Geographic Accessibility of Food Outlets Not Associated with Body Mass Index Change Among Veterans, 2009-14
Zenk, Tarlov, Wing, Matthews, Jones, Tong, & Powell. Health Aff (Millwood). 2017;36(8).

APPENDICES

Appendix Exhibit A. Technical documentation

Design

The Weight and Veterans' Environments Study (WAVES) is a 7-year retrospective longitudinal cohort study of U.S. adults who were military veterans receiving primary healthcare services in the U.S. Department of Veteran Affairs (VA) between 2009 and 2014 and who were followed to date through 2015. This paper links 6 years (2009-2014) of individual-level data from the VA Corporate Data Warehouse, a repository of clinical and administrative data from the electronic health record and other sources, with non-VA data on food outlet locations.

Linking food outlet data with clinical data is complicated in practice because the home address data, clinical data, and food outlet data are updated on different schedules. At the time we obtained the person-level home address geocodes from the VHA Planning Systems Support Group, the data that were available contained best known addresses as of the end of each VA fiscal year (September 30). To operationalize our work, a first decision was to prioritize the second half of the year (July 1-December 31) for BMI measurement in order to maximize the likelihood that patients were at that address when their weight was measured. Additionally, we pursued the general goal of measuring the environment at a time point preceding the outcome measures in each year. Since our study was designed to evaluate the effects of environmental factors (food outlet accessibility) on BMI, we used the basic policy of temporal precedence to make it clear that changes in environmental factors came before changes in BMI and not the other way around. Specifically, we linked individual BMI to food outlet accessibility measures in the 4th quarter of the previous year. For example, 2009 BMI values were joined to supermarket measures derived from 4th quarter 2008 supermarket location data.

Sample

The sample for the analysis in this paper consisted of 1.7 million working-age adults (20-64 years old) residing in metropolitan counties. The sample is derived from a larger study cohort of 3.2 million U.S. military veterans aged 20-80 years who lived in the continental U.S.(1). The exclusion criteria eliminated patients who did not have at least one VA healthcare encounter in the two years prior to baseline; patients who resided in a long-stay nursing home at baseline; patients who did not have at least one home address (not PO Box address) that could be geocoded to the street or ZIP+4 level during the study period; patients without at least one valid and clinically plausible height and weight measurement during the study period; and patients over the age of 65 because of multiple possible lifestyle, mobility, and socioeconomic differences among older versus younger, working-age adults that might manifest in very different relationships between the residential environment and BMI.

Measures

Body mass index (BMI)

The dependent variable in the paper is the patient's BMI (weight in kg / height in m^2) in a given calendar year. Practical challenges in working with electronic medical record data include having no control over data collection periodicity, frequency, or quality. Ancillary rules were needed to address these issues and to impose an annual structure on the data.

BMI is a weight-for-height measure that is intended to standardize weight measurements in a way that accounts for differences in body structure across individuals that are relatively permanent and that are unrelated to things like diet and exercise. In essence, weight measurements are scaled by the square of a person's height in meters. We cleaned the height measures in the medical records by deleting implausible measurements (<48 inches or >85 inches). We then defined each patient's height to be the modal value of all available height measurements taken during the entire study period. When available height measurements had no modal value, we used the mean instead. Although it is possible that a patient's height changes somewhat over time, we felt that difference in observed height measurements for a patient were more likely to reflect measurement errors or data entry errors than genuine changes in a patient's height. Since small error in height measurement can have outsized effects on BMI (because it is squared in the denominator), use of the most frequently appearing value maximized the likelihood that we would be using the patient's true height (rather than a mean or median which, though a good estimate, would be less likely to exactly equal the patient's true height). Taking a modal value each year was not feasible because the majority of patients did not have enough height measurements within a single year to identify a mode. A limitation of our approach is that it does not capture real changes in height arising from the height loss that may occur with aging. Given that our study sample was limited to patients under age 65 at baseline (mean 52 and 43 among men and women, respectively) and the study follow-up time of not more than 6 years, we think undetected measurable height loss likely was infrequent. Still, among those who lose height over the time period, our approach will underestimate BMI loss and overestimate BMI gain. We can think of no reason why unmeasured losses in height would tend to occur more rapidly in patients exposed to time-varying changes in retail food environments and so we do not expect unmeasured height loss to generate bias in our estimates of the effects of the food environment on BMI. Ultimately our approach reflects our judgement that measurement error was a larger threat to accuracy than time-related decreases in height.

We used a multi-step procedure to define an annual weight measurement for each member of the sample. First, we set each patient's weight in a given year equal to the average value of all of the weight measurements available for the patient during the second half of the calendar year. If no valid weight measurement was available during the second half of the year or if the BMI implied by the average weight in the second half of the year was not 15.0-75.0 kg/m^2, we used the average weight value from the entire calendar year. We were able to compute annual BMI measures using the second half of the year approach for 80.7% of 7,441,544 person-year observations in our analysis. We used the full year averaging approach for the remaining 20.3% of the observations.

