The Pennsylvania State University

The Graduate School

THREE ESSAYS ON CONSUMER DEMAND, HEALTH AND FOOD ENVIRONMENT

A Dissertation in

Agricultural, Environmental, and Regional Economics

by

Zefeng Dong

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

December 2019

The dissertation of Zefeng Dong was reviewed and approved* by the following:

Edward C. Jaenicke
Professor of Agricultural Economics
Dissertation Co-Advisor
Chair of Committee
Chair of the Graduate Program

Yizao Liu
Assistant Professor of Agricultural Economics
Dissertation Co-Advisor

Douglas H. Wrenn
Associate Professor of Environmental and Resource Economic

Peter Newberry
Assistant Professor of Economics

*Signatures are on file in the Graduate School

ABSTRACT

This dissertation explores the relationship between consumers' dietary choice, health and healthy behaviors and food environment from three different angles.

Essay 1 investigates the links of dietary choice with physical activity, obesity, type 2 diabetes mellitus and medication usages for these two diseases. This study uses IRI Consumer panel and Medprofiler panel from 2013 to 2017, where only single-member households with valid demographic information over all 5 years are used. I particularly investigate six food categories, fruits and vegetables, snacks and chips, yogurts, regular soft drinks, diet soft drinks and bottled water, and employ yearly expenditure shares of these food categories to denote consumers' dietary choice. The econometric models for analysis are Ordinary Least Square (OLS) models of expenditure share of one food category against physical activity, obesity and medication usage for obesity or type 2 diabetes and medication usage for type 2 diabetes as well as demographics and fixed effects for variation control. Essay 1 finds that the individuals who do exercise most days in a week spend more of their grocery budget on healthy food categories such as fruits and vegetables and yogurts and spend less on unhealthy food categories such as snacks and chips, regular soft drinks and diet soft drinks on average than the individuals who do exercise some days in a week or who rarely or never exercise. Further, Essay 1 finds that the dietary pattern of obese individuals is less healthy on average than that of nonobese individuals. However, when considers the usage of medications for obesity and a mixed method for obesity identification (i.e. identify obesity with self-reported obesity and BMI ≥ 30), the results are mixed. For results

regarding type 2 diabetes, the expenditure shares on food categories are driven by both medication usage and nutrition facts of the food category.

Developed from Essay 1 but instead of using a reduced-form analysis to provide a big picture of the link between diet, physical activity and health, Essay 2 focuses on one product, yogurts, and employ structural demand model to investigate the association between consumer demand on yogurts and physical activity and obesity. Data used for Essay 2 are the IRI Consumer panel and Medprofiler panel in 2013. I consider a mixed multinomial logit model with random coefficients for price and health-related product attributes. Physical activity or obesity is incorporated in the model by the interaction terms of health-related attributes and physical activity or obesity, respectively. Essay 2 finds that individuals who do exercise some days in a week are the most price sensitive on average, followed by the individuals who do exercise most days in a week and the ones who rarely or never exercise. Physically active individuals, on average, prefer healthy yogurts such as plain yogurts and Greek yogurts. Furthermore, Essay 2 finds that the individuals whose BMI $\geq$ 30 (i.e. obese or overweight) are more price sensitive and prefer yogurts with more sugar and protein and less total fat on average than the ones whose BMI $<$ 30. However, when I use self-reported obesity to identify obesity, the regression result regarding price sensitivity is opposite to that using BMI to identify obesity.

In Essay 3, we consider migration as an identification strategy to control the endogeneity of food environment. A first-difference model is employed to examine the association between the change in diet quality and the change in food environments for

movers over the year. This study finds that in some years the improvement in diet quality responds to the change in food environments as a result of migration, but this significant association does not consistently exist in every year. This essay demonstrates that the policy aiming to improve people's diet by improving food environment in neighborhood likely fails.

TABLE OF CONTENTS

LIST OF TABLES

ACKNOWLEDGEMENTS

First, I would like to thank my advisor, Professor Edward Jaenicke, for his support and guidance in my research, study and career development during the past four years. In these days working together with him, I learned how to be a good and honest economist and educator, how to perform economic analysis from the beginning to the end and how to diagnose potential problems in research. I would also like to thank my co-chair, Professor Yizao Liu, for her patient mentoring of how to appropriately approach a research project and how to identify a good research method. Furthermore, I would like to thank my dissertation committee members, Professor Douglas Wrenn and Professor Peter Newberry, for their valuable comments about my dissertation.

Second, I would like to thank my close friends, Xuan Chen, Bo Zhang and Xiaogu Li, as well as other remarkable colleagues in our department, for their help in my research and coursework. Having this great opportunity to know and work together with a cohort of outstanding people lets me know what is excellence and what makes excellence.

Finally, I would like to thank my family, my parents for their support in my 27-year lifetime and my wife, Yixuan, for her accompany and support.

Chapter 1

Introduction

Research on link between food and health is abundant for two big reasons: (a) the diet and health discrepancies between recommended and actual diets and across socioeconomic status are longstanding, persistent problems, and (b) the economic consequences of these discrepancies are substantial. It is a positive sign that the overall diet quality of Americans from 1999-2000 to 2011-2012 improved (Wilson et al., 2016), and a variety of public policies have been implemented aiming to facilitate the intake of healthy foods, lower the sales of soda and provide better food environment in neighborhood (Cawley, 2015). However, academic studies and federal reports show that some critical nationwide health problems stand. The overall consumption of healthy food is below the recommended level of the 2010 Dietary Guidelines for Americans (Wilson et al., 2016). As a result of food insecurity, about 41 million Americans are living with unhealthy dietary patterns, where the probability of health problems such as diabetes and cardiovascular diseases among these individuals are above average (Seligman et al., 2017; Coleman-Jensen et al., 2017). Meanwhile, the obesity rates for both American adults and youths have been increasing in the past 15 years (Hales et al., 2017), where obesity costs every American about $2741 for medical expenses in 2005 (Cawley & Meyerhoefer, 2012). These health-related problems in the U.S., its economic impact and its links with grocery markets motivate me to conduct research and work on this dissertation.

The backbone of this dissertation is an investigation into the links between consumer behavior towards grocery products and health and health-related behaviors. The relationship between health and food, which is affected by household-level socioeconomic constraints, external environmental factors, and individual behavioral tendencies, is so complex that it might be better to approach this question from different angles (Bleich et al., 2015). The three essays that comprise my dissertation all investigate food behavior and health, but do so from two different angles. The first two essays investigate the role of individuals' health and healthy behaviors on their dietary choice, where the first essay is a reduced-form analysis that investigates the association of dietary choice with physical activity, obesity, type 2 diabetes mellitus and medication usages for obesity and type 2 diabetes and the second essay is a structural model demand analysis that investigates the link between consumer demand on yogurts and physical activity and obesity. The third and last essay investigates how households' diet quality might be affected by their neighborhood food environments. More specifically, this essay relies on a shock to the food environment (i.e., migration) to appropriately identify the food environment's impact on diet quality.

All three essays utilize some version of food-purchase scanner data for U.S. households. Essay 1 and 2 uses Consumer Panel and Medprofiler Panel from Information Resources, Inc. (IRI) obtained via a collaborative agreement with USDA's Economic research service; and Essay 3 again uses the IRI Consumer Panel and Retailer Panel. Collectively, all datasets contain comprehensive information on household-level purchases record, detailed product characteristics and household demographics that allow us to track the consumer purchases of the same group of people over multiple years across states. The IRI Medprofiler Panel does not only provide yearly health-related survey data that consist of individual-level medical demographics and lifestyle information, but also allows us to match up with the consumer data and demographic information and investigate the link between health and diet.

While the focus of the three essays are different, there are several commonalities in the data and variables of interest. One essential and common question in three essays is how to define the healthfulness of food purchases or how to evaluate diet quality. The first method is to select some representative food categories that are considered as healthy or unhealthy foods based on literature or the nature of the foods. Literature suggests that fruits and vegetables and bottled water are healthy foods and snacks and chips, regular and diet soft drinks are unhealthy foods. Therefore, I employ the expenditure share of one of these healthy or unhealthy food categories as a proxy of diet quality in Essay 1 and 3. In Essay 1, I select six food categories, including three healthy food categories, fruits and vegetables, yogurts and bottled water, and three unhealthy food categories, snacks and chips, regular soft drinks and diet soft drinks. In Essay 3, the expenditure share of fruits and vegetables is employed as a measure of diet quality, mainly for robustness check as an alternative measure to other diet quality indices. The second method of diet quality measure is relying on the amount of key nutrients per unit. In Essay 2, I consider yogurts, one food category but with good heterogeneity in health-related product attributes, and interest in consumers’ responses to the variation of healthfulness over alternatives. I select calorie, sugar, calcium, protein and total fat levels as proxies for the healthfulness of a yogurt product. The third method for diet quality evaluation is to use employ a diet quality index. In Essay 3, due to our limit access to comprehensive nutrition data of products at UPC level, I adopt an approach to measure diet quality of a household or individual without the need of nutrition data, called "USDA Score". It is a method to evaluate diet quality based on household expenditures on various food categories that requires a comprehensive account of all grocery products purchased over an appropriate period of time. It compares recommended versus actual expenditure shares across 24 general food categories. The IRI Consumer data provide detailed product characteristics for most purchased products at Universal Product Code (UPC) level, which allows me to categorize each purchased product into one of the food groups.

One other important set of variables used in the Essay 1 and 2 is individual-level health demographics and health-related behaviors. I use IRI Medprofiler individual-level information regarding obesity status, type II diabetes status, body mass index (BMI), and frequency of doing physical activity to represent household health behavior. The IRI Medprofiler data collect health information for more than 49,000 households based on questions about each household members' health and lifestyle.

Thus, each of the three essays in this proposed dissertation are empirical investigations that focus on micro-level dietary choices related to health.

Essay 1 employs the IRI Consumer Panel and Medprofiler Panel from 2013 to 2017 to investigate the associations of dietary choice with physical activity, obesity, type 2 diabetes mellitus and medication usage. I employ the annual expenditure shares of six food categories, fruits and vegetables, snacks and chips, yogurts, regular soft drinks, diet soft drinks and bottled water, to measure consumers' dietary choice.

In the first section of Essay 1, I employ Ordinary Least Square (OLS) models to estimate the link between expenditure share of a food category and the frequency of physical activity.

In the second section, I employ OLS models to estimate the link of expenditure share of a food category with obesity, BMI and medication usage for obesity. The obesity is self-reported information collected via a question in the IRI Medprofiler survey. In the same question, participants were also asked for whether they are taking any Rx, OTC or dual medications for obesity. Besides self-reported obesity, I also employ BMI $\geq$ 30 as a proxy to identify obesity as suggested by CDC (April 11, 2017), where BMI is calculated based on self-reported weight and height documented in IRI Medprofiler Panel. Similar exercise is performed for the analysis of dietary choice and type 2 diabetes mellitus, but this model is less complex because I only rely on the self-reported type 2 diabetes and medication usage for identification and do not consider any other proxy.

In this study, I find that physically active individuals spend more of their grocery budget on fruits and vegetables and yogurts and spend less on snacks and chips, regular and diet soft drinks on average than the inactive ones. The disparity in expenditure share on bottled water between physically active and inactive individuals is minimal. These results imply that physically active individuals are likely to have a healthier dietary pattern than inactive one.

Regarding the regressions of obesity and medication usage, I find that obese consumers spend less of their grocery budget on fruits and vegetables and yogurts and spend more on snacks and chips and diet soft drinks on average than non-obese ones. No statistically significant association between obesity and expenditure shares on regular soft drinks and bottled water is found. When I consider a mixed method of obesity identification and medication usage, I find that for the individuals who have BMI ≥ 30 and are self-reported obese but not a medication user or have BMI ≥ 30 but self-reported healthy, they tend to spend less on fruits and vegetables and yogurts and more on snacks and chips and diet soft drinks than the individuals who are healthy in term of BMI and self-reported obesity (i.e. the baseline case) on average. The individuals who are self-reported taking medication for obesity and have BMI ≥ 30 behave similar with healthy individuals towards most food categories except yogurts and diet soft drinks. The disparity in expenditure shares of yogurts, regular soft drinks, diet soft drinks and bottled water between individuals whose BMI < 30 and healthy individuals is minimal. As a conclusion, when only one identification strategy for obesity is considered, the regression results are straightforward, obese individuals having poor diet quality on average. However, when considers a mixed method for obesity identification, using both self-reported obesity and BMI together, the results are mixed and vary over food categories.

For the results of type 2 diabetes mellitus, I find that individuals who have been diagnosed with type 2 diabetes mellitus and are taking Rx, OTC or dual diabetic medications spend less of grocery budget on fruits and vegetables, yogurts and regular soft drinks and spend more on snacks and chips and diet soft drinks than the healthy individuals on average. On one side, these individuals who are taking diabetic medications maintain an unhealthy dietary pattern, spending more on snacks and chips and less on fruits and vegetables. On the other side, they spend less on regular soft drinks. It might because they lower the consumption of sugar-intensive foods such as fruits, yogurts and regular soft drinks, trying to lower sugar intake, and consume more sugar-free alternatives such as diet soft drinks. The individuals who have type 2 diabetes but not taking any medication behave similar to healthy individuals in term of most of food categories investigated in Essay 1 except diet soft drinks. Their type 2 diabetes might still under control but they are actively seeking an improvement in dietary pattern by consuming more sugar-free diet soft drinks.

Essay 2 estimates consumer demand on yogurts and its association with physical activity and obesity. In this study, I select yogurt as the product because it has good heterogeneity in health-related product attributes over alternatives, its healthfulness can be measured with accessible nutrition information labelled on nutrition fact panel such as the amount of sugar, calorie and total fat and it has been widely documented in literature. I employ mixed logit models where the coefficients of price and product characteristics are assumed to be random coefficients following normal distributions. Mixed logit model allows us to count the variation of consumers' taste (Train, 2009). In the model, I incorporate interaction terms of health-related product attributes and physical activity or obesity to capture the disparity in consumers' response to price and healthfulness of yogurts between physical active and inactive individuals or between obese and nonobese individuals.

In the model of physical activity, I define two dichotomous variables for physical activity, where one denotes doing exercise most days in a week and another denotes doing exercise some days in a week, respectively. I find that the individuals who do exercise some days in a week are the most price sensitive on average, followed by the individuals who do exercise most days in a week and the individuals who rarely or never exercise. Furthermore, we also find physically active individual prefer yogurts with less sugar and calcium and more protein on average than the ones who rarely or never exercise.

For the identification strategy of obesity, I follow Essay 1 and employ two methods. The first method is to use the IRI Medprofiler self-reported obesity and medication usage. The second method is to use BMI ≥ 30 as suggested by CDC (April 11, 2017). Differed from Essay 1, I treat them as independent methods for obesity identification and do not use them together in one model. However, the regression results of these two methods are contradictory. The results of the models with self-reported obesity and medication usage show that the individuals who are taking medications for obesity are the most price sensitive on average, followed by the individuals who are self-reported obese and not taking any medications and the individuals who are healthy. They also show that the self-reported obese individuals prefer yogurts with less sugar and protein on average than the healthy ones, while the individuals who are taking medications prefer yogurts with more sugar and protein and less total fat on average than the healthy individuals. One the other side, when we use BMI ≥ 30 to identify obesity and do not consider medication usage, we find that individuals whose BMI ≥ 30 are more price sensitive and prefer yogurts with more sugar and protein and less total fat on average than the ones whose BMI < 30. The results with self-reported obesity are substantially similar with Wang et al. (2017), but they use soda, which is an unhealthy product, but I use yogurts, a relatively healthier product than soda.

Essay 3 employs the IRI Consumer Panel and Retailer Panel to explore the relationship between a household’s food environment and its diet quality. Abundant literature demonstrates the association between an individual’s or household’s diet quality with a variety of aspects of food environment such as numbers of supermarket or grocery stores in the neighborhood (Morland et al., 2002; Morland and Evenson, 2009; Rundle et al., 2009) and food prices or affordability of healthy foods to household (Minaker et al., 2013; Ghosh-Dastidar et al., 2014; Hillier et al., 2015). However, the results are mixed concerning the consistency of these associations and the strength of the effects. One of the biggest reasons for these mixed results is that most studies fail to account for potential endogeneity in the food environment. For example, the food environment may be associated with some unobserved neighborhood characteristics that might affect retailers' decision of entering a market. If so, then food-environment measures, and estimated impacts, may suffer from potential endogeneity.

One recent study, Allcott et al. (2019) address this potential endogeneity. In general, the idea of controlling this endogeneity relies on an exogenous shock to people's food environments. For example, exogenous household migration, or the unanticipated entry and exist of grocery stores in local market, may lead to exogenous changes in a household's food environment.

This third essay follows Allcott et al. (2019) to investigate how households' healthfulness of food purchases respond to an exogenous shock on food environment. The preferred model used in this essay is a first-difference model of movers across zip code. However, the results line with Allcott et al. (2019) and I do not find significant evidence that the link between food environments and diet quality exists consistently across cycle.

References

Allcott, H., Diamond, R., Dube, J-P., Handbury, J., Rahkovsky, I., & Schnell, M. (2019). Food Deserts and the Causes of Nutritional Inequality. *Quarterly Journal of Economics*, 134(4).

Bleich, S. N., Jones-Smith, J., Wolfson, J. A., Zhu, X., & Story, M. (2015). The complex relationship between diet and health. *Health Affairs*, 34(11), 1813-1820.

Cawley, J., & Meyerhoefer, C. (2012). The medical care costs of obesity: an instrumental variables approach. *Journal of Health Economics*, 31(1), 219-230.

Cawley, J. (2015). An economy of scales: A selective review of obesity's economic causes, consequences, and solutions. *Journal of Health Economics*, 43, 244-268.

Centers for Disease Control and Prevention (CDC). (2017, April 11). *Defining Adult Overweight and Obesity.* Retrieved on June 20, 2019 from https://www.cdc.gov/obesity/adult/defining.html.

Coleman-Jensen, A., Rabbitt, M. P., Gregory, C. A., & Singh, A. (2017). *Household Food Security in the United States in 2016*, ERR-237, U.S. Department of Agriculture, Economic Research Service.

Ghosh-Dastidar, B., Cohen, D., Hunter, G., Zenk, S. N., Huang, C., Beckman, R., & Dubowitz, T. (2014). Distance to store, food prices, and obesity in urban food deserts. *American Journal of Preventive Medicine*, 47(5), 587-595.

Hales, C. M., Carroll, M. D., Fryar, C. D., & Ogden, C. L. (2017). *Prevalence of obesity among adults and youth: United States, 2015-2016*. US Department of Health and

Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics.

Hillier, A., Smith, T., Cannuscio, C. C., Karpyn, A., & Glanz, K. (2015). A discrete choice approach to modeling food store access. *Environment and Planning B: Planning and Design*, 42(2), 263-278.

Minaker, L. M., Raine, K. D., Wild, T. C., Nykiforuk, C. I., Thompson, M. E., & Frank, L. D. (2013). Objective food environments and health outcomes. *American Journal of Preventive Medicine*, 45(3), 289-296.

Morland, K. B., & Evenson, K. R. (2009). Obesity prevalence and the local food environment. *Health & Place*, 15(2), 491-495.

Morland, K., Wing, S., Roux, A. D., & Poole, C. (2002). Neighborhood characteristics associated with the location of food stores and food service places. *American Journal of Preventive Medicine*, 22(1), 23-29.

Rundle, A., Neckerman, K. M., Freeman, L., Lovasi, G. S., Purciel, M., Quinn, J., ... & Weiss, C. (2009). Neighborhood food environment and walkability predict obesity in New York City. *Environmental Health Perspectives*, 117(3), 442.

Seligman, H. K., Berkowitz, S. A., & Basu, S. (2017, October 2). Why Paying for Nutrition Saves Money on Health Care? *The Time Magazine*. Retrieved on July 23, 2018 from http://time.com/4962475/nutrition-snap-health-care/.

Train, K. E. (2009). *Discrete Choice Methods with Simulation*. Cambridge university press.

Wang, E., Rojas, C., & Colantuoni, F. (2017). Heterogeneous Behavior, Obesity, and Storability in the Demand for Soft Drinks. *American Journal of Agricultural Economics*, 99(1), 18-33.

Wilson, M. M., Reedy, J., & Krebs-Smith, S. M. (2016). American diet quality: Where it is, where it is heading, and what it could be. *Journal of the Academy of Nutrition and Dietetics*, 116(2), 302-310.

Chapter 2

The Links between Dietary Choice and Physical Activity, Health and Medication Usage

Introduction

In the most recent decades, obesity has become a nation-wise public concern in the U.S. and about 40 percent of American adults are affected by obesity in 2015-2016. (Hales et al., 2017). The rise of this concern is not only because obesity associates with some diseases such as cardiovascular diseases, type 2 diabetes and cancer (Hubert et al., 1983; Weyer et al., 2001; Calle et al., 2003; Kahn et al., 2006; Van Gaal et al., 2006 Vucenik et al., 2012; Hales et al., 2017), but also because it incurs about $147 billion medical cost per year by 2008, where the medical cost for an obese individual is about 42 percent higher than an individual with normal weight (Finkelstein et al., 2009). Several weight-control strategies are introduced and widely promoted to public including healthy diet, physical activity and pharmacotherapy (Pi-Sunyer, et al., 1998). However, the effectiveness of different weight loss strategies varies (Skender et al. 1996; Sallis et al. 2009; Lin et al., 2013) and according to some federal reports, for the implementation of some weight-control strategies such as diet and exercise, the actual level did not meet the recommended level. The average healthy eating index increases over time from 49 in 1999-2000 to 59 in 2011-2012 (Wilson et al., 2016), but the excessive intakes of sodium and added sugar are prevalent in the recent two decades (Yang et al., 2014; Dong et al., 2017). There is an upward but flat trend in the percentage of American adults who meet the federal recommended level of aerobic physical activity in the 2008 Physical Activity Guidelines, but even at the peak, which is in 2018, only about 53 percent of American adults meet the recommended level (CDC, 2019, May 30a). These statistics show a concurrent upward trend of diet and exercise in U.S. population over time, but the link between these two weight-control strategies from an economic angle is not clear yet in literature. Therefore, this study attempts to explore how physical activity associates with consumers' willingness to spend on a healthy or unhealthy food category. Furthermore, we also interest in how diet relates with obesity and type 2 diabetes mellitus and the medication usages for these two diseases.

Some evidence in literature demonstrates that individuals are less likely to change preference on products or change overall diet quality in a short period of time, even under the influence of some external or internal factors such as migration and being diagnosed with type 2 diabetes (Bronnenberg et al., 2012; Oster, 2018; Allcott et al., 2019), but the association between consumer preference and diseases such as obesity is known (Wang et al., 2017). These diseases are relatively an exogenous factor for consumers making purchase decisions compared with physical activity because it is an objective fact and an individual characteristic at that moment. Physical activity is likely happening concurrently with the shopping behavior within a short period of time and therefore affects product choice. Chen et al. (2002) and Lin et al. (2013) find there is no significant correlation between exercise and the product prices of milk or between exercise and overall diet quality, respectively. However, it is unclear about how physical activity links with consumers' willingness to purchase on some particular products.

The second purpose of this study is to explore the links of dietary choice with obesity and type 2 diabetes and the disparity in dietary choice between medication users and non-users. About 27 million Americans diagnosed with type 2 diabetes mellitus, a disease that cause patients having abnormal high blood glycemic level and can cause heart disease, vision loss or kidney disease when not received appropriate medical treatments (CDC, 2019, May 30b). According to our IRI Medprofiler data, about 90 percent of type 2 diabetes patients are taking Rx, OTC or dual medications. However, for some patients not taking any medications, type 2 diabetes might not be fatal in short run as long as the patients manage their diet, blood sugar and physical activity well (CDC, 2018, April 24; CDC, 2019, March 4; CDC, 2019, March 20). On the other side, Obesity as a chronical disease affects patients gradually. Some people can be identified as overweight based on their body mass index (BMI) but might still unaware of having obesity. Prescription medications for obesity from doctors will be recommended only when healthy diet and regular exercise fail to control weight, and even patients are taking prescription medications, the importance of diet and exercise is still irreplaceable (Yanovski & Yanovski, 2014). Therefore, when and why an individual chooses to use or not use any medications for obesity are more difficult to specific than that for type 2 diabetes mellitus. Furthermore, obesity can be caused by a variety of reasons such as imbalance diet, physical inactivity or genetic reason (CDC, 2017, August 29). This disparity in motivation and necessity of medication usage between obesity and type 2 diabetes leads to potential differences in consumer preference and willingness to spend on healthy and unhealthy food products. This is another question this study explores.

To measure consumers' dietary choice, we select six representative food categories, fruits and vegetables (F&V), snacks and chips (S&C), yogurts, regular soft drinks (RSD), diet soft drinks (DSD) and bottled water and employ the annual expenditure share of one of these food categories to denote consumers' willingness to spend on healthy or unhealthy food.

This study employs IRI Consumer Panel and Medprofiler Panel from 2013 to 2017. The IRI Consumer Panel consists of about 70,000 American households' grocery purchase record at Universal Product Code (UPC) level for almost every store visit, UPC-level product characteristics and yearly household demographic information. The IRI Medprofiler Panel consists of individual-level medical demographics and lifestyle information on yearly basis for almost all household members of about half of the households in IRI Consumer Panel. We only consider single-member households who only have one adult member and who only fill in the IRI Medprofiler survey once in a year, in order to ensure that the person who purchased a product is the one who reports health and lifestyle information. To match the trip-level purchases information with the yearly Medprofiler survey, we aggregate household expenditure of each food category to yearly level and calculate the expenditure share of household yearly expenditure on a food category over total yearly expenditure on all grocery products.

The analysis section consists of two parts. For the first part, we perform Ordinary Least Square (OLS) regressions of the expenditure shares of six food categories against the frequency of physical activity and demographic variables. In the second part, we perform OLS models of the expenditure shares of food categories against obesity and medication usage for obesity or type 2 diabetes and medication usage for type 2 diabetes.

One method to identify obesity relies on the self-reported medical survey documented in IRI Medprofiler data, where participants themselves report whether they have obesity and whether they are taking any Rx, OTC or dual medications for obesity. The alternative method is to employ body mass index (BMI) as a proxy for obesity and define the individuals whose BMI larger than or equal to 30 as obese individuals and those whose BMI larger than or equal to 40 as morbidly obese individuals. This follows a standard released by CDC (April 11, 2017). The calculation of BMI is based on weight and height reported by the participants in the IRI Medprofiler. Therefore, the implementation of this method to identify obesity does not require individuals to have any knowledge regarding obesity and therefore can avoid participants' misspecification for obesity that might occur in the Medprofiler self-report survey. However, this method fails to distinguish lean mass from fat mass and might misspecify obesity for some muscular individuals (Buckhauser et al., 2009; Rippe, et al., 2012; Cawley, 2015). Due to the drawbacks of both methods for obesity identification, we consider a mixed method, which use both self-report obesity and BMI together as proxies. It allows us to observe the disparity in dietary choice between individuals who know they are obese and those who are obese based on BMI but do not admit they are obese together in one model. It also allows us to observe the impact of medication usage on willingness to spend on healthy and unhealthy foods.

In the following section, we first clarify our research objectives. In the third section, we conduct a literature review and discuss about the contribution of this study to literature. In the fourth section, we introduce the data used and the definition and summary statistics of important variables. The fifth section shows the OLS models used for estimation and key variables in each model. The sixth section presents the regression results and interpretation for results. Then we conduct robustness check where we split the sample by physical activity or BMI, regress the models with sub-samples and compare the sub-sample regression results with our main model results. In the last part, we discuss about the regression results, where we compare our results with literature, and concludes the study and present potential further work.

Research Objectives

The purpose of this study is to estimate the associations between the expenditure shares on food categories and physical activity, obesity and type 2 diabetes mellitus and medication usages for these two diseases with econometrics models.

We particularly interest in some food categories often considered as healthy foods such as fruits and vegetables, yogurts and bottled water and some foods considered as unhealthy such as snacks and chips, sugar and diet soft drinks. It allows us to observe the disparity in consumer preference on the healthfulness of foods between individuals by their health and healthy behaviors.

We firstly consider physical activity, a variable of healthy behaviors meant to capture individuals' willingness to make effort for weight loss and health improvement. We are interested in when individuals are regularly doing exercise, whether their diet quality, or says willingness to spend on healthy and unhealthy foods, differs from those physically inactive one. This part allows us to observe whether there is any link between diet and exercise, these two weight control strategies.

Second, this study explores the association of dietary choice with obesity and type 2 diabetes and medication usages for these two diseases. It allows us to understand whether there is any link between a disease and dietary choice, and most importantly whether there is a disparity in dietary choice between patients not taking medications, patients taking any medications and healthy individuals.

Literature Review

This study mainly contributes to the literature that links health and health-related behaviors with diet quality or product choice.

The existing literature about the link between physical activity and diet relies on first-hand survey, experimental data or national health survey data such as National Health and Nutrition Examination Survey (NHANES). These datasets provide reliable information about individual medical and health demographics, but they do not have reliable consumer purchase information over an appropriate period of time and over regions. Gillman et al. (2001) use cross-sectional survey data with 1,322 male participants and find that participants with higher intensity of physical activity are likely to consume less unhealthy meats, saturated fat, trans-fat, and dietary cholesterol. Lin et al (2013) employ the NHANES data to investigate the influence of weight management strategies on individuals' diet quality and find that physically activeness and regular consuming lower-calorie foods are associated with high diet quality.

Some studies consider healthy diet and physical activity as two independent healthy behaviors and investigate their socioeconomic determinants. Øvrum (2011) investigates the links between socioeconomic status and lifestyle choices (i.e. physical activity and consumption on fruits and vegetables) and demonstrates negative associations between physical activity and a set of proxies such as whether the participants often work overtime or frequently watching TV. Fan and Jin (2014) measure self-control capability with Totter score and suggest that lack of self-control associates with negative health behaviors such as poor eating and exercise, and incidence of obesity. They also find more fruits and vegetables intake is easier for obese or overweight consumers than doing intensive physical activity. Lopez-Garcia et al. (2015) collect about 4000 adults' yogurt consumer data over about 3 or 4 years and find that the frequency of consuming yogurt during a week associates with the intake of calcium and sugar but does not significantly associate with physical activity, sleep, BMI or energy intake.

The literature that links physical activity and diet or product choice mainly approaches from public health and measures diet quality with consumption of a food category, diet quality index or frequency of healthy food purchases, and rarely consider from an economic perspective. In this study, we measure diet quality with expenditure shares on heathy and unhealthy food categories, while this also implies consumers' willingness to spend for a food category.

Furthermore, this study contributes to literature that links obesity, type 2 diabetes and medication usages for these two diseases with dietary choice. Few studies particularly investigate the link between obesity and food market. Wang et al. (2017) employ a dynamic discrete choice demand model that considers consumers' stockpiling behavior to estimate consumer demand on soft drinks and examine the effect of obesity rate in a market on soft drink consumption. They find that consumers in a market (i.e.

county) with an obesity rate lower than 25th percentile are more price sensitive than markets with an obesity rate higher than 25th percentile.

Other studies of similar topics investigate first the association between the consumption of food products or nutrients and the risk of obesity or obesity, and second the link between dietary patterns and obesity, and third the effectiveness of various weight loss strategies including diet, exercise and both diet and exercise.

For the first cluster of studies that focus on the link between food or nutrient intake and obesity, Bandini et al. (1999) find that most participants underreport the actual consumption of energy, but also find that obese adolescents consume more high-calorie and low-nutrient-dense foods such as candy and baked foods than non-obese adolescents. Ludwig et al. (2001) find the association between sugar-sweetened beverages intake and children obesity. Gillis and Bar-Or (2003) find that obese children and adolescents are more likely to consume meats, grain products, sugar-sweetened beverages, potato chips and food away from home than nonobese ones. Bray et al. (2004) suggest that there is a link between the consumption of high-fructose corn syrup in beverages and the incidence of obesity. Swinburn et al. (2004) review literature and find individuals are more likely to be obese when maintains a diet with less non-starch polysaccharides and fiber or a diet with high energy-dense and micronutrient-poor foods, or when often consumes sugar-sweetened beverages and fruit juices. Hu and Malik (2010) based on existing literature find that there are links of the consumption of sugar-sweetened beverages with the incidences of obesity, long-term weight gain, type 2 diabetes and cardiovascular disease. Della Torre et al. (2015) review 32 studies and find that majority of these studies demonstrate the link between the consumption of sugar sweetened beverages and risk of obesity or obesity. Anari et al., (2017) find that the consumption of sugar-sweetened beverages associates with the increase in incidence of abdominal obesity among diabetic patients.

The second cluster of literature investigates link between dietary patterns and obesity. Skov et al. (1999) find that an ad libitum fat-reduced diet or replacing foods with richer carbohydrate with foods with richer protein has better performance in individual weight loss. Spieth et al. (2000) find that a low-glycemic index diet is more effective in weight loss than a reduced-fat diet for obesity patients. Foster et al. (2003) find that a low-carbohydrate diet has a better performance in weight loss and the prevention of coronary heart diseases and obesity than a conventional diet. Samaha et al. (2003) and Yancy et al. (2004) also find a low-carbohydrate diet is superior to a low-fat diet in weight loss.

For the third cluster, the effectiveness of several weight loss strategies such as diet-induced weight loss and exercise-induced weight loss are examined. Skender et al. (1996) find that a healthy diet is more effective in weight loss than exercise, and performing both strategies is more effective in weight loss than only performing one of these two strategies, but the variation of outcome is considerable. Sallis et al. (2009) also find that a diet-induced weight loss is effective for obesity control but not as effective as an exercise-induced weight loss.

Most studies that link type 2 diabetes and dietary choice focus on how healthy dietary patterns improve glycemic control or prevent type 2 diabetes. Several studies document the benefit of various healthy diet patterns for type 2 diabetes including a low-carbohydrate diet or a low-carbohydrate Mediterranean diet (Boden et al., 2005; Yancy et al., 2005; Schröder, 2007; Westman et al., 2008; Esposito et al., 2009; Elhayany et al., 2010; Esposito et al., 2010; Salas-Salvado et al., 2011), a vegetarian or vegan diet (Jenkins et al., 2003; Barnard et al., 2006; Barnard et al., 2009; Tonstad et al., 2009; Kahleova et al., 2010), a diet based on American Diabetic Association guidelines (Barnard et al., 2006; Barnard et al., 2009; Elhayany et al., 2010), a high protein weight loss diet and a low protein weight loss diet (Parker et al., 2002), and so forth.

Any one of above diet patterns benefits type 2 diabetes patients by lowering carbohydrate intake and reducing the concentration of glucose, but effectiveness of dietary patterns varies. Parker et al. (2002) find that a high-protein weight loss diet is superior to a lower-protein diet in lowering the risk of type 2 diabetes. Gannon and Nuttall (2004) examine how several dietary patterns differed in levels of carbohydrates and proteins affect blood glucose, and suggest a low-biologically-available glucose diet is the most effective dietary pattern in reducing glucose concentration for type 2 diabetes patients without medical treatment. Esposito et al. (2009) find a low-carbohydrate Mediterranean-style diet has a better performance in glycemic control than a low-fat diet. Elhayany et al. (2010) investigate how a low carbohydrate Mediterranean diet, a traditional Mediterranean diet and the 2003 American Diabetes Association (ADA) diet contribute to the reduction of glycemic, respectively, and find a lower carbohydrate Mediterranean is superior to other two. Kahleova et al. (2010) find that a calorie-restricted vegetarian diet has better performance than a conventional diabetic diet in the improvement of insulin sensitivity, and a combination of diet and exercise is the most preferable way. Salas-Salvado et al. (2011) compare two Mediterranean diets and a low-fat diet for diabetes, and find that a Mediterranean diet without restriction on calorie intake is the most effective diet in the prevention of diabetes among three dietary patterns.

The second strand of literature regarding type 2 diabetes investigates the link between type 2 diabetes and general diet quality, where diet quality is measured with various indices such as Healthy Eating Index, alternative Healthy Eating Index, Recommended Food Score, alternative Mediterranean Diet Score Dietary Approaches to Stop Hypertension Score. One common conclusion is that a good diet quality associates with a lower risk of type 2 diabetes (Fung et al., 2007; Liese et al., 2009; De Koning et al., 2011).

