The Role of Food Access in Meeting Some Dietary Guidelines: A Natural Experiment

by

Gayaneh Kyureghian

Research Assistant Professor
Department of Food Science and Technology
University of Nebraska-Lincoln,
Lincoln, NE 68583

Rodolfo M. Nayga, Jr.

Professor and Tyson Endowed Chair
Department of Agricultural Economics and Agribusiness
University of Arkansas
Fayetteville, AR 72701

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's 2012 AAEA Annual Meeting, Seattle, Washington, August 12-14, 2012

The Role of Food Access in Meeting Some Dietary Guidelines: A Natural Experiment

Abstract

We investigate how the supply of retail food outlets affects the household purchase of fruits and vegetables. We are particularly interested in the specific effects of the increased supply on household of fruit and vegetable purchases in underserved areas. Difference-in-difference type fixed-effect OLS regressions are used to estimate these effects for the pooled sample and for the subsamples by income level, in two settings – pooled cross-section and panel data settings. The findings indicate that the increased availability is not associated with an increase in fruit and vegetables quantity purchased. No discernible effects are detected for underserved areas and for income-differentiated subsamples. The results indicate that the policy intervention to increase the number of local food outlets not increase fruit and vegetable consumption in general, and the target underserved communities will likely not be affected.

Keywords: Food deserts, Fruits and Vegetables, Food Availability, Difference-in-Difference, Natural Experiment

JEL codes: I18, R23

This work was partially funded by RIDGE grant from the Economic Research Service, USDA. This work was also partly supported by the National Research Foundation of Korea -Grant funded by the Korean Government (NRF-2011-330-B00074).

Introduction

Poor food choices have been shown to contribute to the rise of major chronic diseases, including overweight and obesity[1]. Consequently, the Dietary Guidelines for Americans, 2010, emphasizes the need to shift food intake patterns to a more plant-based diet that emphasizes nutritious food such as fruits and vegetables (F&V). Despite these efforts, only 42% and less than 60% of Americans meet the recommendations for F&V consumption, respectively[2]. Despite the inconclusive empirical evidence, the popularly held belief is that the improvement of local shopping opportunities will lead to improved diet (Cummins *et al.*, 2005). The objective of this research is to investigate the empirical evidence of a causal relationship between retail food availability and food choice.

Limited access to nutritious food, frequently held accountable for F&V under-consumption in the academic literature and popular press, is commonly associated with what is widely referred to as "food deserts". Among various food retail outlets, supermarkets and groceries have received much attention primarily due the price affordability and wide assortments of F&V these entities typically offer (Larson, Story and Nelson, 2009; Larsen and Gilliland, 2009). The literature findings on food access and food choice are mixed, often contradictory, possibly due to socio-demographic and geographic coverage, scope of research and identification issues. For example, using data from regional surveys, Blanchard and Lyson (2002) find that residents of food deserts are 23.4 % less likely to consume the recommended level of F&V compared to non-food desert residents. Pearson *et al.* (2005) find no evidence of associations between the distance to the nearest supermarket and the difficulty of grocery shopping with either fruit or vegetable consumption. Bodor *et al.* (2007) find a positive association between the availability of fresh produce in the vicinity and vegetable intake, but not fruit intake. Rose and Richards (2004) find that easy access to supermarket shopping is positively associated with household fruit consumption, while the distance to the supermarket was inversely related to it. They did not find similar associations for vegetables.

In a comprehensive review of literature on disparities in access to healthy food, Larson, Story and Nelson (2009) report that although the majority of studies suggest direct relationship between the presence of supermarkets and meeting the dietary guidelines for F&V, especially for black adults, no such evidence was found for the youth. Likewise, in a systematic review of food deserts, Beaulac, Krisjansson and Cummins (2009) report mixed results concerning the availability and quality of healthy foods in disadvantaged areas. Bitler and Haider (2011) provide a comprehensive analysis and discussion of the empirical literature on food deserts.

Michimi and Wimberly (2010) use national level cross sectional data for 7 years to demonstrate an inverse association between the odds of consuming F&V five times a day and the distance to supermarket in metropolitan areas, but not in non-metropolitan areas. Kyureghian,

[1] Centers for Disease Control and Prevention, 2010.
[2] The Dietary Guidelines for Americans, 2010.

Nayga and Bhattacharya (2012) use national level cross sectional purchase data for 2008 to report positive association between the county level supermarket density and F&V purchases in a cross-sectional sample, but find negative associations for low-income households in underserved areas. Even though these studies use a geographic and socio-demographic coverage at the national level, the cross sectional nature of the data precludes from distinguish time trend or 'common time' effect on F&V purchases from the supply shock effect. It also makes it impossible to disentangle the effects due to the systematic differences among households that could give rise to changes in F&V purchases but that have nothing to do with the supply shock or intervention. From this point of view 'before and after' data from a 'natural' experiment setting would seem to be more appropriate (Cummins *et al.*, 2005). For example, Cummins *et al.*, 2005, conduct a 'before and after' survey and report that there is weak evidence for an effect of intervention based on data from a natural experiment.

