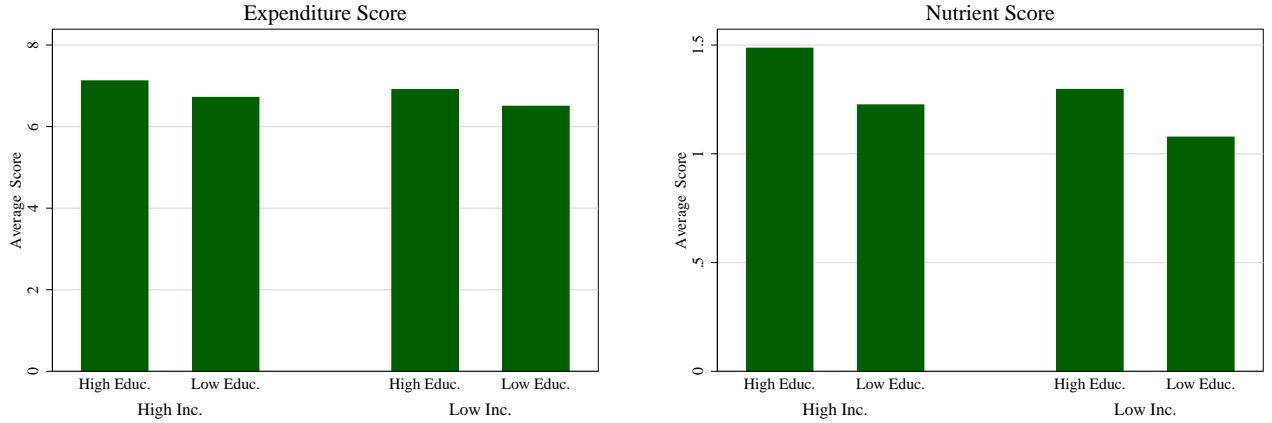
	Dep. Var.: Ln(Expenditure Score)				Dep. Var.: Ln(Nutrient Score)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ln(Income)	0.0424*** (0.0013)		0.0241*** (0.0014)	0.0426*** (0.0024)	0.146*** (0.0028)		0.0893*** (0.0030)	0.0636*** (0.0021)
Ln(Education)		0.247*** (0.0060)	0.203*** (0.0065)	0.0743*** (0.0024)		0.798*** (0.013)	0.635*** (0.014)	0.0939*** (0.0021)
Observations	3440297	3440297	3440297	3440297	3440297	3440297	3440297	3440297
R^2	0.061	0.064	0.066	0.066	0.022	0.026	0.029	0.029
Std Coef	No	No	No	Yes	No	No	No	Yes

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Observations are at the household-month level. Standard errors are clustered by household. All regressions include year-month fixed effects and controls for household demographics, including household size dummies, average head of household age, a dummy for marital status of household heads, dummies for households with either a female or male household head, a dummy for the presence of children, and dummies for whether the household reports being white, black, Asian, or Hispanic.

Figure 1: Expenditure and Nutrient Scores Across Households



Notes: The figure above presents mean household expenditure and nutrient scores across households with different socioeconomic statuses. Households are considered high income if their size-adjusted household income falls above the median level across all households and low income otherwise. Households are considered high education if the average years of education for their household head(s) falls above the median across all households and low education otherwise. These results are for January 2010; they are representative of the other months in the Homescan data.

3.2 Disparities in Access

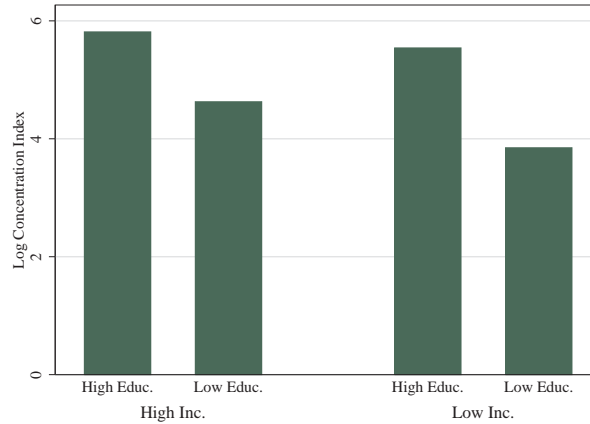
We now turn to documenting the disparities in the availability of healthy foods across locations. We start by looking at simple concentration indexes that reflect the spatial distribution of retail food stores surrounding the census tracts where households in our dataset reside. The concentration indexes are kernel densities based on store location from the TDLinx data. Let d_{sl} denote the distance between store s and the centroid of census tract l , and let S_t denote the universe of stores in our sample. We calculate the concentration kernel density for census tract l in time t as a Gaussian kernel with a bandwidth of 20km:²³

$$Concentration\ Index_{lt} = \sum_{s=1}^{S_t} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{d_{sl}}{20} \right)^2}$$

²³Our results are robust to the choice of bandwidth and kernel specification.

Figure 2 shows how store concentration indexes for 2010 vary with census tract demographics from the U.S. Census.²⁴ We see that there is spatial correlation between income, education, and the concentration index: wealthier and more educated census tracts have a higher concentration of stores in their vicinity.

Figure 2: Store Concentration Indexes Across Tracts



Notes: The figure above presents mean concentration indexes across tracts with different socioeconomic statuses. Tracts are considered high income if their mean household income falls above the median level across all tracts and low income otherwise. Tracts are considered high education if their share of college educated residents falls above the median across all tracts and low-education otherwise. These results are for 2010; they are representative of the other years in the TDLinx sample.

While kernel densities of the number of stores allow us to examine store concentrations, this measure ignores the fact that all stores are not equal. Importantly, stores differ in the products they sell. To account for spatial disparities in nutritional availability across markets, we use the RMS data to compute healthfulness indexes for each of the stores in the RMS panel that we are able to match to location information in the TDLinx data.

To summarize the nutritional content of the products sold in a given store in a given month, we use store-level variants of the expenditure and nutrient scores defined in Section 3.1 for households. The indexes reflect the category-level expenditure shares and per calorie nutrients that a representative household would purchase in store s in month t . The household is nationally representative in that they purchase all of the products sold in a store during a month using proportions derived from the national sales of those products in that month. Let U_t denote the universe of UPCs sold nationally in month t , S_t the set of stores in the sample in month t , and v_{ust} the total sales of UPC u in store s in month t . The expenditure score for store s in month t can be written as

$$Expenditure\ Score_{st} = \left[\sum_{c \in C_{healthful}} (sh_{cst} - sh_{c\bar{h}}^{CNPP})^2 | sh_{cst} < sh_{c\bar{h}}^{CNPP} + \sum_{c \in C_{unhealthful}} (sh_{cst} - sh_{c\bar{h}}^{CNPP})^2 | sh_{cst} > sh_{c\bar{h}}^{CNPP} \right]^{-1}$$

where c again indexes the CNPP food categories. sh_{cst} is the representative household's predicted percent expen-

²⁴The patterns are qualitatively similar for years from 2006 to 2011.

ditures on category c in store s in month t , calculated as

$$sh_{cst} = \sum_{u \in U_{cst}} \left(\frac{v_{ut}}{\sum_{u \in U_{st}} v_{ut}} \right)$$

Here, U_{cst} is the set of CNPP-category c UPCs with positive sales in store s in month t , $U_{st} = \{U_t | v_{ust} > 0\}$ is the set of UPCs with positive sales in store s in month t , and $v_{ut} = \sum_{s \in S_t} v_{ust}$ is the total value of sales of UPC u across all stores S_t in the national RMS sample in month t .²⁵ We look at the distance from this representative household’s category expenditure share from the CNPP’s recommended category c expenditure share for a “typical” household, consisting of a male of age 19-50, a female of age 19-50, one child of age 6-8, and one child of age 9-11. We denote this modal household recommended expenditure share in category c as sh_{ch}^{CNPP} .²⁶

Similarly, the nutrient score for store s in month t can be written as

$$\begin{aligned} Nutrient\ Score_{est} = & \left[\sum_{j \in J_{healthful}} \left(\frac{nutr_{jst} - nutr_j^{FDA}}{nutr_j^{FDA}} \right)^2 | nutr_{jst} < nutr_j^{FDA} \right. \\ & \left. + \sum_{j \in J_{unhealthful}} \left(\frac{nutr_{jst} - nutr_j^{FDA}}{nutr_j^{FDA}} \right)^2 | nutr_{jst} > nutr_j^{FDA} \right]^{-1} \end{aligned}$$

Here, $nutr_j^{FDA}$ is the FDA’s recommendation for the per calorie consumption of nutrient j and $nutr_{jst}$ is the per calorie amount of nutrient j that we expect to be purchased by a representative household in store s in month t , calculated as

$$nutr_{jst} = \left(\frac{\sum_{u \in U_{st}} v_{ut} n_u^j}{\sum_{u \in U_{st}} v_{ut} cal_u} \right)$$

where n_u^j is the amount of nutrient j in UPC u and cal_u denotes the quantity of calories in UPC u .²⁷

In Figure 3, we see that the extent of the variation in the nutritional quality of available products across stores depends on which measure of food quality we are using. There is almost no variation in the average expenditure scores of stores across neighborhoods with different socioeconomic characteristics.²⁸ There is more variation in nutrient scores, but it is still limited compared to the degree of variation that we observed across households with different socioeconomic characteristics. Still, we see that stores in high-income neighborhoods stock foods with higher nutrient scores than stores in low-income neighborhoods. Store nutrient scores are lowest in neighborhoods with low income and low education.²⁹

²⁵Note that neither of the store-level indexes defined above use any information on the quantity of sales of products in a store-month. We use national weights, rather than store-sales weights, in order to capture the relative importance of products to a nationally-representative consumer rather than a store-specific representative consumer. Indexes based on store-sales weights will be biased towards the tastes of the customers visiting that store and, therefore, are going to be mechanically correlated with the demographics of the local community around a store. Using national weights, we are able to control for the relative importance of UPCs to the typical consumer, without introducing this local bias.