Geographic accessibility of retail food outlets

Following our review of validation studies (2), we purchased food store data from InfoUSA and fast food restaurant data from Dun & Bradstreet. After cleaning the home address geocodes and retail food outlet data in order to maximize their accuracy and utility (e.g., reclassifying some records by store type, deduplicating records)(2), we constructed annual (4^{th} quarter) measures of the geographic accessibility of chain supermarkets [standard industrial classification (SIC) codes 541101-541109 (excluding 541103, convenience stores) and >$2M annual sales, or name listed in Supermarket News Top 75 Retailers and Wholesalers in any year between 2010 and 2014]; non-chain supermarkets [SIC codes 541101-541109 (excluding 541103) and >$2M annual sales but name not listed in Supermarket News]; supercenters and other non-membership mass merchandisers (SIC code 53 and Walmart, Kmart, Target, or Meijer in name); chain fast food restaurants including pizza [SIC code 58120601 or 581203 and name listed in National Restaurant News Top 200 between 2007 and 2013 or name listed in Quick Service or Fast Casual or Quick Service Restaurant Top 50 between 2007 and 2013 (but not coffee shops: 58120304)]; and non-chain fast food restaurants (SIC code 58120602 or 581203 and name not in National Restaurant News or Quick Serve Restaurant lists). Most U.S. households shop at mass merchandisers or supermarkets, particularly chain stores(3,4).

We defined the relative accessibility of supermarkets to fast food restaurants as the percentage of food outlets (supermarkets and fast food restaurants) that were supermarkets. Mass merchandisers were not included in the relative accessibility measure because of mixed conceptual and empirical evidence for their potential impact on BMI.

We used a raster approach, inspired by "smartmaps" (5) to construct our measures of geographic accessibility of retail food outlets. Specifically, we divided the continental U.S. into 30m x 30m grid cells with approximately 9 billion cells. Retail food outlet accessibility measures are based on each grid cell's centroid (geometric center) and calculated as the number of outlets within a 1-mile (1mi) radius and a 3-mile (3mi) radius. For each study year, we assigned the value of each retail food outlet measure to each patient based on the cell in which his or her home geocode was located.

A patient's retail food outlet measures can vary over time for two reasons: (a) individual migration and (b) neighborhood change. Environmental variation over time because of individual migration occurs whenever a patient moves to a new home address and the new address has a different number of nearby retail food outlets. We considered a patient to have moved if home geocodes based on addresses from adjacent years were more than 0.25 miles apart. A patient's environment may change without any migration because of the opening and closing of retail food outlets, which we refer to as neighborhood change. Accessibility within 3 miles may be particularly relevant for food stores, given that multiple studies show that individuals travel between two and four miles from home to shop for groceries (4,6-10). Accessibility within 1 mile may be more relevant for fast food restaurants where individuals often purchase prepared foods or snacks for home consumption.

To avoid strong functional form assumptions about the relationship between the number of nearby food outlets and BMI, we grouped the members of our sample into discrete categories of food outlet accessibility. When categorizing the food outlet variables, we considered several options. Our goal was to compare having different levels of food outlets (e.g., a little, some, a lot) to having no food outlet. The variable distributions shaped whether the variable was dichotomized or categorized based on tertiles or quartiles. We used a hierarchy of decision rules. When more than 50% of people had none of an outlet within the (1- or 3-mile radius) area, we created binary variables (0, 1 or more). This prevented the construction of scarcely populated categories. For other outlets, we created 4-category variables. When at least 10% (and <50%) of the people had no outlet within that distance, we derived a 4-level variable: 0, and then tertiles of the non-zero distribution. When <10% of people had none of the outlet within the specified area, we categorized the variable based on quartiles of the entire distribution to avoid having a scarcely populated reference category. For the relative accessibility there was an additional category of no supermarket or fast food restaurant within that distance because we conceived of having neither a supermarket or a fast food outlet as having potentially different effects than having no supermarket (or a low number of supermarkets within 3 miles) but at least one fast food restaurant. The Table shows how each food outlet was categorized at 1mi and 3mi.

Table. Approach for categorizing food outlets

	Dichotomous (0, 1 or more)	0 and tertiles of the remaining non-zero distribution of values	Quartiles of the distribution of values
Chain supermarkets, 1 mi	X		
Non-chain supermarkets, 1 mi	X		
Mass merchandisers, 1 mi	X		
Grocery stores, 1 mi	X		
Chain fast food restaurants, 1 mi		X	
Non-chain fast food restaurants, 1 mi		X	
Convenience stores, 1 mi		X	
Relative accessibility supermarkets to fast food, 1 mi		X	
Chain supermarkets, 1 mi		X	
Non-chain supermarkets, 3 mi		X	
Mass merchandisers, 3 mi		X	
Grocery stores, 3 mi		X	
Chain fast food restaurants, 3 mi		X	
Non-chain fast food restaurants, 3 mi			X
Convenience stores, 3 mi			X
Relative accessibility supermarkets to fast food, 3 mi			X

Covariates
Individual-level time-invariant variables included baseline age and race/ethnicity (non-Hispanic white, non-Hispanic black, Hispanic, non-Hispanic other/unknown). We supplemented missing race and ethnicity information in the VA data with Medicare data on race from the VA-CMS Data Repository (11,12).