The natural experiment setting constitutes a unique experiment that can help increase our understanding of the linkage between food availability and food choice. Given the difficulty to conduct natural experiments on national scale, it is hard to obtain appropriate control and treatment groups to derive proper statistical inference. Moreover, identifying and isolating changes in purchase solely due to increase in availability is confounded by the impact of changes in other factors such as changes in the household size, marital status, educational attainment, employment status, income, etc. We will use a standard difference-in-difference approach to model the association between increase in availability and choice.

In this study we set to test two hypotheses: (i) increased food availability induces increase F&V purchases; and (ii) the increase of food availability in underserved areas (food deserts) induces increase in F&V purchases. For this research we use the Nielsen HomeScan purchase data from 2005 to 2006 to estimate the causal effect of increases in the supply of food retail outlets on F&V purchases. We will compare the evolution of purchases by households that were 'treated' (exposed to increased number of groceries, supermarkets, supercenters and price clubs) with the purchases of households that did not face improved shopping opportunities from 2005 to 2006 ('control' group). We are using difference-in-difference (DD) and triple difference (DDD) estimation methods to insure that the change in F&V consumption is attributable to the improved availability and rule out spurious effects due to other changes. Our results suggest that there is little evidence to support the popular belief that improved food retail environment would indeed induce increased F&V purchases, and therefore would eventually improve diets.

Data and Summary Statistics

Data for this study come from three sources – the Nielsen HomeScan Panel; County Business Patterns, U.S. Census Bureau; and Population Estimates, U.S. Census Bureau. We draw on 2005 and 2006 County Business Patterns and Population Estimates, U.S. Census Bureau, to delineate

the food retail environment and the population/area estimates for the geographical units in our analysis. We then align the information on food access with actual household purchase data from the Nielsen panel from the same areas or counties.

Nielsen, one of the largest commercial supplier of scanner data, started collecting in-home household scanner data in 1989. The panel members, selected from all 48 contiguous states, are supplied with handheld scanners to scan Universal Product Codes (UPCs) of all purchases and to upload this information on a weekly basis. The data are categorized in five datasets by food type: frozen foods, produce and meat products with UPCs, random-weight products without a UPC, dairy products, dry grocery products, and alcohol and cigarettes. Each record in the data set contains a household identification number, purchase date, a set of variables that combined provide a complete description of each product (product type variables), quantity purchased, price, etc. Additionally, information is available concerning demographic variables (county of residence, race, ethnicity, household income, household size, household head education and employment status, household composition, etc.) associated with the respective households.

Although the Nielsen HomeScan data are collected from a nationally representative cross section of households over time, the data set is referred to as panel as a sizable portion of the households continue the membership from one year to the next. There were 38,802 and 37,719 households participating in the Nielsen panel for 2005 and 2006, respectively. Altogether there were 46,301 households that participated in one year or both. A panel of 30,255 households or approximately 78% and 80% of 2005 and 2006 cross-sections, respectively, participated in both years.

County Level Variables

The food accessibility data, obtained from the County Business Patterns, U.S. Census Bureau, include the number of establishments of the following store formats: supercenters and price clubs (hereafter Supercenters, including North American Industry Classification System (NAICS) codes 44511 and 452910) and quick-service restaurants (hereafter QS, including NAICS code 7222) for 3,091 counties. For each county a binary variable 'Improved Supercenters' was created that equals to unity if the number of establishments (NAICS codes 44511 and 452910) has increased from 2005 to 2006. A similar variable was created for the quick-service restaurants – 'Improved QS'. Many studies on food deserts use commuting distance of 10 miles or more as an indicator for food deserts. Following these studies, we identified counties that have 1 or less supercenters per 314^3 square miles as 'Underserved'.

[3] An alternative definition of 'Underserved' – differentiated by Metropolitan or Micropolitan Areas, with 1and 3 miles or more for metro areas and 10 miles for non-metro areas, was used for the robustness check of the results. The results were close to the reported ones and, therefore, are not reported in the Results section.

Household Level Variables

An indicator variable 'Better Sup' was created to indicate if the household exposed to better retail conditions in 2006 due to increase in the number of Supercenters in the residence county from 2005 to 2006, or due to moving to a new county with a higher number of Supercenters in 2006. A similar variable was created to indicate an increased access to QS restaurants – 'Better QS'. In case the households moved, the Supercenters and QS variables were adjusted by expressing the levels by 100 square miles. A binary variable 'Moved' was created to capture the change due to moving to a new place. Quantity and Price measure the quantity in ounces and price per ounce of fresh and processed F&V. PIR – poverty income ratio, is the ratio of household income to poverty threshold issued by the U.S. Department of Health and Human Services for each year. Households with PIR less than 1.35, from 1.35 to 1.85, and greater than 1.85 are combined in income groups 'Income 1', 'Income 2' and 'Income 3', respectively[4].