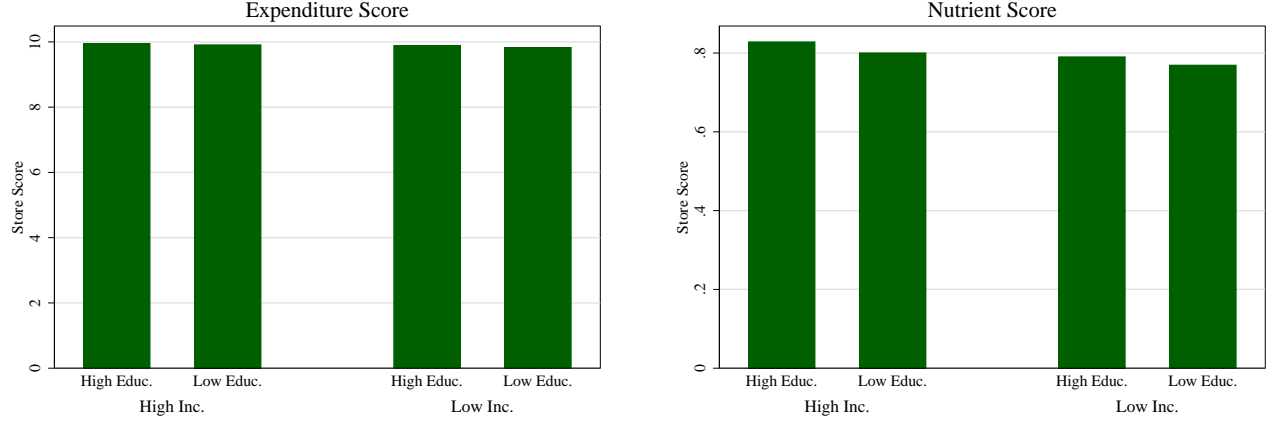
²⁶We drop store expenditure scores that are more than twice the distance between the 90th and 50th percentiles (less than 0.5% of store-month scores).

²⁷As with the expenditure scores, we drop store nutrient scores that are more than twice the distance between the 90th and 50th percentiles (approximately 5% of store-month scores).

²⁸The lack of differences in the average expenditure score across stores in different neighborhoods does not imply that expenditure scores do not vary across stores at all. The differences in expenditure scores are actually quite pronounced when we look across store type instead of store location. Nielsen categorizes each store in the RMS data into one of four channels: food, convenience, drug, or mass merchandise. Looking to Figure A.1 in the appendix, we see that food stores have higher expenditure scores than convenience stores, for example.

²⁹We see similar results at the neighborhood level. We calculate kernel densities of the healthfulness and nutrient scores of the stores around each census tract centroid and find very little variation in the expenditure scores and only a small amount of variation in the nutrient scores of stores in the vicinity of high- and low-socioeconomic status census tracts.

Figure 3: Expenditure and Nutrient Scores Across Stores: Available Products



Notes: The figure above presents mean store expenditure and nutrient scores across tracts with different socioeconomic statuses. Tracts are considered high-income if their mean household income falls above the median level across all tracts and low income otherwise. Tracts are considered high-education if their share of college-educated residents falls above the median across all tracts and low-education otherwise. These results are for January 2010; they are representative of the other months in the Kilts sample.

We formalize these results by regressing the store-level nutrition availability indexes on store-specific, market-level variables in Table 2. In Figures 2 and 3, we define neighborhood socioeconomic characteristics at the tract level. Here, we treat space continuously, looking at how the socioeconomic statuses of residents in the general vicinity of a store covaries with the nutritional quality of the products available in that store. We measure the average socioeconomic status in the vicinity of a store with the kernel densities of median income and college-educated share of the tracts surrounding a store, using a Gaussian kernel with a bandwidth of 20km.³⁰ Letting L denote the set of census tracts, p_l the socioeconomic characteristic in census tract l in 2010, and d_{sl} the distance (in km) between store s and the centroid of census tract l , the relevant socioeconomic kernel density around store s is given by $\sum_{l=1}^L p_l w_{sl} / \sum_{l=1}^L w_{sl}$ where $w_{sl} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{d_{sl}}{20} \right)^2}$.

In the first and fourth columns of Table 2, we regress the log expenditure and nutrient scores for each store in each month on kernel densities of household income and education. The results confirm what we saw in the bar charts: the nutrient scores of stores are correlated with the socioeconomic status of local residents, whereas the expenditure scores are not. Stores in wealthier and more educated neighborhoods tend to offer a range of products whose macro-nutrient content, on the whole, better accords with the FDA recommendations. In the subsequent columns, we control for DMA (a Nielsen market definition of similar geographic scope to a Metropolitan Statistical Area), store chain, and chain interacted with DMA. Interestingly, the differences in nutrient scores across neighborhoods with different college-educated shares persist both when we look within local markets and within chains in these markets. Chains appear to be changing their product offerings across stores even within the same DMA.

³⁰Our results are robust to using bandwidths of 5km, 10km, and 40km.

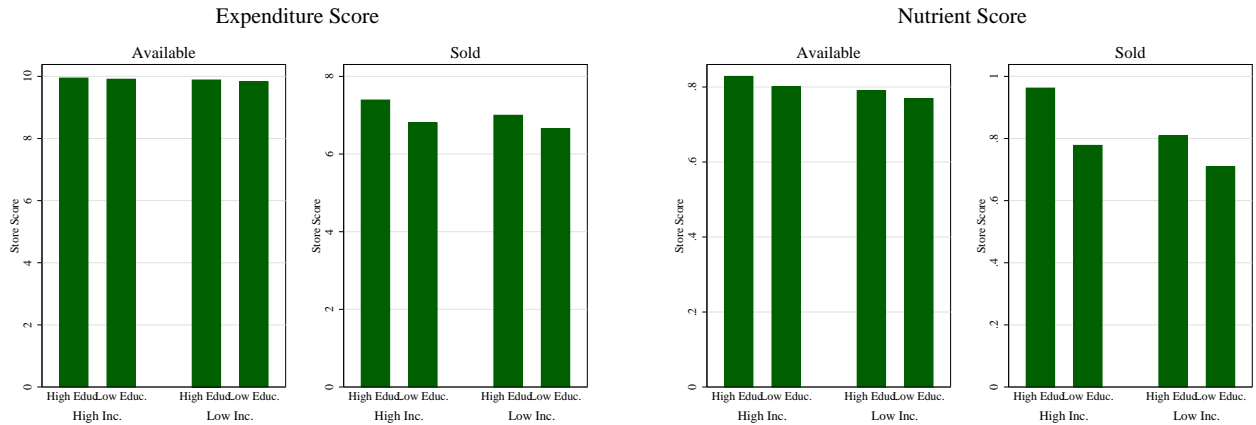
Table 2: Neighborhood Characteristics and Nutritional Quality of Product Offerings

	Dep Var: Ln(Exp Score, Natl. Wgts)			Dep Var: Ln(Nutr Score, Natl. Wgts)		
	(1)	(2)	(3)	(4)	(5)	(6)
Median Household Income Dens	0.0171 (0.0093)	0.161*** (0.024)	-0.00523 (0.012)	0.0874*** (0.0083)	0.184*** (0.019)	0.0239*** (0.0060)
College-educated Share Dens	0.00603 (0.011)	0.00435 (0.018)	0.0539*** (0.011)	0.0320*** (0.0084)	-0.0461** (0.014)	0.0319*** (0.0052)
R^2	0.092	0.200	0.707	0.152	0.203	0.803
FEs	None	DMA	DMAxCh	None	DMA	DMAxCh
Obs	1239023	1239023	1239023	1239021	1239021	1239021

Notes: Observations are at the store-month level and all regressions include year-month fixed effects. Standard errors are clustered by store. All variables are standardized. DMA refers to designated market area and DMAxCh is the intersection of DMA and store chain.

We demonstrated that stores surrounded by high-socioeconomic status neighborhoods tend to offer products that are at least as healthy, if not healthier, than those available in low-socioeconomic status neighborhoods. Recall that the store healthfulness and nutrient scores we used in this analysis employed national sales weights and, therefore, reflect the scores that would be achieved by a consumer purchasing all of the products that were ever sold in the store in a given month, allocating their expenditure or calories between products in the same proportion as we observe these products being sold in the national RMS sample in the same month. It is worthwhile noting that we observe much larger disparities in scores based instead on the expenditure and calorie allocations that we observe in the sales of each store. Figure 4 compares the differences that we observe in the healthfulness of the products available to the differences that we observe in the healthfulness of the typical bundle of products actually purchased across neighborhoods with different socioeconomic compositions.

Figure 4: Expenditure and Nutrient Scores Across Stores: Available versus Sold



Notes: The figure above presents mean store expenditure and nutrient scores across tracts with different socioeconomic statuses. Tracts are considered high income if their mean household income falls above the median level across all tracts and low income otherwise. Tracts are considered high education if their share of college-educated residents falls above the median across all tracts and low education otherwise. In each subfigure (expenditure score/nutrient score), the plot on the left ("available") replicates the availability indexes presented in Figure 3 above, while the plots on the right ("sold") reflect the sales-weighted scores that are calculated based on both the types and proportions of products sold in stores. These results are for January 2010; they are representative of the other months in the Kilts sample.

The relative magnitudes of the differences in the healthfulness of products sold and available in stores across socioeconomically diverse neighborhoods indicate that it is unlikely that differences in product availability drive the observed differences in sales. We confirm this in Table 3 where we see that stores in higher income and more educated neighborhoods tend to sell more healthful bundles of products, even controlling for the availability of

products. In fact, adding the availability control has almost no impact on the correlation between store sales-weighted expenditure scores and neighborhood characteristics. This is not surprising given the small amount of variation we observe in the national sales-weighted (availability) expenditure scores in Figure 3 and Table 2 above.

In general, these results suggest that nutritional disparities in the products sold across stores cannot be explained by any constraint imposed by differences in the availability of the nutritious food products. This does not imply that access, more broadly defined, cannot explain the differences in product sales. Stores with identical expenditure or nutrient scores for the products offered may provide different levels of access to nutritious foods because one store offers lower quality versions of these products at higher prices. The manner in which healthful products are presented, including their shelf space and department cleanliness, may also make these products relatively less attractive in certain stores. Our analysis below will control for all differences in access across stores in order to obtain an upper bound on the role that these factors play in explaining socioeconomic differences in household purchases.

Table 3: Neighborhood Characteristics and Nutritional Quality of Store Sales

	Dep Var: Ln(Exp. Score, Store Weights)		Dep Var: Ln(Nutr Score, Store Weights)	
	(1)	(2)	(3)	(4)
Median Household Income Dens	0.115*** (0.011)	0.104*** (0.0080)	0.108*** (0.0096)	0.0317*** (0.0044)
College-educated Share Dens	-0.112*** (0.011)	-0.116*** (0.0081)	-0.0198* (0.0100)	-0.0478*** (0.0046)
Ln(Relevant Score, Natl. Wgts)		0.643*** (0.020)		0.876*** (0.0032)
R^2	0.024	0.359	0.044	0.650
Obs	1239023	1239023	1239022	1239021

Notes: Observations are at the store-month level and all regressions include year-month fixed effects. Standard errors are clustered by store. All variables are standardized. In columns (1) and (2), relevant score is expenditure score; in columns (3) and (4), relevant score refers to nutrient score.

