HHS Public Access

Author manuscript

J Epidemiol Community Health. Author manuscript; available in PMC 2018 December 01.

Published in final edited form as:

J Epidemiol Community Health. 2017 December ; 71(12): 1185–1190. doi:10.1136/jech-2017-209158.

Understanding bias in relationships between the food environment and diet quality: The Coronary Artery Risk Development in Young Adults (CARDIA) Study

Penny Gordon-Larsen, PhD,

Department of Nutrition, University of North Carolina at Chapel Hill, Carolina Population Center, 137 East Franklin Street, 6th floor, Campus Box #8120, Chapel Hill, NC 27514, USA

Pasquale E. Rummo,

Department of Nutrition, University of North Carolina at Chapel Hill, Carolina Population Center, Chapel Hill, NC, USA, United States. Present address: New York University School of Medicine, Department of Population Health, Section on Health Choice, Policy and Evaluation, 227 E. 30th Street, #649, New York, NY, USA

David K. Guilkey,

Department of Economics, University of North Carolina at Chapel Hill, Carolina Population Center, Chapel Hill, NC, USA

Shu Wen Ng,

Department of Nutrition, University of North Carolina at Chapel Hill, Carolina Population Center, Chapel Hill, NC, USA

Katie A. Meyer,

Department of Nutrition, University of North Carolina at Chapel Hill, Carolina Population Center, Chapel Hill, NC, USA

Barry M. Popkin,

Department of Nutrition, University of North Carolina at Chapel Hill, Carolina Population Center, Chapel Hill, NC, USA

Jared P. Reis, and

Corresponding author: Tel.: 919-962-6110, pglarsen@email.unc.edu.

COMPETING INTEREST

Competing interest: None declared.

ETHICS APPROVAL

Ethics approval was obtained by the University of North Carolina at Chapel Hill Institutional Review Board (IRB: #11-1393).

CONTRIBUTIONS

P. E. Rummo had full access to the data and takes responsibility for the integrity of the data and the accuracy of the data analysis. P. E. Rummo performed the statistical analysis, and analyzed and interpreted the data. P. E. Rummo and P. Gordon-Larsen drafted the article. D. K. Guilkey, S. W. Ng, K. A. Meyer, B. M. Popkin, J. Shikany, J. Reis, and P. Gordon-Larsen critically revised the article for important intellectual content. P. Gordon-Larsen acquired the data, obtained the funding, approved the final draft of the article, and supervised the study. All authors contributed to the study concept and design.

Division of Cardiovascular Sciences, National Heart, Lung, and Blood Institute, Bethesda, Maryland, USA

James M. Shikany

Division of Preventive Medicine, University of Alabama at Birmingham, Birmingham, AL, USA

Pasquale E. Rummo: prummo@live.unc.edu; David K. Guilkey: david_guilkey@unc.edu; Shu Wen Ng: shuwen@unc.edu; Katie A. Meyer: ktmeyer@email.unc.edu; Barry M. Popkin: popkin@unc.edu; Jared P. Reis: reisjp@nhlbi.nih.gov; James M. Shikany: jshikany@uabmc.edu

Abstract

Background—The relationship between food environment exposures and diet behaviors is unclear, possibly because the majority of studies ignore potential residual confounding.

Methods—We used 20 years (1985–6; 1992–3; 2005–6) of data from the Coronary Artery Risk Development in Young Adults (CARDIA) study across four U.S. cities (Birmingham, AL; Chicago, IL; Minneapolis, MN; Oakland, CA) and instrumental-variables (IV) regression to obtain causal estimates of longitudinal associations between the percentage of neighborhood food outlets (per total food outlets within 1-km network distance of respondent residence) and an *a priori* diet quality score, with higher scores indicating higher diet quality. To assess the presence and magnitude of bias related to residual confounding, we compared results from causal models (IV regression) to non-causal models, including ordinary least squares regression, which does not account for residual confounding at all; and fixed-effects regression, which only controls for time-invariant unmeasured characteristics.

Results—The mean diet quality score across follow-up was 63.4 (SD=12.7). A 10% increase in fast food restaurants (relative to full-service restaurants) was associated with a lower diet quality score over time using IV regression (β=−1.01; 95% CI=−1.99, −0.04); estimates were attenuated using non-causal models. The percentage of neighborhood convenience and grocery stores (relative to supermarkets) was not associated with diet quality in any model, but estimates from non-causal models were similarly attenuated compared to causal models.

Conclusion—Ignoring residual confounding may generate biased estimated effects of neighborhood food outlets on diet outcomes, and may have contributed to weak findings in the food environment literature.

