



AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
<http://ageconsearch.umn.edu>
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

How Millennial Food Purchase Decisions Compare to Previous Generations*

Annemarie Kuhns

USDA, Economic Research Service

amkuhns@ers.usda.gov

Michelle Saksena

USDA, Economic Research Service

mjsaksena@ers.usda.gov

***The views expressed are those of the author(s) and should not be attributed to the Economic Research Service, USDA, or Information Resources, Inc. (IRI).**

*Selected Paper prepared for presentation at the 2016 Agricultural & Applied Economics Association
Annual Meeting, Boston, Massachusetts, July 31-August 2*

Abstract

Millennials, those born between 1980 and 2000, have captured the attention of researchers, media, and the food industry alike, as their tastes and preferences are increasingly shaping what is being purchased at the grocery store. Market analysis has shown that this generation is demanding healthier and fresher items and spending fewer of their food expenditures at restaurants. However, to our knowledge, no research has specifically examined how millennials' purchasing decisions differ after controlling for a robust set of demographic and socioeconomic (SES) variables. The goal of our research is to investigate whether the purchasing decisions of millennial households differ significantly from the rest of the population, looking both at the healthfulness of food purchases as well as the shopping environment used to purchase food-at-home. Overall, our study finds that being a millennial had a small and positive effect on diet quality when we measure diet quality as the deviation from the Dietary Guidelines for Americans. However, once we correct for overconsumption of healthy foods and underconsumption of unhealthy foods, this difference disappears suggesting that millennials are better at complying to the recommended guidelines.

Introduction

Consumers' food choices and dietary quality are central to the study of health outcomes and their related costs, longevity, food access, and food security. As an extension of previous research, this project looks to explore how food purchasing decisions differ based on the head of household's generational cohort. Specifically, this approach assesses consumers' shopping baskets based on their expenditure shares in key food categories, such as fruits, vegetables, prepared foods, meat, etc. This research is an extension of Volpe and Okrent's (2012) work on the healthfulness of consumers' grocery purchases and Todd's (2014) research on diet quality among working-age adults.

Using the concept of Age-Period-Cohort models we attempt to estimate a millennial effect for healthy food purchase behavior. As a work around to the identifiability problem, we first use propensity score matching to control for period effects. Using the subset of households, which were matched, we run a FGLS regression with cluster robust errors. Three different measures for diet quality are used as the dependent variable and we compare the results of each.

Overall, we find that millennials purchase slightly healthier items than non-millennials when using the CNPPScore1 to measure diet quality, which penalizes for deviations from food consumption of each food category. This implies that millennials may do a better job of conforming to dietary guidelines. At first glance, this does not seem to correspond to our initial summary statistics, which indicate that millennials purchase slightly less healthy foods than other cohorts. But, when observing the results from the other two regressions millennials were found

to not exhibit significantly different behavior after correcting for under- and over- consumption. When comparing the results from all three regressions, there is possible evidence that older cohorts may be under eating unhealthy foods, but also undereating healthy foods.

Background

Considerable research has been devoted to cataloguing what foods Americans are eating and how this affects overall diet quality (Kennedy et al., 1999; Hiza et al., 2013). Most recently, several articles have focused on the increasing prevalence of food away from home (FAFH) consumption and how ultimately, this has deteriorated diet quality even further (Mancino and Kinsey, 2008; Todd, Mancino and Lin, 2010; Mancino et al., 2010). While increased FAFH consumption may be a contributing factor to a reduction of diet quality, the majority of the average household's food budget is allocated to food at home (FAH) purchases. Specifically, the Bureau of Labor Statistics' Consumer Expenditure survey states that the average American consumer spent \$3,983 on FAH and \$2,904 on FAFH in 2014. Furthermore, during the most recent recession, the reduction in FAFH consumption only accounted for 20 percent of the overall improvement in diet quality during the recession (Todd 2014). This, in the least, warrants a closer look at how the impact of FAH consumption influences dietary quality for individuals.

Volpe and Okrent (2012) do such an analysis looking at the impact of FAH expenditure on dietary quality over time. When disaggregating by income, race, and geographical location, they found small, but significant differences among races and geographical locations. And, overall, all groups were in need of dietary quality improvements. Similarly, Hiza et al. (2013) found small differences in diet quality across demographic groups and found that all subgroups were in need of dietary quality improvements. In this article we extend Volpe and Okrent's

(2012) work on the health impact of FAH purchases by exploring potential cohort differences in FAH consumption. In particular, we are interested in discerning whether millennials exhibit different FAH purchasing behaviors than the generations before them.

It is not unreasonable to expect millennials to exhibit different consumer behavior than other generations. And, it would seem that both anecdotally and market research suggest that millennials do in fact shop differently than older individuals, placing more importance on convenience and experiential attributes. For example, millennials shop more frequently at gas stations, are more likely to buy organic food, and prefer artisanal alcohol beverages (Tuttle, 2015). Many arguments can be made as to why consumer behavior changes with time. Closer scrutiny of these factors is nonetheless a deserved inquiry. For instance, how much of millennial behavior is due to the individual's age? Or, how do contemporaneous conditions influence an individual's consumption choice at that given moment? Finally, how does the cumulative effect of past concurrent conditions uniquely imprint consumption behavior of that same individual?

Our analysis is motivated by concepts highlighted by Age-Period-Cohort models (APC) popularized in the sociology and demography literature (Feinberg and Mason, 1985; Hobcraft, Menken and Preston, 1985; O'Brien, 2014).¹ The driving concept behind APC models is that each cohort experiences events uniquely as a result of a combination of physiological factors (age effect), current environmental conditions (period effect), and past experiences (cohort effect). We are interested in how the manifestation of this concept uniquely shapes cohort specific consumer FAH preferences.

¹ This is not an exhaustive list of references. APC models are very common in sociology and demography. There are accordingly many articles that are APC applications. To the best of our knowledge, the three references listed have been found to be the most tautological.

APC models suffer fundamentally from an “identifiability” problem. That is, each factor is deterministic of the others. More precisely, $Cohort = Period - Age$. To overcome this problem, longitudinal data is needed to disentangle all three factors. In our analysis we use pooled cross-sectional data and in doing so, impose limitations on the interpretation of the effect of the coefficient on the millennial indicator variable because it captures both age and cohort effects (Mason and Wolfinger, 2001). Nonetheless, the compound effect of age and cohort factors is still informative and still illuminates consumer behavior differences.²

Data and Methods

The primary model we estimate is the Feasible Generalized Least Squares (FGLS) estimator with cluster robust errors.^{3, 4} To capture the compositional “millennial effect” we include an indicator variable for millennials. Our model includes individual and community-level variables, which we believe impact dietary quality and food acquisition behavior. The full model is as follows,

$$(1) Y_{i,c} = \alpha + \delta I_i^{mil} + x_i' \beta_i + z_c' \rho_c + \epsilon_i.$$

Where $Y_{i,c}$ is a measure of diet quality, I_i^{mil} is the indicator variable for millennial, $x_i \beta_i$ is a vector of individual-level variables and $z_c \rho_c$ is a vector of community variables.

² Further research plans are to apply our analysis to longitudinal data so that we can isolate pure cohort and age effects.

³ When performing the Breusch-Pagan test on the least squares estimation for heteroskedasticity, our Chi-Squared statistic was 1499.51, so we strongly reject the null hypothesis of homoskedasticity.

