Applied Economic Perspectives and Policy (2018) volume 00, number 0, pp. 1–41.
doi:10.1093/aepp/ppy018

Submitted Article

Documenting the Link Between Poor Food Access and Less Healthy Product Assortment Across the U.S.

Lauren Chenarides* and Edward C. Jaenicke

Lauren Chenarides is an assistant professor in the Morrison School of Agribusiness, W.P. Carey School of Business, Arizona State University. Edward C. Jaenicke is a professor in the Department of Agricultural Economics, Sociology, and Education, Pennsylvania State University, University Park, Pennsylvania.
*Correspondence may be sent to: Lauren.Chenarides@asu.edu

Submitted 30 May 2018; accepted 31 July 2018.

Abstract *This paper, which investigates food access and underserved areas in the U.S., differs from most existing research by explicitly examining food retailers' in-store product availability, a retailing outcome that might exacerbate or mitigate consumer impacts of residing in areas with poor food access. Using detailed retailer scanner data from 2010–2015 and measures of food access, we investigate the relationship between healthy product assortments and food access. Our results consistently show a negative and significant relationship between census tracts with poor food access and healthy product assortments featuring fruits and vegetables. This finding provides rigorous documentation for what other researchers have claimed, namely that consumers who reside in underserved communities and are already burdened with poor food access endure the hardship of facing less healthy assortments of food items.*

Key words: food access, product assortment, scanner data.

JEL codes: Q18, I18, L11, L81.

The USDA's Economic Research Service (ERS) estimates that, in the United States, 23.5 million people reside in areas with severe food access limitations. Within these neighborhoods, nearly half of the households are low-income (Ver Ploeg et al. 2009). These neighborhoods are extreme examples of poor food access and are commonly referred to as "food deserts."[1]

[1]*A commonly used definition is one that meets both low-income and low-access criteria. For example, Dutko et al. (2012) define a food desert census tract as having both (a) a poverty rate greater than or equal to 20% or a median family income does not exceed 80% statewide (rural/urban) or metro-area (urban) median family income, and (b) at least 500 people or 33% of the population located more than 1 mile (urban) or 10 miles (rural) from the nearest supermarket or large grocery store. For the purposes of this study, the precise definition for food deserts will not be crucial because our empirical investigation does not contain a specific variable for food deserts, but rather considers several census-tract measures related to food access.*

Existing research on areas with food-access challenges focuses mainly on explaining consumer purchasing behaviors, consumer well-being, and public health impacts associated with living in underserved communities. The negative outcomes associated with living in neighborhoods with poor food access have been attributed to the relatively high density of convenience stores and other non-traditional food outlets (i.e., non-supermarkets) when compared to more richly-served markets. Current research claims that consumers who shop at these non-traditional food retailers are faced with higher prices and potentially a lack of nutritious products (Jetter and Cassady 2006; Blanchard and Matthews 2007; Alviola et al. 2013). Moreover, studies examine how food access differences may contribute to health disparities, with some but by no means all studies linking higher BMI and obesity with the prevalence of convenience stores and fast food restaurants (e.g., Larson et al. 2009; Ver Ploeg et al. 2009; Ver Ploeg et al. 2012; Chen et al. 2016).

While consumer impacts from poor food access have been at the forefront of many research agendas, the evolution of the food retailing landscape also plays a major role in the resulting food-access gap across the United States. Faced with an increasing number of non-traditional outlets such as supercenters, dollar stores, and even convenience stores, traditional supermarkets may no longer serve as viable food retail options in many underserved areas where high entry costs can make it undesirable for chain grocery stores to enter these markets. Reflecting entry, exit, and equilibrium concepts, some economic research investigates behavior in the food retailing industry to show that the food-retailing landscape, especially the presence and/or location of large food stores, is the equilibrium outcome of certain demand-and supply-side factors (Ellickson 2007; Bonanno et al. 2012). In these models, extreme food inaccessibility is one such equilibrium outcome, where some of the most significant indicators include a number of demand-side factors such as total area income, poverty levels, and the number of SNAP participants (Bonanno et al. 2012). Despite the commonalities between the concepts of food accessibility and food availability, these two terms describe separate dimensions of "food access," where *accessibility* refers to a more geographically-motivated concept that emphasizes the location of food outlets as well as the ease in which it takes to travel to them (an extensive measure), while *availability* refers to how sufficient the supply of healthy food items is within that location (an intensive measure; Caspi et al. 2012).

Once the retailer's supply-side perspective is included, a number of straightforward but currently unanswered research questions emerge about how the food retailing landscape and the area's underlying demographic attributes influence a whole suite of firm decisions and marketing outcomes. Using detailed micro-level scanner data, this paper examines both the extensive and intensive aspects of food access and food availability across the U.S. food retailing landscape.

Much of the conversation around food access has focused on the extensive aspects of food store location and firm concentration across the United States, rather than an intensive approach, which refers to the availability of products within the set of stores. This paper makes a contribution to the literature by documenting in-store product availability across the U.S. food retailing landscape. More specifically, given the existing food retailing landscape and set of food retailers across the United States, we focus on the question of whether food-retailing firms located in underserved communities offer less-healthy product assortments based on fruit and vegetable offerings.[2] While a wide number of studies claim or assume this relationship exists, the link between poor access (an extensive measure) and less healthy product assortments (an intensive measure) has not been thoroughly documented. If it is indeed true that food retailers in poor food access areas offer less healthy product assortments, then this food-retailing outcome could exacerbate consumer impacts of residing in underserved areas in the United States.

We use a combination of Information Resources, Inc. (IRI)'s InfoScan data and the USDA-ERS's Food Access Research Atlas (FARA) to calculate census tract-level measures of healthy product assortments and food access. Exploiting the panel structure of our dataset, we estimate the relationship between healthy product assortments and food access, controlling for a range of other market-level characteristics and demographics variables. In an equilibrium setting, one might expect that supply and demand factors, including demographic variables, would fully explain geographic variation in product assortment. Instead, we find that poor food access plays a significant role in explaining less healthy product assortments even after accounting for a number of characteristics. To our knowledge, this research is among the first studies (along with Handbury et al. 2015) to explicitly document the relationship between low food access and less healthy product assortments.

Literature Review

Food Access and Policy

Research on food access often supports or investigates policies that address the negative effects on consumers living in underserved communities, including higher associated rates of obesity and other health disparities. This broad research agenda is abundant and arguably fits within three general classifications. One class investigates the relationship between the local food environment and health outcomes. While this research offers mixed results, many studies do find a link between low food access and obesity, diabetes, and other health-related outcomes (Wrigley et al. 2002; Moore and Diez Roux 2006; Chen and Florax 2010; Thomsen et al. 2015; Chen et al. 2016), whereas other studies find no statistical link between food access and obesity, body mass, or weight (Burdette and Whitaker 2004; Sturm and Datar 2005; Jeffery et al. 2006; Ford and Dzewaltowski 2010; Slack et al. 2014). A second class documents significant differences in the food environment, where low food access areas are found to have different retail outlet mixes, including higher densities of convenience stores or fast food restaurants and lower densities of supermarkets, and then identifies how these differences might create higher prices, less product selection, and fewer nutritious products (Alwitt and Donley 1997; Moore and Diez Roux 2006; Blanchard and Matthews 2007; Alviola et al. 2013). These first two streams of research can be related. For example, Neckerman et al. (2009) find that better access to supermarkets is associated with a better quality diet, lower rates of obesity, and lower BMI.

[2] *More precise definitions of healthy product assortments will be introduced and thoroughly discussed in subsequent sections.*

A third class of food access research focuses on the particularly complex issue of policy prescriptions and evaluation. Ver Ploeg et al. (2012) note the persistence of food-access problems and focuses on supermarket availability as a key indicator of food security. Cummins et al. (2014) directly test this assumption via a pilot study through the Pennsylvania Fresh Food Financing Initiative that evaluates the impacts of opening a new supermarket in Philadelphia. The study finds that, although there is increased access, shoppers do not markedly change the amount of fruit and vegetable consumption. This finding supports an earlier national-level study that indicates supermarket densities in urban areas "do not have significant effects on fruit and vegetable purchases" (Kyureghian et al. 2013, 86).

These studies suggest that, although supermarket access is an important force, the larger question is how accessibility and marketing outcomes interact to influence consumer purchasing behavior, and ultimately consumer welfare and health outcomes. Some case study-based research examines behavior and recommends tailored policy prescriptions. However, rather than rely on small-scale studies that may only present part of the picture, some research encourages policymakers to look at larger-scale studies, such as the one we present in this paper, as these studies may provide context around targeted policy analysis (Cummins and Macintyre 2002). Finally, some recent economic studies, including Currie et al. (2010), Anderson and Matsa (2011), Courtemanche and Carden (2011), Volpe et al. (2013), Caillavet et al. (2015), Handbury, et al. (2015), Fan et al. (forthcoming), and Allcott et al. (2017), provide good examples of large-scale studies that address food access and policy.

Food Retailing Behavior and Product Assortment

Research that examines the food retailing industry and broadly describes firms' entry decisions highlights the increased differentiation of store formats. This research suggests that, while the food retailing industry is still dominated by a small number of large-scale, high-quality food retail stores, the remaining portion of the market is left to be dispersed among smaller, lower-quality "fringe" grocery stores (Ellickson 2006; Bonanno and Lopez 2009). This stratification between high- and low-quality stores becomes especially relevant when we consider intensive measures associated with poor food access, such as product availability and product assortment within the low-quality stores. In particular, due to space limitations within fringe stores, for example, the portion of selling space that may be dedicated to the assortment of fresh fruits and vegetables can be restricted (Ver Ploeg et al. 2009).

The underlying conceptual framework related to product assortment is rooted in rational economic theory, which suggests that more choice never reduces well-being, especially in non-complex decision making circumstances where consumers can rank their preferences (Botti and Iyengar 2006). Hausman and Leonard (2002), for example, use detailed scanner data to estimate the net benefit to consumers after the introduction of a new line of bath tissue, and find that consumers are significantly better off after the introduction. In part, this positive welfare effect is due to the decrease in price of competing brands, but also the value to consumers is attributed to the increase in variety due to the availability of a new brand. Research in psychology supports this claim qualitatively by arguing that more choice allows consumers to feel in control of their own fate, have a clearer sense of self-determination, and enhance the evaluation of their ultimate decision, leading to decisions and attitudes that will be consistent over time (Botti and Iyengar 2006; Chernev 2011).

Within the context of food retailing, utility-maximizing consumers are expected to create a bundle of differentiated products that matches their pre-defined preferences (Mussa and Rosen 1978; Botti and Iyengar 2006). Unfortunately, results are mixed in studies examining consumer perceptions and preferences for variety. On one hand, some research finds that consumers generally prefer stores with higher assortment levels (Fox et al. 2004; Borle et al. 2005). Yet, other recent research concludes that more is not always better.[3] Too many alternatives can lead to a consumer feeling overwhelmed and unconfident, even to the point of opting out (Broniarczyk and Hoyer 2006; Iyengar and Lepper 2000).[4] More recent empirical studies have tested these experimental results using a variety of product assortment measures (e.g., number of brands versus number of stock keeping units, which further divide brands by packaging), and the results vary (Briesch et al. 2009; Chernev 2011). Table 1 presents an overview of variations of product assortment measures used in the marketing literature. While there is room for debate, the rest of this paper assumes that consumers who face lower product assortments, all else being equal, are, in fact, worse off, and in the long-run low product assortment would negatively affect their well-being.