4 Theoretical Framework

We have demonstrated both that there are large socioeconomic disparities in the nutritional content of household grocery purchases and that there are spatial disparities in the concentration and offerings of retail outlets. The direction of causality here is undetermined. It is plausible that the disparities in nutritional consumption are due entirely to the fact that lower income and less educated households have access to different products than higher income and more educated households (that is, any systematic variation in the content of grocery purchases would disappear if all households lived in the same location). It is also plausible that these spatial disparities are due to households sorting into locations where they have access to the food products they prefer to purchase or, more likely, that households sort by income and education into locations based on factors unrelated to their taste for grocery products (e.g., housing prices, proximity to employment opportunities) and spatial disparities in product availability arise because stores are catering to local demand. In reality, there are likely feedback effects between household demand and access.

We now introduce a simple theoretical framework in which local tastes and retail costs both influence the spatial distribution of retail food products. We use this framework to motivate the empirical approach we take to identify the causal link between access and the nutritional quality of household purchases in Section 5.

The model describes an economy with many locations populated by an equal number of immobile households. Households can be of either high or low socioeconomic status, with locations differing in the proportion of their population from each socioeconomic group. Two types of foods, healthful and unhealthful, are freely traded between locations on a wholesale market. Healthful foods take more labor to produce than unhealthful foods, so they sell at a higher wholesale price. Retailers in each location pay a fixed cost to purchase the technology to produce a differentiated food product from the relevant input. Only healthful (unhealthful) food inputs can be converted into healthful (unhealthful) food products. The production of a single unit of a differentiated food product requires a single unit of the relevant freely-traded input plus a single unit of shelf space. For simplicity, we assume that households are immobile and can only shop in retail stores in their location.³¹ Retail is monopolistically competitive, so the number of healthful and unhealthful food products a store stocks will depend on the demand for each type of product in the retailer's location.

We demonstrate two mechanisms through which a correlation between the spatial distribution of healthful foods and the spatial distribution of socioeconomic class can emerge. First, we allow for high-socioeconomic individuals to have a stronger taste for healthful food products than low-socioeconomic individuals.³² Assuming that there are fixed costs in the distribution of differentiated food products, these heterogeneous tastes and the spatial sorting of households by demographic class will result in firms in high-socioeconomic neighborhoods offering more healthful food products than firms in low-socioeconomic neighborhoods. The second mechanism works through supply, rather than demand. The assumption that healthful foods sell at a higher wholesale price than unhealthful food products, along with the assumed fixed shelf-space requirement, implies a complementarity between the healthfulness of the food products a retailer sells and the rental cost of shelf space in the market where they are located. If we further assume that retail rents are increasing in the high-socioeconomic share of the neighborhood population, firms in high-socioeconomic share locations will have a comparative advantage in the production of high-quality goods.

The theory delivers two key results. First, it confirms that the socioeconomic disparities in the availability and purchases of healthful food products are overdetermined. Each mechanism alone is sufficient to generate the socioeconomic disparities in the healthfulness of food purchases across households and in the healthfulness of food availability across neighborhoods documented in Section 3. Second, the theory identifies an important distinction between the two mechanisms. Conditional on household location, the correlation between the healthfulness of household food purchases and socioeconomic status is due solely to differences in tastes across households. If the spatial disparities in nutritional consumption are entirely due to preference externalities, the model predicts that the socioeconomic disparities in nutritional consumption within a location should be as large as they are between locations. If the estimated disparities within locations are smaller than those across locations, then the difference between the two can be interpreted as an upper bound for the role that access, as opposed to tastes, plays in explaining the socioeconomic disparities in nutritional consumption across households. That is, if retail environments were equalized across locations, we could not expect the resulting nutritional gap between high- and low-socioeconomic households to be any less than the estimated disparity between high- and low-socioeconomic

³¹This assumption is innocuous for the purpose of distinguishing the role access plays in determining household's grocery purchases. Household mobility would be relevant in considering counterfactuals, however, since households may migrate across locations in response to changes in economic activity.

³²To keep the model tractable, we abstract from other reasons why households of different socioeconomic characteristics but the same choice set might purchase different products. For example, we assume that all households have the spending ability and, more importantly, can purchase products in continuous quantities. In doing so, we rule out the possibility that low-socioeconomic status households may purchase fewer healthful food products because they are, in general, available only in discrete quantities at high prices and, therefore, do not fit within a more constrained budget. To the extent that these factors generate differences in demand across socioeconomic groups facing the same choice set, they could be considered complementary to the heterogeneous taste mechanism that we use here.

households who currently live in the same retail environment.

The rest of this section provides the details of the model. The reader who is satisfied with the intuition alone may proceed directly to Section 5.

4.1 Set-up

There are M locations indexed by l . Each location l has a population of size N composed of heterogeneous individuals whose socioeconomic status, indexed by h , can take one of two values, low (L) or high (H). We rank locations by their share of high-socioeconomic status households, with higher l locations having larger shares of high-socioeconomic status households. We assume that the share of high-socioeconomic status households in a neighborhood is exogenously determined.

4.1.1 Demand

Consider a representative consumer for socioeconomic status h . For simplicity, we assume that the consumer is immobile and can only shop at retail stores in his location. The preferences of the representative consumer are given by a nested-CES utility function over a continuum of grocery varieties indexed by u . The nests are defined by the healthfulness of the product u , denoted by $q(u) \in \mathbb{Q}$. Let \mathbb{U}_q denote the set of products of the same healthfulness. A consumer of status h in location l will select their grocery purchases, $x(u)$, to maximize utility over the products available in location l , \mathbb{U}_l , subject to a budget constraint. The budget constraint is defined by local grocery prices, $p(u, l)$, and the per-capita grocery expenditure, Y , which we normalize to one. That is,

$$\max_{x(u)} X_h = \left[\int_{q \in \mathbb{Q}} \alpha_h(q) \left(\int_{u \in \mathbb{U}_q} x(u)^{\rho_w} du \right)^{\frac{\rho_a}{\rho_w}} \right]^{\frac{1}{\rho_a}} \quad \text{subject to} \quad \sum_{u \in \mathbb{U}_l} p(u, l)x(u) \leq Y = 1$$

where $\rho_a \in (0, 1)$ reflects the degree of perceived horizontal differentiation between varieties of different healthfulnesses and $\rho_w \in (0, 1)$ reflects the degree of perceived horizontal differentiation between varieties of the same healthfulness. where we assume that $\rho_a > \rho_w$. The elasticity of substitution between varieties of different healthfulnesses and between varieties of the same healthfulness can be expressed as $\sigma_a = 1/(1 - \rho_a)$ and $\sigma_w = 1/(1 - \rho_w)$, respectively. We assume that varieties are also differentiated vertically by their degree of healthfulness, so the amount of utility a consumer with socioeconomic status h gets from a unit of consumption of a given variety is scaled up (or down) by their taste for healthfulness, denoted by $\alpha_h(q(u)) > 0$.

The demand of a status h consumer in market l can be characterized by their expenditure share on product u :

$$x_h(u, l) = \left(\frac{p(u, l)}{P(q, l)} \right)^{-\sigma_w} \left(\frac{P(q, l)/\alpha_h(q)}{P_h(l)} \right)^{-\sigma_a}$$

where $P(q, l)$ denotes the price index for products of healthfulness q available in market l ($\mathbb{U}_{q,l} = \mathbb{U}_q \cap \mathbb{U}_l$), defined as

$$P(q, l) = \left[\int_{u \in \mathbb{U}_{q,l}} (p(u, l))^{1-\sigma_w} \right]^{\frac{1}{1-\sigma_w}}$$

and $P_h(l)$ denotes the aggregate taste-adjusted price index that consumers of type h face in market l , defined as

$$P_h(l) = \left[\int_{q \in \mathbb{Q}} \left(\frac{P(q, l)}{\alpha_h(q)} \right)^{1-\sigma_a} \right]^{\frac{1}{1-\sigma_a}}$$

A household h 's total expenditure on all varieties of quality q is given by

$$x_h(q, l) = \left(\frac{P(q, l)/\alpha_h(q)}{P_h(l)} \right)^{-\sigma_a}$$

The relative expenditure of high-socioeconomic households to low-socioeconomic households on products of the same healthfulness in the same location can be expressed as

$$\frac{\partial x_H(q, l)/x_L(q, l)}{\partial q} = \sigma_a \left(\frac{\alpha_H(q)}{\alpha_L(q)} \right)^{\sigma_a} \left(\frac{P_H(l)}{P_L(l)} \right)^{\sigma_a} \left(\frac{\alpha'_H(q)}{\alpha_H(q)} - \frac{\alpha'_L(q)}{\alpha_L(q)} \right)$$

High-socioeconomic households will spend relatively more than low-socioeconomic households on healthful products when $\frac{\alpha'_H(q)}{\alpha_H(q)} > \frac{\alpha'_L(q)}{\alpha_L(q)}$ for all q . We assume that this inequality holds in all cases where tastes vary with socioeconomic status.³³

4.1.2 Supply

In order to distribute x units of a food product of healthfulness q to a neighborhood with a λ_l share of high-socioeconomic residents, we assume that a firm must incur a fixed cost f ; a per unit wholesale cost that can vary with product healthfulness, $w(q)$; and a per unit shelf-space cost that can vary with the share of high-socioeconomic residents, $s(\lambda_l)$. To reflect higher rents in higher-socioeconomic neighborhoods, we assume that shelf-space costs are increasing in the share of high-socioeconomic status individuals living in the location. We denote the total marginal cost of retail by $c(q, l) = w(q) + s(\lambda_l)$. We assume that there are no economies of scope, so each retailer sells only one variety in any one location l . Taking the behavior of competitors as given, the optimal price charged by a firm producing variety u of healthfulness q in location l is the price that maximizes profits. That is, the firm solves the following problem

$$\max_{p(u, l)} \pi(u, l) = (p(u, l) - c(q, l)) x(u, l) - f$$

where $x(u, l)$ denotes the demand for variety u in location l , with

$$x(u, l) = \lambda_l x_H(u, l) + (1 - \lambda_l) x_L(u, l)$$

where we have normalized the population in each location to one. For all varieties u of quality q sold in location l , the optimal pricing strategy is a proportional mark-up over marginal cost:

$$p(u, l) = \frac{c(q, l)}{\rho_w}$$

³³To keep the model tractable, we abstract from other reasons why households of different socioeconomic characteristics but the same choice set might purchase different products. For example, we assume that all households have the spending ability and, more importantly, can purchase products in continuous quantities. In doing so, we rule out the possibility that low-socioeconomic status households may purchase fewer healthful food products because they are, in general, available only in discrete quantities at high prices and, therefore, do not fit within a more constrained budget. To the extent that these factors generate differences in demand across socioeconomic groups facing the same choice set, they could be considered complementary to the heterogeneous taste mechanism that we use here.