⁴ We use pooled cross-sectional data in our analysis. Thus, to control for within correlation we apply cluster robust standard errors where the unit is the household.

We use Information Resource Inc.'s (IRI) 2012 Consumer Network data. The 2012 sample includes over 100,000 distinct households from across the United States. Each household is classified by various demographic and socioeconomic identifiers. The IRI data allows us to analyze households' food expenditures as the dataset consists of each household's self-scanned food purchases. IRI data provide information on what was purchased, how much was purchased, how much they paid for it, and where they purchased it. The data is also supplemented with demographic information on each household. We utilize the data as a pooled cross-section.

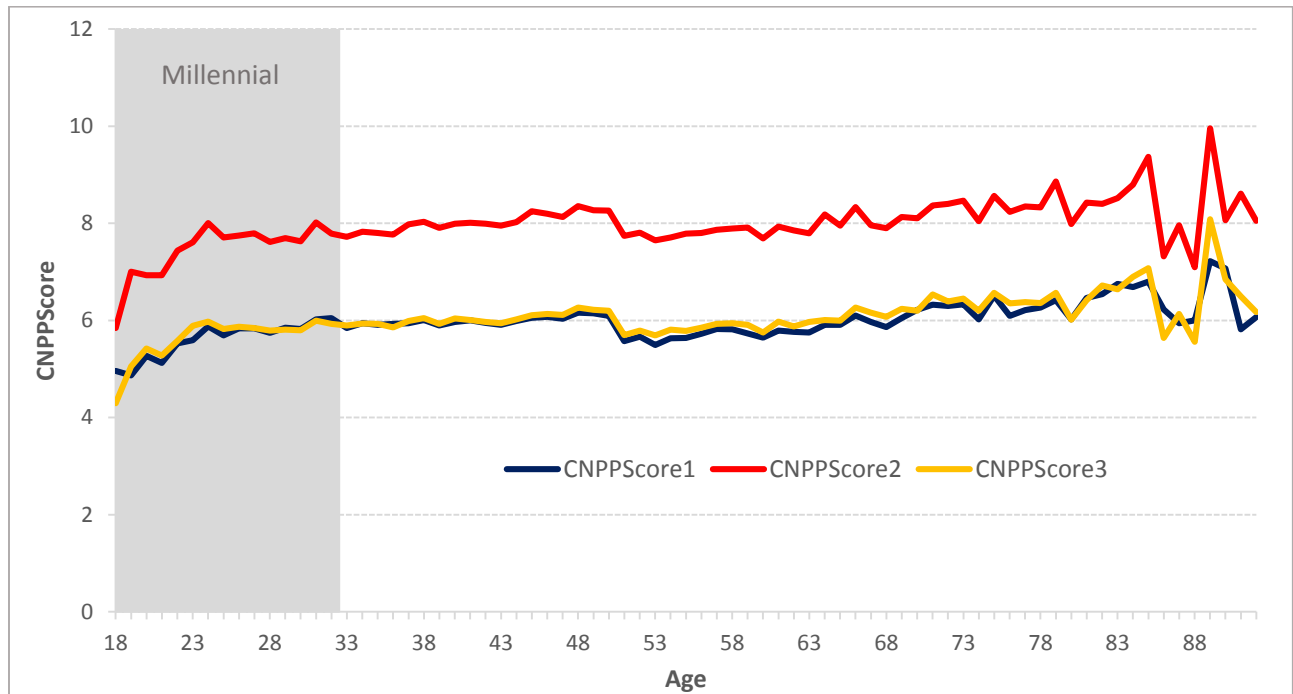
Individual Variables

Using the demographic data available from our IRI panel, we create a *millennial* dummy variable. The *millennial* binary was generated by classifying each head of household doing the grocery shopping as either being born after 1979 or before 1979. Those born after, 1980 to 2000, are classified as millennial shoppers (CEA, 2014). The millennial subset is roughly 10 percent of the total IRI panel. Whereas, nearly 20 percent of the total U.S. population are millennials, indicating that our data may underrepresent this portion of the population or that they still live in households with their parents, who are the primary shoppers.

Looking at the sample, we uncover many interesting patterns in our data.⁵ First, breaking the sample out by age, we can see that households with millennial shoppers generally have lower CNPPScores (chart 1). That is, they, on average, purchase a less healthy basket of goods than shoppers in other generational cohorts.

⁵ A full comparison of the millennial versus non-millennial sample can be found in Appendix A, Table 2.

Chart 1. Average CNPPScore, by Age



In addition, millennials tend to make, on average, fewer trips to the grocery store per month than other age groups. As shown in chart 2, millennials made at most 2.5 trips per month, whereas, shoppers in their sixties made over 3.5 trips per month. This could suggest that millennials eat away from home more relative to other shoppers or that they make larger purchases at the grocery store during each trip.⁶ However, we can see (chart 3) that millennials and those over 65, spend, on average, less per month on groceries than consumers who are between 33 and 65.

⁶ The data we use does not include food away from home purchases. However, looking at the U.S. Bureau of Labor Statistics' Consumer Expenditure survey, in 2014 Millennials spent, on average, 46 percent of their food dollars per month away from home. Whereas, the average American consumer spent 42 percent of their food dollars away from.