From the perspective of the retailer, the significance of product assortment is closely linked with the line of literature on store competition, particularly as it relates to product variety and services offered. For example, Richards and Hamilton (2006) find that product variety, within a given product line, has a positively significant effect on sales. Research by Feenstra (1994) and Handbury and Weinstein (2014), who account for product variety in price indices, also supports the argument that increased product assortment is welfare improving. Retailers understand the significance of proper assortment planning as it is one of the key drivers affecting customer loyalty, but it is especially challenging for retailers due to the combinations of possible assortments, retailer constraints, and changes in consumer preferences and market factors (Bauer et al. 2012). Despite the types of constraints affecting product assortment planning, research recognizes that non-price competition is a major driving force in how firms attract customers.[5]

Our research draws from these research fields to investigate food access and product assortment, which represent extensive and intensive food-retailing measures. Drawing from the marketing literature, we specify a primary definition of product assortment aimed at fruits and vegetables. Our

[3]*For retailers, some studies show that a reduction in product assortment (i.e., a decrease in the number of stock keeping units offered) results in an increase in sales as well as an increase in cost savings (Boatwright and Nunes 2001; Hoch et al. 1994). In addition, the results from these studies suggest that the number of items on the shelves may not outweigh unobservable consumer perceptions, in addition to cost-savings for the retailer.*

[4]*These findings depend on how assortment is defined; the opposite is true for small assortments.*

[5]*Marketing research states that product assortment relies on three major inputs: retailer constraints, environmental factors, and consumer preferences (Mantrala et al. 2009). Retailer constraints may be structurally related, such as the limitations of physical space or distribution requirements, or they may depend on the retailer's branding agenda, such as the retailer's image, market position, or market scope. Environmental constraints are exogenous factors that may be related to changing consumer trends or changes to the economic conditions, which can also influence the presence and variety of products.*

Table 1 Select Literature Review of Product Assortment Measures

Study	Assortment Measure	Product Category	Store Format Considered	Data Description	Study Area
Bauer et al. (2012)	Measure perceived assortment which reflects: (1) the variety of different brands; (2) the variety of different flavors; (3) the variety of different package sizes; and (4) the variety of different quality ranges	Yogurt, fruit juices, salty snacks	Grocery	Survey and focus group of grocery patrons	Switzerland
Boatwright and Nunes (2001)	Number of stock-keeping units (SKU)	42 of the 47 top-selling food and non-food product categories	Grocery	Online grocer experimental data	Northeast United States
Briesch et al. (2009)	Calculate a composite measure of assortment that accounts for the number of SKUs, brands, and sizes that the retailer offers; availability of the household's preferred brands; and the proportion of items that are unique to the retailer (a proxy for private-label items)	10 food and non-food categories: chocolate candy, carbonated beverages, coffee, diapers, dog food, household cleaners, laundry detergent, salty snacks, sanitary napkins, and shampoo	Grocery and non-grocery (drug, club, mass merchandiser)	Multichannel panel collected over a 104-week period between October 1995 and October 1997	Chicago, IL, United States
Broniarczyk et al. (1998)	Compare number of unique SKUs to measures of perceived assortment including (1) the total space devoted to the category, and (2) the availability of a consumer's favorite SKU	6 food and non-food categories: popcorn, canned pasta, yogurt, toilet tissue, detergent, peanut butter	N/A	Two components: (1) experimental lab study, and (2) a simulated shopping trip	United States

Chong et al. (2001)	Number of SKUs	Yogurt	Grocery	IRI household scanner data	Single, undisclosed metropolitan area in the United States
Dhar et al. (2001)	(1) Number of brands offered in each category (breadth); (2) Number of SKUs (depth)	19 food categories sold in 106 major supermarket chains	Grocery	Nielsen retail scanner data	106 major grocery retail chains, United States
Jetter and Cassady (2006)	The presence of a particular item within a given product category	19 food items, considering two alternatives: Thrifty Food Plan-compliant and a healthier option	Grocery	Retailer inventory survey collected over a 12-month period between June 2003 and April 2004	Los Angeles and Sacramento, CA, United States
Kahn and Lehmann (1991)	(1) Number of distinct options or distinct categories; (2) The availability of a more highly valued option; (3) Number of unique alternatives	36 snack foods (6 product, 6 varieties)	N/A	Experimental lab study	United States
Kwak et al. (2015)	Number of SKUs, all of which are members of the same product category	Yogurt	Grocery	ERIM household scanner panel over a 2.5-year period	Sioux Falls, SD, United States

store-level scanner data and store characteristics data allow us to consider this fruit and vegetable-based product-assortment measure, as well as some related measures for robustness checks.

Modeling the Product Assortment-Food Access Relationship

Extrapolating from the product assortment planning literature and also the retailer equilibrium research of Ellickson (2007) and others, one might expect equilibrium outcomes, such as product assortment, to be fully explained by consumer demographics, retailer constraints, and market-level characteristics. However, our central question involves whether food access across the United States has been overlooked in this relationship. We hypothesize that food access plays a significant role in explaining product assortment outcomes even after accounting for consumer demographics and some market characteristics.

We begin our examination with a standard definition of food access and a measure of product assortment that is closely linked to health. That is, we measure product assortment as the number of unique products in the fruit and vegetable categories within each census tract. If, for example, two stores are located within one census tract and both stores sell the same product, the item would be counted once, even though it appears in both stores. This measurement reflects the overall depth of available products generally deemed as healthy. For the food-access variable, we use a standard measure of food access: the proportion of the households who live farther than a half mile from the nearest supermarket and report not having access to a vehicle.

To investigate the statistical relationship between product assortment and food access, we rely on an econometric model to estimate the link between poor food access and our health-related product assortment variable. We construct a linear model of the following form,

$$y_{it} = \beta_0 + \beta_X \boldsymbol{X}_{it} + \alpha_Z \boldsymbol{Z}_i + \upsilon_{it} \tag{1}$$

where y_{it} is the measure of product assortment, i = census tract and t = {2010, 2011, 2012, 2013, 2014, 2015}, $\boldsymbol{X}_{it}$ is a vector of time-varying variables, and $\boldsymbol{Z}_i$ is a vector of time-invariant variables.

Due to the infrequent collection of certain archived public data, including our measure of food access, not all of our data vary across the study period. In addition, some of our data are not available at the census-tract level. Given this lack of temporal and spatial variation in certain variables of our model, coupled with our desire to estimate the relationship between the time-invariant measure of food access with the time-varying product assortment outcomes, we decompose our unobserved effect, υ_{it}, such that $\upsilon_{it} = \mu_j + \varepsilon_{it}$ where μ_j is the unobserved time-invariant county-specific effect that captures county-specific time-invariant unobserved supply and demand factors, and ε_{it} is an idiosyncratic disturbance term.

Because of the nature of food retailing and the underlying strategic competitive forces at play, the relationship between food access and product assortment might be poorly identified because of a number of confounding factors, even after accounting for fixed effects. We thus consider the possibility that our estimated model linking product assortment and food access might need to account for endogeneity in the observed relationships.

Therefore, one estimator that was considered at great length was the Hausman and Taylor (1981) estimator, which combines elements of both random effects and fixed effects estimation for panel data. Because this estimator uses the exogenous variables from the model itself to derive instruments, it obviates the need to find external instruments that are exogenous and correlated with the food access measure, which can be especially challenging to find (Abbott and Klaiber 2011). However, due to lack of variation across time by the exogenous time-variant regressors, the Hausman and Taylor instruments do not pass our instrument test, and therefore we had to discard this approach. Regardless, the results we present are intended to describe the statistical relationship between product assortment and food access rather than explain a causality between the two.

Data

Our intensive analysis of in-store product availability relies on data compiled from four main sources: the IRI InfoScan panel; Nielsen's TDLinx store characteristics database; the USDA-ERS Food Access Research Atlas (FARA); and the U.S. Census American Community Survey (ACS).

Dependent Variable

To construct our dependent variable, product assortment, at the census-tract level, we use six years (2010–2015) of weekly store-level scanner data ("InfoScan") from IRI. The InfoScan dataset reports average weekly prices, total weekly quantity sold, and product descriptions for food items sold at participating food retailers. IRI monitors food transactions according to individual universal product codes (UPCs). In this way, our measure of product assortment is analogous to the definition used by Dhar et al. (2001) and represents product mix *depth*, which is equivalent to the total number of unique products, including variations by size or flavor, within a product category.[6] In addition, we make one important distinction in our measurement of product assortment. That is, rather than using the UPC associated with each product, we use the manufacturer-provided UPC, referred to as the European Article Number (EAN). Unlike UPCs, which can be recycled, EANs are truly unique to the product and cannot be reused after a product is discontinued. As such, we collect nationally-branded EANs sold within each census tract as a basis of constructing our product assortment variable and its variations.

Because of prior research that links poor food access to health, we focus primarily on a specification for a product assortment variable that is strongly associated with the availability of healthy food options. For a particular set of store types, we construct *AssortFV*, which we define as the total number of nationally-branded EANs for each census tract during the years 2010 to 2015 that are in the fruit and vegetable category.[7] Broadly, we define

[6] *Alternative measures of product assortment are calculated in our robustness checks to represent* breadth *(also referred to as* width *in the marketing literature) of a retailer's product mix, which represents the number of brands within a product category the retailer sells.*

[7] *A previous version of estimation results considered the ratio between the total number of nationally-branded UPCs in the fruit and vegetable category divided by the total number of UPCs outside the fruit and vegetable category. This measure was meant to describe how prevalent fruit and vegetable products*

the fruit and vegetable category to include bagged produce, frozen fruits and vegetables, shelf-stable or canned fruit, and shelf-stable or canned vegetables. The measure reflects only nationally-branded products and not store brand products because of better data reliability.[8] *AssortFV* captures the ability to find a specific fruit or vegetable item, considered to be healthy options, in a given census tract.

Calculations of *AssortFV* are determined from products that are sold in the following store formats: grocery, convenience stores, dollar stores, mass merchandisers (excluding club stores), and drug stores, the breakdown of which is presented in table 2.[9] Later, we compare results stemming from the *AssortFV* definition to results using several other definitions of product assortment.

Food Access and Socio-demographic Covariates

To construct our food-access variable, we use publicly-available data from the 2015 USDA-ERS Food Access Research Atlas (FARA), a census tract-level dataset that identifies areas with poor food access. Measures of food access in the FARA are derived by calculating distances between geographic centroids of Census Block Groups and the nearest "supermarket." The set of supermarkets, supercenters, and large grocery stores within the United States is compiled after merging the 2010 STARS directory of stores and the 2010 TDLinx directory of stores (Ver Ploeg and Dutko 2013). The choice of constructing census-tract, rather than ZIP-Code level measures of product assortment is driven primarily by the use of the FARA. Another publicly-available dataset that provides information on food access (e.g., the USDA-ERS Food Environment Atlas) is recorded at the county level, which we consider to be too large of a region to address our research question.[10]

The FARA defines low food access as "being far from a supermarket, supercenter, or large grocery store (supermarket for short)." If a significant portion of the population in each tract is far from a supermarket, then that tract is considered to have low access (Ver Ploeg and Dutko 2013). These data include population estimates of spatial access to affordable and nutritious food from the 2010 Census, income and vehicle availability data from the 2006–2010 American Community Survey (ACS), and a 2010 directory of supermarkets. We define *LANVhalf* as the percentage of housing units living at least a half mile from the nearest supermarket and report having no access to a vehicle. Later, we explore alternative measures of food access and the results are presented in our sensitivity analysis.

are relative to all other products. After receiving feedback from an anonymous reviewer, we decided to forgo this measure for a more straightforward measurement as presented in this manuscript. The main results of the two specifications are consistent.

[8] *Estimation results that include private-label products are available upon request.*

[9] *Only one major club store appears in the InfoScan panel. Therefore, our measures of product assortment do not reflect products that may be available at club stores. We have estimation results that account for product availability at the one club store, and the core results presented below remain unchanged. Therefore, due to the differences in available information provided across channel types, we do not present estimation results that reflect products at club stores in this manuscript.*

[10] *Alternative sources on business listings are available from other public agencies, such as the County Business Patterns (NAICS), or proprietary sources, such as Dun & Bradstreet, which also account for the location of local restaurants; however, due to limitations with these data (e.g., CBP) or access issues (e.g., D&B), we have chosen to use FARA over the other two listed.*

Table 2 Retailer Format Types—Considerations for Constructing Dependent Variable (*AssortFV*)

IRI Channel	TDLinx Retailer Format Type	Included in Assortment Variations
Grocery	Conventional Supermarket	Yes
	Limited Assortment Supermarket	Yes
	Warehouse Grocery	Yes
	Superette	Yes
	Natural/Gourmet Supermarket	Yes
Convenience	Conventional Convenience	Yes
	Military Convenience	No
Mass Merchandiser	Supercenter	Yes
	Conventional Mass Merchandiser	Yes
	General Merchandiser	No
	Military Exchange	No
Dollar	Dollar Store	Yes
Drug	Conventional Drug	Yes
	Small and Rx Only	Yes
Wholesale Club	Conventional Club	No

Note: *EANs are treated identically in the construction of each assortment specification regardless of what retailer format the item is sold, as long as the retailer format indicated above indicates "Yes."*

Because we hypothesize that food access plays a significant role in explaining product assortment, an equilibrium outcome, we also include a set of covariates that capture consumer demographics and market-level characteristics. To motivate our model specification, we turn to the literature on food retailer assortment planning and store choice. Assortment planning makes two important considerations: store profitability and retailer patronage. Drivers of store profitability are dependent on consumers choosing stores based on a given store's prices, location, and assortment. For example, Cleary et al. (2018) use a set of demand- and supply-side factors, including education, race, ethnicity, age, adjacency to a metropolitan area, and the size of a county (in square miles), as factors that impact the profitability (variable profits and fixed costs) of food retailers. Other literature that examines store choice (e.g., Hillier et al. 2015) acknowledges that shopping behavior is affected by race, ethnicity, age, education, socio-economic status, and vehicle ownership. Accordingly, these factors are included as covariates in our model.