BACKGROUND

Observational research suggests that the food environment is associated with diet outcomes, [1, 2] and that these relationships differ across types of food outlets, perhaps due to differences in the nutrient and energy density of foods sold at each outlet.[3, 4] However, other studies report no associations between food environment exposures and dietary behaviors,[5–7] and past experiments placing supermarkets in food deserts or banning new fast food restaurants did not successfully change food purchases or obesity.[8, 9] An important limitation of the current literature is the fact that the majority of research ignores potential residual confounding.

For example, some food environment studies use ordinary least squares (OLS) regression, [10, 11] which may generate biased estimates if omitted variables (e.g., selective placement

of food outlets) influence food environment exposures and health outcomes[12] More recently, longitudinal studies in the food environment literature have used fixed effects (FE) regression[13–15] which quantifies within-person associations, effectively using each individual as his/her own control. FE regression controls for all measured and unmeasured time-invariant confounding, but not for unmeasured characteristics that change over time (e.g., residential preferences).[12]

A causal approach to correct for residual confounding bias is instrumental-variables (IV) regression,[16, 17] which uses surrogates for exposures to eliminate the correlation between exposures and unmeasured characteristics. In the presence of time-varying residual confounding, IV regression theoretically provides less biased estimates relative to non-causal models (e.g., OLS and FE regression). Although the literature suggests that associations between environmental exposures and health outcomes are sensitive to residual confounding,[18] to our knowledge no studies have examined the potential for unobserved bias in the context of the food environment and diet behaviors.

Based on these gaps in the literature, we used 20 years of dietary data from the Coronary Artery Risk Development in Young Adults (CARDIA) study and a causal approach (IV regression) to estimate longitudinal associations between the relative availability of neighborhood food outlets and an *a priori* diet quality score. We then compared results to non-causal models (OLS and FE regression) to assess the presence and extent of bias related to residual confounding. Based on previous work,[19–21] we hypothesized that estimates from non-causal models would be attenuated relative to causal models.

METHODS

Study Sample

CARDIA is a prospective study of the development of cardiometabolic disease among white and black young adults. In 1985–86, 5,115 CARDIA participants were recruited from four US metropolitan areas (Birmingham, AL; Chicago, IL; Minneapolis, MN; Oakland, CA), with balanced enrollment by age (18–24 y or 25–30 y), race (white or black), gender, and education (≤HS or >HS). Follow-up examinations were conducted in 1987–1988 (Year 2), 1990–1991 (Year 5), 1992–1993 (Year 7), 1995–1996 (Year 10), 2000–2001 (Year 15), 2005–2006 (Year 20), and 2010–2011 (Year 25), with retention of 91%, 86%, 81%, 79%, 74%, 72%, and 72% of participants, respectively.

Individual-level data

Dietary intake was assessed at Years 0, 7, and 20 using the CARDIA Diet History, an interviewer-administered, open-ended validated questionnaire,[22] where participants self-reported type and frequency of dietary consumption in the past month. Responses were assigned to 46 food groups based on previously established methods (Supplemental file 1).[23, 24]

A standardized questionnaire was used to collect self-reported individual-level sociodemographics at each exam, including age, gender, race (black, white), current educational attainment (years), marital status, and number of children. Total family income

(categorical responses) was collected starting in Year 5, so we used income values from Year 5 as a proxy for baseline values.

Neighborhood-level data

We extracted data related to counts of physical activity (PA) facilities and food outlets at each exam year from the Dun & Bradstreet (D&B) Duns Market Identifiers File (Dun & Bradstreet, Inc., Short Hills, NJ),[25] a commercial data set of US businesses with fair reliability and validity.[26, 27] Food outlets were classified according to 8-digit Standard Industrial Classification codes in Years 7, 10, 15, 20, and 25. Only 4-digit codes were available at baseline, necessitating a second step that involved matching business names (Supplemental file 2).

We used data from several commercial sources to calculate measures related to neighborhood sociodemographics, employment, street connectivity, and consumer and housing prices (Supplemental file 2). Using a Geographic Information System, we matched neighborhood-level measures to CARDIA respondents' residential addresses at baseline and Years 7, 10, 15, 20, and 25.

Analytic Sample

Participants who resided in one of the baseline cities at each exam year were eligible for the current study (n=4,316, 2,462, and 1,202 at Years 0, 7, and 20, respectively). We excluded one participant who withdrew from the study and two participants who changed gender. We also excluded person-observations with missing diet data (n=3, 83, and 149 at Years 0, 7, and 20, respectively). Using multilevel mixed-effects linear regression (-mixed- in Stata 14.0) with baseline study center, gender, race, age, and exam year, we imputed values for missing individual-level income data [n=758 (17.6%), 56 (2.3%), and 34 (2.8%) at Years 0, 7, and 20, respectively] and marital status (n=6 (0.1%) at baseline). We also imputed values for missing data related to counts of PA and food outlets, road connectivity, and neighborhood sociodemographics (n=2, 2, and 4 at Year 7) using the mean of non-missing values across exam years. Our final sample size was 4,310, 2,377, and 1,053 individuals at exam years 0, 7, and 20, respectively (n=7740 person-observations).