Chart 2. Average Number of Trips to the Grocery Store per Month, by Age

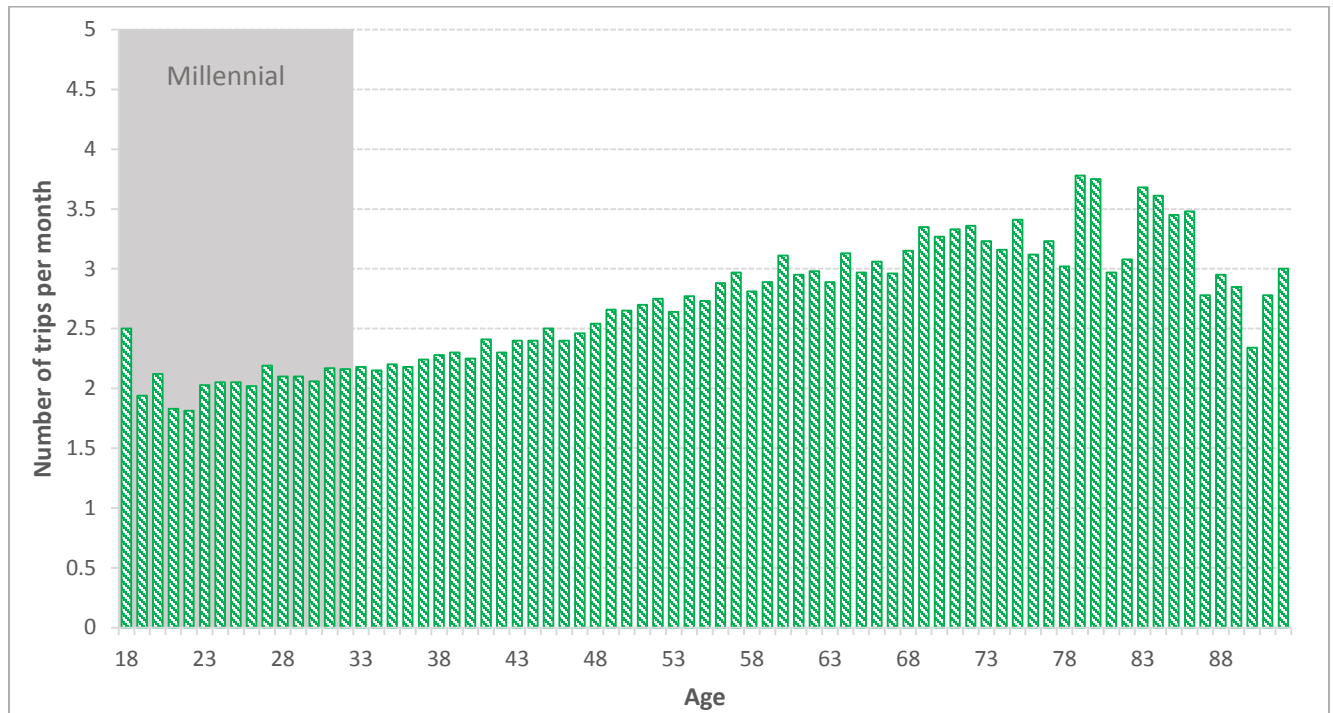
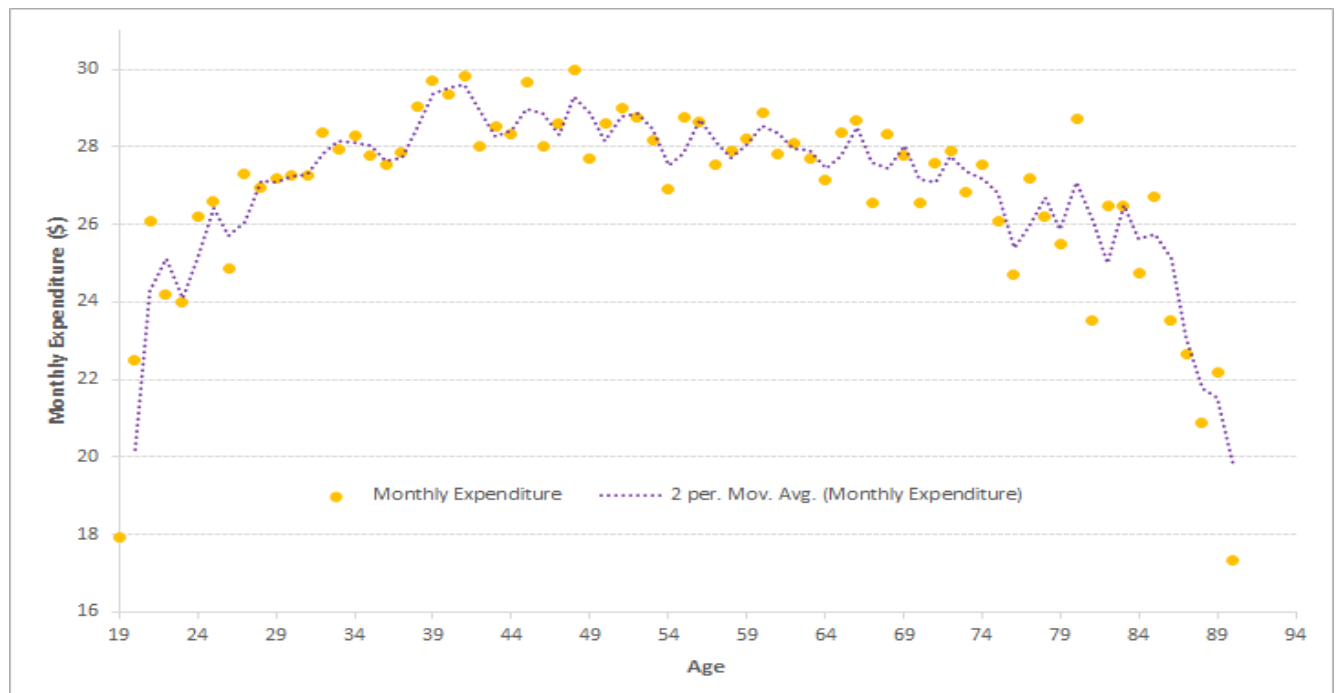
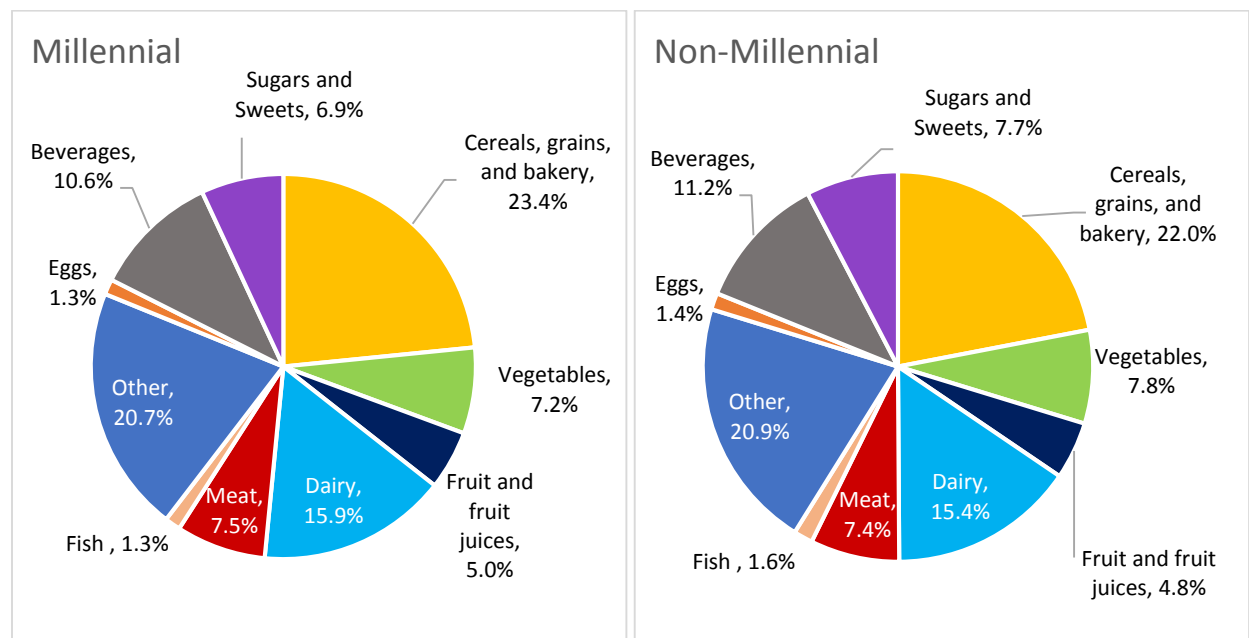


Chart 3. Average Monthly Expenditure, by Age



Using our data, we can also break out the share of expenditures by category (chart 4). Millennials spend a greater share of their food at home expenditures on cereals, grains, and bakery products, fruit and fruit juices, and dairy. However, they spend a smaller percentage of their expenditures on vegetables, meats, fish, other foods, eggs, beverages, and sugars and sweets.⁷ The largest difference occurs in the cereals, grains, and rice category where millennials spend 23.4 percent of their food at home spending and all others spend 22 percent.