We can use this optimal price to rewrite the price index for quality q in location l as

$$P(q, l) = (N(q, l))^{\frac{1}{1-\sigma_w}} \left(\frac{c(q, l)}{\rho_w} \right) \quad (1)$$

where $N(q, l)$ is the number of varieties of healthfulness q distributed to location l . The price index for household type h in location l is

$$P_h(l) = \left[\int_{q \in \mathbb{Q}} \left(\frac{P(q, l)}{\alpha_h(q)} \right)^{1-\sigma_a} \right]^{\frac{1}{1-\sigma_a}} = \frac{1}{\rho_w} \left[\int_{q \in \mathbb{Q}} \left(\frac{(N(q, l))^{\frac{1}{1-\sigma_w}} c(q, l)}{\alpha_h(q)} \right)^{1-\sigma_a} \right]^{\frac{1}{1-\sigma_a}}$$

Therefore, the quantity of sales of any firm selling a variety of healthfulness q in location l is given by

$$x(q, l) = (N(q, l))^{\frac{\sigma_w + \sigma_a}{1-\sigma_w}} \left(\frac{c(q, l)}{\rho_w} \right)^{\sigma_a} [\lambda_l (\alpha_H(q) P_H(l))^{\sigma_a} + (1 - \lambda_l) (\alpha_L(q) P_L(l))^{\sigma_a}] \quad (2)$$

4.1.3 Equilibrium

We assume that there is free entry into retailing, so active firms earn zero profits. This implies that the scale of firm sales in any given market is given by

$$x(q, l) = \frac{f}{c(q, l)} (\sigma_w - 1) \quad (3)$$

4.2 Comparative Statics

4.2.1 Equilibrium Pattern of Product Availability and Consumption Across Locations

Taken together, the zero profit condition (Equation (3)), the aggregate demand condition (Equation (2)), and the healthfulness-location-specific price index (Equation (1)), implicitly defines the number of varieties of healthfulness q in each location l as a function of the fixed and marginal costs of producing each variety, the local share of households in each socioeconomic class, and the model parameters:

$$N(q, l) = \underbrace{\Gamma (c(q, l))^K}_{Cost} \underbrace{[\lambda_l (\alpha_H(q) P_H(l))^{\sigma_a} + (1 - \lambda_l) (\alpha_L(q) P_L(l))^{\sigma_a}]^{\frac{1-\sigma_w}{\sigma_w + \sigma_a}}}_{Demand} \quad (4)$$

where $\Gamma = \left[f (\sigma_w - 1) \left(\frac{\sigma_w - 1}{\sigma_w} \right)^{\sigma_a} \right]^{\frac{\sigma_w - 1}{\sigma_w + \sigma_a}} > 0$ and $K = \frac{(1 - \sigma_w)(1 + \sigma_a)}{\sigma_a} < 0$. Given the distribution of socioeconomic classes across locations and the retail technology, the pattern of product availability is determined by two forces, each reflected by an individual term in the above expression for product availability. The first, labeled *Cost*, reflects the role that costs play in determining the healthfulness distribution in different locations. The second, labeled *Demand*, reflects the role played by differences in tastes across socioeconomic groups combined with differences in the share of socioeconomic classes in each location's population.

We now demonstrate that each of these mechanisms could individually explain the qualitative patterns that we observe in product availability across neighborhoods and purchases across households. We are interested in showing that the number of healthful, relative to unhealthful, varieties available in a location is increasing in the share of high-socioeconomic households in the location (*i.e.*, that $\frac{N(q, l)}{N(q', l)} > \frac{N(q, l')}{N(q', l')}$ for $\lambda > \lambda'$). If tastes are weakly supermodular in quality and household socioeconomic status, high-socioeconomic status households will

spend at least as much on high-quality food products as low-socioeconomic status households in the same location. Therefore, if the healthfulness of available products is increasing in the share of high-socioeconomic households in a neighborhood, it follows that high-socioeconomic households will spend more on healthful food products. Even if high-socioeconomic and low-socioeconomic households share the same tastes, all households will spend more on healthful foods in locations where more of these are available. Since high-socioeconomic status households are, by definition, disproportionately located in high-socioeconomic status locations, on average high-socioeconomic households will spend more on healthful food products.

We start by turning both mechanisms off. That is, we assume that **tastes are identical** across consumers, *i.e.*, $\alpha_H(q) = \alpha_L(q) = \alpha(q)$ for all q , and that **wholesale costs are equal** across products of different healthfulnesses, *i.e.* $w(q) = w$ for all q . If wholesale costs are equal across products, then the healthfulness of the varieties available in each location will be determined by the taste shifter, $\alpha(q)$:

$$N(q, l) = \Gamma (c(l))^K (\alpha(q)P(l))^{1-\sigma_w} \quad (5)$$

Since tastes are assumed to be identical across consumers, the distribution of healthfulness of available varieties will be identical across locations. To see this, note that the relative number of varieties of two healthfulness levels, q and q' , in location l can be written as the ratio of the common taste shifter for varieties of quality q relative to q' . That is,

$$\frac{N(q, l)}{N(q', l)} = \left(\frac{\alpha(q)}{\alpha(q')} \right)^{1-\sigma_w} \quad (6)$$

Since tastes are identical across households and the distribution of healthful products available is identical across locations, Marshallian demand must be also identical across households, regardless of their socioeconomic status or location.

If we assume that **tastes are identical** (and, for simplicity, do not vary with product quality), *i.e.* $\alpha_H(q) = \alpha_L(q) = \alpha$ for all q , but allow **wholesale costs to vary** with healthfulness, then the zero profit condition reduces to

$$N(q, l) = \Gamma (c(q, l))^K (\alpha P(l))^{1-\sigma_w} \quad (7)$$

Taking the derivative with respect to healthfulness q and location l and imposing that retail costs are equal to the sum of wholesale and shelf costs, *i.e.*, $c(q, l) = w(q) + s(\lambda_l)$, we see that as long as wholesale costs are increasing in quality and shelf-space costs are increasing in λ_l , the healthfulness- and location-specific variety counts are supermodular in quality (q) and the high-socioeconomic share of households (λ_l):

$$\frac{\partial N(q, l)}{\partial q \partial \lambda_l} = \Gamma K (\alpha P(l))^{1-\sigma_w} \frac{w'(q) s'(\lambda_l)}{(w(q) + s(\lambda_l))^{2-K}} > 0 \text{ for } w'(q), s'(\lambda_l) > 0.$$

This result implies that high-socioeconomic status households are more likely to live in locations with a greater variety of healthful food products. The ratio of the price of healthful relative to unhealthful food products will be identical across locations, so households in locations with a greater variety of healthful food products available will purchase relatively more of these products. As a result, we expect to see high-socioeconomic status

households spending more on healthful food products, on average, even if they have the same preferences as low-socioeconomic status households. That is, socioeconomic disparities in access to healthful and unhealthful food products alone can generate socioeconomic disparities in household purchases.

If we instead assume that **the cost functions are identical** across locations, *i.e.* $c(q, l) = c(q)$ for all l , but allow for **tastes to vary** with socio-economic status, the zero profit condition becomes:

$$N(q, l) = \Gamma(c(q))^K [\lambda_l (\alpha_H(q) P_H(l))^{\sigma_a} + (1 - \lambda_l) (\alpha_L(q) P_L(l))^{\sigma_a}]^{\frac{1 - \sigma_w}{\sigma_w + \sigma_a}} \quad (8)$$

To characterize how the quality distribution is determined by demand, we start by considering the simplest case and compare two locations, l and l' , which are populated entirely by high-socioeconomic and low-socioeconomic consumers, respectively. The ratio of the product counts across the two locations at any given quality level q is given by

$$\frac{N(q, l)}{N(q, l')} = \left(\frac{\alpha_H(q) P_H(l)}{\alpha_L(q) P_L(l')} \right)^{\frac{\sigma_a(1 - \sigma_w)}{\sigma_w + \sigma_a}} \quad (9)$$

since $\lambda_l = 1$ and $\lambda_{l'} = 0$. Taking the derivative of this function with respect to healthfulness we see that the ratio of varieties available for a given healthfulness level across the two locations will be increasing in healthfulness as long as $\frac{\alpha'_L(q)}{\alpha_L(q)} < \frac{\alpha'_H(q)}{\alpha_H(q)}$. This is the same condition required for the relative expenditure share of high-socioeconomic to low-socioeconomic households to be increasing in quality:

$$\frac{\partial \frac{N(q, l)}{N(q, l')}}{\partial q} = A \frac{N(q, l)}{N(q, l')} \left(\frac{\alpha'_H(q)}{\alpha_H(q)} - \frac{\alpha'_L(q)}{\alpha_L(q)} \right) > 0 \text{ for } \frac{\alpha'_H(q)}{\alpha_H(q)} > \frac{\alpha'_L(q)}{\alpha_L(q)} \quad (10)$$

for $A = \left(\frac{\sigma_a(\sigma_w - 1)}{\sigma_w + \sigma_a} \right) < 0$.

Now, consider two locations with intermediate, but non-equal, shares of high-socioeconomic status households. When costs are identical across locations, the zero profit condition implies that the scale of firms producing varieties of the same healthfulness is also identical across locations. The number of varieties available at each healthfulness level will be determined solely by demand for products at that healthfulness level. Since demand for healthful varieties is increasing in socioeconomic status, and all households earn the same income, we must therefore have that locations with more high-socioeconomic status households can support a greater variety of healthful food products.

4.2.2 Upper Bound for the Impact of Access on Consumption

We have demonstrated that two separate forces can each individually explain the distribution of product availability and consumption that we observe across locations. The correlation between access and household purchases demonstrated in the previous literature, however, is insufficient to determine the role that differences in access play in driving differences in consumer behavior (or vice versa). In what follows, we show that by comparing the differences in household purchases across locations to those within locations, we can identify an upper bound on the role that access plays in generating these differences. The critical result is that demand alone determines differences in purchases across households with different socioeconomic statuses in the same location.