We augment the FARA dataset with supplemental data from several other public sources. These data are collected either at the census-tract or county levels. Tables 3 and 4 provide a complete list of variables and summary statistics. We use the ACS five-year estimates over the period 2007–2015 to define the census-tract level measures of factors that draw from the aforementioned literature. These include population density (*PopDen*), per capita income (*PCInc*), share of female population (*ShFemale*), population share with a high school degree (*ShHiSch*), population share with some college (*ShSomeCol*), population share with a bachelor's degree or higher (*ShBach*), population share between the ages of 18 and 24 (*Sh1824*), population share between the ages of 25 and 64 (*Sh2564*), population share over age 65 (*SH65plus*), population share that owns a car (*ShOwnCar*), population

Table 3 Model Variable List

Variable	Description	Measurement	Time-Varying	Data Source
AssortFV	Number of unique nationally branded EANs sold in the fruit and vegetable category.	Level	Yes	InfoScan
AssortFVBrand	Number of unique national brand product lines sold in the fruit and vegetable category.	Level	Yes	InfoScan
AssortFV_SQFT	Number of unique nationally branded EANs sold in the fruit and vegetable category divided by total square footage of food retail selling space.	Ratio	Yes	InfoScan & TDLinx
LALI	Proportion of population who are further than one (1) mile from the nearest supermarket and live in a low-income census tract.	Share (00's)	No	FARA
LANVhalf	Proportion of households who are further than a half (0.5) mile from the nearest supermarket and report not having access to a vehicle	Share (00's)	No	FARA
LANVmile	Proportion of households who are further than one (1) mile from the nearest supermarket and report not having access to a vehicle.	Share (00's)	No	FARA
LAPOP	Proportion of population who are further than one (1) mile from the nearest supermarket.	Share (00's)	No	FARA
PCInc	Per capita income, adjusted for inflation.	Level (000's)	Yes	ACS
PopDen	Population per square mile.	Level (000's per sq. mile)	Yes	ACS
Sh1824	Proportion of population between the ages of 18 and 24.	Share (00's)	Yes	ACS
Sh2564	Proportion of population between the ages of 25 and 64.	Share (00's)	Yes	ACS
Sh65Plus	Proportion of population age 65 and older.	Share (00's)	Yes	ACS
ShBach	Proportion of population with a bachelor's degree or higher.	Share (00's)	Yes	ACS
ShBlack	Proportion of population identifying as black.	Share (00's)	Yes	ACS
ShFemale	Proportion of female population.	Share (00's)	Yes	ACS
ShForgn	Proportion of population who are foreign-born.	Share (00's)	Yes	ACS
ShHiSch	Proportion of population with a high school degree.	Share (00's)	Yes	ACS
ShHisp	Proportion of population identifying as Hispanic.	Share (00's)	Yes	ACS
ShSomeCol	Proportion of population with some college education.	Share (00's)	Yes	ACS
Urban	Binary indicator for urban census tract; A census tract is urban if the geographic centroid of the tract has more than 2,500 people; all other tracts are rural.	Binary	No	FARA

Note: *All variables are measured at the census-tract level.*

Table 4 Summary Statistics for 2010–2015

	2010[a]					2011[b]					2012[c]				
	N	mean	sd	min	max	N	mean	sd	min	max	N	mean	sd	min	max
Assortment Measures															
AssortFV	30,846	646.02	719.99	1.00	4,666.00	31,079	648.89	734.93	1.00	4,853.00	31,225	685.90	784.33	1.00	5,013.00
AssortFVBrand	30,846	129.95	125.59	1.00	816.00	31,079	128.77	125.67	1.00	815.00	31,225	130.62	127.92	1.00	839.00
AssortFV_SQFT	30,140	21.82	84.07	0.00	3,260.00	30,410	21.44	83.44	0.00	3,266.00	30,589	21.94	84.34	0.00	3,203.00
AssortFV_TZ	31,045	641.88	719.53	0.00	4,666.00	31,272	644.89	734.41	0.00	4,853.00	31,412	681.82	783.77	0.00	5,013.00
Food Access Measures															
LALI	72,043	39.20	39.60	0.00	100.00	72,031	39.19	39.60	0.00	100.00	72,023	39.19	39.59	0.00	100.00
LANVhalf	71,938	4.56	5.94	0.00	100.00	71,926	4.56	5.94	0.00	100.00	71,918	4.56	5.94	0.00	100.00
LANVmile	71,938	1.92	3.51	0.00	100.00	71,926	1.92	3.51	0.00	100.00	71,918	1.92	3.51	0.00	100.00
LAPOP	72,043	12.46	15.77	0.00	100.00	72,031	12.46	15.78	0.00	100.00	72,023	12.46	15.78	0.00	100.00
Socio-Demographic Measures															
PCInc	71,856	27.20	14.17	0.15	293.61	71,880	27.74	14.46	0.13	225.52	71,882	27.85	14.50	0.02	252.91
PopDen	72,246	5.16	11.63	0.00	520.87	72,234	5.19	11.62	0.00	452.83	72,226	5.22	11.70	0.00	484.98
Sh1824	71,896	9.86	8.33	0.00	100.00	71,925	9.95	8.44	0.00	100.00	71,923	9.97	8.42	0.00	100.00
Sh2564	71,896	53.05	7.72	0.00	100.00	71,925	53.04	7.66	0.00	100.00	71,923	53.00	7.60	0.00	100.00
Sh65Plus	71,896	13.35	7.66	0.00	100.00	71,925	13.53	7.57	0.00	100.00	71,923	13.79	7.61	0.00	100.00
ShBach	71,896	18.25	13.22	0.00	100.00	71,925	18.44	13.27	0.00	100.00	71,923	18.66	13.38	0.00	100.00
ShBlack	71,896	13.69	22.62	0.00	100.00	71,925	13.77	22.49	0.00	100.00	71,923	13.82	22.44	0.00	100.00
ShFemale	71,896	50.88	4.98	0.00	100.00	71,925	50.85	4.97	0.00	100.00	71,923	50.83	4.91	0.00	100.00
ShForgn	71,896	12.07	13.78	0.00	100.00	71,925	12.12	13.72	0.00	100.00	71,923	12.15	13.66	0.00	100.00
ShHiSch	71,896	19.51	7.94	0.00	100.00	71,925	19.37	7.95	0.00	100.00	71,923	19.19	7.93	0.00	100.00
ShHisp	71,896	14.72	20.91	0.00	100.00	71,925	15.02	21.01	0.00	100.00	71,923	15.28	21.10	0.00	100.00

Continued

Table 4 *Continued*

	2010[a]					2011[b]					2012[c]				
	N	mean	sd	min	max	N	mean	sd	min	max	N	mean	sd	min	max
ShSomeCol	71,896	18.53	5.83	0.00	100.00	71,925	18.85	5.89	0.00	100.00	71,923	19.21	5.90	0.00	100.00
Urban	72,246	0.76	0.43	0.00	1.00	72,234	0.76	0.43	0.00	1.00	72,226	0.76	0.43	0.00	1.00
	2013[d]					**2014[e]**					**2015[f]**				
	N	**mean**	**sd**	**min**	**max**	**N**	**mean**	**sd**	**min**	**max**	**N**	**mean**	**sd**	**min**	**max**
Assortment Measures															
AssortFV	31,651	699.92	825.79	1.00	4,645.00	32,186	746.60	901.27	1.00	5,159.00	32,061	690.75	815.68	1.00	4,976.00
AssortFVBrand	31,651	130.70	131.38	1.00	722.00	32,186	139.98	144.84	1.00	815.00	32,061	135.00	137.94	1.00	768.00
AssortFV_SQFT	30,988	24.84	104.11	0.00	3,413.00	31,537	26.04	110.71	0.00	3,104.00	31,469	21.61	79.37	0.01	2,469.00
AssortFV_TZ	31,852	695.51	825.04	0.00	4,645.00	32,393	741.83	900.36	0.00	5,159.00	32,269	686.30	814.92	0.00	4,976.00
Food Access Measures															
LALI	72,023	39.19	39.59	0.00	100.00	72,022	39.19	39.60	0.00	100.00	72,019	39.19	39.59	0.00	100.00
LANVhalf	71,918	4.56	5.94	0.00	100.00	71,917	4.56	5.94	0.00	100.00	71,914	4.56	5.94	0.00	100.00
LANVmile	71,918	1.92	3.51	0.00	100.00	71,917	1.92	3.51	0.00	100.00	71,914	1.92	3.51	0.00	100.00
LAPOP	72,023	12.46	15.78	0.00	100.00	72,022	12.46	15.78	0.00	100.00	72,019	12.46	15.77	0.00	100.00
Socio-Demographic Measures															
PCInc	71,875	27.94	14.52	0.02	253.44	71,880	28.32	14.76	0.08	247.85	71,871	28.68	15.02	0.13	254.20
PopDen	72,226	5.26	11.78	0.00	479.99	72,225	5.31	11.90	0.00	455.64	72,222	5.36	12.01	0.00	454.71
Sh1824	71,928	9.98	8.40	0.00	100.00	71,933	9.95	8.41	0.00	100.00	71,917	9.88	8.35	0.00	100.00
Sh2564	71,928	52.93	7.60	0.00	100.00	71,933	52.86	7.61	0.00	100.00	71,917	52.75	7.58	0.00	100.00

Sh65Plus	71,928	14.08	7.67	0.00	100.00	71,933	14.40	7.73	0.00	100.00	71,917	14.77	7.80	0.00	100.00
ShBach	71,928	18.92	13.49	0.00	100.00	71,933	19.27	13.64	0.00	100.00	71,917	19.63	13.77	0.00	100.00
ShBlack	71,928	13.83	22.34	0.00	100.00	71,933	13.85	22.25	0.00	100.00	71,917	13.86	22.14	0.00	100.00
ShFemale	71,928	50.81	4.86	0.00	100.00	71,933	50.80	4.84	0.00	100.00	71,917	50.80	4.74	0.00	100.00
ShForgn	71,928	12.20	13.62	0.00	100.00	71,933	12.28	13.61	0.00	100.00	71,917	12.36	13.58	0.00	100.00
ShHiSch	71,928	19.19	7.90	0.00	100.00	71,933	19.19	7.88	0.00	100.00	71,917	19.16	7.87	0.00	100.00
ShHisp	71,928	15.50	21.17	0.00	100.00	71,933	15.71	21.22	0.00	100.00	71,917	15.91	21.25	0.00	100.00
ShSomeCol	71,928	19.34	5.91	0.00	100.00	71,933	19.44	5.90	0.00	100.00	71,917	19.54	5.87	0.00	100.00
Urban	72,226	0.76	0.43	0.00	1.00	72,225	0.76	0.43	0.00	1.00	72,222	0.76	0.43	0.00	1.00

Note: *Results of two-sample t-tests indicate that the mean values for all pairwise combinations across years are statistically different from each other. Different letters denote significant differences at the 0.05 level. For example, the mean value for AssortFV in 2010 marked with a superscript [a] is statistically different from years marked with [b,c,d,e, and f]. In the table above, each year is denoted with a different letter superscript, which means that the mean values for all pairwise combinations across years are statistically different from each other.*

Table 5 Market Coverage Overview

Year	Gazateer	ACS	FARA	TDLinx[a]		IRI[a]	
				Matched	Missing	Matched	Missing
2010	72,246	72,246	72,246	62,666	9,580	30,846	41,400
2011	72,246	72,234	72,234	62,849	9,385	31,079	41,167
2012	72,246	72,226	72,226	63,076	9,150	31,225	41,021
2013	72,246	72,226	72,226	63,215	9,011	31,651	40,595
2014	72,246	72,225	72,225	63,414	8,811	32,186	40,060
2015	72,246	72,222	72,222	63,525	8,697	32,061	40,185
Total Obs.	433,476	433,379	433,379	378,745	54,634	189,048	244,428
% of Total				87%	13%	44%	56%

Note: *Superscript [a] indicates that the table above reflects all channel types except military stores, which include grocery, mass merchandisers, convenience, dollar, drug, and club stores.*

share with Hispanic ethnicity (*ShHisp*), population share identified as black (*ShBlack*), and population share who are foreign-born (*ShForgn*).

Using each of these sources, we construct a six-year panel dataset composed of product assortment, food access, and demographic variables at the census-tract level. After removing census-tract observations with missing information on product assortment measures, we are left with an unbalanced panel of 189,048 observations spanning six years. Table 5 presents the number of census tract-level assortment measures, by year: 30,846 (2010), 31,079 (2011), 31,225 (2012), 31,651 (2013), 32,186 (2014), 32,061 (2015), which accounts for approximately 44% of all census tracts at the national level for each year in our panel. In other words, our analysis captures the intensity of in-store product availability within the set of stores that service 44% of census tracts.

Limitations

According to the Economic Research Report published by the ERS, there are three important considerations researchers should make when using the IRI InfoScan panel (Muth et al. 2016). First, data restrictions between ERS and IRI may affect market coverage. Second, some food retailers choose to geographically aggregate their store sales information. Finally, some, but not necessarily the same, food retailers choose not to release sales information for private-label products. Because our research objective is to measure the statistical relationship between food access and product assortment using the IRI InfoScan panel available through ERS, it is important to acknowledge that these three factors have the potential to influence our analysis and the estimation results. In this section, we discuss these three limitations in detail and describe our approach at addressing them in our estimation.