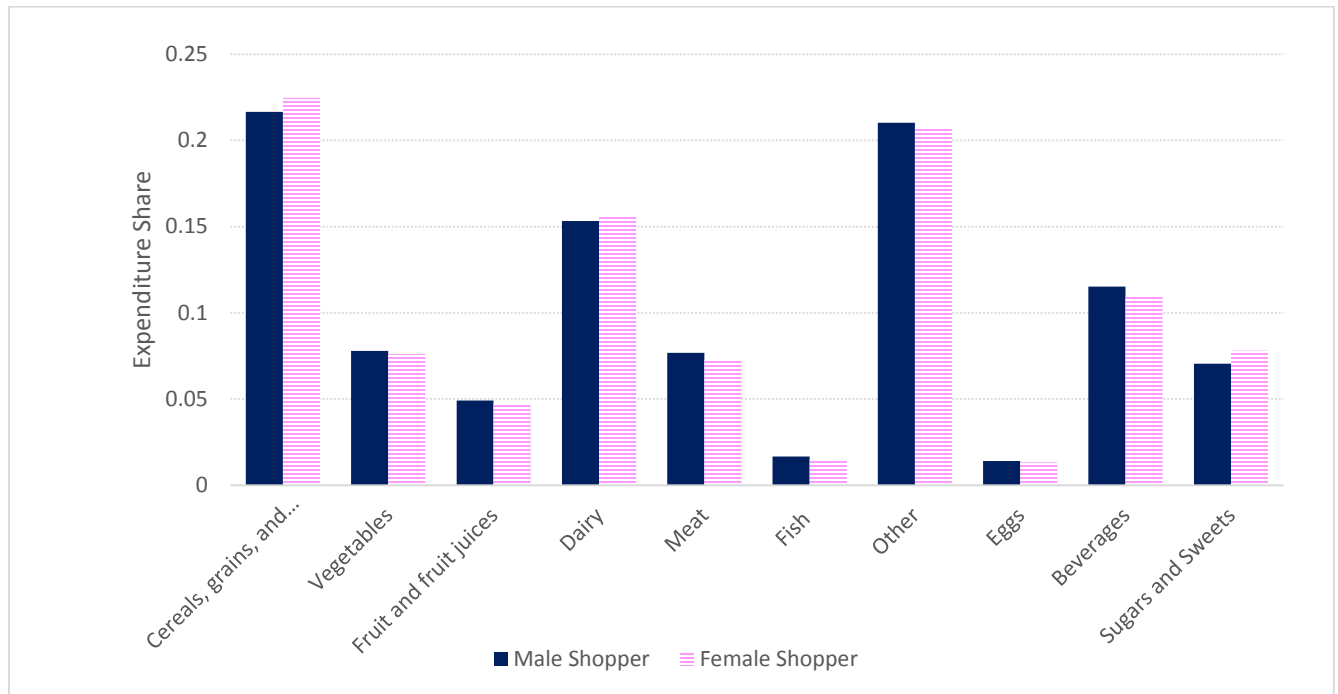
Chart 4. Average Expenditure Shares, by Food Category



In addition to differences in expenditures shares by age, there are also differences based on the gender of the primary shopper. Chart 5 shows that male shoppers tend to, on average, purchase more vegetables, fruits, and proteins than female shoppers, who purchase more cereals, dairy, and sugars and sweets.

⁷ The other food category includes soups, frozen and shelf-stable entrees, and fats and oils.

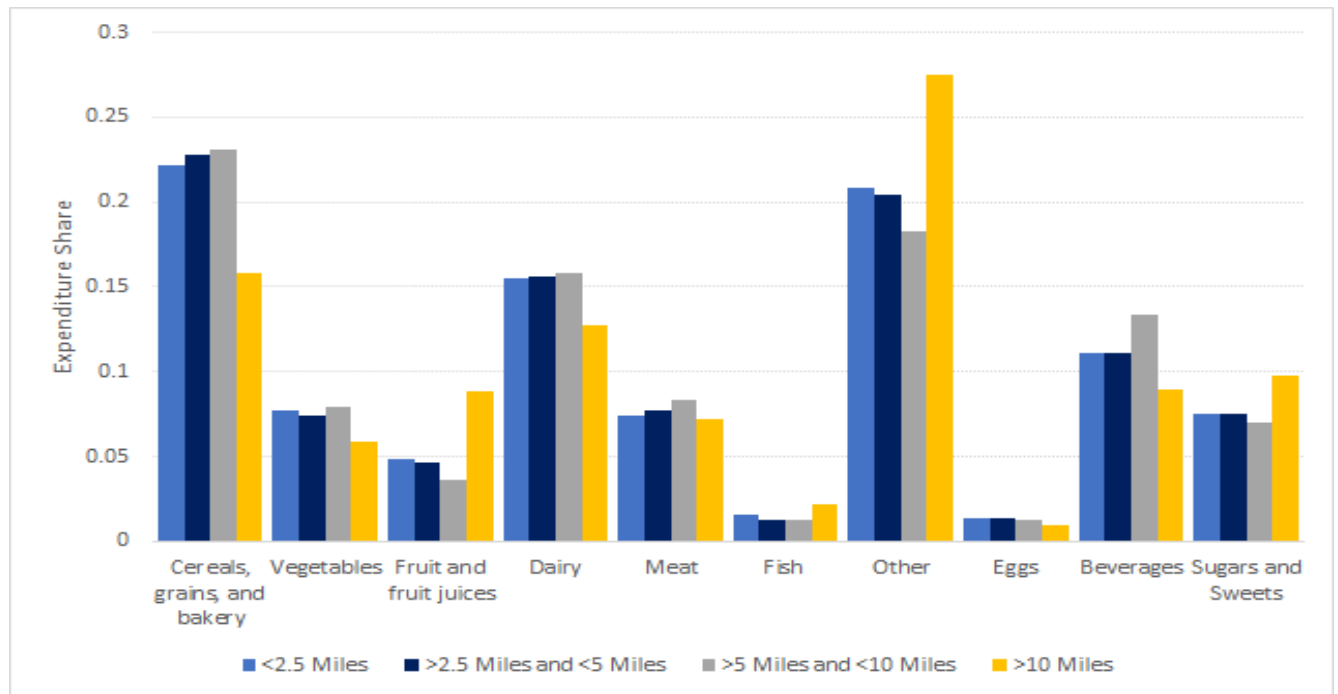
Chart 5. Average Expenditure Shares, by Shopper Sex



In addition to adding a millennial dummy, we augment our data with geographic variables to account for market concentration. Geographical Information System (GIS) modeling was employed to establish the geographic location of each household as well as access to food stores. Each household's food environment is measured using geospatial proximity methods by calculating each household's distance to Nielsen's Trade Dimensions (TDLinx) grocery stores. The TDLinx stores were chosen as the target store dataset since IRI store locations are a subset of the data. This allows for a more complete picture of the surrounding food shopping environment. GIS methods are used to calculate the distances from the IRI household locations to the nearest grocery stores. Euclidean distance is then employed to quantify proximity between household address and the closest food stores in ArcGIS software. We use the distance from the household to the closest grocery store as well as the proportion of the grocery store to total stores within a 1-mile band as measures of the household's food environment. Observing the

household's distance from a grocery store, our *distance* variable, we can see that there are differences among average expenditures shares (chart 6).