The descriptive statistics of the full set of variables for the panel are in Table 1. Households in the panel sample purchase slightly more F&V than households in the cross-sectional sample in general. In both samples the purchase quantity decreased from 2005 to 2006, while the mean price increased from \$0.100/oz to \$0.106/oz. In 2006 approximately 80% of both panel and cross-sectional samples are in the highest income groups, with 9% and 11% being in the lowest income groups, respectively.

Approximately 34% of both samples in both years resided in counties that experienced an increase in Supercenters in 2006. 33% of households were exposed to increased Supercenters availability due to either increase in the number of establishments in their respective counties (if stayed) or due to moving to counties with higher number of establishments. Approximately 15% of households reside in counties that were under-served in 2005.

The demographic information indicates that approximately 59% of households were married, with 2.3 members, on average. The panel sample seems to have slightly higher number of household size than the cross-section sample. 37% and 36% of female household heads are full-time employed in the cross-section and panel samples, respectively. The full-time employment rates for male household heads are slightly lower in the panel sample than in the cross-sectional sample with both declining in 2006. The educational attainment for female household heads is consistently higher than that of male household heads. In 2005 and 2006, 76.8% and 78.4% of households in the cross-sectional sample and 79.3% and 80.6% of households in the panel sample, respectively, had no children under 18 in the household.

[4] Natural cutoffs of PIR $\leq$ 1.35 and PIR $\leq$ 1.85 used for the eligibility for participation in the National Food Stamp Program (the 2005 and 2006 HHS Poverty Guidelines, U.S. Department of Health and Human Services.)

Econometric Specification

The baseline model specification is a difference-in-difference type fixed-effect OLS regression model (Conley and Taber, 2011; Donald and Lang, 2007). Let i index a household observed in county j at time t.

$$Q_{ijt} = \alpha_j + \beta_t + D_{jt}\gamma + Z_{ijt}\delta + u_{ijt} \qquad (1)$$

where the dependent variable Q_{ijt} is the quantity (oz) of F&V purchased by household i in county j at year t; α_j are time-invariant county fixed effects; β_t are year fixed effect; Z_{ijt} are household-specific regressor (for example, demographic variables); D_{jt} are group-time effects; and u_{ijt} is the idiosyncratic error.

To address the data issues described above, the baseline model (1) is modified and estimated by methods developed for both cross-sectional and panel data structures (Wooldridge, 2002).

The data structure of pooled cross sections over time can be used for estimating 'treatment' or 'intervention' effects in natural experiment settings such as the supply shock in this analysis (Wooldridge, 2007, p. 129). In this setting the data are considered to be independently, but not identically distributed. The independence requirement - the incidence of some households being in both cross sections is considered purely incidental and random, may be violated as 78% of the 2005 cross section continues to the next year. Hence, the need for dual analysis at panel and cross-sectional levels. The 'treatment' is the increase in Supercenters in some counties from 2005 to 2006 at the county level – all households in the county are assumed to have been exposed to a uniformly improved availability. The households residing in these counties are the treatment group, households residing in the rest of the counties comprise the control group. The changes in F&V purchases for both treatment and control groups in both pre-treatment and post-treatment periods are modeled in this scenario with the purpose to isolate and estimate the change in the dependent variable due to the treatment.

The variable that measures the treatment is 'Improved Supercenters'. A similar variable – 'Improved QS', indicates an increase in quick service restaurants that was included to account for the 'food swamp' effect mentioned in the previous literature (Congress, 2009; Larson, Story and Nelson, 2009). To identify the treatment effect, other factors that might contribute to the change in the dependent variable, separately or combined, but have nothing to do with the treatment, such as income, household size, marital status, education level and employment status change, have been included in the model as well. The baseline model (1) is modified to estimate

$$
\begin{aligned}
Q_{ijt} = \alpha_j + \beta_t + Z_{ijt}\delta \\
&+ \gamma_1(Year\_2006) \times (Improved\ Supercenters)_{jt} \\
&+ \gamma_2(Year\_2006) \times (Improved\ QS)_{jt} + \gamma_3(Year\_2006) \times (Undersearved)_{jt} \\
&+ \gamma_4(Improved\ Supercenters) \times (Improved\ QS)_{jt} \\
&+ \gamma_5(Undersearved) \times (Improved\ Supercenters)_{jt} + u_{ijt}
\end{aligned} \tag{2}
$$

The parameter of interest in (2) is $\gamma_1$which measures the change in F&V purchases due to increased availability, as expressed in hypothesis 1. Statistically insignificant $\gamma_1$would indicate that making new Supercenters available would not induce more F&V purchase or consumption.