Statistical Analysis

Outcome specification—We re-created an *a priori* diet quality score using a previously established approach.[23, 24] Briefly, based on their hypothesized relationships with health, we classified food groups (n=46) as beneficial (n=20), adverse (n=13), or neutral (n=13). At each exam, consumption (servings/week) of non-neutral foods was categorized into quintiles ranging from lowest to highest consumption and given a score of 0 to 4 for beneficial food groups and 4 to 0 for adverse food groups; for example, intake in the highest quintile of positively-rated foods received a score of 4 and vice-versa for negatively-rated foods. For food groups with a large proportion of non-consumption, we categorized non-consumers as 0 and created quintiles for consumers only. The *a priori* diet quality score has a theoretical maximum of 132; a higher score indicates higher diet quality. The actual mean diet quality score in our analytic sample was 63.3 (SD=13, range=24–107) at baseline.

Exposure specification—We created our explanatory variables (*Y* vector in equations below) by calculating food outlet counts within a 1-km street network distance from respondents' residences, with the intent of capturing resources accessible by walking.[28] We then calculated the relative (versus absolute) availability of each food outlet type by calculating the percentage of convenience stores, grocery stores, and supermarkets per total food stores (sum of total convenience stores, grocery stores, and supermarkets), and the percentage of fast food restaurants and full-service restaurants per total restaurants (sum of total fast food and full-service restaurants). For example, a 10% increase in fast food restaurants equals a 10% decrease in full-service restaurants.

Given that associations between the percentage of food outlets and diet may differ with variation in the denominator (i.e., having fewer or more alternatives might influence choice of food outlet), we modeled the total number of food outlets as endogenous. Endogenous variables (including exposures) are related to, and determined by, other variables in the model.[29]

Covariates—We treated marital status and number of children as endogenous variables (*W* vectors in equations below). We also adjusted for non-instrument exogenous variables (*X* vectors in equations below), including age and age-squared (continuous), race (white, black), gender, educational attainment (<HS, ≥HS), income (≤$42,500, >$42,500), baseline study center, year, and market-level food prices. Exogenous variables are theoretically and statistically associated with endogenous variables, and not determined by other variables in the model.

Instrumental variables regression

Instrumental variables: Potential instruments (*Z* vectors in equations below) should be theoretically and statistically associated with endogenous variables; not correlated with unmeasured characteristics (i.e., error terms); and have no direct theoretical or statistical association with the outcome.[16, 17] We considered several potential instruments, including neighborhood sociodemographics; employment density; count of PA facilities; and street connectivity. Prior studies have shown that these factors influence variation in neighborhood food outlets,[4, 25] providing empirical support for our set of instruments. We also assumed that our instruments did not have a direct association with diet quality (outside of associations with food environment variables); however, sociodemographic variables predicted diet outcomes in multivariable-adjusted OLS models, and thus were excluded in our final IV model.

Empirical model: The following is the general specification for our model (Supplemental file 3):

$$W_{it}=\alpha_1 Z_{it}+\beta_1 X_{it}+\mu_{1i}+\varepsilon_{1it} \quad 1)$$

$$Y_{it}=\alpha_2 Z_{it}+\beta_2 X_{it}+\gamma_1 W_{it}+\mu_{2i}+\varepsilon_{2it} \quad 2)$$

$$D_{it}=\delta_1 Y_{it}+\beta_3 X_{it}+\gamma_2 W_{it}+\mu_{3i}+\varepsilon_{3it} \quad 3)$$

In equation 1, W_{it} represented endogenous variables, which influenced both diet and food environment variables, and were also influenced by exogenous variables in the model; Z_{it} represented exogenous IVs, which influenced food environment variables but not diet quality; and X_{it} represented other exogenous variables (e.g., food prices), which influenced both diet and food environment variables. In equation 2, Y_{it} represented endogenous food environment variables. In equation 3, D_{it} represented diet quality. Across equations, i equaled 1, …, N participants; t equaled 1, …, T_i years; and μ_i and ε_{it} represented unobserved time-invariant and time-varying error components, respectively. The equations capture both the direct and indirect effects of vectors on endogenous variables (e.g., α_2 represents the direct effect of Z_{it} on Y_{it}, and α_1 represents the indirect effect of Z_{it} on Y_{it} via W_{it}).

We used the generalized method of moments (GMM) estimator, which is based on choosing a parameter value that minimizes a quadratic function of the moment conditions.[30] Unlike two-stage least squares, GMM estimation allows for a cluster-corrected weighting matrix, which is more efficient than other IV estimators.