Chart 6. Average Expenditure Shares, by Distance



Next, we augmented our data with a variable to measure the healthfulness of each household's monthly shopping basket. In the literature, there are many examples that choose to narrowly define diet quality as a function of a singular food item or nutrient.⁸ While this practice alleviates computational difficulties, it also places limitations on the interpretation of the analysis since diet quality is determined by consumption of an entire suite of foods over time. In doing so, this overestimates the effect of consumption of the health promoting food or nutrient of concern on dietary quality because it does not take into account other compensatory consumption behavior nor does it take into consideration that food items are compositional items, and may have both health promoting and health deteriorating nutrients. For example, Lakdawalla et al.

⁸ See Bertail and Caillavet, 2005; Richards and Patterson, 2005; Mancino et al., 2008; Lin and Yen, 2008 for examples.

(2005) found that a decrease in consumption of ground beef and orange juice would have beneficial effects on reducing fat and sugar consumption, but would also diminish iron, folate and vitamin C consumption levels. Likewise, Okrent and Alston (2011) found that fruit and vegetable consumption would increase with a corresponding price decrease, but the effects would be negated by a total increase in calorie consumption. Zhen et al. (2013) found that a half-cent per ounce increase in price for sugar sweetened beverages would reduce total calorie consumption by 23 calories, but would cause a simultaneous increase in sodium and fat consumption. With these issues in mind, a more inclusive measure of dietary quality would be more forthright.

In order to address the issues illustrated above, we measure the healthfulness, or dietary quality, of food purchases by applying three of the measures conceived by Volpe and Okrent (2012). To begin, we constructed monthly shopping baskets by IRI Panel household. For each monthly basket, we then calculated USDAScore1, USDAScore2 and USDAScore3.

$$(2) \text{USDAScore1}_{icq} = (\sum_c (\text{expshare}_{icq} - \text{USDAexpshare}_{ic})^2)^{-1}$$

$$(3) \text{USDAScore2}_{icq} = (\sum_c (\text{expshare}_{icq} - \text{USDAexpshare}_{ic})^2 | \text{expshare}_{icq} > 0)^{-1},$$

$$(4) \text{USDAScore3}_{icq} = \left[\left(\sum_c (\text{expshare}_{icq} - \text{USDAexpshare}_{ic})^2 \right) | \text{expshare}_{icq} > \text{USDAexpshare}_{ic}, c \in \text{limited} \right] + \left[\left(\sum_c (\text{expshare}_{icq} - \text{USDAexpshare}_{ic})^2 \right) | \text{expshare}_{icq} < \text{USDAexpshare}_{ic}, c \in \text{increased} \right]^{-1}$$

where

$$(5) \text{share}_{igq} = \frac{\text{exp}_{igq}}{\sum_{g=1}^{24} \text{exp}_{igq}},$$

The scores are calculated in compliance with the 2010 *Dietary Guidelines for Americans* (DGA), a joint publication from the U.S. Department of Health and Human Services and the U.S. Department of Agriculture. The DGA is regularly updated every five years to reflect the most up-to-date nutrition research and serves as guidance for Americans on how to make healthy eating choices.

From (2), (3), (4), and (5) *expshare* is the expenditure share each household spent on each food group and *USDAexpshare* is the expenditure share recommended for each household based on the age and gender of household members. *USDA Score1* assigns penalties for any deviation from the recommended amounts, *USDA Score2* excludes food categories for which no purchases are recorded in a given month, and *USDA Score3* does not penalize for too little of a food recommended for limited consumption or too much of a food recommended for increased consumption. Assuming here that there is a mistake or items were purchased or consumed elsewhere. The *USDA Score* is meant to be interpreted ordinal; the higher the *USDA Score*, the more closely the household is adhering to the dietary guidelines.

Ultimately, we are interested in measuring diet quality because of its impact on health, specifically with regards to obesity and other related non-communicable diseases. While measuring health outcomes is outside the purview of our analysis, it is well established that there are distinct contrasts with regards to obesity rates across racial and gender groups. This may be attributed to distinguishing food patterns by race and gender. Specifically, non-Hispanic blacks have an age-adjusted obesity rate of 47.8 percent while non-Hispanic Asians only have a 10.8% rate. This difference is even more pronounced when categorizing by race and gender. For

example, non-Hispanic black women had an age-adjusted obesity rate of 56.6 percent while non-Hispanic Asian men only had an obesity rate of 10.0 percent. Hispanic and Non-Hispanic white populations had the second and third highest obesity rates overall (Ogden et al. 2014).

Several other demographic variables are of interest to our model: marital status, per capita income, and education level. Looking at our data, 46 percent of participating households are married. IRI reports income as a range. Therefore, to calculate per capita income, we first take the mean of the income range and then divided it by the number of individuals within the household.⁹ The average per capita income in our dataset is \$27,029. Additionally, we include household-level dummy variables for education level. We include a dummy for high school, *hs*, which is equal to 1 if either head, male or female, completed high school or the equivalent. We additionally add another dummy, *both_hs*, to indicate whether both heads of the household completed high school. We added similar dummies for college educated households.

Community (Period) Variables

In our analysis we include community variables which we believe to influence FAH expenditure patterns. We include per capita violent crime rates in our estimation to serve as a proxy for community disamenities or disadvantages that could influence FAH purchases. High violent crime rates may dissuade large supermarkets-- which offer more healthy and affordable food items-- from locating in these areas. In fact, wealthy neighborhoods have more than four times as many large grocery stores than in poor neighborhoods (Morland et al., 2002).

⁹ The average mean household income in our sample is \$61,905. This is consistent with the BLS' Consumer Expenditure Survey, which reports that the average American's before tax income in 2012 was \$65,596.

Crime data was obtained from the 2012 National Crime Victimization Survey maintained by the Bureau of Justice Statistics. We found crime rates with various levels of geographical specificity. Per capita crime rates were calculated and scaled up for every 100,000 people by county, metropolitan statistical area, and state. We used the most localized measure of crime-- county being the most specific-- for each observation when available.

Obesity rates are much higher among black populations compared to any other racial group. It has also been established that an inverse relationship exists between socioeconomic status (SES) and obesity prevalence (McLaren, 2007). We include a variable indicating concentration of black populations because it may be also be a proxy for community disadvantage. Specifically, blacks are more likely to have low SES and tend to live in low-income neighborhoods. In fact, Jargowsky (1997) found that even when controlling for individual SES, blacks are still more likely to live in economically disadvantaged neighborhoods.

Diez-Roux et al. (1999) find that individuals living in low SES communities tend to have poorer eating habits. A contributing factor may be that these communities-- often with high black populations-- generally have less access to supermarkets, which impact healthy food purchasing patterns. Research also shows that Blacks are more responsive to improvements in neighborhood amenities. Morland et al. (2002) found that marginal fruit and vegetable consumption increased 32 percent with every additional supermarket. And, this was not the case in wealthy, predominantly white neighborhoods. However, this is likely due to a high saturation of supermarkets in these communities.

We include a variable indicating the rate of participation in food assistance programs. The percentage of households participating in food assistance also gives us a measure of community disadvantage and food insecurity. Evidence shows that Supplemental Nutrition

Assistance Program (SNAP) participants have lower diet quality than non-eligible individuals (Condone et. al., 2015).

Black population rates and SNAP participation rates were acquired from the Integrated Public Use Microdata Series (IPUMS) for 2012. IPUMS is microdata collected from census, the American Community Survey and the Current Population Survey. These data are maintained through the Minnesota Population Center at the University of Minnesota.