The third strand of literature investigates how a mixed weight control strategy, doing diet and exercise together, relates with diet 2 diabetes mellitus. Barnard et al. (2006) and Barnard et al. (2009) show the positive impact of a low-fat vegan diet and a diet based on ADA guidelines on the reduction of glycemic and lipid among type 2 diabetes patients. Boden et al. (2005) suggest a healthy diet with low carbohydrate, high protein and high fat improve blood glucose and weight loss. Esposito et al. (2010) conduct a literature review about Mediterranean diet and type 2 diabetes, and find that all studies using randomized controlled trials show the positive impact of a Mediterranean diet on glycemic control as well as the prevention of cardiovascular diseases. Jenkins et al. (2003) show the benefit of a vegetarian diet on type 2 diabetes patients by analyzing the components of a vegetarian diet. Hu et al. (2001) suggest that the practice of a healthier diet and lifestyle habits can reduce the risk of type 2 diabetes. Halton et al. (2008) show that taking fat and protein from vegetable sources is likely to help the control of diabetes. Schröder (2007) reviews literature and find that the positive influence of a Mediterranean diet on weight loss and prevention of type 2 diabetes. Tonstad et al. (2009) suggest the vegetarian diets lower the risk of obesity and type 2 diabetes and lower BMI. Yancy et al. (2005) find that a low-carbohydrate, ketogenic diet lowers blood glucose for type 2 diabetes patients. Westman et al. (2008) find that a diet with low carbohydrate improves the capacity of glycemic control.

In this literature review section, we find abundant studies investigate the link between diet or dietary choice with exercise, obesity or type 2 diabetes with experiments or health-related databases such as NHANES data, and not so many studies particularly focus on a certain product or food category or approach this question from an economic or a market angle. Sugar-sweetened beverages, or say soft drinks, are one of the popular food categories being widely investigated, and the association between the consumption of those unhealthy products such as sugar-sweetened beverage as well as potato chips and obesity or type 2 diabetes is documented in literature, for example, Ludwig et al. (2001), Gillis and Bar-Or (2003), Bray et al. (2004), Swinburn et al. (2004), Hu and Malik (2010), Della Torre et al. (2015) and Anari et al., (2017). Instead of measuring the consumption by volume, we measure by expenditure share on one unhealthy food category to denote potential intake and willingness to spend. This means that we are interested in how many percentages of one consumer’s total expenditures on grocery consumers are willing to spend on a healthy or an unhealthy food category. This allows us to better understand the relationship between the willingness to spend, or say potential intake, and exercise, obesity or type 2 diabetes while controlling the spatial and temporal variations and demographics with panel data. Besides soft drinks, we also investigate several others food categories including fruits and vegetables, yogurts, snacks and chips, diet soda and bottled water. Because of our high-quality IRI data, we can observe the grocery purchases for a household over multiple year, which allows us to investigate more than just one product or one dietary pattern. This also contributes to literature by providing more information about the link between the dietary choice of one particular food category and the usage of medications for obesity and type 2 diabetes.

Data

This study employs the IRI (i.e. Information Resources, Incorporated) Consumer Panel and Medprofiler Panel from 2013 to 2017 acquired from USDA Economic Research Service. The IRI Consumer Panel consists of more than 120,000 American households' grocery purchases record for every time they visit a store across 49 states, demographic characteristics of those households, and UPC-level product characteristics. The trip visit data consist of UPC barcode, price, quantity and at sales or not for each product purchased in a store visit and date and type of retail channel of the trip. The household demographic data consist of comprehensive information about household on a yearly basis including each household member's gender and birthday, household heads' employment status and education level and household census tract, zip code, state, market, type of housing and so forth. The UPC-level product information consists of basic product characteristics on yearly basis such as brand, package size and package type as well as nutrition fact panel that provides the amount of sodium, sugar, calorie and protein and other nutrients.

From 2013 to 2017, IRI collected medical demographic and health-related behavior information of almost all household members from about 30,000 households of the Consumer Panel with s yearly-basis survey. The Medprofiler Panel consists of medical demographic such as whether the participant had been diagnosed with one disease, for instance, obesity, type 2 diabetes mellitus and cardiovascular disease, and whether the participant had been taking any Rx, OTC or dual medications for this disease. It also consists of health-related lifestyle information, for example, how often the participant does physical activity in a week, most days in week, some days in a week or rarely/never.

To measure dietary choice and consumers' spending on a food category, we use the expenditure share on a food category. The food categories this study investigates are shown in Table 2-1, where we include some healthy foods such as fruits and vegetables, yogurts and bottled water, and some unhealthy foods such as snacks and chips, regular and diet soft drinks.

Table **2-1**: Food categories.

Food Categories	Included Sub-groups
Fruits & Vegetables (F&V)	Uniform Weighted Apples, Grapefruits, Oranges, and other Fruits, Uniform Weighted Dark Green, Orange, Starchy and Other Vegetables
Snacks & Chips (S&C)	Crackers, Salted Snack, Cookies, Chewy Snack, Nut Snacks, Potato Chips and Tortilla/Tostada Chips
Yogurts	Refrigerated Yogurt, Yogurt Drinks
Regular Soft Drink (RSD)	Carbonated Soda, Regular Soft Drink
Diet Soft Drink (DSD)	Low Calorie Soft Drink
Bottled Water	Distilled Water, Non-Carbonated Water, Bottled Water

The summary statistics are shown in Table 2-2. In order to match up household-level consumer data with individual-level medical and lifestyle data, we only select single-member households. Because some panelists in the Consumer and Medprofiler Panels are allowed to quit from the panel after any year. To obtain a balanced panel over five years, we only keep the single-member households that appear in every year of the five years, which left about 3,500 single-member households. About 30 percent are male but we control the variation in gender as well as other important demographics in our models. About 50 percent enrolled or graduated from a college or university. About 35 percent are employed on full-time basis, and about 15 percent are employed on part-time basis. About 10 percent are African American, about two percent are Hispanic, about two percent are Asian, and the rest are non-Hispanic White.

Regarding the health-related information, we allow their behaviors to change from year to year in the five years. About 40 percent of observations are physically active, doing physical activity in most days in a week. About 20 percent of the single-member households are self-reported as obesity, and only 15 percent of the obese individuals report taking any Rx, OTC, or dual medications for obesity. To identify obesity with a relatively objective method, besides the self-reported medical survey, we also calculate their BMIs based on self-reported height and weight by the IRI Medprofiler participants. According to the CDC, an American adult with a BMI larger than or equal to 30 is overweight and obese (CDC, April 11, 2017), and about 38 percent of observations can be identified as obese based on BMI ≥ 30. The average BMI over the entire sample is about 29. In our data, about 13 percent have type 2 diabetes mellitus, and more than 90 percent of those participants with type 2 diabetes mellitus are taking Rx, OTC, or dual medications.

For the dietary choice, about 6 percent of a single-member households' annual expenditure on grocery products is spent on fruits and vegetables (F&V) on average; about 8 percent is spent on snacks and chips (S&C) on average; about 3 percent is spent on yogurts on average; about 2 percent is spent on regular soft drinks (RSD) on average; about 1 percent is spent on diet soft drinks (DSD) on average; and about 1 percent is spent on bottled water (BW) on average.

Table **2-2**: Statistical summary.

Variables	Description	Value
N	Number of Individual/Household	17,420
		Percentage
Gender	A binary variable that equals to 1 if the individual is male, and equals to 0 otherwise.	0.290 (0.454)
College	A binary variable that equals to 1 if enrolled or graduated from a college or university, and equals to zero otherwise.	0.510 (0.500)
Age	Female	63.8 (10.8)
	Male	62.6 (10.2)
FTEmp	A binary variable that equals to 1 if employed on full-time basis, and equals to 0 otherwise.	0.348 (0.476)
PTEmp	A binary variable that equals to 1 if employed on part-time basis, and equals to 0 otherwise.	0.149 (0.356)
Asian	A binary variable that equals to 1 if the individual is Asian, and equals to 0 otherwise.	0.017 (0.130)
AfrAm	A binary variable that equals to 1 if the individual is African American, and equals to 0 otherwise.	0.097 (0.296)

Hispanic	A binary variable that equals to 1 if the individual is Hispanic, and equals to 0 otherwise.	0.017 (0.129)
Physical Activity (PA)	A binary variable that equals to 1 if the individual does physical activity in most days in a week, and equals to 0 otherwise.	0.390 (0.488)
BMI ≥ 30	A binary variable that equals to 1 if the individual's BMI is larger than or equal to 30, and equals to 0 otherwise.	0.375 (0.484)
Self-reported Obesity	A binary variable that equals to 1 if the individual self-reports being diagnosed with or suffering obesity, and equals to 0 otherwise.	0.190 (0.392)
Self-reported Obesity and Taking Medications	A binary variable that equals to 1 if the individual self-reports being diagnosed with or suffering and is taking Rx, OTC or Dual medications, and equals to 0 otherwise.	0.028 (0.166)
Self-reported Type 2 Diabetes Mellitus	A binary variable that equals to 1 if the individual self-reports being diagnosed with or suffering type 2 diabetes, and equals to 0 otherwise.	0.131 (0.338)
Self-reported Type 2 Diabetes and Taking Medications	A binary variable that equals to 1 if the individual self-reports being diagnosed with or suffering type 2 diabetes and is taking Rx, OTC or Dual medications, and equals to 0 otherwise.	0.120 (0.325)
		Means
Income	Household annual income.	48770.26 (43668.21)
BMI	Body mass index.	29.29 (7.37)
ExpshFV	Household annual expenditure shares of fruits and vegetables over all grocery expenditure.	0.061 (0.053)
ExpshSC	Household annual expenditure shares of snacks and chips over all grocery expenditure.	0.080 (0.052)
ExpshYG	Household annual expenditure shares of yogurts over all grocery expenditure.	0.032 (0.047)
ExpshRSD	Household annual expenditure shares of regular soft drinks over all grocery expenditure.	0.022 (0.036)
ExpshDSD	Household annual expenditure shares of diet soft drinks over all grocery expenditure.	0.014 (0.033)

ExpshBW	Household annual expenditure shares of bottled water over all grocery expenditure.	0.011 (0.024)

Notes: Values in parenthesis are standard deviation. This statistical summary is for five-year data for time-varying variables.

Methodology

To investigate the link between the expenditure share of a food category and physical activity, obesity or type 2 diabetes, we first aggregate trip-level household expenditures on each food category and all grocery products over the year and calculate yearly expenditure share on a food category over grocery products Y_{its} for single-member household i in year t in state s. Then we construct a linear model of the yearly expenditure share on a food category Y_{its} regressed against the variables of interest and some demographic variables that capture the variation of individuals.

First, for the model of physical activity, we employ an Ordinary Least Square (OLS) model to estimate whether physical activity associates with household expenditure share on a food category. The model can be written as

$$Y_{its} = \delta + \alpha PA_{its} + \beta X_{its} + \gamma_t + \theta_s + \varepsilon_{its}, \tag{2.1}$$

where Y_{its} is a variable that denotes household i's expenditure shares on a food category at year t in state s, PA_{its} is a variable that equals one if the single-member household i in year t and state s is physically active and equals zero otherwise, and X_{its} is a vector of demographic characteristics that capture the heterogeneity of households. Additionally, γ_t is a time-specific fixed effect; θ_s is a state fixed effect; ε_{its} is error term; and δ, α and β are coefficients.

For the models of obesity or type 2 diabetes, we replace PA_{its} in Equation (2.1) with H_{its}, a dichotomous variable that denotes whether the individual i has been diagnosed with a disease. The OLS model can be written as

$$Y_{its} = \delta + \alpha H_{its} + \beta X_{its} + \gamma_t + \theta_s + \varepsilon_{its}, \tag{2.2}$$

where for the model of obesity, H_{its} is a dichotomous variable that equals one if individual i in state s at year t is obese and equals zero otherwise. There are two methods to identify obesity. The first one relies on the self-reported IRI Medprofiler data, where participants report whether they are suffering obesity and whether they are taking Rx, OTC or dual medications for obesity. The second one is to identify obesity with BMI where BMI is calculated based on self-reported weight and height documented in the IRI Medprofiler panel. According to CDC (2017, April 11), a BMI larger than or equal to 30 is a standard to identify overweight, and a BMI larger than or equal to 40 is a standard to identify severe or morbid obesity. For the model of type 2 diabetes, we only consider the self-reported IRI Medprofiler data and define H_{its} as a dichotomous variable that equals one if household i in state s at year t is self-reported being diagnosed with type 2 diabetes mellitus or taking any Rx, OTC or dual medications for type 2 diabetes and equals zero otherwise.

However, when both self-reported survey and BMI can be used to identify obesity, we find that some individuals whose BMI < 30 report they are obese, and some individuals whose BMI ≥ 30 report they are healthy. Self-reported survey is subjectively reported by participants once a year. For chronical diseases such as obesity, participants might be unaware of having such a disease until it has been so severe that bothers their living. Hence, the self-reported obesity fails to prevent the misspecification of obesity by participants themselves. On the other hand, the calculation of BMI relies on the self-reported weight and height. The drawback of this method is that it fails to distinguish lean mass from fat mass and therefore overestimates obesity rate (Buckhauser et al., 2009; Rippe, et al., 2012; Cawley, 2015). These reasons lead us to consider these two identification strategies for obesity together in one model, where we create multiple dichotomous variables that categorize individuals by self-reported obesity and BMI. Furthermore, since the self-reported medication usage for obesity is also provided by the IRI Medprofiler survey, we also consider it together with obesity. Hence, in order to consider multiple conditions in one model, we replace H_{its} with a vector of dichotomous variables. The model of obesity with a mixed method to identify obesity and medication usage can be written as

$$\begin{aligned} Y_{its} = \delta + \alpha_1 H_{its}^{M,BMI\geq 30} + \alpha_2 H_{its}^{O,BMI\geq 30} + \alpha_3 H_{its}^{H,BMI\geq 30} + \alpha_4 H_{its}^{M,BMI<30} \\ + \alpha_5 H_{its}^{O,BMI<30} + \beta X_{its} + \gamma_t + \theta_s + \varepsilon_{its}, \end{aligned} \tag{2.3}$$

where $H_{its}^{M,BMI\geq 30}$ is a dichotomous variable that equals one if individual i in state s at year t is self-reported taking Rx, OTC or dual medications for obesity and having a BMI ≥ 30 and equals zero otherwise; $H_{its}^{O,BMI\geq 30}$ is a dichotomous variable that equals one if individual i in state s at year t is self-reported obesity, not taking any medications for obesity and having a BMI ≥ 30 and equals zero otherwise; $H_{its}^{H,BMI\geq 30}$ is a dichotomous variable that equals one if individual i in state s at year t is self-reported nonobese, not taking any medications for obesity (i.e. self-reported healthy) and having a BMI ≥ 30 and equals zero otherwise; $H_{its}^{M,BMI<30}$ is a dichotomous variable that equals one if individual i in state s at year t is self-reported taking Rx, OTC or dual medications for obesity and having a BMI < 30 and equals zero otherwise; $H_{its}^{O,BMI<30}$ is a dichotomous variable that equals one if individual i in state s at year t is self-reported obesity, not taking any medications for obesity and having a BMI < 30 and equals zero otherwise; the baseline case is healthy and having BMI < 30. α_1, α_2, α_3, α_4 and α_5 are coefficients.

We design a similar model for type 2 diabetes. However, we do not consider any identification method other than self-reported IRI Medprofiler survey. The model of type 2 diabetes and medication usage can be written as

$$Y_{its} = \delta + \alpha_1 H_{its}^{D} + \alpha_2 H_{its}^{M} + \beta X_{its} + \gamma_t + \theta_s + \varepsilon_{its}, \tag{2.4}$$

where H_{its}^{D} is a dichotomous variable that equals one if individual i in state s at year t is being diagnosed with type 2 diabetes but not taking any Rx, OTC or dual medications and equals zero otherwise; H_{its}^{M} is a dichotomous variable that equals one if individual i in state s at year t is taking Rx, OTC or dual medications and equals zero otherwise. α_1 and α_2 are coefficients.

Results

Physical Activity and Dietary Choices

To examine the link between physical activity and dietary choice, we first perform OLS regressions of the expenditure shares of food categories against physical activity and demographics as Equation (2.1). The results are shown in Table 2-3. Recall that PA (i.e. physical activity) is a dichotomous variable equals one if the individual reports doing exercise in most days in a week and equals zero otherwise. We find that PA positively associates with expenditure shares of healthy foods such as fruits and vegetables (F&V) and yogurts and negatively associates with expenditure shares of unhealthy foods such as snacks and chips (S&C), regular soft drinks (RSD) and diet soft drinks (DSD). There is no significant relationship between physical activity and the expenditure share of bottled water (BW). For the interpretation of the results, using the first column of the F&V model as an example, the expenditure shares are 0.94 percentage points higher, on average, for physically active individuals. Demographic characteristics such as gender, education, employment, income and race also significantly relate with expenditure shares in almost all cases.

Table **2-3**: The OLS regressions of food expenditure shares against physical activity and demographics.

	F &V		S & C		Yogurts	
Intercept	5.79^{***}	10.74^{***}	8.08^{***}	7.42^{***}	2.45^{***}	2.21^{***}
	(0.08)	(0.32)	(0.08)	(0.34)	(0.07)	(0.30)
PA	0.94^{***}	0.89^{***}	-0.69^{***}	-0.67^{***}	0.82^{***}	0.81^{***}
	(0.08)	(0.08)	(0.08)	(0.08)	(0.07)	(0.07)
Male	-0.80^{***}	-0.83^{***}	0.41^{***}	0.44^{***}	-1.22^{***}	-1.22^{***}
	(0.09)	(0.08)	(0.09)	(0.09)	(0.08)	(0.08)
College	0.75^{***}	0.58^{***}	-0.23^{***}	-0.23^{***}	0.83^{***}	0.86^{***}
	(0.08)	(0.08)	(0.08)	(0.08)	(0.07)	(0.07)
FTEmp	-0.59^{***}	-0.81^{***}	0.40^{***}	0.41^{***}	0.43^{***}	0.41^{***}
	(0.09)	(0.09)	(0.09)	(0.09)	(0.08)	(0.08)
PTEmp	-0.67^{***}	-0.71^{***}	0.51^{***}	0.51^{***}	0.12	0.09
	(0.12)	(0.11)	(0.12)	(0.12)	(0.10)	(0.10)
Income	0.03	0.26^{***}	-0.17^{*}	-0.22^{**}	0.55^{***}	0.45^{***}
	(0.10)	(0.09)	(0.10)	(0.10)	(0.09)	(0.09)
Asian	1.24^{***}	1.38^{***}	0.59^{*}	0.91^{***}	−0.12	−0.25
	(0.31)	(0.29)	(0.30)	(0.31)	(0.27)	(0.27)
AfrAm	0.65^{***}	0.52^{***}	0.92^{***}	0.80^{***}	-1.17^{***}	-1.17^{***}
	(0.14)	(0.13)	(0.13)	(0.14)	(0.12)	(0.12)
Hispanic	−0.09	−0.15	−0.14	0.04	−0.19	−0.11
	(0.31)	(0.28)	(0.30)	(0.30)	(0.27)	(0.27)
Year FE		X		X		X
State FE		X		X		X
Adj R^2	0.022	0.185	0.010	0.032	0.043	0.061

N	17420

Notes: The values in parenthesis are standard error. *** is p-value ≤ 0.01, ** is p-value ≤ 0.05 and * is p-value ≤ 0.1. F&V is fruits and vegetables and S&C is snacks and chips. In each model, the dependent variable is the expenditure share times 100, and the income is measured with 100,000.

Table **2-3**: The OLS regressions of food expenditure shares against physical activity and demographics (Cont.).

	RSD		DSD		BW	
Intercept	2.17***	2.05***	1.03***	1.25***	1.04***	0.97
	(0.05)	(0.24)	(0.05)	(0.21)	(0.04)	(0.16)
PA	−0.40***	−0.40***	−0.14***	−0.13***	−0.001	−0.01
	(0.06)	(0.06)	(0.05)	(0.05)	(0.04)	(0.04)
Male	0.70***	0.71***	0.32***	0.33***	−0.12***	−0.13***
	(0.06)	(0.06)	(0.05)	(0.05)	(0.04)	(0.04)
College	−0.23***	−0.21***	0.34***	0.34***	−0.10**	−0.09**
	(0.06)	(0.06)	(0.05)	(0.05)	(0.04)	(0.04)
FTEmp	0.42***	0.40***	0.30***	0.28***	0.22***	0.25***
	(0.06)	(0.06)	(0.06)	(0.06)	(0.04)	(0.04)
PTEmp	0.32***	0.33***	0.19***	0.19***	0.04	0.05
	(0.08)	(0.08)	(0.07)	(0.07)	(0.05)	(0.05)
Income	−0.40***	−0.44***	0.26***	0.30***	0.01	−0.04
	(0.07)	(0.07)	(0.06)	(0.06)	(0.05)	(0.05)
Asian	−0.27	−0.11	−0.63***	−0.45**	−0.34**	−0.34**
	(0.21)	(0.21)	(0.19)	(0.19)	(0.14)	(0.14)
AfrAm	1.16***	1.14***	−1.00***	−1.01***	0.69***	0.67***
	(0.09)	(0.09)	(0.08)	(0.08)	(0.06)	(0.06)
Hispanic	0.52**	0.52**	−0.03	0.02	1.64***	1.55***
	(0.21)	(0.21)	(0.19)	(0.19)	(0.14)	(0.14)
Year FE		X		X		X
State FE		X		X		X
Adj R^2	0.024	0.037	0.019	0.035	0.017	0.030
N			17420			

Notes: The values in parenthesis are standard error. *** is p-value ≤ 0.01, ** is p-value ≤ 0.05 and * is p-value ≤ 0.1. RSD is regular soft drinks, DSD is diet soft drinks and BW is bottled water. In each model, the dependent variable is the expenditure share times 100, and the income is measured with 100,000.

Obesity, Type 2 Diabetes, Medication Usage and Dietary Choice

In this section, we explore the relationship of dietary choice with obesity and type 2 diabetes. Table 2-4 shows numbers of observations by BMI and responses for self-reported obesity for 3484 individuals over five years as shown in Equation (2.2). About 61 percent of observations are healthy in term of BMI (i.e. BMI < 30) and self-reported medical survey, where about 20 percent of observations (i.e. 3458) claim they are not obese or taking any obesity medications based on self-reported survey, but are obese based on BMI (i.e. BMI ≥ 30). About 2500 observations are obese based on both self-reported survey and BMI, but not taking any Rx, OTC or dual medications, and about 400 observations are taking Rx, OTC or Dual medications and having BMI ≥ 30. However, there is a group of observations who are not obese based on BMI but claim they are suffering obesity or taking medications for obesity. Using BMI as a standard to identify obesity is controversial because muscular individuals with a lean mass level above the average will have high BMI, mean BMI fails to distinguish lean mass from fat mass (Buckhauser et al., 2009; Rippe, et al., 2012; Cawley, 2015). Therefore, we employ both BMI and self-reported obesity to identify the health status regarding obesity of an individual. We define healthy individuals who claim they are healthy by themselves and have BMI < 30 as the baseline case in the OLS regression, and create dichotomous variables for every scenario regarding BMI, self-reported obesity and medication usage to capture different situations.

A similar exercise is also performed for the analysis of type 2 diabetes and dietary choice as shown in Table 2-5. Significantly differed from obesity, about 90 percent of type 2 diabetes patients claim they are taking Rx, OTC or dual medications, where only about 15 percent of obese patients claim taking medications based on self-report survey. Similarly, we employ OLS regressions for analysis, and treat the no type 2 diabetes observations are the baseline case.

Table **2-4**: The distribution of observations by self-reported obesity and BMI.

	BMI<30	BMI≥30	Total
Self-Reported Taking Rx, OTC or Dual Medications	78	418	493
Self-Reported Sufferer and Not Taking Medications	266	2546	2812
Self-Reported Healthy	10654	3458	14112
Total	10998	6422	17420

Notes: The numbers of observations are numbers of participants over five years, meaning one individual has five data points in total and is allowed to be in different slots from year to year.

Table **2-5**: The numbers of observations who are diagnosed with type 2 diabetes, who are taking medications and who are healthy.

Variables	Counts
Type 2 Diabetes Taking Rx, OTC or Dual Medications	2089
Type 2 Diabetes but Not Taking Medications	198
No Type 2 Diabetes	15133
Total Observations	17420

Notes: The numbers of observations are numbers of participants over five years, meaning one individual has five data points in total and is allowed to be in different slots from year to year.

The regression results of the simplified models of Equation (2.2) for obesity are shown in Table 2-6. The three columns of each food category denote the models that use self-reported obesity, BMI $\geq$ 30 and BMI $\geq$ 40 to identify overweight or obesity, respectively. Using the regressions of F&V as an example, the individuals who are self-reported obese or taking any Rx, OTC or dual medications spend about 0.61 percentage less on fruits and vegetables on average than the healthy ones. The sign of coefficients is identical for each food category over identification methods of obesity. The absolute value of coefficient of BMI ≥ 40 is larger than the other two variables in most cases. This might imply that individuals with larger BMI tend to spend less on F&V and yogurts and spend more on S&C and DSD on average.

These three coefficients of obesity in separate models are not comparable because the coefficients of other variables and error terms vary over models. This leads us to develop a new model with multiple dichotomous variables that specify self-reported obesity, BMI and medication usage in one model so that self-reported obesity and BMI are comparable.

Table **2-6**: The OLS regressions of food expenditure shares against obesity and demographics.

	F&V			S&C		
Intercept	11.15***	11.31***	11.10***	7.13***	6.98***	7.16***
	(0.32)	(0.32)	(0.32)	(0.34)	(0.34)	(0.34)
Obesity	−0.61***			0.38***		
	(0.09)			(0.10)		
BMI ≥ 30		−0.70***			0.55***	
		(0.08)			(0.08)	
BMI ≥ 40			−0.94***			0.66***
			(0.13)			(0.14)
Male	−0.90***	−0.89***	−0.89***	0.48***	0.48***	0.49***
	(0.08)	(0.08)	(0.08)	(0.09)	(0.09)	(0.09)
College	0.65***	0.61***	0.64***	−0.28***	−0.24***	−0.27***
	(0.11)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
FTEmp	−0.81***	−0.80***	−0.80***	0.41***	0.41***	0.40***
	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)
PTEmp	−0.71***	−0.71***	−0.70***	0.50***	0.51***	0.50***
	(0.29)	(0.11)	(0.11)	(0.12)	(0.12)	(0.12)
Income	0.31***	0.30***	0.30***	−0.27***	−0.25**	−0.25**
	(0.13)	(0.09)	(0.09)	(0.10)	(0.10)	(0.10)
Asian	1.24***	1.17***	1.26***	1.00***	1.07***	0.99***
	(0.29)	(0.29)	(0.29)	(0.31)	(0.31)	(0.31)
AfrAm	0.49***	0.57***	0.52***	0.82***	0.76***	0.80***
	(0.13)	(0.13)	(0.13)	(0.14)	(0.14)	(0.14)
Hispanic	−0.20	−0.19	−0.15	0.07	0.07	0.04

	(0.29)	(0.28)	(0.29)	(0.30)	(0.30)	(0.30)
Year & State FEs	X	X	X	X	X	X
R^2	0.180	0.182	0.181	0.029	0.031	0.030
N			17420			

Notes: The values in parenthesis are standard error. *** is the p-value ≤ 0.01, ** is the p-value ≤ 0.05 and * is the p-value ≤ 0.1. F&V is fruits and vegetables and S&C is snacks and chips. In each model, the dependent variable is the expenditure share times 100, and the income is measured with 100,000. Obesity denotes self-reported suffering obesity or taking any Rx, OTC or dual medications for obesity.

Table **2-6**: The OLS regressions of food expenditure shares of against obesity and demographics (Cont.).

	Yogurts			RSD		
Intercept	2.60***	2.76***	2.53***	1.93***	1.91***	1.93***
	(0.30)	(0.31)	(0.30)	(0.24)	(0.24)	(0.24)
Obesity	−0.63***			0.02		
	(0.09)			(0.07)		
BMI ≥ 30		−0.72***			0.04	
		(0.07)			(0.06)	
BMI ≥ 40			−0.80***			−0.06
			(0.13)			(0.10)
Male	−1.29***	−1.28***	−1.28***	0.73***	0.73***	0.73***
	(0.08)	(0.08)	(0.08)	(0.06)	(0.06)	(0.06)
College	0.92***	0.87***	0.91***	−0.24***	−0.24***	−0.24***
	(0.07)	(0.07)	(0.07)	(0.06)	(0.06)	(0.06)
FTEmp	0.40***	0.41***	0.41***	0.40***	0.40***	0.40***
	(0.08)	(0.08)	(0.08)	(0.06)	(0.06)	(0.06)
PTEmp	0.08	0.08	0.10	0.32***	0.32***	0.32***
	(0.10)	(0.10)	(0.10)	(0.08)	(0.08)	(0.08)
Income	0.50***	0.49***	0.50***	−0.47***	−0.47***	−0.47***
	(0.09)	(0.09)	(0.09)	(0.07)	(0.07)	(0.07)
Asian	−0.38	−0.46*	−0.36	−0.07	−0.06	−0.07
	(0.27)	(0.27)	(0.27)	(0.21)	(0.21)	(0.21)
AfrAm	−1.19***	−1.11***	−1.17***	1.14***	1.14***	1.15***
	(0.12)	(0.12)	(0.12)	(0.09)	(0.09)	(0.09)
Hispanic	−0.16	−0.15	−0.12	0.54**	0.54**	0.54**
	(0.27)	(0.27)	(0.27)	(0.21)	(0.21)	(0.21)
Year & State FEs	X	X	X	X	X	X
R^2	0.057	0.060	0.057	0.035	0.035	0.035
N				17420		

Notes: The values in parenthesis are standard error. *** is the p-value ≤ 0.01, ** is the p-value ≤ 0.05 and * is the p-value ≤ 0.1. RSD is regular soft drinks. In each model, the dependent variable is the expenditure share times 100, and the income is measured with

100,000. Obesity denotes self-reported suffering obesity or taking any Rx, OTC or dual medications for obesity.

Table **2-6**: The OLS regressions of food expenditure shares of against obesity and demographics (Cont.).

	DSD			BW		
Intercept	1.04***	0.97***	1.15***	0.95***	0.94***	0.97***
	(0.21)	(0.21)	(0.21)	(0.16)	(0.16)	(0.16)
Obesity	0.83***			0.09*		
	(0.06)			(0.05)		
BMI ≥ 30		0.60***			0.07*	
		(0.05)			(0.04)	
BMI ≥ 40			0.82***			−0.01
			(0.09)			(0.07)
Male	0.37***	0.35***	0.35***	−0.13***	−0.13***	−0.13***
	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)
College	0.33***	0.37***	0.33***	−0.09**	−0.09**	−0.09**
	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)
FTEmp	0.28***	0.28***	0.27***	0.25***	0.25***	0.25***
	(0.05)	(0.06)	(0.06)	(0.04)	(0.04)	(0.04)
PTEmp	0.22***	0.21***	0.20***	0.05	0.05	0.05
	(0.07)	(0.07)	(0.07)	(0.05)	(0.05)	(0.05)
Income	0.32***	0.32***	0.32***	−0.04	−0.04	−0.04
	(0.06)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)
Asian	−0.36*	−0.33*	−0.41**	−0.33**	−0.32**	−0.33**
	(0.19)	(0.19)	(0.19)	(0.14)	(0.14)	(0.14)
AfrAm	−1.00***	−1.07***	−1.02***	0.67***	0.66***	0.67***
	(0.09)	(0.09)	(0.09)	(0.06)	(0.06)	(0.06)
Hispanic	0.05	0.03	0.00	1.56***	1.55***	1.55***
	(0.19)	(0.19)	(0.19)	(0.14)	(0.14)	(0.14)
Year & State FEs	X	X	X	X	X	X
R^2	0.044	0.042	0.039	0.031	0.030	0.030
N	17420					

Notes: The values in parenthesis are standard error. *** is p-value ≤ 0.01, ** is p-value ≤ 0.05 and * is p-value ≤ 0.1. DSD is diet soft drinks and BW is bottled water. In each model, the dependent variable is the expenditure share times 100, and the income is measured with 100,000. Obesity denotes self-reported suffering obesity or taking any Rx, OTC or dual medications for obesity.

The regression results for the Equation (2.3) are shown in Table 2-7. This model allows us to identify individuals by self-reported obesity, BMI and obesity medication usage simultaneously. The coefficients regarding obesity and medication usage denote whether expenditure share on a food category for observations in that group is significantly differed from the baseline group on average. Using the first column of the first model as a case for interpretation, on average, the individuals who take Rx, OTC or dual medications and have BMI≥30 are likely to spend about 0.38 percentage point less of their total expenditure on fruits and vegetables than healthy individuals (i.e. individuals who are self-reported not taking any obesity medications, not suffering obesity and having BMI < 30). Comparing the first and second coefficients of interest in the first column, on average, the individuals who are self-reported suffering obesity but not taking any medications and having BMI ≥ 30 spend 0.54 percentage point less of their expenditure on fruits and vegetables than those who are self-report taking medications for obesity and have BMI ≥ 30. The group of individuals who spends the least of their expenditure on fruits and vegetables is those who taking medications for obesity but have BMI < 30 for the model with year and state fixed effects.

As a summary for the models with year and state fixed effects, the results are mixed. The individuals who are self-reported taking obesity medications and have BMI < 30 tend to spend less of their total grocery expenditure on healthy foods such as fruits and vegetables and are likely to spend more on unhealthy foods such as regular soft drinks. However, in the meanwhile, they tend to spend less on unhealthy foods such as snacks and chips. There is a conflict where some individuals prefer some unhealthy foods but do not prefer some other unhealthy foods, but this contradictory relation only exist for who are taking medication with BMI < 30.

For households who are self-reported healthy and have BMI $\geq$ 30 and who are self-reported suffering obesity but not taking any medication and have BMI $\geq$ 30, despite their expenditure share on regular soft drinks is not significantly differing from the baseline, their dietary patterns are less healthy in general, spending more on snacks and chips and diet soft drinks and spending less on fruits and vegetables and yogurts. However, compared with the baseline, the households who are self-reported sufferers of obesity tend to have a less healthy dietary pattern on average, spending more on unhealthy foods and spending less on healthy foods. The households who are taking any medications and have BMI $\geq$ 30 tend to have a similar dietary pattern with the baseline, but they spend less on fruits and vegetables (even not statistically significant) and yogurts and more on diet soft drinks. Meanwhile, the households who are self-reported sufferer and not taking any medications but have BMI < 30 tend to have a similar dietary pattern with the baseline, but they tend to spend less on fruits and vegetables (not statistically significant) and spend more diet soft drinks. Besides aforementioned general trend by groups, we can also observe that there are no statistically significantly differences in the expenditure shares on regular soft drinks and bottled water over groups.

Furthermore, we also perform F statistics to examine whether a coefficient of one group is significantly differed from one other group for the models with year and state fixed effects. The F statistics are shown in Table 2-8. Using 95 percent significance level, in some models, there is not a pair of coefficients significantly correlated, for example, the models of fruits and vegetables, diet soft drinks and bottled water. In the model of yogurts, only one pair of coefficients is significantly correlated. Only in the models of snacks and chips and regular soft drinks there is significantly differences in the expenditure shares on that food categories between coefficients.

Table **2-7**: The OLS regressions of food expenditure shares against obesity, BMI, medication usage and demographics.

	F&V		S&C		Yogurts	
Intercept	6.45***	11.33***	7.62***	6.99***	3.03***	2.77***
	(0.08)	(0.32)	(0.08)	(0.34)	(0.07)	(0.31)
$H^{M,BMI\geq30}$	−0.38	−0.41*	0.36	0.42	−0.79***	−0.79***
	(0.26)	(0.24)	(0.26)	(0.26)	(0.23)	(0.23)
$H^{O,BMI\geq30}$	−0.92***	−0.85***	0.65***	0.64***	−0.82***	−0.85***
	(0.12)	(0.10)	(0.11)	(0.11)	(0.10)	(0.10)
$H^{H,BMI\geq30}$	−0.73***	−0.66***	0.54***	0.48***	−0.64***	−0.63***
	(0.10)	(0.10)	(0.10)	(0.10)	(0.09)	(0.09)
$H^{M,BMI<30}$	−0.87	−1.18**	−1.37**	−1.26**	−0.60	−0.51
	(0.60)	(0.55)	(0.59)	(0.59)	(0.52)	(0.52)
$H^{O,BMI<30}$	−0.69**	−0.50*	−0.21	−0.21	−0.14	−0.16
	(0.32)	(0.30)	(0.32)	(0.32)	(0.28)	(0.28)
Male	−0.86***	−0.89***	0.45***	0.48***	−1.28***	−1.29***
	(0.09)	(0.08)	(0.09)	(0.09)	(0.08)	(0.08)
College	0.77***	0.61***	−0.25***	−0.25***	0.86***	0.88***
	(0.09)	(0.08)	(0.08)	(0.08)	(0.07)	(0.07)
FTEmp	−0.59***	−0.80***	0.38***	0.41***	0.44***	0.41***
	(0.09)	(0.09)	(0.09)	(0.09)	(0.08)	(0.08)
PTEmp	−0.68***	−0.72***	0.51***	0.51***	0.11	0.08
	(0.12)	(0.11)	(0.12)	(0.12)	(0.10)	(0.10)
Income	0.07	0.30***	−0.20**	−0.25**	0.58***	0.49***
	(0.10)	(0.09)	(0.10)	(0.10)	(0.09)	(0.09)
Asian	1.02***	1.17***	0.76**	1.07***	−0.33	−0.46*
	(0.31)	(0.29)	(0.31)	(0.31)	(0.27)	(0.27)
AfrAm	0.70***	0.56***	0.89***	0.78***	−1.12***	−1.12***
	(0.14)	(0.13)	(0.13)	(0.14)	(0.12)	(0.12)
Hispanic	−0.17	−0.20	−0.10	0.07	−0.24	−0.16
	(0.31)	(0.28)	(0.30)	(0.30)	(0.27)	(0.27)
Year FEs		X		X		X
State FEs		X		X		X
R^2	0.021	0.183	0.010	0.031	0.042	0.060
N			17420			

Notes: The values in parenthesis are standard error. *** is p-value ≤ 0.01, ** is p-value ≤ 0.05 and * is p-value ≤ 0.1. F&V is fruits and vegetables, S&C is snacks and chips, RSD is regular soft drinks and DSD is diet soft drinks. In each model, the dependent variable is the expenditure share times 100, and the income is measured with 100,000.