Individual-level, time-varying covariates included marital status (married, separated or divorced, widowed, single, unknown) and ten chronic health conditions associated in prior research with both BMI and independently with diet and/or physical activity (breast cancer, cerebrovascular disease, colon cancer, congestive heart failure, depression, diabetes, hyperlipidemia, hypertension, myocardial infarction, and osteoporosis).

We included several area-level covariates: census division (New England, Middle Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, Mountain, Pacific) and urbanicity measured at the county level (large central metro, large fringe metro, medium metro, and small metro)(1,13). County urban-rural classification codes were available for 2006 and 2013 only; thus, we assigned 2006 NCHS urban-rural classification codes to patients' residential location for years 2009-2012 and 2013 NCHS urban-rural classification codes to patients' residential location for years 2013-14.

We also adjusted for small-area demographics. Based on census tract of residence, we assigned each patient local tract-level demographic information using the American Community Survey 5-year estimates of SES (percent of residents with annual incomes below the federal poverty line and median household income, both categorized into deciles of the distribution of the values for all continental U.S. census tracts) and population density (number of residents per land area, categorized into quartiles based on all continental U.S. census tracts). Given the delay in annual releases of successive 5-year ACS estimates, we used a 2-year lag based on the ACS 5-year midpoint for linking patient measures to ACS measures (e.g., 2009 patient BMI linked to 2005-2009 ACS data, midpoint 2007; 2014 patient BMI linked to 2010-2014 ACS data).

In addition, we controlled for accessibility of parks (1mi), fitness facilities (1mi), and other retail food outlets (1mi or 3mi depending on the model): grocery stores (SIC codes 541101-541109 [excluding 541103], <$2M annual sales, and name not listed in Supermarket News) and convenience stores (SIC codes 541103, 554101, 554103). We obtained grocery store, convenience store, and fitness facility data from InfoUSA. We combined data from TeleAtlas and NAVTEQ to derive the park measures.

Data analysis
We estimated pooled (all years) cross-sectional models with year fixed effects and longitudinal models that also incorporated individual-level fixed effects to examine associations between food outlet accessibility and BMI.

Cross-sectional models
To understand our statistical modeling strategy in more detail, let BMI_{it} be the BMI associated with patient i in study year t as described above. X_i is a vector of patient-level time-invariant characteristics. Z_{it} is a vector of time-varying patient-level characteristics and characteristics of

the patient's environment (small area demographics, accessibility of parks, convenience stores, etc). FE_{it} is a vector of time-varying food environment variables. Depending on the model, FE_{it} may include indicator variables for several different levels of supermarket counts, fast food restaurant counts, mass merchandiser counts, and relative accessibility measures. And depending on the model, these measures may be defined on either a 1-mile radius or a 3-mile radius around the patient's place of residence. With that notation as background, we fit the following regressions using OLS:

$$BMI_{it} = X_i\alpha + Z_{it}\beta + FE_{it}\delta + \theta_t + \epsilon_{it}$$

In the model, θ_t is a year-specific intercept and ϵ_{it} is an error term. We estimated standard errors that allowed for observations to be correlated within counties. δ is the vector of coefficients on the food outlet variables. These coefficients measure the cross-sectional association between the food outlet variables and BMI, after controlling for time period effects, time-varying covariates, and time-invariant covariates. Under the strong assumption that there are no unmeasured variables that are associated with both BMI and the food outlet variables, δ captures the causal effects of the geographic accessibility of food outlets on BMI.

Fixed effects models

Residential self-selection bias or omitted variable bias is an important threat to the validity of the cross-sectional regression models. A basic worry is that people decide where to live partly because of their preferences for different food environments. It is possible that a person's food environment preferences are associated with his or her BMI. Together these two points raise concerns that the coefficient on the food outlet variables in the cross-sectional regressions may be biased in ways that make food outlet access look like a more important determinant of BMI than it really is. For example, unmeasured lifestyle preferences that make people like living near fast food restaurants and might also lead them to have higher BMI. In that case, the cross-sectional regression coefficient on measures of fast food restaurant accessibility will reflect both the causal effects of the restaurants and the unmeasured lifestyle factors. The results will imply that fast food restaurant accessibility increases BMI even though most of the relationship may have nothing to do with the restaurants themselves and will merely reflect lifestyle differences between people who choose to live near vs. far from fast food restaurants. It is important to note, though, that this concern arises from any unmeasured factor associated with both where someone lives (e.g., discrimination) and the associated environmental exposures.

To avoid these kinds of confounding interpretations, we took advantage of the longitudinal structure of our data to estimate person fixed effects regression models. These models isolate the causal effect of the food outlet variables among patients who experience a change in their residential food environment. The key assumption required in this type of analysis is that the confounding factors that threaten the validity of the cross-sectional models are time invariant over the study time period. That is, these models work under the assumption that the lifestyle factors that (partially) shape residential choices do not themselves change over time. Arguments like this one apply to any unmeasured confounding patient characteristic that does not change over the study time period. Like the cross-sectional models, the person fixed effects models also allow for a flexible time trend that may which may capture changes in economic conditions, market environments, and health behaviors that could confound food outlet-BMI associations.