Propensity Score Matching

Volpe and Okrent (2012) find small differences in diet quality across demographic groups, particularly in regards to income. Similarly, diet quality was found to be very comparable across racial groups where the dietary quality of Whites and Asians were only slightly better than Blacks and other races. This however, does not reconcile with the variation of obesity rates by race where it is known that Blacks have far higher rates of obesity than any other race (Ogden et. al., 2014). Volpe and Okrent (2012) portend that this may be due to differences in FAFH purchases, which is not collected in their dataset. While this is a highly reasonable assumption, it is in need of further investigation. Differences in diet quality based solely on FAFH purchases may potentially be underestimated because we do not control for all economic and demographic conditions that impact diet quality.

To do so, we employ propensity score matching (PSM) techniques in a nonconventional way. PSM is typically used in the advent of suspected selection bias. That is, there is non-random sorting of individuals into treated and non-treated groups. However, since our “treatment” variable is the millennial indicator-- and birth year is not an individual choice-- there is no inherent selection bias. Rather, we use PSM as a sort of robustness check using it to trim the dataset to control for community exogenous factors such that in the primary regression we

are able to extrapolate a millennial effect using information only from non-millennial and millennial individuals with like backgrounds. We take advantage of the conditional independence assumption since we know a priori, that birth year is independent of the other regressors in the treatment assignment function. What results is that we directly control for factors which contribute to the period effect in the APC model. Though we cannot parse out the cohort and age effect, as mentioned earlier, this does give us an opportunity to directly control one of the effects, which solves the identifiability problem.¹⁰ The auxiliary function which determines the treatment assignment is as follows,

$$(6) I_{Mil} = \begin{cases} 1 & \text{if } \mathbf{z}'\boldsymbol{\rho} + \nu > 0 \\ \text{otherwise} & \end{cases}$$

where \mathbf{z} is a vector of period effect variables that are community specific, $\boldsymbol{\rho}$ is a vector of coefficients and ν is the disturbance term. Note that \mathbf{z} does not include age despite the fact that millennial status is deterministic of birth year. This is attributable to the fact that we want to maintain that I_{Mil} is independent of \mathbf{z} . What results from the regression analog of (6) are propensity scores obtained from the following probit model,

$$(7) Pr(I_{Mil} = 1|z) = \Phi(\mathbf{z}'\boldsymbol{\rho})$$

Where Pr (propensity score) is the probability of being a millennial conditional on period/community related variables and Φ is the CDF of the standard normal distribution. After

¹⁰ Our future plans are to use this methodology along with a panel estimation to both segregate the cohort and age effects and solve the identifiability problem.

matching, we are left with two groups: millennials and non-millennials, where both groups share similar distributions.¹¹ It is from this subsample that we estimate (1) using FGLS.

Results

Across all three regressions, the millennial effect had a small and positive effect on diet quality, but was only statistically significant when *CNPPScore1* is used as the dependent variable.

Shopper_sex indicates whether the shopper was female. The results show a consistent, but small and significant negative effect when the shopper is female. Similarly, nonwhite shoppers seem to buy foods that are slightly lower in diet quality than their white shopper counterparts. These results were also highly significant across all three regressions. The per capita violent crime rate also shows a very small, but highly significant negative effect on diet quality and this is true for all regressions. Perhaps unexpectedly, the concentration of grocery stores within a mile of the shopper's residence had a negative effect on diet quality. Marriage had one of the highest effects on diet quality relatively speaking, where married individuals purchase foods with diet quality just over 0.4 units on average more than non-married shoppers. The marriage coefficients were significant across all regressions. Unsurprisingly, monthly food expenditure was significant and positively associated with diet quality, but this effect was very small. For example, the regression using *CNPPScore2* which has the largest coefficient shows that a \$100 increases in food expenditure per month would improve diet quality by only 1.2 units. Household income per capita had a small and positive effect on diet quality, but was only significant for regressions (2)

¹¹ Since we use pooled cross-section data, thus each household has multiple observations we restrict the matching process to one observation per household. We are able to do so because millennial status and the community variables are time invariant.

and (3). Households with two employed heads of households show a significant and positive effect on diet quality. The effect was just over 0.10 units across all regressions. This was also the case for households with two college educated heads of households, though the effect was slightly higher ranging from an increase of 0.143 to 0.187 in diet quality on average. Much like the coefficients from grocery store concentration, distance from grocery stores shows a positive effect on diet quality. The effects are significant and small across all regressions. Since the furthest recorded distance to a grocery store in this sample is 31.1 miles, this household would have a 1.97 units increase in diet quality. Finally, the number of trips to the grocery store had a positive and significant effect on diet quality regardless of the dependent variable, increasing diet quality anywhere from 0.275 to 0.378 units on average.

Discussion

The millennial indicator is our main variable of interest. Table 2 shows that when diet quality is measured using the *CNPPScore1* formula, millennials are healthier than older cohorts albeit, the difference is small, but significant. This is consistent with previous research which has found that diet quality has remained essentially unchanged for the past 30 years (Gregory et. al., 2014). From (1), *CNPPScore1* is the inverse of the deviation from the recommended expenditure share squared, thus it does not make a distinction between overconsumption or underconsumption. Rather it is a measure of “compliance” to the recommended shares. From the results, it would appear that millennials conform to the recommended shares more than older cohorts, but this does not necessarily mean that millennials are purchasing healthier foods. Because of noise build into the calculation of *CNPPScore1*, it may mask health seeking behavior and the penalty

is more pronounced the further away from the recommended purchases regardless of direction. Specifically, individuals who purchase more than the recommended healthy foods are penalized equally as those individuals who under-purchase the same foods with the same magnitude and this is also the case for underconsumption of unhealthy foods. *CNPPScore2* corrects for zero consumption of unhealthy foods by excluding them in the calculation, however, zero consumption of healthy foods are also excluded. The total effect is that *CNPPScore2* superficially increases the individual's score. Because *CNPPScore2* does not definitively correct for the noise in *CNPPScore1*, it maybe be the case that older cohorts are under purchasing unhealthy foods, which would correctly increase their scores or that other cohorts are not consuming certain healthy foods, incorrectly inflating their score effectively equalizing millennial and non-millennial scores. To further decipher these effects, *CNPPScore3* corrects for underconsumption of healthy foods. However, despite this correction there is little difference in the magnitude of the coefficient on the millennial variable and it still remains insignificant. When the regression results are interpreted together, a possible explanation is that older cohorts are under eating unhealthy foods relative to millennials enough to compensate for under consumption of healthy foods, effectively implying that there is no statistical difference between the diet quality of millennials and non-millennials.¹²

In general, all the coefficients are small suggesting that diet quality generally remains consistent across demographic groups. Some of the notable results show that women shoppers purchased foods with lower diet quality than male shoppers. This is consistent with the fact that overweight and obesity is more pervasive among women than men (Ogden et. al., 2014).

¹² To address measurement issues, our future research efforts will utilize the Health Eating Index (HEI) to measure diet quality to, which we are unable to calculate using the IRI data.

Additionally, since women are still considered to be the predominant household food provider, they may be shopping for other members of their household, particularly children, thus buying unhealthier foods to appeal to childhood tastes. Male shoppers may be more likely to be shopping for themselves or for adult-only households and may be less compelled to purchase unhealthy foods for other members of their household. Likewise, nonwhites who also suffer from larger rates of overweight and obesity purchased lower quality foods than whites.