Table **2-7**: The OLS regressions of food expenditure shares against obesity, BMI, medication usage and demographics (Cont.).

	RSD		DSD		BW	
Intercept	2.01***	1.91***	0.70***	0.94***	1.01***	0.94***
	(0.06)	(0.24)	(0.05)	(0.21)	(0.04)	(0.16)
$H^{M,BMI\geq30}$	0.09	0.09	0.83***	0.91***	0.16	0.17
	(0.18)	(0.18)	(0.16)	(0.16)	(0.12)	(0.12)
$H^{O,BMI\geq30}$	−0.01	−0.01	0.90***	0.93***	0.07	0.08
	(0.08)	(0.08)	(0.07)	(0.07)	(0.05)	(0.05)
$H^{H,BMI\geq30}$	0.11	0.09	0.36***	0.37***	0.05	0.05
	(0.07)	(0.41)	(0.06)	(0.06)	(0.05)	(0.05)
$H^{M,BMI<30}$	0.78*	0.69*	−0.28	−0.33	0.04	−0.04
	(0.41)	(0.22)	(0.37)	(0.34)	(0.28)	(0.28)
$H^{O,BMI<30}$	0.17	0.16	1.19***	1.22***	0.27*	0.26*
	(0.22)	(0.06)	(0.20)	(0.20)	(0.15)	(0.15)
Male	0.72***	0.73***	0.36***	0.37***	−0.12***	−0.13***
	(0.06)	(0.06)	(0.05)	(0.05)	(0.04)	(0.04)
College	−0.25***	−0.24***	0.35***	0.35***	−0.09**	−0.09**
	(0.06)	(0.06)	(0.05)	(0.05)	(0.04)	(0.04)
FTEmp	0.42***	0.40***	0.30***	0.28***	0.22***	0.25***
	(0.06)	(0.08)	(0.06)	(0.06)	(0.04)	(0.04)
PTEmp	0.31***	0.32***	0.21***	0.22***	0.04	0.05
	(0.08)	(0.07)	(0.07)	(0.07)	(0.05)	(0.05)
Income	−0.44***	−0.47***	0.30***	0.34***	0.01	−0.03
	(0.07)	(0.21)	(0.06)	(0.06)	(0.05)	(0.05)
Asian	−0.23	−0.06	−0.52***	−0.34*	−0.33**	−0.32**
	(0.21)	(0.10)	(0.19)	(0.19)	(0.14)	(0.14)
AfrAm	1.15***	1.13***	−1.03***	−1.04***	0.68***	0.66***
	(0.09)	(0.21)	(0.08)	(0.08)	(0.06)	(0.06)
Hispanic	0.54**	0.54**	0.02	0.05	1.65***	1.55***
	(0.21)	(0.09)	(0.19)	(0.18)	(0.14)	(0.14)
Year FEs		X		X		X
State FEs		X		X		X
R^2	0.022	0.035	0.031	0.047	0.018	0.031
N			17420			

Notes: The values in parenthesis are standard error. *** is p-value ≤ 0.01, ** is p-value ≤ 0.05 and * is p-value ≤ 0.1. F&V is fruits and vegetables, S&C is snacks and chips, RSD is regular soft drinks and DSD is diet soft drinks. In each model, the dependent variable is the expenditure share times 100, and the income is measured with 100,000.

Table **2-8**: The F statistics between coefficients in the regression of obesity.

	$H^{M,BMI\geq30}$	$H^{O,BMI\geq30}$	$H^{H,BMI\geq30}$	$H^{M,BMI<30}$	$H^{H,BMI<30}$

F&V					
$H^{M,BMI\geq30}$	-	3.06*	1.01	1.66	0.05
$H^{O,BMI\geq30}$		-	2.36	0.34	1.34
$H^{H,BMI\geq30}$			-	0.87	0.29
$H^{M,BMI<30}$				-	1.20
$H^{O,BMI<30}$					-
S&C					
$H^{M,BMI\geq30}$	-	0.63	0.05	7.01***	2.43
$H^{O,BMI\geq30}$		-	1.39	10.26***	6.52**
$H^{H,BMI\geq30}$			-	8.67***	4.42**
$H^{M,BMI<30}$				-	2.54
$H^{O,BMI<30}$					-
Yogurts					
$H^{M,BMI\geq30}$	-	0.05	0.47	0.25	3.14*
$H^{O,BMI\geq30}$		-	3.28*	0.42	5.51**
$H^{H,BMI\geq30}$			-	0.05	2.66
$H^{M,BMI<30}$				-	0.35
$H^{O,BMI<30}$					-
RSD					
$H^{M,BMI\geq30}$	-	0.02	10.66***	9.90***	1.55
$H^{O,BMI\geq30}$		-	45.49***	11.82***	1.95
$H^{H,BMI\geq30}$			-	3.64*	17.80***
$H^{M,BMI<30}$				-	14.22***
$H^{O,BMI<30}$					-
DSD					
$H^{M,BMI\geq30}$	-	0.25	0.00	1.88	0.07
$H^{O,BMI\geq30}$		-	1.07	2.89*	0.53
$H^{H,BMI\geq30}$			-	2.16	0.09
$H^{M,BMI<30}$				-	1.35
$H^{O,BMI<30}$					-
BW					
$H^{M,BMI\geq30}$	-	0.45	0.79	0.46	0.23
$H^{O,BMI\geq30}$		-	0.16	0.17	1.29
$H^{H,BMI\geq30}$			-	0.11	1.74
$H^{M,BMI<30}$				-	0.88
$H^{O,BMI<30}$					-

Notes: $H^{M,BMI\geq30}$ is self-reported taking medications and BMI ≥ 30, $H^{O,BMI\geq30}$ is self-reported sufferers not taking any medications and BMI ≥ 30, $H^{H,BMI\geq30}$ is self-reported healthy and BMI ≥ 30, $H^{M,BMI<30}$ is self-reported taking medications and BMI < 30 and $H^{O,BMI<30}$ is self-reported sufferers not taking medications and BMI < 30. *** is p-value ≤ 0.01, ** is p-value ≤ 0.05 and * is p-value ≤ 0.1. F&V is fruits and vegetables, S&C is snacks and chips, RSD is regular soft drinks, DSD is diet soft drinks and BW is bottled water. The models contain yearly and state fixed effect.

For the simplified models of Equation (2.2) for type 2 diabetes, the results are shown in Table 2-9. Using the first column as an example, the individuals who have been diagnosed with type 2 diabetes or taking any Rx, OTC or dual diabetic medications spend about 0.33 percentage less on F&V on average than the healthy ones. Patients with type 2 diabetes mellitus tend to spend more S&C and DSD and less on F&V, yogurts and RSD on average. This model does not allow us to differentiate individuals who have been taking medications and who have not, and therefore cannot rule out the impact of medication intake on food choice.

Table **2-9**: The OLS regressions of food expenditure shares against type 2 diabetes and demographics.

	F&V	S&C	Yogurts	RSD	DSD	BW
Intercept	11.09***	7.10***	2.64***	1.99***	0.99***	0.97***
	(0.32)	(0.34)	(0.30)	(0.24)	(0.21)	(0.16)
Diabetes 2	−0.33***	0.54***	−0.89***	−0.32***	1.14***	−0.03
	(0.11)	(0.12)	(0.10)	(0.08)	(0.07)	(0.06)
Male	−0.85***	0.44***	−1.21***	0.75***	0.27***	−0.13***
	(0.08)	(0.09)	(0.08)	(0.06)	(0.05)	(0.04)
College	0.64***	−0.26***	0.89***	−0.25***	0.36***	−0.09**
	(0.08)	(0.08)	(0.07)	(0.06)	(0.05)	(0.04)
FTEmp	−0.83***	0.45***	0.34***	0.38***	0.36***	0.25***
	(0.09)	(0.09)	(0.08)	(0.06)	(0.06)	(0.04)
PTEmp	−070***	0.51***	0.07	0.31***	0.24***	0.05
	(0.11)	(0.12)	(0.10)	(0.08)	(0.07)	(0.05)
Income	0.33***	−0.27***	0.50***	−0.48***	0.32***	−0.04
	(0.09)	(0.10)	(0.09)	(0.07)	(0.06)	(0.04)
Asian	1.30***	0.96***	−0.31	−0.06	−0.46**	−0.33**
	(0.29)	(0.31)	(0.27)	(0.21)	(0.19)	(0.14)
AfrAm	0.52***	0.77***	−1.11***	1.17***	−1.10***	0.67***
	(0.13)	(0.14)	(0.12)	(0.09)	(0.09)	(0.06)
Hispanic	−0.17	0.05	−0.13	0.54**	0.01	1.55***
	(0.29)	(0.30)	(0.27)	(0.21)	(0.19)	(0.14)
Year FE	X	X	X	X	X	X
State FE	X	X	X	X	X	X
R^2	0.179	0.030	0.058	0.036	0.047	0.030
N			17420			

Notes: The values in parenthesis are standard error. *** is p-value ≤ 0.01, ** is p-value ≤ 0.05 and * is p-value ≤ 0.1. F&V is fruits and vegetables, S&C is snacks and chips, RSD is regular soft drinks and DSD is diet soft drinks. Diabetes 2 denotes either suffering type 2 diabetes mellitus or taking Rx, OTC or dual medications for type 2 diabetes. In each model, the dependent variable is the expenditure share times 100, and the income is measured with 100,000.

The regressions of type 2 diabetes and dietary choice considering medication usage as shown by Equation (2.4), are shown in Table 2-10. Individuals who are taking Rx, OTC or dual medications tend to consume less fruits and vegetables, yogurts and regular soft drinks, and more snacks and chips and diet soft drinks than the baseline case (i.e. individuals with no type 2 diabetes mellitus). Type 2 diabetes patients who are not taking medications are not significantly differed from the baseline in dietary choice in most cases, except in the model of diet soft drinks where they spend more than baseline but still less than individuals who taking medications. F statistics are also performed to study whether there is a difference between the two coefficients of type 2 diabetes and medication usage. The F statistics are shown in Table 2-11. Even though in the models of fruits and vegetables and regular soft drinks F statistics are significant, the coefficient of having type 2 diabetes but not taking any medications is not significant. F statistics fail to present meaningful interpretation because the coefficients of having type 2 diabetes not taking medications are not significant in the OLS regressions.

Table **2-10**: The OLS regressions of food expenditure shares against type 2 diabetes, medication usage and demographics.

	F&V		S&C		Yogurts	
Intercept	6.17^{***}	11.10^{***}	7.75^{***}	7.10^{***}	2.85^{***}	2.65^{***}
	(0.08)	(0.32)	(0.08)	(0.34)	(0.07)	(0.30)
H^D	-0.53^{***}	-0.40^{***}	0.62^{***}	0.54^{***}	-0.92^{***}	-0.94^{***}
	(0.12)	(0.11)	(0.12)	(0.12)	(0.11)	(0.11)

H^M	0.18	0.33	0.52	0.53	−0.34	−0.37
	(0.38)	(0.35)	(0.37)	(0.37)	(0.33)	(0.33)
Male	-0.81^{***}	-0.85^{***}	0.41^{***}	0.44^{***}	-1.20^{***}	-1.21^{***}
	(0.09)	(0.08)	(0.09)	(0.09)	(0.08)	(0.08)
College	0.80^{***}	0.64^{***}	-0.26^{***}	-0.26^{***}	0.88^{***}	0.89^{***}
	(0.09)	(0.08)	(0.08)	(0.08)	(0.07)	(0.07)
FTEmp	-0.63^{***}	-0.83^{***}	0.43^{***}	0.45^{***}	0.37^{***}	0.34^{***}
	(0.09)	(0.09)	(0.09)	(0.09)	(0.08)	(0.08)
PTEmp	-0.68^{***}	-0.70^{***}	0.52^{***}	0.51^{***}	0.09	0.07
	(0.12)	(0.11)	(0.12)	(0.12)	(0.10)	(0.10)
Income	0.11	0.33^{***}	-0.22^{**}	-0.27^{***}	0.61^{***}	0.51^{***}
	(0.10)	(0.09)	(0.10)	(0.10)	(0.09)	(0.09)
Asian	1.18^{***}	1.31^{***}	0.63^{**}	0.96^{***}	−0.17	−0.30
	(0.31)	(0.29)	(0.30)	(0.31)	(0.27)	(0.27)
AfrAm	0.67^{***}	0.53^{***}	0.89^{***}	0.77^{***}	-1.11^{***}	-1.11^{***}
	(0.14)	(0.13)	(0.13)	(0.14)	(0.12)	(0.12)
Hispanic	−0.11	−0.17	−0.13	0.05	−0.20	−0.12
	(0.31)	(0.29)	(0.30)	(0.30)	(0.27)	(0.27)
Year FE		X		X		X
State FE		X		X		X
R^2	0.016	0.179	0.008	0.030	0.040	0.058
N			17420			

Notes: H^D equals one if individual has been diagnosed with type 2 diabetes and is not taking any medications and equals zero otherwise. H^M equals one if individual is taking any medications and equals zero otherwise. The values in parenthesis are standard error. *** is p-value ≤ 0.01, ** is p-value ≤ 0.05 and * is p-value ≤ 0.1. F&V is fruits and vegetables and S&C is snacks and chips. In each model, the dependent variable is the expenditure share times 100, and the income is measured with 100,000.

Table **2-10**: The OLS Regressions of food expenditure shares against type 2 diabetes, medication usage and demographics (Cont.).

	RSD		DSD		BW	
Intercept	2.09^{***}	1.99^{***}	0.80^{***}	0.98^{***}	1.04^{***}	0.97^{***}
	(0.05)	(0.24)	(0.05)	(0.21)	(0.04)	(0.16)
H^D	-0.32^{***}	-0.32^{***}	1.19^{***}	1.20^{***}	−0.01	−0.02
	(0.08)	(0.08)	(0.08)	(0.08)	(0.06)	(0.06)
H^M	−0.32	−0.31	0.45^{**}	0.47^{**}	−0.07	−0.04
	(0.26)	(0.26)	(0.23)	(0.23)	(0.17)	(0.17)
Male	0.74^{***}	0.75^{***}	0.26^{***}	0.27^{***}	-0.12^{***}	-0.13^{***}
	(0.06)	(0.06)	(0.05)	(0.05)	(0.04)	(0.04)
College	-0.27^{***}	-0.25^{***}	0.37^{***}	0.36^{***}	-0.09^{**}	-0.09^{**}
	(0.06)	(0.06)	(0.05)	(0.05)	(0.04)	(0.04)
FTEmp	0.40^{***}	0.38^{***}	0.38^{***}	0.36^{***}	0.22^{***}	0.25^{***}

	(0.06)	(0.06)	(0.06)	(0.06)	(0.04)	(0.04)
PTEmp	0.30^{***}	0.31^{***}	0.24^{***}	0.24^{***}	0.04	0.05
	(0.08)	(0.08)	(0.07)	(0.07)	(0.0.5)	(0.05)
Income	-0.45^{***}	-0.48^{***}	0.28^{***}	0.32^{***}	0.005	−0.04
	(0.07)	(0.07)	(0.06)	(0.06)	(0.05)	(0.05)
Asian	−0.23	−0.06	-0.65^{***}	-0.47^{**}	-0.34^{**}	-0.33^{**}
	(0.21)	(0.21)	(.019)	(0.19)	(0.14)	(0.14)
AfrAm	1.19^{***}	1.17^{***}	-1.09^{***}	-1.11^{***}	0.68^{***}	0.67^{***}
	(0.09)	(0.10)	(0.08)	(0.09)	(0.06)	(0.06)
Hispanic	0.54^{**}	0.54^{**}	−0.04	0.002	1.64^{***}	1.55^{***}
	(0.21)	(0.21)	(0.19)	(0.19)	(0.14)	(0.14)
Year FE		X		X		X
State FE		X		X		X
R^2	0.022	0.036	0.033	0.048	0.017	0.030
N			17420			

Notes: H^D equals one if the individual has been diagnosed with type 2 diabetes and is not taking any medications and equals zero otherwise. H^M equals one if the individual is taking any medications and equals zero otherwise. The values in parenthesis are standard error. *** is p-value ≤ 0.01, ** is p-value ≤ 0.05 and * is p-value ≤ 0.1. RSD is regular soft drinks, DSD is diet soft drinks and BW is bottled water. In each model, the dependent variable is the expenditure share times 100, and the income is measured with 100,000.

Table **2-11**: The F statistics between coefficients in the regressions of type 2 diabetes.

	F Statistics
F&V	4.03^{**}
S&C	0.00
Yogurts	2.83^{*}
RSD	9.54^{***}
DSD	0.00
BW	0.01

Notes: *** is p-value ≤ 0.01, ** is p-value ≤ 0.05 and * is p-value ≤ 0.1. F&V is fruits and vegetables, S&C is snacks and chips, RSD is regular soft drinks, DSD is diet soft drinks and BW is bottled water. The models include yearly and state fixed effects.

Robustness Check

For the robustness check, we split the sample by a BMI of 30 for Equation (2.1), the model of physical activity, and split the sample by physical activity for the simplified

Equation (2.2) model of obesity and type 2 diabetes. The results of the OLS regressions of food expenditure shares against physical activity and demographic variables where sample is split by a BMI of 30 are shown in Table 2-12. In most cases, the sign of coefficients is substantially similar over BMI ≥ 30 and BMI < 30 in each food category except DSD and BW. The absolute values of PA coefficients for the sample of BMI < 30 are lower than the ones for the sample of BMI ≥ 30. This might imply compared with individuals with a BMI ≥ 30, individuals with lower BMI spend more on healthy foods such as F&V and yogurts and spend less on unhealthy foods such as S&C and RSD on average. There is a significantly association between physical activity and expenditure shares on DSD and BW for the individuals with BMI ≥ 30, but this association does not show significance for the sample with BMI < 30.

Table **2-12**: The OLS regressions of food expenditure shares against physical activity, sample split by BMI.

	F&V		S&C		Yogurts	
	BMI≥30	BMI<30	BMI≥30	BMI<30	BMI≥30	BMI<30
Intercept	9.88***	11.30***	6.55***	7.87***	2.41***	2.16***
	(0.45)	(0.43)	(0.54)	(0.44)	(0.42)	(0.41)
PA	0.41***	0.92***	−0.46***	−0.62***	0.46***	0.78***
	(0.12)	(0.10)	(0.15)	(0.10)	(0.12)	(0.09)
Male	−0.60***	−0.97***	0.17	0.60***	−1.06***	−1.34***
	(0.12)	(0.11)	(0.14)	(0.11)	(0.11)	(0.10)
College	0.29***	0.74***	−0.09	−0.28***	0.76***	0.91***
	(0.11)	(0.11)	(0.13)	(0.11)	(0.10)	(0.10)
FTEmp	−0.81***	−0.83***	0.49***	0.32***	0.49***	0.39***
	(0.12)	(0.12)	(0.15)	(0.12)	(0.12)	(0.11)
PTEmp	−0.71***	−0.70***	0.79***	0.36**	0.10	0.09
	(0.16)	(0.14)	(0.19)	(0.15)	(0.15)	(0.14)
Income	0.48***	0.13	0.02	−0.29**	0.17	0.51***
	(0.15)	(0.12)	(0.18)	(0.12)	(0.14)	(0.11)
Asian	1.10	1.33***	−0.58	1.13***	−1.19*	−0.29
	(0.69)	(0.33)	(0.83)	(0.33)	(0.65)	(0.31)
AfrAm	0.72***	0.44**	0.72***	0.88***	−0.84***	−1.30***
	(0.17)	(0.18)	(0.20)	(0.19)	(0.16)	(0.18)

Hispanic	1.71^{***}	-1.02^{***}	-0.32	0.20	-0.51	0.01
	(0.43)	(0.37)	(0.52)	(0.37)	(0.41)	(0.35)
Year FE	X	X	X	X	X	X
State FE	X	X	X	X	X	X
R^2	0.200	0.190	0.056	0.034	0.067	0.066
N	6422	10998	6422	10998	6422	10998

Notes: The values in parenthesis are standard error. *** is p-value ≤ 0.01, ** is p-value ≤ 0.05 and * is p-value ≤ 0.1. PA denotes the variable of physical activity. F&V is fruits and vegetables and S&C is snacks and chips. In each model, the dependent variable is the expenditure share times 100, and the income is measured with 100,000.

Table **2-12**: The OLS regressions of food expenditure shares against physical activity, sample split by BMI (Cont.).

	RSD		DSD		BW	
	BMI≥30	BMI<30	BMI≥30	BMI<30	BMI≥30	BMI<30
Intercept	2.00^{***}	2.11^{***}	2.24^{***}	0.56^{**}	1.23^{***}	0.81^{***}
	(0.39)	(0.30)	(0.36)	(0.26)	(0.22)	(0.22)
PA	-0.35^{***}	-0.44^{***}	-0.21^{**}	0.09	0.11^{*}	-0.04
	(0.11)	(0.07)	(0.10)	(0.06)	(0.06)	(0.05)
Male	0.76^{***}	0.71^{***}	0.49^{***}	0.27^{***}	0.05	-0.22^{***}
	(0.10)	(0.07)	(0.10)	(0.07)	(0.06)	(0.06)
College	-0.15	-0.26^{***}	0.39^{***}	0.36^{***}	-0.05	-0.12^{**}
	(0.10)	(0.07)	(0.09)	(0.06)	(0.05)	(0.05)
FTEmp	0.32^{***}	0.45^{***}	0.20^{**}	0.33^{***}	0.05	0.36^{***}
	(0.11)	(0.08)	(0.10)	(0.07)	(0.06)	(0.06)
PTEmp	0.25^{*}	0.35^{***}	0.07	0.29^{***}	0.10	0.02
	(0.14)	(0.10)	(0.13)	(0.09)	(0.08)	(0.07)
Income	-0.50^{***}	-0.40^{***}	0.12	0.39^{***}	0.04	-0.07
	(0.13)	(0.08)	(0.12)	(0.07)	(0.07)	(0.06)
Asian	0.01	-0.16	1.14^{**}	-0.53^{***}	-0.37	-0.30^{*}
	(0.60)	(0.22)	(0.55)	(0.20)	(0.34)	(0.17)
AfrAm	1.11^{***}	1.18^{***}	-1.33^{***}	-0.80^{***}	0.51^{***}	0.79^{***}
	(0.15)	(0.13)	(0.13)	(0.11)	(0.08)	(0.09)
Hispanic	1.22^{***}	0.27	-0.10	0.14	0.08	2.25^{***}
	(0.38)	(0.25)	(0.34)	(0.23)	(0.21)	(0.19)
Year FE	X	X	X	X	X	X
State FE	X	X	X	X	X	X
R^2	0.051	0.046	0.063	0.037	0.033	0.045
N	6422	10998	6422	10998	6422	10998

Notes: The values in parenthesis are standard error. *** is p-value ≤ 0.01, ** is p-value ≤ 0.05 and * is p-value ≤ 0.1. PA denotes the variable of physical activity. RSD denotes regular soft drinks, DSD denotes diet soft drink and BW denotes bottled water. In each model, the dependent variable is the expenditure share times 100, and the income is measured with 100,000.

We perform similar exercise to the simplified model of Equation (2.2), where we split the sample by physical activity and separately estimate the OLS models of food expenditure shares against obesity or BMI ≥ 30 with these two samples as shown in Table 2-13. The general trend is similar over models and samples for any one food category. Individuals who have obesity or taking any medications for obesity or BMI ≥ 30 spend more share of their grocery expenditure on S&C and DSD and spend less share on F&V and yogurts on average than the healthy individuals. The regression results for the models of RSD and BW are mixed. For those who are physically active, obese individuals spend more on RSD on average than non-obese ones, while for physically inactive individuals, obese ones spend less on RSD on average than non-obese ones. The coefficients of BMI and obesity are marginally significantly associated with expenditure share on BW for physical active individuals, but there is no significant association between expenditure share on BW and obesity or BMI ≥ 30 for physically inactive ones. These results regarding RSD and BW are similar with the main results presented in the result section if we consider 5 percent as the significance level.

We also split the sample by physical activeness and estimate the models of type 2 diabetes separately with two sub-samples. The results are shown in Table 2-14. The general trend is similar to the main results shown in the result section for most food categories such as S&C, yogurts, RSD, DSD and BW. The primary difference is that there is no significant association between type 2 diabetes and expenditure shares on F&V for physical active individuals. Comparing physical inactive and diabetic consumers, physical active consumers with type 2 diabetes spend more on S&C, RSD and DSD and less on yogurts on average. However, coefficients are difficult to compare over models, but the sign of coefficients is comparable over models.

Table **2-13**: The OLS regressions of food expenditure shares against obesity, sample split by physical activity.

	F&V				S&C			
	PA=1	PA=0	PA=1	PA=0	PA=1	PA=0	PA=1	PA=0
Intercept	12.91^{***}	10.07^{***}	13.10^{***}	10.15^{***}	7.04^{***}	7.14^{***}	6.90^{***}	7.05^{***}
	(0.60)	(0.37)	(0.60)	(0.37)	(0.57)	(0.43)	(0.57)	(0.43)
Obesity	-0.83^{***}	-0.28^{***}			0.18	0.26^{**}		
	(0.21)	(0.10)			(0.20)	(0.12)		
BMI≥30			-0.96^{***}	-0.32^{***}			0.59^{***}	0.35^{***}
			(0.16)	(0.09)			(0.15)	(0.10)
Male	-1.01^{***}	-0.79^{***}	-0.99^{***}	-0.79^{***}	0.46^{***}	0.47^{***}	0.46^{***}	0.47^{***}
	(0.15)	(0.09)	(0.15)	(0.09)	(0.14)	(0.11)	(0.14)	(0.11)
College	0.70^{***}	0.52^{***}	0.63^{***}	0.50^{***}	−0.13	-0.29^{***}	−0.09	-0.28^{***}
	(0.14)	(0.09)	(0.14)	(0.09)	(0.13)	(0.11)	(0.13)	(0.11)
FTEmp	-0.71^{***}	-0.88^{***}	-0.68^{***}	-0.88^{***}	0.29^{**}	0.53^{***}	0.28^{*}	0.53^{***}
	(0.15)	(0.10)	(0.15)	(0.10)	(0.15)	(0.12)	(0.15)	(0.12)
PTEmp	-0.95^{***}	-0.55^{***}	-0.95^{***}	-0.55^{***}	0.54^{***}	0.52^{***}	0.54^{***}	0.52^{***}
	(0.19)	(0.13)	(0.19)	(0.13)	(0.18)	(0.15)	(0.18)	(0.15)
Income	0.18	0.28^{**}	0.17	0.28^{**}	−0.23	−0.22	−0.21	−0.21
	(0.15)	(0.12)	(0.15)	(0.12)	(0.15)	(0.14)	(0.15)	(0.14)
Asian	1.01^{*}	1.58^{***}	0.96^{*}	1.53^{***}	1.00^{**}	0.86^{**}	1.04^{**}	0.92^{**}
	(0.53)	(0.33)	(0.53)	(0.33)	(0.50)	(0.39)	(0.50)	(0.39)
AfrAm	0.55^{**}	0.54^{***}	0.68^{***}	0.57^{***}	0.61^{***}	0.89^{***}	0.52^{**}	0.86^{***}
	(0.23)	(0.15)	(0.24)	(0.15)	(0.22)	(0.17)	(0.23)	(0.17)
Hispanic	-0.90^{*}	0.33	-0.90^{*}	0.34	0.08	−0.03	0.08	−0.03
	(0.53)	(0.33)	(0.53)	(0.33)	(0.51)	(0.38)	(0.51)	(0.38)
Year FE	X	X	X	X	X	X	X	X
State FE	X	X	X	X	X	X	X	X
R^2	0.207	0.172	0.209	0.173	0.038	0.033	0.040	0.033
N	6799	10621	6799	10621	6799	10621	6799	10621

Notes: The values in parenthesis are standard error. *** is p-value ≤ 0.01, ** is p-value ≤ 0.05 and * is p-value ≤ 0.1. PA denotes the variable of physical activity. Obesity denotes self-reported obesity or taking any Rx, OTC or dual medications for obesity. F&V is fruits and vegetables and S&C is snacks and chips. In each model, the dependent variable is the expenditure share times 100, and the income is measured with 100,000.

Table **2-13**: The OLS regressions of food expenditure shares against obesity, sample split by physical activity (Cont.).

	Yogurts				RSD			

	PA=1	PA=0	PA=1	PA=0	PA=1	PA=0	PA=1	PA=0
Intercept	3.28^{***}	2.25^{***}	3.45^{***}	2.35^{***}	2.00^{***}	1.89^{***}	1.98^{***}	1.90^{***}
	(0.56)	(0.35)	(0.56)	(0.35)	(0.34)	(0.32)	(0.34)	(0.32)
Obesity	-0.86^{***}	-0.34^{***}			0.24^{*}	-0.16^{*}		
	(0.20)	(0.10)			(0.12)	(0.09)		
BMI≥30			-0.90^{***}	-0.41^{***}			0.15^{*}	-0.12
			(0.15)	(0.08)			(0.09)	(0.08)
Male	-1.48^{***}	-1.14^{***}	-1.46^{***}	-1.14^{***}	0.49^{***}	0.85^{***}	0.48^{***}	0.86^{***}
	(0.14)	(0.09)	(0.14)	(0.09)	(0.09)	(0.08)	(0.09)	(0.08)
College	0.73^{***}	0.96^{***}	0.68^{***}	0.94^{***}	-0.20^{**}	-0.22^{***}	-0.19^{**}	-0.23^{***}
	(0.13)	(0.09)	(0.13)	(0.09)	(0.08)	(0.08)	(0.08)	(0.08)
FTEmp	0.46^{***}	0.40^{***}	0.48^{***}	0.40^{***}	0.26^{***}	0.48^{***}	0.25^{***}	0.48^{***}
	(0.14)	(0.10)	(0.14)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)
PTEmp	0.06	0.10	0.06	0.10	0.34^{***}	0.30^{***}	0.34^{***}	0.30^{***}
	(0.18)	(0.12)	(0.18)	(0.12)	(0.11)	(0.11)	(0.11)	(0.11)
Income	0.52^{***}	0.33^{***}	0.51^{***}	0.33^{***}	-0.38^{***}	-0.47^{***}	-0.38^{***}	-0.47^{***}
	(0.14)	(0.12)	(0.14)	(0.12)	(0.09)	(0.11)	(0.09)	(0.11)
Asian	-0.77	-0.12	-0.80	-0.18	-0.51^{*}	0.10	-0.51^{*}	0.09
	(0.49)	(0.32)	(0.49)	(0.32)	(0.30)	(0.29)	(0.30)	(0.29)
AfrAm	-1.19^{***}	-1.20^{***}	-1.07^{***}	-1.16^{***}	0.57^{***}	1.45^{***}	0.55^{***}	1.47^{***}
	(0.22)	(0.14)	(0.22)	(0.14)	(0.13)	(0.13)	(0.14)	(0.13)
Hispanic	-0.20	-0.15	-0.20	-0.14	-0.05	0.83^{***}	-0.05	0.83^{***}
	(0.50)	(0.31)	(0.50)	(0.31)	(0.30)	(0.28)	(0.30)	(0.28)
Year FE	X	X	X	X	X	X	X	X
State FE	X	X	X	X	X	X	X	X
R^2	0.062	0.058	0.065	0.059	0.039	0.044	0.039	0.044
N	6799	10621	6799	10621	6799	10621	6799	10621

Notes: The values in parenthesis are standard error. *** is p-value ≤ 0.01, ** is p-value ≤ 0.05 and * is p-value ≤ 0.1. PA denotes the variable of physical activity. Obesity denotes self-reported obesity or taking any Rx, OTC or dual medications for obesity. RSD denotes regular soft drinks. In each model, the dependent variable is the expenditure share times 100, and the income is measured with 100,000.

Table **2-13**: The OLS regressions of food expenditure shares against obesity, sample split by physical activity (Cont.).

	DSD				BW			
	PA=1	PA=0	PA=1	PA=0	PA=1	PA=0	PA=1	PA=0
Intercept	0.59^{*}	1.32^{***}	0.55	1.20^{***}	0.57^{*}	1.18^{***}	0.55^{*}	1.18^{***}
	(0.35)	(0.27)	(0.35)	(0.27)	(0.32)	(0.17)	(0.32)	(0.17)
Obesity	0.91^{***}	0.79^{***}			0.19^{*}	0.05		
	(0.13)	(0.07)			(0.11)	(0.05)		
BMI≥30			0.40^{***}	0.69^{***}			0.15^{*}	0.02
			(0.09)	(0.06)			(0.09)	(0.04)
Male	0.25^{***}	0.45^{***}	0.23^{***}	0.43^{***}	-0.05	-0.16^{***}	-0.05	-0.16^{***}
	(0.09)	(0.07)	(0.09)	(0.07)	(0.08)	(0.04)	(0.08)	(0.04)
College	0.28^{***}	0.35^{***}	0.31^{***}	0.39^{***}	0.01	-0.17^{***}	0.02	-0.17^{***}
	(0.08)	(0.07)	(0.08)	(0.07)	(0.08)	(0.04)	(0.08)	(0.04)

FTEmp	0.19^{**}	0.34^{***}	0.18^{*}	0.34^{***}	0.34^{***}	0.21^{***}	0.33^{***}	0.21^{***}
	(0.90)	(0.07)	(0.09)	(0.07)	(0.08)	(0.05)	(0.08)	(0.05)
PTEmp	0.18	0.24^{**}	0.17	0.23^{**}	0.02	0.10^{*}	0.01	0.10^{*}
	(0.11)	(0.09)	(0.11)	(0.09)	(0.10)	(0.06)	(0.10)	(0.06)
Income	0.50^{***}	0.18^{**}	0.49^{***}	0.17^{*}	−0.12	0.02	−0.11	0.01
	(0.09)	(0.09)	(0.09)	(0.09)	(0.08)	(0.06)	(0.08)	(0.06)
Asian	−0.46	−0.28	−0.47	−0.20	-0.58^{*}	−0.15	-0.57^{**}	−0.15
	(0.31)	(0.24)	(0.31)	(0.24)	(0.28)	(0.15)	(0.28)	(0.16)
AfrAm	-0.95^{***}	-1.05^{***}	-0.98^{***}	-1.12^{***}	0.88^{***}	0.52^{***}	0.86^{***}	0.52^{***}
	(0.14)	(0.11)	(0.14)	(0.11)	(0.13)	(0.07)	(0.13)	(0.07)
Hispanic	-0.55^{*}	0.37	-0.56^{*}	0.35	3.40^{***}	0.47^{***}	3.40^{***}	0.47^{***}
	(0.31)	(0.24)	(0.31)	(0.24)	(0.28)	(0.15)	(0.28)	(0.15)
Year FE	X	X	X	X	X	X	X	X
State FE	X	X	X	X	X	X	X	X
R^2	0.049	0.052	0.044	0.052	0.058	0.028	0.058	0.028
N	6799	10621	6799	10621	6799	10621	6799	10621

Notes: The values in parenthesis are standard error. *** is p-value ≤ 0.01, ** is p-value ≤ 0.05 and * is p-value ≤ 0.1. PA denotes the variable of physical activity. Obesity denotes self-reported obesity or taking any Rx, OTC or dual medications for obesity. DSD denotes diet soft drinks and BW denotes bottled water. In each model, the dependent variable is the expenditure share times 100, and the income is measured with 100,000.

Table **2-14**: The OLS regressions of food expenditure shares against type 2 diabetes, sample split by physical activity.