Market Coverage

Although the use of the InfoScan data gives us the ability to measure the assortment of bar-coded products at the census-tract level, it is also true that the InfoScan panel used in this analysis is a subset of the entire data available through IRI (Muth et al. 2016). As such, the data does not capture all

Table 6 Cases for Describing Possible Selection Bias

	TDLinx			
IRI	**At least one store reported**	**Zero stores reported[a]**	**Missing census tract**	**Total Obs.**
Positive value for assortment	Case I 185,133 (42.7%)	Case II 66 (0.02%)	Case III 3,849 (0.9%)	189,048(43.6%)
Missing value for assortment	Case IV 192,351 (44.4%)	Case V 1,195 (0.3%)	Case VI 50,785 (11.7%)	244,331(56.4%)
Total Obs.	377,484 (87.1%)	1,261 (0.3%)	54,634 (12.6%)	433,379

Note*: Product assortment and store counts are measured at the census-tract level. Superscript [a] indicates that "Zero stores" means a census tract contains zero stores in the set of channel types from the TDLinx data. Channel types include grocery, mass merchandisers, convenience, dollar, and drug stores, excluding club stores.*

available store information across the United States. Therefore, we suspect that some census tracts have missing values of assortment when, in actuality, they should be positive. In other words, the assortment measures for the census tracts that are captured by our data are a subset of all the census tracts in the United States. If the subset is truly random, then we would not expect there to be any selection bias, and OLS would yield unbiased parameter estimates. However, if there is selection bias, and if we do not correct for this, then OLS would yield biased estimates. Table 6 presents six possible cases that will help us frame our discussion of possible selection bias. We discuss each of these in turn below.

For each census tract, we are able to observe when that census tract has a positive value for product assortment (Cases I, II, and III). However, there are many census tracts where IRI InfoScan does not record transaction information. For these tracts, we do not have positive values for assortment and therefore we report these as missing (Cases IV, V, and VI). A zero value of product assortment would not appear in our data because InfoScan records sales transactions rather than store inventory. To some degree, we can rely on the TDLinx dataset to identify the extent to which selection may be a cause for concern in our analysis, as TDLinx presents an extensive cross-section of food retailers, including food retail channels (e.g., independent stores) that may not appear in the IRI data. As it turns out, 44.4% of census tracts fall into Case IV (missing values in IRI but positive values in TDLinx), 0.3% fall into Case V (missing values in IRI but zero values in TDLinx), and 11.7% fall into Case VI (missing values in both IRI and TDLinx). With this information, we first address whether there is any evidence that IRI systematically, i.e., non-randomly, selects the census tracts that appear in the subset of data used in our analysis and discuss our approach below.

The food retailers in the InfoScan panel obtained by ERS include only the subset of retailers that have agreed to provide transactions data for all of their stores, either at an individual store level (Store) or at a geographic-region level defined separately by each retailer. IRI refers to this subset of retailers as the "census" retailers. IRI also collects an additional sample, referred to as a "statistically representative sample," of randomly drawn stores across the United States. This sample is proprietary and not available to researchers (Muth et al. 2016). If census retailers are selected in a specific way, then we might conclude that the corresponding census tracts where participating retailers report transactions are non-randomly drawn. Accordingly, the most pertinent information for the specification of our selection equation includes factors that could explain why some retailers chose to participate in the subset of stores that IRI considers as its census retailers.

Little information is provided regarding those factors that directly influence why some retailers agree to participate as census retailers and provide sales data while others do not. We do know, however, that there are commonalities among stores within each channel type. The 2016 ERS report outlines data coverage by channel type and we can use this information to specify our selection equation. From the report, other than a minimum store volume of \$2 million required for grocery stores, all other stores are accepted as census retailers for channel types outside of grocery regardless of store volume (Muth et al. 2016). Drug and convenience stores appear to have the highest frequency of complete sales information, whereas only one major club store chain agrees to provide transactions data. The four major U.S. mass merchandisers appear in the data, although sales information may differ according to the geographic aggregation decided by each retailer.

A simple test for selection bias is conducted by regressing an indicator variable, z_{it}, on possible census-tract factors related to retailers, T_{it}, and demographics, M_{it}, that might explain the census tract-based selection of IRI stores that appear in the InfoScan panel. The selection equation takes the following form

$$z_{it} = \gamma_0 + \gamma_T \boldsymbol{T}_{it} + \gamma_M \boldsymbol{M}_{it} + \nu_{it} \tag{2}$$

where z_{it}, is binary and receives a value of one if the census tract is represented in the InfoScan panel and a zero value otherwise. Taking into consideration the volume requirements for grocery stores to appear in InfoScan, the density of convenience and drug stores represented in the data, and the participation of major food retail chains, we use a set of food environment and socio-demographic variables to specify the selection equation (2). If we find evidence that there is systematic bias, then we can correct for this econometrically by estimating a first-stage probit equation with these covariates, collecting the inverse mills ratio (IMR), and using the IMR as a coefficient in the second-stage fixed-effects model.

To investigate the degree of incompatibility between IRI and TDLinx, we also consider the cases when we observe positive values for assortment (Cases I, II, and III). Due to differences between these two datasets, stores may appear in one dataset and not the other, or vice versa. Even where we observe positive values of assortment and the presence of stores in TDLinx (Case I), the degree of coverage in IRI data may be sparser relative to the density of stores reported by TDLinx. In general, however, the TDLinx data show a more complete picture of the food retailing landscape. We therefore identify certain areas of the United States as high coverage tracts (where the match percentage between the two datasets is high) and low coverage tracts (where the match percentage between the two datasets is low). As part of our sensitivity analysis, we compare the estimation results using "High" and "Low" coverage with results using the entire national panel.[11] In addition, there are census tracts where we observe a positive value for product assortment and TDLinx either reports zero stores (Case II) or does not report any information for that census tract (Case III). Although these instances are rare, we do not explicitly correct for these two cases and acknowledge that this remains a limitation in our analysis.

Finally, a small portion of census tracts show that the value of product assortment from IRI InfoScan is missing yet TDLinx reports zero stores are present (Case V). As another robustness check, we replace our product assortment measures with zero for these instances.

Geographic Aggregation

Some food retail chains may choose to participate as a census retailer on the condition that their sales information is released at a geographically aggregated level. This set of retailers decides on the level of aggregation, or retailer marketing area (RMA). As noted in Muth et al. (2016), a total of eight supermarket and club stores appear as RMA stores. While the sales information for retailers participating in the Store-level data is unique to each store, the associated sales information for RMA retailers is aggregated across the RMA and cannot be attributed to any specific store. In addition, some RMA retailers may combine a larger geographic area compared to other RMA retailers. In this way, one retailer's RMA may include an entire state, while another retailer's RMA may include only a few counties within one state. While the count and location of individual stores within the RMA set of retailers are included in InfoScan, we account for this discrepancy in sales reporting as best we can.

The reason to combine both store and RMA retailers, rather than to exclude all RMA retailers, is because several major chains appear as RMA retailers. Therefore, we make an important assumption in our construction of product assortment. That is, all stores for any retailer reporting sales at the RMA-level carry the same assortment of products. In other words, we make the assumption that any transaction recorded by an RMA retailer would have appeared in all the stores within that RMA. While this may be a strong assumption, it is also true that food retailers manage inventory at a regional level, so it is not unlikely that stores within a region carry the same products and/or coordinate with regional vendors (Dhar et al. 2001; Hesse and Rodrigue 2004). To our knowledge, this paper is among the first to combine both datasets, data discrepancies and all, to conduct a broad national-level analysis.[12]

Private-label Products

One of the limitations inherent in the InfoScan panel is the underrepresentation of private-label (PL) products. For all non-PL (e.g., national brands or generic) transactions, each product sold and recorded in InfoScan is

[11] *There is an important distinction to make between the ERS report and our analysis. Muth et al. (2016) compare the coverage, as measured by store counts, against the U.S. Census (NAICS) and find that the number of stores in the InfoScan panel accounts for approximately 44% of stores in NAICS. In our analysis, we compare census-tract coverage with TDLinx and we find a similar coverage rate of 42.9% of census tracts (Case I and II).*

[12] *A detailed diagram illustrating the construction of the dataset is presented in Supplementary Appendix I.*

associated with a unique UPC (or EAN). Only some retailers, however, have agreed to provide PL sales information to IRI. For these retailers, it is up to the discretion of the retailer whether they provide UPC-level PL information. While some retail chains provide point-of-sale transactions for PL UPCs, certain retail chains aggregate these transactions to the brand or category level and only a generic "placeholder" UPC is provided (Muth et al. 2016). Because of the non-uniformity among retailers regarding PL information, we do not consider PL products in our measure of product assortment.

Estimation

The variable of interest in our model is product assortment, measured as the number of unique manufacturer-generated EANs in the fruit and vegetable categories for each census tract-year combination. Product assortment is observed only when at least one transaction from a participating census retailer in a given census tract i during year t is recorded, otherwise it is counted as missing. This relationship can be formalized as follows, with the help of two latent variables, one that reflects the underlying truth for missing assortment values and another that reflects the underlying truth for the inclusion of a store in the InfoScan panel:[13]

$$AssortFV_{it} = \begin{cases} AssortFV^*_{it} & if \;\; z_{it} = 1 \\ missing & otherwise \end{cases} \tag{3}$$

where the model for the latent variable $AssortFV^*_{it}$ for census tract i in year t is a slight modification of equation (2):

$$AssortFV^*_{it} = \beta_0 + \beta_X \boldsymbol{X}_{it} + \delta FoodAccess_i + \mu_j + \varepsilon_{it}. \tag{4}$$

The indicator variable, z_{it}, is binary and receives a value of one if the census tract is represented in the InfoScan panel, and a zero value otherwise:

$$z_{it} = \begin{cases} 1 & if \;\; z^*_{it} = 1 \\ 0 & otherwise \end{cases} \tag{5}$$

where the latent variable, z^*_{it}, can be expressed as a linear model of the form:

$$z^*_{it} = \gamma_0 + \gamma_T \boldsymbol{T}_{it} + \gamma_M \boldsymbol{M}_{it} + \nu_{it}. \tag{6}$$

We rely on TDLinx as the source for many of the census-tract level variables used in the selection equation (6).[14] Elements of $\boldsymbol{T}_{it}$ include the number

[13] *When the indicator variable,* z_{it}*, does not receive a value of one, we cannot make the assumption that assortment is zero, because the value may in fact be positive. Therefore, we note the value as missing.*
[14] *It should be noted that TDLinx covers only 87% of the census tracts in the United States. Additional census-tract level predictors were considered, yet were discarded after considering the selection equation. The selection equation represents the selection of stores into the InfoScan panel, rather than store location in a given census tract. Therefore, demand- and supply-side instruments traditionally employed in store location and entry models (e.g., energy prices, miles to distribution centers, miles of public roadways) do not apply in our application.*

Table 7 Regression Results for Selection Equation

Variable	(1) OLS	(2) FE	(3) Probit
Grocery2M	0.114***	0.119***	0.428***
	(0.001)	(0.004)	(0.004)
Conv	0.013***	0.013***	0.060***
	(0.000)	(0.001)	(0.001)
Drug	0.108***	0.124***	0.417***
	(0.001)	(0.004)	(0.003)
MassMerch	0.153***	0.151***	0.863***
	(0.002)	(0.005)	(0.008)
Dol	0.141***	0.129***	0.569***
	(0.001)	(0.004)	(0.005)
Club	0.082***	0.059***	0.430***
	(0.005)	(0.011)	(0.021)
PCInc	0.002***	0.002***	0.007***
	(0.000)	(0.000)	(0.000)
Urban	0.201***	0.146***	0.675***
	(0.002)	(0.005)	(0.006)
Year=2011	−0.002	−0.002***	0.052***
	(0.002)	(0.001)	(0.008)
Year=2012	−0.008***	−0.008***	0.043***
	(0.002)	(0.001)	(0.008)
Year=2013	−0.010***	−0.010***	0.023***
	(0.002)	(0.001)	(0.008)
Year=2014	−0.008***	−0.008***	0.013
	(0.002)	(0.002)	(0.008)
Year=2015	−0.013***	−0.013***	0.020**
	(0.002)	(0.002)	(0.008)
Constant	0.033***	0.086***	−1.652***
	(0.003)	(0.008)	(0.010)
Observations	378,179	378,179	378,179
R-squared	0.282	0.284	
Number of Counties		3,102	3,102
County FE		YES	

Note*: Standard errors appear in parentheses. Asterisks indicate the following: *** $p < 0.01$. ** $p < 0.05$. and * $p < 0.1$.*

of independent and chain grocery stores with average commodity volume (ACV) greater than $2 million (*Grocery2M*), the number of independent and chain convenience stores (*Conv*), the number of independent and chain mass merchandisers (*MassMerch*), the number of chain dollar stores (*Dol*), the number of chain drug stores (*Drug*), and the number of club stores (*Club*). Elements of $\boldsymbol{M}_{it}$ include average per capital income (*PCInc*) from the ACS and a binary variable indicating that a census tract is located in an area with a population of more than 2,500 people (*Urban*) from FARA.[15]

We estimate equation (3) using ordinary least squares (OLS) and county-level fixed effects (FE) and present our results in table 7. The estimated

[15]*The idea to account for sample selection in the IRI data came from an anonymous reviewer. To our knowledge, our paper is the first to address selection.*

coefficients for all covariates in columns (1) and (2) of table 7 are statistically significant, results that suggest that census tracts are not randomly selected by IRI. Therefore, we proceed by estimating a first-stage probit with these covariates, collecting the IMR, and using the IMR as a coefficient in the second-stage fixed-effects model.

