Does unmeasured confounding influence associations between the retail food environment and body mass index over time? The Coronary Artery Risk Development in Young Adults (CARDIA) study

Pasquale E Rummo,1,2 David K Guilkey,2,3 Shu Wen Ng,1,2 Katie A Meyer,1,2 Barry M Popkin,1,2 Jared P Reis,4 James M Shikany5 and Penny Gordon-Larsen1,2*

1 Department of Nutrition, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, NC, USA, 2Carolina Population Center, Chapel Hill, North Carolina, 3Department of Economics, University of North Carolina at Chapel Hill, NC, USA, 4Division of Cardiovascular Sciences, National Heart, Lung, and Blood Institute, Bethesda, MD, USA and 5Division of Preventive Medicine, University of Alabama at Birmingham, AL, USA

*Corresponding author. Department of Nutrition, University of North Carolina at Chapel Hill, Carolina Population Center, 137 East Franklin Street, Campus Box #8120, Chapel Hill, NC 27514, USA. E-mail: pglarsen@email.unc.edu

Abstract

Background: Findings in the observational retail food environment and obesity literature are inconsistent, potentially due to a lack of adjustment for residual confounding.

Methods: Using data from the CARDIA study (n = 12174 person-observations; 6 examinations; 1985–2011) across four US cities (Birmingham, AL; Chicago, IL; Minneapolis, MN; Oakland, CA), we used instrumental-variables (IV) regression to obtain causal estimates of the longitudinal associations between the percentage of neighbourhood food stores or restaurants (per total food outlets within 1 km network distance of respondent residence) with body mass index (BMI), adjusting for individual-level socio-demographics, health behaviours, city, year, total food outlets and market-level prices. To determine the presence and extent of bias, we compared the magnitude and direction of results with ordinary least squares (OLS) and random effects (RE) regression, which do not control for residual confounding, and with fixed effects (FE) regression, which does not control for time-varying residual confounding.

Results: Relative to neighbourhood supermarkets (which tend to be larger and have healthier options than grocery stores), a higher percentage of grocery stores [mean = 53.4%; standard deviation (SD) = 31.8%] was positively associated with BMI [β = 0.05; 95% confidence interval (CI) = 0.01, 0.10] using IV regression. However, associations were negligible or null using OLS (β = −0.001; 95% CI = −0.01, 0.01), RE (β = −0.003; 95% CI = −0.01, 0.0001) and FE (β = −0.003; 95% CI = −0.01, 0.0002) regression.
Neighbourhood convenience stores and fast-food restaurants were not associated with BMI in any model.

**Conclusions:** Longitudinal associations between neighbourhood food outlets and BMI were greater in magnitude using a causal model, suggesting that weak findings in the literature may be due to residual confounding.

**Key words:** Instrumental-variables regression, neighbourhoods, retail food environment, obesity, weight, endogeneity

**Key Messages**
- Findings in the observational retail food environment and obesity literature are inconsistent, potentially due to a lack of adjustment for residual confounding.
- We sought to assess the presence and extent of bias from residual confounding by comparing estimates derived from causal models (instrumental-variables regression) with estimates derived from non-causal methods, including ordinary least squares and random effects regression, which do not account for residual confounding at all; and fixed effects regression, which only corrects for time-invariant residual confounding.
- Overall, estimates derived from non-causal models were attenuated relative to a causal modelling strategy, which suggests that non-causal models may underestimate the effect of the neighbourhood retail food environment on weight status.
- Using causal model strategies in future studies is important for informing efforts to modify neighbourhood retail food environments to improve health outcomes.