	F&V		S&C		Yogurts	
	PA=1	PA=0	PA=1	PA=0	PA=1	PA=0
Intercept	12.87^{***}	10.07^{***}	6.99^{***}	7.14^{***}	3.31^{***}	2.32^{***}
	(0.60)	(0.37)	(0.57)	(0.43)	(0.56)	(0.35)
Diabetes 2	−0.15	-0.26^{**}	0.93^{***}	0.28^{**}	-1.29^{***}	-0.62^{***}
	(0.25)	(0.12)	(0.24)	(0.14)	(0.23)	(0.11)
Male	-0.98^{***}	-0.76^{***}	0.42^{***}	0.44^{***}	-1.41^{***}	-1.08^{***}
	(0.15)	(0.09)	(0.14)	(0.11)	(0.14)	(0.09)
College	0.69^{***}	0.51^{***}	−0.09	-0.28^{***}	0.68^{***}	0.94^{***}
	(0.14)	(0.09)	(0.13)	(0.11)	(0.13)	(0.09)
FTEmp	-0.71^{***}	-0.90^{***}	0.35^{**}	0.56^{***}	0.39^{***}	0.35^{***}
	(0.15)	(0.10)	(0.15)	(0.12)	(0.14)	(0.10)
PTEmp	-0.94^{***}	-0.55^{***}	0.57^{***}	0.52^{***}	0.03	0.09
	(0.19)	(0.13)	(0.18)	(0.15)	(0.18)	(0.12)
Income	0.20	0.28^{**}	−0.23	−0.21	0.54^{***}	0.32^{***}
	(0.15)	(0.12)	(0.14)	(0.14)	(0.14)	(0.12)
Asian	1.05^{**}	1.62^{***}	0.96^{*}	0.82^{**}	−0.69	−0.06
	(0.53)	(0.33)	(0.50)	(0.39)	(0.49)	(0.32)
AfrAm	0.54^{**}	0.56^{***}	0.52^{**}	0.86^{***}	-1.08^{***}	-1.15^{***}
	(0.24)	(0.15)	(0.23)	(17)	(0.22)	(0.14)
Hispanic	-0.88^{*}	0.35	0.10	−0.05	−0.22	−0.11

	(0.53)	(0.33)	(0.51)	(0.38)	(0.50)	(0.31)
Year FE	X	X	X	X	X	X
State FE	X	X	X	X	X	X
R^2	0.205	0.172	0.040	0.032	0.064	0.060
N	6799	10621	6799	10621	6799	10621

Notes: The values in parenthesis are standard error. *** is p-value $\leq$ 0.01, ** is p-value $\leq$ 0.05 and * is p-value $\leq$ 0.1. PA denotes the variable of physical activity. Diabetes 2 denotes self-reported type 2 diabetes mellitus or taking any Rx, OTC or dual diabetic medications. F&V denotes fruits and vegetables and S&C denotes snacks and chips. In each model, the dependent variable is the expenditure share times 100, and the income is measured with 100,000.

Table **2-14**: The OLS regressions of food expenditure shares against type 2 diabetes, sample split by physical activity (Cont.).

	RSD		DSD		BW	
	PA=1	PA=0	PA=1	PA=0	PA=1	PA=0
Intercept	2.03***	1.95***	0.57	1.24***	0.59*	1.20***
	(0.34)	(0.32)	(0.35)	(0.27)	(0.32)	(0.17)
Diabetes 2	−0.28**	−0.40***	1.13***	1.12***	−0.05	−0.02
	(0.14)	(0.10)	(0.15)	(0.09)	(0.13)	(0.06)
Male	0.49***	0.89***	0.18**	0.33***	−0.05	−0.16***
	(0.09)	(0.08)	(0.09)	(0.07)	(0.08)	(0.04)
College	−0.21***	−0.23***	0.33***	0.38***	0.01	−0.17***
	(0.08)	(0.08)	(0.08)	(0.07)	(0.08)	(0.04)
FTEmp	0.24***	0.45***	0.25***	0.43***	0.33***	0.21***
	(0.09)	(0.09)	(0.09)	(0.07)	(0.08)	(0.05)
PTEmp	0.32***	0.29**	0.20*	0.25***	0.01	0.10
	(0.11)	(0.11)	(0.11)	(0.09)	(0.10)	(0.06)
Income	−0.38***	−0.48***	0.48***	0.19**	−0.12	0.01
	(0.09)	(0.11)	(0.09)	(0.09)	(0.08)	(0.06)
Asian	−0.51*	0.13	−0.54*	−0.39	−0.58**	−0.16
	(0.30)	(0.29)	(0.31)	(0.24)	(0.28)	(0.15)
AfrAm	0.61***	1.48***	−1.04***	−1.15***	0.89***	0.52***
	(0.14)	(0.13)	(0.14)	(0.11)	(0.13)	(0.07)
Hispanic	−0.06	0.85***	−0.54*	0.30	3.39***	0.47***
	(0.30)	(0.28)	(0.31)	(0.24)	(0.28)	(0.15)
Year FE	X	X	X	X	X	X
State FE	X	X	X	X	X	X
R^2	0.039	0.045	0.050	0.056	0.058	0.028
N	6799	10621	6799	10621	6799	10621

Notes: The values in parenthesis are standard error. *** is p-value ≤ 0.01, ** is p-value ≤ 0.05 and * is p-value ≤ 0.1. PA denotes the variable of physical activity. Diabetes 2 denotes self-reported type 2 diabetes mellitus or taking any Rx, OTC or dual diabetic medications. RSD denotes regular soft drinks, DSD denotes diet soft drinks and BW denotes bottled water. In each model, the dependent variable is the expenditure share times 100, and the income is measured with 100,000.

Discussions and Conclusions

In this study, we investigate the links between the expenditure shares of several important food categories and three health-related metrics (i.e. physical activity, obesity and type 2 diabetes) based on household-level data over multiple years. The expenditure share provides us insights of not just people's consumption on a food category as many existing studies investigated, but also people's willingness to spend on a food category. Including both healthy foods and unhealthy foods allow us to observe the disparity in consumers' diet preference between consumers with different health status or undertaking different health-related practice. Differed from literature that only focus on one food category or use survey data, this study employs consumer data over five years. The multiple-year comprehensive consumer data allow us to capture consumer behavior in an entire year as well as the shift of preference over years and allow us to acquire consumer purchases of any food categories, healthy or unhealthy. The health data, IRI Medprofiler panel, allow us to obtain information regarding physical activity, obesity and type 2 diabetes on yearly basis over multiple years and allow us to observe the disparity in dietary choice between patients taking medications and the ones suffering a disease but not taking any medical solution yet. These advantages in term of unique data used for this study do not just fill in the niches in literature, but also provide comprehensive information about the relationship between the healthfulness of food consumption and physical activity, obesity, type 2 diabetes or medication usages.

Regarding the relationship between physical activity and dietary choice, our findings align with previous literature, for example, Gillman et al. (2001) and Lin et al. (2013). We find that physically active individuals would spend more on healthy foods such as fruits and vegetables and spend less on unhealthy foods such as snacks and chips and regular and diet soft drinks. Our findings regarding yogurts seem contradictory with Lopez-Garcia et al. (2015) where they did not find significantly association between the frequency of consuming yogurts during a week and physical activity. However, we are interested in expenditure share on yogurts over the year, where Lopez-Garcia et al. (2015) focus on frequency of yogurt purchases in a week. These two metrics might be associated, but expenditure share can better capture consumers' potential total consumption in a year and their willingness to spend for yogurts, and most importantly it counts product variations in size and package and counts the impact of stockpiling effect that has been discussed for storable products in literature such as Hendel and Nevo (2013), Wang (2015) and Wang et al. (2017).

For the results of obesity and expenditure shares, we find the associations of expenditure shares on some food categories with obesity, a BMI $\geq$ 30 and a BMI $\geq$ 40, but only using one variable to denote obesity or BMI cannot capture obesity, BMI and medication usage together in one model, and the key coefficients are barely comparable over models. Hence, we develop from the simplified model of obesity and define a vector of health-related variables that allow us specify the disparity between self-reported obesity and BMI and the disparity between obesity patients taking Rx, OTC or dual medications and obese individuals suffering obesity but not any taking medications. To prevent and reduce obesity, taking medications sometimes is not a must. Regularly doing exercise and maintaining healthy dietary patterns can also effectively reduce weight gradually (Skender et al., 1996; Sallis et al., 2009). However, differed from individuals who are suffering obesity but not taking medications, we assume the individuals who take medications have a strong willingness to improve health. Because of these complicated relationship between individuals who are not sufferers but taking medications, who are sufferers but not taking medications and who are sufferers and actively taking medications, the results of regressions are mixed. Furthermore, we also consider BMI as a proxy to identify obesity even Rippe, et al. (2012) and Cawley (2015) suggest this approach fails to distinguish lean mass from fat mass, but compared with self-reported obesity, using BMI is a more objective method.

Some results in the regressions of expenditure shares against obesity line up with previous literature, but since we incorporate several variables rather than just one variable of obesity, we obtain more information regarding the disparity in consumer preference on healthy and unhealthy foods between people with different characteristics regarding obesity, BMI and medication usage. In literature, a healthy dietary pattern benefit health, reduce glycemic index and improve weight loss (Skov et al., 1999; Spieth et al., 2000; Foster et al., 2003; Samaha et al., 2003; Yancy et al., 2004), and obese youths or adults tend to consume more sugar-sweetened beverages and potato chips (Ludwig et al., 2001; Gillis and Bar-Or, 2003; Swinburn et al., 2004; Hu and Malik, 2010; Della Torre et al., 2015; Anari et al., 2017). Here in this study, we find that for individuals who are self-reported obese or having BMI ≥ 30 are likely to purchase more unhealthy foods such as snacks and chips and diet soft drinks and purchase less healthy foods such as fruits and vegetables and yogurts. However, we also find that individuals who self-reported having obesity and taking medications but having BMI < 30 spend less on snacks and chips, even lower than the baseline case (i.e. individuals who are self-reported healthy and BMI < 30), where individuals who are self-reported sufferers and not taking any medications and having BMI ≥ 30 and who are self-reported healthy but having BMI ≥ 30 are likely to spend more on snacks and chips than the baseline case. Those individuals with BMI < 30 in most cases can be identified as non-obese, but taking medications even if they have BMI <30 might because their strong intention to maintain healthy diet and prevent excessive intake of some known unhealthy foods such as snacks and chips. The statistically insignificant coefficients at 95 percent significant level in the models of yogurts, regular and diet soft drinks and bottled water might also demonstrate that point. However, some key coefficients are statistically significant in the model with year and state fixed effect, but are not significant while excluding all fixed effects. The variations of temporal and geographic factors can also affect consumer behavior.

Based on the F statistics of the regressions regarding obesity, we find there is a significant disparity in expenditure shares of snacks and chips between individuals whose BMI < 30 and the ones whose have BMI ≥ 30. The expenditure share of yogurts for individuals having self-reported obesity but not taking medications and having BMI ≥ 30 is significantly differed from individuals who are self-reported obesity sufferers not taking medications and having BMI < 30. For the F statistics for the regression of regular soft drinks against obesity variables, several pairs of key metrics are significant. This denotes that the disparity in preference on regular soft drinks over groups by obesity and BMI is considerable. Statistical significance in F statistics is not observed in models of fruits and vegetables, diet soft drinks and bottled waters. This shows that among individuals who believe they have obesity concern or have high BMI, the disparity in willingness to spend on these three food categories over groups is minimal.

However, some results in this section might not exactly line with literature. Wang et al. (2017) find a market with higher incidence of obesity tends to be more price sensitive in soft drinks than a market with lower incidence of obesity, but they do not distinguish diet and regular soft drinks. Our study also finds that obese individuals or individuals with BMI ≥ 30 are likely to consume more regular soft drinks, but we did not observe consistent significance over groups, and our model does not allow us to observe price elasticity. Bandini et al. (1999), Ludwig et al. (2001) and Gillis and Bar-Or (2003) demonstrate the association between soft drinks intake and obesity for children and adolescents, where our study only investigates American adults. Other studies also demonstrate the association between the consumption of soft drinks and obesity, for example, Della Torre et al. (2015). Our study does observe a positive relation of obesity with expenditure shares on regular soft drinks but it is not statistically significant. On the other side, regarding the diet soft drinks, Patel (2012) find that individuals with higher weight prefer diet sodas. This lines with our study where we also find the positive association between expenditure share on diet soft drinks and obesity or BMI ≥ 30. When considering the models of both regular and diet soft drinks, the results might imply that the willingness to spend on regular soft drinks is similar between obese and healthy individuals, but obese individuals, particularly who have BMI ≥ 30, have a stronger preference on diet soft drinks, meant to prevent excessive sugar intake by increasing the consumption on diet alternatives. This might also because consumers' preference on regular soft drinks are consistent over time and groups. Therefore, the impact of a tax aiming to lower the consumption of regular soft drinks might be minimal.

Type 2 diabetes, or say insulin non-dependent diabetes mellitus, differs from obesity in nature. Only a small portion of type 2 diabetes patients (i.e. about ten percent in our data) choose not to take any medications or medical treatments, while the majority of obese individuals (i.e. about 85 percent in our data) have not been actively seeking medical solutions. Medications or insulin might not be a must for some diabetic patients at some points of time as long as they could maintain blood glucose level within an appropriate range by restrictive healthy diet and physical activity, but prescription oral medications or insulin are needed when lowering glucose level to the target level at the beginning is needed or when the situation gets worse (ADA, 2015, October 27). In our study, we perform simplified models to identify the link between type 2 diabetes and expenditure shares on food categories, but this model fails to capture the impact of diabetic medication usage. A developed model with two dichotomous variables that capture diabetic patients taking medications and the ones not taking medications, respectively, is estimate. We find that in most models except the model of diet soft drinks, the coefficient of having type 2 diabetes but not taking any medications are not statistically significant, meant that the individuals who have aware of suffering type 2 diabetes but not taking any medications for some reasons behave like the healthy people, except spending more on diet soft drinks than healthy people. When they report they have type 2 diabetes, we assume they already know they have been diagnosed with type 2 diabetes. Therefore, not taking medications might because, first, it has not been a serious health concern for them, and second, possibly according the prescription of doctor's medications are not needed at the moment. For type 2 diabetes patients who are taking Rx, OTC or dual medications, their expenditure shares on fruits and vegetables, yogurts and regular soft drinks are significantly lower than healthy individuals, and their expenditures shares on snacks and chips and diet soft drinks are significantly higher than healthy individuals. These results are mixed in term of their overall healthfulness of food purchases. On one side, they spend more on healthy alternative of soft drinks (i.e. diet soft drinks) than unhealthy alternative (i.e. regular soft drinks). On another side, more unhealthy foods (i.e. snacks and chips) are purchased, while less healthy foods (i.e. fruits and vegetables and yogurts) are purchased. It is difficult for us to explore what caused the type 2 diabetes mellitus of these patients and why they are maintaining an unhealthy dietary pattern but when they are facing the choice of a particular food category, they prefer the healthy alternative over the unhealthy one. This contradiction needs further exploration. In literature, we can know that a better dietary pattern lowers blood glucose level and prevent diabetes, for example, Salas-Salvado et al. (2011), and we know the positive association between lower risk of type 2 diabetes and higher overall diet quality (Fung et al., 2007; Liese et al., 2009; De Koning et al., 2011). However, there are few studies may observe the contraction observed in our study.

This study provides comprehensive analysis about the association between consumers' spending on multiple food categories and consumers' health or health-related behaviors. The innovative method provides some new information that has not identified in literature. In the meanwhile, it also inspires us to find research questions needed for further studies. Even though we show a potential link between two main weight control strategies, exercise and healthy diet, their substitution pattern is unknown. We do observe physically active individuals spend less on regular and diet soft drinks than physically inactive ones. But consumers might spend more on substitute products of soft drinks, for instance, energy drinks and sports drinks. Energy drinks contain caffeine and added sugars more than daily recommended levels, and excessive intake can cause death and caffeine addiction (CDC, 2016, March 22; Chen, 2018). However, if we only look at the results of the regressions, we might find healthy diet and physical activity are complement, the improvement in one associated with the increase in another, but it is difficult to conclude without the support of rigorous theoretical models.

This study fills the niche of literature that links of diet with obesity and type 2 diabetes by investigating how consumers taking medications have different preference on healthy and unhealthy foods from consumers who are not taking medications but suffering a disease and who are healthy. However, as discussed above, we do observe some contradictory findings with literature. Further studies can be conducted using a different dataset such as NHANES data to explore the relationship between diet and medication usage.

References

Allcott, H., Diamond, R., Dube, J-P., Handbury, J., Rahkovsky, I., & Schnell, M. (2019). Food Deserts and the Causes of Nutritional Inequality. *Quarterly Journal of Economics*, 134(4).

American Diabetes Association (ADA). (2015, October 27). *Facts About Type 2*. Retrieved on June 20, 2019 from http://www.diabetes.org/diabetes-basics/type-2/facts-about-type-2.html.

Anari, R., Amani, R., & Veissi, M. (2017). Sugar-sweetened beverages consumption is associated with abdominal obesity risk in diabetic patients. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 11, S675-S678.

Bandini, L. G., Vu, D., Must, A., Cyr, H., Goldberg, A., & Dietz, W. H. (1999). Comparison of high-calorie, low-nutrient-dense food consumption among obese and non-obese adolescents. *Obesity Research*, 7(5), 438-443.

Barnard, N. D., Cohen, J., Jenkins, D. J., Turner-McGrievy, G., Gloede, L., Green, A., & Ferdowsian, H. (2009). A low-fat vegan diet and a conventional diabetes diet in the treatment of type 2 diabetes: a randomized, controlled, 74-wk clinical trial. *The American Journal of Clinical Nutrition*, 89(5), 1588S-1596S.

Barnard, N. D., Cohen, J., Jenkins, D. J., Turner-McGrievy, G., Gloede, L., Jaster, B., ... & Talpers, S. (2006). A low-fat vegan diet improves glycemic control and cardiovascular risk factors in a randomized clinical trial in individuals with type 2 diabetes. *Diabetes Care*, 29(8), 1777-1783.

Boden, G., Sargrad, K., Homko, C., Mozzoli, M., & Stein, T. P. (2005). Effect of a low-carbohydrate diet on appetite, blood glucose levels, and insulin resistance in obese patients with type 2 diabetes. *Annals of Internal Medicine*, 142(6), 403-411.

Bray, G. A., Nielsen, S. J., & Popkin, B. M. (2004). Consumption of high-fructose corn syrup in beverages may play a role in the epidemic of obesity. *The American Journal of Clinical Nutrition*, 79(4), 537-543.

Bronnenberg, B. J., Dubé, J. P. H., & Gentzkow, M. (2012). The evolution of brand preferences: Evidence from consumer migration. *American Economic Review*, 102(6), 2472-2508.

Burkhauser, R. V., Cawley, J., & Schmeiser, M. D. (2009). *Differences in the US trends in the prevalence of obesity based on body mass index and skinfold thickness* (No. w15005). National Bureau of Economic Research.

Calle, E. E., Rodriguez, C., Walker-Thurmond, K., & Thun, M. J. (2003). Overweight, obesity, and mortality from cancer in a prospectively studied cohort of US adults. *New England Journal of Medicine*, 348(17), 1625-1638.

Cawley, J. (2015). An economy of scales: A selective review of obesity's economic causes, consequences, and solutions. *Journal of Health Economics*, 43, 244-268.

Centers for Disease Control and Prevention (CDC). (2016, March 22). *The Buzz on Energy Drinks*. Retrieved on June 20, 2019 from https://www.cdc.gov/healthyschools/nutrition/energy.htm.

Centers for Disease Control and Prevention (CDC). (2017, April 11). *Defining Adult Overweight and Obesity*. Retrieved on June 20, 2019 from https://www.cdc.gov/obesity/adult/defining.html.

Centers for Disease Control and Prevention (CDC). (2017, August 29). *Adult Obesity Causes & Consequences*. Retrieved on June 20, 2019 from https://www.cdc.gov/obesity/adult/causes.html.

Centers for Disease Control and Prevention (CDC). (2018, April 24). *Get Active!* Retrieved on June 23, 2019 from https://www.cdc.gov/diabetes/managing/active.html.

Centers for Disease Control and Prevention (CDC). (2019, March 4). *Manage Blood Sugar*. Retrieved on June 23, 2019 from https://www.cdc.gov/diabetes/managing/manage-blood-sugar.html.

Centers for Disease Control and Prevention (CDC). (2019, March 20). *Eat Well!* Retrieved on June 23, 2019 from https://www.cdc.gov/diabetes/managing/eat-well.html.

Centers for Disease Control and Prevention (CDC). (2019, May 30a). *National Health Interview Survey: Early Release of Selected Estimates Based on Data From the 2018 National Health Interview Survey*. Retrieved on June 20, 2019 from https://www.cdc.gov/nchs/nhis/releases/released201905.htm#7a.

Centers for Disease Control and Prevention (CDC). (2019, May 30b). *Type 2 Diabetes*. Retrieved on June 20, 2019 from https://www.cdc.gov/diabetes/basics/type2.html.

Chen, S. N., Shogren, J. F., Orazem, P. F., & Crocker, T. D. (2002). Prices and health: identifying the effects of nutrition, exercise, and medication choices on blood pressure. *American Journal of Agricultural Economics*, 84(4), 990-1002.

Chen, X. (2018). The Rational Addiction, Health Information and Dynamic Demand of Energy Drinks. 2018 Annual Meeting, August 5-7, Washington, D.C. 273979, Agricultural and Applied Economics Association.

De Koning, L., Chiuve, S. E., Fung, T. T., Willett, W. C., Rimm, E. B., & Hu, F. B. (2011). Diet-quality scores and the risk of type 2 diabetes in men. *Diabetes Care*, 34(5), 1150-1156.

Della Torre, S. B., Keller, A., Depeyre, J. L., & Kruseman, M. (2016). Sugar-sweetened beverages and obesity risk in children and adolescents: a systematic analysis on how methodological quality may influence conclusions. *Journal of the Academy of Nutrition and Dietetics*, 116(4), 638-659.

Dong, Z., Gao, Z., & McFadden, B. (2017). Impact of food choice on sodium intake patterns from multiple NHANES surveys. *Appetite*, 109, 144-153.

Elhayany, A., Lustman, A., Abel, R., Attal-Singer, J., & Vinker, S. (2010). A low carbohydrate Mediterranean diet improves cardiovascular risk factors and diabetes control among overweight patients with type 2 diabetes mellitus: a 1-year prospective randomized intervention study. *Diabetes, Obesity and Metabolism*, 12(3), 204-209.

Esposito, K., Maiorino, M. I., Ceriello, A., & Giugliano, D. (2010). Prevention and control of type 2 diabetes by Mediterranean diet: a systematic review. *Diabetes Research and Clinical Practice*, 89(2), 97-102.

Esposito, K., Maiorino, M. I., Ciotola, M., Di Palo, C., Scognamiglio, P., Gicchino, M., ... & Giugliano, D. (2009). Effects of a Mediterranean-style diet on the need for antihyperglycemic drug therapy in patients with newly diagnosed type 2 diabetes: a randomized trial. *Annals of Internal Medicine*, 151(5), 306-314.

Fan, M., & Jin, Y. (2013). Do neighborhood parks and playgrounds reduce childhood obesity? *American Journal of Agricultural Economics*, 96(1), 26-42.

Finkelstein, E. A., Trogdon, J. G., Cohen, J. W., & Dietz, W. (2009). Annual medical spending attributable to obesity: payer-and service-specific estimates. *Health Affairs*, 28(5), w822-w831.

Foster, G. D., Wyatt, H. R., Hill, J. O., McGuckin, B. G., Brill, C., Mohammed, B. S., ... & Klein, S. (2003). A randomized trial of a low-carbohydrate diet for obesity. *New England Journal of Medicine*, 348(21), 2082-2090.

Fung, T. T., McCullough, M., Van Dam, R. M., & Hu, F. B. (2007). A prospective study of overall diet quality and risk of type 2 diabetes in women. *Diabetes Care*, 30(7), 1753-1757.

Gannon, M. C., & Nuttall, F. Q. (2006). Control of blood glucose in type 2 diabetes without weight loss by modification of diet composition. *Nutrition & Metabolism*, 3(1), 16.

Gillis, L. J., & Bar-Or, O. (2003). Food away from home, sugar-sweetened drink consumption and juvenile obesity. *Journal of the American College of Nutrition*, 22(6), 539-545.

Gillman, M. W., Pinto, B. M., Tennstedt, S., Glanz, K., Marcus, B., & Friedman, R. H. (2001). Relationships of physical activity with dietary behaviors among adults. *Preventive Medicine*, 32(3), 295-301.

Hales, C. M., Carroll, M. D., Fryar, C. D., & Ogden, C. L. (2017). *Prevalence of obesity among adults and youth: United States, 2015–2016*.

Halton, T. L., Liu, S., Manson, J. E., & Hu, F. B. (2008). Low-carbohydrate-diet score and risk of type 2 diabetes in women. *The American Journal of Clinical Nutrition*, 87(2), 339-346.

Hendel, I., & Nevo, A. (2013). Intertemporal price discrimination in storable goods markets. *American Economic Review*, 103(7), 2722-51.

Hu, F. B., & Malik, V. S. (2010). Sugar-sweetened beverages and risk of obesity and type 2 diabetes: epidemiologic evidence. *Physiology & Behavior*, 100(1), 47-54.

Hu, F. B., Manson, J. E., Stampfer, M. J., Colditz, G., Liu, S., Solomon, C. G., & Willett, W. C. (2001). Diet, lifestyle, and the risk of type 2 diabetes mellitus in women. *New England Journal of Medicine*, 345(11), 790-797.

Hubert, H. B., Feinleib, M., McNamara, P. M., & Castelli, W. P. (1983). Obesity as an independent risk factor for cardiovascular disease: a 26-year follow-up of participants in the Framingham Heart Study. *Circulation*, 67(5), 968-977.

Jenkins, D. J., Kendall, C. W., Marchie, A., Jenkins, A. L., Augustin, L. S., Ludwig, D. S., ... & Anderson, J. W. (2003). Type 2 diabetes and the vegetarian diet. *The American Journal of Clinical Nutrition*, 78(3), 610S-616S.

Kahleova, H., Matoulek, M., Malinska, H., Oliyarnik, O., Kazdova, L., Neskudla, T., ... & Pelikanova, T. (2011). Vegetarian diet improves insulin resistance and oxidative stress markers more than conventional diet in subjects with Type 2 diabetes. *Diabetic Medicine*, 28(5), 549-559.

Kahn, S. E., Hull, R. L., & Utzschneider, K. M. (2006). Mechanisms linking obesity to insulin resistance and type 2 diabetes. *Nature*, 444(7121), 840.

Liese, A. D., Nichols, M., Sun, X., D'Agostino, R. B., & Haffner, S. M. (2009). Adherence to the DASH Diet is inversely associated with incidence of type 2 diabetes: the insulin resistance atherosclerosis study. *Diabetes Care*, 32(8), 1434-1436.

Lin, C. T. J., Gao, Z., & Lee, J. Y. (2013). Associations between self-reported weight management methods with diet quality as measured by the Healthy Eating Index–2005. *Preventive Medicine*, 57(3), 238-243.

Lopez-Garcia, E., Leon-Muñoz, L., Guallar-Castillon, P., & Rodríguez-Artalejo, F. (2015). Habitual yogurt consumption and health-related quality of life: a prospective cohort study. *Journal of the Academy of Nutrition and Dietetics*, 115(1), 31-39.

Ludwig, D. S., Peterson, K. E., & Gortmaker, S. L. (2001). Relation between consumption of sugar-sweetened drinks and childhood obesity: a prospective, observational analysis. *The Lancet*, 357(9255), 505-508.

Oster, E. (2018). Diabetes and diet: purchasing behavior change in response to health information. *American Economic Journal: Applied Economics*, 10(4), 308-48.

Øvrum, A. (2011). Socioeconomic status and lifestyle choices: evidence from latent class analysis. *Health Economics*, 20(8), 971-984.

Parker, B., Noakes, M., Luscombe, N., & Clifton, P. (2002). Effect of a high-protein, high–monounsaturated fat weight loss diet on glycemic control and lipid levels in type 2 diabetes. *Diabetes Care*, 25(3), 425-430.

Patel, K.R. (2012). Essays in Industrial Organization. PhD dissertation, Northwestern University.

Pi-Sunyer, F. X., Becker, D. M., Bouchard, C., Carleton, R. A., Colditz, G. A., Dietz, W. H., ... & Higgins, M. (1998). Clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults: Executive summary. *American Journal of Clinical Nutrition*, 68(4), 899-917.

Rippe, J. M., McInnis, K. J., & Melanson, K. J. (2001). Physician involvement in the management of obesity as a primary medical condition. *Obesity Research*, 9(S11), 302S-311S.

Salas-Salvadó, J., Bulló, M., Babio, N., Martínez-González, M. Á., Ibarrola-Jurado, N., Basora, J., ... & Ruiz-Gutiérrez, V. (2011). Reduction in the incidence of type 2 diabetes with the Mediterranean diet: results of the PREDIMED-Reus nutrition intervention randomized trial. *Diabetes Care*, 34(1), 14-19.

Sallis, J. F., Story, M., & Lou, D. (2009). Study designs and analytic strategies for environmental and policy research on obesity, physical activity, and diet: recommendations from a meeting of experts. *American Journal of Preventive Medicine*, 36(2), S72-S77.

Samaha, F. F., Iqbal, N., Seshadri, P., Chicano, K. L., Daily, D. A., McGrory, J., ... & Stern, L. (2003). A low-carbohydrate as compared with a low-fat diet in severe obesity. *New England Journal of Medicine*, 348(21), 2074-2081.

Schröder, H., Fïto, M., & Covas, M. I. (2007). Association of fast food consumption with energy intake, diet quality, body mass index and the risk of obesity in a representative Mediterranean population. *British Journal of Nutrition*, 98(6), 1274-1280.

Skender, M. L., Goodrick, G. K., Del Junco, D. J., Reeves, R. S., Darnell, L., Gotto Jr, A. M., & Foreyt, J. P. (1996). Comparison of 2-year weight loss trends in behavioral treatments of obesity: diet, exercise, and combination interventions. *Journal of the American Dietetic Association*, 96(4), 342-346.

Skov, A. R., Toubro, S., Rønn, B., Holm, L., & Astrup, A. (1999). Randomized trial on protein vs carbohydrate in ad libitum fat reduced diet for the treatment of obesity. *International Journal of Obesity*, 23(5), 528.

Spieth, L. E., Harnish, J. D., Lenders, C. M., Raezer, L. B., Pereira, M. A., Hangen, S. J., & Ludwig, D. S. (2000). A low–glycemic index diet in the treatment of pediatric obesity. *Archives of Pediatrics & Adolescent Medicine*, 154(9), 947-951.

Swinburn, B. A., Caterson, I., Seidell, J. C., & James, W. P. T. (2004). Diet, nutrition and the prevention of excess weight gain and obesity. *Public Health Nutrition*, 7(1a), 123-146.

Tonstad, S., Butler, T., Yan, R., & Fraser, G. E. (2009). Type of vegetarian diet, body weight, and prevalence of type 2 diabetes. *Diabetes Care*, 32(5), 791-796.

Van Gaal, L. F., Mertens, I. L., & Christophe, E. (2006). Mechanisms linking obesity with cardiovascular disease. *Nature*, 444(7121), 875.

Yanovski, S. Z., & Yanovski, J. A. (2014). Long-term drug treatment for obesity: a systematic and clinical review. *The Journal of the American Medical Association*, 311(1), 74-86.

Vucenik, I., & Stains, J. P. (2012). Obesity and cancer risk: evidence, mechanisms, and recommendations. *Annals of the New York Academy of Sciences*, 1271(1), 37-43.

Wang, E. Y. (2015). The impact of soda taxes on consumer welfare: implications of storability and taste heterogeneity. *The RAND Journal of Economics*, 46(2), 409-441.

Wang, E., Rojas, C., & Colantuoni, F. (2017). Heterogeneous Behavior, Obesity, and Storability in the Demand for Soft Drinks. *American Journal of Agricultural Economics*, 99(1), 18-33.

Westman, E. C., Yancy, W. S., Mavropoulos, J. C., Marquart, M., & McDuffie, J. R. (2008). The effect of a low-carbohydrate, ketogenic diet versus a low-glycemic index diet on glycemic control in type 2 diabetes mellitus. *Nutrition & Metabolism*, 5(1), 36.

Weyer, C., Funahashi, T., Tanaka, S., Hotta, K., Matsuzawa, Y., Pratley, R. E., & Tataranni, P. A. (2001). Hypoadiponectinemia in obesity and type 2 diabetes: close association with insulin resistance and hyperinsulinemia. *The Journal of Clinical Endocrinology & Metabolism*, 86(5), 1930-1935.

Wilson, M. M., Reedy, J., & Krebs-Smith, S. M. (2016). American diet quality: Where it is, where it is heading, and what it could be. *Journal of the Academy of Nutrition and Dietetics*, 116(2), 302-310.

Yancy, W. S., Foy, M., Chalecki, A. M., Vernon, M. C., & Westman, E. C. (2005). A low-carbohydrate, ketogenic diet to treat type 2 diabetes. *Nutrition & Metabolism*, 2(1), 34.

Yancy, W. S., Olsen, M. K., Guyton, J. R., Bakst, R. P., & Westman, E. C. (2004). A low-carbohydrate, ketogenic diet versus a low-fat diet to treat obesity and hyperlipidemia: a randomized, controlled trial. *Annals of Internal Medicine*, 140(10), 769-777.

Yang, Q., Zhang, Z., Gregg, E. W., Flanders, W. D., Merritt, R., & Hu, F. B. (2014). Added sugar intake and cardiovascular diseases mortality among US adults. *JAMA Internal Medicine*, 174(4), 516-524.

Chapter 3

The Associations of Physical Activity and Obesity with Consumer Demand on Yogurts

Introduction

The upward trend of obese population in U.S. during the past decade incurs considerable medical costs and financial burden to American households and the society every year (Finkelstein et al., 2009; Hales et al., 2017). Public concerns for obesity because of its association with the incidence of type 2 diabetes mellitus, cardiovascular disease and cancer (Hubert et al., 1983; Weyer et al., 2001; Calle et al., 2003; Kahn et al., 2006; Van Gaal et al., 2006; Vucenik et al., 2014). People are encouraged to maintain a healthy diet and regularly do physical activity in order to reduce weight and blood glycemic level and prevent obesity or other relevant diseases (Skender et al. 1996; Pi-Sunyer, et al., 1998; Sallis et al. 2009; Lin et al., 2013). Differed from those self-motivated weight control strategies, governments have implemented or have been proposing various policies aiming to reduce unhealthy foods or unhealthy nutrients intake, nudge people to consume more fruits and vegetables or reward individuals for doing more exercise (Cawley, 2015). However, the effectiveness of these government-driven health policies is mixed (Wang, 2015; Debnam, 2017). This inspires us to approach this question from a different angle, investigating the effectiveness of a self-motivated weight control strategy, physical activity, on consumer behavior towards a product with heterogenous health-related product attributes. In addition, we also investigate the association of obesity with product choice.

Regularly doing physical activity is an effective way to reduce weight and improve health (Skender et al. 1996; Sallis et al. 2009), but how physical activity associates with product choice is not clear. The hypothesis of this study is that individuals who are physically active prefer products with a better diet quality. Bonanno (2012) finds that in markets with more population practicing sports, the probability for consumers to purchase functional yogurts versus conventional yogurts is higher than that in other markets with smaller population practicing sports. Some existing studies demonstrate the correlation between doing physical activity and positive overall diet quality (Yngve et al., 1999; Gillman et al., 2001; Lin et al., 2013). Evidence from federal government reports also shows an upward trend of the percentage of American adults who meet minimal aerobic physical activity guideline over time, increased from 43.5 percent in 2008 to 54.1 percent in 2017 (CDC, 2019, May 13), and an upward trend of Americans' diet quality on average, gradually increased over time (Wilson et al., 2016). However, in literature, the method of measuring physical activity varies, which is also a challenge for our study, how to quantify physical activity. One primary purpose of physical activity is to reduce weight by losing energy. The effectiveness of physical activity can be affected by intensity, time and type of physical activity and is also affected by other social factors such as school education and family effect (Berniell et al., 2013; Cawley et al. 2007, Cawley et al. 2013). In this study, we rely on the IRI Medprofiler data that collect participants' frequency of doing physical activity in a week, and we assume the disparity in intensity and type of physical activity over individuals is minimal.

The second topic of this study is the link between obesity and product choice. According to a study based on the National Health and Nutrition Examination Survey (NHANES), in 2015-2016, about 40 percent of American adults and 20 percent of American youths are obese. Even though the change over the year from 2013-2014 to 2015-2016 is not significant, there is a gradual upward trend in the prevalence of obesity over the year, increased from about 31 percentage in 1999-2000 to almost 40 percentage in 2015-2016 (Hales et al., 2017). It confused us that the incidence of obesity increases over time when average diet quality and physical activity improved, which forces us to switch from federal reports to academic studies in order to find any meaningful insight about the association between obesity and dietary choice. Some existing studies demonstrate a link between obesity and overall diet quality or the total consumption of some specific food categories. For example, literature shows that obese children or adolescents consume more high-calorie and low-nutrient-dense foods such as candy or more meats, sugar-sweetened beverages and potato chips than non-obese ones (Bandini et al., 1999; Gillis & Bar-Or, 2003). There is one study, Wang et al. (2017), that particularly investigates the link between obesity and consumer demand of one product. Wang et al. (2017) find that consumers in markets with lower incidence of obesity are more price sensitive to soda price than the ones in markets with higher incidence of obesity. We follow Wang et al. (2017), but differed from Wang et al., (2017), we employ individual-level approach to examine the disparity in consumer preference on the healthfulness of a product between obese and non-obese individuals. Furthermore, we investigate yogurt, a healthier grocery product than soda. It is interesting to know how the healthfulness of a product affects the association between obesity and product choice.