The results of the first-stage probit are presented in table 7, column (3). The discussion of these coefficients is brief, as they are not directly used for inference. A closer look at the signs of the coefficients indicates that census tracts are more likely to appear in the InfoScan panel when the census tract has a higher number of stores across all channel types. Census tracts are also more likely to appear in the InfoScan panel when the census tract has higher per capita income and is located in an urban county.

Baseline Model

After collecting the IMR, the second-stage estimation equation becomes

$$AssortFV_{it} = \beta_0 + \beta_X \boldsymbol{X}_{it} + \delta LANVhalf_i + \sigma\lambda_{it} + \mu_j + \varepsilon_{it} \tag{7}$$

where λ_{it} is the IMR collected from the first-stage probit regression, and the set of demographic shifters in $\boldsymbol{X}_{it}$ includes one variable from our first-stage regression (*PCInc*), as well as variables for educational attainment (*ShHiSch*, *ShSomeCol*, *ShBach*), gender (*ShFemale*), race (*ShBlack*), ethnicity (*ShHisp*, *ShForgn*), and age (*Sh1824*, *Sh2564*, *Sh65Plus*). We also interact *PCInc* with variables of educational attainment to reflect non-linearities between the effects of education on income.

We estimate equation (7) using OLS and FE, and our results are presented in table 8.[16] For comparison, we present both sets of results in columns (1) and (2), respectively. Although there is minimal difference, OLS may be underestimating the association between poor food access and product assortment.[17] Therefore, we focus the discussion of the baseline results on the estimates presented in column (2) of table 8.

Our variable of interest, *LANVhalf*, shows a negative and highly significant relationship with product assortment, indicating that, on average, levels of healthy product assortment decrease as the share of households without a vehicle who live farther than a half a mile from a supermarket increases. In other words, a one percentage point increase in the share of poor access households is associated with approximately ten fewer products ($\delta = -10.566$). Because we have conditioned our product assortment measure to look at nationally-branded fruit and vegetable EANs, this result suggests that consumers living in already underserved neighborhoods face an additional burden, namely the lack of availability of universally-accepted healthy food items.

While the magnitude of the coefficient may appear relatively small as a percentage of the average number of bar-coded fruit and vegetable items, this negative result has a direct and non-trivial economic consequence. The

[16] *Although not presented in table 8, including the IMR as a covariate increases the model fit substantially, from roughly 0.041 to 0.146.*

[17] *We estimate the model using fixed and random effects, and test the consistency of our parameter estimates using a Hausman test. In every case, we fail to reject the null hypothesis that random effects provide consistent estimates.*

Table 8 Second Stage Regression Results for Baseline Model (*AssortFV*)

Variable	**(1) OLS**	**(2) FE**	**(3) FE w/Temporal Interactions**
LANVhalf	−10.577***	−10.566***	−8.812***
	(0.954)	(0.870)	(0.840)
LANVhalf x Year=2011			0.075
			(0.116)
LANVhalf x Year=2012			−0.387*
			(0.202)
LANVhalf x Year=2013			−1.479***
			(0.440)
LANVhalf x Year=2014			−3.671***
			(0.548)
LANVhalf x Year=2015			−4.801***
			(0.538)
PopDen	−9.200***	−7.334***	−7.345***
	(2.203)	(2.705)	(2.707)
ShHiSch	2.725	−0.408	−0.368
	(2.873)	(2.210)	(2.207)
ShSomeCol	9.353***	9.566***	9.702***
	(1.989)	(1.789)	(1.786)
ShBach	13.083***	12.367***	12.413***
	(1.726)	(1.286)	(1.284)
PCInc	4.700**	3.522*	3.512*
	(2.180)	(1.923)	(1.922)
PCInc x ShHiSch	0.318***	0.315***	0.316***
	(0.080)	(0.067)	(0.067)
PCInc x ShSomeCol	−0.042	−0.004	−0.006
	(0.059)	(0.049)	(0.049)
PCInc x ShBach	−0.143***	−0.140***	−0.140***
	(0.029)	(0.026)	(0.026)
ShFemale	0.944	1.176	1.166
	(1.002)	(1.007)	(1.007)
ShBlack	−1.024***	−1.218***	−1.221***
	(0.301)	(0.275)	(0.275)
ShHisp	−1.390**	−1.038**	−1.031**
	(0.597)	(0.439)	(0.439)
ShForgn	4.360***	2.254***	2.256***
	(1.104)	(0.618)	(0.618)
Sh1824	−3.164***	−5.062***	−5.061***
	(0.877)	(0.816)	(0.816)
Sh2564	−13.425***	−14.307***	−14.358***
	(1.535)	(1.327)	(1.327)
Sh65plus	−9.982***	−13.434***	−13.503***
	(1.479)	(1.233)	(1.233)
Year=2011	−3.483**	−2.470**	−2.795**
	(1.371)	(1.203)	(1.323)
Year=2012	32.380***	33.591***	35.235***
	(2.329)	(2.085)	(2.398)
Year=2013	43.412***	47.500***	53.929***
	(3.397)	(3.003)	(3.982)
Year=2014	83.258***	89.385***	105.455***

Continued

Table 8 *Continued*

Variable	(1) OLS	(2) FE	(3) FE w/Temporal Interactions
	(4.357)	(4.075)	(5.553)
Year=2015	25.795***	33.160***	54.148***
	(5.752)	(5.032)	(5.857)
IMR (λ)	−615.191***	−617.147***	−617.107***
	(11.326)	(11.849)	(11.851)
Constant	1,277.061***	1,472.251***	1,465.125***
	(106.808)	(90.557)	(90.552)
Observations	185,077	185,077	185,077
R-squared	0.158	0.145	0.146
Number of Counties		2,738	2,738
County FE		YES	YES

Notes: *Robust standard errors in parentheses. Standard errors are clustered at the county level. Asterisks indicate the following: ***p < 0.01. **p < 0.05. and *p < 0.1.*

parameter estimate on *LANVhalf* relates levels of poor food access with product assortment at the mean. In places with higher than average densities of low access communities, such as parts of Washington, D.C., Philadelphia, and Louisiana, where the share of households living in poor access census tracts may be greater than 20%, the negative relationship between product assortment and poor food access could translate into an average of 200 fewer nationally-branded fruit and vegetable items. We explore this result in detail by performing myriad sensitivity and robustness checks after investigating the other results from our estimation of equation (7).

According to the results from our baseline model, where the product assortment measure is the total number of nationally-branded fruit and vegetable EANs for each census tract (*AssortFV*), the parameter estimates on the demographic variables in $\boldsymbol{X}_{it}$ have mixed signs, but are, for the most part, highly significant. Census tracts where the proportion of households with higher levels of educational attainment (*ShSomeCol, ShBach*) are associated with higher levels of healthful product assortment. A one percentage point increase in the share of population with some college education is associated with nine additional products (β=9.566), while those with a bachelor's degree or higher is associated with approximately twelve extra items (β=12.367). Similarly, we see a positive relationship between per capita income (*PCInc*) and product assortment, at a value of over three additional items (β=3.522). The interaction of per capita income with the same levels of educational attainment that yield positive and significant estimates (*PCInc x ShSomeCol*, *PCInc x ShBach*) are negative, yet only the interaction term between per capita income and the share of population with a bachelor's degree or more (*PCInc x ShBach*) is significant, implying that the highest level of educational attainment dampens the positive relationship that per capita income has with product assortment.

Higher proportions of foreign-born populations (*ShForgn*) are associated with increased levels of product assortment, with roughly two additional products (β=2.254) per one percentage point increase in foreign-born population. Increased proportions of black (*ShBlack*) and Hispanic (*ShHisp*) populations show a negative and significant relationship with product assortment, where a one percentage point increase in the respective populations is associated with a decrease of one item, on average (β=−1.218, β=−1.038). The coefficients on various age groups (*Sh1824, Sh2564, Sh65Plus*) also show a negative and significant relationship with product assortment, where the largest is attributed to the share of those between twenty-five and sixty-four years contributing to approximately fourteen fewer products (β=−14.307) compared to the population share between eighteen and twenty four (β=−5.062) and the population share aged sixty five and older (β=−13.434). The outside age group, which was not included in our model due to collinearity, is the share of the population under eighteen years of age.

We do not explicitly control for urbanization in our baseline model specification, though we do include population density (*PopDen*) as a covariate. The parameter estimate for population density is negative and highly significant (β=−7.334). Under the assumption that product assortment is an equilibrium outcome, this result implies that IRI food outlets located in more densely-populated census tracts, controlling for other demographics, sell a more limited assortment of fruit and vegetable products, on average. Conversely, the less densely populated an area is, the more expansive the assortment. To the extent that we can generalize this result, stores located in less densely-populated census tracts may be drawing consumers from a wider radius than stores located in more densely populated census tracts. At the time of data collection, the computational intensity to construct a dataset that accounts for products available in contiguous census tracts inhibited our ability to explore this result further. Although it may be a limitation of this paper, future research should consider additional robustness checks that consider varying product assortment to include contiguous geographical regions. Nonetheless, because population density may play a different role depending on levels of urbanization, we present regression results by subsampling the data between urban and non-urban census tracts. These results are presented in a subsequent section.

Temporal Interactions with Food Access

Exploiting the panel structure of the data, we interact our measure of food access (*LANVhalf*) with each year in our study period. In this way, we are interested in the marginal yearly statistical associations between food access and product assortment (*AssortFV*). Results of two-sample t-tests indicate that the mean values for all pairwise combinations across years are statistically different from each other. Our results for the temporal interactions in table 8 column (3) show not only that the relationship between *LANVhalf* and healthy product assortment remains negative each year, but also that it increases over time.

For 2010, a one percentage point increase in *LANVhalf*, or the share of households living farther than a half mile from the nearest supermarket and report not having access to a vehicle, corresponds to a decrease of roughly nine nationally-branded fruit and vegetable products (δ=−8.812). This relationship increases over time, and by 2015, increasing proportions of low access households are associated with a decrease of approximately thirteen fruit and vegetable items (δ=−13.613). The parameter estimate for the temporal interaction in year 2011 is not statistically significant, yet the estimates are highly significant and increasingly negative in years 2012, 2013, 2014, and 2015.

Two possible trends could explain these time-related results. First, an increasing negative coefficient could mean that overall food availability for bar-coded fruit and vegetable items is decreasing. Because our dependent variable is the number of unique nationally-branded fruit and vegetable EANs, food retailers within a given census tract may collectively be tightening their offerings of healthy bar-coded options, in particular frozen and packaged nationally-branded fruit and vegetable items. In other words, the increasingly negative relationship may be driven by the fact that stores are responding to changes in consumer preferences, which may suggest a distaste for these particular products, which we observe as lower levels of healthy product assortment. In this way, the negative relationship between poor food access and healthy product assortment grows over time because food retailers are merely meeting the demands of the consumers in the neighborhood. Another look at table 4 reveals that the number of unique fruit and vegetable products available appears to oscillate over time, peaking in 2014, and then dropping in 2015. This trend is consistent within the low access and non-low access subsampled tracts, yet average values of product assortment appear to increase more consistently over time in non-low access tracts, compared with low-access tracts, with the exception of 2015. In this year, the drop in mean product assortment values is steeper in low-access tracts (−8%) versus non-low access tracts (−7%).[18]

Another interpretation of these results focuses more specifically on food access, where access implies the presence of a store. Estimates that indicate a diminishing number of fruit and vegetable items from 2010 to 2015 within low-access census tracts may make sense if we see that the number of stores entering a census tract is less than those exiting. To the extent that we can rely on TDLinx to track the number of store entries and exits year-over-year, we present these statistics in table 9. The number of stores entering and exiting low-access tracts is presented in columns (1) and (2), respectively, and non-low access tracts in columns (4) and (5), respectively. This table reports, however, that the annual total number of stores entering census tracts is *greater* than those exiting for both low-access (column (3)) and non-low access tracts (column (6)).