**Background**
In response to inequities in access to healthy food choices, policy makers have sought to modify the retail food environment in low-income areas.\(^1,2\) Theoretically, such efforts would influence where residents shop, what they consume and ultimately weight status. However, such experiments have not been successful in reducing obesity,\(^1,3-7\) despite some mixed supporting evidence from observational research. For example, findings from observational studies suggest a positive association between density of fast-food restaurants, convenience stores and grocery stores with body mass index (BMI),\(^8-13\) and a negative association between full-service restaurants and supermarkets with BMI.\(^8-10,13\) Yet Cobb et al., in a systematic review, reported that associations between the retail food environment and obesity are predominantly null.\(^10\)

Inconsistencies in observational research may be due to a lack of adjustment for unmeasured confounding such as: unmeasured preferences for residing near certain food outlet types; placement of food outlets in areas with higher demand;\(^14\) reverse causality; or differential measurement error. Non-causal methods (i.e. any model that ignores time-invariant and time-varying residual confounding), such as ordinary least squares (OLS) and random effects (RE) regression, implicitly assume that omitted variables (e.g. residential preferences) are independent of explanatory variables, and thus may produce biased estimates in the presence of residual confounding.\(^15\) Fixed effects (FE) regression controls for observed and unobserved time-invariant characteristics\(^15\) but ignores unobserved time-varying characteristics. In contrast, instrumental-variables (IV) regression is a causal approach that corrects for time-varying and time-invariant residual confounding by using proxies for exposures and eliminating the correlation between exposures and unmeasured characteristics.\(^16,17\) A few cross-sectional studies on fast-food restaurant availability and BMI have used IV regression finding estimates of greater magnitude relative to OLS regression\(^18-20\) but these studies did not address possible substitution effects (e.g. higher relative availability of full-service versus fast-food restaurants).

To address these gaps, we used 25 years of data from the Coronary Artery Risk Development in Young Adults (CARDIA) study and IV regression to quantify associations between different types of neighbourhood food outlets and BMI over time, while accounting for correlation between measured exposures and unmeasured characteristics. We compared the magnitude and direction of estimates from a causal approach (IV regression) with estimates derived from non-causal models (OLS, RE and FE regression) to assess the extent of bias. Based on previous work,\(^18-20\) we...
hypothesized that non-causal models would underestimate the impact of the retail food environment on obesity, possibly due to a lack of adjustment for unobserved bias.

Methods

Study sample

CARDIA is a prospective study of the development and risk factors of cardiometabolic disease in Black and White young adults. In 1985–86, 5115 CARDIA participants were recruited from four US metropolitan areas (Birmingham, AL; Chicago, IL; Minneapolis, MN; Oakland, CA); enrolment was balanced by age (18–24 years or 25–30 years), race (White or Black), gender and education (≤ high school or > high school). Follow-up examinations were conducted in 1987–88 (Year 2), 1990–91 (Year 5), 1992–93 (Year 7), 1995–96 (Year 10), 2000–01 (Year 15), 2005–06 (Year 20) and 2010–11 (Year 25), with retention of 91%, 86%, 81%, 79%, 74%, 72% and 72% of participants, respectively.

Individual-level data

Self-reported socio-demographics were collected at each examination, using a standardized questionnaire, including age, gender, race (Black, White), current educational attainment (years), marital status and number of children. Total family income (categorical responses) was collected starting with Year 5, so we used income values from Year 5 as a proxy for baseline values (no other Year 5 data were used).

Self-reported physical activity (PA) was assessed at each examination using the CARDIA PA History questionnaire, which captures frequency of participation in 13 categories of exercise in the previous 12 months. Alcohol consumption in the past year was assessed using a self-report questionnaire.21 Outcome variables. Height and weight were measured to the nearest 0.5 cm and 0.1 kg, respectively, by trained study staff and used to calculate BMI (kg/m²). Waist circumference (WC) was measured in duplicate at the minimum abdominal girth.

Neighbourhood-level data

Using Dun & Bradstreet (D&B) Duns Market Identifiers File (Dun & Bradstreet, Inc., Short Hills, NJ), a commercial dataset of US businesses with fair reliability and validity,23–25 we obtained the counts of PA facilities and food outlets at each examination year. We classified food outlets according to 8-digit Standard Industrial Classification (SIC) codes in Years 7, 10, 15, 20 and 25 (Appendix 1, available as Supplementary data at IJE online). Only 4-digit codes were available in 1986, so we used matched business names and a prediction model to supplement classification at baseline (Appendix 2, available as Supplementary data at IJE online).