We investigate yogurt, mainly because of its good heterogeneity of health-related product attributes and the accessibility of these health-related attributes from nutrition fact panel. Yogurt as a dairy product is considered as a healthy food because its consumption associates with the intake of several healthy nutrients such as calcium, vitamin D and protein (Keast et al., 2015) and the intake of yogurts helps to reduce the risk of obesity and type 2 diabetes mellitus (Chen et al., 2014; Martinez-Gonzalez et al., 2014). A few researchers do hold an opposite opinion about the healthfulness of yogurts (Lopez-Garcia et al., 2015). Another reason for us to select yogurt is because of the good variation of brands, varieties and product characteristics (i.e. flavor and types of add-in items) and also because of its large sales every year and its importance to retailers (Villas-Boas, 2007). The yogurt market is an oligopoly market where Dannon and General Mills have about 62 percent of yogurt sales and private labels (i.e. brands owned and distributed by retailers) account for about 15 percent of the market in the U.S. (Villas-Boas, 2007). Besides brand owned by Dannon and General Mills, there are also many brands owned by independent manufactures, for example, Chobani, Fage and Stonefield. Yogurts can be categorized by the type of milks into nonfat, low-fat and whole milk and can also be categorized by the type of yogurts into Greek and conventional yogurts. Furthermore, there is a considerable variation in nutrition of yogurts over flavors and add-in items. This heterogeneity in product nutritional attributes allows us to observe how consumers' consciousness of healthy lifestyle associate with their preference on products over the healthfulness of alternatives. However, due to the non-disclosure agreement for the IRI data usage, we are not allowed to expose any brand names or retailer names so we use English alphabets and Arabic numbers to denote manufactures, brands and varieties.

A few studies explore the link between yogurts consumption and physical activity. Lopez-Garcia et al. (2015) find there is not significant association between the frequency of consuming yogurts and physical activity. Bonanno (2012) find a positive link between population practicing sports and probability of purchasing functional yogurts versus conventional yogurts in markets, meaning that physically active individuals are more likely to purchase a yogurt that is thought to be a healthier alternative of yogurts. Our study is substantially similar with Bonanno (2012) in term of the research on physical activity and consumer demand on yogurts. However, we employ micro-level data that consist of consumer purchase information and medical demographics, which allows us to identify the link between consumer demand and physical activity at individual level. Further, we measure the healthfulness of yogurts by nutrition and not by whether it is a functional yogurt or a conventional yogurt. This is also a challenge, how to measure diet quality of a yogurt.

Several methods of healthfulness measure for one product are documented in literature, but we use the amount of several important nutrients to differentiate the healthfulness of products. Bonanno (2012) considers functional yogurts as a healthy alternative versus conventional ones because in functional yogurts probiotics benefits digestive symptom and improve immune system, and Omega-3 benefit cardiovascular health (Guarner et al., 2005; Guyonnet et al., 2007; Dawcynski et al., 2013). However, based on the nutrition facts of functional and conventional yogurts documented in Bonanno (2013), it is difficult for us to find strong evidence that functional yogurts are superior to conventional yogurts in term of nutrition. Zhu et al. (2015) employ diet quality indices such as the Nutrition Profile Index, NuVal scores and an index created by principal component analysis of nutrients to denote the healthfulness of cereals. In our study, we employ the amount of some important nutrients per ounce for yogurts, where we consider sugar, protein, calcium and total fat. This allows us to capture more variation of alternatives and show interpretable results.

This study primarily investigates the associations of physical activity and obesity with consumer demand on yogurts. The data are the IRI Consumer and Medprofiler panels in 2013, where the IRI Consumer panel consists of household-level purchase record for every store visit over the year, product characteristics at UPC (i.e. Universal Product Code) level and detailed household demographic information, and the IRI Medprofiler panel consists of individual-level medical demographic and lifestyle information. To exact match up household-level consumer data with individual-level medical data, we only select single-member adult households who live alone over the year and have only one record in the IRI Medprofiler in 2013.

We select the top 30 varieties of the best sales 10 yogurt brands and private label brand, where we differentiate products by fat level, flavor and Greek. We do not distinguish different private labels brands. To obtain an effective but not computational burdensome sample, we select the top12 big yogurt markets in the U.S. including New York, Los Angeles, Baltimore/Washington DC, Chicago, Boston, Philadelphia, Miami/Ft. Lauderdale, Minneapolis/St. Paul, Phoenix/Tucson, Dallas/Ft. Worth, San Francisco/Oakland and Detroit.

To estimate consumer demand, we employ a mixed multinomial logit model. The mixed logit model allows us to capture the variation of consumer taste (Train, 2009). It is a variety-choice discrete choice demand model where we incorporate physical activity or obesity into the model by constructing interaction terms of physical activity or obesity with price and health-related product attributes. It allows us to observe the disparity in consumers' price sensitivity and preference on the healthfulness of yogurts between physically active and inactive individuals or between obese and non-obese individuals.

Research Objective

The research objective of this study is to estimate consumer demand on yogurts and investigate whether consumer demand on yogurts associates with physical activity and obesity. We construct a discrete choice demand model to estimate the disparity in price sensitivity and healthfulness of yogurts between physical active and inactive individuals and between obese and nonobese individuals. It allows us to understand how product-level choice links with obesity and physical activity, where obesity represents an objective health status and physical activity represents a healthy behavior for weight loss and healthy improvement. It also allows us to estimate the price elasticity over products and over individuals with different health status and with different frequency of physical activity.

Literature Review

This study contributes to three strands of literature. It first contributes to the literature that links physical activity with food choice.

One study, Bonanno (2012), particularly investigates the impact of exercise on probability of yogurt purchases. Bonanno (2012) mainly estimates consumer demand on yogurts in Italian market and measure the impact of health-related demographics characteristics on the probability to purchase functional yogurts versus conventional yogurts. He finds that as one percentage increase in the population practicing sports, the probability to purchase a functional yogurt versus a conventional yogurt decreases by about 3.8 percentage on average.

Other studies primarily investigate the link between exercise and overall diet quality. Gillman et al. (2001) conduct a cross-sectional survey with 1,322 male participants and find that as the increase in the intensity of physical activity, participants tend to consume less unhealthy meats, saturated fat, trans-fat and dietary cholesterol on average. They primarily focus on fruits, vegetables, whole grain foods, whole-fat dairy foods, and red and processed meats and find association between physical activity and a health dietary pattern. Lin et al (2013) investigate the association between weight management strategies and individuals' diet quality, where they use healthy eating index 2005 (HEI-2005) and several HEI-2005 components to measure diet quality. They find that doing physical activity and consuming foods with fewer calories together associate with HEI-2005, and only doing exercise but not eating fewer calories does not significantly associate with either HEI-2005 or the milk score of HEI-2005.

Some studies treat healthy diet and physical activity as two independent healthy behaviors and investigate their determinants separately. Øvrum (2011) investigates whether physical activity and consumption on fruits and vegetables associate with socioeconomic factors. Fan and Jin (2014) measure the self-control capability with Totter score and suggest that there is an association between lack of self-control and negative healthy behaviors such as poor eating and exercise and incidence of obesity. They also find that consuming more fruits and vegetables is easier for obese or overweight consumers than doing more physical activity.

A cluster of literature constructs theoretical models to investigate the relationship between food intake, physical activity and other resources input and output over a life-time span. One framework named SLOTH developed by Cawley (2004) is used in some studies. This SLOTH framework (Sleeping, Leisure, Occupation, Transport, and Household unpaid work) considers a model that maximizes a household's overall utility when optimally distributes its time and money for sleeping, leisure, occupation, transport and household unpaid work. Humphreys and Ruseski (2011) follow this SLOTH model to investigate the determinants of physical activity and find that individual characteristics such as income, gender, race, marital status and having children affect people's participation and duration of physical activity. Yaniv et al. (2009) use an idea similar with Cawley (2004) and propose a theoretical rational choice model where consumers optimally utilize their time in cooking, eating and doing physical activity. They conclude that imposing a fat tax and imposing a thin subsidy plays different roles in determining obesity for weight-conscious and non-weight conscious individuals. Dragone and

Savorelli (2012) also illustrate a theoretical model about food consumption, health and social pressure and suggest that lifting ideal weight improves people's welfare.

Second, this study contributes to the strand of literature that links obesity and food choice. A few studies are similar to our study. Patel (2012) employs a static demand model to estimate the link between demand on soda and body weight and obesity rate. Patel (2012) finds that consumers who have higher body weight or who are obese are less price elastic to soda products and their preference on diet sodas is superior to regular soda. Wang et al. (2017) employ dynamic discrete choice models to estimate the association between consumer demand on soda and obesity rate in markets. They find that consumers in markets with a lower obesity rate are more price elastic on average than the ones in markets with a higher obesity rate.

Other studies only explore the association between obesity and consumption of food categories. Bandini et al. (1999) find obese adolescents consume more candy, baked foods than non-obese ones. Ludwig et al. (2001) find the incidence of obesity associates with the intake of sugar-sweetened beverages for children. Gillis and Bar-Or (2003) find for obese youths, the consumption of sugar-sweetened beverages and potato chips is higher than that of non-obese ones. Bray et al. (2004) find an association between high-fructose corn syrup intake from beverages and obesity. Anari et al., (2017) find that as the increase in sugar-sweetened beverage consumption, the rate of abdominal obesity among diabetic patients increases. Based on reviewing existing literature, Swinburn et al. (2004), Hu and Malik (2010) and Della Torre et al. (2015) suggest there is evidences demonstrated by multiple studies that the consumption of sugar-sweetened beverages associates with the incidence of several diseases such as obesity, type 2 diabetes and cardiovascular diseases.

Finally, this study contributes to the strand of literature of yogurt markets and consumer demand. The benefits of often consuming yogurts are demonstrated in literature. Keast et al. (2015) based on NHANES 2005-2008 data find that yogurt consumption positively associates with the intake of some healthy nutrients such as calcium, vitamin D and protein and negatively associates with the intake of several types of fat. Lopez-Garcia et al. (2015) collect about 4000 adults' yogurt consumption during about 3 to 4 years, and find that the frequency of consuming yogurt during a week associates with the intake of some nutrients such as calcium and sugar, but it does not significantly associate with physical activity, sleep, body mass index (BMI) or energy intake. However, some yogurts might not as healthy as people think because of their excessive amount of sugar (Ware, 2018).

The first cluster of studies on yogurt market investigates yogurt market in a nation or using yogurt market as a case to demonstrate some questions of industrial organization. Carlucci et al. (2013) employ a hedonic pricing model to investigate yogurt prices in Italy. They find that yogurt products with fiber, probiotics or calcium tend to price higher than the ones without these nutrients or without probiotics. Hassan and Monier-Dilhan (2006) investigate the value of public-quality labels on yogurt product and other five products in France. Bonanno et al. (2015), using Italian yogurt market as a case, find that when several health claims were not allowed to be printed on product package in Europe since December 20, 2006, the truthfulness of the health claim directly affect consumers' and producers' welfare. Villas-Boas (2007) uses yogurt market as an example to analyze vertical relationships between manufacturers and retailers.

The second cluster of studies is the demand analysis considering the healthfulness of yogurt products, which is similar with this essay. As described by Bonanno (2012), some companies developed some particular types of yogurts, claiming their special functions that can facilitate digestive process and potentially benefit health. To distinguish the functional yogurt from conventional products, they are labelled with particular health claims such as strengthening the intestinal tract and reducing the absorption of cholesterol. Bonanno (2012) employs nested logit demand model to analyze the Italian yogurt market and identify price elasticity by brands and role of functional product. They find that consumers who are more valuing health lifestyle are likely to purchase functional yogurts versus conventional yogurts. Bonanno (2013) further examines the substitution pattern between conventional and functional yogurts and concludes that functional yogurts, on average, are less elastic on market but more profitable for producers than conventional yogurts. Bonanno (2016) employ hedonic price model to examine the correlation between health-related attributes and yogurt prices. He finds that among all health-related attributes, the food labels of probiotic, special health claim, reduced carbohydrates, organic and natural positively associate with product price, while the labels related with fiber, omega 3 and lactose free negatively associate with product price. Among all product attributes that are unrelated with health, the yogurt price is positive associated with the labels related with kids, Greek, local, fruit and flavored, and positively associated with the presence of fruit, honey, sauce and syrup on the bottom. The price is negatively associated with package size and being as prestirred. These studies provide rich information about yogurt markets, healthy yogurt products and the impact of health labelling information on packages. We consider the demand estimation as the backbone of this study, but we particularly interest in what roles physical activity and obesity play in consumer demand on yogurts.

The research questions this study investigated on are important and have been explored by Bonanno (2012) and Wang et al. (2017) with market-level approach. One big reason why there are not so many studies being conducted is because of the limitation of the data. Aforementioned several studies that investigate either the link between diet and physical activity or the link between obesity and physical activity rely on first-hand surveys, NHANES data or a combination of Nielsen or IRI consumer data with nation-level health survey data. The health survey such as first-hand surveys of health and diet and NHANES data lack of a continuous tracking of consumers' consumption or purchases over an appropriate period of time but have comprehensive record of health or diet information. On the other hand, Nielsen or IRI consumer data provide comprehensive data of a considerable number of American households over multiple years, but in most version used in existing literature, they do not contain any information about medical characteristics. Therefore, authors must acquire medical demographics and lifestyle information from different datasets such as American Community Survey, where they are allowed to obtain valid information about the incidence of obesity or population doing exercise in as many regions as possible over the country. This limit these studies to approach this question from market level. The unique data used in this study, IRI Consumer panel and Medprofiler panel, allow us to match the household purchase records over multiple years with individual medical and exercise data by individuals or by households, and therefore, observe how consumers' response to prices and product characteristics varies by their physical activity or by their health.

Empirical Models

This study employs a discrete choice demand model, Mixed Multinomial Logit Model, to estimate consumer demand on yogurts. The discrete choice framework can better describe the decision-making process of consumers purchasing yogurts over time. The random coefficients of mixed logit model that vary over choices simulate the variation of consumer tastes, allowing for unrestricted substitution patterns (Train, 2009).

First, we interest in the link between physical activity and product choice. We assume the indirect utility function of a household i for choosing a product $j \in J$ at time t in market m as

$$U_{ijtm} = \alpha_{itm} p_{jtm} + \beta_{itm} X_j + \gamma PA_i \left(p_{jtm} + X_j\right) + \tau_j + \xi_{ijtm} + \epsilon_{ijtm}, \tag{3.1}$$

where p_{jtm} is a variable that denotes the price of product j at time t in market m, X_j is a vector of observable products characteristics that are invariant over household, time or market and PA_i is a variable of physical activity that equals one if the household i does physical activity most days in a week and equals zero otherwise. The PA_i here does not vary over time because we only use the data in Year 2013 for analysis and the IRI Medprofiler data are collected on yearly basis. The interaction term, $PA_i p_{jtm}$, captures how physically active and inactive households respond to the change in prices. The interaction term, $PA_i X_j$, captures how physically active and inactive households respond the variation of product health-related attributes. α_{it} and β_{it} are random coefficients that vary over consumer and time and follow normal distributions. γ is a vector of non-random coefficients. τ_j is brand fixed effect. ξ_{ijtm} is unobservable product characteristics that associate with price. ϵ_{ijtm} is the error term.

To address the endogeneity issue of price caused by unobservable product characteristics ξ_{ijtm}, we employ the control function method proposed by Petrin and Train (2010). To employ this method, firstly obtain error terms of auxiliary regression of an endogenous variable against instruments and exogenous variables and secondly use these error terms as independent variables in main regression Equation (3.1). The auxiliary regression can be written as

$$p_{jtm} = \mu_1 Z_{tm} + \mu_2 X_j + \tau_j + \rho_{jtm}, \tag{3.2}$$

where Z_{tm} is a vector of instrument variables. μ_1 and μ_2 are coefficients. X_j is a vector of product characteristics. τ_j is a vector of brand fixed effects. ρ_{jtm} is the error term of auxiliary regression.

Considering the control function approach, we incorporate ρ_{jtm} is Equation (3.1) and rewrite it as

$$U_{ijtm} = \alpha_{it} p_{jtm} + \beta_{it} X_j + \gamma PA_i\left(p_{jtm} + X_j\right) + \rho_{jtm} + \tau_j + \xi_{ijtm} + \epsilon_{ijtm}. \tag{3.3}$$

For the instruments of prices, we employ the average price of all milk products in the same market and week and the average price of all cheese products in the same market and week as instruments.

For robustness check, we revise the model based on Equation (3.3). In one model following Equation (3.3), we redefine PA_i as a vector of two dichotomous variables and can be denoted as $PA_i = [PA_i^3 \quad PA_i^2]$, where PA_i^2 is a variable that equals one if household i does physical activity some days in a week and equals zero otherwise, and PA_i^3 is a variable that equals one if household i does physical activity most days a in week and equals zero otherwise. The baseline here is the households who rarely or never exercise. In another version of Equation (3.3) for robustness check, we assume τ_j is a choice fixed effect rather than a brand fixed effect.

Second, we also interest in the association between obesity and product choice on yogurts. We construct a discrete choice demand model similar to Equation (3.3). The indirect utility function for a household i choosing a product $j \in J$ at time t in market m is

$$
\begin{aligned}
U_{ijtm} = \alpha_{itm} p_{jtm} + \beta_{itm} X_j + \gamma_1 O_i\left(p_{jtm} + X_j\right) + \gamma_2 M_i\left(p_{jtm} + X_j\right) + \rho_{jtm} \\
+ \tau_j + \xi_{ijtm} + \epsilon_{ijtm},
\end{aligned}
\tag{3.4}
$$

where O_i is a dichotomous variable that equals one if household i is self-reported obese but not taking any Rx, OTC or dual medications and equals zero otherwise. M_i is a dichotomous variable that equals one if household i is self-reported obese and is taking Rx, OTC or dual medications and equals zero otherwise. The baseline is healthy individuals who are not obesity and not taking any medications for obesity. α_{itm} and β_{itm} are random coefficients, and γ_1 and γ_2 are vectors of non-random coefficients. The interaction terms $O_i p_{jtm}$ and $O_i X_j$ capture the disparity in consumers' responses to product price and health-related attributes between obese without medication usage and non-obese individuals, respectively. The interaction terms $M_i p_{jtm}$ and $M_i X_j$ capture the disparity in consumers' responses to prices and health-related product attributes between obese using any medications and non-obese individuals, respectively.

The O_i and M_i are obtained from the IRI Medprofilers data and are self-reported by participants. Another method to identify obesity is to use BMI. Individuals with a BMI larger than or equal to 30 are identified as obese as suggested by CDC (April 11, 2017). Notes that this this approach does not allow us to identify individual medication usage. The indirect utility function for a household i choosing a product $j \in J$ at time t in market m can be rewritten based on Equation (3.4) as

$$(3.5) \quad U_{ijtm} = \alpha_{itm} p_{jtm} + \beta_{itm} X_j + \gamma BMI_i \left(p_{jtm} + X_j\right) + \rho_{jtm} + \tau_j + \xi_{ijtm} + \epsilon_{ijtm},$$

where BMI_i is a dichotomous variable that equals one if household i has a BMI ≥ 30 and equals zero otherwise.

As ϵ_{ijt} is following independent and identically distributed extreme value distribution, the probability conditional on the set of random and non-random coefficients Ψ_{it}, where $\Psi_{it} = [\alpha_{it} \quad \beta_{it} \quad \gamma]$ for Equation (3.3), (3.4) and (3.5) and $\gamma = [\gamma_1 \quad \gamma_2]$ for Equation (3.4), can be written as

$$\mathcal{L}_{itm}(\Psi_{itm}) = \frac{exp\left[\alpha_{itm} p_{jtm} + \beta_{itm} X_j + \gamma H_i \left(p_{jtm} + X_j\right) + \rho_{jtm} + \tau_j\right]}{\sum_j \left[\alpha_{itm} p_{jtm} + \beta_{itm} X_j + \gamma H_i \left(p_{jtm} + X_j\right) + \rho_{jtm} + \tau_j\right]},$$

where H_i denotes PA_i for Equation (3.3), $[O_i \quad M_i]$ for Equation (3.4) and BMI_i for Equation (3.5). γ denotes $[\gamma_1 \quad \gamma_2]$ for Equation (3.4).

The unconditional choice probability is the integral of $\mathcal{L}_{\mathrm{it}}(\Psi_{itm})$ over Ψ_{itm}, which can be written as

$$(3.6) \qquad Pr_{itm}(\Psi_{itm}) = \int \mathcal{L}_{\mathrm{itm}}(\Psi_{itm}) f(\Psi_{itm}) d\Psi_{itm}.$$

The Equation (3.6) is the mixed logit probability, and we estimate with simulated maximum likelihood.

To calculate price elasticity, we follow Train (2009) and Richard and Bonnet (2016) to calculate own-price elasticities. We employ the Equation (3.7) as following

$$e_{itm}^{j} = \frac{p_{itmj}}{Pr_{itm}^{j}} \int (\alpha_{itm} + \gamma^{p}) \mathcal{L}_{itm}^{j}(\Psi)(1 - \mathcal{L}_{itm}^{j}(\Psi)) f(\Psi) d\Psi, \tag{3.7}$$

where γ^{p} is the coefficient of the interaction term of price and physical activity, obesity, medication usage or BMI, $\mathcal{L}_{itm}^{j}(\Psi)$ denotes the conditional probability for alternative j, and $f(\Psi)$ is a density function. After estimation, we can obtain the mean and standard deviation of the random coefficient α_{itm} to recover the normal distribution of α_{itm}.

Data

The data used for this study are the IRI (i.e. Information Resources, Incorporated) Consumer panel and Medprofiler panel in 2013 acquired from USDA's Economic Research Service. The IRI Consumer panel consists of household-trip-level grocery purchases record for more than 120,000 American households across 49 states over the year. It reports date and type of retail channel of almost all store visits in a year as well as UPC barcode, price and quantity of every product purchased in a store visit. The IRI data provide comprehensive product characteristics at UPC level, including brand, package size and type, volume per unit, flavor and almost all information on nutrition fact panel. The data also document household characteristics on yearly basis, for instance, household heads' birthday, gender, employment status and education level as well as household's size, income and zip code and state of residence.

The IRI Medprofiler panel collects individual-level medical demographic and lifestyle information on yearly basis from almost all household members of about half of the households in the IRI Consumer panel. Since we only consider one-year data in Year 2013, we assume households' medical and lifestyle information remain consistent over the year. In the IRI Medprofiler panel, participants were asked how often he or she does exercise in a week. Is it most days in a week, some days in a week or rarely/never? They were also asked whether they had been diagnosed with obesity and whether they had been taking Rx, OTC or dual medications for obesity. Furthermore, the IRI Medprofiler panel collects individuals' height and weight, which allows us to calculate individual BMI.

To merge household-level consumer data with individual-level medical data, we only use single-member households who live alone during the year and only fill in the Medprofiler survey once in the year. The observations with missing information or answered "refused to answer" for the questions of physical activity and obesity are deleted.

The summary statistics for households is shown in Table 3-1. For comparison, the statistical summary of all relevant demographics, medical and lifestyle information of our IRI Consumer and Medprofiler data is also shown in Table 3-1. The statistical summary of demographics is on household level for all households appeared in the 2013 IRI Medprofiler data, and the statistical summary of medical and lifestyle information is on individual level for all individuals with valid information in the 2013 IRI Medprofiler data. The average income for all single-member households used in this study is about $60,000.

Table **3-1**: Summary statistic of single-member households and entire sample.

Variable	Definition	Used Sample	Entire Sample
		Mean/ Frequency	Mean/ Frequency
Demographics			
Income	Average household income.	\$60417.6 (50676.4)	\$70161.4 (58057.6)
African American	Percentage of African American.	0.184 (0.387)	0.102 (0.302)
Asian	Percentage of Asian American.	0.024 (0.152)	0.029 (0.168)
Hispanic	Percentage of Hispanic American.	0.045 (0.207)	0.054 (0.226)
College Education	Percentage of individuals who enrolled in college or graduated from college.	0.615 (0.487)	0.542 (0.498)
Full-Time Employed	Percentage of individuals who are employed on full-time basis.	0.472 (0.499)	0.594 (0.491)
Part-Time Employed	Percentage of individuals who are employed on part-time basis.	0.153 (0.360)	0.224 (0.417)
Male	Percentage of male.	0.257 (0.437)	0.736 (0.441)
N	Number of households.	1177	47040
Medical and Lifestyle Characteristics			
PA3	Percentage of individuals who do exercise most days in a week.	0.406 (0.491)	0.369 (0.483)
PA2	Percentage of individuals who do exercise some days in a week.	0.367 (0.481)	0.361 (0.480)
Obesity	Percentage of individuals who are self-reported obese and not taking any Rx, OTC or dual medications for obesity.	0.137 (0.344)	0.153 (0.360)
Medication Usage	Percentage of individuals who are taking any Rx, OTC or dual medications for obesity.	0.040 (0.200)	0.026 (0.160)
BMI $\geq$ 30	Percentage of individuals whose BMI is larger than or equal to 30.	0.353 (0.478)	0.350 (0.477)
BMI	Average BMI.	28.8 (7.1)	28.7 (7.4)
N	Number of participants.	1177	92800

Notes: Values in parenthesis are standard deviation. For the summary statistics of the entire sample, it counts the demographics of either household head in the households. For example, for the percentage of male in the entire sample, it counts the percentage of households has a male head. For the medical and lifestyle characteristics, the summary statistics is individual-level for all adults with valid information in the IRI Medprofiler panel in 2013.

In order to understand the disparity in demographics over physical activity, obesity and BMI, the statistical summary of demographics is shown in Table 3-2.

Table **3-2**: The individual and medical demographics by physical activity and BMI ≥ 30.

Variable	PA3	PA2	PA1	BMI ≥ 30	BMI < 30
Income	65631.5	60463.9	51146.8	55344.7	63149.7
	(53481.3)	(50968.3)	(43404.2)	(47407.7)	(52161.2)
African American	0.167	0.210	0.171	0.247	0.150
	(0.373)	(0.408)	(0.377)	(0.432)	(0.357)
Asian	0.019	0.028	0.026	0.012	0.030
	(0.136)	(0.408)	(0.159)	(0.110)	(0.171)
Hispanic	0.044	0.040	0.056	0.044	0.046
	(0.205)	(0.195)	(0.230)	(0.205)	(0.209)
College Education	0.641	0.643	0.534	0.545	0.652
	(0.480)	(0.480)	(0.500)	(0.499)	(0.477)
Full-Time Employed	0.486	0.455	0.476	0.477	0.470
	(0.500)	(0.500)	(0.500)	(0.500)	(0.499)
Part-Time Employed	0.152	0.182	0.108	0.130	0.165
	(0.360)	(0.386)	(0.311)	(0.336)	(0.372)
Male	0.248	0.256	0.275	0.213	0.281
	(0.433)	(0.437)	(0.447)	(0.410)	(0.500)
PA3				0.267	0.482
				(0.443)	(0.500)
PA2				0.386	0.353
				(0.487)	(0.478)
Obesity	0.065	0.131	0.275	0.342	0.027
	(0.246)	(0.337)	(0.447)	(0.475)	(0.163)
Medication Usage	0.023	0.058	0.041	0.088	0.014
	(0.150)	(0.235)	(0.198)	(0.284)	(0.119)
BMI ≥ 30	0.228	0.368	0.528		
	(0.420)	(0.483)	(0.500)		
BMI	26.6	29.3	31.7	36.2	24.8
	(5.2)	(7.2)	(8.4)	(6.2)	(3.2)
N	479	429	269	409	768

Notes: Values in parenthesis are standard deviation. PA3 denotes individuals who do exercise most days in a week. PA2 denotes individuals who do exercise some days in a week. PA1 denotes individuals who rarely or never exercise.

In this study, we use refrigerated yogurts as a case. In order to guarantee an effective and not computational burdensome sample, we select the top 12 yogurt markets in the IRI Consumer Panel, including New York, Los Angeles, Baltimore/Washington DC, Chicago, Boston, Philadelphia, Miami/Ft. Lauderdale, Minneapolis/St. Paul, Phoenix/Tucson, Dallas/Ft. Worth, San Francisco/Oakland and Detroit.

Also, we only consider 10 major yogurt brands and one private label brand so that the estimation is efficient without losing the representativeness of the sample. Different private label brands are not differentiated. Because the heterogeneity of yogurts in nutrients over varieties cannot be ignored. For example, the disparity in fat and calorie amount between nonfat and low-fat milk is not minimal and the disparity in sugar amount between Greek and conventional yogurts is also not minimal. Therefore, we also differentiate yogurts by plain versus flavor, nonfat versus low-fat versus whole milk, and Greek versus conventional yogurts. Furthermore, only the brand-varieties with sales in volume larger than 10,000 oz in the year are selected, which are the top 30 brand-varieties.

To measure the healthfulness of yogurts, several key nutrients are selected including sugar, calorie, calcium, protein and total fat. The statistical summary regarding the product characteristics for the chosen products is shown in Table 3-3. The average price per ounce for the chosen yogurts is about \$0.27. We particularly interest in five important nutrients, sugar, calorie, calcium, protein and total fat, and their average amounts per ounce are 2.93 g, 21.76 g, 329.97 mg, 1.55 g and 0.19 g, respectively. About ten percent of the yogurts purchased are plain yogurts. About 52 percent of yogurts purchased are nonfat yogurts and about 38 percent of yogurts are low-fat. About half of the purchased yogurts are Greek yogurts.

Table **3-3**: Statistical summary of the chosen yogurts.

Variable	Definition	Mean / Frequency
Price	US Dollar per ounce.	0.27
		(0.08)
Sugar	Gram (g) per ounce.	2.93
		(1.10)
Calorie	Calorie per ounce.	21.76
		(6.33)
Calcium	Milligram (g) per ounce.	329.97
		(119.86)
Protein	Gram (g) per ounce.	1.55
		(0.68)
Total Fat	Gram (g) per ounce.	0.19
		(0.29)
Plain	Percentage of choices is plain yogurts.	9.6%
Nonfat	Percentage of choices is nonfat yogurts.	52.4%
Low-fat	Percentage of choices is low-fat yogurts.	38.3%
Greek	Percentage of choices is Greek yogurts.	50.5%
N	Number of observations.	32,254

Notes: Values in parenthesis are standard deviation.

The Table 3-4 shows statistical summary of product characteristics by brand and choice. The average prices of plain yogurts, non-Greek and private labels are lower than average price of the entire sample. No particular pattern for the prices of nonfat and low-fat yogurts. Compared with low-fat yogurts, nonfat yogurts have lower sugar and calorie on average. Greek yogurts are more intensive in protein than conventional yogurts.

Table **3-4**: The choice sets and summary statistics by choices and manufactures.

	Price	Sugar	Calorie	Calcium	Protein	Total Fat
Manufacture A						
Brand A1.1	0.25	4.19	26.55	0.59	0.88	0.30

Flavored, Low-fat, Non-Greek	(0.02)	(0.18)	(0.83)	(0.05)	(0.03)	(0.03)
Brand A1.2	0.36	1.67	21.21	0.26	2.13	0.00
Flavored, Nonfat, Greek	(0.03)	(0.19)	(0.91)	(0.05)	(0.05)	(0.00)
Brand A1.3	0.30	2.30	16.94	0.33	0.87	0.00
Flavored, Nonfat, Non-Greek	(0.02)	(0.06)	(0.16)	(0.00)	(0.02)	(0.00)
Brand A2.1	0.48	5.63	36.75	0.33	1.25	0.84
Flavored, Low-fat, Non-Greek	(0.03)	(0.00)	(0.93)	(0.02)	(0.00)	(0.05)
Manufactural Mean	0.35	3.45	25.36	0.38	1.28	0.28
	(0.09)	(1.57)	(7.44)	(0.13)	(0.51)	(0.34)
Manufacture B						
Brand B1.1	0.18	3.43	20.32	0.36	0.93	0.38
Flavored, Low-fat, Non-Greek	(0.08)	(0.99)	(5.62)	(0.16)	(0.26)	(0.17)
Brand B1.2	0.05	0.69	6.36	0.16	0.44	0.24
Plain, Low-fat, Non-Greek	(0.04)	(0.36)	(2.50)	(0.09)	(0.23)	(0.07)
Brand B1.3	0.09	0.92	5.60	0.22	0.58	0.00
Plain, Nonfat, Non-Greek	(0.07)	(0.51)	(3.42)	(0.13)	(0.33)	(0.00)
Brand B2.1	0.37	3.50	32.02	0.30	2.20	1.00
Flavored, Low-fat, Greek	(0.03)	(0.00)	(0.05)	(0.00)	(0.00)	(0.00)
Brand B2.2	0.32	3.76	24.52	0.29	2.36	0.00
Flavored, Nonfat, Greek	(0.04)	(0.14)	(0.70)	(0.00)	(0.06)	(0.00)
Brand B2.3	0.04	0.23	3.75	0.08	0.69	0.00
Plain, Nonfat, Greek	(0.02)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Brand B3.1	0.12	4.59	28.85	0.38	1.01	0.50
Flavored, Low-fat, Non-Greek	(0.01)	(0.12)	(0.36)	(0.00)	(0.03)	(0.01)
Brand B3.2	0.14	3.92	25.82	0.32	2.34	0.00
Flavored, Nonfat, Greek	(0.07)	(0.37)	(1.71)	(0.03)	(0.11)	(0.00)
Brand B3.3	0.12	1.87	17.22	0.37	1.08	0.00
Flavored, Nonfat, Non-Greek	(0.01)	(0.02)	(0.51)	(0.00)	(0.06)	(0.00)
Brand B4.1	0.14	1.82	11.82	0.22	0.74	0.00
Flavored, Nonfat, Non-Greek	(0.06)	(0.23)	(1.34)	(0.02)	(0.08)	(0.00)
Brand B4.2	0.18	1.45	15.69	0.29	2.34	0.00
Flavored, Nonfat, Greek	(0.08)	(0.09)	(0.66)	(0.01)	(0.11)	(0.00)
Manufactural Mean	0.16	2.38	17.45	0.27	1.34	0.19
	(0.11)	(1.48)	(9.59)	(0.11)	(0.77)	(0.31)
Manufacture C						
Brand C1.1	0.28	3.10	26.83	0.25	2.21	0.51
Flavored, Low-fat, Greek	(0.03)	(0.11)	(0.75)	(0.01)	(0.05)	(0.02)
Brand C1.2	0.28	3.02	22.70	0.27	2.27	0.01
Flavored, Nonfat, Greek	(0.02)	(0.12)	(0.68)	(0.01)	(0.07)	(0.01)
Brand C1.3	0.17	0.70	10.15	0.21	1.78	0.00
Plain, Nonfat, Greek	(0.07)	(0.28)	(3.33)	(0.06)	(0.62)	(0.00)

Brand C2.1	0.38	4.05	39.57	0.30	2.30	1.31
Flavored, Low-fat, Greek	(0.03)	(0.22)	(2.96)	(0.00)	(0.12)	(0.29)
Manufactural Mean	0.28	2.72	24.81	0.26	2.14	0.46
	(0.09)	(1.25)	(10.75)	(0.05)	(0.38)	(0.55)
Manufacture D						
Brand D1.1	0.27	2.64	23.29	0.19	2.01	0.52
Plain, Low-fat, Greek	(0.08)	(0.94)	(5.42)	(0.03)	(0.35)	(0.11)
Brand D1.2	0.23	2.05	15.66	0.22	1.77	0.00
Plain, Nonfat, Greek	(0.07)	(0.67)	(4.04)	(0.04)	(0.38)	(0.00)
Manufactural Mean	0.25	2.34	19.47	0.21	1.89	0.26
	(0.07)	(0.87)	(6.12)	(0.04)	(0.38)	(0.27)
Manufacture E						
Brand E1.1	0.15	3.03	19.74	0.32	0.89	0.32
Flavored, Low-fat, Non-Greek	(0.06)	(0.88)	(5.43)	(0.09)	(0.23)	(0.10)
Brand E1.2	0.21	3.28	18.01	0.37	0.94	0.00
Flavored, Nonfat, Non-Greek	(0.09)	(1.03)	(5.73)	(0.12)	(0.29)	(0.00)
Brand E1.3	0.30	3.23	24.04	0.31	2.58	0.00
Flavored, Nonfat, Greek	(0.10)	(0.55)	(3.12)	(0.04)	(0.26)	(0.00)
Brand E1.4	0.07	0.99	7.76	0.26	0.62	0.16
Plain, Low-fat, Non-Greek	(0.06)	(0.46)	(3.03)	(0.10)	(0.23)	(0.07)
Manufactural Mean	0.18	2.64	17.39	0.31	1.26	0.12
	(0.11)	(1.22)	(7.48)	(0.10)	(0.82)	(0.15)
Private Labels						
Brand PL.1	0.23	3.15	20.38	0.30	0.93	0.28
Flavored, Low-fat, Non-Greek	(0.07)	(1.15)	(5.10)	(0.06)	(0.19)	(0.05)
Brand PL.2	0.18	1.97	13.77	0.34	0.91	0.00
Flavored, Nonfat, Non-Greek	(0.08)	(0.18)	(1.28)	(0.04)	(0.11)	(0.00)
Brand PL.3	0.23	2.64	20.14	0.32	2.14	0.00
Flavored, Nonfat, Greek	(0.09)	(0.76)	(4.12)	(0.14)	(0.41)	(0.00)
Brand PL.4	0.08	1.07	7.90	0.21	0.59	0.01
Plain, Nonfat, Non-Greek	(0.05)	(0.45)	(3.73)	(0.08)	(0.26)	(0.01)
Brand PL.5	0.19	0.54	8.33	0.17	1.38	0.00
Plain, Nonfat, Greek	(0.10)	(0.25)	(3.64)	(0.07)	(0.56)	(0.00)
Manufactural Mean	0.18	1.88	14.11	0.27	1.19	0.06
	(0.10)	(1.17)	(6.62)	(0.11)	(0.64)	(0.11)

Notes: For some nonfat varieties, the total fat is not restrictive zero caused by the imputation of missing values or added-in flavors or items. Values in parenthesis are standard deviation. The unit for price is US dollar per ounce and the unit for gram per ounce.