Both access and tastes could be at play in generating the socioeconomic disparities that we observe in purchases

across households living in different locations. To see this, note that the expenditures of a household of socio-economic status h on products of a given healthfulness q are determined both by their taste for that healthfulness $\alpha_h(q)$, and by the price index of products of that healthfulness in their location:

$$x_h(q, l) = (\alpha_h(q))^{\sigma_a} \left(\frac{P(q, l)}{P_h(l)} \right)^{1-\sigma_a} \quad (11)$$

We saw above that high-socioeconomic status individuals purchase more healthful food products either because there are more of these products available in the locations where they live and/or because they have a stronger taste for these products. To see this mathematically, note that the average expenditure share of healthfulness q varieties for high-socioeconomic relative to low-socioeconomic status individuals living across two locations, l and l' , is given by

$$\begin{aligned} \frac{x_H(q)}{x_L(q)} &= \left(\frac{\lambda_l x_H(q, l) + \lambda_{l'} x_H(q, l')}{(1 - \lambda_l) x_L(q, l) + (1 - \lambda_{l'}) x_L(q, l')} \right) \left(\frac{2 - \lambda_l - \lambda_{l'}}{\lambda_l + \lambda_{l'}} \right) \\ &= \underbrace{\left(\frac{\alpha_H(q)}{\alpha_L(q)} \right)^{\sigma_a}}_{Tastes} \underbrace{\left(\frac{\lambda_l \left(\frac{P(q, l)}{P_H(l)} \right)^{1-\sigma_a} + \lambda_{l'} \left(\frac{P(q, l')}{P_H(l')} \right)^{1-\sigma_a}}{(1 - \lambda_l) \left(\frac{P(q, l)}{P_L(l)} \right)^{1-\sigma_a} + (1 - \lambda_{l'}) \left(\frac{P(q, l')}{P_L(l')} \right)^{1-\sigma_a}} \right)}_{Availability} \left(\frac{2 - \lambda_l - \lambda_{l'}}{\lambda_l + \lambda_{l'}} \right) \quad (12) \end{aligned}$$

The first term reflects taste differences alone. The second term reflects differences in access that, as we outlined above, could be the result of either firms catering to local tastes or to supply-side factors, such as the complementarities between healthfulness and local distribution costs proposed above. These differences in local product availability are reflected through the local price indexes, with $P(q, l)$ decreasing in the number of healthfulness q varieties that are available in location l . There are relatively more healthful varieties available in a location l where there are more high-socioeconomic status individuals, so the local healthfulness q price index will be lower, relative to the overall price index a household faces in a location ($P_H(l)$ or $P_L(l)$), in high- λ_l locations relative to locations with a lower share of high-socioeconomic status residents. This correlation implies that the numerator of the availability term is increasing in quality (since $1 - \sigma_a < 0$), whereas the denominator is falling in quality.

If we instead look at the average expenditure share of healthfulness q varieties for high-socioeconomic relative to low-socioeconomic status individuals living in the same location, l , this availability term no longer varies with product quality:

$$\frac{x_H(q, l)}{x_L(q, l)} = \left(\frac{\alpha_H(q)}{\alpha_L(q)} \right)^{\sigma_a} \left(\frac{P_L(l)}{P_H(l)} \right)^{1-\sigma_a} \quad (13)$$

Any systematic variation that we observe in the healthfulness consumed by high-socioeconomic relative to low-socioeconomic status individuals living in the same location must be attributed to tastes alone. In the context of this model, the within-location variation in healthfulness only provides a lower bound for the role of tastes, because tastes could also explain part (or all) of the differences in availability. This model is highly stylized, so there are various additional reasons why within-location socioeconomic disparities in healthfulness may reflect more than differences in tastes alone. Important factors that the model abstracts from include the mobility of both products and households between locations, unobserved heterogeneity in tastes across households within the same socioeconomic class, and differences in the mobility of households and the availability of products within locations. We will address each of these below, but it is worth noting that these biases will tend to lead us to further overestimate the role of product availability in explaining the overall socioeconomic disparities in purchases.

5 Role of Access in Explaining Consumption Disparities

In order to determine the role of access, we examine the extent to which the nutritional quality of household purchases varies across households in different socioeconomic groups that live in the same location or shop in the same store. We assume that households living in the same census tract or shopping at the same store have access to the same choice set of stores or products, respectively. We attribute any systematic differences that we observe in purchases across socioeconomic groups, controlling for residential tract or store location, to differences in purchase decisions, as opposed to retail environments. Since household tastes may play a role in determining a given household's retail environment - that is, households sort into neighborhoods that provide access to the products that they prefer to consume and stores sort into locations with high demand for their products - our within-location estimates underestimate the role of household demand. Therefore, the difference in our across-location and our within-location estimates can be interpreted as an upper-bound for the extent of the existing socioeconomic nutritional disparities that are due to differential access.

5.1 Controlling for Location

In the analysis that follows, we control for access to see whether the nutritional disparities remain. In columns (1) and (4) of Table 4, we replicate the standardized regression analysis from columns (4) and (8) of Table 1 for the sample of households with non-missing county and census tract information. In subsequent columns, we add controls for household location, using either county or census tract fixed effects. Looking first to the results for the nutrient score, we see that the impact of income on healthfulness is reduced by approximately one third when we control for county fixed effects and again by another third when we control for census tract fixed effects. The relationship between education and the nutrient score, however, is more persistent: the coefficient on education remains surprisingly stable regardless of the access controls included. The results are quantitatively more pronounced for the nutrient score, although the results are qualitatively similar for both indexes. Differential access explains between one third to one half of the nutritional disparities across different income groups, but only 10% of the disparities across education groups.

Table 4: Consumer Characteristics and Nutritional Quality of Purchases: Controlling for Location

	Dep. Var.: Ln(Expenditure Score)			Dep. Var.: Ln(Nutrient Score)		
	(1)	(2)	(3)	(4)	(5)	(6)
Ln(Income)	0.0241*** (0.0014)	0.0210*** (0.0015)	0.0182*** (0.0015)	0.0888*** (0.0031)	0.0638*** (0.0032)	0.0445*** (0.0033)
Ln(Education)	0.202*** (0.0067)	0.196*** (0.0067)	0.175*** (0.0073)	0.634*** (0.015)	0.619*** (0.015)	0.566*** (0.016)
Observations	3274436	3274436	3274436	3274436	3274436	3274436
R^2	0.066	0.083	0.248	0.029	0.049	0.179
Location Cntrl	No	County	Tract	No	County	Tract

Standard errors in parentheses
 * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Observations are at the household-month level. Standard errors are clustered by household. All variables are standardized. All regressions include year-month fixed effects and controls for household demographics, including household size dummies, average head of household age, a dummy for marital status of household heads, dummies for households with either a female or male household head, a dummy for the presence of children, and dummies for whether the household reports being white, black, Asian, or Hispanic.

These results are visually depicted in Figures 5 and 6. The figures display the coefficients on income and education when the same analysis as shown in Table 4 is done using income and education dummies instead of

levels. The points in Figure 5 are the coefficient estimates on the income dummies in the specification without household location controls plotted against the relevant income levels. The solid line depicts the smoothed value of these estimates. The dashed lines reflect the smoothed kernel density of the coefficient estimates with county or tract controls. We see that for both the expenditure and the nutrient score, adding location controls dampens the correlation between income and nutritional quality. As before, the impact of location controls on the relationship between income and quality is more pronounced when quality is measured using the nutrient score. Looking to Figure 6, we see that the relationship between education and each measure of quality is more persistent. For both the expenditure and the nutrient score, the addition of county or census tract fixed effects does little to reduce the correlation between education and quality.

Figure 5: Income Effects with Geographic Controls

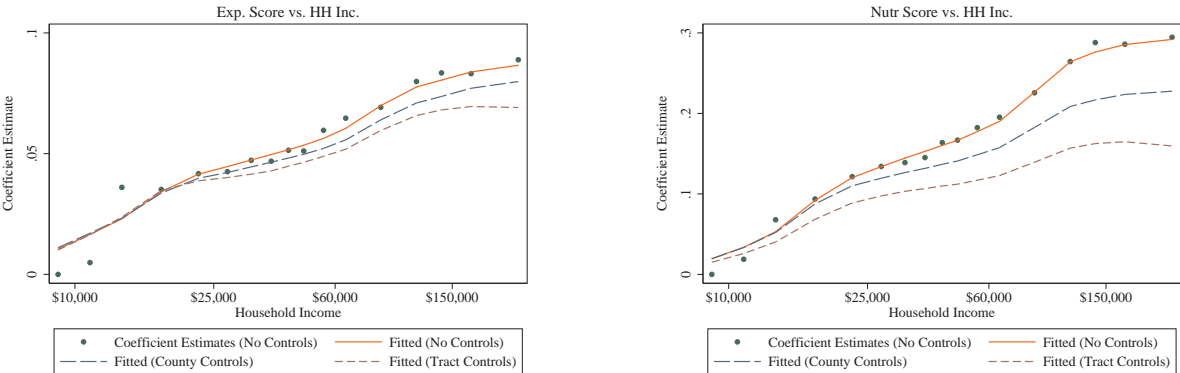
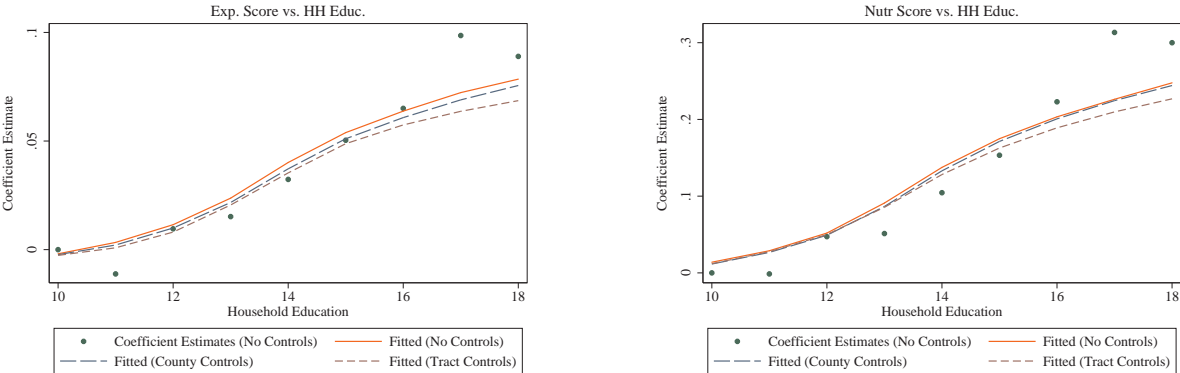


Figure 6: Education Effects with Geographic Controls



5.2 Controlling for Store

One concern with the within-location analysis is that households living in the same neighborhood may still have differential access. Even within a census tract, distance to retail outlets varies depending on the location of the household, and factors such as car ownership or proximity to public transportation may differentially impact the ability of households to travel to stores. To entirely remove the impact of access, we now turn to a within-store analysis. By including fixed effects for the store in which a purchase is observed, we can explore how the nutritional quality of purchases varies with the characteristics of households shopping in the same store. For this analysis,

data is at the household-store-month level.