We also used data from several commercial sources to calculate measures related to neighbourhood socio-demographics, employment density, street connectivity and consumer prices (Appendix 2). Using a geographic information system (GIS), we matched neighbourhood-level measures to CARDIA respondents’ residential addresses at baseline and Years 7, 10, 15, 20 and 25.

Analytical sample

Participants who resided in one of the four baseline cities in each examination year were eligible for the current study (n = 4316, 2462, 1728, 1481, 1202 and 1119 at baseline and Years 7, 10, 15, 20 and 25, respectively). We excluded one participant who withdrew from the study and two participants who changed gender. We also excluded women who were pregnant at the time of examination (n = 6, 33, 9, 4, 3 and 1 at baseline and Years 7, 10, 15, 20 and 25, respectively) and those with missing BMI data (n = 13, 23, 15, 5, 10 and 3 at baseline and Years 7, 10, 15, 20 and 25, respectively). Our final sample sizes were 4294, 2404, 1728, 1481, 1202 and 1119 individuals at baseline and Years 7, 10, 15, 20 and 25, respectively (n = 12 174 person-observations).

Using multilevel mixed effects linear regression (-mixed- in Stata 14.0) with baseline study centre, gender, race, age and year, we imputed missing values for individual-level income (n = 755, 55, 25, 26, 34, and 31 at baseline and Years 7, 10, 15, 20 and 25, respectively), marital status (n = 6 at baseline), alcohol intake (n = 2, 12, 18, 4, 21 and 11 at baseline and Years 7, 10, 15, 20 and 25, respectively) and PA (n = 1, 47, 23, 6, 12 and 312 at baseline and Years 7, 10, 15, 20 and 25, respectively). Using the mean of non-missing values across all years, we also imputed missing values for census-derived socio-demographics (n = 4), food outlets (n = 4) and road connectivity (n = 5) at baseline and Years 7, 10 and 15.

To account for potential selection bias due to out-migration over time, we used gender, race and baseline study centre to predict the probability of being in the sample at the end of follow-up. We used the inverse of the probability to weight all models (-pweight-).

Statistical analysis

Exposure specification

To create our explanatory variables (Y vector in equations below), we used the count of each food outlet type within
A 1-km street network distance from respondents’ residences, which captures walking distance to food outlets. We calculated the percentage of convenience stores, grocery stores and supermarkets out of total food stores (sum of convenience stores, grocery stores and supermarkets). We also calculated the percentage of fast-food restaurants and full-service restaurants out of total restaurants (sum of fast-food and full-service restaurants). Thus, modelling a 10% increase in one type of food store (or restaurant) equals a 10% decrease in the other food stores (or restaurants). We also modelled the total count of food outlets as endogenous variables, to account for variation in the denominator of our central exposure variables (i.e., having fewer or more alternatives might influence choice of food outlet). Endogenous variables (including exposures) are related to and determined by other variables in the model.

Covariates
We adjusted for several exogenous variables (X vectors in equations below), including age and age-squared (continuous), race (White, Black), gender, educational attainment (< high school, ≥ high school), income (≤ $42,500, > $42,500), baseline study centre, year and market-level cigarette and fast-food prices (Appendix 2). Exogenous variables are theoretically and statistically associated with endogenous variables, and not determined by other variables in the model.

Based on previously established methods, we calculated total PA intensity scores (exercise units) using a summary of the frequency and intensity of participants’ moderate and vigorous activities. We treated total PA, alcohol intake (yes/no), marital status (yes/no) and number of children as endogenous (W vectors in equations below).

Instrumental variables
Valid instruments (Z vectors in equations below) should be theoretically and statistically associated with endogenous variables, and have no direct associations with the outcome (outside their influence on endogenous variables) nor with error terms in regression equations. Our set of instruments included: population density; percentage neighbourhood White population; percentage neighbourhood population ≤ 18 years; distance to nearest employment subcentre; count of public and fee-based PA facilities; market-level wine and beer prices; and street connectivity (Appendix 2). We theorized that this set of variables was directly associated with neighbourhood food outlets and other endogenous variables, but not directly associated with BMI or error terms in the model.