Results

First, we employ mixed logit model to estimate consumer demand on yogurts without interaction terms. The results of Equation (3.3) are shown in Table 3-5. We regress three models, where Model 1 does not consider any interaction terms, Model 2 considers interaction terms of intensive physical activity (i.e. doing exercise most days in a week) and price and nutrition and Model 3 considers interaction terms of two levels of physical activity and price and nutrition. The baseline case for Model 2 is the individuals who do exercise some days in a week and who rarely or never exercise, and the baseline case for Model 3 is the individuals who rarely or never exercise. Since for logit models, the value of the coefficient does not directly link with the odd ratio, it is, therefore, not interpretable, but Table 3-5 does present the disparity in consumers' price sensitivity and preference on the healthfulness of yogurts over individuals with different levels of physical activity. The individuals who do exercise some days in a week is the most price sensitive on average, followed by the ones who do exercise most days in a week and the ones who rarely or never exercise. The most physically active group of individuals prefer yogurts with less sugar and calcium and more protein on average than other groups but its preference on total fat does not significantly differ from the baseline (i.e. rarely or never exercise). For the individuals who do exercise some days in a week, their sensitivity to nutrition in yogurts does not statistically significantly differ from that of who rarely or never exercise.

In Table 3-5, all three models employ the average market price of yogurt products in the same market and week as an instrument for yogurt price. For robustness check, we replace the instrument with the average market price of cheese in the same market and week, and find the change in regression results is minimal.

Table **3-5**: The regression results of mixed logit models with physical activity.

	Model 1		Model 2		Model 3	
	Coeff	SD	Coeff	SD	Coeff	SD
Price	-5.58^{***}	0.00	-5.29^{***}	0.00	-3.71^{***}	0.00
	(0.20)	(3.72)	(0.21)	(0.00)	(0.25)	(3.73)
Sugar	0.10^{***}	0.00	0.13^{***}	0.00	0.12^{***}	0.00
	(0.02)	(0.27)	(0.02)	(0.27)	(0.02)	(0.27)
Calcium	0.00^{***}	0.00	0.00^{***}	0.00	0.002^{***}	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Protein	0.43^{***}	0.00	0.38^{***}	0.00	0.36^{***}	0.00
	(0.06)	(0.31)	(0.06)	(0.31)	(0.07)	(0.31)
Total Fat	0.85^{***}	0.08	0.87^{***}	0.07	0.79^{***}	0.08
	(0.21)	(0.70)	(0.21)	(0.70)	(0.21)	(0.70)
Price×PA2					-2.50^{***}	
					(0.22)	
Sugar×PA2					0.01	
					(0.02)	
Ca×PA2					-0.0004^{*}	
					(0.0002)	
Protein×PA2					0.05	
					(0.03)	
Fat×PA2					-0.004^{*}	
					(0.062)	
Price×PA3			-0.61^{***}		-2.17^{***}	
			(0.16)		(0.21)	
Sugar×PA3			-0.06^{***}		-0.05^{***}	
			(0.02)		(0.02)	
Ca×PA3			-0.00^{***}		-0.00^{***}	
			(0.00)		(0.000)	
Protein×PA3			0.13^{***}		0.16^{***}	
			(0.02)		(0.03)	
Fat×PA3			−0.08		−0.00	
			(0.05)		(0.06)	
IV Terms	Milk		Milk		Milk	
Choice FEs	Yes		Yes		Yes	
Log-L		−89861		−89699		−89618
N			967,620			

Notes: Values in parenthesis are standard error. *** denotes p-value ≤ 0.01; ** denotes p-value ≤ 0.05; * denotes p-value ≤ 0.1. PA3 denotes physical activity, and it equals to one if the individual does physical activity in most days in a week and equals to zero otherwise. PA3 denotes physical activity, and it equals to one if the individual does physical activity in some days in a week and equals to zero otherwise. Cal denotes calorie. Ca denotes calcium. Fat denotes total fat. Milk is the average price of all milk products in the same week and market. Cheese is the average price of all cheese products in the same week and market.

To investigate the link between consumer demand on yogurts and obesity, we follow Equation (3.4) and employ interaction terms of obesity and medication usage with health-related product attributes, respectively. The Obesity dichotomous variable denotes the individual is self-reported obese and not taking any Rx, OTC or dual medications, while the Med dichotomous variable denotes the individual is taking medication for obesity. The baseline is self-reported non-obese. Model 4 employs brand fixed effects that only capture the variation of brands, and Model 5 employs choice fixed effects that capture the variation of brands and varieties. Indirectly, the choice fixed effects also capture the variations of nutrition because of the differences in nutrition over varieties. Therefore, the regressed coefficients of nutrients vary over Model 4 and 5.

The results are shown in Table 3-6. Using Model 5 as an example for interpretation, we find that individuals who are not suffering obesity and not taking any medication for obesity are likely to purchase yogurts with more sugar, calcium, protein and total fat on average when we capture the variation of varieties with fixed effects. The individuals who are taking medications for obesity are the least price sensitive on average, followed by the ones who are self-reported obese and not taking medications and the baseline case. Self-reported obese individuals prefer yogurts with less sugar and protein on average than the baseline, but they are not significantly different in the preference in calcium and total fat from the healthy individuals. For the individuals who are taking any Rx, OTC or dual medications for obesity, they prefer yogurts with more sugar and protein and less total fat on average than the baseline, but they are not sensitive to the amount of calcium.

The self-reported obesity and medication usage denote that the individuals admit they have obesity by themselves. This method is sensitive to participants' own belief and knowledge regarding obesity. The misunderstanding about obesity and medication for obesity, lack of corresponding professional knowledge about obesity and overestimation or underestimation of the weight and health might easily affect the results. Therefore, we also employ another approach, using BMI to identify obesity.

Table **3-6**: The regression results of mixed logit models of obesity and medication usage.

	Model 4		Model 5	
	Coeff	SD	Coeff	SD
Price	-2.58^{***}	0.01	-5.76^{***}	0.00
	(1.60)	(2.71)	(0.20)	(3.73)
Sugar	0.59^{***}	0.00	0.10^{***}	0.00
	(0.01)	(0.25)	(0.02)	(0.27)
Calcium	0.00^{***}	0.00	0.002^{***}	0.00
	(0.00)	(0.00)	(0.000)	(0.00)
Protein	-0.06^{***}	0.00	0.44^{***}	0.00
	(0.02)	(0.26)	(0.06)	(0.31)
Total Fat	-0.83^{***}	1.74^{***}	0.89^{***}	0.07
	(0.06)	(0.08)	(0.02)	(0.70)
Price × Obesity	0.66^{***}		0.62^{***}	
	(0.23)		(0.23)	
Price × Med	1.11^{***}		1.12^{***}	
	(0.39)		(0.39)	
Sugar × Obesity	-0.06^{**}		-0.06^{**}	
	(0.02)		(0.02)	
Sugar × Med	0.14^{***}		0.15^{***}	
	(0.04)		(0.04)	
Ca × Obesity	−0.00		−0.0002	
	(0.08)		(0.0003)	
Ca × Med	−0.00		0.0001	
	(0.00)		(0.0004)	
Protein × Obesity	-0.10^{***}		-0.11^{***}	

	(0.03)	(0.03)
Protein × Med	0.25^{***}	0.26^{***}
	(0.05)	(0.05)
Fat × Obesity	-0.14^{*}	-0.06
	(0.08)	(0.07)
Fat × Med	-0.57^{***}	-0.47^{***}
	(0.13)	(0.11)
IV Terms	Milk	Milk
Choice FEs		Yes
Brand FEs	Yes	
Log-L	−91756	−89754
N	967,620	

Notes: Values in parenthesis are standard error. *** denotes p-value ≤ 0.01; ** denotes p-value ≤ 0.05; * denotes p-value ≤ 0.1. Obesity is a binary variable that equals to one if the individual is self-reported obese but not taking any Rx, OTC or dual medications, and equals to zero otherwise. Med is a binary variable that equals to one if the individual is self-reported obese and is taking Rx, OTC or dual medications, and equals to zero otherwise. Cal denotes calorie. Ca denotes calcium. Fat denote total fat. Milk is the average price of all milk products in the same week-market.

Besides the self-reported obesity and medication usage, BMI can also be used to identify obesity. CDC suggests an individual with a BMI larger than or equal to 30 can be classified as overweight or obese (CDC, April 11, 2017). BMI is calculated with the self-reported weight and height documented in the IRI Medprofiler Panel. The regressions of Equation (3.5) are shown in Table 3-7. Here we also perform two models where one uses choice fixed effects and another uses brand fixed effects. The baseline case is the individuals whose BMI < 30.

The Model 7, a model with choice fixed effects is used as an example for interpretation because more variation over choices is captured and the model has better goodness of fit. The individuals whose BMI ≥ 30 are more price sensitive and prefer yogurts with more sugar and protein and less total fat on average than the ones whose BMI < 30. This result is contradictory to the results in Table 3-6. While both models, Model 5 and Model 7, are investigating the link between obesity and consumer demand on yogurts, but using different identification strategies for obesity produces the opposite results.

Table **3-7**: The regression results of mixed logit models of BMI ≥ 30.

	Model 6		Model 7	
	Coeff	SD	Coeff	SD
Price	-2.22^{***}	0.01	-5.44^{***}	0.00
	(0.16)	(2.72)	(0.20)	(3.72)
Sugar	0.56^{***}	0.00	0.07^{***}	0.00
	(0.01)	(0.25)	(0.02)	(0.27)
Calcium	0.00^{***}	0.00	0.002^{***}	0.00
	(0.00)	(0.00)	(0.000)	(0.00)
Protein	-0.08^{***}	0.00	0.41^{***}	0.00
	(0.02)	(0.26)	(0.06)	(0.31)
Total Fat	-0.84^{***}	1.74^{***}	0.90^{***}	0.07
	(0.06)	(0.08)	(0.21)	(0.70)
Price × BMI ≥ 30	-0.77^{***}		-0.67^{***}	
	(0.17)		(0.17)	
Sugar × BMI ≥ 30	0.09^{***}		0.08^{***}	
	(0.02)		(0.02)	
Ca × BMI ≥ 30	−0.00		−0.0001	
	(0.00)		(0.0002)	
Protein × BMI ≥ 30	0.06^{***}		0.06^{**}	
	(0.02)		(0.02)	
Fat × BMI ≥ 30	-0.11^{*}		-0.12^{**}	
	(0.06)		(0.05)	
IV Terms	Milk		Milk	
Choice FEs			Yes	
Brand FEs	Yes			
Log-L		−91812		−89815
N	967,620			

Notes: Values in parenthesis are standard error. *** denotes p-value ≤ 0.01; ** denotes p-value ≤ 0.05; * denotes p-value ≤ 0.1. BMI ≥ 30 is a dichotomous variable that equals to one if individual's BMI is larger than or equal to 30 and equals to zero otherwise. Cal denotes calorie. Ca denotes calcium. Fat denote total fat. Milk is the average price of all milk products in the same week-market.

We employ the regression of Model 3 to estimate the choice-level own-price elasticity following Equation (3.7). Since the coefficient of price is a random coefficient following a normal distribution, we perform simulations 100 times to approximately estimate the integration. The statistical summary of average choice-level own-price elasticities is shown in Table 3-8. The average own-price elasticities for most brand-varieties are larger than one. Most of brand-varieties produced by Manufacture B are price inelastic on average for consumers. Plain yogurts have low own-price elasticities on average, while Greek yogurts have high own-price elasticities on average.

Table **3-8**: The average choice-level own-price elasticities of yogurts.

Brand	Characteristics	Own-Price Elasticity
Manufacture A		
Brand A1.1	Flavored, Low-fat, Non-Greek	−1.183
		(0.257)
Brand A1.2	Flavored, Nonfat, Greek	−1.812
		(0.364)
Brand A1.3	Flavored, Nonfat, Non-Greek	−1.442
		(0.284)
Brand A2.1	Flavored, Low-fat, Non-Greek	−2.630
		(0.500)
Manufacture B		
Brand B1.1	Flavored, Low-fat, Non-Greek	−0.986
		(0.469)
Brand B1.2	Plain, Low-fat, Non-Greek	−0.295
		(0.236)
Brand B1.3	Plain, Nonfat, Non-Greek	−0.516
		(0.391)
Brand B2.1	Flavored, Low-fat, Greek	−1.932
		(0.377)
Brand B2.2	Flavored, Nonfat, Greek	−1.631
		(0.347)
Brand B2.3	Plain, Nonfat, Greek	−0.218
		(0.131)
Brand B3.1	Flavored, Low-fat, Non-Greek	−0.606
		(0.111)
Brand B3.2	Flavored, Nonfat, Greek	−0.759

		(0.414)
Brand B3.3	Flavored, Nonfat, Non-Greek	−0.645
		(0.128)
Brand B4.1	Flavored, Nonfat, Non-Greek	−0.714
		(0.312)
Brand B4.2	Flavored, Nonfat, Greek	−0.971
		(0.471)
Manufacture C		
Brand C1.1	Flavored, Low-fat, Greek	−1.456
		(0.298)
Brand C1.2	Flavored, Nonfat, Greek	−1.368
		(0.267)
Brand C1.3	Plain, Nonfat, Greek	−0.919
		(0.429)
Brand C2.1	Flavored, Low-fat, Greek	−2.104
		(0.406)
Manufacture D		
Brand D1.1	Plain, Low-fat, Greek	−1.458
		(0.491)
Brand D1.2	Plain, Nonfat, Greek	−1.196
		(0.402)
Manufacture E		
Brand E1.1	Flavored, Low-fat, Non-Greek	−0.816
		(0.382)
Brand E1.2	Flavored, Nonfat, Non-Greek	−1.160
		(0.511)
Brand E1.3	Flavored, Nonfat, Greek	−1.614
		(0.600)
Brand E1.4	Plain, Low-fat, Non-Greek	−0.390
		(0.354)
Private Labels		
Brand PL.1	Flavored, Low-fat, Non-Greek	−1.206
		(0.431)
Brand PL.2	Flavored, Nonfat, Non-Greek	−0.977
		(0.443)
Brand PL.3	Flavored, Nonfat, Greek	−1.261
		(0.563)
Brand PL.4	Plain, Nonfat, Non-Greek	−0.433
		(0.285)
Brand PL.5	Plain, Nonfat, Greek	−1.027
		(0.563)

Notes: Values in parenthesis are standard deviation.

Furthermore, we calculate the average choice-level own-price elasticities by physical activity and BMI of 30 as shown in Table 3-9. The own-price elasticities by physical activity are estimated based on Model 3 and the own-price elasticities by BMI ≥ 30 are estimated based on Model 7. The estimation is following Equation (3.7). The average choice-level own-price elasticities over physical activity and the ones over BMI < 30 and BMI ≥ 30 are substantially similar with the corresponding regression results of Model 3 and Model 7, respectively.

Table **3-9**: The average choice-level own-price elasticities by physical activity and BMI of 30.

Variable	Own-price Elasticity
Physical Activity	
PA3	−1.200
	(0.682)
PA2	−1.269
	(0.721)
PA1	−0.753
	(0.425)
BMI ≥ 30	
BMI ≥ 30	−1.249
	(0.707)
BMI < 30	−1.108
	(0.630)

Notes: Values in parenthesis are standard deviation.

Discussions and Conclusions

This study estimates consumer demand on yogurts and its association with the frequency of physical activity and obesity. It contributes to the literature by first analyzing the links of consumer demand on a grocery product with exercise and obesity using micro-level data and methodology. Bonanno (2012) finds the markets with more population practicing sports tend to have a higher odds ratio to purchase functional yogurts versus conventional yogurts on average. Wang et al. (2017) find a negative association between the obesity rate in a market and average price elasticity on soda. Differed from these two market-level studies, we approach the question with micro-level data and method, aiming to understand how consumers with different health-related characteristics behave on the grocery market towards one particular product. This study selects yogurt because of its good heterogeneity in health-related product attributes over alternatives and its wide presentation in literature.

We find that individuals who are physically active are more price sensitive and prefer yogurts with less sugar and calcium and more protein on average than the ones who rarely or never exercise. There is also a difference in price sensitivity and healthfulness preference on yogurts as the frequency of exercise varying. Based on average own-price elasticity over physical activity, the individuals who do exercise some days in a week are the most price elastic on average, followed by the individuals who do exercise most days in a week and the ones who rarely or never exercise. Since price elasticity directly links with product unit price, the reason why the most physical active group of individuals is the second most price sensitive might because they prefer yogurts with lower prices, for example, plain yogurts. We find that for the choices made by the one who do exercise most days in a week, about 12 percent are plain yogurts and about 54 percent are Greek yogurts, where the percentages of choosing plain yogurts are about nine percent and five percent for individuals who do exercise some days in a week and the ones who rarely or never exercise, respectively, and the percentages of choosing

Greek yogurts are about 47 percent for both these two groups. Plain yogurts, as shown in summary statistics, contain less sugar and calorie on average and might be considered as one of the healthiest alternatives among all types of yogurts. The physically active individuals' preference on a healthy but less expensive product might lead to their low average own-price elasticity. This point is substantially similar with Bonanno (2012), where he finds a positive association between population practicing sports and consumption on functional yogurts versus conventional yogurts. Functional yogurts are also considered as a healthier yogurt variety but one with higher average market price (Guarner et al., 2005; Guyonnet et al., 2007; Bonanno, 2013; Dawcynski et al., 2013). The commonality finding is that physically active individuals prefer healthier yogurt varieties on average, but our finding regarding consumers' willingness to buy for a more expensive but healthier variety is mixed and is not consistent with Bonanno (2012). We do observe more consumption of a type of healthier but relatively expensive yogurts, Greek yogurts, from those physically active consumers, but we also observe more consumption of plain yogurts, which contain lower sugar and calorie on average and have lower average market price than others.

In this study, we employ two methods to identify obesity. One method is to use the medical demographics documented in the IRI Medprofiler panel, where participants were asked whether they had obesity and whether they had been taking any Rx, OTC or dual medications for obesity. Another is to calculate BMI with weight and height reported in IRI Medprofiler panel. We select one method at a time for one regression and find the regression results from models using these two methods are contradictory. In the model with self-reported obesity and medication usage, we find that individuals who are self-reported obese but not taking any medications are less price sensitive on average than the healthy individuals, and the individuals who are self-reported taking medications for obesity are less price sensitive on average than the self-reported obese individuals. Furthermore, obese individuals prefer yogurts with less sugar and protein on average than the healthy ones, while the individuals who are taking obesity medications prefer yogurts with more sugar and protein and less total fat on average than the healthy ones. However, with BMI $\geq$ 30 as the identification strategy for obesity, we find that obese individuals (i.e. BMI $\geq$ 30) are more price sensitive and prefer yogurts with more sugar and protein and less total fat on average than the ones whose BMI $<$ 30.

These two obesity identification strategies can be interpreted separately because they represent two different meanings. The one identifying obesity with self-reported obesity-related survey question represents that an individual admits, or say thinks, he or she is obese, even as shown in the statistical summary, there is a cohort of consumers who think they are obese but their BMIs are lower than 30. The regression results, therefore, show that when knowing themselves is obese, how consumers respond to the price change and variation of product healthfulness. However, besides official diagnosis with obesity by doctors, whether obese or not is almost all determined by participants' own belief. The lack of appropriate knowledge about obesity and medical assistance from professional personals might cause the misspecification of obesity. The second one employs a more objective standard, BMI, based on a method for obesity identification suggested by CDC (April 11, 2017), where obesity is determined by the BMI calculated by authors based on weight and height reported by participants and is not determined by participants' own belief. However, Buckhauser et al. (2009), Rippe et al. (2012) and

Cawley (2015) suggest that BMI fails to distinguish lean mass from fat mass, and therefore fails to correctly identify obesity for muscular individuals and overestimates the incidence of obesity.

One study conducts a similar study, Wang et al. (2017), and finds that in markets with lower obesity rate, the average own-price elasticity for soda is higher in absolute value than that in the markets with higher obesity rate. It is reasonable for unhealthy products such as soda that there is a negative relationship between obesity rate and price elasticity. In our study, we also observe that consumers with BMI < 30 are more price elastic with respect to yogurt products on average than the ones with BMI ≥ 30. However, due to difference in the nature of the product and methodology, it is different to compare our results with Wang et al. (2017). Further studies with more products and other identification method for obesity are needed.

Furthermore, fat level is an easier way for consumers to differentiate yogurt alternatives than the amount of sugar as well as other nutrition information that only printed on the nutrition fact panel. Fat level is often printed on the front of the package or as a part of the product name for yogurts. Some consumers might not understand nutrition fact panel and a simplified nutritional label lowers consumers' information cost, meaning the fat amount labelled on the front of yogurt package is more accessible and understandable for the majority of consumers (Rothman et al. 2006; Visschers et al., 2010; Zhu et al., 2015). It might lead obese individuals to consume more nonfat or low-fat yogurts in order to prevent excessive intake of fat and improve obesity.

There is a discrepancy in own-price elasticity of yogurts between our study and literature. Bonanno (2013) summarizes the average own-price elasticity of yogurts shown in recent studies as well as obtained by his study. Our results regarding choice-level own-price elasticity are close with the results of Di Giacomo (2008) and Richard et al. (2013), where other studies including Bonanno (2013) and Villas-Boas (2007) report higher absolute values than that shown in our study. This discrepancy might be caused by reasons from two aspects. The first possible aspect is the model used in this study, mixed logit. Due to the big heterogeneity in product attributes for products outside the choice set, the products that are not selected being in the choice set are excluded in the study. This increases the probability any products in the choice set purchased. The second possible aspect is the data. We use individual-level data and investigate consumer demand on yogurts at micro-level, where for most of the studies, their data is aggregated to market-level, which allows them to count the effect of exercise or obesity at market-level. Furthermore, since we are not only interested in consumer demand but also interested in health and exercise, we must keep individuals who have well-documented consumer data and medical data at the same time. This force us to shrink the sample size and inevitably causes an imbalanced panel. As shown in the summary statistics of demographics for participants in this study and the ones in the original IRI Consumer and Medprofiler panels, we can clearly see that the participants used in our study have higher education, more African Americans, lower employment rate and more females on average than the ones of the entire IRI Consumer panelists. While we assume all individuals are identical and rational consumers, the uncatchable endogenous factors are likely to affect consumer behavior toward yogurts.

In this study, we investigate whether consumer demand on yogurts link with physical activity and obesity, but we fail to answer any questions related with causality.

Further study can construct a structure model based on economic theories, which might answer some questions such as the substitution pattern between eating healthy and doing exercise. Literature only shows that consuming more fruits and vegetables is easier to be implemented than doing more exercise (Jin & Fan, 2013), and it is difficult to change people's diet in short run even after migration or diagnosed with type 2 diabetes mellitus (Oster, 2018; Allcott et al., 2019). Hence, it is interesting to explore whether there can be a truly substitution between diet and exercise.

References

Allcott, H., Diamond, R., Dube, J-P., Handbury, J., Rahkovsky, I., & Schnell, M. (2019). Food Deserts and the Causes of Nutritional Inequality. *Quarterly Journal of Economics*, 134(4).

Anari, R., Amani, R., & Veissi, M. (2017). Sugar-sweetened beverages consumption is associated with abdominal obesity risk in diabetic patients. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 11, S675-S678.

Bandini, L. G., Vu, D., Must, A., Cyr, H., Goldberg, A., & Dietz, W. H. (1999). Comparison of high-calorie, low-nutrient-dense food consumption among obese and non-obese adolescents. *Obesity Research*, 7(5), 438-443.

Berniell, L., & Valdés, N. (2013). Spillovers of health education at school on parents' physical activity. *Health Economics*, 22(9), 1004-1020.

Bonanno, A. (2012). Some like it healthy: demand for functional and conventional yogurts in the Italian market. *Agribusiness*, 28(1), 67-85.

Bonanno, A. (2013). Functional foods as differentiated products: the Italian yogurt market. *European Review of Agricultural Economics*, 40(1), 45-71.

Bonanno, A. (2016). A hedonic valuation of health and nonhealth attributes in the US yogurt market. *Agribusiness*, 32(3), 299-313.

Bonanno, A., Huang, R., & Liu, Y. (2015). Simulating welfare effects of the European nutrition and health claims' regulation: the Italian yogurt market. *European Review of Agricultural Economics*, 42(3), 499-533.

Bray, G. A., Nielsen, S. J., & Popkin, B. M. (2004). Consumption of high-fructose corn syrup in beverages may play a role in the epidemic of obesity. *The American Journal of Clinical Nutrition*, 79(4), 537-543.

Calle, E. E., Rodriguez, C., Walker-Thurmond, K., & Thun, M. J. (2003). Overweight, obesity, and mortality from cancer in a prospectively studied cohort of US adults. *New England Journal of Medicine*, 348(17), 1625-1638.

Carlucci, D., Stasi, A., Nardone, G., & Seccia, A. (2013). Explaining price variability in the Italian yogurt market: a hedonic analysis. *Agribusiness*, 29(2), 194-206.

Cawley, J. (2004). An economic framework for understanding physical activity and eating behaviors. *American Journal of Preventive Medicine*, 27(3), 117-125.

Cawley, J. (2015). An economy of scales: A selective review of obesity's economic causes, consequences, and solutions. *Journal of Health Economics*, 43, 244-268.

Cawley, J., Frisvold, D., & Meyerhoefer, C. (2013). The impact of physical education on obesity among elementary school children. *Journal of Health Economics*, 32(4), 743-755.

Cawley, J., Meyerhoefer, C., & Newhouse, D. (2007). The impact of state physical education requirements on youth physical activity and overweight. *Health Economics*, 16(12), 1287-1301.

Centers for Disease Control and Prevention (CDC) (2016, June 16). *Defining Adult Overweight and Obesity*. Retrieved on September 12, 2018 from https://www.cdc.gov/obesity/adult/defining.html.

Centers for Disease Control and Prevention (CDC). (2017, April 11). *Defining Adult Overweight and Obesity*. Retrieved on June 20, 2019 from https://www.cdc.gov/obesity/adult/defining.html.

Centers for Disease Control and Prevention (CDC) (2019, May 13). *2008 Physical Activity Guidelines for Americans*. Retrieved on June 1, 2019 from https://www.cdc.gov/physicalactivity/downloads/trends-in-the-prevalence-of-physical-activity-508.pdf.

Chen, M., Sun, Q., Giovannucci, E., Mozaffarian, D., Manson, J. E., Willett, W. C., & Hu, F. B. (2014). Dairy consumption and risk of type 2 diabetes: 3 cohorts of US adults and an updated meta-analysis. *BMC Medicine*, 12(1), 215.

Dawczynski, C., Massey, K. A., Ness, C., Kiehntopf, M., Stepanow, S., Platzer, M., ... & Jahreis, G. (2013). Randomized placebo-controlled intervention with n-3 LC-PUFA-supplemented yoghurt: effects on circulating eicosanoids and cardiovascular risk factors. *Clinical Nutrition*, 32(5), 686-696.

Debnam, J. (2017). Selection Effects and Heterogeneous Demand Responses to the Berkeley Soda Tax Vote. *American Journal of Agricultural Economics*, 99(5), 1172-1187.

Della Torre, S. B., Keller, A., Depeyre, J. L., & Kruseman, M. (2016). Sugar-sweetened beverages and obesity risk in children and adolescents: a systematic analysis on how methodological quality may influence conclusions. *Journal of the Academy of Nutrition and Dietetics*, 116(4), 638-659.

Di Giacomo, M. (2008). GMM estimation of a structural demand model for yogurt and the effects of the introduction of new brands. *Empirical Economics*, 34(3), 537.

Dragone, D., & Savorelli, L. (2012). Thinness and obesity: A model of food consumption, health concerns, and social pressure. *Journal of Health Economics*, 31(1), 243-256.

Fan, M., & Jin, Y. (2013). Obesity and self-control: Food consumption, physical activity, and weight-loss intention. *Applied Economic Perspectives and Policy*, 36(1), 125-145.

Finkelstein, E. A., Linnan, L. A., Tate, D. F., & Birken, B. E. (2007). A pilot study testing the effect of different levels of financial incentives on weight loss among overweight employees. *Journal of Occupational and Environmental Medicine*, 49(9), 981-989.

Gillis, L. J., & Bar-Or, O. (2003). Food away from home, sugar-sweetened drink consumption and juvenile obesity. *Journal of the American College of Nutrition*, 22(6), 539-545.

Gillman, M. W., Pinto, B. M., Tennstedt, S., Glanz, K., Marcus, B., & Friedman, R. H. (2001). Relationships of physical activity with dietary behaviors among adults. *Preventive Medicine*, 32(3), 295-301.

Guarner, F., Perdigon, G., Corthier, G., Salminen, S., Koletzko, B., & Morelli, L. (2005). Should yoghurt cultures be considered probiotic? *British Journal of Nutrition*, 93(6), 783-786.

Guyonnet, D., Chassany, O., Ducrotte, P., Picard, C., Mouret, M., Mercier, C. H., & Matuchansky, C. (2007). Effect of a fermented milk containing Bifidobacterium animalis DN-173 010 on the health-related quality of life and symptoms in irritable bowel syndrome in adults in primary care: a multicentre, randomized, double-blind, controlled trial. *Alimentary Pharmacology & Therapeutics*, 26(3), 475-486.

Hales, C. M., Carroll, M. D., Fryar, C. D., & Ogden, C. L. (2017). *Prevalence of obesity among adults and youth: United States, 2015-2016*. US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics.

Hassan, D., & Monier-Dilhan, S. (2006). National brands and store brands: competition through public quality labels. *Agribusiness*, 22(1), 21-30.

Hu, F. B., & Malik, V. S. (2010). Sugar-sweetened beverages and risk of obesity and type 2 diabetes: epidemiologic evidence. *Physiology & Behavior*, 100(1), 47-54.

Hubert, H. B., Feinleib, M., McNamara, P. M., & Castelli, W. P. (1983). Obesity as an independent risk factor for cardiovascular disease: a 26-year follow-up of participants in the Framingham Heart Study. *Circulation*, 67(5), 968-977.

Humphreys, B. R., & Ruseski, J. E. (2011). An economic analysis of participation and time spent in physical activity. *The B.E. Journal of Economic Analysis & Policy*, 11(1).

Kahn, S. E., Hull, R. L., & Utzschneider, K. M. (2006). Mechanisms linking obesity to insulin resistance and type 2 diabetes. *Nature*, 444(7121), 840-846.

Keast, D. R., Hill Gallant, K. M., Albertson, A. M., Gugger, C. K., & Holschuh, N. M. (2015). Associations between yogurt, dairy, calcium, and vitamin D intake and obesity among US children aged 8–18 years: NHANES, 2005–2008. *Nutrients*, 7(3), 1577-1593.

Lin, C. T. J., Gao, Z., & Lee, J. Y. (2013). Associations between self-reported weight management methods with diet quality as measured by the Healthy Eating Index–2005. *Preventive Medicine*, 57(3), 238-243.

Lopez-Garcia, E., Leon-Muñoz, L., Guallar-Castillon, P., & Rodríguez-Artalejo, F. (2015). Habitual yogurt consumption and health-related quality of life: a prospective cohort study. *Journal of the Academy of Nutrition and Dietetics*, 115(1), 31-39.

Ludwig, D. S., Peterson, K. E., & Gortmaker, S. L. (2001). Relation between consumption of sugar-sweetened drinks and childhood obesity: a prospective, observational analysis. *The Lancet*, 357(9255), 505-508.

Martinez-Gonzalez, M. A., Sayon-Orea, C., Ruiz-Canela, M., De la Fuente, C., Gea, A., & Bes-Rastrollo, M. (2014). Yogurt consumption, weight change and risk of overweight/obesity: the SUN cohort study. *Nutrition, Metabolism and Cardiovascular Diseases*, 24(11), 1189-1196.

Oster, E. (2018). Diabetes and diet: purchasing behavior change in response to health information. *American Economic Journal: Applied Economics*, 10(4), 308-48.

Øvrum, A. (2011). Socioeconomic status and lifestyle choices: evidence from latent class analysis. *Health Economics*, 20(8), 971-984.

Patel, K.R. (2012). Essays in Industrial Organization. PhD dissertation, Northwestern University.

Petrin, A., & Train, K. (2010). A control function approach to endogeneity in consumer choice models. *Journal of Marketing Research*, 47(1), 3-13.

Pi-Sunyer, F. X., Becker, D. M., Bouchard, C., Carleton, R. A., Colditz, G. A., Dietz, W. H., ... & Higgins, M. (1998). Clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults: Executive summary. *American Journal of Clinical Nutrition*, 68(4), 899-917.

Richards, T. J., Allender, W. J., & Hamilton, S. F. (2013). Rivalry in price and location by differentiated product manufacturers. *American Journal of Agricultural Economics*, 95(3), 650-668.

Richards, T. J., & Bonnet, C. (2016). Models of consumer demand for differentiated products (No. 16-741). TSE Working Papers.

Rothman, A. J., Bartels, R. D., Wlaschin, J., & Salovey, P. (2006). The strategic use of gain-and loss-framed messages to promote healthy behavior: How theory can inform practice. *Journal of Communication*, 56(suppl_1), S202-S220.

Sallis, J. F., Story, M., & Lou, D. (2009). Study designs and analytic strategies for environmental and policy research on obesity, physical activity, and diet: recommendations from a meeting of experts. *American Journal of Preventive Medicine*, 36(2), S72-S77.

Skender, M. L., Goodrick, G. K., Del Junco, D. J., Reeves, R. S., Darnell, L., Gotto Jr, A. M., & Foreyt, J. P. (1996). Comparison of 2-year weight loss trends in behavioral treatments of obesity: diet, exercise, and combination interventions. *Journal of the American Dietetic Association*, 96(4), 342-346.

Swinburn, B. A., Caterson, I., Seidell, J. C., & James, W. P. T. (2004). Diet, nutrition and the prevention of excess weight gain and obesity. *Public Health Nutrition*, 7(1a), 123-146.

Train, K. E. (2009). *Discrete Choice Methods with Simulation*. Cambridge university press.

Van Gaal, L. F., Mertens, I. L., & Christophe, E. (2006). Mechanisms linking obesity with cardiovascular disease. *Nature*, 444(7121), 875.

Villas-Boas, S. B. (2007). Vertical relationships between manufacturers and retailers: Inference with limited data. *The Review of Economic Studies*, 74(2), 625-652.

Visschers, V. H., Hess, R., & Siegrist, M. (2010). Health motivation and product design determine consumers' visual attention to nutrition information on food products. *Public Health Nutrition*, 13(7), 1099-1106.

Vucenik, I., Jones, L. P., & McLenithan, J. C. (2014). Linking obesity, metabolism and cancer. *Metabolic Syndrome*, 1-21.

Wang, E. Y. (2015). The impact of soda taxes on consumer welfare: implications of storability and taste heterogeneity. *The RAND Journal of Economics*, 46(2), 409-441.

Wang, E., Rojas, C., & Colantuoni, F. (2017). Heterogeneous Behavior, Obesity, and Storability in the Demand for Soft Drinks. *American Journal of Agricultural Economics*, 99(1), 18-33.

Ware, M. (2018, January 11). *Everything you need to know about yogurt*. Medical News Today. Retrieved on July 18, 2018 from https://www.medicalnewstoday.com/articles/295714.php.