Here, we calculate expenditure and nutrient scores for the purchases that households make in specific stores in each month. We then regress these household-store-month scores against household demographics, time fixed effects, and store controls. The results of this analysis, shown in Table 5, paint a similar picture as the within-location analysis presented above. The healthfulness of store-month-level household purchases is increasing in both income and education. When we control for access by looking at the variation within stores of the same type (i.e., grocery, drug, mass-merchandise, or convenience) the correlation between the nutrient score and income falls slightly, but not by a statistically significant margin. Looking to the expenditure score, we see that the correlation between the expenditure score and income actually increases when we control for store type.

In Section 3.2, we saw that the nutrient scores of the products available in stores vary even across stores in the same chain. Therefore, to hold the expenditure and nutrient scores of a household’s shopping environment fixed, we need to control for the exact store in which they are shopping. When we include store fixed effects, the correlation between the expenditure score and income falls slightly, while the correlation between the nutrient score and income falls to a little over 50% of its original value. This indicates that at least half of the observed disparity between the store-specific shopping bundles purchased by households with different incomes can be explained by tastes. We stress that the remaining component could be explained by either tastes or access - households may shop at different stores either because they are more accessible or because they offer products better suited to the household’s tastes. Access plays a smaller role in explaining the relationship between nutritional quality and household education. Moving from columns (1) to (4) and from columns (5) to (8), we see that the correlations between expenditure and nutrient scores and household education each only fall by around 10%.

Table 5: Consumer Characteristics and Nutritional Quality of Purchases: Controlling for Store

	Dep. Var.: Ln(Expenditure Score)				Dep. Var.: Ln(Nutrient Score)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ln(Income)	0.0282*** (0.0034)	0.0351*** (0.0026)	0.0317*** (0.0026)	0.0265*** (0.0021)	0.0804*** (0.0048)	0.0793*** (0.0045)	0.0589*** (0.0044)	0.0453*** (0.0038)
Ln(Education)	0.188*** (0.016)	0.186*** (0.012)	0.169*** (0.012)	0.165*** (0.0099)	0.470*** (0.023)	0.467*** (0.021)	0.448*** (0.021)	0.434*** (0.018)
Observations	4224012	4224012	4224012	4224012	4224012	4224012	4224012	4224012
R^2	0.021	0.358	0.371	0.438	0.022	0.121	0.135	0.185
Store Cntrl	No	Channel	Parent	Store	No	Channel	Parent	Store

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Observations are at the household-store-month level. Standard errors (in parentheses) are clustered by household. Observations are weighted by the dollar value of purchases represented in each household-stores-month observation.

These results are visually depicted in Figures 7 and 8, where we have replicated the regressions in Table 5 with household income and education dummies in place of levels and plotted the kernel of the coefficient estimates from the four different specifications. The points and solid lines represent the point estimates and kernel of these estimates from the specifications in columns (1) and (5) of Table 5. The dashed and dotted lines represent the kernel of the point estimates from columns (2) through (4) and (6) through (8), where we subsequently add more detailed controls for retail outlet. It is clear from Figure 7 that the healthfulness scores of household-store-specific bundles are not monotonic in income. The relationship becomes more monotonic once we control for channel fixed effects, indicating that the curvature of the regression coefficients without these controls is due to compositional differences in the types of stores where high- and low-income households shop. Overall, the inclusion of store

controls moves the correlation between income and nutritional quality closer to zero. For both the expenditure and the nutrient score, this result is most noticeable for the highest levels of income, where the correlation between income and quality is greatest in the absence of controls. Looking to Figure 8, we see that the relationship between education and quality is again more persistent. When we measure quality using the nutrient score, the inclusion of stores controls has barely any effect on the correlation between the nutrient score for household-store food purchases and education at all levels of education.

Figure 7: Income Effects with Store Controls

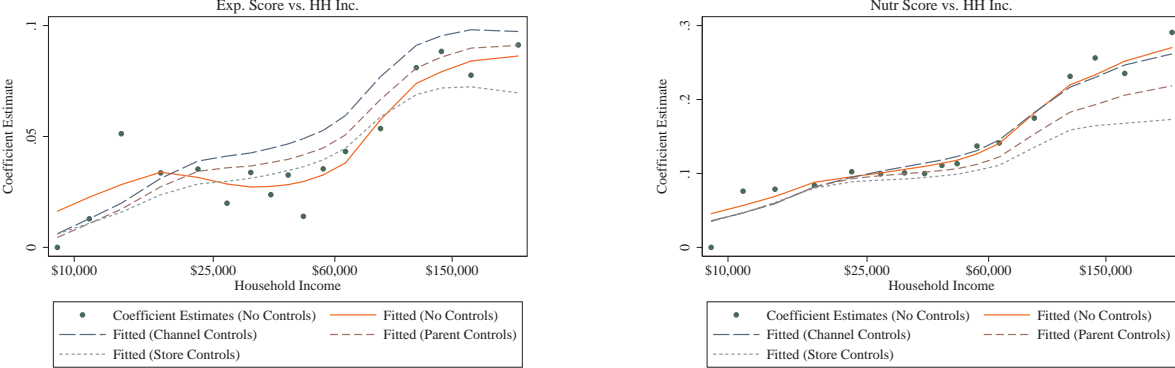
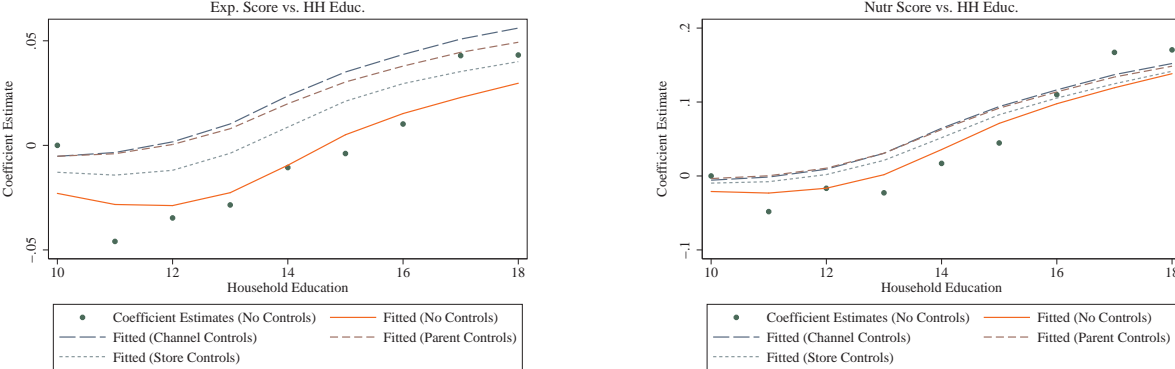


Figure 8: Education Effects with Store Controls



5.3 Discussion

In the analysis above, we find that, conditional on education, the correlation between household income and the nutritional content of households’ purchases is cut in half when we control for either their residential location or the store in which they are shopping. The effects of education conditional on income, however, are much more persistent: only 10% of the existing disparities in consumption across education groups can be attributed to differences in access. This suggests that over half of the socioeconomic disparities in nutritional consumption across income groups and nearly all of the socioeconomic disparities in nutritional consumption across education groups would remain even if the spatial disparities in access to nutritious foods were resolved.

The fact that socioeconomic disparities persist, even looking across households shopping in the same store, indicates that differences in demand across socioeconomic groups yield empirically relevant disparities above and

beyond those that could also be attributed to the sorting of households by income and education across residential locations or stores. This suggests that resolving disparities in access to healthful food products would not resolve these disparities, at least not in the short run. In the longer run, it is possible that improved access to healthful foods could impact demand indirectly by providing low-income and less educated households increased exposure to more healthful food products. Further analysis is required to understand which factors are most important in explaining why demand varies across socioeconomic groups shopping in the same stores.

The fact that the socioeconomic disparities diminish when we control for household location does not necessarily indicate that access alone explains this portion of the disparity. As discussed above, tastes could be reflected in household’s store or location choices. If households are well sorted spatially or across stores by income or education, we are less likely to see within-location or within-store disparities in purchases. The first reason for this is mechanical. It is possible that the spatial sorting by income leaves little variation in income within the households in the same county or census tract. Sampling error in household purchases, which results in noisy measures of the nutritional content of these purchases, could potentially outweigh the residual variation in income after controlling for residential or purchase location, resulting in attenuation bias.³⁴ The second reason is that the retail environment itself is determined by household tastes. If neighborhoods are segregated by income and stores sort spatially to cater to local tastes, then we might not expect to see much variation in the choice sets of households living in the same location. To the extent that this is the case, we expect there to be less scope for differences in the healthfulness of households’ purchases within locations (or stores) than across them. These possibilities suggest that, if anything, the results above overstate the role of access in generating disparities across income and education groups and contribute to our belief that we have identified an upper bound on the role of access in explaining nutritional disparities in consumption.

6 Response of Household Purchases to a Changing Retail Environment

Before concluding, we take an alternative, and more direct, approach to thinking about the potential impact that improved access would have on household consumption. Specifically, we look at the responsiveness of household purchases to changes in the availability of healthful foods in their area.

Over the six years in our sample, we observe changes in the retail environments of households. The retail environment of a household can change for three reasons: 1) the household moves to a different census tract with different access, 2) the stores in a household’s neighborhood change the products they offer, and 3) stores enter and exit a household’s neighborhood. To capture changes in the retail environment, we use time-varying kernel densities of store concentration and store nutritional quality. The concentration indexes are as before, where we use a kernel density of store indicators to account for differences in the distance-weighted number of stores. Similarly, we construct kernel densities of the store quality measures, both for the expenditure and the nutrient score, to measure differences in the distance-weighted availability of recommended products.³⁵

³⁴One might also be concerned that the disparities that we estimate controlling for household location and store choice are identified from only a small subset of the sample that lives in the same areas and shops in the same stores. We investigate this possibility. The distributions of income and education residualized from other demographics and month and year effects are extremely similar to the distributions of income and education residualized from other demographics, month and year effects, and location or store effects. Therefore, we are identifying the “within-location” and “within-store” effects over a similar support of income and education as used in the regressions without location or store controls.