As noted, our measure of product assortment looks at the number of unique nationally-branded fruit and vegetable items within a given census tract, irrespective of the specific retail chain or channel. Therefore, the number of store entries and exits alone may not necessarily explain the decreasing and significant relationship between poor food access and healthy product assortment over time. To this end, we also consider the composition of stores entering and exiting, in particular if the increase in store entries is being driven by convenience stores and/or drug stores, rather than supermarkets and mass merchandisers, which are traditionally known for offering a wider assortment of fruit and vegetable items. Looking at the number of store entries and exits by channel type according to TDLinx, more grocery stores and mass merchandisers are exiting low-access tracts over time, as shown in column (3) of table 9. Although we observe an increase in the number of total store entries in low-access tracts, a large portion of the entries are convenience stores—see columns (9) and (10). This is also true in

[18]*These summary tables are presented in Supplementary Appendix II.A and Supplementary Appendix II.B.*

Table 9 Number of Store Entry and Exits over 2010–2015 Study Period according to TDLinx

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Store Counts								Percent of Total			
	Low Access			Non-Low Access			Total		Low Access		Non-Low Access	
	Entries	**Exits**	**Difference**	**Entries**	**Exits**	**Difference**	**Entries**	**Exits**	**Entries**	**Exits**	**Entries**	**Exits**
						2010–2011						
Total	2,600	1,668	932	8,862	5,908	2,954	11,462	7,576	23%	22%	77%	78%
Grocery	356	285	71	1,608	1,037	571	1,964	1,322	14%	17%	18%	18%
Mass Merch	258	28	230	726	81	645	984	109	10%	2%	8%	1%
Conv	1,388	835	553	4,355	3,161	1,194	5,743	3,996	53%	50%	49%	54%
Dollar	274	232	42	955	706	249	1,229	938	11%	14%	11%	12%
Drug	324	288	36	1,218	923	295	1,542	1,211	12%	17%	14%	16%
						2011–2012						
Total	2,547	1,278	1,269	8,543	4,262	4,281	11,090	5,540	23%	23%	77%	77%
Grocery	294	354	−60	1,319	1,194	125	1,613	1,548	12%	28%	15%	28%
Mass Merch	73	52	21	234	116	118	307	168	3%	4%	3%	3%
Conv	1,290	721	569	4,017	2,437	1,580	5,307	3,158	51%	56%	47%	57%
Dollar	355	50	305	1,061	138	923	1,416	188	14%	4%	12%	3%
Drug	535	101	434	1,912	377	1,535	2,447	478	21%	8%	22%	9%
						2012–2013						
Total	2,318	1,429	889	7,723	4,717	3,006	10,041	6,146	23%	23%	77%	77%
Grocery	331	336	−5	1,350	1,168	182	1,681	1,504	14%	24%	17%	25%
Mass Merch	47	46	1	193	117	76	240	163	2%	3%	2%	2%
Conv	1,024	793	231	3,275	2,646	629	4,299	3,439	44%	55%	42%	56%
Dollar	401	45	356	1,219	146	1,073	1,620	191	17%	3%	16%	3%
Drug	515	209	306	1,686	640	1,046	2,201	849	22%	15%	22%	14%

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Store Counts								Percent of Total			
	Low Access			Non-Low Access			Total		Low Access		Non-Low Access	
	Entries	Exits	Difference	Entries	Exits	Difference	Entries	Exits	Entries	Exits	Entries	Exits
2013–2014												
Total	2,606	1,606	1,000	8,654	5,740	2,914	11,260	7,346	23%	22%	77%	78%
Grocery	349	367	−18	1,451	1,295	156	1,800	1,662	13%	23%	17%	23%
Mass Merch	110	73	37	319	219	100	429	292	4%	5%	4%	4%
Conv	1,290	673	617	3,893	2,691	1,202	5,183	3,364	50%	42%	45%	47%
Dollar	402	73	329	1,315	221	1,094	1,717	294	15%	5%	15%	4%
Drug	455	420	35	1,676	1,314	362	2,131	1,734	17%	26%	19%	23%
2014–2015												
Total	2,245	1,809	436	8,056	6,157	1,899	10,301	7,966	22%	23%	78%	77%
Grocery	321	444	−123	1,452	1,413	39	1,773	1,857	14%	25%	18%	23%
Mass Merch	55	101	−46	235	420	−185	290	521	2%	6%	3%	7%
Conv	1,104	794	310	3,647	2,719	928	4,751	3,513	49%	44%	45%	44%
Dollar	370	155	215	1,369	528	841	1,739	683	16%	9%	17%	9%
Drug	395	315	80	1,353	1,077	276	1,748	1,392	18%	17%	17%	17%

Note: *Low access tracts are measured as low access at a half mile and at least 33% of households report not having access to a vehicle (n = 13,204); Non-low access tracts otherwise (n = 59,840). Channel types include grocery, mass merchandisers, convenience, dollar, and drug stores, excluding club stores.*

non-low access tracts, yet when we compare the proportion of grocery and mass merchandiser exits, these formats are exiting low-access tracts at a slightly higher rate than non-low access tracts over time, columns (11) and (12). Therefore, the composition of new entrant store formats along with store formats exiting could be driving the increasingly negative sign that is observed on *LANVhalf*. Finally, new entrants may be selling the same types of products (i.e., an undifferentiated set of national-brand items), compared with non-low access tracts, where extant stores could be selling a more differentiated set of national-brand items. Although our analysis does take a granular look at product assortment and store formats, our data are limited in regards to the depth of product availability within specific retailer channel types.

In the section that follows we explore variations to our preferred model – the baseline with temporal interactions presented in this section.

Sensitivity Analyses and Robustness Checks

To test the robustness of the results, we re-estimate our model after altering various elements in several ways: (i) we vary the specification of product assortment; (ii) we vary the way we define poor food access; (iii) we split the data into two subsamples based on the degree of match between the IRI and TDLinx data sets; (iv) we split the data into urban and non-urban area subsamples. Except for (i) we re-estimate the baseline results with *AssortFV* as the definition of product assortment. In addition, regression results tables for each robustness check are abridged, yet a full set of covariates are included as well as the IMR, unless otherwise indicated.

Alternative Measures of Assortment

We test three alternative specifications of product assortment: (*a*) we define *AssortFVBrand* as the total number of unique fruit and vegetable brands sold within a census tract during the year. This measure most closely represents the product category *breadth* as described in Dhar et al. (2001). (*b*) We define *AssortFV_SQFT* as the total number of unique fruit and vegetable EANs divided by the total area of selling space within each census tract. For this measure, we use TDLinx to calculate the total square footage of food retail chains at the census-tract level. This variation of assortment is meant to reflect constraints faced by retailers located within a given census tract.[19] (*c*) We replace *AssortFV* with zero if TDLinx does not report a store for the respective census tract-year combination and IRI reports that tract as missing (table 6, Case V). Because we replace *AssortFV* with zero to consider census tracts that truly have no stores, we estimate this model as a censored Tobit.[20]

Our results, presented in table 10, columns (1)–(3), yield similar findings, implying that poor food access has a consistently negative and significant

[19] *TDLinx store database has information on the total square footage of each food retail outlet at the census tract level. We use this information to aggregate up to the census tract level by summing the total square footage of retail space within each census tract. The number we get from this calculation is what we use for the denominator of* AssortFV_SQFT.
[20] *Table 10, column (3) is estimated as a Tobit.*

Table 10 Regression Results for Sensitivity Analysis: Changes to Assortment

Variable	(1) AssortFVBrand	(2) AssortFV_SQFT	(3) AssortFV_TZ
LANVhalf	−1.535***	−0.537***	−8.384***
	(0.143)	(0.119)	(0.814)
LANVhalf x Year=2011	0.041*	−0.039	0.137
	(0.022)	(0.035)	(1.104)
LANVhalf x Year=2012	−0.066	0.002	−0.278
	(0.041)	(0.038)	(1.100)
LANVhalf x Year=2013	−0.205***	−0.098	−1.327
	(0.072)	(0.111)	(1.100)
LANVhalf x Year=2014	−0.691***	−0.183	−3.481***
	(0.092)	(0.115)	(1.094)
LANVhalf x Year=2015	−0.954***	−0.131	−4.584***
	(0.104)	(0.113)	(1.096)
Constant	263.586***	−2.967	1,143.893***
	(15.086)	(10.262)	(48.163)
Observations	185,077	185,019	186,182
R-squared	0.160	0.058	
Number of Counties	2,738	2,738	2,739
County FE	YES	YES	NO

Notes: *Robust standard errors in parentheses. Standard errors are clustered at the county level. Asterisks indicate the following: ***$p < 0.01$. **$p < 0.05$. and *$p < 0.1$.*

relationship with product assortment, even after accounting for alternative definitions of product assortment. The coefficient estimates of *LANVhalf* on *AssortFVBrand* in column (1) shows that increasing the share of poor food access households by one percentage point is associated with a decrease of almost two national brands, on average, for the reference year 2010 ($\delta = -1.535$), and also increases over time, with the most negative association in 2015. When we account for total selling space at the census-tract level (*AssortFV_SQFT*), the coefficient on *LANVhalf* continues to have a statistically significant relationship with product assortment, though it is much smaller ($\delta = -0.537$). Finally, the relationship between poor food access and *AssortFV_TZ* looks similar to that from our preferred model ($\delta = -8.384$) in 2010 and worsens over time.

Taken together with our results from the baseline model, these results suggest that, even after changing our measurement of product assortment, poor food access continues to have a negative association with the availability of food items in a statistically significant way.

Alternative Measures of Food Access

For our sensitivity analysis that looks at three alternative measures of food access, we estimate equation (7), replacing *LANVhalf* with food access measured as: (*a*) the proportion of households who report not having access to a vehicle and who live farther than one mile from the nearest supermarket (*LANVmile*), (*b*) the proportion of population, regardless of income status, who live farther than one mile from the nearest supermarket (*LAPOP*), and (*c*) the proportion of population living in low-income tracts and who live farther than one mile from the nearest supermarket (*LALI*).[21] According to Ver Ploeg et al. (2015), households without a vehicle travel 0.9 miles, on average, to their preferred food store. Therefore, we choose the half-mile limit as our lower bound (*LANVhalf*) and one-mile limit as our upper bound (*LANVmile*) of the Ver Ploeg et al. (2015) measure. The first alternative specification of food access, *LANVmile*, is similar to our baseline measure, while the second and third specifications of food access, *LAPOP* and *LALI*, do not account for vehicle access.

We present the results for the alternative specifications of food access in table 11, and our results show negative and highly significant estimates on product assortment. A one percentage point increase in the share of households who report not having access to a vehicle and who also live farther than one mile from the nearest supermarket corresponds with a decrease in approximately eleven fruit and vegetable items ($\delta = -11.870$) for the reference year. Yearly interactions are only significant in 2014 and 2015, though the relationship worsens over time ($\delta = -15.020$; $\delta = -20.572$).

When we vary the measure of food access to exclude vehicle access (*LAPOP* or *LALI*), the coefficient on poor food access in the reference year remains negative and highly significant. In these two alternative measures, yearly interactions are almost always significant, yet the signs on the coefficients vary. During years 2011–2013, the negative coefficient of *LAPOP* decreases slightly, yet remains negative. The relationship worsens in years 2014 (ns) and 2015. We see similar results with the third variation (*LALI*). Overall, our main results from our preferred model are robust to different specifications of food access.

High Coverage vs. Low Coverage

As part of our sensitivity analysis, we define census tracts according to their coverage rate between IRI and TDLinx. TDLinx indicates the presence of a food store in 87% of U.S. census tracts, while IRI reports the presence of stores in approximately 44% of those tracts (table 4, Cases I and II). Of the percentage of tracts that are matched (i.e., tracts where TDLinx and IRI both report stores), we can observe the number of stores that TDLinx reports as well as the number of stores that IRI reports within each tract. For example, for some tracts, TDLinx and IRI report having the same number of stores or some arbitrarily high percentage, so we refer to these tracts as having "high coverage." Alternatively, TDLinx may report several more stores that IRI reports, which we refer to as having "low coverage." We estimate the same model as in equation (7), but run separate estimations for tracts with high coverage and tracts with low coverage. We test various coverage rate thresholds, and present results with high coverage, where the threshold falls above 75%, and low coverage, where the threshold falls below 50%. For these subsamples based on coverage, we do not include the IMR as a covariate.

Results for these two subsamples are presented in table 12 columns (2) and (3): one with census tracts representing high coverage, and one with

[21] *Low-income tracts are tracts such that the poverty rate of 20% or higher, or the median family income, is less than 80% of the median family income for the state or metro-area. The original food desert measure identifies areas that are low-income and low-access, where low access is measured using a 1 mile demarcation in urban areas and 10 miles in rural areas, according to the proximity to the nearest supermarket. According to this measure, an estimated 18.3 million people live low-income and low-access tracts in 2010 (Ver Ploeg et al. 2009).*

Table 11 Regression Results for Sensitivity Analysis: Changes to Food Access Measure

Variable	**(1)**	**(2)**	**(3)**
LANVmile	−11.870***		
	(1.458)		
LANVmile x Year=2011	0.107		
	(0.223)		
LANVmile x Year=2012	0.372		
	(0.372)		
LANVmile x Year=2013	0.169		
	(0.936)		
LANVmile x Year=2014	−3.150***		
	(1.119)		
LANVmile x Year=2015	−8.702***		
	(1.009)		
LAPOP		−2.125***	
		(0.390)	
LAPOP x Year=2011		0.128**	
		(0.054)	
LAPOP x Year=2012		0.378***	
		(0.093)	
LAPOP x Year=2013		0.368*	
		(0.205)	
LAPOP x Year=2014		−0.256	
		(0.250)	
LAPOP x Year=2015		−2.590***	
		(0.237)	
LALI			−0.812***
			(0.172)
LALI x Year=2011			0.009
			(0.024)
LALI x Year=2012			0.112***
			(0.041)
LALI x Year=2013			0.285***
			(0.088)
LALI x Year=2014			0.121
			(0.106)
LALI x Year=2015			−0.752***
			(0.104)
Constant	1,482.649***	1,499.161***	1,522.485***
	(90.675)	(90.621)	(90.860)
Observations	185,077	185,090	185,090
R-squared	0.144	0.143	0.143
Number of Counties	2,738	2,738	2,738
County FE	YES	YES	YES