Empirical model
The general specification for the IV model (Supplementary Figure 1, available as Supplementary data at IJE online) is shown below:

\[ W_{it} = \alpha_{1} Z_{it} + \beta_{1} X_{it} + \mu_{i} + \epsilon_{1it} \]  
\[ Y_{it} = \alpha_{2} Z_{it} + \beta_{2} X_{it} + \gamma_{1} W_{it} + \mu_{2i} + \epsilon_{2it} \]  
\[ B_{it} = \delta_{1} Y_{it} + \beta_{3} X_{it} + \gamma_{2} W_{it} + \mu_{3i} + \epsilon_{3it} \]

In equation 1, \( W_{it} \) represents a vector of endogenous variables, which influence BMI and retail food environment variables, and are also influenced by exogenous variables; \( Z_{it} \) represents a vector of exogenous instrumental variables; and \( X_{it} \) represents a vector of non-instrument exogenous variables. In equation 2, \( Y_{it} \) represents a vector of endogenous retail food environment variables. In equation 3, \( B_{it} \) is BMI at each examination. Across equations, \( i \) equals 1, . . . , \( N \) participants; \( t \) equals 1, . . . , \( T \), years; and \( \mu_{i} \) and \( \epsilon_{it} \) represent unobserved time-invariant and time-varying error components, respectively. The equations capture both the direct and the indirect effects of vectors on endogenous variables (e.g., \( x_{2} \) represents the direct effect of \( Z_{it} \) on \( Y_{it} \), and \( x_{1} \) represents the indirect effect of \( Z_{it} \) on \( Y_{it} \) via \( W_{it} \)).

Estimators and empirical tests of IV assumptions
We used a generalized method of moments (GMM) estimator for IV regression, which is a single-equation estimation approach based on a two-stage least-squares estimator. The GMM estimator allows for a cluster-corrected weighting matrix, which is more efficient than other IV estimators. We used -ivregress- with the ‘gmm’ option in Stata (version 14.0).

We used the Sargan-Hansen J test of over-identifying restrictions to test the assumption that our IVs were exogenous (i.e. not related to or determined by other variables also in the model). Failure to reject the null hypothesis (\( P < 0.05 \)) indicates that our IVs were exogenous and that it was valid to exclude them as predictors of BMI. We used the Durbin-Wu-Hausman test to evaluate whether our theoretically endogenous variables were in fact endogenous (i.e. related to and determined by other variables in the model). Rejecting the null hypothesis (\( P < 0.10 \)) implies that our assumption about endogeneity was correct. We obtained goodness-of-fit statistics to evaluate the explanatory power of our IVs. An F statistic with a critical value greater than 10 indicates that our IVs were strong predictors of endogenous variables. We used the -estat- post-estimation command for all empirical tests.

We then compared IV estimates with non-causal estimators, including: OLS regression (with robust variance) and RE regression, which do not account for endogeneity (i.e.
unmeasured confounding, reverse causality and differential measurement error); and FE regression, which controls for time-invariant endogeneity only (Supplementary Table 1, available as Supplementary data at IJE online).

We adjusted for all covariates in each model. We did not include food purchasing and consumption measures because these constructs are on the causal pathway and adjustment would theoretically attenuate estimated effects.

We considered comparing IV estimates with Heckman selection models, but we were unable to identify an exclusion restriction (i.e. a variable that predicts the probability of being obese, but not linear BMI); as well as propensity score-matching methods, but this approach does not account for unobserved bias.

Sensitivity analyses
To determine whether estimates from the central analysis were robust to our measure of obesity, we replicated all analyses with WC as the outcome. We considered using lagged IVs and endogenous variables, but decided that loss of explanatory power and uneven intervals between examinations justified using contemporaneous exposure and outcome variables.

Results
Mean BMI was 24.5 kg/m$^2$ (SD = 5.1) and 31.0 kg/m$^2$ (SD = 8.0) at baseline and Year 25, respectively, with a mean of 27.3 kg/m$^2$ (SD = 6.9) across follow-up (Table 1). Over time, the percentage of neighbourhood full-service restaurants, convenience stores and supermarkets increased, the percentage of fast-food restaurants and grocery stores decreased and total food outlet counts increased.