³⁵As before, we use a Gaussian kernel with a bandwidth of 20km. Letting S_t denote the universe of stores in time t , H_{slt} the expenditure score of store s in census tract l in time t , and d_{sl} the distance between store s and the centroid of census tract l , the expenditure score kernel density for census tract l in time t is given by $\sum_{s=1}^{S_t} \frac{H_{slt}}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{d_{sl}}{20}\right)^2}$. Similarly, letting N_{slt} denote the nutrient score of store s in census

In Table 6, we examine how household purchases in our sample respond to changes in these measures of access. In columns (1) and (5) the analysis is analogous to what was presented in Table 4, where we explore how the quality of monthly household purchases varies with income and education when controlling for access with continuous measures of the density and healthfulness of the local retail environment rather than with household location fixed effects. We start by looking in the cross section. As before, both measures of household purchase quality are significantly related to income and education. The expenditure score is positively related to store concentration and the distance-weighted store-level expenditure scores, though the magnitudes of these coefficients are small, especially once one takes into account how little variation there is in the expenditure scores across stores in high- and low-socioeconomic status neighborhoods. The nutrient score is significantly related to store concentration but not the distance-weighted store-level nutrient scores. Households in areas with more stores come closer to meeting the FDA's nutrient recommendations.

Table 6: Response of Nutritional Quality of Purchases to Changes in Retail Access

	Dep. Var: Ln(Expenditure Score)				Dep. Var: Ln(Nutrient Score)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ln(Income)	0.0235*** (0.0015)				0.0717*** (0.0032)			
Ln(Education)	0.199*** (0.0068)				0.615*** (0.015)			
Ln(Store Concentration)	0.00140* (0.00071)	-0.000923 (0.0027)	-0.000897 (0.0027)	0.00662 (0.0066)	0.0409*** (0.0016)	0.00565 (0.0063)	0.00507 (0.0063)	-0.0134 (0.017)
Ln(Avg. Store Score)	0.0558* (0.024)	0.0149 (0.025)	0.0178 (0.025)	0.00120 (0.025)	0.0104 (0.017)	0.0633*** (0.013)	0.0691*** (0.013)	0.0636*** (0.014)
Ln(Store Conc.)*Ln(Inc)			-0.00149 (0.0012)	-0.00191 (0.0013)			0.00441*** (0.00092)	0.00447*** (0.00098)
Ln(Store Conc.)*Ln(Edu)			-0.0151 (0.0081)	-0.0196* (0.0088)			0.0216*** (0.0060)	0.0197** (0.0063)
Ln(Avg. Store Score)*Ln(Inc)			0.00966*** (0.0027)	0.00969*** (0.0029)			0.0357*** (0.0068)	0.0369*** (0.0072)
Ln(Avg. Store Score)*Ln(Edu)			0.0241 (0.018)	0.0341 (0.020)			0.149*** (0.033)	0.146*** (0.035)
Observations	3187956	3187956	3187956	2877746	3187956	3187956	3187956	2877746
R ²	0.066	0.435	0.435	0.438	0.032	0.327	0.327	0.329
E w.r.t Store Concen	0.00140	-0.000923	0.00121	0.00933	0.0409	0.00565	0.00104	-0.0172
E w.r.t Corr. Store Score	0.0558	0.0149	0.0112	-0.00638	0.0104	0.0633	0.0390	0.0332
Demographic Controls	Yes	No	No	No	Yes	No	No	No
Household Fixed Effects	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Non-Movers Only	No	No	No	Yes	No	No	No	Yes

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Observations are at the household-month level. Standard errors are clustered by household. All regressions include year-month fixed effects. Log income and education are both demeaned. Demographic controls include household size dummies, average head of household age, a dummy for marital status of household heads, dummies for households with either a female or male household head, a dummy for the presence of children, and dummies for whether the household reports being white, black, Asian, or Hispanic.

We control for household demographics in the cross-sectional analysis, but households may sort spatially by unobservable characteristics that are correlated with tastes for healthy foods. To the extent that stores are sorted according to these unobservable characteristics, the coefficients on the store density and nutritional scores of the neighborhood will be biased upwards. It is also possible that households with a taste for healthful food products tend to sort into residential neighborhoods with limited retail activity and a greater density of convenience stores and gas stations than full-service grocery stores, in which case our coefficients will be biased downwards. To deal

tract l in time t , the nutrient score kernel density for census tract l in time t is given by $\sum_{s=1}^{S_t} \frac{N_{s|t}}{\sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{d_{sl}}{20}\right)^2}$.

with these potential issues, we add household fixed effects in the remaining columns for each dependent variable. By controlling for the household, the coefficients are identified by the time-series variation in purchases and retail environment.³⁶ In columns (2) and (6), we cannot reject that household purchases respond neither to changes in the concentration of retail outlets nor to the average expenditure score of the stores in their vicinity. Household nutrient scores do, however, respond positively to the average nutrient score of the stores in their vicinity.

In columns (3) and (7), we add interactions of these access kernel densities with income and education to explore whether the response of a household to changes in their retail environment varies with these two dimensions of their socioeconomic status. In column (3), we see that the statistically insignificant average response of the expenditure scores of household purchases masks a statistically significant difference in the responses of households at different income levels: higher-income households improve their expenditure scores when offered a more nutritionally-balanced mix of food groups in their neighborhood stores. We see similar socioeconomic disparities in how household nutrient scores respond to changes in both the density and nutritional quality of the products offered in local stores in column (7).

Even controlling for household fixed effects, one might be concerned about households sorting into different locations based on their tastes. In columns (4) and (8) of Table 6 we limit the sample to only those households who report living in the same census tract for all years that they are in the panel. The results are very consistent across specifications. This indicates that the variation in household retail environments driving our results is due to either the entry or exit of stores or changes in the product offerings of incumbent stores. Though this variation is not exogenous to the overall market in which these stores are located, these shifts in aggregate demand are more likely the result of households moving into or out of the neighborhood, rather than shifts in the individual demand of the incumbent households whose responses we are measuring.

To get a better sense of what the magnitudes of these coefficients imply, we consider how low-income and low-education households would respond to a change in their retail environment equivalent to moving from the average low-income, low-education neighborhood to the average high-income, high-education neighborhood. We focus on a household with income and education at the 25th percentile in each dimension - 13 years of education and \$32,500 annual income. The elasticities of household expenditure and nutrient scores implied by the coefficients from each regression specification are presented in the bottom row of Table 6.³⁷

Moving from a low-income and low-education neighborhood to a high-income and high-education neighborhood would result in an increase of 1.96 in the log store concentration index, an increase of 0.005 in the average log store expenditure index, and an increase of 0.053 in the average log store nutrient index. Combined with the estimated elasticities displayed in columns (3) and (7), these improvements in access imply that a low-income, low-education household's expenditure and nutrient scores would improve by 0.002 and 0.004 log units, respectively, if they were to move from a low-socioeconomic to a high-socioeconomic status neighborhood. Comparing these changes to the socioeconomic disparities that we see in household scores in Figure 1, we see that 3% of the gap in expenditure scores and 1% of the gap in the nutrient scores would be removed by closing the gap in access to healthy foods.

Overall, these results indicate that encouraging entry of stores offering healthy foods alone will do little to

³⁶Since demographics are nearly constant across our sample period for a given household, we no longer control for income, education, and other household characteristics.

³⁷Note that log income and education are demeaned in these regressions, so the elasticities are calculated as $\beta_0 + \beta_1 (\ln 13 - \ln \overline{Educ}) + \beta_2 (\ln 32500 - \ln \overline{Inc})$, where β_0 , β_1 , and β_2 are the coefficients on the density, the density interacted with demeaned education, and the density interacted with demeaned income, respectively; \overline{Educ} is the sample mean education level (14.3 years); and \overline{Inc} is the sample mean income (\$50,852).

resolve the socioeconomic nutritional disparities. One possibility for the differential responses across socioeconomic groups is that stores offering more healthful products may also charge higher prices for healthful foods, deterring lower socioeconomic status households from purchasing these items. If this is the case, policies aimed at improving access to healthful foods will only be effective if they pair improved access with subsidies or tax breaks to encourage entry with pricing controls. We plan to explore the role of differential price sensitivities and budget constraints in explaining nutritional disparities in future work.

7 Conclusion

Despite the absence of evidence drawing a causal link between disparities in retail access and disparities in nutritional consumption, much of the literature on food deserts has assumed that equalizing access will decrease nutritional disparities across different demographic groups. Such an assumption underlies policies which aim to improve the quality of food purchases by increasing the availability of healthful products in areas with unhealthful consumption. Contrary to this assumption, our analysis suggests that disparities in nutritional consumption are not driven by differential access to healthy food products. Even when looking at purchases made within the same store, much of the disparities that we observe when looking across stores remain. We also observe a limited response of household purchases to changes in retail access that have occurred in the past. Taken together, our results provide strong evidence that policies which aim to reduce nutritional disparities by improving access to healthful foods will leave much of the disparity unresolved.