Nonwhites may be more likely to live in urban areas where because of land constraints, have less access to large grocery stores which is the primary source of healthy food purchases. While the per capita crime rate effect was negative and significant, the effect was very small, suggesting that crime is a small deterrent for healthy food purchases. Since we use the crime rate essentially as a proxy for regional disamenities the results would suggest that poor community infrastructure has little effect on diet quality. This is consistent with recent research which has shown that prevalence of overweight and obesity eventually infiltrates into groups with higher SES, equalizing obesity rates across SES groups (Zhang and Wang, 2004; Chang and Lauderdale, 2005; Jolliffe, 2011). And, in fact our other results support this as regional food stamp participation, percentage of black population and per capita income have small and mostly insignificant effects on diet quality.

Following other research, we observe a positive effect on diet quality when at least one head of the household had a college degree, but there is no additional positive effect if both household heads had earned a college degree among households with two household heads. Additionally, when both household heads were employed, there was also an increase in diet quality of food purchases. This may lay credence to the idea that more educated, employed households have both the means and nutritional literacy to make better food choices.

Unsurprisingly, the number of trips to food stores was positively correlated to diet quality. This may be due to the fact that more trips to the grocery store may indicate that the household is purchasing a higher number of perishable goods, which typically are more nutritious than shelf-stable foods. Additionally, perishable foods may also cost more, which is why we observe a positive relationship with monthly grocery expenditure and diet quality.

Lastly, we found a surprising negative result on the effect of distance to food stores. Our results indicate that households farther away from food stores purchased healthier foods than those closer. One possible reason for this effect is that our food store variable includes all types of food stores including convenience stores. Since this is the case, households that are closer to convenience stores may be more likely to purchase unhealthy foods, while those households which are not located in areas with high concentration of convenience stores (typically suburban settings) maybe more likely to do the majority of their food purchases at large grocery stores which have more variety of healthy foods.

References

- Biing-Hwan, L. (2005, February). *Nutrition and Health Characteristics of Low-Income Populations* (United States of America, United States Department of Agriculture, Economic Research Service). Retrieved March 25, 2016.
- Caliendo, Marco, and Sabine Kopeinig. "Some practical guidance for the implementation of propensity score matching." *Journal of Economic Surveys* 22, no. 1 (2008): 31-72.
- Chang, Virginia W., and Diane S. Lauderdale. "Income disparities in body mass index and obesity in the United States, 1971-2002." *Archives of internal medicine* 165, no. 18 (2005): 2122-2128.
- Condon, Elizabeth, Susan Drilea, Keri Jowers, Carolyn Lichtenstein, James Mabli, Emily Madden, and Katherine Niland. (2015). Diet Quality of Americans by SNAP Participation Status: Data from the National Health and Nutrition Examination Survey, 2007–2010. Prepared by Walter R. McDonald & Associates, Inc. and Mathematica Policy Research for the Food and Nutrition Service.
- Council of Economic Advisors. (October 2014). *15 Economic Facts about Millennials*. Retrieved October, 30, 2015.
- Diez-Roux, Ana V., F. Javier Nieto, Laura Caulfield, Hermann A. Tyroler, Robert L. Watson, and Moyses Szklo. "Neighbourhood differences in diet: the Atherosclerosis Risk in Communities (ARIC) Study." *Journal of Epidemiology and Community Health* 53, no. 1 (1999): 55-63.
- Fienberg, Stephen E., and William M. Mason. *Specification and implementation of age, period and cohort models*. Springer New York, 1985.
- Gregory, Christian A., Travis A. Smith, and Minh Wendt. *How Americans rate their diet quality: an increasingly realistic perspective*. Economic Research Service, US Department of Agriculture, 2011.a
- Gregory, Christian, Ilya M. Rahkovsky, and Tobenna Anekwe. "Consumers' use of nutrition information when eating out." *USDA-ERS Economic Information Bulletin* 127 (2014).
- Hazel A.B. Hiza, Kellie O. Casavale, Patricia M. Guenther, Carole A. Davis, Diet Quality of Americans Differs by Age, Sex, Race/Ethnicity, Income, and Education Level, *Journal of the Academy of Nutrition and Dietetics*, Volume 113, Issue 2, February 2013, Pages 297-306, ISSN 2212-2672, <http://dx.doi.org/10.1016/j.jand.2012.08.011>.
- Hobcraft, John, Jane Menken, and Samuel Preston. *Age, period, and cohort effects in demography: a review*. Springer New York, 1985.
- Jargowsky, Paul A. *Poverty and Place: Ghettos, Barrios, and the American City: Ghettos, Barrios, and the American City*. Russell Sage Foundation, 1997.
- Jolliffe, Dean. "Overweight and poor? On the relationship between income and the body mass index." *Economics & Human Biology* 9, no. 4 (2011): 342-355.

Kennedy, E., Bowman, S. A., Lino, M., Gerrior, S. A., & Basiotis, P. P. (1999). Diet quality of Americans: Healthy eating index. *America's eating habits: Changes and consequences*, 97-110.

McLaren, Lindsay. "Socioeconomic status and obesity." *Epidemiologic reviews* 29, no. 1 (2007): 29-48.

Mancino, L., and J.D. Kinsey. 2008. Is Dietary Knowledge Enough? Hunger, Stress, and Other Roadblocks to Healthy Eating, ERR-62, U.S. Department of Agriculture, Economic Research Service, Aug.

Mancino, L., J.E. Todd, J. Guthrie, and B. Lin. 2010. How Food Away From Home Affects Children's Diet Quality, ERR-104, U.S. Department of Agriculture, Economic Research Service, Oct.

Mason, William M., and Nicholas H. Wolfinger. "Cohort analysis." *California Center for Population Research* (2001).

Morland, Kimberly, Steve Wing, Ana Diez Roux, and Charles Poole. "Neighborhood characteristics associated with the location of food stores and food service places." *American Journal of Preventive Medicine* 22, no. 1 (2002): 23-29.

O'Brien, Robert. *Age-period-cohort models: Approaches and analyses with aggregate data*. CRC Press, 2014.

Ogden CL, Carroll MD, Kit BK, Flegal KM. Prevalence of Childhood and Adult Obesity in the United States, 2011-2012. *JAMA*.2014;311(8):806-814. doi:10.1001/jama.2014.732.

Okrent, Abigail M., and Julian M. Alston. "The effects of farm commodity and retail food policies on obesity and economic welfare in the United States." *American Journal of Agricultural Economics* 94, no. 3 (2012): 611-646.

Stata, A. "STATA Treatment-Effects Reference Manual: Potential Outcomes/Counterfactual Outcomes Release 13." (2013).

Todd, Jessica E. *Changes in Eating Patterns and Diet Quality Among Working-Age Adults, 2005-10*, ERR-161. U.S. Department of Agriculture, Economic Research Service, January 2014

Todd, J.E., L. Mancino, and B. Lin. 2010. The Impact of Food Away from Home on Adult Diet Quality, ERR-90. U.S. Department of Agriculture, Economic Research Service, Feb.