An alternative data structure is panel when households in both 2005 and 2006 cross sections are retained (Wooldridge, 2007, p. 265). In this setting, the 'treatment' is the increase in availability for some households from 2005 to 2006 either due to increase in the number of Supercenters in the county or due to moving to a new county with a higher number of Supercenters. Therefore the treatment is at the household level – it is possible that a household in a county is 'treated' while the rest of the households in the county are not. These households are the treatment group, the rest of the households are the control group.

The variable that measures this treatment is 'Better Sup'. A similar variable – 'Better QS', indicates a better access to quick service restaurants. The same set of identifying controls is added to the modified model (3). The parameter of interest in (3) is again γ_1.

$$
\begin{aligned}
Q_{ijt} = \alpha_j + \beta_t + Z_{ijt}\delta + \gamma_1(Year\_2006) \times (Better\ Sup)_{jt} \\
&+ \gamma_2(Year\_2006) \times (Better\ QS)_{jt} + \gamma_3(Year\_2006) \times (Undersearved)_{jt} \\
&+ \gamma_4(Better\ Sup) \times (Better\ QS)_{jt} + \gamma_5(Undersearved) \times (Better\ Sup)_{jt} \\
&+ u_{ijt}
\end{aligned} \tag{3}
$$

To test hypothesis 2 of this research we turn to the triple difference models. The DDD models are further fine-tuning the parameter estimates of interest to allow for the food desert interpretation. The intention is to check if the starting point makes a difference. In other words, adding the eleventh Supercenter to a county is going to have the same effect on F&V consumption as adding the first to a similar county? It is important to realize that this hypothesis builds on the first hypothesis – does increased availability translate into increased consumption? Hence the triple difference specification that compares the treated group in 2006 that is also underserved to the rest of the sample. The corresponding modifications to DD models therefore are estimated as

$$
\begin{aligned}
Q_{ijt} = \alpha_j + \beta_t + Z_{ijt}\delta \\
&+ \gamma_1(Year\_2006) \times (Improved\ Supercenters)_{jt} \\
&+ \gamma_2(Year\_2006) \times (Improved\ QS)_{jt} + \gamma_3(Year\_2006) \times (Undersearved)_{jt} \\
&+ \gamma_4(Improved\ Supercenters) \times (Improved\ QS)_{jt} \\
&+ \gamma_5(Undersearved) \times (Improved\ Supercenters)_{jt} \\
&+ \delta_1(Year\_2006) \times (Undersearved) \times (Improved\ Supercenters)_{jt} \\
&+ \delta_2(Year\_2006) \times (Moved) \times (Improved\ Supercenters)_{jt} \\
&+ \delta_3(Year\_2006) \times (Undersearved) \times (Improved\ QS)_{jt} + u_{ijt}
\end{aligned} \tag{4}
$$

$$
\begin{aligned}
Q_{ijt} = \alpha_j + \beta_t + Z_{ijt}\delta + \gamma_1(Year\_2006) \times (Better\ Sup)_{jt} \\
&+ \gamma_2(Year\_2006) \times (Better\ QS)_{jt} + \gamma_3(Year\_2006) \times (Undersearved)_{jt} \\
&+ \gamma_4(Better\ Sup) \times (Better\ QS)_{jt} + \gamma_5(Undersearved) \times (Better\ Sup)_{jt} \\
&+ \delta_1(Year\_2006) \times (Undersearved) \times (Better\ Sup)_{jt} \\
&+ \delta_2(Year\_2006) \times (Moved) \times (Better\ Sup)_{jt} \\
&+ \delta_3(Year\_2006) \times (Undersearved) \times (Better\ QS)_{jt} + u_{ijt}
\end{aligned} \tag{5}
$$

The parameter of interest in (4) and (5) is δ_1, which measures the relative increase or decrease in F&V if the treatment is administered in 2006 in under-served areas. The models in both settings were analyzed using difference-in-difference type fixed-effect OLS regression estimations. For panel data, in the special case of only two time periods the difference-in-difference and fixed effect OLS estimates are identical (Wooldridge, 2002).

Empirical Findings

Tables 2 and 3 show the OLS estimates from the cross-sectional and panel analyses, respectively. The results for DD and DDD specifications are presented for the pooled sample (columns 1 and 2) and for subsamples by income group (columns 3 to 8). The dependent variable is the natural logarithm of the number of ounces of F&V purchased. The parameter estimates and test statistics that comprise the formal tests of the hypotheses we set to test in this paper are boldfaced in each model. All standard errors are corrected for heteroskedasticity, but not for intragroup correlation (Donald and Lang, 2007; Moulton, 1990; Moulton, 1986; Conley and Taber, 2011).

In general, the results do not support either hypothesis this research addresses. The results demonstrate that there is no indication that food availability has any causal effect on the food choice.