References

- Anderson, Michael L and David A Matsa**, “Are restaurants really supersizing America?,” *American Economic Journal: Applied Economics*, 2011, pp. 152–188.
- Aussenberg, Randy Alison**, “SNAP and Related Nutrition Provisions of the 2014 Farm Bill (P.L. 113-79),” *Congressional Research Service Report*, 2014.
- Beaulac, Julie, Elizabeth Kristjansson, and Steven Cummins**, “Peer Reviewed: A Systematic Review of Food Deserts, 1966-2007,” *Preventing chronic disease*, 2009, 6 (3).
- Bitler, Marianne and Steven J. Haider**, “An Economic View of Food Deserts in the United States.,” *Journal of Policy Analysis and Management*, 2011, 30 (1), 153–176.
- Bodor, J Nicholas, Donald Rose, Thomas A Farley, Christopher Swalm, and Susanne K Scott**, “Neighbourhood fruit and vegetable availability and consumption: the role of small food stores in an urban environment,” *Public Health Nutrition*, 2008, 11 (04), 413–420.
- Carlson, Andrea, Mark Lino, and Thomas V. Fungwe**, “The low-cost, moderate-cost, and liberal food plans,” Technical Report, United States Department of Agriculture, Center for Nutrition Policy and Promotion 2007.
- CDC**, “State Initiatives Supporting Healthier Food Retail: An Overview of the National Landscape,” 2011.
- Chen, Susan, Raymond JGM Florax, Samantha Snyder, and Christopher C. Miller**, “Obesity and access to chain grocers,” *Economic geography*, 2010, 86 (4), 431–452.
- Courtemanche, Charles and Art Carden**, “Supersizing supercenters? The impact of Walmart Supercenters on body mass index and obesity,” *Journal of Urban Economics*, 2011, 69 (2), 165–181.
- Cummins, Steven, Anne Findlay, Mark Petticrew, and Leigh Sparks**, “Healthy cities: The impact of food retail-led regeneration on food access, choice and retail structure,” *Built Environment (1978-)*, 2005, pp. 288–301.
- , **Ellen Flint, and Stephen A Matthews**, “New neighborhood grocery store increased awareness of food access but did not alter dietary habits or obesity,” *Health Affairs*, 2014, 33 (2), 283–291.
- Currie, Janet, Stefano Della Vigna, Enrico Moretti, and Vikram Pathania**, “The Effect of Fast Food Restaurants on Obesity and Weight Gain,” *American Economic Journal: Economic Policy*, 2010, pp. 32–63.
- Cutler, David M. and Adriana Lleras-Muney**, “Understanding differences in health behaviors by education,” *Journal of health economics*, 2010, 29 (1), 1–28.
- Harding, Matthew and Michael Lovenheim**, “The effect of prices on nutrition: comparing the impact of product- and nutrient-specific taxes,” Technical Report, National Bureau of Economic Research 2014.
- Jones, Eugene**, “An analysis of consumer food shopping behavior using supermarket scanner data: differences by income and location,” *American Journal of Agricultural Economics*, 1997, pp. 1437–1443.
- Kyureghian, Gayaneh, Rodolfo M Nayga, and Suparna Bhattacharya**, “The Effect of Food Store Access and Income on Household Purchases of Fruits and Vegetables: A Mixed Effects Analysis,” *Applied Economic Perspectives and Policy*, 2013, 35 (1), 69–88.

- Morland, Kimberly, Steve Wing, Ana Diez Roux, and Charles Poole**, “Neighborhood characteristics associated with the location of food stores and food service places,” *American Journal of Preventive Medicine*, 2002, 22 (1), 23–29.
- Pearson, Tim, Jean Russell, Michael J Campbell, and Margo E Barker**, “Do “food deserts” influence fruit and vegetable consumption? A cross-sectional study,” *Appetite*, 2005, 45 (2), 195–197.
- Ploeg, Michele Ver, Vince Breneman, Tracey Farrigan, Karen Hamrick, David Hopkins, Phil Kaufman, Biing-Hwan Lin, Mark Nord, Travis Smith, Ryan Williams, Kelly Kinnison, Carol Olander, Anita Singh, Elizabeth Tuckermanty, Rachel Krantz-Kent, Curtis Polen, Howard McGowan, and Stella Kim**, “Access to Affordable and Nutritious Food: Measuring and Understanding Food Deserts and Their Consequences,” *Report to Congress, Economic Research Service*, June 2009.
- Rose, Donald and Rickelle Richards**, “Food store access and household fruit and vegetable use among participants in the US Food Stamp Program,” *Public Health Nutrition*, 2004, 7 (08), 1081–1088.
- Sharkey, Joseph R, Cassandra M Johnson, and Wesley R Dean**, “Food access and perceptions of the community and household food environment as correlates of fruit and vegetable intake among rural seniors,” *BMC Geriatrics*, 2010, 10 (1), 32.
- Song, Hee-Jung, Joel Gittelsohn, Miyong Kim, Sonali Suratkar, Sangita Sharma, and Jean Anliker**, “A corner store intervention in a low-income urban community is associated with increased availability and sales of some healthy foods,” *Public Health Nutrition*, 2009, 12 (11), 2060–2067.
- Volpe, Richard, Abigail Okrent, and Ephraim Leibtag**, “The Effect of Supercenter-format Stores on the Healthfulness of Consumers’ Grocery Purchases,” *American Journal of Agricultural Economics*, January 2013, 95, 568–589.
- and — , *Assessing the Healthfulness of Consumers’ Grocery Purchases*, BiblioGov, 2013.
- Weatherspoon, Dave, James Oehmke, Assa Dembélé, Marcus Coleman, Thasanee Satimanon, and Lorraine Weatherspoon**, “Price and expenditure elasticities for fresh fruits in an urban food desert,” *Urban Studies*, 2013, 50 (1), 88–106.
- Wrigley, Neil, Daniel Warm, and Barrie Margetts**, “Deprivation, diet, and food-retail access: findings from the Leedsfood deserts’ study,” *Environment and Planning A*, 2003, 35 (1), 151–188.

Appendix

A Supplementary Tables and Figures

Table A.1: Distribution of Household Income by Year

Income category	Year					
	2006	2007	2008	2009	2010	2011
Under 5,000	0.01	0.01	0.01	0.01	0.01	0.01
5,000-7,999	0.01	0.01	0.01	0.01	0.01	0.01
8,000-9,999	0.01	0.01	0.01	0.01	0.01	0.01
10,000-11,999	0.02	0.02	0.01	0.01	0.01	0.01
12,000-14,999	0.03	0.02	0.02	0.02	0.02	0.02
15,000-19,999	0.05	0.04	0.04	0.04	0.04	0.04
20,000-24,999	0.07	0.06	0.06	0.06	0.06	0.06
25,000-29,999	0.07	0.06	0.06	0.06	0.06	0.06
30,000-34,999	0.07	0.07	0.07	0.07	0.07	0.06
35,000-39,999	0.06	0.07	0.06	0.06	0.06	0.06
40,000-44,999	0.07	0.06	0.06	0.06	0.06	0.06
45,000-49,999	0.07	0.06	0.06	0.06	0.06	0.06
50,000-59,999	0.10	0.11	0.11	0.11	0.11	0.10
60,000-69,999	0.09	0.09	0.09	0.09	0.08	0.08
70,000-99,999	0.16	0.18	0.19	0.19	0.20	0.20
100,000 +	0.11	0.12	0.14	0.14	0.14	0.15
Total counts	37786	63350	61440	60506	60658	62092

Table A.2: Distribution of Male Household Head Education by Year

Year	Grade School	Some High School	Graduated High School	Some College	Graduated College	Post College Grad	Total Counts
2006	0.013	0.050	0.253	0.292	0.265	0.127	27439
2007	0.010	0.046	0.255	0.294	0.273	0.121	47786
2008	0.010	0.045	0.254	0.291	0.277	0.123	46199
2009	0.009	0.042	0.256	0.288	0.280	0.124	45280
2010	0.009	0.041	0.253	0.286	0.286	0.126	45465
2011	0.008	0.040	0.245	0.285	0.294	0.128	46565

Table A.3: Distribution of Female Household Head Education distribution by Year

Year	Grade School	Some High School	Graduated High School	Some College	Graduated College	Post College Grad	Total Counts
2006	0.005	0.031	0.277	0.315	0.264	0.108	33963
2007	0.005	0.026	0.268	0.320	0.278	0.103	57317
2008	0.004	0.025	0.264	0.319	0.280	0.107	55634
2009	0.004	0.023	0.263	0.314	0.287	0.109	54699
2010	0.004	0.022	0.256	0.311	0.296	0.111	54747
2011	0.004	0.021	0.247	0.309	0.303	0.116	56135

Table A.4: Consumer Characteristics and Nutritional Quality of Purchases: Full Regression Results

	Dep. Var.: Ln(Expenditure Score)				Dep. Var.: Ln(Nutrient Score)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ln(Income)	0.0424*** (0.0013)		0.0241*** (0.0014)	0.0426*** (0.0024)	0.146*** (0.0028)		0.0893*** (0.0030)	0.0636*** (0.0021)
ln(Avg. HH Head Age)	0.0328*** (0.0036)	0.0392*** (0.0035)	0.0437*** (0.0035)	0.0285*** (0.0023)	0.0440*** (0.0079)	0.0616*** (0.0078)	0.0783*** (0.0078)	0.0206*** (0.0021)
HH Heads Married	0.0436*** (0.0034)	0.0458*** (0.0033)	0.0417*** (0.0033)	0.0557*** (0.0045)	0.103*** (0.0073)	0.113*** (0.0072)	0.0972*** (0.0072)	0.0525*** (0.0039)
Female HH Head Only	-0.0526*** (0.0040)	-0.0690*** (0.0040)	-0.0634*** (0.0040)	-0.0743*** (0.0047)	0.0885*** (0.0086)	0.0337*** (0.0086)	0.0544*** (0.0086)	0.0257*** (0.0040)
Male HH Head Only	0.0340*** (0.0047)	0.0210*** (0.0047)	0.0224*** (0.0047)	0.0178*** (0.0037)	-0.111*** (0.010)	-0.153*** (0.010)	-0.148*** (0.010)	-0.0475*** (0.0032)
Kids Present	0.0258*** (0.0024)	0.0179*** (0.0024)	0.0210*** (0.0024)	0.0250*** (0.0028)	0.0838*** (0.0055)	0.0574*** (0.0054)	0.0686*** (0.0054)	0.0330*** (0.0026)
Race: White	0.00789* (0.0038)	0.0101** (0.0038)	0.00888* (0.0038)	0.00892* (0.0038)	0.0589*** (0.0085)	0.0667*** (0.0084)	0.0620*** (0.0084)	0.0251*** (0.0034)
Race: Black	0.00323 (0.0045)	0.00389 (0.0045)	0.00164 (0.0045)	0.00129 (0.0035)	-0.0863*** (0.0099)	-0.0830*** (0.0098)	-0.0913*** (0.0098)	-0.0289*** (0.0031)
Race: Asian	-0.00872 (0.0061)	-0.0155* (0.0061)	-0.0199** (0.0061)	-0.00855** (0.0026)	0.00766 (0.013)	-0.0113 (0.013)	-0.0272* (0.013)	-0.00473* (0.0022)
Hispanic	0.0130*** (0.0036)	0.0153*** (0.0036)	0.0142*** (0.0036)	0.00867*** (0.0022)	0.0402*** (0.0080)	0.0480*** (0.0080)	0.0439*** (0.0079)	0.0108*** (0.0020)
Ln(Education)		0.247*** (0.0060)	0.203*** (0.0065)	0.0743*** (0.0024)		0.798*** (0.013)	0.635*** (0.014)	0.0939*** (0.0021)
Observations	3440297	3440297	3440297	3440297	3440297	3440297	3440297	3440297
R ²	0.061	0.064	0.066	0.066	0.022	0.026	0.029	0.029
Std Coef	No	No	No	Yes	No	No	No	Yes

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Observations are at the household-month level. Standard errors are clustered by household. All regressions include year-month fixed effects and household size dummies.

Figure A.1: Availability Healthfulness and Nutrient Scores Across Channels