Note: *Robust standard errors in parentheses. Standard errors are clustered at the county level. Asterisks indicate the following: ***p < 0.01. **p < 0.05. and *p < 0.1.*

census tracts representing low coverage. It is important to note that low-coverage tracts outnumber high coverage tracts. Table 12 shows that our low-coverage and high-coverage results are very similar to our full sample baseline results. In addition, our main result, a statistically negative estimate

Table 12 Regression Results for Sensitivity Analysis: Sub-Samples

Variable	(1) Full	(2) High Coverage	(3) Low Coverage	(4) Urban	(5) Non-Urban
LANVhalf	−8.812***	−18.128***	−8.121***	−7.423***	−2.430
	(0.840)	(1.900)	(0.948)	(0.864)	(4.650)
LANVhalf x Year=2011	0.075	0.143	0.254	0.117	0.005
	(0.116)	(0.516)	(0.257)	(0.124)	(0.651)
LANVhalf x Year=2012	−0.387*	−1.602**	−0.082	−0.426**	−0.474
	(0.202)	(0.748)	(0.401)	(0.215)	(0.898)
LANVhalf x Year=2013	−1.479***	−0.715	−1.197**	−1.627***	0.101
	(0.440)	(1.548)	(0.569)	(0.449)	(2.395)
LANVhalf x Year=2014	−3.671***	−3.370*	−2.339***	−3.952***	−1.490
	(0.548)	(1.723)	(0.675)	(0.565)	(2.565)
LANVhalf x Year=2015	−4.801***	−7.494***	−3.159***	−4.778***	−5.655**
	(0.538)	(1.487)	(0.681)	(0.556)	(2.467)
PopDen	−7.345***	−12.539**	−4.912***	−6.433***	29.998
	(2.707)	(5.743)	(1.674)	(2.234)	(57.394)
ShHiSch	−0.368	3.645	0.136	1.489	−6.114
	(2.207)	(3.938)	(2.581)	(2.325)	(6.782)
ShSomeCol	9.702***	9.299***	10.880***	10.437***	13.013*
	(1.786)	(3.347)	(2.197)	(1.911)	(6.887)
ShBach	12.413***	8.813***	11.125***	13.631***	14.004***
	(1.284)	(2.212)	(1.537)	(1.375)	(4.326)
PCInc	3.512*	4.943	1.734	1.234	14.007
	(1.922)	(3.506)	(2.815)	(1.898)	(9.518)
PCInc x ShHiSch	0.316***	0.138	0.223**	0.272***	0.193
	(0.067)	(0.116)	(0.094)	(0.069)	(0.261)

Continued

Table 12 *Continued*

Variable	(1) Full	(2) High Coverage	(3) Low Coverage	(4) Urban	(5) Non-Urban
PCInc x ShSomeCol	−0.006	−0.100	−0.093	0.030	−0.119
	(0.049)	(0.078)	(0.075)	(0.052)	(0.233)
PCInc x ShBach	−0.140***	−0.124***	−0.076**	−0.114***	−0.375***
	(0.026)	(0.047)	(0.038)	(0.025)	(0.132)
ShFemale	1.166	6.308***	3.747***	3.143***	1.755
	(1.007)	(1.865)	(1.198)	(1.025)	(3.157)
ShBlack	−1.221***	−0.716	−0.286	−1.398***	1.723
	(0.275)	(0.635)	(0.370)	(0.281)	(1.474)
ShHisp	−1.031**	−1.222	1.063*	−0.912**	−0.508
	(0.439)	(0.768)	(0.559)	(0.458)	(2.241)
ShForgn	2.256***	3.063**	2.995***	2.207***	10.261***
	(0.618)	(1.243)	(0.885)	(0.613)	(3.964)
Sh1824	−5.061***	−1.422	−0.597	−5.790***	11.503***
	(0.816)	(1.635)	(1.158)	(0.797)	(3.189)
Sh2564	−14.358***	−6.688***	−8.805***	−16.657***	0.288
	(1.327)	(2.455)	(1.745)	(1.386)	(3.992)
Sh65plus	−13.503***	−7.332***	−5.050***	−15.411***	−1.627
	(1.233)	(2.033)	(1.793)	(1.312)	(3.978)
Year=2011	−2.795**	9.202***	0.669	−2.831*	−5.192
	(1.323)	(3.362)	(2.530)	(1.473)	(3.909)
Year=2012	35.235***	61.371***	31.676***	33.004***	41.622***
	(2.398)	(5.112)	(3.645)	(2.682)	(6.046)
Year=2013	53.929***	112.602***	21.857***	49.254***	58.111***
	(3.982)	(9.118)	(5.244)	(4.205)	(14.377)
Year=2014	105.455***	198.653***	52.367***	101.856***	101.801***

	(5.553)	(11.883)	(6.918)	(5.998)	(15.700)
Year=2015	54.148***	133.935***	26.455***	61.670***	4.664
	(5.857)	(10.332)	(6.925)	(6.377)	(15.235)
IMR (λ)	−617.107***			−723.100***	−463.544***
	(11.851)			(16.901)	(23.986)
Constant	1,465.125***	554.751***	395.536***	1,523.002***	232.741
	(90.552)	(173.523)	(142.076)	(93.930)	(344.328)
Observations	185,077	56,596	78,764	157,161	27,916
R-squared	0.146	0.037	0.029	0.159	0.116
Number of Counties	2,738	1,356	2,630	2,083	2,191
County FE	YES	YES	YES	YES	YES

Note: *Robust standard errors in parentheses. Standard errors are clustered at the county level. Asterisks indicate the following: ***$p < 0.01$. **$p < 0.05$. and *$p < 0.1$.*

for *LANVhalf*, holds for both the low-coverage and high-coverage subsamples. The low-coverage estimates appear to be most similar to our preferred model with temporal interactions.

Urban vs. Non-urban

It is very possible that our food-access variable (*LANVhalf*) is sensitive to whether the underlying census tract is in an urban setting. In other words, if food retailers and households behave very differently in urban settings, we might find that our main result, a negative association between poor food access and levels of healthy product assortment, does not hold outside of urban cases. To investigate this issue, we split our data sample into urban and non-urban census tracts and re-estimate our baseline model. After splitting our data, we have 157,142 observations for urban census tracts and 27,916 for non-urban census tracts.

We estimate equation (7) using FE and our results for our urban and non-urban subsamples are presented in table 12 columns (4) and (5). We include the full set of results, rather than just the abridged results for *LANVhalf*, to show that many demographic variables play different roles in urban and non-urban census tracts. More specifically, while table 12 shows that the sign of estimated coefficients are often the same for urban and non-urban subsamples, many of the demographic variables are only statistically significant in one of the two subsamples.

More importantly, table 12 shows that our main result holds in the urban but not the non-urban subsample. For the non-urban subsample, the estimated coefficient on *LANVhalf* is not statistically different from zero, except for year 2015. While the lack of statistical significance may be partly due to fewer observations, it might also suggest that the population share in non-urban settings classified as having low food assess (using a half mile radius) has no statistical relationship with a fruit and vegetable-based product assortment measure. In other words, retailers in these census tracts are generally offering the same relative amount of fruits and vegetables in low food access areas as they are in non-low access areas. Nonetheless, because of the sparsity of data available to describe the relationship between product assortment and food access in non-urban census tracts, our results are most likely explaining a relationship between the two in urban or semi-urban areas.

Conclusion

This research investigates health-related product assortment and poor food access and finds the generally robust result that product assortment decreases as measures of poor food access increase. Our results consistently document a negative and significant relationship between product assortment and food access and support what other researchers have claimed: for households living in underserved communities that already face the burden of shopping at a restricted set of food retail outlets, they also face the additional burden of fewer healthy food options within the stores in their own neighborhood.

Despite the complex nature of retailer and consumer behavior in underserved areas, we are confident in our results for two main reasons: (*a*) the significant impact of a low-access variable is present even after accounting for demographic factors, which also play a significant role, and (*b*) our findings are robust to variations in the product assortment dependent variable, variations in the food access measure, and most constructions of the data sample. The only data subsample where our main result fails to hold with statistical significance is for a non-urban data subset. For this case, the number of usable observations is much less than other subsamples.

To explore the full implications of our observed inverse relationship between product assortment and food access, we must consider our result as an equilibrium outcome of retail supply and consumer demand factors. Our main result suggests that the low-assortment outcome found in areas defined by having poor food access is the natural result of equilibrium forces. With this in mind, our results can provide insight into several important economic implications.

First, when products are available and conveniently located, it is easier for people to have the opportunity to purchase healthier foods. If, in a given neighborhood, healthy food items are not available, then those living in underserved communities are faced with two choices: shop from the available set of products, or travel outside their neighborhood to find better options. In this way, new policies, such as healthy-retailer initiatives, may be a low-cost way to bring new products to low-access communities, as long as stores are present. Some common features of this policy are that they target existing retail establishments, partner with existing income-support programs, and focus on local efforts. Organizations such as the Philadelphia-based Food Trust have already identified several city-based strategies including one such pilot program, the Healthy Corner Store Initiative, that incorporates marketing and educational techniques to expand choice to help improve habits in areas where supermarkets are scarce.

A second economic consequence relates to the importance of fruit and vegetable product assortments, both to retailers and consumers. Fresh produce has been shown to be one of the most significant drivers of store traffic and store sales in grocery retail (e.g., Blattberg and Fox 1995; Dhar et al. 2001). In addition, the presence of food products that appear to be highly valued, such as fruit and vegetable items, has a greater impact on a consumer's perceived variety than a less-preferred item (Chernev 2011). As a result, the more favorable a consumer's perception is on the availability of store offerings, the more loyal the customer will be to the store that offers them. In other words, fruit and vegetable offerings are extremely important for retailers because they attract customers and build loyalty. While we do not explicitly model the availability of fresh fruits and vegetables, some of the products captured in our measure of product assortment do include some bar-coded fresh items (e.g., bagged lettuce and bagged apples).

The extent to which fresh fruits and vegetables, mostly random weight items, are excluded from our analysis should not discount our results. Processed vegetables and fruit, including flash-frozen vegetables or canned items, contain key nutrients required for optimal health. Households living in underserved communities who also report a lack of access to transportation rely on shopping at nearby small grocers, convenience stores, or drug stores, where processed fruits and vegetables may be the only options. While policies that promote the consumption of fresh fruits and vegetables are of high merit, the amount of resources required to implement these programs must not be overlooked. A highly-cited barrier for purchasing healthy foods is a lack of knowledge to prepare them (e.g., Bonanno et al.

2014; Farahmand et al. 2015). The resources that would have to be allocated to educate households who are unfamiliar with the preparation and storage of fresh fruits and vegetables (i.e., food utilization) is costly. Therefore, having better access to processed fruits and vegetables, often with cooking instructions on the packaging, can be highly valuable for these households.

Finally, our results may relate to important conclusions in recent studies about how product offerings and variety affect consumer welfare. For example, Handbury and Weinstein (2014) find that "consumers spend less, on average, to get the same amount of consumption utility in larger cities pg. 259," where the number of available products can be substantially larger. Rather than focus on all products or cities like that important study, our investigation examines product availability in the fruit and vegetable category and uses census tracts as the unit of analysis. Our conclusion is that fruit and vegetable product offerings are linked (inversely) with food access at the census-tract level. Extending our conclusion into the Handbury and Weinstein framework would imply that these fewer offerings may lead to lower purchasing power and, thus, lower welfare for households in areas of low food access.

Acknowledgment

This research is funded, in part, by the USDA-NIFA-AFRI Predoctoral Fellowship Grant Program number 2015- 67011-22790.

Supplementary Material

Supplementary material is available at *Applied Economics Perspectives and Policy* online.

Disclaimer

The analysis, findings, and conclusions expressed in this paper also should not be attributed to either Nielsen or Information Resources, Inc. (IRI).

References

Abbott, J.K., and H.A. Klaiber. 2011. An Embarrassment of Riches: Confronting Omitted Variable Bias and Multi-Scale Capitalization in Hedonic Price Models. *Review of Economics and Statistics* 93 (4): 1331–42.

Allcott, H., R. Diamond, and J.P. Dubé. 2017. The Geography of Poverty and Nutrition: Food Deserts and Food Choices Across the United States. Working Paper No. w24094, National Bureau of Economic Research.

Alviola, P.A., R.M. IV, Nayga, Jr, and M. Thomsen. 2013. Food Deserts and Childhood Obesity. *Applied Economic Perspectives and Policy* 35 (1): 106–24.

Alwitt, L.F., and T.D. Donley. 1997. Retail Stores in Poor Urban Neighborhoods. *Journal of Consumer Affairs* 31 (1): 139–64.

Anderson, M.L., and D.A. Matsa. 2011. Are Restaurants Really Supersizing America? *American Economic Journal: Applied Economics* 3 (1): 152–88.