Pooled Cross-Sectional DD and DDD Models

The results in DD models (Table 2) exhibit no statistical significance. This indicates that the increased number of Supercenters is not associated with increased levels of F&V purchase in the pooled sample or income-based subsamples. The results in the DD model in Income 1 group indicate that households residing in counties with increased number of QS, counties experiencing growth in retail outlets (both Supercenters and QS), and under-served counties experiencing growth in Supercenters purchase 1.07, 7.22 and 2.06 ounces less of F&V, respectively. Similar results are found for the DDD model in the Income 1 subsample. Low-income households in under-served and growing counties buy significantly less produce (7.58 and 2.08 ounces, respectively) than households in the control group.

The parameter estimate of interest in the DDD model is not statistically significant: $\delta_1 = 0.027, p = 0.3633$. The implication of this finding is that an increase in the food retail availability in food deserts does not induce an increase in F&V purchases. No significance in the parameters of interest is found in analysis by income group indicating no difference in the amount of F&V purchase due to the increase in retail outlets.

Panel DD and DDD Model

The parameter estimates of interest in the panel models, namely the parameter estimates to variables that indicate (i) the households that have better access in 2006 than in 2005 purchase more, and (ii) the households that are from underserved areas and have better access in 2006 than in 2005 purchase more, are presented in Table 3. The results indicate that none of γ_1 and δ_1 are significant in any of the models (Table 3).

Based on the estimation results from the pooled sample and Income 1 group subsample, households that were exposed to better retail environment due to moving purchase significantly more (2.29oz) compared to the control group. So did households from counties that experienced increase in Supercenters and QS outlets (pooled sample and Income 1 subsample only). In the pooled model households that resided in under-served areas and that faced increased QS availability purchased significantly less F&V ($\delta_3 = -0.052, p = 0.0775$).

Conclusions

The literature findings on the linkage of diets with metabolic diseases motivated a host of empirical studies seeking to shed light on the reasons giving rise to particular food choices or patterns of it. A favorite factor considered by the professional community and public press is the neighborhood food environment or food availability. Despite the logical appeal, there does not seem to be conclusive evidence in the literature on the existence or the nature of such linkage. This research aims to shed light on some aspects of this question. In particular we seek to answer two questions: (1) is there a positive relationship between the number of stores available in each location and F&V consumption by the households in that location? and (2) is there a pronounced relationship in low-income, under-served neighborhoods?

We use national level purchase data for two years – 2005 and 2006, in a difference-in-difference type fixed effects OLS estimation to model the relationship between food availability and food choice. The analyses were replicated for cross-sectional and panel samples from these two years. Our results suggest no statistically significant association between food access and food choice. No statistical significance emerged when estimating the same models for subsamples determined by different income levels. Based on the current results, the conclusion we offer is that the objective of improving the population's diet through increasing F&V consumption may be attained through increasing food retail outlet availability.

The shortcomings of this research may motivate the need for the future research that will help with a more comprehensive analysis. The current research disregards the bordering effects – it views counties separately, not in the context of clusters. The replication of the current analysis with Metropolitan Statistical Areas as the unit of residence may help with confirming the robustness of our findings. Given the separate findings in the literature, it also may prove helpful to conduct this analysis separately for fruits and vegetable, and separately for fresh and non-fresh produce.

Another shortcoming that could possibly confound the true effects is disregarding the magnitude or the length of the treatment – we account for an increase, but not the size of the increase, or the time span of the exposition to the supply shock. Our indication variable for the increase does not account for how many more stores were added to the neighborhood, or how big the new stores are. The current data do not allow for estimating the improvement of the current stores as well, which may be useful information to account.

References

Beaulac, J., E. Kristjansson, and S. Cummins. 2009. A Systematic Review of Food Deserts, 1966-2007. *Preventing Chronic Disease* 6(3):A105-A115.

Bitler, M., and S.J. Haider. 2011. An Economic View of Food Deserts in the United States. *Journal of Policy Analysis and Management* 30(1):153-176.

Blanchard, T.C., and T.A. Lyson. 2002. Retail Concentration, Food Deserts, and Food Disadvantaged Communities in Rural America. Article ed. Mississippi State University.

Bodor, J.N., D. Rose, T.A. Farley, C. Swalm, and S.K. Scott. 2007. Neighborhood Fruit and Vegetable Availability and Consumption: the Role of Small Food Stores in an Urban Environment. *Public Health Nutrition* 11(4):413-420.

Conley, T.G., and C.R.Taber. 2011. Inference with "Difference in Differences" with Small Number of Policy Changes. *The Review of Economics and Statistics* 93(1): 113-125.

Cummins, S., M. Petticrew, C. Higgins, A. Findlay and L. Sparks. 2005. Large Scale Food Retailing as an Intervention for Diet and Health: Quasi-Experimental Evaluation of a Natural Experiment. *Journal of Epidemiology and Community Health* 59: 1035-1040.

Donald, S.G., and K. Lang. 2007. Inference with Difference-in-Difference and Other Panel Data. *The Review of Economics and Statistics* 89(2): 221-233.