Weyer, C., Funahashi, T., Tanaka, S., Hotta, K., Matsuzawa, Y., Pratley, R. E., & Tataranni, P. A. (2001). Hypoadiponectinemia in obesity and type 2 diabetes: close association with insulin resistance and hyperinsulinemia. *The Journal of Clinical Endocrinology & Metabolism*, 86(5), 1930-1935.

Wilson, M. M., Reedy, J., & Krebs-Smith, S. M. (2016). American diet quality: Where it is, where it is heading, and what it could be. *Journal of the Academy of Nutrition and Dietetics*, 116(2), 302-310.

Yaniv, G., Rosin, O., & Tobol, Y. (2009). Junk-food, home cooking, physical activity and obesity: The effect of the fat tax and the thin subsidy. *Journal of Public Economics*, 93(5-6), 823-830.

Yngve, A., Sjöström, M., Warm, D., Margetts, B., Rodrigo, C. P., & Nissinen, A. (1999). Effective promotion of healthy nutrition and physical activity in Europe requires skilled and competent people; European Master's Programme in Public Health Nutrition. *Public Health Nutrition*, 2(3a), 452-452.

Zhu, C., Lopez, R. A., & Liu, X. (2015). Information cost and consumer choices of healthy foods. *American Journal of Agricultural Economics*, 98(1), 41-53.

Chapter 4

Investigating the Role of the Food Environment on Households' Food-Purchase Healthfulness Using Migration as an Identification Strategy

Introduction

Increasing concern in the U.S. about food-related chronic diseases such as obesity and the disparities across both socioeconomic status and geography has generated numerous studies and reports. For example, the 1990 National Nutrition Monitoring and Related Research Act authorized the federal government to periodically release Dietary Guidelines for Americans (USDA and HHS, 2010). Yet, these guidelines have not reversed the current negative health trends. The disparities of these food-related diseases, especially across geography, have sparked increased interest in the role that the food environment plays in consumer food choices.

Research investigating the association between food environment and households' dietary quality has often (but not always) documented the link. For instance, Inagami et al. (2006) find that the number of grocery stores in neighborhoods, as well as the distance from home to grocery stores, affects households' body mass index (BMI). Holsten (2009) links the food environment to the consumption of fruits and vegetables and individuals' BMIs. In these and other studies, the food environment is quantified by examining numbers of food outlets in neighborhood, distances from home to nearby grocery stores, variation of food categories in stores, availability of healthy foods, presence of regular and fast food restaurants, and healthy food prices (Morland et al., 2002; Bodor et al., 2007; Boder et al., 2008; Morland & Evenson, 2009; Rundle et al., 2009; Ding et al., 2012; Minaker et al., 2013; Ghosh-Dastidar et al., 2014; Couch et al., 2014; Hillier et al., 2015). Sometimes these studies include other factors that might explain a family's or a youth's dietary pattern or health from household itself: for example, some studies include mothers' nutritional knowledge (Campbell et al., 2013), and parent's behavior of how to feed their children and parent's awareness of consuming healthy food (Couch et al., 2014). However, most of aforementioned studies fail to account for the possible endogeneity of the food environment, and thus the effect of the food environment on dietary outcomes is not well identified.

Endogeneity of the food environment is possible or even probable because food retailers' and consumers' selections are linked. On one hand, people intentionally select their residential location for various reasons such housing, family, and employment (Ihrke, June 2014). Their demographic characteristics are likely similar to the neighborhood characteristics. On the other hand, supermarkets, convenience stores, and other food retailers select entry into a market for the purpose of profit maximization. As the result, the numbers of food outlets differ across population density, urbanization and region, and they are also associated with average income in neighborhoods. Full-service restaurants as well as supermarkets and convenience stores are more likely to locate in high- and median-income neighborhoods, while fast-food restaurants are not (Chung and Myers, 1999; Morland, et al., 2002; Powell et al., 2007; Larson et al., 2009). Even among supermarkets, large chain stores do not locate as often as non-chain stores in low-income neighborhood in metropolitan area (Chung and Myers, 1999). Hence, the characteristics of local food environments are not unilateral determined by either households or food outlets, and studies that investigate the role of the food environment on consumers' purchasing behavior must acknowledge this endogeneity. Under these circumstances, a valid identification strategy must involve finding an exogenous shock to households' food environments, and then to examining whether the households' healthfulness of food purchases therefore changes.

Two recent papers, Allcott et al. (2017) and Handbury et al. (2017), address the issue of endogeneity by exploiting an exogenous shock on household's food environment. These two papers merged into one paper, Allcott et al. (2019), but keep the majority of the contents from previous working papers. Handbury et al. (2017) consider three scenarios of changes in food environment: migration across census tracts, the entry or exit of stores, and modification of food categories in stores. They conclude that the disparities in household food consumption are minimally affected by the changes in retail environment. Allcott et al. (2017) investigate two cases, the entry of new supermarkets and households' migration across zip codes and counties, and they find similar conclusions as Handbury et al. (2017). We only focus on one exogenous shock, migration, and investigate this possible influence from a number of different perspectives. We assume that migration, while intentionally decided by households, is not based on the food relative food environments. Thus, the relative difference in the before and after food environments for movers identifies the impact of the food environment on diet quality. To diminish the peer effect of neighborhood, we employ the channel information that is aggregated to county level, and we do not consider movers who move across census tract but still live in the same county.

In this paper, we follow Allcott et al.'s (2019) migration method to investigate an exogenous change in the household food environment but explore a number of different measures for diet quality and the food environment. To measure households' healthfulness of food purchases, we mainly employ monetary index of diet quality such as USDA Scores and Expenditure Score, which denote the overall healthfulness of purchases by comparing the difference between households' expenditure shares of all food categories and the corresponding recommended expenditures. The algorithm for USDA scores is introduced by Volpe and Okrent (2012), and the algorithm for Expenditure Score is introduced by Handbury et al. (2017). Also, the total expenditure and expenditure share on F&V are used to indicate healthfulness. The data used is the purchase record and demographic information of Nielsen Homescan data from 2007 to 2008, where the demographic information including the residential location is on yearly basis. The county-level stores information is collected from Nielsen TDLink data. Hence, we identify whether a household move by its annually reported county and state at the end of each year. For the econometrics analysis, we first perform an ordinary least square (OLS) model to capture the relationship between food environment and households' diet quality. Then using first-difference models for movers we can examine whether switching to a new food environment affect the quality of households' food baskets.

Data and Method

Data

The data used in this study are the Nielsen Homescan Data in 2007 and 2008 acquired by USDA's Economic Research Service. It records more than 40,000 American households' grocery purchases over 48 states on a trip basis, including the information related to purchases such as price, quantity, product brand and so forth at UPC level. It also provides information about the channel and store such as store name, county and state. The data collected households' demographic information on a yearly basis, including households' annual income, race and household members' employment and education, and their residential census tract, county and state. Based on the household county and state reported in the annual demographic file, we identify whether that household moved between each pair of years. Therefore, we calculate households' diet quality using annual year purchases on natural year basis, so that it can be comparable over household and time. After removing observations with missing values or observations that do not exist in both years in a paired-year sample, the numbers of movers and non-movers are shown in Table 4-1. The movers across county are about 533 households, which is about one percent of the sample in each period. The movers across state only are 257, about 0.7 of entire households from 2007 to 2008.

Table **4-1**: Numbers of movers and non-movers over the year.

	Mover	Non-Mover
Across County	533	38414
Across State	257	38690
Total	38947	

Notes: The numbers of households moved across county include those moved across state.

Food store information, including the numbers of stores per channel aggregated at the county level, is obtained from the Nielsen TDLink data. The types of channels consist of conventional supermarkets, limited assortment supermarkets, warehouse grocery, supercenters, club stores, convenience stores, military/general merchandise/mass merchandisers, natural/gourmet supermarket, superette, and dollar stores. The numbers of conventional and limited assortment supermarkets and warehouse grocery are aggregated up as the number of supermarkets, and convenience store is defined as the convenience store in this study. We take the means of supermarkets and convenience stores over quarters as numbers of stores in each year.

The coding scheme of demographic and food environment variables are shown in Table 4-2, and their statistical summaries are shown in Table 4-3. There are about 90 supermarkets on average in each household's county. The mean number of convenience stores for each household is about 356. Instead of using the numbers of supermarkets and convenience stores to represent the food environment, this study also employs the ratio of numbers of supermarkets to numbers of convenience stores, SM/CS. Supermarkets are considered to be a better channel for providing healthy food than convenience stores because supermarkets sell a variety of healthy foods with relatively low prices. Hence, a lower SM/CS ratio denotes that the food environment is more likely to provide local households healthy food. However, as a relative measure, the SM/CS ratio fails to represent variation and quantity of food outlets in neighborhood. The average annual household income is about $62,000 and the median is $55,000, which is slightly higher than the median annual household income in 2010, which is 51,444, according to a US Census statistic report in 2012. The average number of household members is about two individuals. About 30 percent of households' male head and about 35 percent of households' female head have college or higher degree. In term of employment, about 45 percent of households their male head are full-time employed, and about 6 percent of households are part-time employed. About 38 percent and 15 percent of households, their female head are full-time and part-time employed, respectively. About 10 percent of households are African Americans, about 3 percent are Asians and about 5 percent are Hispanic households. About 11 percent households live in rural county.

Table **4-2**: Coding scheme of demographic and food environment variables.

Variables	Coding
SMs	The number of supermarkets in county.
CS	The number of convenience stores in county.
SM/CS	The ratio of numbers of supermarket to numbers of convenience store in county.
Move	Dichotomous variable that indicates whether that household moved across county or state. It equals to one if household moves over the year and equals to zero for the rest.
Income	Household annual income. Income=2,500 if household income is from 0 to $4999; Income=6,500 if household income is from $5,000 to $7,999; Income=9,000 if household income is from $8,000 to $10,000; Income=11,000 if household income is from $10,000 to $11,999; Income=13,500 if household income is from $12,000 to $14,999; Income=17,500 if household income is from $15,000 to $19,999; Income=22,500 if household income is from $20,000 to $24,999; Income=27,500 if household income is from $25,000 to $29,999; Income=32,500 if household income is from $30,000 to $34,999; Income=37,500 if household income is from $35,000 to $39,999; Income=42,500 if household income is from $40,000 to $44,999; Income=47,500 if household income is from $45,000 to $49,999; Income=55,000 if household income is from $50,000 to $59,999; Income=65,000 if household income is from $60,000 to $69,999; Income=85,000 if household income is from $70,000 to $99,999; Income=112,500 if household income is from $100,000 to $124,999;

	Income=137,500 if household income is from $125,000 to $149,999; Income=175,000 if household income is from $150,000 to $199,999; Income=300,000 if household income is equal to or above $200,000;
CollegeM	Equals to one if the male head has a college or higher degree, and equals to zero if he has not or if there is not male head.
CollegeF	Equals to one if the female head has a college or higher degree, and equals to zero if he has not or if there is not female head.
FTempM	Whether male head is full-time employed or not. Equals to one if he is, and equals to zero for the rest.
PTempM	Whether male head is part-time employed or not. Equals to one if he is, and equals to zero for the rest.
FTempF	Whether female head is full-time employed or not. Equals to one if she is, and equals to zero for the rest.
PTempF	Whether female head is part-time employed or not. Equals to one if she is, and equals to zero for the rest.
HHSize	Household size. Integral from 1 to 9.
AfrAm	Equals to one if the household head is African American.
Asian	Equals to one if the household head is Asian.
Hispanic	Equals to one if the household head is Hispanic.
Rural	Equals to one if the household's residential county is rural, and equals to zero if it is not.

Table **4-3**: Statistical summary of demographic and food environment variables.

Variables	Entire Sample	Movers Across County	Movers Across State
		Mean	
Income	61763.4	61102.6	66554.5
	(43474.3)	(47922.4)	(51628.2)
HHSize	2.3	2.3	2.2
	(1.3)	(1.3)	(1.4)
SMs	88.2	87.8	91.8
	(135.6)	(133.6)	(135.8)
CStores	356.7	357.5	342.6
	(457.7)	(445.6)	(390.5)
SM/CStore	0.2	0.2	0.2
	(0.1)	(0.1)	(0.1)
		Percentage	
CollegeM	30.3%	35.5%	32.7%
CollegeF	35.0%	39.8%	42.4%
FTempF	37.5%	40.0%	37.4%
PTempF	15.3%	11.8%	10.9%
FTempM	45.1%	44.7%	41.2%
PTempM	5.8%	4.1%	4.3%
AfrAm	9.5%	9.8%	10.1%
Asian	2.6%	4.3%	2.7%

Hispanic	4.8%	4.9%	3.1%
Rural	11.1%	10.7%	6.6%
N	77894	533	257

Notes: the values in parenthesis are the standard deviations. The number of observations in the entire sample contains households in 2007 and 2008, so it counts one household twice. The numbers of households in paired-year sample is double counted, where the real number of households is 38947.

Diet Quality Measurement

The Nielsen Homescan data we used for this study lack of detailed information of products nutritional fact panel, therefore we can only measure households' healthfulness of food baskets depending on the monetary approach. The first approach is to use USDA scores described by Volpe and Okrent (2012) and Volpe et al. (2013), or called Expenditure Score described by Handbury et al. (2017). The second approach is to use the expenditure share and total expenditure of F&V.

The methodology of USDA scores or Expenditure Score is firstly aggregating grocery expenditure up to 24 food categories and computing the expenditure share of each food category, secondly comparing the household's shopping basket expenditure shares with their corresponding recommended expenditure shares by food category, and finally reporting the square of summation of these differences (Carlson et al., 2007; Volpe and Okrent, 2012). Its algorithm is shown as Equation (1) and (2). The list of food categories is shown in Table 4-4. This approach has been employed to measure diet quality across the U.S. on a variety of issues documented by Volpe and Okrent (2012), Adjemian and Volpe (2012), Volpe et al. (2013), Rudi and Cakir (2014), Chen et al. (2016) and Handbury et al. (2017). Volpe and Okrent (2012) introduce the initial algorithm of USDA score, and Volpe et al. (2013) introduce a following supplementary algorithm. In this additional algorithm, food groups are intentionally categorized by authors into healthy and unhealthy foods. They include all fruits and vegetables, whole grain products, low-fat dairy products, low-fat meat, poultry, fish, nuts, eggs, oils and water in the healthy foods, and include the rest in unhealthy food. This process of categorizing healthy and unhealthy foods is controversial, so we will not report this approach in this study.

Two equations for calibrating USDA score are shown as following.

(4.1) $$USDA\ Score\ 1_{ict} = [\sum_c (e_{ict} - \bar{e}_{ic})^2]^{-1}$$

(4.2) $$USDA\ Score\ 2_{ict} = [\sum_c (e_{ict} - \bar{e}_{ic})^2\ |e_{ict} > 0]^{-1}$$

where e_{ict} denotes the actual household i's expenditure share of food category c at time t, and $\bar{e}_{ic}$ denotes the recommendation expenditure share of food category c for household i. Equation (4.1) is the basic naïve approach. Equation (4.2) excludes the food categories that the household has not ever purchased. As shown above, this score is the inverse of the difference between real expenditure shares and recommendations for all food categories, so a higher value means that household's expenditure shares are closer to the recommendation levels. Hence, by comparing the USDA scores between households, we can understand the relative dietary quality. However, absolute USDA scores are difficult for representing an absolute healthfulness for a household.

In term of the construction of recommended levels for all food groups, two methods are employed in literature. One is to simply used the household-level described in Volpe and Okrent (2012). The recommended expenditure shares shown in Table 4-4. However, the food choice is likely to vary across households, because of their differences in numbers of household members and demographic characteristics such as gender, age and ethnicity. Children tend to consume more dairy products and vegetables, and boys' consumption of animal products increases as they growing but this does not apply to girl (Nicklaus et al., 2005). The alternative method described by Handbury et al. (2017) provides more variation by constructing the recommendations of expenditure shares for each household based on its number of members, and gender and age of each member. Each household members are weighted as $w_{adult} = \frac{0.5+0.5n_{adult}}{0.5n_{adult}+0.5n_{adult}^2+0.3n_{kid}n_{adult}}$ for adults and $w_{kid} = \frac{0.3}{0.5n_{adult}+0.5+0.3n_{kid}}$ for children, such that $w_{adult}n_{adult} + w_{kid}n_{kid} = 1$ holds. The source of recommended expenditure shares is documented by Carlson et al. (2007) as individual-level recommended expenditure shares of Liberal Food Plans. Indeed, the food choice for a household can be influence by many factors beyond gender and age, but similar approach using an index to measure diet quality is demonstrated to be valid and reliable (Guenther et al., 2008; Guenther et al., 2013; Guenther et al., 2014).

Table **4-4**: Food categories and their recommended expenditure shares by QFAHPD.

CNPP food category	Household shopping basket expenditure share USDA Food Plan (%)	Household-specified shopping basket expenditure share (%)
Grains		
All whole-grain products	10.09	8.30 (3.42)
Non-whole-grain breads, cereals, rice, pasta, pies, pastries, snacks, and flours	6.10	3.68 (1.87)
Vegetables		
All potato products	1.77	1.81 (0.85)
Dark-green vegetables	5.59	5.21

		(3.63)
Orange vegetables	2.61	1.55 (0.55)
Canned and dry beans, lentils, and peas (legumes)	8.32	5.21 (2.69)
Other vegetables	8.66	8.01 (4.11)
Fruits		
Whole fruits	16.49	14.09 (6.82)
Fruit juices	1.86	1.41 (0.83)
Milk products		
Whole-milk products	0.86	2.42 (2.38)
Lower fat and skim milk and low-fat yogurt	8.77	8.63 (3.27)
All cheese (including cheese soup and sauce)	0.60	0.40 (0.35)
Meat and beans		
Beef, pork, veal, lamb, and game	5.31	5.21 (2.27)
Chicken, turkey, and game birds	2.69	2.60 (1.20)
Fish and fish products	11.92	7.10 (3.70)
Bacon, sausages, and luncheon meats (including spreads)	0.91	0.39 (0.27)
Nuts, nut butters, and seeds	3.16	2.81 (1.46)
Eggs and egg mixtures	0.12	0.13 (0.06)
Other foods		
Fats and condiments	1.79	1.32 (0.67)
Coffee and tea	0.02	0.04 (0.03)
Soft drinks, sodas, fruit drinks, and ades (including rice beverages)	1.33	0.80 (0.56)
Sugars, sweets, and candies	0.41	0.39 (0.28)
Soups	0.51	0.71 (0.45)
Frozen or refrigerated entrees (including pizza, fish sticks, and frozen meals)	0.18	0.05 (0.05)

Notes: QFAHPD denotes Quarterly Food-at-Home Price Database. The expenditure shares are calculated by Volpe and Okrent (2012) based on Carlson et al. (2007). This category consists of whole grain breads, rice, pasta, pastries, flours, cereals, hot cereal mixes, popcorn, and other whole grain snacks. This category consists of whole-milk, yogurt and cream, and milk drinks and desserts. The low-fat beef, pork, veal, lamb and game are categorized as healthful foods by Volpe et al. (2013). This category consists of table fats, oils, and salad dressings, and gravies, sauces, condiments and spices. This category consists of ready-to-serve and condensed soups, and dry soups. This household-specified household recommended expenditure shares are calculated by authors following Handbury et al. (2017) based on the Liberal Food Plans documented in Carlson et al. (2007). The values in parenthesis are standard deviations.

Our second approach we used to measure diet quality is to use households' expenditure share of all F&V since F&V are widely considered as the proxy of healthy diet (Eikenberry and Smith, 2004; Evans et al., 2015). Here for identification, we also use the total expenditure of both all F&V and fresh F&V where the fresh F&V only consists of fresh fruit, dark-green vegetables, orange vegetables and other vegetables.

The statistical summary of these diet quality indices and variables of interest are shown in Table 4-5. According to our calculation, the average USDA Score 1 and 2 are about 8.3 and 8.9, respectively, where the means of these scores documented in Volpe and Okrent (2012) are about 7.8 and 9.3, respectively. This disparity might because of different categorization for food groups. All these diet quality measurements are statistically correlated between each other, and the results of Pearson correlation coefficients are shown in Table 4-6.

Table **4-5**: Statistical summary of diet quality measurements.

Diet Quality Measures	Entire Sample	Move Across County		Move Across State	
		Movers	Non-movers	Movers	Non-movers
USDA Score 1	8.34	8.34	8.34	8.59	8.34
	(2.90)	(2.83)	(2.90)	(2.96)	(2.90)
USDA Score 2	8.88	8.92	8.88	9.11	8.88

	(3.05)	(2.95)	(3.05)	(3.08)	(3.05)
Exp Score	9.47	9.56	9.47	9.69	9.47
	(3.92)	(3.86)	(3.92)	(3.93)	(3.92)
Exp Share of Fresh F&V	0.10	0.10	0.10	0.10	0.10
	(0.06)	(0.06)	(0.06)	(0.05)	(0.06)
Total Exp of Fresh F&V	183.26	165.98	183.50	172.87	183.33
	(150.11)	(141.33)	(150.22)	(138.29)	(150.19)
Exp Share of All F&V	0.13	0.13	0.13	0.13	0.13
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
Total Exp of All F&V	235.93	218.49	236.17	225.65	236.00
	(181.99)	(172.77)	(182.11)	(168.99)	(182.08)
N	77894	1,066	76,828	514	77,380

Notes: The number of observations in the entire sample contains households in 2007 and 2008. The numbers of households in paired-year sample is double counted, and the real numbers of all households are 38947. The values in parenthesis are standard deviations.

Table **4-6**: The correlation between diet quality measurements.

	USDA Score 1	USDA Score 2	Exp. Score	Fresh F&V Exp Share	Fresh F&V Exp	All F&V Exp Share	All F&V Exp
USDA Score 1	1.00	0.96	0.80	0.68	0.51	0.65	0.48
USDA Score 2		1.00	0.78	0.67	0.45	0.65	0.43
Exp. Score			1.00	0.49	0.50	0.48	0.49
Fresh F&V Exp Share				1.00	0.65	0.87	0.56
Fresh F&V Exp					1.00	0.56	0.93
All F&V Exp Share						1.00	0.61
All F&V Exp							1.00

Econometrics Analysis

In the econometrics analysis, this study first examines the correlation between food environment and households' healthfulness of food purchases via ordinary least square (OLS) models. For this "naïve" model, which does not account for the endogeneity of the local food environment, the OLS equation is

(4.3) $$H_{it} = \alpha F_{it} + \beta_1 X_{it} + \beta_2 D_i + \varepsilon_{it},$$

where H_{it} is household i's healthfulness of food purchase at year t. F_{it} is a vector of food environment for household i at year t. X_{it} is a vector of household i's demographic information that varies over time such as income and employment. D_i is a vector of household i's time-invariant characteristics such as race and gender. α, β_1 and β_2 are coefficients. ε_i is the error term.

To identify the causal effect of disparity in food environment, we derive a first-difference model for paired-year sample, so this model becomes a version of a treatment effect model where household migration causes the treatment, i.e., an exogenous change in the food environment. At year *t*, a household's food-purchase healthfulness is

(4.4) $$H_{it} = \alpha F_{it} + \phi Imp_t + \gamma M_{it} + \delta F_{it} M_{it} + \theta F_{it} Imp_t + \beta_1 X_{it} + \beta_2 D_i + \lambda_t + \varepsilon_{it},$$

where Imp_t is a dichotomous variable that denotes whether households' food environment improved or not over the year, and Imp_t and $Imp_{\mathrm{t-1}}$ equals one if household ith's food environment improved over the year. M_{it} is a dichotomous variable, and in a pair-year sample, $M_{it} = 1$ and $M_{i,t-1} = 1$ if household i moved from time $t - 1$ to t, and equals zero otherwise. α, ϕ, γ, δ, θ, β_1 and β_2 are coefficients. λ_t is a year fixed-effect.

Then we can rewrite Equation (4.4) for the past period, $t-1$, as

$$H_{i,t-1} = \alpha F_{i,t-1} + \phi Imp_{t-1} + \gamma M_{i,t-1} + \delta F_{i,t-1}M_{i,t-1} + \theta F_{i,t-1}Imp_{t-1} + \beta_1 X_{i,t-1} + \beta_2 D_i + \varepsilon_{i,t-1}. \tag{4.5}$$

When we subtract Equation (4.2) from Equation (4.1), all the time-invariant variables drop out including M_{it} and D_i. Furthermore, we assume the change of food environment in the identical location is minimal over the year, so we only keep observations whose i =movers. Finally, we obtain the first-difference model,

$$\Delta H_i = \alpha \Delta F_i + \theta \Delta F_i Imp_i + \beta \Delta X_i + \epsilon_i, \tag{4.6}$$

where ΔH_i is the change in household i's healthfulness of food purchases, ΔF_i is a vector that denotes the change in food environment, Imp_i is a dichotomous variable and equals to one if ΔF_i is positive, and ΔX_i is a vector of changes in household i's time-variant demographic characteristics. ϵ_i is the error term. Here, the interaction term $\Delta F_i Imp_i$ captures the effect of the improvement in food environment, and ΔF_i capture the effect for those households whose food environment did not improve. The demographic characteristics consist of income and employment. The clustered standard error on migration level is employed. For example, for household moved across state, their regression standard error is clustered to state level.

Results

The naïve models where the USDA Score 2, Expenditure Score or the F&V expenditure share is regressed on the food environment measure and household characteristics for the paired-year sample of 2007-2008 are shown in Table 4-7, 4-8 and

4-9, respectively. As shown in the first column in Table 4-7, the SM/CS ratios are positive correlated with households' diet quality. For instance, for one unit increase in SM/CS, the USDA Score 2 increases 0.48. This result suggests that diet quality, measured with either the USDA Score 2 or the F&V expenditure share, is associated with an improved food environment: namely, households who have better access to supermarkets relative to convenience stores are more likely to have better diet quality. Table 4-7 also presents two additional regressions where the number of supermarkets and convenience stores are both included instead of their ratio. When living in a neighborhood with large numbers of convenience stores, households' diet quality tends to lower, and they are more likely to consume less fruit and vegetable, but their marginal effects are minimal. The opposite result partially holds for supermarket: households F&V expenditure share increases as the number of supermarkets increases, but supermarket's impact on the USDA Score 2 is statistically insignificant. Also, as shown in Table 4-7, the diet quality of rural households is worse than those who live outside rural area. This provides evidence that the existence of rural "food dessert" issue can limit rural households' consumption of fruit and vegetable.

The results of the regression of Expenditure Score as shown in Table 4-8 are not consistent with the other two models. It shows that for one unit increase in SM/CS, the Expenditure Score decreases by 0.45, and the increases in both numbers of supermarkets and convenience stores lead to the decrease of Expenditure Score. When we perform a more naïve model, where Expenditure Score is only regressed on food environment variables, then the results are consistent with the other two. After adding up demographic variables in the model, we find that some households' characteristics, such as income, play important roles on determining the relationship between food environment and Expenditure Score. In the meanwhile, this phenomenon does not show in the other two cases. This might result from the special setting of Expenditure Score where its recommendation levels are computed based on household characteristics such as household size as well as member's age and gender. Hence, endogeneity of the food environment and household characteristics might cause these problematic results, which also demonstrates the importance of controlling endogeneity in similar study.

Most of household's demographic characteristics are associated with households' diet quality and purchases of fruit and vegetables. All else equal, households that are non-white, well-educated, and have higher income tend to have a higher USDA Score 2. On the other hand, households with larger sizes and those where the female or male head is employed on a full- or part-time basis have lower diet quality and less expenditure on fruit and vegetable. With limited household income, the trade-off between quantity and quality forces household heads to decrease the overall expenditure on expensive foods such as those "healthy foods" and to evenly distribute resources on every member. Household heads who need to spend time to work might have less leisure time and available time for trips to supermarkets in a week, and therefore they might prefer to visit convenience stores in close neighborhood instead of a supermarket which requires more shopping time and longer traveling distance.

Table **4-7**: The regression of USDA Score 2 on food environment and household characteristics using 2007-2008 paired-year sample.

Variable	Coeff.	SE	Coeff.	SE
Intercept	8.59***	0.04	8.76***	0.03
SM/CS	0.48***	0.13		

SMs			0.0002	0.00
CStores			-0.0002***	0.00
Income	0.00***	0.00	0.00***	0.00
CollegeM	0.52***	0.03	0.53***	0.03
CollegeF	0.44***	0.02	0.44***	0.02
FTempF	-0.45***	0.03	-0.45***	0.03
PTempF	-0.02	0.03	-0.03	0.03
FTempM	-0.45***	0.03	-0.45***	0.03
PTempM	0.05	0.05	0.05	0.05
HHSize	-0.12***	0.01	-0.12***	0.01
AfrAm	0.14***	0.04	0.17***	0.04
Asian	0.31***	0.07	0.36***	0.07
Hispanic	0.21***	0.05	0.26***	0.05
Rural	-0.07**	0.04	-0.14***	0.04
year2	0.11***	0.02	0.11***	0.02
Adj R^2		0.033		0.033
N				77894

Notes: *** denotes p-value≤0.01, ** denotes p-value≤0.05 and * denotes p-value≤0.1.

Table **4-8**: The regression of Expenditure Score on food environment and household characteristics using 2007-2008 paired-year sample.

Variable	Coeff.	SE	Coeff.	SE
Intercept	7.34***	0.05	7.37***	0.04
SM/CS	-0.45***	0.17		
SMs			-0.0003	0.0002
CStores			-0.0003***	0.0001
Income	0.00***	0.00	0.00***	0.00
CollegeM	1.02***	0.03	1.02***	0.03
CollegeF	0.21***	0.03	0.20***	0.03
FTempF	-0.49***	0.03	-0.49***	0.03
PTempF	0.03	0.04	0.02	0.04
FTempM	0.73***	0.03	0.72***	0.03
PTempM	1.49***	0.06	1.49***	0.06
HHSize	0.34***	0.01	0.34***	0.01
AfrAm	-0.30***	0.05	-0.24***	0.05
Asian	0.03	0.08	0.09	0.08
Hispanic	0.40***	0.06	0.49***	0.06
Rural	0.33***	0.04	0.23***	0.04
year2	0.05	0.03	0.05*	0.03
Adj R^2		0.113		0.115
N			77894	

Notes: *** denotes p-value≤0.01, ** denotes p-value≤0.05 and * denotes p-value≤0.1.

Table **4-9**: The regression of all F&V expenditure share on food environment and household characteristics using 2007-2008 paired-year sample.

Variable	Coeff.	SE	Coeff.	SE
Intercept	0.13***	0.00	0.13***	0.00
SM/CS	0.02***	0.00		
SMs			0.00001***	0.00
CStores			-0.000003***	0.00
Income	0.00***	0.00	0.00***	0.00
CollegeM	0.01***	0.00	0.01***	0.00
CollegeF	0.01***	0.00	0.01***	0.00
FTempF	-0.01***	0.00	-0.01***	0.00
PTempF	-0.00***	0.00	-0.00***	0.00
FTempM	-0.01***	0.00	-0.01***	0.00
PTempM	-0.00**	0.00	-0.00**	0.00
HHSize	-0.00***	0.00	-0.00***	0.00
AfrAm	0.01***	0.00	0.01***	0.00
Asian	0.02***	0.00	0.02***	0.00
Hispanic	-0.00***	0.00	0.00***	0.00
Rural	-0.00***	0.00	-0.00***	0.00
year2	0.00	0.00	-0.00	0.00
Adj R^2		0.049		0.048
N				77894

Notes: *** denotes p-value≤0.01, ** denotes p-value≤0.05, and * denotes p-value≤0.1.

The results of the first-difference models just for moved households are shown in Table 4-10, 4-11, 4-12, 4-13, 4-14 and 4-15. Table 4-10, 4-11 and 4-12 shows the results of USDA score 2, Expenditure Score and all F&V expenditure share, respectively, using SM/CS ratio to denote food environment. In these models, the food-environment measure is included on its own but also interacted with a dummy variable indicating that move was to an improved food environment. In general, the first-difference food-environment measure is only statistically significant for the interacted variable, a result that generally suggested only moves to an improved food environment have an impact on diet quality. Moreover, this result, discussed in more detail next, is not always found.

Using the results shown in the first column of Table 4-10 without-demographic covariates, one can see that the estimated coefficient for SM/CS without interaction is not statistically different from zero. On the other hand, the estimated coefficient for SM/CS interacted with the improved food environment dummy (Imp) is positive and significant. Taken together, these results suggest that a change the supermarket-convenience store ratio does cause improved diet quality when the change reflected an improved food environment. For those whose SM/CS ratio improved, one-unit increase in the ratio increases the USDA Score 2 by 5.37. While the results suggest that higher SM/CS ratio leads a better diet quality in neighborhood is found in all cases, the result is most often not statistically significant. That is, the result is not significant in the case of migration across state or in the cases of regressions of all F&V expenditure share found in Table 4-12. The general trend of the regressions of USDA Score 2 and Expenditure Score is similar regarding coefficients of food environment, but all these coefficients are not significant in Table 4-11. Except the full-employment of female or male head, all demographic variables in Table 4-10, 4-11 and 4-12 are not significant. The results regarding demographic information align with aforementioned OLS models, which is that finding a new full-time job actually makes entire family's diet quality worse.

Table **4-10**: The first-difference model of USDA Score 2 on the ratio of stores using 2007-2008 paired-year sample.

Variables	Across County		Across State	
	Coeff.	SE	Coeff.	SE
Without Demo				
ΔSM/CS*Imp	5.37***	1.68	3.97	2.81
ΔSM/CS	-1.93	1.29	-1.99	1.96
R^2	0.019		0.012	
With Demo				

ΔSM/CS*Imp	5.30***	1.69	4.08	2.84
ΔSM/CS	-1.91	1.29	-1.90	2.02
ΔFTempF	-0.46*	0.24	-0.44	0.40
ΔFTempM	-0.04	0.30	-0.09	0.65
ΔPTempF	0.25	0.26	0.39	0.31
ΔPTempM	-0.29	0.38	0.13	0.57
ΔIncome	0.00	0.00	0.00	0.00
R^2	0.028		0.022	
N	533		257	

Notes: *** denotes p-value≤0.01, ** denotes p-value≤0.05, and * denotes p-value≤0.1.

Table **4-11**: The first-difference model of Expenditure Score on the ratio of stores using 2007-2008 paired-year sample.

Variables	Across County		Across State	
	Coeff.	SE	Coeff.	SE
Without Demo				
ΔSM/CS*Imp	2.49	1.89	2.03	3.04
ΔSM/CS	-0.48	1.29	-0.51	1.85
R^2	0.004		0.003	
With Demo				
ΔSM/CS*Imp	2.47	1.85	2.20	2.97
ΔSM/CS	-0.43	1.27	-0.30	1.97
ΔFTempF	-0.75***	0.28	-0.94**	0.40
ΔFTempM	1.47***	0.54	2.30**	1.10
ΔPTempF	0.31	0.32	0.51	0.40
ΔPTempM	0.58	0.56	0.99	0.96
ΔIncome	0.00	0.00	0.00	0.00
R^2	0.049		0.088	
N	533		257	

Notes: *** denotes p-value≤0.01, ** denotes p-value≤0.05, and * denotes p-value≤0.1.

Table **4-12**: The first-difference model of all F&V expenditure share using 2007-2008 paired-year sample.

Variables	Across County		Across State	
	Coeff.	SE	Coeff.	SE
Without Demo				
ΔSM/CS*Imp	0.04	0.03	0.01	0.06
ΔSM/CS	-0.02	0.02	-0.01	0.05

R^2	0.002		0.0004	
With Demo				
ΔSM/CS*Imp	0.04	0.03	0.02	0.06
ΔSM/CS	-0.02	0.02	-0.01	0.05
ΔFTempF	-0.01	0.01	-0.02	0.02
ΔFTempM	0.00	0.01	-0.01	0.01
ΔPTempF	0.00	0.01	0.00	0.01
ΔPTempM	0.01	0.01	0.01	0.01
ΔIncome	0.00	0.00	0.00	0.00
R^2	0.010		0.014	
N	533		257	

Table 4-13, 4-14 and 4-15 show the regression results where numbers of supermarkets and convenience stores represent the local food environment instead of the supermarket-convenience store ratio. In general, only a couple results suggest that an improved food environment causes improved diet quality. As shown in Table 4-13, for movers who move across county to a new neighborhood, neither the change in supermarket numbers nor the change in convenience store numbers has a statistically significant impact on diet quality as measured by the USDA Score 2, the Expenditure Score, or the F&V expenditure share. Moreover, this null result holds for moves that both decrease or improve the food environment.

However, for movers who move across state lines, the change in the food environment can have a causal impact on diet quality when the F&V expenditure share is sued to reflect diet quality. More specifically, when the F&V expenditure share denotes households' diet quality, in all cases, households moving to a new environment with more supermarkets and less convenience stores tend to spend more on fruit and vegetable. In term of statistical significance however, we can conclude only a move across state lines to new location with more convenience stores causes less expenditure share on fruits and vegetables. This also holds when we use total expenditure instead of expenditure share to measure diet quality.