The basic form of the fixed effects model that we work with is:

$$BMI_{it} = Z_{it}\beta + FE_{it}\delta + \theta_t + \lambda_i + \epsilon_{it}$$

In this model, λ_i represents a full set of person fixed effects. The time-invariant covariates contained in X_i are absorbed into the person specific intercepts, along with any unmeasured time-invariant factors that may have generated omitted variable bias in the cross-sectional models. In these models, δ represents the causal effects of the food outlet variables under the assumption that there are no unmeasured time-varying confounders that are associated with both BMI and changes in food outlet accessibility.

These person fixed effect models exploit two conceptually different sources of within-person change in food outlet accessibility in sequential years: change due to individual migration (i.e., a person moving to a new address with a net change in the prevalence of food outlets) and neighborhood change for non-migrants (i.e., the openings and closings of food outlets), which can affect patients whose home address does not change (i.e., non-migrants). The distinction may be important because residential self-selection bias may still be a problem among migrants. For example, patients may decide to move to a neighborhood where supermarkets are more accessible because of a negative health event that makes it harder for them to travel. In that case, even the fixed effects model may be biased because the (unobserved) change in underlying health status may affect both BMI and food outlet accessibility. To explore these concerns, we also estimated separate fixed effects models for non-migrants. Fixed effects models applied to a sample of non-migrants rely only on within-person variation from neighborhood evolution, which may be less prone to bias from time-varying factors that may prompt people to move to a new environment while also changing their BMI.

To test our main hypotheses, we ran each model twice, once for retail food outlet accessibility within 1 mile and again for retail food outlet accessibility within 3 miles of patients' homes. Cross-sectional and fixed effects models include several time-varying individual- and area-level covariates (see Measures). To examine whether associations differed by area economic characteristics, we added interaction terms between the food outlet access variables and area poverty level to the main effects models. Census tract poverty level was categorized using nationwide census tract tertiles as low (0-8.26%; mean=5.00), medium (8.27-17.71%; mean=12.54), or high (17.71-100%; mean=29.19). All models accounted for clustering of patients within counties at baseline using a Huber-White cluster robust variance matrix. Because men comprise almost 90% of the sample, we estimated separate models for men and women.

References

(1) National Center for Health Statistics. NCHS Urban-Rural Classification Scheme for Counties. 2015; Available at: https://www.cdc.gov/nchs/data_access/urban_rural.htm, 2017.

(2) Jones K, Zenk SN, Tarlov E, Powell L, Matthews SA, Horoi I. A step-by-step approach to improve data quality when using commercial business lists to characterize retail food environments. BMC Res Notes 2017;10(35).

(3) Stern D, Robinson WR, Ng SW, Gordon-Larsen P, Popkin BM. US Household Food Shopping Patterns: Dynamic Shifts Since 2000 And Socioeconomic Predictors. Health Aff 2015;34(11):1840-1848.

(4) Ver Ploeg M, Mancino L, Todd JE, Clay DM, Scharadin B. 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. Economic Information Bulletin 2015;138.

(5) Hurvitz PM, Moudon AV. Home Versus Nonhome Neighborhood. Am J Prev Med 2012;42(4):411.

(6) Zenk S, Schulz A, Israel B, Mentz G, Miranda P, Opperman A, et al. Food shopping behaviours and exposure to discrimination. Public Health Nutr 2014;17(5):1167-1176.

(7) Ghosh-Dastidar B, Cohen D, Hunter G, Zenk S, Huang C, Beckman R, et al. Distance to Store, Food Prices, and Obesity in Urban Food Deserts. Am J Prev Med 2014;47(5):587-595.

(8) Drewnowski A, Aggarwal A, Hurvitz PM, Monsivais P, Moudon AV. Obesity and Supermarket Access: Proximity or Price? Am J Public Health 2012 08;102(8):74-80.

(9) Liu J, Han B, Cohen D. Beyond Neighborhood Food Environments: Distance Traveled to Food Establishments in 5 US Cities, 2009-2011. Prev Chronic Dis 2015;12(8):E126.

(10) Fuller D, Cummins S, Matthews SA. Does transportation mode modify associations between distance to food store, fruit and vegetable consumption, and BMI in low-income neighborhoods? Am J Clin Nutr 2013;97(1):167-172.

(11) US Department of Veterans' Affairs. Available at: www.virec.research.va.gov/Index-VACMS.htm. Accessed 09/24, 2016.

(12) Stroupe KT, Tarlov E, Zhang Q, Haywood T, Owens A, Hynes DM. Use of Medicare and DOD data for improving VA race data quality. J Rehabil Res Dev 2010;47(1938-1352; 0748-7711; 8):781-796.