Bauer, J.C., A.J. Kotouc, and T. Rudolph. 2012. What Constitutes a "Good Assortment"? A Scale for Measuring Consumers' Perceptions of an Assortment

Offered in a Grocery Category. *Journal of Retailing and Consumer Services* 19 (1): 11–26.
Blanchard, T.C., and T.L. Matthews. 2007. Retail Concentration, Food Deserts, and Food-Disadvantaged Communities in Rural America. In *Remaking the North American Food System: Strategies for Sustainability*, ed. C. Clare Hinrichs and Thomas A. Lyson, 201–15. Lincoln: University of Nebraska Press.
Blattberg, R.C., and E.J. Fox. 1995. *Category Management: A Series of Implementation Guides, vols I–IV*. Washington, DC: Food Marketing Institute.
Boatwright, P., and J.C. Nunes. 2001. Reducing Assortment: An Attribute-Based Approach. *Journal of Marketing* 65 (3): 50–63.
Bonanno, A., and R.A. Lopez. 2009. Competition Effects of Supermarket Services. *American Journal of Agricultural Economics* 91 (3): 555–68.
Bonanno, A., L. Chenarides, and S.J. Goetz. 2012. Limited Food Access as an Equilibrium Outcome: An Empirical Analysis. Paper presented at the 2012 AAEA/EAAE Food Environment Symposium, Boston, MA.
Bonanno, A., L. Chenarides, A. Palmer, and K. Clancy. 2014. Perceived Barriers to Purchasing Healthy Foods vs. Access in Underserved Areas across the Northeast. Poster presented at the 2014 Agricultural and Applied Economics Association Annual Meeting, Minneapolis, Minnesota.
Borle, S., P. Boatwright, J.B. Kadane, J.C. Nunes, and S. Galit. 2005. The Effect of Product Assortment Changes on Customer Retention. *Marketing Science* 24 (4): 616–22.
Botti, S., and S.S. Iyengar. 2006. The Dark Side of Choice: When Choice Impairs Social Welfare. *Journal of Public Policy & Marketing* 25 (1): 24–38.
Briesch, R.A., P.K. Chintagunta, and E.J. Fox. 2009. How Does Assortment Affect Grocery Store Choice? *Journal of Marketing Research* 46 (2): 176–89.
Broniarczyk, S.M., and W.D. Hoyer. 2006. *Retail Assortment: More ≠ Better. Retailing in the 21st Century*, 225–38. Berlin, Heidelberg: Springer.
Broniarczyk, S. M., W. D. Hoyer, and L. McAlister. 1998. Consumers' Perceptions of the Assortment Offered in a Grocery Category: The Impact of Item Reduction. *Journal of Marketing Research* 35 (2): 166–176.
Burdette, H.L., and R.C. Whitaker. 2004. Neighborhood Playgrounds, Fast Food Restaurants, and Crime: Relationships to Overweight in Low-Income Preschool Children. *Preventive Medicine* 38 (1): 57–63.
Caillavet, F., G. Kyureghian, R.M. Nayga, Jr, C. Ferrant, and P. Chauvin. 2015. Does Healthy Food Access Matter in a French Urban Setting? *American Journal of Agricultural Economics* 97 (5): 1400–16.
Caspi, C.E., I. Kawachi, S.V. Subramanian, G. Adamkiewicz, and G. Sorensen. 2012. The Relationship between Diet and Perceived and Objective Access to Supermarkets among Low-Income Housing Residents. *Social Science & Medicine* 75 (7): 1254–62.
Chen, D., E.C. Jaenicke, and R.J. Volpe. 2016. Food Environments and Obesity: Household Diet Expenditure versus Food Deserts. *American Journal of Public Health* 106 (5): 881–8.
Chen, S.E., and R.J.G.M. Florax. 2010. Zoning for Health: The Obesity Epidemic and Opportunities for Local Policy Intervention. *The Journal of Nutrition* 140 (6): 1181–4.
Chernev, A. 2011. Product Assortment and Consumer Choice: An Interdisciplinary Review. *Foundations and Trends® in Marketing* 6 (1): 1–61.
Chong, J.K., T.C. Ho, and C. Tang. 2001. A Modeling Framework for Category Assortment Planning. *Manufacturing & Service Operations Management* 3 (3): 191–210.
Cleary, R., A. Bonanno, L. Chenarides, and S.J. Goetz. 2018. Store Profitability and Public Policies to Improve Food Access in Non-Metro U.S. Counties. *Food Policy* 75: 158–70.
Courtemanche, C., and A. Carden. 2011. Supersizing Supercenters? The Impact of Walmart Supercenters on Body Mass Index and Obesity. *Journal of Urban Economics* 69 (2): 165–81.

Cummins, S., and S. Macintyre. 2002. “Food Deserts”—Evidence and Assumption in Health Policy Making. *BMJ (Clinical Research Ed.)* 325 (7361): 436–438.

Cummins, S., E. Flint, and S.A. Matthews. 2014. New Neighborhood Grocery Store Increased Awareness of Food Access but Did Not Alter Dietary Habits or Obesity. *Health Affairs (Project Hope)* 33 (2): 283–91.

Currie, J., S. DellaVigna, E. Moretti, and V. Pathania. 2010. The Effect of Fast Food Restaurants on Obesity and Weight Gain. *American Economic Journal: Economic Policy* 2 (3): 32–63.

Dhar, S.K., S.J. Hoch, and N. Kumar. 2001. Effective Category Management Depends on the Role of the Category. *Journal of Retailing* 77: 165–84.

Dutko, P., M. Ver Ploeg, and T.L. Farrigan. 2012. *Characteristics and Influential Factors of Food Deserts*. Washington DC: U.S. Department of Agriculture, Economic Research Service, Economic Research Report No. 140.

Ellickson, P.B. 2006. Quality Competition in Retailing: A Structural Analysis. *International Journal of Industrial Organization* 24 (3): 521–40.

——. 2007. Does Sutton Apply to Supermarkets? *The RAND Journal of Economics* 38 (1): 43–59.

Fan, L., K. Baylis, C. Gundersen, and M. Ver Ploeg. Forthcoming. Does a Nutritious Diet Cost More in Food Deserts? *Agricultural Economics*.

Farahmand, M., P. Amiri, F. Ramezani Tehrani, A.A. Momenan, P. Mirmiran, and F. Azizi. 2015. What Are the Main Barriers to Healthy Eating among Families? A Qualitative Exploration of Perceptions and Experiences of Tehranian Men. *Appetite* 89: 291–7.

Feenstra, R.C. 1994. New Product Varieties and the Measurement of International Prices. *The American Economic Review* 84 (1): 157–77.

Ford, P.B., and D.A. Dzewaltowski. 2010. Limited Supermarket Availability Is Not Associated with Obesity Risk among Participants in the Kansas WIC Program. *Obesity* 18 (10): 1944–51.

Fox, E.J., A.L. Montgomery, and L.M. Lodish. 2004. Consumer Shopping and Spending across Retail Formats. *The Journal of Business* 77 (S2): S25–60.

Handbury, J., and D.E. Weinstein. 2014. Goods Prices and Availability in Cities. *The Review of Economic Studies* 82 (1): 258–96.

Handbury, J., I. Rahkovsky, and M. Schnell. 2015. Is the Focus on Food Deserts Fruitless? Retail Access and Food Purchases Across the Socioeconomic Spectrum. Working Paper No. w21126, National Bureau of Economic Research.

Hausman, J.A., and G.K. Leonard. 2002. The Competitive Effects of a New Product Introduction: A Case Study. *The Journal of Industrial Economics* 50 (3): 237–63.

Hausman, J.A., and W.E. Taylor. 1981. Panel Data and Unobservable Individual Effects. *Econometrica: Journal of the Econometric Society* 1377–98.

Hesse, M., and J.P. Rodrigue. 2004. The Transport Geography of Logistics and Freight Distribution. *Journal of Transport Geography* 12 (3): 171–84.

Hillier, A., T. Smith, C.C. Cannuscio, A. Karpyn, and K. Glanz. 2015. A Discrete Choice Approach to Modeling Food Store Access. *Environment and Planning B: Planning and Design* 42 (2): 263–78.

Hoch, S.J., X. Dreze, and M.E. Purk. 1994. EDLP, Hi-Lo, and Margin Arithmetic. *The Journal of Marketing* 16–27.

Iyengar, S.S., and M.R. Lepper. 2000. When Choice Is Demotivating: Can One Desire Too Much of a Good Thing? *Journal of Personality and Social Psychology* 79 (6): 995.

Jeffery, R.W., J. Baxter, M. McGuire, and J. Linde. 2006. Are Fast Food Restaurants an Environmental Risk Factor for Obesity? *International Journal of Behavioral Nutrition and Physical Activity* 3 (2). https://ijbnpa.biomedcentral.com/articles/10.1186/1479-5868-3-2.

Jetter, K., and D.L. Cassady. 2006. The Availability and Cost of Healthier Food Alternatives. *American Journal of Preventive Medicine* 30 (1): 38–44.

Kahn, B.E., and D.R. Lehmann. 1991. Modeling Choice among Assortments. *Journal of Retailing* 67 (3): 274–99.

Kwak, K., S.D. Duvvuri, and G.J. Russell. 2015. An Analysis of Assortment Choice in Grocery Retailing. *Journal of Retailing* 91 (1): 19–33.

Kyureghian, G., R.M. Nayga, Jr, and S. Bhattacharya. 2013. The Effect of Food Store Access and Income on Household Purchases of Fruits and Vegetables: A Mixed Effects Analysis. *Applied Economic Perspectives and Policy* 35 (1): 69–88.

Larson, N.I., M.T. Story, and M.C. Nelson. 2009. Neighborhood Environments: Disparities in Access to Healthy Foods in the U.S. *American Journal of Preventive Medicine* 36 (1): 74–81.

Mantrala, M.K., M. Levy, B.E. Kahn, E.J. Fox, P. Gaidarev, B. Dankworth, and D. Shah. 2009. Why Is Assortment Planning so Difficult for Retailers? A Framework and Research Agenda. *Journal of Retailing* 85 (1): 71–83.

Moore, L.V., and A.V. Diez Roux. 2006. Associations of Neighborhood Characteristics with the Location and Type of Food Stores. *American Journal of Public Health* 96 (2): 325–31.

Mussa, M., and S. Rosen. 1978. Monopoly and Product Quality. *Journal of Economic Theory* 18 (2): 301–17.

Muth, M.K., M. Sweitzer, D. Brown, K. Capogrossi, S. Karns, D. Levin, A. Okrent, P. Siegel, and C. Zhen. 2016. Understanding IRI Household-Based and Store-Based Scanner Data. Technical Bulletin 1942, Washington DC: U.S. Department of Agriculture, Economic Research Service.

Neckerman, K.M., M. Bader, M. Purciel, and P. Yousefzadeh. 2009. Measuring Food Access in Urban Areas. Working Paper.

Richards, T.J., and S.F. Hamilton. 2006. Rivalry in Price and Variety among Supermarket Retailers. *American Journal of Agricultural Economics* 88 (3): 710–26.

Slack, T., C.A. Myers, C.K. Martin, and S.B. Heymsfield. 2014. The Geographic Concentration of U.S. Adult Obesity Prevalence and Associated Social, Economic, and Environmental Factors. *Obesity* 22 (3): 868–74.

Sturm, R., and A. Datar. 2005. Body Mass Index in Elementary School Children, Metropolitan Area Food Prices and Food Outlet Density. *Public Health* 119 (12): 1059–68.

Thomsen, M.R., R.M. Nayga, Jr., P.A. Alviola, IV, and H.L. Rouse. 2015. The Effect of Food Deserts on the Body Mass Index of Elementary Schoolchildren. *American Journal of Agricultural Economics* 98 (1): 1–18.

Ver Ploeg, M., and P. Dutko. 2013. *Economic Research Service Documentation on the Food Access Research Atlas.* Washington DC: U.S. Department of Agriculture.

Ver Ploeg, M., V. Breneman, P. Dutko, R. Williams, S. Snyder, C. Dicken, and P. Kaufman. 2012. Access to Affordable and Nutritious Food: Updated Estimates of Distance to Supermarkets Using 2010 Data. Washington DC: U.S. Department of Agriculture, Economic Research Service, Economic Research Report No. 143.

Ver Ploeg, M., V. Breneman, T. Farrigan, K. Hamrick, D. Hopkins, P. Kaufman, B.H. Lin, et al. 2009. Access to Affordable and Nutritious Food: Measuring and Understanding Food Deserts and Their Consequences. Washington DC: U.S. Department of Agriculture, Economic Research Service, Report to Congress.

Ver Ploeg, M., L. Mancino, J.E. Todd, D.M. Clay, and B. Scharadin. 2015. Where Do Americans Usually Shop for Food and How Do They Travel to Get There? Initial Findings from the National Household Food Acquisition and Purchase Survey. Washington DC: U.S. Department of Agriculture, Economic Research Service, Economic Information Bulletin No. 138.

Volpe, R., A. Okrent, and E. Leibtag. 2013. The Effect of Supercenter-Format Stores on the Healthfulness of Consumers' Grocery Purchases. *American Journal of Agricultural Economics* 95 (3): 568–89.

Wrigley, N., D. Warm, B. Margetts, and A. Whelan. 2002. Assessing the Impact of Improved Retail Access on Diet in a "Food Desert": A Preliminary Report. *Urban Studies* 39 (11): 2061–82.