Empirical tests of IV assumptions and specification: We used the Durbin-Wu-Hausman test to evaluate the assumption that our endogenous food environment variables were related to and influenced by other variables in the model. Failure to reject the null hypothesis implies that our food environment exposures were not exogenous, and our assumption was invalid. To test the assumption that our instruments were uncorrelated with the error terms, we used the Sargan-Hansen J test of overidentifying restrictions. Failure to reject the null hypothesis ($p>0.05$) supports our assumption that our IVs were not correlated with error terms in the model and that it was valid to exclude instruments as direct predictors of the outcome. To test the strength of the set of instruments combined to predict food environment exposures, we evaluated the magnitude of F-statistics from reduced form regressions; a critical value <10 indicates that instruments were weak.[31]

We used -ivregress- with the “gmm” option and the -estat- post-estimation command to report goodness-of-fit statistics and perform tests of endogeneity and overidentifying restrictions for IV regression in Stata (version 14.0). We also used a probit model to derive inverse probability weights to account for potential selection bias due to out-migration from the four cities between baseline and Year 20. We used gender, race, and baseline study center to predict the probability of being in the sample at Year 20, and used the inverse of the probability to weight all models (-pweight-).

To evaluate whether our results differed from non-causal methods, we compared the magnitude and direction of estimates from IV regression to OLS regression (with robust variance), which does not account for residual confounding at all; and FE regression, which accounts for time-invariant unmeasured characteristics only.[12] We used -reg- with the “vce” option for OLS regression and -xtreg- with the “fe” option for FE regression, and adjusted for all covariates.

Sensitivity analyses—To determine whether estimates from the diet quality analysis were robust to our *a priori* classification, we also compared IV regression results to empirically-derived diet pattern scores. Using principle components analysis with orthogonal rotation, we derived two uncorrelated dietary patterns with the 46 food groups: one with high factor loadings of healthier foods, labeled "Prudent," and another with high factor loadings of unhealthier foods, labeled "Western". We also used IV regression to examine associations between the food environment and servings per week of fruits, vegetables, whole grains, red/processed meat, and sugar-sweetened beverages (SSBs) (Supplemental file 1). To determine whether estimates were robust to missingness, we compared IV models with imputed values to those without imputed values.

RESULTS

Participants' *a priori* diet quality score increased from 61.5 (SD=12.6) at baseline to 67.3 (SD=12.5) in Year 20, with a mean of 63.4 (SD=12.7) over time (Table 1). Intake of fruits and vegetables increased between baseline and year 20, while intake of red/processed meat and SSBs decreased; whole grain consumption remained relatively stable. The percentage of fast food restaurants and grocery stores decreased over time, while the percentage of full-service restaurants, convenience stores, and supermarkets increased.

In the central IV regression analysis, we rejected the null hypothesis of the Durbin-Wu-Hausman test ($P<0.001$), and we failed to reject the null hypothesis of the test of overidentifying restrictions ($P=0.59$), which implies our model was properly specified. *F*-statistic values for were greater than 10,[31] suggesting that our combined instruments strongly predicted our explanatory variables (Table 2). Results were robust to the inclusion and exclusion of individual instruments, where identification was possible.

Although the magnitude of point estimates of food environment variables was small across all models, it was relatively larger using causal (IV regression) versus non-causal (OLS and FE regression) models (Table 3). Assuming a linear relationship, a 10% increase in the percentage of fast food restaurants (relative to full-service restaurants) was associated with a 1.01 (95% CI: −1.99, −0.04) decrease in the *a priori* diet quality score over time using IV regression; whereas, the percentage of fast food restaurants and diet quality were not related using non-causal models. The percentage of convenience or grocery stores (relative to supermarkets) was not associated with diet quality using any model.

Sensitivity analyses

Point estimates derived from IV regression models with empirically-derived diet pattern scores were similar in direction to those obtained in the central analysis (Supplemental file 4). However, the magnitude of point estimates was more similar to non-causal models. The food environment was not associated with weekly servings of any single food group. Estimates from IV models without imputed values were nearly identical to those with imputed values.

DISCUSSION

Although some researchers propose that improvements to the neighborhood food environment can promote healthy lifestyles, there are substantial challenges in studying how neighborhoods influence health. In addition to a largely cross-sectional literature with a narrow focus on grocery stores,[2, 32] previous studies using non-causal models do not explicitly account for residual confounding, and other methods (e.g., propensity score matching) do not control for unobserved bias.[33] We sought to resolve these methodological limitations by estimating longitudinal associations between neighborhood food outlets and diet quality using a causal model (IV regression), which corrects for time-varying and time-invariant residual confounding (assuming instruments are valid). We also compared the magnitude and direction of causal estimates to those derived from non-causal models, including OLS regression, which does not account for residual confounding at all, and FE regression, which only accounts for time-invariant unmeasured characteristics.