Kyureghian, G., R.M. Nayga, Jr., and S. Bhattacharya. 2012. The Effect of Food Store Access and Income on Household Purchases of Fruits and Vegetables: A Mixed Effect Analysis at the County and MSA levels. Working Paper, University of Nebraska-Lincoln.

Larsen, K., and J. Gilliland. 2009. A Farmers' Market in a Food Desert: Evaluating Impacts on the Price and Availability of Healthy Food. *Health & Place* 15: 1158-1162.

Larson, N.I., M.T. Story, and M.C. Nelson. 2009. Neighborhood Environments: Disparities in Access to Healthy Foods in the U.S. *American Journal of Preventative Medicine* 36(1):74-81.e10.

Michimi, A., and M.C. Wimberly. 2010. Associations of Supermarket Accessibility with Obesity and Fruit and Vegetable Consumption in Conterminous United States. *International Journal of Health Geographics* 9(49):1-14.

Moulton, B.R. 1990. An Illustration of a Pitfall in Estimating the Effect of Aggregate Variables on Micro Units. *The Review of Economics and Statistics* 72: 334-338.

Moulton, B.R. 1986. Random Group Effects and the Precision of Regression Estimates. *Journal of Econometrics* 32: 385-397.

Pearson, T., J. Russell, M.J. Campbell, and M.E. Barker. 2005. Do 'Food Deserts' Influence Fruit and Vegetable Consumption? – A Cross-Sectional Study. *Appetite* 45:195-197.

Rose, D., and R. Richards. 2004. Food Store Access and Household Use among Participants in the US Food Stamp Program. *Public Health Nutrition* 7(8):1081-1088.

United States Department of Agriculture. 2009. Access to Affordable and Nutritious Food: Measuring and Understanding Food Deserts and Their Consequences. Report to Congress. Accessed at*: http://www.ers.usda.gov/publications/AboutPDF*.

Wooldridge, J. M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Massachusetts Institute of Technology Press.

Table 1. Descriptive Statistics of Variables in the Cross-Sectional and Panel Samples.

Variable	Cross-Section 2005 (38,802 HHDs)		Cross-Section 2006 (37,719 HHDs)		Panel 2005 (30,255 HHDs)		Panel 2006 (30,255 HHDs)	
	Mean	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**
F&V Quantity (oz)	6378.50	4605.03	6065.80	4402.10	6556.28	4639.83	6259.68	4446.81
Price (oz)	0.100	0.06	0.106	0.06	0.100	0.06	0.106	0.06
Poverty Income Ratio (PIR)	3.791	2.21	4.049	2.79	3.843	2.21	4.021	2.77
Income								
Income 1 (= 1 if PIR < 1.35)	0.084	0.28	0.091	0.29	0.077	0.27	0.091	0.29
Income 2 (= 1 if 1.35 ≤ PIR < 1.85)	0.131	0.34	0.107	0.31	0.129	0.36	0.110	0.31
Income 3 (= 1 if 1.85 ≤ PIR)	0.786	0.41	0.802	0.40	0.794	0.40	0.799	0.40
Improved Sup (= 1 if in 2006 county has increased number of Supercenters)	0.338	0.47	0.339	0.47	0.337	0.47	0.337	0.47
Improved QS (= 1 if in 2006 county has increased number of QS)	0.631	0.48	0.631	0.48	0.629	0.48	0.630	0.48
Better Sup (= 1 if in 2006 household has access to increased number of Sup)	0.283	0.45	0.289	0.45	0.331	0.47	0.331	0.47
Better QS (= 1 if in 2006 household has access to increased number of QS)	0.584	0.49	0.589	0.49	0.622	0.49	0.622	0.49
Underserved (= 1 if county has less than 1 Supercenters per 314 sq. mile)	0.147	0.36	0.145	0.35	0.148	0.36	0.148	0.36
DD								
(Year 2006) × (Better Sup)					0	0	0.331	0.47
(Year 2006) × (Better QS)					0	0	0.622	0.49
(Year 2006) × (Underserved)					0	0	0.148	0.36
(Better Sup) × (Better QS)					0.225	0.42	0.225	0.42
(Underserved) × (Better Sup)					0.047	0.21	0.047	0.21
(Year 2006) × (Improved Sup)	0	0	0.339	0.47				
(Year 2006) × (Improved QS)	0	0	0.631	0.48				
(Year 2006) × (Underserved)	0	0	0.145	0.35				
(Improved Sup) × (Improved QS)	0.230	0.42	0.231	0.42				
(Underserved) × (Improved Sup)	0.048	0.21	0.047	0.21				
DDD								
(Year 2006) × (Underserved) × (Better Sup)					0	0	0.047	0.21
(Year 2006) × (Better Sup) × (Moved)					0	0	0.002	0.04
(Year 2006) × (Underserved) × (Better QS)					0	0	0.081	0.27
(Year 2006) × (Underserved) × (Improved Sup)	0	0	0.047	0.21				
(Year 2006) × (Improved Sup) × (Moved)	0	0	0.074	0.26				
(Year 2006) × (Underserved) × (Improved QS)	0	0	0.080	0.27				
Married (= 1 if married)	0.589	0.49	0.591	0.49	0.587	0.49	0.587	0.49
Household Size	2.345	1.30	2.302	1.26	2.268	1.24	2.243	1.22

Table 1. – Continued.