We failed to reject the null hypothesis of the test of over-identifying restrictions ($P = 0.667$), and rejected the null hypothesis of the Durbin-Wu-Hausman test ($P = 0.001$). Taken together, these results suggest our model was appropriately specified. The $F$ statistic value for each endogenous variable was greater than 10 (Table 2), suggesting that our combined IVs strongly identified endogenous variables.

Estimates of retail food environment exposures in relation to BMI were approximately 10–20 times smaller in magnitude using non-causal (versus causal) models (Table 3). For example, a 10% increase in the percentage of grocery stores (relative to supermarkets) was associated with a 0.50 kg/m$^2$ (95% CI: 0.10, 1.00; $P = 0.026$) increase in BMI over time using IV regression (assuming a linear
relationship). On the other hand, a 10% increase in the percentage of grocery stores was associated with a negligible decrease in BMI using RE regression ($\beta = -0.03$; 95% CI: -0.10, -0.001; $P = 0.037$) and FE regression ($\beta = -0.03$; 95% CI: -0.10, -0.002; $P = 0.031$).

The percentages of convenience stores (relative to supermarkets) and fast-food restaurants (relative to full-service restaurants) were not associated with BMI in any model, but the magnitude of coefficients was also larger using IV regression.

### Sensitivity analyses

The magnitude and direction of estimates derived from models with WC were similar to those obtained in BMI analyses (Supplementary Table 2, available as Supplementary data at IJE online). Goodness-of-fit statistics (Supplementary Table 3, available as Supplementary data at IJE online) and empirical tests of overidentifying restrictions ($P = 0.646$) and endogeneity ($P = 0.001$) were also similar to BMI analyses.

### Discussion

With clinic-based, anthropometric measures and detailed neighbourhood environment data, we used IV regression to estimate causal effects of the retail food environment on BMI over time. We also compared the magnitude and direction of causal IV estimates with non-causal models, including OLS and RE regression, which do not account for residual confounding, and with FE regression, which only corrects for unmeasured time-invariant characteristics. Controlling for unmeasured characteristics with causal models in neighbourhood environment studies is important because omitted variables (e.g. unmeasured preferences) may bias relationships between environmental variables and health outcomes.

Although selection bias usually biases OLS estimates upwards, we found that longitudinal associations between food outlets and BMI were attenuated using non-causal (versus causal) models. The smaller magnitude of non-causal model estimates also suggests that the error terms corresponding to retail food environment exposures and BMI were negatively correlated, possibly due to a mismatch between unmeasured preferences and environment (e.g. individuals with a preference for locating near supermarkets might locate in areas with few supermarkets for reasons unrelated to the retail food environment). Furthermore, the observed differences between FE and IV regression suggest that bias may be time-varying, such as unmeasured preferences for larger residences over time.

Overall, our findings are consistent with previous studies showing that using IV regression resulted in stronger associations between environment variables and health outcomes than did OLS regression. Given the rich empirical literature comparing causal and non-causal methodologies across several other disciplines, we argue that our findings would consistently apply to future studies and are not a unique feature of the CARDIA study.

Our causal model results suggest that the percentage of grocery stores (relative to supermarkets) was positively albeit weakly-associated with BMI over time. Others suggest that grocery stores (which are larger and have higher sales) have a lower ratio of healthy to unhealthy shelf space than do supermarkets. Therefore, it is hypothesically possible that decreasing the number of smaller grocery stores while simultaneously increasing the number of supermarkets, possibly via changes to zoning ordinances, may contribute to reducing population-level BMI (though we acknowledge that such efforts are not trivial). On the other hand, natural intervention studies suggest that modifying the retail food environment may not meaningfully reduce obesity whereas price interventions to improve healthy eating have been more successful. Although we posit that changes to BMI would operate through changes in food consumption, in an earlier study we did not find an association between the availability of grocery stores and diet outcomes (unpublished); however, it is possible that a shorter follow-up period and a smaller sample size undermined our ability to detect statistically significant associations.