Compared to IV regression, we found that associations between the neighborhood food environment and diet quality were attenuated using non-causal models. Our results were consistent with previous studies comparing IV and OLS estimates of associations between environmental exposures and BMI across several disciplines.[19–21, 34, 35] Given the differences in magnitude between IV and FE regression, it also appears that time-varying unmeasured characteristics were an important source of bias. IV regression producing larger estimates may be due to a mismatch between demand and availability of certain types of food outlets (e.g., high demand for fast food among individuals locating in neighborhoods with few fast food restaurants). Although a few studies report no associations between neighborhood fast food restaurants and diet outcomes,[7, 36] our findings suggest that these relationships may be understated in the literature due to residual confounding.

We observed an inverse association between the percentage of fast food restaurants and diet quality using IV regression. However, we did not observe similar associations between the percentage of convenience or grocery stores and diet quality, nor with weekly servings of single food groups. These findings suggest that the distribution of fast food restaurants affects overall diet quality, possibly via changes across multiple food groups. For example, the relationship between fast food restaurants and single food groups may be too small to detect alone, while combined changes may be large enough to matter (e.g., relatively high consumption of adverse foods and low consumption of beneficial foods).

Our study shows that it may be possible to correct for residual confounding using a single-equation estimator with valid instruments.[16, 37] Our sensitivity analyses with empirically-derived dietary pattern scores and single food groups were also a strength, and the latter allowed us to consider which food groups may (or may not) have driven negative or null associations with diet quality. Although we were able to use street-network buffers and relative (vs. absolute) measures of food outlets to quantify the residential food environment, we were not able to capture what foods were sold in each food outlet type or food-purchasing behavior. We also observed small classification errors in D&B, but we explicitly corrected for conspicuous errors, which were probably random in nature. Our measures of dietary intake were also based on self-report, which is prone to recall bias and error.

In sum, we found that the relative availability of fast food restaurants was negatively associated with diet quality over time using a causal approach (IV regression), and that estimates from non-causal models (OLS and FE regression) were attenuated relative to causal models, potentially due to residual confounding. Our results emphasize the importance of assessing potential bias in studies of the food environment and health, and future observational studies may benefit from using causal methods that explicitly account for bias. Estimates of neighborhood associations with health in the literature are often weak, so ignoring potential bias could impact the accuracy and interpretation of findings. Furthermore, interventions and policies related to the food environment have focused on absolute changes to food outlets and have been largely unsuccessful,[8, 9, 38–40] but our results highlight the need to consider relative changes to food outlets.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

FUNDING

This work was supported by the National Heart, Lung, and Blood Institute (grants R01HL104580 and R01HL114091). The Coronary Artery Risk Development in Young Adults Study 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 NIA and NHLBI (grants HHSN268201300025C, HHSN268201300026C, HHSN268201300027C, HHSN268201300028C, HHSN268201300029C, HHSN268200900041C, and AG0005). 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 R24 HD050924), to the Nutrition Obesity Research Center, University of North Carolina for support from the National Institute for Diabetes and Digestive and Kidney Diseases (grant P30DK56350), and to the Center for Environmental Health Sciences, University of North Carolina for support from the National Institute for Environmental Health Sciences (grant P30ES010126).

The authors would like to acknowledge Marc Peterson, of the University of North Carolina, Carolina Population Center (CPC) and the CPC Spatial Analysis Unit for creation of the environmental variables.

References

1. Black C, Moon G, Baird J. Dietary inequalities: What is the evidence for the effect of the neighbourhood food environment? Health & place. 2014; 27:229–42. [PubMed: 24200470]
2. Gustafson A, Hankins S, Jilcott S. Measures of the consumer food store environment: a systematic review of the evidence 2000–2011. Journal of community health. 2012; 37(4):897–911. [PubMed: 22160660]
3. Stern D, Ng SW, Popkin BM. The nutrient content of US household food purchases by store type. American journal of preventive medicine. 2016; 50(2):180–90. [PubMed: 26437868]
4. Walker RE, Keane CR, Burke JG. Disparities and access to healthy food in the United States: a review of food deserts literature. Health & place. 2010; 16(5):876–84. [PubMed: 20462784]
5. Duffey KJ, Gordon-Larsen P, Steffen LM, Jacobs DR, Popkin BM. Regular consumption from fast food establishments relative to other restaurants is differentially associated with metabolic outcomes in young adults. The Journal of nutrition. 2009; 139(11):2113–8. [PubMed: 19776183]
6. Pearce J, Hiscock R, Blakely T, Witten K. The contextual effects of neighbourhood access to supermarkets and convenience stores on individual fruit and vegetable consumption. Journal of Epidemiology and Community Health. 2008; 62(3):198–201. [PubMed: 18272733]