Variable	Cross-Section 2005 (38,802 HHDs)		Cross-Section 2006 (37,719 HHDs)		Panel 2005 (30,255 HHDs)		Panel 2006 (30,255 HHDs)	
	Mean	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**
Moved					0.025	0.16	0.025	0.16
Female Head Employment								
Under 30 hours	0.103	0.30	0.101	0.30	0.103	0.30	0.100	0.30
30-34 hours	0.044	0.20	0.042	0.20	0.043	0.20	0.041	0.20
35+ hours	0.370	0.48	0.372	0.48	0.362	0.48	0.357	0.48
Not employed for pay	0.381	0.49	0.384	0.49	0.387	0.49	0.398	0.49
No Female Head	0.104	0.315	0.101	0.30	0.106	0.31	0.105	0.31
Male Head Employment								
Under 30 hours	0.033	0.18	0.032	0.18	0.034	0.18	0.033	0.18
30-34 hours	0.017	0.13	0.018	0.13	0.016	0.13	0.017	0.13
35+ hours	0.440	0.50	0.437	0.50	0.426	0.50	0.416	0.49
Not employed for pay	0.238	0.43	0.240	0.43	0.247	0.43	0.257	0.44
No Male Head	0.273	0.45	0.274	0.45	0.276	0.45	0.277	0.45
Female Head Education								
Grade School	0.005	0.07	0.005	0.07	0.005	0.07	0.005	0.07
Some High School	0.028	0.17	0.027	0.16	0.028	0.16	0.028	0.16
Graduated High School	0.251	0.43	0.250	0.43	0.258	0.44	0.259	0.44
Some College	0.287	0.45	0.283	0.45	0.281	0.45	0.279	0.45
Graduated College	0.233	0.42	0.237	0.43	0.230	0.42	0.230	0.42
Post College Graduate	0.091	0.29	0.097	0.30	0.093	0.29	0.094	0.29
No Female Head or Unknown	0.104	0.31	0.101	0.30	0.106	0.31	0.105	0.31
Male Head Education								
Grade School	0.011	0.10	0.010	0.10	0.011	0.10	0.010	0.10
Some High School	0.039	0.19	0.037	0.19	0.037	0.19	0.037	0.19
Graduated High School	0.188	0.39	0.184	0.39	0.187	0.39	0.187	0.39
Some College	0.214	0.41	0.212	0.41	0.211	0.41	0.209	0.41
Graduated College	0.187	0.39	0.193	0.39	0.188	0.39	0.188	0.39
Post College Graduate	0.088	0.28	0.092	0.29	0.091	0.29	0.091	0.29
No Female Head or Unknown	0.273	0.45	0.274	0.45	0.276	0.45	0.277	0.45
Age and Presence of Children								
Under 6 only	0.028	0.16	0.026	0.16	0.023	0.15	0.021	0.14
6-12 only	0.053	0.22	0.051	0.22	0.049	0.22	0.046	0.21
13-17 only	0.077	0.27	0.073	0.26	0.073	0.26	0.070	0.25
Under 6 & 6-12	0.025	0.16	0.022	0.15	0.020	0.14	0.018	0.13
Under 6 & 13-17	0.004	0.07	0.004	0.06	0.004	0.060	0.003	0.056
6-12 & 13-17	0.039	0.19	0.035	0.18	0.033	0.18	0.032	0.18
Under 6 & 6-12 & 13-17	0.006	0.08	0.005	0.07	0.005	0.07	0.004	0.07
No Children Under 18	0.768	0.42	0.784	0.41	0.793	0.41	0.806	0.40

Table 2. Impact of Improved Grocery Access on Household Fruit and Vegetable Purchases – Difference-in-Difference and Triple Difference Analysis of Cross Sectional Data.