(13) Ingram DD, Franco SJ. 2013 NCHS Urban-Rural Classification Scheme for Counties. Vital Health Stat 2 2014 04(166):1-73

Appendix Exhibit B. Descriptive statistics for the total sample and non-migrants at baseline by sex

		Men (n=1,522,803)		Women (n=183,618)	
		Total sample	Non-migrant sample	Total sample	Non-migrant sample
		n=1,522,803	n=1,034,375	n=183,618	n=112,670
		% or Mean (SD)	% or Mean (SD)	% or Mean (SD)	% or Mean (SD)
Body mass index	Mean (SD)	30.2 (6.0)	30.3 (6.0)	29.5 (6.4)	29.6 (6.4)
Body weight status, %	Underweight or normal weight (BMI ≤ 24.9)	18.2	17.6	26.5	25.9
	Overweight (25 ≥ BMI ≤ 29.9)	35.9	35.9	31.1	31.2
	Obese (BMI ≥ 30)	45.9	46.5	42.4	42.9
Age	Mean (SD)	51.8 (11.5)	52.5 (11.3)	43.4 (11.5)	44.5 (11.3)
Marital status, %	Unknown	1.4	1.6	2.1	2.4
	Married	48.8	53.0	33.3	36.6
	Separated or divorced	26.2	23.8	31.6	30.1
	Widowed	1.8	1.7	2.2	2.2
	Single	21.8	19.9	30.8	28.7
Race/ethnicity, %	Non-Hispanic white	60.5	60.9	50.1	49.8
	Non-Hispanic black	22.5	20.8	32.1	31.1
	Hispanic	6.0	6.0	6.0	5.8
	Other	2.5	2.6	3.3	3.3
	Unknown	8.5	9.7	8.5	10.1
Medical diagnoses, %	Breast cancer	0.0	0.0	1.3	1.4
	Cerebrovascular disease	2.7	2.7	1.2	1.3
	Colon cancer	0.4	0.4	0.2	0.2
	Congestive heart failure	3.1	3.1	0.8	0.8
	Depression	20.1	18.2	29.2	27.1
	Diabetes	19.1	19.3	8.0	8.1

	Hyperlipidemia	32.4	33.1	17.2	19.1
	Hypertension	41.3	41.4	21.8	22.2
	Myocardial infarction	1.7	1.6	0.4	0.3
	Osteoporosis	0.5	0.5	1.6	1.8
Urbanicity, %	Large central metro	29.9	28.9	30.2	28.9
	Large fringe metro	24.0	24.4	24.1	24.3
	Medium metro	29.9	30.2	30.8	31.2
	Small metro	16.3	16.5	15.0	15.5
Census Division, %	New England	3.7	3.9	2.5	2.6
	Middle Atlantic	9.5	9.9	7.3	7.6
	East North Central	13.4	13.3	10.5	10.1
	West North Central	5.8	5.8	4.8	4.8
	South Atlantic	24.7	24.8	30.8	31.3
	East South Central	7.1	7.2	7.3	7.5
	West South Central	14.0	14.0	15.8	15.7
	Mountain	8.4	8.2	9.0	8.8
	Pacific Alaska	13.3	13.0	12.1	11.6
Median household income, Census tract	Mean (SD)	52334.3 (21346.8)	53374.4 (21462.8)	53192.7 (20672.6)	54160.5 (20930.6)
Poverty rate, Census tract	Mean (SD)	14.9 (11.5)	14.3 (11.0)	14.4 (10.8)	14.1 (10.5)
Population density (per square mile), Census tract	Mean (SD)	4139.5 (8866.6)	4050.6 (8957.9)	4034.3 (8274.4)	3957.9 (8525.6)
Chain supermarkets, 1mi[1]	1 or more stores	41.9	41.1	42.9	41.5
Non-chain supermarkets, 1mi[1]	1 or more stores	25.5	24.7	23.3	22.4
Mass merchandisers, 1mi[1]	1 or more stores	14.7	14.5	16.2	15.5
Grocery stores, 1mi[1]	1 or more stores	48.2	46.8	47.5	45.8
Chain fast food restaurants, 1mi[2]	0 restaurants	19.4	19.8	20.0	20.3
	1-2 restaurants	24.6	24.3	25.3	24.6
	3-6 restaurants	21.9	20.7	22.6	21.1
	7 or more restaurants	34.2	35.2	32.1	33.9