7. Richardson AS, Boone-Heinonen J, Popkin BM, Gordon-Larsen P. Neighborhood fast food restaurants and fast food consumption: a national study. BMC Public Health. 2011; 11:543.doi: 10.1186/1471-2458-11-543 [PubMed: 21740571]
8. Cummins S, Flint E, Matthews SA. New neighborhood grocery store increased awareness of food access but did not alter dietary habits or obesity. Health Affairs. 2014; 33(2):283–91. [PubMed: 24493772]
9. Sturm R, Cohen DA. Zoning for health? The year-old ban on new fast-food restaurants in South LA. Health Affairs. 2009; 28(6):w1088–w97. [PubMed: 19808703]
10. Hollands S, Campbell MK, Gilliland J, Sarma S. Association between neighbourhood fast-food and full-service restaurant density and body mass index: A cross-sectional study of Canadian adults. Can J Public Health. 2014; 105(3):e172–e8. [PubMed: 25165835]
11. Powell LM, Nguyen BT, Han E. Energy intake from restaurants: demographics and socioeconomics, 2003–2008. Am J Prev Med. 2012; 43(5):498–504. DOI: 10.1016/j.amepre.2012.07.041 [PubMed: 23079172]
12. Allison, PD. Fixed Effects Regression Methods for Longitudinal Data Using SAS. SAS Institute; 2005.
13. Boone-Heinonen J, Gordon-Larsen P, Kiefe CI, Shikany JM, Lewis CE, Popkin BM. Fast food restaurants and food stores: longitudinal associations with diet in young to middle-aged adults: the CARDIA study. Arch Intern Med. 2011; 171(13):1162–70. DOI: 10.1001/archinternmed.2011.283 [PubMed: 21747011]
14. Nguyen BT, Powell LM. The impact of restaurant consumption among US adults: effects on energy and nutrient intakes. Public Health Nutrition. 2014; (Supplement -1):1–8. FirstView. DOI: 10.1017/S1368980014001153 [PubMed: 24650538]
15. Powell LM, Nguyen BT. Fast-food and full-service restaurant consumption among children and adolescents: effect on energy, beverage, and nutrient intake. JAMA Pediatr. 2013; 167(1):14–20. DOI: 10.1001/jamapediatrics.2013.417 [PubMed: 23128151]
16. Greenland S. An introduction to instrumental variables for epidemiologists. International journal of epidemiology. 2000; 29(4):722–9. [PubMed: 10922351]
17. Hernán MA, Robins JM. Estimating causal effects from epidemiological data. Journal of epidemiology and community health. 2006; 60(7):578–86. [PubMed: 16790829]
18. Boone-Heinonen J, Gordon-Larsen P. Obesogenic environments in youth: concepts and methods from a longitudinal national sample. American journal of preventive medicine. 2012; 42(5):e37–e46. [PubMed: 22516502]
19. Chen SE, Florax RJ, Snyder SD. Obesity and fast food in urban markets: a new approach using geo-referenced micro data. Health economics. 2013; 22(7):835–56. [PubMed: 22911977]
20. Dunn RA. The effect of fast-food availability on obesity: an analysis by gender, race, and residential location. American Journal of Agricultural Economics. 2010:aaq041.
21. Dunn RA, Sharkey JR, Horel S. The effect of fast-food availability on fast-food consumption and obesity among rural residents: an analysis by race/ethnicity. Economics & Human Biology. 2012; 10(1):1–13. [PubMed: 22094047]
22. Liu K, Slattery M, Jacobs D Jr, Cutter G, McDonald A, Van Horn L, et al. A study of the reliability and comparative validity of the cardia dietary history. Ethnicity & disease. 1994; 4(1):15–27. [PubMed: 7742729]
23. Nettleton JA, Schulze MB, Jiang R, Jenny NS, Burke GL, Jacobs DR. A priori–defined dietary patterns and markers of cardiovascular disease risk in the Multi-Ethnic Study of Atherosclerosis (MESA). The American journal of clinical nutrition. 2008; 88(1):185–94. [PubMed: 18614740]
24. Sijtsma FP, Meyer KA, Steffen LM, Shikany JM, Van Horn L, Harnack L, et al. Longitudinal trends in diet and effects of sex, race, and education on dietary quality score change: the Coronary Artery Risk Development in Young Adults study. The American journal of clinical nutrition. 2012; 95(3):580–6. [PubMed: 22301926]
25. Schuetz J, Kolko J, Meltzer R. Are poor neighborhoods “retail deserts”? Regional Science and Urban Economics. 2012; 42(1):269–85.