Tuttle, Brad. "10 Things Millennials Buy Far More Often Than Everyone Else." *Money*. July 31, 2015. Accessed March 25, 2016. <http://time.com/money/3979425/millennials-consumers-boomers-gen-x/>.

United States Bureau of Labor Statistics. Consumer Expenditure Survey (2014). <http://www.bls.gov/cex/#tables>.

Volpe, R.J., and A. Okrent. 2012. "Assessing the Healthfulness of Consumers' Grocery Purchases." *EIB-102*, U.S. Department of Agriculture, Economic Research Service, November 2012.

Zhang, Qi, and Youfa Wang. "Trends in the association between obesity and socioeconomic status in US adults: 1971 to 2000." *Obesity Research* 12, no. 10 (2004): 1622-1632.

Zhen, Chen, Eric A. Finkelstein, James M. Nonnemaker, Shawn A. Karns, and Jessica E. Todd. "Predicting the effects of sugar-sweetened beverage taxes on food and beverage demand in a large demand system." *American Journal of Agricultural Economics* (2013): aat049.

Appendix A

Table 1. Summary Statistics, Entire versus Random Weight Sample

Entire Panel: N = 1,005,233					Urban, Random Weight Panel: N=127,838			
Variable	Mean	Std Dev	Min	Max	Mean	Std Dev	Min	Max
Millennial	0.09392	0.29172	0	1	0.15251	0.35951	0	1
shopper_sex	0.63793	0.48060	0	1	0.65106	0.47664	0	1
hhsiz	2.57616	1.33885	1	8	2.65572	1.36816	1	8
HHinc_per_capita	28228.97	14806.93	1249.9	59999	27029.17	14689.70	1249.9	59999
poverty	0.05125	0.22051	0	1	0.05499	0.22796	0	1
FullTime_M	0.22435	0.41716	0	1	0.17897	0.38333	0	1
FullTime_F	0.39316	0.48845	0	1	0.35684	0.47907	0	1
Male_Age	53.92747	13.75981	18	97	49.01951	13.72070	18	97
Female_Age	52.84049	13.78570	17	97	47.95295	13.74114	18	96
Mean_HHInc	63327.29	30982.56	9999	100000	61905.73	30664.42	9999	100000
Age	53.39093	13.83800	17	97	48.56569	13.84633	18	96
Nonwhite	0.20977	0.40714	0	1	0.22104	0.41495	0	1
Married	0.66690	0.47132	0	1	0.64166	0.47952	0	1
F_Employed	0.51745	0.49970	0	1	0.55020	0.49748	0	1
M_Employed	0.52922	0.49915	0	1	0.56550	0.49569	0	1
trips	3.94411	2.60950	1	32	2.59564	1.60567	1	20
dollarspaid_mt	61.60134	64.91562	0.01	1712.2	27.96934	24.80300	0.09	450
Mean_PPO1	0.19207	0.12591	0.00046	1.5	0.17815	0.11651	0.00150	1.5
Mean_PPO2	0.23946	0.14339	0.00067	1.5	0.23069	0.15145	0.01188	1.49875
Mean_PPO3	0.18561	0.13547	0.00305	1.49833	0.17643	0.14586	0.00773	1.49833

Mean_PPO4	0.18057	0.11572	0.00036	1.49833	0.17511	0.12723	0.00594	1.4975
Mean_PPO5	0.30108	0.26396	0.00500	1.5	0.26075	0.28147	0.00781	1.5
Mean_PPO6	0.19383	0.14355	0.00042	1.5	0.18656	0.15156	0.00869	1.5
Mean_PPO7	0.20665	0.12926	0.01160	1.5	0.19542	0.13980	0.01160	1.49875
CNPPScore1	6.31114	2.53189	0.89786	17.19565	5.90686	2.32395	0.89786	17.15615
CNPPScore2	8.31221	3.67442	0.98534	23.38197	7.95370	3.53262	1	23.35480
CNPPScore3	6.43270	2.65069	0.91459	17.48693	5.98915	2.40254	0.91459	17.45589
Distance	0.93190	0.79693	0.00320	31.14266	0.93275	0.80505	0.00800	31.14266
Ratio_Grocery_All	0.17114	0.17317	0	1	0.16810	0.17149	0	1
Num_Grocery_1mi	2.37992	6.46780	0	124	2.26480	6.18485	0	114
Num_All_1mi	10.23140	17.49145	0	283	9.94947	16.43337	0	268
NoHS_M	0.03671	0.18806	0	1	0.02781	0.16443	0	1
NoHS_F	0.02183	0.14614	0	1	0.01647	0.12729	0	1
HS_M	0.40446	0.49079	0	1	0.40109	0.49012	0	1
HS_F	0.49605	0.49998	0	1	0.47603	0.49943	0	1
College_M	0.21759	0.41261	0	1	0.23134	0.42169	0	1
College_F	0.28655	0.45215	0	1	0.31416	0.46418	0	1
GradSchool_M	0.09090	0.28746	0	1	0.08318	0.27615	0	1
GradSchool_F	0.10359	0.30473	0	1	0.10230	0.30304	0	1
Uilocal					7.81305	1.51108	3.9	12.5

Table 2. FGLS regression results using post-match subsample

	(1)	(2)	(3)
VARIABLES	cnppscore1	cnppscore2	cnppscore3
millennial	0.131***	0.0411	0.0482

	(0.0468)	(0.0658)	(0.0450)
shopper_sex	-0.209***	-0.277***	-0.258***
	(0.0419)	(0.0610)	(0.0417)
nonwhite	-0.145***	-0.223***	-0.0708*
	(0.0393)	(0.0558)	(0.0395)
black_rt	0.120	0.250	0.200
	(0.130)	(0.183)	(0.128)
violentcrimetotal_pc	-0.000582***	-0.000758***	-0.000788***
	(0.000173)	(0.000258)	(0.000173)
foodstmp_rt	0.169	0.185	0.297
	(0.246)	(0.353)	(0.243)
num_grocery_1mi	-0.00476**	-0.0119***	-0.00672***
	(0.00208)	(0.00304)	(0.00217)
married	0.438***	0.413***	0.403***
	(0.0450)	(0.0661)	(0.0455)
dollarspaid_mt	0.00799***	0.0120***	0.00926***
	(0.000524)	(0.000774)	(0.000537)
hhinc_per_capita	0.00188	0.00292*	0.00473***
	(0.00120)	(0.00171)	(0.00121)
employed	-0.0213	0.00987	-0.0244
	(0.0463)	(0.0668)	(0.0468)
both_employed	0.126***	0.114**	0.100**
	(0.0397)	(0.0569)	(0.0391)
hs	-0.0503	0.0517	-0.0659
	(0.0475)	(0.0676)	(0.0481)

both_hs	0.0481	0.179**	0.0777
	(0.0507)	(0.0747)	(0.0510)
college	0.168***	0.187***	0.143***
	(0.0472)	(0.0684)	(0.0476)
both_college	-0.00120	-0.0368	-0.0647
	(0.0667)	(0.0937)	(0.0659)
distance	0.0335*	0.0638**	0.0409**
	(0.0193)	(0.0289)	(0.0191)
trips	0.295***	0.378***	0.275***
	(0.0101)	(0.0147)	(0.0101)
Constant	4.697***	6.345***	4.794***
	(0.0896)	(0.132)	(0.0897)
Observations	48,624	48,624	48,624
R-squared	0.085	0.063	0.079

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1