Non-chain fast food restaurants, 1mi[2]	0 restaurants	23.5	24.4	25.3	26.1
	1-2 restaurants	25.8	25.8	27.0	26.2
	3-6 restaurants	24.2	22.7	22.2	20.5
	7 or more restaurants	26.5	27.1	25.6	27.2
Convenience stores, 1mi[2]	0 stores	22.1	23.1	23.6	24.6
	1-2 stores	24.5	24.6	26.2	26.0
	3-5 stores	30.0	28.3	28.6	26.6
	6 or more stores	23.4	23.9	21.6	22.9
Relative accessibility of supermarkets to fast food restaurants {Supermarkets /(Supermarkets + Fast Food Restaurants) * 100}, 1mi[3]	Low (0)	26.9	27.5	28.7	29.1
	Low-mid (0.4-9.1%)	12.2	11.5	11.8	10.9
	Mid-high (9.1-16.7%)	21.6	21.2	21.8	21.0
	High (16.8-100%)	19.0	19.2	19.1	19.2
	No supermarket or fast food restaurant	20.3	20.7	18.6	19.8
Parks, 1mi	0 parks	17.8	18.3	19.0	19.4
	1 park	22.4	22.4	22.2	21.8
	2-3 parks	26.3	25.3	24.0	23.0
	4 or more parks	33.5	34.1	34.8	35.9
Fitness facilities, 1mi	0 facilities	27.5	28.1	28.7	29.3
	1-2 facilities	17.2	17.2	17.9	17.5
	3-4 facilities	26.8	25.8	26.1	24.6
	5 or more facilities	28.5	28.9	27.2	28.6
Chain supermarkets, 3mi[4]	0 stores	22.3	23.3	23.6	24.7
	1-2 stores	33.6	33.7	35.7	35.4
	3-6 stores	24.6	23.4	24.4	23.1
	7 or more stores	19.5	19.6	16.2	16.8
Non-chain supermarkets, 3mi[4]	0 stores	30.0	21.4	21.2	21.5
	1 store	20.6	20.5	20.8	20.2
	2-3 stores	21.6	20.5	19.4	18.3
	4 or more stores	36.8	37.6	38.5	40.0

Mass merchandisers, 3mi[4]	0 stores	25.0	25.2	26.3	26.7
	1 store	20.1	20.0	21.7	21.4
	2 stores	18.3	18.1	19.4	18.6
	3 or more stores	36.6	36.8	32.6	33.3
Chain fast food restaurants, 3mi[4]	0 restaurants	28.3	29.6	27.7	29.3
	1-14 restaurants	29.8	30.2	32.0	32.2
	15-32 restaurants	30.7	29.1	31.7	29.6
	33 or more restaurants	11.2	11.0	8.6	8.9
Non-chain fast food restaurants, 3mi[5]	0-5 restaurants	23.5	24.0	21.0	22.3
	6-18 restaurants	25.0	26.2	28.0	29.3
	19-39 restaurants	24.8	24.8	26.5	25.9
	40 or more restaurants	26.7	25.1	24.4	22.5
Convenience stores, 3mi[5]	0-6 stores	23.6	24.1	20.3	21.5
	7-18 stores	23.4	24.5	25.4	26.6
	19-36 stores	24.6	24.7	27.0	26.6
	37 or more stores	28.4	26.7	27.3	25.3
Grocery stores, 3mi[4]	0 stores	27.0	28.1	28.2	29.3
	1-3 stores	26.4	26.6	29.3	29.1
	4-11 stores	28.3	26.6	26.3	24.7
	12 or more stores	18.4	18.6	16.3	17.0
Relative accessibility of supermarkets to fast food restaurants, 3mi[6]	Low (0-7.4%)	23.2	23.3	24.1	24.0
	Mid-low (7.4-10.5%)	24.6	24.5	25.2	24.6
	Mid-high (10.5-14.3%)	23.1	23.0	23.5	23.3
	High (14.3-100%)	23.5	24.0	23.1	23.8
	No supermarket or fast food restaurant	5.6	5.3	4.2	4.3

Authors' analysis of participant data from the VA corporate Data Warehouse, 2009-2014; Census tract demographic data from US Census Bureau (2005-2009, 2006-2010, 2007-2011, 2008-2012, 2009-2013, 2010-2014); Food store data from InfoUSA (2008-2013); Fast food restaurant data from Dun & Bradstreet (2008-2013).

[1]For food outlets for which less than 50% of the sample had an outlet within 1 mile (chain supermarkets, non-chain supermarkets, mass merchandisers, grocery stores), we used a binary variable (0, 1 or more).

[2]For food outlets for which at least 10% of the sample had no outlet within 1 mile (chain fast food restaurants, non-chain fast food restaurants, convenience stores), we used a 4-category variable, constructed as 0 and then tertiles of the non-zero distribution of values.

[3]A 5-category variable was used for relative accessibility: no supermarket or fast food restaurant, low or no supermarket (but at least one fast food restaurant), and then low-mid, mid-high, and high based on tertiles of the remaining non-zero distribution.

[4]For food outlets for which at least 10% of the sample had no outlet within 3 miles (chain supermarkets, non-chain supermarkets, mass merchandisers, chain fast food restaurants, grocery stores), we used a 4-category variable, constructed as 0 and then tertiles of the non-zero distribution of values.

[5]For other food outlets (non-chain fast food restaurants, convenience stores), we used a 4-category variable based on quartiles of the distribution of values.

[6]A 5-category variable was used for relative accessibility: no supermarket or fast food restaurant, and then low, low-mid, mid-high, and high based on quartiles of the remaining distribution of values.