26. Han E, Powell LM, Zenk SN, Rimkus L, Ohri-Vachaspati P, Chaloupka FJ. Classification bias in commercial business lists for retail food stores in the US. Int J Behav Nutr Phys Act. 2012; 9(1): 46. [PubMed: 22512874]
27. Liese AD, Colabianchi N, Lamichhane AP, Barnes TL, Hibbert JD, Porter DE, et al. Validation of 3 food outlet databases: completeness and geospatial accuracy in rural and urban food environments. American Journal of Epidemiology. 2010; 172(11):1324–33. [PubMed: 20961970]
28. Oliver LN, Schuurman N, Hall AW. Comparing circular and network buffers to examine the influence of land use on walking for leisure and errands. International journal of health geographics. 2007; 6(1):1. [PubMed: 17214903]
29. Gunasekara FI, Carter K, Blakely T. Glossary for econometrics and epidemiology. Journal of epidemiology and community health. 2008; 62(10):858–61. [PubMed: 18791041]
30. Baum CF, Schaffer ME, Stillman S. Instrumental variables and GMM: Estimation and testing. Stata journal. 2003; 3(1):1–31.
31. Stock JH, Wright JH, Yogo M. A survey of weak instruments and weak identification in generalized method of moments. Journal of Business & Economic Statistics. 2012
32. Caspi CE, Sorensen G, Subramanian SV, Kawachi I. The local food environment and diet: a systematic review. Health Place. 2012; 18(5):1172–87. DOI: 10.1016/j.healthplace.2012.05.006 [PubMed: 22717379]
33. Hogan JW, Lancaster T. Instrumental variables and inverse probability weighting for causal inference from longitudinal observational studies. Statistical Methods in Medical Research. 2004; 13(1):17–48. [PubMed: 14746439]
34. Easterly W. Inequality does cause underdevelopment: Insights from a new instrument. Journal of Development Economics. 2007; 84(2):755–76.
35. Gabel M, Scheve K. Estimating the effect of elite communications on public opinion using instrumental variables. American Journal of Political Science. 2007; 51(4):1013–28.
36. Jeffery RW, Baxter J, McGuire M, Linde J. Are fast food restaurants an environmental risk factor for obesity? Int J Behav Nutr Phys Act. 2006; 3:2.doi: 10.1186/1479-5868-3-2 [PubMed: 16436207]
37. Zick CD, Hanson H, Fan JX, Smith KR, Kowaleski-Jones L, Brown BB, et al. Re-visiting the relationship between neighbourhood environment and BMI: an instrumental variables approach to correcting for residential selection bias. Int J Behav Nutr Phys Act. 2013; 10(1):27. [PubMed: 23425701]
38. Cummins S, Petticrew M, Higgins C, Findlay A, Sparks L. Large scale food retailing as an intervention for diet and health: quasi-experimental evaluation of a natural experiment. Journal of Epidemiology and Community Health. 2005; 59(12):1035–40. [PubMed: 16286490]
39. Sadler RC, Gilliland JA, Arku G. A food retail-based intervention on food security and consumption. International journal of environmental research and public health. 2013; 10(8):3325–46. [PubMed: 23921626]
40. Wrigley N, Warm D, Margetts B. Deprivation, diet, and food-retail access: findings from the Leedsfood deserts' study. Environment and Planning A. 2003; 35(1):151–88.

SUMMARY BOXES

What is already known on this subject?

In two or three sentences explain what the state of scientific knowledge was in this area before you did your study and why this study needed to be done. Be clear and specific.

- The majority of food environment and diet studies use non-causal modeling approaches (e.g., ordinary least squares regression), which ignore potential residual confounding and may contribute to mixed findings in the literature.
- A causal approach to correct for residual confounding bias is instrumental-variables (IV) regression, which theoretically provides less biased estimates relative to non-causal models.
- To our knowledge, no studies have examined the potential for unobserved bias in the context of the food environment and diet outcomes.

What this study adds?

Give a simple answer to the question “What do we now know as a result of this study that we did not know before?”. Be brief, succinct, specific, and accurate. You might use the last sentence to summarize any implications for practice, research, policy, or public health.

- We found that the relative availability of fast food restaurants was negatively associated with diet quality over time using a causal approach (IV regression), and that estimates from non-causal models (ordinary least squares and foxed effects regression) were attenuated relative to causal models, potentially due to residual confounding.
- Our results emphasize the importance of assessing potential bias in studies of the food environment and health, and future observational studies may benefit from using causal methods that explicitly account for bias.