	Full Sample		Income 1 PIR < 1.35		Income 2 1.35 ≤ PIR < 1.85		Income 3 PIR ≥ 1.85	
	DD	DDD	DD	DDD	DD	DDD	DD	DDD
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(Year_2006) × (Improved Supercenters)	**-0.001** **(0.011)**	0.050*** (0.012)	**0.031** **(0.041)**	0.076 (0.047)	**0.009** **(0.032)**	0.067* (0.036)	**-0.006** **(0.012)**	0.045*** (0.013)
(Year_2006) × (Improved QS)	0.002 (0.011)	0.007 (0.012)	-0.064* (0.039)	-0.057 (0.044)	0.035 (0.030)	0.052 (0.034)	0.006 (0.012)	0.012 (0.013)
(Year_2006) × (Underserved)	0.003 (0.014)	0.013 (0.022)	-0.064 (0.044)	-0.045 (0.066)	0.042 (0.037)	0.114* (0.058)	0.008 (0.015)	0.016 (0.025)
(Improved Supercenters) × (Improved QS)	0.152 (0.290)	0.102 (0.292)	-1.977*** (0.384)	-2.026*** (0.392)	-0.315 (0.644)	-0.382 (0.651)	0.043 (0.294)	-0.010 (0.297)
(Underserved) × (Improved Supercenters)	-0.009 (0.110)	-0.013 (0.110)	-0.722** (0.339)	-0.733** (0.339)	0.247 (0.305)	0.262 (0.305)	0.040 (0.126)	0.035 (0.126)
(Year_2006) × (Underserved) × (Improved Supercenters)		**0.027** **(0.030)**		**-0.035** **(0.093)**		**-0.070** **(0.081)**		**0.037** **(0.033)**
(Year_2006) × (Improved Supercenters) × (Moved)		-0.258* (0.017)		-0.186*** (0.071)		-0.247*** (0.060)		-0.264*** (0.019)
(Year_2006) × (Underserved) × (Improved QS)		-0.039 (0.029)		-0.023 (0.089)		-0.088 (0.075)		-0.042 (0.032)
T-2006	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Improved Sup	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Improved QS	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Underserved	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
T-2006 × Household Size	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
T-2006 × PIR	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-Value	11.11 (<.0001)	11.24 (<.0001)	2.47 (<.0001)	2.48 (<.0001)	3.32 (<.0001)	3.33 (<.0001)	9.86 (<.0001)	9.99 (<.0001)
R-Square	0.2731	0.2757	0.4107	0.4116	0.41822	0.4198	0.2842	0.2871
Adj R-Square	0.2485	0.2512	0.2446	0.2454	0.2922	0.2939	0.2554	0.2584
Sample Size	76,160	76,160	6,641	6,641	9,047	9,047	60,445	60,445

Table 3. Impact of Improved Grocery Access on Household Fruit and Vegetable Purchases – Difference-in-Difference and Triple Difference Analysis of Panel Data.

	Full Sample		Income 1 PIR < 1.35		Income 2 1.35 ≤ PIR < 1.85		Income 3 PIR ≥ 1.85	
	DD	DDD	DD	DDD	DD	DDD	DD	DDD
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(Year_2006) × (Better Sup)	**0.002** **(0.012)**	0.001 (0.013)	**0.000** **(0.043)**	0.002 (0.049)	**0.026** **(0.034)**	0.031 (0.038)	**-0.003** **(0.013)**	-0.004 (0.014)
(Year_2006) × (Better QS)	-0.001 (0.011)	0.007 (0.012)	-0.021 (0.039)	-0.023 (0.045)	0.023 (0.032)	0.040 (0.035)	0.002 (0.013)	0.009 (0.014)
(Year_2006) × (Underserved)	0.004 (0.014)	0.029 (0.023)	-0.058 (0.044)	-0.059 (0.067)	0.047 (0.038)	0.104* (0.059)	0.004 (0.016)	0.029 (0.025)
(Better Sup) × (Better QS)	-0.156** (0.076)	-0.157** (0.076)	-1.216*** (0.233)	-1.217*** (0.232)	-0.111 (0.168)	-0.111 (0.168)	-0.085 (0.090)	-0.086 (0.090)
(Underserved) × (Better Sup)	-0.167** (0.090)	-0.165* (0.092)	-0.447 (0.292)	-0.439 (0.297)	0.103 (0.178)	0.115 (0.185)	-0.174 (0.115)	-0.174 (0.115)
(Year_2006) × (Underserved) × (Better Sup)		**0.012** **(0.031)**		**-0.017** **(0.093)**		**-0.007** **(0.083)**		**0.017** **(0.035)**
(Year_2006) × (Better Sup) × (Moved)		0.012 (0.079)		0.827*** (0.221)		0.187 (0.238)		0.016 (0.087)
(Year_2006) × (Underserved) × (Better QS)		-0.052* (0.029)		0.012 (0.087)		-0.099 (0.074)		-0.053 (0.033)
T-2006	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Better Sup	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Better QS	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Underserved	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
T-2006 × Household Size	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
T-2006 × PIR	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-Value	11.21 (<.0001)	30.01 (<.0001)	2.81 (<.0001)	2.81 (<.0001)	3.55 (<.0001)	3.55 (<.0001)	9.86 (<.0001)	9.85 (<.0001)
R-Square	0.3086	0.3086	0.4725	0.4727	0.4644	0.4645	0.3164	0.3164
Adj R-Sq	0.2810	0.2810	0.3047	0.3043	0.3337	0.3335	0.2843	0.2843
Sample Size	60,159	60,159	5,042	5,042	7,191	7,191	47,926	47,926