Although our instruments strongly identified obesity outcomes, we acknowledge that there are many challenges with causal models, including availability of longitudinal data, lack of temporal variation in retail food environment exposures and difficulties in identifying valid and robust IVs. The latter can be partially addressed with full-information IV regression, which is preferable in the

### Table 2. Goodness-of-fit statistics for evaluating strength of identification of endogenous variables with body mass index: CARDIA baseline and Years 7–25 (1985/86–2010/11)

<table>
<thead>
<tr>
<th></th>
<th>$F$ statistic</th>
<th>$P$-value $^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Convenience stores, % per total food stores</td>
<td>106.0</td>
<td>$&lt; 0.001$</td>
</tr>
<tr>
<td>Grocery stores, % per total food stores</td>
<td>87.9</td>
<td>$&lt; 0.001$</td>
</tr>
<tr>
<td>Fast-food restaurants, % per total restaurants</td>
<td>78.9</td>
<td>$&lt; 0.001$</td>
</tr>
<tr>
<td>Total food stores, count</td>
<td>272.3</td>
<td>$&lt; 0.001$</td>
</tr>
<tr>
<td>Total restaurants, count</td>
<td>183.9</td>
<td>$&lt; 0.001$</td>
</tr>
<tr>
<td>Marital status (yes, no)</td>
<td>12.6</td>
<td>$&lt; 0.001$</td>
</tr>
<tr>
<td>Number of children</td>
<td>17.3</td>
<td>$&lt; 0.001$</td>
</tr>
<tr>
<td>Alcohol intake (yes, no)</td>
<td>29.1</td>
<td>$&lt; 0.001$</td>
</tr>
<tr>
<td>Physical activity (exercise units)</td>
<td>19.7</td>
<td>$&lt; 0.001$</td>
</tr>
</tbody>
</table>

$^a$Rejecting the $F$ test indicates that our set of instruments provides good identification for that endogenous variable.
presence of weak instruments.\textsuperscript{41} We also acknowledge that the retail food environment is only one risk factor for weight gain, and additional risk factors should be considered in future research, including factors related to food availability and prices in school, professional and recreational environments (i.e. not retail food outlets). We also lacked data related to zoning ordinances and land use policies, which may restrict the placement of healthy food outlets (e.g. supermarkets) in neighbourhoods,\textsuperscript{42} especially in low-income areas,\textsuperscript{43} though we controlled for neighbour-

### Table 3. Beta coefficients (95% confidence intervals)\textsuperscript{a} for the associations between each type of food store or restaurant\textsuperscript{b} and body mass index,\textsuperscript{c} using ordinary least squares, fixed effects, and instrumental variables regression: CARDIA baseline and Years 7–25 (1985/86–2010/11)