TABLE 1

Descriptive statistics for study participants over the study period[1]: CARDIA exam years 0 (1985/86), 7 (1992/3), and 20 (2005/06)

Exam year	**Year 0**	**Year 7**	**Year 20**
N	**4318**	**2374**	**1060**
Individual-level socio-demographics (% or mean (SD))			
White	44.0	41.8	30.1
Female	53.9	55.0	59.2
Education ≤high school	41.8	29.2	24.2
Income ≤$45,000	79.8	74.3	52.6
Marital status (yes)	21.7	38.3	39.9
Children (yes)	34.1	57.5	74.1
Age, yr	24.8 (3.7)	32.0 (3.7)	45.2 (3.7)
Individual-level diet outcomes (mean (SD))			
a priori diet quality score	61.5 (12.6)	65.2 (12.3)	67.3 (12.6)
Prudent diet pattern score[3]	−0.04 (0.99)	−0.08 (0.97)	−0.15 (0.97)
Western diet pattern score[4]	0.06 (1.03)	0.10 (1.09)	0.16 (1.18)
Fruit (svgs/wk)	1.5 (1.8)	1.7 (2.0)	1.7 (1.8)
Vegetables (svgs/wk)	2.5 (2.3)	2.9 (2.5)	2.8 (2.3)
Whole grains (svgs/wk)	1.5 (1.6)	1.7 (1.7)	1.5 (1.6)
Red/processed meat (svgs/wk)	4.3 (4.0)	3.7 (3.5)	2.9 (2.7)
Sugar-sweetened beverages (svgs/wk)	1.7 (2.3)	1.5 (2.3)	1.2 (2.6)
Neighborhood-level food outlets within 1km (mean (SD))[2]			
Fast food restaurants, % per total restaurants	65.0 (46.0)	46.0 (34.7)	41.7 (31.1)
Full-service restaurants, % per total restaurants	2.6 (9.9)	39.8 (33.4)	41.5 (31.1)
Convenience stores, % per total food stores	29.6 (28.3)	37.8 (25.8)	38.5 (28.0)
Grocery stores, % per total food stores	60.4 (33.2)	53.9 (28.0)	45.4 (29.6)
Supermarkets, % per total food stores	1.5 (6.3)	3.3 (8.3)	5.4 (11.2)
Total restaurants, count	2.8 (4.5)	10.5 (17.7)	11.9 (26.0)
Total food stores, count	5.4 (4.3)	11.3 (8.4)	8.1 (8.7)

[1] Data are shown for each of the three CARDIA exam periods included in the analysis, with Year 0 being the study baseline.

[2] Values represent the mean for all CARDIA participants per year and thus do not equal 100% for total restaurants or total food stores.

[3] The “Western” diet pattern score has a theoretical mean and standard deviation of 0 and 1, respectively; a higher score indicates lower diet quality (range in analytic sample: −3.0 to 17.8).

[4] The “Prudent” diet pattern score has a theoretical mean and standard deviation of 0 and 1, respectively; a higher score indicates higher diet quality (range in analytic sample: −3.3 to 9.5).

TABLE 2

Goodness-of-fit statistics for evaluating strength of identification of endogenous variables: CARDIA exam years 0, 7, and 20

	***F*-statistic**	***p*-value[1]**
Fast food restaurants, % per total restaurants	87.4	<0.001
Convenience stores, % per total food stores	24.5	<0.001
Grocery stores, % per total food stores	55.1	<0.001
Total food outlets, count	288.5	<0.001
Marital status	14.3	<0.001
Number of children	9.9	<0.001

[1] Rejecting the *F*-test indicates that our set of instruments provides good identification for that endogenous variable.

TABLE 3

Beta coefficients (95% confidence intervals)[1] for the associations between each type of food store or restaurant[2] and *a priori* diet quality score[3], using ordinary least squares, fixed effects, and instrumental variables regression: CARDIA exam years 0 (1985/86), 7 (1992/3), and 20 (2005/06)

	OLS regression[4]	FE regression[5]	IV regression[6]
N	***7,740***	***7,740***	***7,740***
Full-service restaurants, % per total restaurants [7]	0.00	0.00	0.00
Fast food restaurants, % per total restaurants	−0.01 (−0.01, 0.0004)	−0.003 (−0.01, 0.01)	−0.10 (−0.20, −0.004)
Supermarkets, % per total food stores [7]	0.00	0.00	0.00
Convenience stores, % per total food stores	−0.002 (−0.02, 0.01)	−0.001 (−0.02, 0.01)	0.14 (−0.21, 0.49)
Grocery stores, % per total food stores	0.003 (−0.01, 0.02)	0.005 (−0.01, 0.02)	−0.03 (−0.18, 0.12)
Total food outlets, count	0.03 (−0.02, 0.05)	0.03 (0.01, 0.04)	−0.02 (−0.09, 0.04)

[1] Multivariable-adjusted models were adjusted for individual-level age, gender, race, educational attainment, income, children, marital status, exam year, and C2ER food prices

[2] Calculated within a 1-km network buffer of participants' residences.

[3] *a priori* diet quality score, mean (SD): 63.4 (12.7)

[4] Ordinary least squares regression (OLS) using Stata's -reg- command with robust variance.

[5] Repeated measures fixed-effects (FE) regression using Stata's -xtreg- command with the 'fe' option.

[6] Instrumental variables (IV) regression using Stata's -ivregress- command with the 'gmm' option.

[7] Omitted from the model (referent)