Appendix Exhibit C. Cross-sectional and longitudinal associations between BMI and geographic accessibility of food outlets (1 mile and 3 miles) by sex

	Men			Women		
	Cross-sectional	Longitudinal		Cross-sectional	Longitudinal	
	Total sample	Total sample	Non-migrant sample	Total sample	Total sample	Non-migrant sample
Persons	n=1,522,803	n=1,522,803	n=1,034,375	n=183,618	n=183,618	n=112,670
Person-year observations	n=6,668,033	n=6,668,033	n=4,229,727	n=773,511	n=773,511	n=424,329
	Coefficient (SE)	**Coefficient (SE)**	**Coefficient (SE)**	**Coefficient (SE)**	**Coefficient (SE)**	**Coefficient (SE)**
1 mile						
Chain supermarkets						
0 stores	1.000	1.000	1.000	1.000	1.000	1.000
1 or more stores	-0.010	0.007*	0.001	-0.034	0.009	-0.002
	(0.014)	(0.003)	(0.005)	(0.036)	(0.011)	(0.018)
Non-chain supermarkets						
0 stores	1.000	1.000	1.000	1.000	1.000	1.000
1 or more stores	-0.027	-0.007	-0.002	-0.020	-0.005	-0.033
	(0.017)	(0.004)	(0.005)	(0.039)	(0.011)	(0.021)
Mass merchandisers						
0 stores	1.000	1.000	1.000	1.000	1.000	1.000
1 or more stores	0.123***	0.011*	-0.004	0.081	0.019	-0.000
	(0.016)	(0.005)	(0.007)	(0.043)	(0.014)	(0.025)
Chain fast food restaurants						
0 restaurants	1.000	1.000	1.000	1.000	1.000	1.000
1-2 restaurants	0.027	0.016***	0.018**	0.059	0.009	0.009
	(0.014)	(0.005)	(0.006)	(0.039)	(0.014)	(0.022)
3-6 restaurants	0.057**	0.015**	0.012	0.139**	0.025	0.035
	(0.018)	(0.006)	(0.009)	(0.051)	(0.017)	(0.027)
7+ restaurants	0.085***	0.025***	0.019	0.168**	0.007	0.030
	(0.025)	(0.007)	(0.012)	(0.062)	(0.021)	(0.034)

Non-chain fast food restaurants						
0 restaurants	1.000	1.000	1.000	1.000	1.000	1.000
1-2 restaurants	-0.014	0.013***	0.013**	-0.035	-0.025*	-0.032
	(0.014)	(0.004)	(0.005)	(0.037)	(0.011)	(0.019)
3-6 restaurants	-0.062**	0.013**	0.015*	-0.130**	-0.013	0.002
	(0.020)	(0.005)	(0.007)	(0.047)	(0.014)	(0.025)
7+ restaurants	-0.206***	0.010	0.013	-0.426***	-0.024	-0.015
	(0.028)	(0.006)	(0.009)	(0.069)	(0.020)	(0.034)
3 miles						
Chain supermarkets						
0 stores	1.000	1.000	1.000	1.000	1.000	1.000
1-2 stores	0.006	0.003	-0.005	-0.078	0.012	-0.024
	(0.020)	(0.006)	(0.008)	(0.050)	(0.018)	(0.027)
3-6 stores	-0.039	0.003	-0.001	-0.106	0.016	-0.033
	(0.029)	(0.007)	(0.010)	(0.063)	(0.021)	(0.031)
7+ stores	-0.078	-0.001	-0.013	-0.156	0.004	-0.048
	(0.045)	(0.008)	(0.011)	(0.082)	(0.025)	(0.039)
Non-chain supermarkets						
0 stores	1.000	1.000	1.000	1.000	1.000	1.000
1 store	-0.001	-0.003	-0.004	0.073*	0.025*	0.018
	(0.017)	(0.004)	(0.005)	(0.035)	(0.012)	(0.016)
2-3 stores	0.010	-0.010*	-0.010	0.104*	0.026	0.030
	(0.024)	(0.005)	(0.006)	(0.046)	(0.015)	(0.021)
4+ stores	-0.080*	-0.017**	-0.015	-0.000	0.006	0.008
	(0.037)	(0.006)	(0.009)	(0.069)	(0.020)	(0.029)
Mass merchandisers						
0 stores	1.000	1.000	1.000	1.000	1.000	1.000
1 store	0.122***	0.009*	0.007	0.158***	-0.009	-0.026
	(0.019)	(0.005)	(0.007)	(0.045)	(0.015)	(0.025)
2 stores	0.196***	0.021***	0.016	0.161**	-0.010	-0.018
	(0.022)	(0.006)	(0.009)	(0.054)	(0.017)	(0.031)
3+ stores	0.283***	0.026***	0.012	0.338***	0.019	0.020
	(0.031)	(0.007)	(0.011)	(0.065)	(0.018)	(0.034)

Chain fast food restaurants						
0 restaurants	1.000	1.000	1.000	1.000	1.000	1.000
1-14 restaurants	0.024	0.012	-0.004	0.062	-0.041	-0.063
	(0.020)	(0.007)	(0.012)	(0.069)	(0.026)	(0.051)
15-32 restaurants	0.011	0.011	-0.003	0.069	-0.050	-0.017
	(0.029)	(0.009)	(0.014)	(0.088)	(0.030)	(0.058)
33+ restaurants	0.043	0.011	-0.003	0.111	-0.026	0.015
	(0.038)	(0.010)	(0.016)	(0.106)	(0.034)	(0.067)
Non-chain fast food restaurants						
0-5 restaurants	1.000	1.000	1.000	1.000	1.000	1.000
6-18 restaurants	-0.091***	0.010	0.014	-0.232***	-0.018	-0.009
	(0.022)	(0.005)	(0.007)	(0.057)	(0.016)	(0.022)
19-39 restaurants	-0.228***	0.005	0.009	-0.411***	-0.010	-0.009
	(0.033)	(0.007)	(0.009)	(0.078)	(0.022)	(0.031)
40+ restaurants	-0.370***	-0.002	0.002	-0.871***	-0.027	0.007
	(0.044)	(0.008)	(0.011)	(0.104)	(0.027)	(0.039)