<table>
<thead>
<tr>
<th></th>
<th>IV regression\textsuperscript{d}</th>
<th>P-value</th>
<th>OLS regression\textsuperscript{e}</th>
<th>P-value</th>
<th>RE regression\textsuperscript{f}</th>
<th>P-value</th>
<th>FE regression\textsuperscript{g}</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>N (person-observations)</td>
<td>12 174</td>
<td></td>
<td></td>
<td></td>
<td>12 174</td>
<td></td>
<td>12 174</td>
<td></td>
</tr>
<tr>
<td>Full-service restaurants, % per total restaurants\textsuperscript{b}</td>
<td>0.00</td>
<td></td>
<td>0.00</td>
<td></td>
<td>0.00</td>
<td></td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Fast-food restaurants, % per total restaurants\textsuperscript{b}</td>
<td>-0.01</td>
<td>0.700</td>
<td>-0.001</td>
<td>0.710</td>
<td>-0.001</td>
<td>0.333</td>
<td>-0.001</td>
<td>0.391</td>
</tr>
<tr>
<td>Supermarkets, % per total food stores\textsuperscript{b}</td>
<td>0.00</td>
<td></td>
<td>0.00</td>
<td></td>
<td>0.00</td>
<td></td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Convenience stores, % per total food stores\textsuperscript{b}</td>
<td>0.02</td>
<td>0.457</td>
<td>-0.002</td>
<td>0.854</td>
<td>-0.003</td>
<td>0.117</td>
<td>-0.002</td>
<td>0.123</td>
</tr>
<tr>
<td>Grocery stores, % per total food stores\textsuperscript{b}</td>
<td>0.05</td>
<td>0.026</td>
<td>-0.001</td>
<td>0.752</td>
<td>-0.003</td>
<td>0.037</td>
<td>-0.003</td>
<td>0.301</td>
</tr>
<tr>
<td>Total restaurants, count \textsuperscript{b}</td>
<td>(0.01, 0.10)*</td>
<td></td>
<td></td>
<td></td>
<td>(0.01, 0.001)</td>
<td></td>
<td>(0.01, -0.0001)</td>
<td></td>
</tr>
<tr>
<td>Total food stores, count \textsuperscript{b}</td>
<td>(0.02, 0.06)</td>
<td></td>
<td></td>
<td></td>
<td>(0.02, -0.004)*</td>
<td></td>
<td>(0.02, -0.004)*</td>
<td></td>
</tr>
<tr>
<td>Total food stores, count \textsuperscript{b}</td>
<td>0.01</td>
<td>0.414</td>
<td>0.001</td>
<td>0.481</td>
<td>0.003</td>
<td>0.721</td>
<td>-0.10</td>
<td>0.280</td>
</tr>
<tr>
<td>Total food stores, count \textsuperscript{b}</td>
<td>(0.02, 0.05)</td>
<td></td>
<td></td>
<td></td>
<td>(0.01, 0.02)</td>
<td></td>
<td>(0.01, 0.02)</td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{a}Multivariable-adjusted models were adjusted for individual-level age, gender, race, educational attainment, income, children, marital status, examination year and market-level food prices.

\textsuperscript{b}Calculated within a 1-km network buffer of participants’ residences.

\textsuperscript{c}Body mass index, mean (SD): 27.3 (6.9) kg/m\textsuperscript{2}.

\textsuperscript{d}Instrumental-variables regression using Stata’s -ivregress- command with the ‘gmm’ option.

\textsuperscript{e}Ordinary least squares regression using Stata’s -reg- command with robust variance.

\textsuperscript{f}Repeated measures random effects regression using Stata’s -xtreg- command with the ‘re’ option.

\textsuperscript{g}Repeated measures fixed effects regression using Stata’s -xtreg- command with the ‘fe’ option.

\textsuperscript{h}Omitted from the model (referent).

\textsuperscript{i}Indicates the estimate is statistically significant at the \(P < 0.05\) level.

Supplementary Data

Supplementary data are available at IJE online.

Funding

This work was supported by the National Heart, Lung, and Blood Institute [grant numbers R01HL104580 and R01HL114091]. The Coronary Artery Risk Development in Young Adults study (CARDIA) is supported by contracts from the National Heart, Lung, and Blood Institute, the Intramural Research Program of the National Institute on Aging and an intra-agency agreement between National Institute on Aging and Human Development [grant numbers R01HL104580 and R01HL114091]. The authors are grateful to: the Carolina Population Center, University of North Carolina at Chapel Hill, for general support from the Eunice Kennedy Shriver National Institute of Child Health and Human Development [grant number P2C HD050924]; the Nutrition Obesity Research Center, University of North Carolina, for support from the National Institute for Diabetes and Digestive and Kidney Diseases [grant...
number P30DK56350); and to the Center for Environmental Health Sciences, University of North Carolina, for support from the National Institute for Environmental Health Sciences [grant number P30ES010126].

Acknowledgments
The authors would like to acknowledge CARDIA Chief Reviewer Jane Boone-Heinonen, PhD, whose thoughtful suggestions improved the paper, as well as Marc Peterson, of the University of North Carolina, Carolina Population Center (CPC) and the CPC Spatial Analysis Unit, for creation of the environmental variables. P.E.R. had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of data analysis.

Conflict of interest: None declared. The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institutes of Health; or the U.S. Department of Health and Human Services.

References
3. Cummins S, Flint E, Matthews SA. New neighbourhood grocery store increased awareness of food access but did not alter dietary habits or obesity. Health Aff (Millwood) 2014;33:283–91.