Authors' analysis of participant BMI from the VA corporate Data Warehouse, 2009-2014; Urbanicity data from National Center for Health Statistics (2006, 2013); Census tract demographic data from US Census Bureau (2005-2009, 2006-2010, 2007-2011, 2008-2012, 2009-2013, 2010-2014); Food store data from InfoUSA (2008-2013); Fast food restaurant data from Dun & Bradstreet (2008-2013); Park data from TeleAtlas and NAVTEQ (2010, 2014); and Fitness facility data from InfoUSA (2008-2013).

Note: Covariates for cross-sectional and longitudinal models included year, marital status, multiple health conditions, region, population density, median household income, poverty, and accessibility of grocery stores, convenience stores, parks, and fitness facilities. Cross-sectional models also controlled for baseline age and race/ethnicity.

*$p \leq 0.05$

**$p \leq 0.01$

***$p \leq 0.001$

Appendix Exhibit D. Cross-sectional and longitudinal associations between BMI and relative accessibility of supermarkets to fast food restaurants within 1 mile and 3 miles by sex

	Men			Women		
	Cross-sectional	Longitudinal		Cross-sectional	Longitudinal	
	Total sample	Total sample	Non-migrant sample	Total sample	Total sample	Non-migrant sample
Persons	n=1,522,803	n=1,522,803	n=1,034,375	n=183,618	n=183,618	n=112,670
Person-year observations	n=6,668,033	n=6,668,033	n=4,229,727	n=773,511	n=773,511	n=424,329
	Coefficient (SE)	**Coefficient (SE)**	**Coefficient (SE)**	**Coefficient (SE)**	**Coefficient (SE)**	**Coefficient (SE)**
1 mile						
Low relative accessibility (0)	1.000	1.000	1.000	1.000	1.000	1.000
Low-mid relative accessibility (0.4-9.1%)	-0.037*	0.009	0.008	-0.094	0.003	0.006
	(0.018)	(0.005)	(0.007)	(0.049)	(0.014)	(0.024)
Mid-high relative accessibility (9.1-16.7%)	-0.000	0.008*	0.004	0.045	0.013	0.011
	(0.015)	(0.004)	(0.006)	(0.041)	(0.012)	(0.019)
High relative accessibility (16.8-100%)	0.004	0.002	<0.001	0.062	0.012	0.007
	(0.016)	(0.004)	(0.006)	(0.039)	(0.013)	(0.022)
No supermarkets or fast food restaurants	-0.045**	-0.018***	-0.015*	-0.058	0.015	0.035
	(0.015)	(0.005)	(0.006)	(0.041)	(0.014)	(0.023)
3 miles						
Low relative accessibility (0-7.4%)	1.000	1.000	1.000	1.000	1.000	1.000
Low-mid relative accessibility (7.4-10.5%)	-0.017	-0.002	-0.005	0.043	-0.003	-0.005
	(0.016)	(0.003)	(0.004)	(0.035)	(0.010)	(0.013)
Mid-high relative accessibility (10.5-14.3%)	-0.007	-0.002	-0.010	0.100*	-0.005	-0.007
	(0.021)	(0.004)	(0.005)	(0.044)	(0.012)	(0.017)
High relative accessibility (14.3-100%)	0.021	-0.001	-0.005	0.140**	0.025	-0.003
	(0.022)	(0.005)	(0.006)	(0.044)	(0.013)	(0.020)
No supermarkets or fast food restaurants	-0.071**	-0.021**	-0.004	-0.063	0.021	0.059
	(0.024)	(0.008)	(0.012)	(0.077)	(0.027)	(0.048)

Authors’ analysis of participant BMI from the VA corporate Data Warehouse, 2009-2014; Urbanicity data from National Center for Health Statistics (2006, 2013); Census tract demographic data from US Census Bureau (2005-2009, 2006-2010, 2007-2011, 2008-2012, 2009-2013, 2010-2014); Food store data from InfoUSA (2008-2013); Fast food restaurant data from Dun & Bradstreet (2008-2013); Park data from TeleAtlas and NAVTEQ (2010, 2014); and Fitness facility data from InfoUSA (2008-2013).

Note: Covariates for cross-sectional and longitudinal models included year, marital status, multiple health conditions, region, population density, median household income, poverty, and accessibility of grocery stores, convenience stores, mass merchandisers, parks, and fitness facilities. Cross-sectional models also controlled for baseline age and race/ethnicity.

*$p \leq 0.05$

**$p \leq 0.01$

***$p \leq 0.001$