Table **4-13**: The first-difference model of USDA Score 2 on numbers of stores using 2007-2008 paired-year sample.

	Across County		Across State	
Variables	Coeff.	SE	Coeff.	SE
Without Demo				
ΔSM*ImpSM	-0.0020	0.0023	0.0010	0.0030
ΔCS*ImpCS	0.0004	0.0006	-0.0007	0.0010
ΔSM	0.0004	0.0015	-0.0014	0.0026
ΔCS	-0.0004	0.0005	0.0003	0.0010
R^2	0.008		0.006	
With Demo				
ΔSM*ImpSM	-0.0021	0.0023	0.0008	0.0031
ΔCS*ImpCS	0.0004	0.0006	-0.0006	0.0011
ΔSM	0.0004	0.0015	-0.0011	0.0027
ΔCS	-0.0004	0.0005	0.0003	0.0010
ΔFTempF	-0.4283*	0.2469	-0.3972	0.4577
ΔFTempM	-0.0622	0.3030	-0.0799	0.6516
ΔPTempF	0.2695	0.2585	0.3533	0.3483
ΔPTempM	-0.2467	0.3799	0.2095	0.5939
ΔIncome	0.0000	0.0000	0.0000	0.0000
R^2	0.017		0.014	
N	533		257	

Table **4-14**: The first-difference model of Expenditure Score on numbers of stores using 2007-2008 paired-year sample.

	Across County		**Across State**	
Variables	Coeff.	SE	Coeff.	SE
Without Demo				
ΔSM*ImpSM	-0.0005	0.0025	0.0033	0.0026
ΔCS*ImpCS	-0.0005	0.0006	-0.0016	0.0010
ΔSM	-0.0001	0.0017	-0.0025	0.0024
ΔCS	0.0001	0.0004	0.0009	0.0008
R^2	0.005		0.008	

With Demo				
ΔSM*ImpSM	-0.0005	0.0023	0.0029	0.0026
ΔCS*ImpCS	-0.0005	0.0006	-0.0016	0.0010
ΔSM	-0.0001	0.0015	-0.0016	0.0023
ΔCS	0.0001	0.0004	0.0008	0.0009
ΔFTempF	-0.7662***	0.2842	-0.9393*	0.4764
ΔFTempM	1.4577***	0.5446	2.3388**	1.1191
ΔPTempF	0.2805	0.3239	0.4494	0.4001
ΔPTempM	0.6526	0.5683	1.0812	1.0146
ΔIncome	0.0000	0.0000	-0.0000	0.0000
R^2	0.005		0.092	
N	533		257	

Notes: *** denotes p-value≤0.01, ** denotes p-value≤0.05, and * denotes p-value≤0.1.

Table **4-15**: The first-difference model of all F&V expenditure share on numbers of stores using 2007-2008 paired-year sample.

	Across County		**Across State**	
Variables	Coeff.	SE	Coeff.	SE
Without Demo				
ΔSM*ImpSM	0.00002	0.00005	0.00011	0.00007
ΔCS*ImpCS	-0.00002	0.00001	-0.00004**	0.00002
ΔSM	-0.00002	0.00002	-0.00007	0.00004
ΔCS	0.00000	0.00001	0.00002	0.00001
R^2	0.009		0.020	
With Demo				
ΔSM*ImpSM	0.00001	0.00005	0.00010	0.00007
ΔCS*ImpCS	-0.00001	0.00001	-0.00004**	0.00002
ΔSM	-0.00001	0.00002	-0.00007	0.00005
ΔCS	0.00000	0.00001	0.00002	0.00002
ΔFTempF	-0.01003	0.00856	-0.01822	0.01889
ΔFTempM	0.00018	0.00714	-0.00635	0.01409
ΔPTempF	-0.00157	0.00702	-0.00514	0.01334
ΔPTempM	0.01010	0.00667	0.01143	0.01252
ΔIncome	0.00000	0.00000	0.00000	0.00000
R^2	0.016		0.033	
N	533		257	

Notes: *** denotes p-value≤0.01, ** denotes p-value≤0.05, and * denotes p-value≤0.1.

Robustness Check

As a check, we calculate different measures to represent household's healthfulness of food purchases instead of using aforementioned three measures. The corresponding regressions results of are shown in the Table 4-16, 4-17, 4-18 and 4-19. These different models provide similar results. In most cases, only a few of the variables of interest are statistically significant.

Furthermore, instead of using data in year 2007 and 2008, we also employed paired-year data of 2008 and 2009, and 2009 and 2010. The potential influence at macro-level on households' food and migration choices is difficult for us to identify when we use multiple period data. Hence, we only present the results of one period.

Table **4-16**: The regression results of the first-difference models with time-variant demographic characteristics.

		Move Across County		Move Across State	
Dep Variables	Indep Variables	Coeff.	S.E.	Coeff.	S.E.
USDA Score1	ΔSM/CS * Imp	4.20***	2.80	3.21	1.42
	ΔSM/CS	-1.94*	1.84	-2.12	1.05
	ΔFTempF	-0.39*	0.35	-0.30	0.20
	ΔFTempM	-0.06	0.65	-0.06	0.31
	ΔPTempF	0.26	0.29	0.51*	0.23
	ΔPTempM	-0.27	0.58	-0.01	0.34
	ΔIncome	0.00	0.00	0.00	0.00
Fresh F&V Share	ΔSM/CS * Imp	0.03	0.05	0.03	0.03
	ΔSM/CS	-0.02	0.04	-0.02	0.02
	ΔFTempF	0.00	0.01	-0.01	0.00
	ΔFTempM	0.00	0.01	0.00	0.01
	ΔPTempF	0.00	0.01	-0.01	0.00
	ΔPTempM	0.01**	0.01	0.02*	0.01
	ΔIncome	0.00	0.00	0.00	0.00
Fresh F&V Exp	ΔSM/CS * Imp	8.20	134.04	34.50	84.13
	ΔSM/CS	16.26	104.54	-42.11	66.71

	ΔFTempF	-14.13	17.52	0.13	11.59
	ΔFTempM	-1.21	33.01	-8.74	16.63
	ΔPTempF	5.10	13.19	6.64	12.56
	ΔPTempM	10.06	28.31	37.87	18.88
	ΔIncome	0.00	0.00	0.00	0.00
All F&V Exp	ΔSM/CS * Imp	-10.31	172.77	14.50	99.47
	ΔSM/CS	38.18	138.63	-4.69	80.69
	ΔFTempF	-17.82	20.68	7.19	15.02
	ΔFTempM	-12.23	47.00	-21.16	21.54
	ΔPTempF	9.60	18.99	27.84	15.85
	ΔPTempM	-7.90	32.58	7.67	21.02
	ΔIncome	0.00	0.00	0.00	0.00

Notes: *** denotes p-value≤0.01, ** denotes p-value≤0.05, and * denotes p-value≤0.1.

Table **4-17**: The regression results of the first-difference models without time-variant demographic characteristics.

		Move Across County		Move Across State	
Dep Variables	Indep Variables	Coeff.	S.E.	Coeff.	S.E.
USDAScore2	ΔSM/CS * Imp	5.37***	1.68	3.97	2.81
	ΔSM/CS	-1.93	1.29	-1.99	1.96
Fresh F&V Share	ΔSM/CS * Imp	0.04	0.03	0.03	0.05
	ΔSM/CS	-0.02	0.02	-0.02	0.04
Fresh F&V Exp	ΔSM/CS * Imp	13.87	83.63	39.95	128.56
	ΔSM/CS	16.06	66.73	-39.51	105.46
All F&V Exp	ΔSM/CS * Imp	-4.13	98.51	15.46	166.62
	ΔSM/CS	38.20	80.34	-3.36	139.11

Notes: *** denotes p-value≤0.01, ** denotes p-value≤0.05, and * denotes p-value≤0.1.

Table **4-18**: The regression results of the first-difference models with time-variant demographic characteristics.

Dep Variables	Indep Variables	Move Across County		Move Across State	
		Coeff.	S.E.	Coeff.	S.E.
USDA Score1	ΔSM * SMImp	-0.00	0.00	0.00	0.00
	ΔCS * CSImp	0.00	0.00	-0.00	0.00
	ΔSM	0.00	0.00	-0.00	0.00
	ΔCS	-0.00	0.00	0.00	0.00
	ΔFTempF	-0.37*	0.21	-0.27	0.41
	ΔFTempM	-0.08	0.31	-0.05	0.64

	ΔPTempF	0.28	0.23	0.47	0.32
	ΔPTempM	-0.23	0.34	0.05	0.62
	ΔIncome	0.00	0.00	-0.00	0.00
Fresh F&V Share	ΔSM * SMImp	-0.00	0.00	0.00*	0.00
	ΔCS * CSImp	-0.00	0.00	-0.00**	0.00
	ΔSM	-0.00	0.00	-0.00**	0.00
	ΔCS	-0.00	0.00	0.00**	0.00
	ΔFTempF	-0.00	0.00	-0.01	0.01
	ΔFTempM	0.00	0.01	-0.00	0.01
	ΔPTempF	-0.00	0.00	-0.01	0.01
	ΔPTempM	0.01**	0.01	0.02*	0.01
	ΔIncome	0.00	0.00	0.00	0.00
Fresh F&V Exp	ΔSM * SMImp	0.06	0.11	0.30**	0.13
	ΔCS * CSImp	-0.05*	0.03	-0.15***	0.04
	ΔSM	0.03	0.07	-0.22**	0.10
	ΔCS	-0.00	0.02	0.07*	0.03
	ΔFTempF	-14.20	11.64	1.97	16.24
	ΔFTempM	-0.55	16.44	-6.48	34.26
	ΔPTempF	3.31	12.64	1.37	14.79
	ΔPTempM	15.16	18.82	46.05	30.62
	ΔIncome	0.00	0.00	0.00	0.00
All F&V Exp	ΔSM * SMImp	0.09	0.12	0.27	0.16
	ΔCS * CSImp	-0.07**	0.03	-0.15***	0.05
	ΔSM	0.05	0.09	-0.13	0.13
	ΔCS	-0.00	0.03	0.04	0.04
	ΔFTempF	-18.59	15.03	10.37	19.72
	ΔFTempM	-11.43	21.29	-16.50	47.95
	ΔPTempF	6.86	15.83	22.72	20.69
	ΔPTempM	-0.39	20.50	20.48	33.65
	ΔIncome	0.00	0.00	-0.00	0.00

Notes: *** denotes p-value≤0.01, ** denotes p-value≤0.05, and * denotes p-value≤0.1.

Table **4-19**: The regression results of the first-difference models without time-variant demographic characteristics.

Dep Variables	Indep Variables	Move Across County		Move Across State	
		Coeff.	S.E.	Coeff.	S.E.
USDA Score1	ΔSM * SMImp	-0.00	0.00	0.00	0.00
	ΔCS * CSImp	0.00	0.00	-0.00	0.00

	ΔSM	0.00	0.00	-0.00	0.00
	ΔCS	-0.00	0.00	0.00	0.00
Fresh F&V Exp Share	ΔSM * SMImp	-0.00	0.00	0.00*	0.00
	ΔCS * CSImp	-0.00	0.00	-0.00**	0.00
	ΔSM	-0.00	0.00	-0.00**	0.00
	ΔCS	-0.00	0.00	0.00**	0.00
Fresh F&V Tot Exp	ΔSM * SMImp	0.08	0.11	0.31**	0.14
	ΔCS * CSImp	-0.05*	0.03	-0.15***	0.05
	ΔSM	0.02	0.07	-0.22**	0.10
	ΔCS	-0.00	0.02	0.07*	0.04
All F&V Tot Exp	ΔSM * SMImp	0.11	0.12	0.28*	0.16
	ΔCS * CSImp	-0.07**	0.03	-0.15***	0.06
	ΔSM	0.05	0.09	-0.14	0.13
	ΔCS	-0.00	0.03	0.05	0.05

Notes: *** denotes p-value≤0.01, ** denotes p-value≤0.05, and * denotes p-value≤0.1.

Discussions and Conclusions

This study explores a first-difference approach to examine whether an exogenous shock on households' food environment influences the healthfulness of their food purchases. This study primary addresses an issue, the endogeneity of food environment, which has been ignored all but a couple studies linking the food environment and an individual's or household's diet quality. Because the numbers of food outlets are not only determined by firms themselves, but also indirectly associate by neighborhood demographic characteristics, many of these association-based results are hampered by endogeneity concerns. To control this endogeneity, we follow Handbury et al. (2017), Allcott et al. (2017) and Allcott et al. (2019) and consider migration as the exogenous shock to households' food environment. We explore along this approach from more diverse perspectives so that we can provide more information.

The results in the majority cases align with Handbury et al. (2017), Allcott et al. (2017) and Allcott et al. (2019), which is that exogenous change of households' food environment will not affect their diet quality. However, this study also finds that for movers across county whose food environment improved over that year, their average USDA score 2 is significantly higher than before the move. The same result does not hold for households moving to a worse food environment. Furthermore, for movers across state who moved to a new environment with more convenience stores, the move leads to a decrease in diet quality. This result suggests that convenience stores can play a negative role in diet quality. However, a move to an area with more supermarkets than their previous location does not seem to affect diet quality. In addition, effect of the move varies for the scale of the move, i.e., across county versus across state lines.

Based on aforementioned results of this study, we can conclude that the impact of changing food environment on households' healthfulness of food purchases is minimal in general. Hence, many policies about the improvement of food environment might not be as effective as policymakers expected. This study provides evidence that if we design policy from an appropriate angle, households can benefit from it via reforming their shopping habit. We understand that more supermarkets in neighborhood encourages households to purchase more healthy foods, but our results show this difficult and expensive policy may have minimal impact on diet quality. However, policies aimed at convenience stores might fare better. Perhaps policies that encourage convenience stores to provide more fruit and vegetable or policies that encourage entry of healthy convenience stores may be more effective. A good example of this sort of policy is Philadelphia's Healthy Corner Store Initiative. Future research may explore this policy's impact on diet quality more closely.

References

Adjemian, M. K., & Volpe, R. J. (2012, August). Alcohol consumption and food-at-home dietary quality in the United States. In Agricultural & Applied Economics Association Annual Meeting, Seattle, Washington.

Allcott, H., Diamond, R., & Dubé, J. P. (2017). The geography of poverty and nutrition: Food deserts and food choices across the United States (No. w24094). National Bureau of Economic Research.

Allcott, H., Diamond, R., Dube, J-P., Handbury, J., Rahkovsky, I., & Schnell, M. (2019). Food Deserts and the Causes of Nutritional Inequality. *Quarterly Journal of Economics*, 134(4).

American Cancer Society (2015). Excess body weight: A major health issue in America. Retrieved Nov 15, 2017, from https://www.cancer.org/cancer/cancer-causes/diet-physical-activity/body-weight-and-cancer-risk/health-issues.html

Ayala, G. X., Baquero, B., & Klinger, S. (2008). A systematic review of the relationship between acculturation and diet among Latinos in the United States: implications for future research. Journal of the American Dietetic Association, 108(8), 1330-1344.

Basu, S., Wimer, C., & Seligman, H. K. (2016). Cost of living, healthy food acquisition, and the Supplemental Nutrition Assistance Program.

Bodor, J. N., Rose, D., Farley, T. A., Swalm, C., & Scott, S. K. (2008). Neighbourhood fruit and vegetable availability and consumption: the role of small food stores in an urban environment. *Public Health Nutrition*, 11(4), 413-420.

Bonanno, A. (2012). Food deserts: Demand, supply, and economic theory. *Choices*, 27(3), 1-5.

Bonanno, A., Chenarides, L., & Volpe III, R. (2013, May). The Size vs. Health Trade-off in Lower-Income Households' Food Choices: The Case of Fluid Milk. In Agricultural & Applied Economics Association's 2013 AAEA & CAES Joint Annual Meeting, Washington, DC, August (pp. 4-6).

Bureau of Labor Statistics (2016). Table 1800. Region of residence: Annual expenditure means, shares, standard errors, and coefficients of variation, Consumer Expenditure Survey, 2016. Retrieved Nov 15, 2017, from https://www.bls.gov/cex/2016/combined/region.pdf.

Burns, C. (2004). Effect of migration on food habits of Somali women living as refugees in Australia. *Ecology of Food and Nutrition*, 43(3), 213-229.

Camhi, S. M., Evans, E. W., Hayman, L. L., Lichtenstein, A. H., & Must, A. (2015). Healthy eating index and metabolically healthy obesity in US adolescents and adults. *Preventive Medicine*, 77, 23-27.

Campbell, K. J., Abbott, G., Spence, A. C., Crawford, D. A., McNaughton, S. A., & Ball, K. (2013). Home food availability mediates associations between mothers' nutrition knowledge and child diet. *Appetite*, 71, 1-6.

Carlson, A., Lino, M., & Fungwe, T. V. (2007). The low-cost, moderate-cost, and liberal food plans, 2007 (No. 45850). United States Department of Agriculture, Center for Nutrition Policy and Promotion.

Caspi, C. E., Sorensen, G., Subramanian, S. V., & Kawachi, I. (2012). The local food environment and diet: a systematic review. *Health & Place*, 18(5), 1172-1187.

Centers for Disease Control and Prevention (2013). State Indicator Report on Fruits and Vegetables, 2013. Atlanta, GA: Centers for Disease Control and Prevention, U.S. Department of Health and Human Services.

Centers for Disease Control and Prevention (2014, March 6). General Food Environment Resources. Retrieved Nov 15, 2017, from https://www.cdc.gov/healthyplaces/healthtopics/healthy food/general.htm.

Centers for Disease Control and Prevention (2017, May 8). Data, Trend and Maps. National Center for Chronic Disease Prevention and Health Promotion, Division of Nutrition, Physical Activity, and Obesity. Retrieved Nov 15, 2017, from https://www.cdc.gov/nccdphp/dnpao/data-trends-maps/index.html.

Chaparro, M. P., Langellier, B. A., Wang, M. C., Koleilat, M., & Whaley, S. E. (2015). Effects of parental nativity and length of stay in the US on fruit and vegetable intake among WIC-enrolled preschool-aged children. *Journal of Immigrant and Minority Health*, 17(2), 333-338.

Chen, D., Jaenicke, E. C., & Volpe, R. J. (2016). Food environments and obesity: household diet expenditure versus food deserts. *American Journal of Public Health*, 106(5), 881-888.

Chung, C., & Myers, S. L. (1999). Do the poor pay more for food? An analysis of grocery store availability and food price disparities. *Journal of Consumer Affairs*, 33(2), 276-296.

CNPP, USDA (1995). The Healthy Eating Index. https://www.cnpp.usda.gov/sites/default/files/healthy_eating_index/HEI89-90report.pdf

Couch, S. C., Glanz, K., Zhou, C., Sallis, J. F., & Saelens, B. E. (2014). Home food environment in relation to children's diet quality and weight status. *Journal of the Academy of Nutrition and Dietetics*, 114(10), 1569-1579.

Dean, W. R., & Sharkey, J. R. (2011). Rural and urban differences in the associations between characteristics of the community food environment and fruit and vegetable intake. *Journal of Nutrition Education and Behavior,* 43(6), 426-433.

Ding, D., Sallis, J. F., Norman, G. J., Saelens, B. E., Harris, S. K., Kerr, J., ... & Glanz, K. (2012). Community food environment, home food environment, and fruit and vegetable intake of children and adolescents. *Journal of Nutrition Education and Behavior*, 44(6), 634-638.

Eikenberry, N., & Smith, C. (2004). Healthful eating: perceptions, motivations, barriers, and promoters in low-income Minnesota communities. *Journal of the American Dietetic Association*, 104(7), 1158-1161.

Ervin, R. B. (2008). Healthy Eating Index scores among adults, 60 years of age and over, by sociodemographic and health characteristics: United States, 1999–2002. Adv Data, 395, 1-16.

Ervin, R. B. (2011). Healthy Eating Index–2005 total and component scores for adults aged 20 and over: National Health and Nutrition Examination Survey, 2003–2004. *Natl Health Stat Report*, 44, 1-9.

Evans, R., Kawabata, M., & Thomas, S. (2015). Prediction of fruit and vegetable intake: The importance of contextualizing motivation. *British Journal of Health Psychology*, 20(3), 534-548.

Feskanich, D., Rockett, H. R., & Colditz, G. A. (2004). Modifying the Healthy Eating Index to assess diet quality in children and adolescents. *Journal of the American Dietetic Association*, 104(9), 1375-1383.

Ford, C. N., Poti, J. M., Ng, S. W., & Popkin, B. M. (2017). SSB taxes and diet quality in US preschoolers: estimated changes in the 2010 Healthy Eating Index. *Pediatric Obesity*, 12(2), 146-154.

Fungwe, T., Guenther, P. M., Juan, W., Hiza, H., & Lino, M. (2009). The quality of children’s diets in 2003–04 as measured by the healthy eating index-2005. *Nutrition Insight*, 43.

Ghosh-Dastidar, B., Cohen, D., Hunter, G., Zenk, S. N., Huang, C., Beckman, R., & Dubowitz, T. (2014). Distance to store, food prices, and obesity in urban food deserts. *American Journal of Preventive Medicine*, 47(5), 587-595.

Glanz, K., & Yaroch, A. L. (2004). Strategies for increasing fruit and vegetable intake in grocery stores and communities: policy, pricing, and environmental change. *Preventive Medicine*, 39, 75-80.

Gregory, C. A., Smith, T. A., & Wendt, M. (2011). How Americans rate their diet quality: an increasingly realistic perspective. Economic Research Service, US Department of Agriculture.

Guenther, P. M., Casavale, K. O., Reedy, J., Kirkpatrick, S. I., Hiza, H. A., Kuczynski, K. J., ... & Krebs-Smith, S. M. (2013). Update of the healthy eating index: HEI-2010. *Journal of the Academy of Nutrition and Dietetics*, 113(4), 569-580.

Guenther, P. M., Kirkpatrick, S. I., Reedy, J., Krebs-Smith, S. M., Buckman, D. W., Dodd, K. W., ... & Carroll, R. J. (2013). The Healthy Eating Index-2010 Is a Valid and Reliable Measure of Diet Quality According to the 2010 Dietary Guidelines for Americans1–3. *The Journal of Nutrition*, 144(3), 399-407.

Guenther, P. M., Reedy, J., & Krebs-Smith, S. M. (2008). Development of the healthy eating index-2005. *Journal of the American Dietetic Association*, 108(11), 1896-1901.

Guo, X., Warden, B. A., Paeratakul, S., & Bray, G. A. (2004). Healthy eating index and obesity. *European Journal of Clinical Nutrition*, 58(12), 1580-1586.

Hales CM, Carroll MD, Fryar CD, Ogden CL (2017). Prevalence of obesity among adults and youth: United States, 2015–2016. NCHS data brief, no 288. Hyattsville, MD: National Center for Health Statistics.

Handbury, J., Rahkovsky, I., & Schnell, M. (2017). *Is the focus on food deserts fruitless? Retail access and food purchases across the socioeconomic spectrum*. Working paper.

Herforth, A., & Ahmed, S. (2015). The food environment, its effects on dietary consumption, and potential for measurement within agriculture-nutrition interventions. *Food Security*, 7(3), 505-520.

Hillier, A., Smith, T., Cannuscio, C. C., Karpyn, A., & Glanz, K. (2015). A discrete choice approach to modeling food store access. *Environment and Planning B: Planning and Design*, 42(2), 263-278.

Himmelgreen, D. A., Pérez-Escamilla, R., Martinez, D., Bretnall, A., Eells, B., Peng, Y., & Bermúdez, A. (2004). The longer you stay, the bigger you get: length of time and language use in the US are associated with obesity in Puerto Rican women. *American Journal of Physical Anthropology*, 125(1), 90-96.

Holmboe-Ottesen, G., & Wandel, M. (2012). Changes in dietary habits after migration and consequences for health: a focus on South Asians in Europe. *Food & Nutrition Research*, 56(1), 18891.

Holsten, J. E. (2009). Obesity and the community food environment: a systematic review. *Public Health Nutrition*, 12(3), 397-405.

Ihrke, D. (June 2014). Reason for Moving: 2012 to 2013. Current population reports. Washington, DC: US Census Bureau, 20, 574.

Inagami, S., Cohen, D. A., Finch, B. K., & Asch, S. M. (2006). You are where you shop: grocery store locations, weight, and neighborhoods. *American Journal of Preventive Medicine*, 31(1), 10-17.

Jiménez-Cruz, A., Wojcicki, J. M., Bacardí-Gascón, M., Castellón-Zaragoza, A., García-Gallardo, J. L., Schwartz, N., & Heyman, M. B. (2011). Maternal BMI and migration status as predictors of childhood obesity in Mexico. *Nutricion* *hospitalaria: organo oficial de la Sociedad Espanola de Nutricion Parenteral y Enteral*, 26(1), 187.

Kant, A. K. (1996). Indexes of overall diet quality: a review. *Journal of the American Dietetic Association*, 96(8), 785-791.

Kennedy, E., Ohls, J., Carlson, S., & Fleming, K. (1995). The healthy eating index: design and applications. *Journal of the American Dietetic Association*, 95(10), 1103-1108.

Kim, J., & Chan, M. M. (2004). Acculturation and dietary habits of Korean Americans. *British Journal of Nutrition*, 91(3), 469-478.

Kirkpatrick, S. I., Reedy, J., Butler, E. N., Dodd, K. W., Subar, A. F., Thompson, F. E., & McKinnon, R. A. (2014). Dietary assessment in food environment research: a systematic review. *American Journal of Preventive Medicine*, 46(1), 94-102.

Krebs-Smith, S. M., Guenther, P. M., Subar, A. F., Kirkpatrick, S. I., & Dodd, K. W. (2010). Americans do not meet federal dietary recommendations. *The Journal of Nutrition*, 140(10), 1832-1838.

Kudo, Y., Falciglia, G. A., & Couch, S. C. (2000). Evolution of meal patterns and food choices of Japanese-American females born in the United States. *European Journal of Clinical Nutrition*, 54(8), 665.

Larson, N. I., Story, M. T., & Nelson, M. C. (2009). Neighborhood environments: disparities in access to healthy foods in the US. *American Journal of Preventive Medicine*, 36(1), 74-81.

Lee, S. K., Sobal, J., & Frongillo, E. A. (1999). Acculturation, food consumption, and diet-related factors among Korean Americans. *Journal of Nutrition Education*, 31(6), 321-330.

Lesser, I. A., Gasevic, D., & Lear, S. A. (2014). The association between acculturation and dietary patterns of South Asian immigrants. *PloS One*, 9(2), e88495.

Lin, B. H. (2017, April 3). Shares of Food Commodities Purchased Away From Home Vary by Commodity. United States Department of Agriculture, Economic Research Service. Retrieved Nov 15, 2017, from https://www.ers.usda.gov/amber-waves/2017/april/shares-of-food-commodities-purchased-away-from-home-vary-by-commodity/.

Lv, N., & Cason, K. L. (2004). Dietary pattern change and acculturation of Chinese Americans in Pennsylvania. *Journal of the American Dietetic Association*, 104(5), 771-778.

Lytle, L. A., & Sokol, R. L. (2017). Measures of the food environment: A systematic review of the field, 2007–2015. *Health & Place*, 44, 18-34.

Mancino, L., Kuchler, F., & Leibtag, E. (2008). Getting consumers to eat more whole-grains: the role of policy, information, and food manufacturers. *Food Policy*, 33(6), 489-496.

McCullough, M. L., & Willett, W. C. (2006). Evaluating adherence to recommended diets in adults: the Alternate Healthy Eating Index. *Public Health Nutrition*, 9(1a), 152-157.

Minaker, L. M., Raine, K. D., Wild, T. C., Nykiforuk, C. I., Thompson, M. E., & Frank, L. D. (2013). Objective food environments and health outcomes. *American Journal of Preventive Medicine*, 45(3), 289-296.

Montez, J. K., & Eschbach, K. (2008). Country of birth and language are uniquely associated with intakes of fat, fiber, and fruits and vegetables among Mexican-American women in the United States. *Journal of the American Dietetic Association*, 108(3), 473-480.

Morland, K. B., & Evenson, K. R. (2009). Obesity prevalence and the local food environment. *Health & Place*, 15(2), 491-495.

Morland, K., Wing, S., Roux, A. D., & Poole, C. (2002). Neighborhood characteristics associated with the location of food stores and food service places. *American Journal of Preventive Medicine*, 22(1), 23-29.

Neuhouser, M. L., Thompson, B., Coronado, G. D., & Solomon, C. C. (2004). Higher fat intake and lower fruit and vegetables intakes are associated with greater acculturation among Mexicans living in Washington State. *Journal of the American Dietetic Association*, 104(1), 51-57.

Nunn, M. E., Braunstein, N. S., Kaye, E. K., Dietrich, T., Garcia, R. I., & Henshaw, M. M. (2009). Healthy eating index is a predictor of early childhood caries. *Journal of Dental Research*, 88(4), 361-366.

Opare-Obisaw, C., Fianu, D. A., & Awadzi, K. (2000). Changes in family food habits: the role of migration. *International Journal of Consumer Studies*, 24(3), 145-149.

Pan, Y. L., Dixon, Z., Himburg, S., & Huffman, F. (1999). Asian students change their eating patterns after living in the United States. *Journal of the American Dietetic Association*, 99(1), 54-57.

Powell, L. M., Slater, S., Mirtcheva, D., Bao, Y., & Chaloupka, F. J. (2007). Food store availability and neighborhood characteristics in the United States. P*reventive Medicine*, 44(3), 189-195.

Reedy, J., Krebs-Smith, S. M., Miller, P. E., Liese, A. D., Kahle, L. L., Park, Y., & Subar, A. F. (2014). Higher diet quality is associated with decreased risk of all-cause, cardiovascular disease, and cancer mortality among older adults. *The Journal of Nutrition*, 144(6), 881-889.

Rehm, C. D., Monsivais, P., & Drewnowski, A. (2015). Relation between diet cost and Healthy Eating Index 2010 scores among adults in the United States 2007–2010. *Preventive Medicine*, 73, 70-75.

Renzaho, A., & Burns, C. (2006). Post-migration food habits of sub-Saharan African migrants in Victoria: A cross-sectional study. *Nutrition & Dietetics*, 63(2), 91-102.

Rudi, J., & Cakir, M. (2015, March). The Effect of Grocery Shopping Frequency on the Healthfulness of Food Purchases. In 143rd Joint EAAE/AAEA Seminar.

Rundle, A., Neckerman, K. M., Freeman, L., Lovasi, G. S., Purciel, M., Quinn, J., ... & Weiss, C. (2009). Neighborhood food environment and walkability predict obesity in New York City. *Environmental Health Perspectives*, 117(3), 442.

Savoca, M. R., Arcury, T. A., Leng, X., Bell, R. A., Chen, H., Anderson, A., ... & Quandt, S. A. (2009). The diet quality of rural older adults in the South as measured by healthy eating index-2005 varies by ethnicity. *Journal of the American Dietetic Association*, 109(12), 2063-2067.

Schwingshackl, L., & Hoffmann, G. (2015). Diet quality as assessed by the Healthy Eating Index, the Alternate Healthy Eating Index, the Dietary Approaches to Stop Hypertension score, and health outcomes: a systematic review and meta-analysis of cohort studies. *Journal of the Academy of Nutrition and Dietetics*, 115(5), 780-800.

Semega, J., Fontenot, K., & Kollar, M. (2017, September). Income and Poverty in the United States: 2016. United States Census Bureau. Retrieved Nov 15, 2017, from https://www. census.gov/content/dam/Census/library/publications/2017/demo/P60-259.pdf.

Shatenstein, B., Nadon, S., Godin, C., & Ferland, G. (2005). Diet quality of Montreal-area adults needs improvement: estimates from a self-administered food frequency questionnaire furnishing a dietary indicator score. *Journal of the American Dietetic Association*, 105(8), 1251-1260.

Tande, D. L., Magel, R., & Strand, B. N. (2010). Healthy Eating Index and abdominal obesity. *Public Health Nutrition*, 13(2), 208-214.

Todd, J. E., Leibtag, E., & Penberthy, C. (2011, June). Geographic differences in the relative price of healthy foods. EIB-78, U.S. Department of Agriculture, Economic Research Service.

Tseng, M., Wright, D. J., & Fang, C. Y. (2015). Acculturation and dietary change among Chinese immigrant women in the United States. *Journal of Immigrant and Minority Health*, 17(2), 400-407.

U.S. Department of Agriculture and U.S. Department of Health and Human Services (2010, December). Dietary Guidelines for Americans, 2010. 7th Edition, Washington, DC: U.S. Government Printing Office.

Van Hook, J., Baker, E., Altman, C. E., & Frisco, M. L. (2012). Canaries in a coalmine: Immigration and overweight among Mexican-origin children in the US and Mexico. *Social Science & Medicine*, 74(2), 125-134.

Variyam, J. N., & Blaylock, J. (1998). Unlocking the mystery between nutrition knowledge and diet quality. *Food Review-Washington DC*., 21, 21-28.

Volpe, R., & Okrent, A. M. (2012). Assessing the healthfulness of consumers' grocery purchases. United States Department of Agriculture, Economic Research Service.

Volpe, R., Okrent, A., & Leibtag, E. (2013). The effect of supercenter-format stores on the healthfulness of consumers' grocery purchases. *American Journal of Agricultural Economics*, 95(3), 568-589.

Wu, B., & Smith, C. (2016). Acculturation and environmental factors influencing dietary behaviors and body mass index of Chinese students in the United States. *Appetite*, 103, 324-335.

Ye, X., Scott, T., Gao, X., Maras, J. E., Bakun, P. J., & Tucker, K. L. (2013). Mediterranean diet, healthy eating index 2005, and cognitive function in middle-aged and older Puerto Rican adults. *Journal of the Academy of Nutrition and Dietetics*, 113(2), 276-281.

Yoo, S., Baranowski, T., Missaghian, M., Baranowski, J., Cullen, K., Fisher, J. O., ... & Nicklas, T. (2006). Food-purchasing patterns for home: a grocery store-intercept survey. *Public Health Nutrition*, 9(3), 384-393.

Zoellner, J., You, W., Connell, C., Smith-Ray, R. L., Allen, K., Tucker, K. L., ... & Estabrooks, P. (2011). Health literacy is associated with healthy eating index scores and sugar-sweetened beverage intake: findings from the rural Lower Mississippi Delta. *Journal of the American Dietetic Association*, 111(7), 1012-1020.

Chapter 5

Conclusions

This dissertation consists of three essays on different topics but with a common goal, exploring potential links between dietary choice, health and healthy behaviors and food environment. It contributes to literature from several perspectives. Essay 1 provides a comprehensive analysis about the links between dietary choice and physical activity, obesity, type 2 diabetes and medication usage. Essay 2 investigates the associations of the consumer demand on yogurts with physical activity and obesity with micro-level method and data and estimates the average own-price elasticities by physical activity and obesity. Essay 3 employs a new method to control the endogeneity of food environment and estimates the effect of an exogenous shock on the healthfulness of American households' food baskets.

Essay 1 estimates the associations of consumers' expenditure shares of several food categories with physical activity, obesity and type 2 diabetes and medication usages for these two diseases. Compared with existing studies on similar topics, Essay 1 measures diet quality with expenditure shares of some representative food categories, including fruits and vegetables, snack and chips, yogurts, regular soft drinks, diet soft drinks and bottled water, and presents the link between diet quality, health and healthy behaviors from several perspectives. In this essay, I find that physically active individuals are likely to have a better diet quality on average, spending more share of their grocery budget on healthy foods such as fruits and vegetables and yogurts and spending less share on unhealthy foods such as snacks and chips, regular soft drinks and diet soft drinks. I do not find a significant link between physical activity and expenditure share of bottled water. The endogeneity of physical activity leads to an underestimation of the association between physical activity and expenditure shares of fruits and vegetables, snacks and chips, yogurts and diet soft drinks. Furthermore, I find the incidence of obesity positively associates with the expenditure shares of unhealthy foods such as snack and chips and diet soft drinks and negatively associates with the expenditure shares of healthy foods such as fruits and vegetables and yogurts. There is no statistically significant association between obesity and expenditure share of regular soft drinks and bottled water. However, when considering medication usage for obesity and using a mixed method for obesity identification that identify obesity with self-reported obesity and body mass index (BMI) together in one model, the results are mixed. Also, I find that the incidence of type 2 diabetes mellitus positively associates with the expenditure shares of snacks and chips and diet soft drinks and negatively associates with the expenditure shares of fruits and vegetables, yogurts and diet soft drinks. When distinguishing diabetic medication users and nonusers, I find that individuals who have been diagnosed with type 2 diabetes but not taking any Rx, OTC or dual medications behave similar with healthy individuals in expenditure distribution on grocery.

Essay 2 estimates consumer demand on yogurts and its associations with physical activity and obesity. I find that individuals who do exercise most days in a week are more price sensitive on average than the ones who rarely or never exercise but they are less price sensitive on average than the ones who do exercise some days in a week. Physically active consumers prefer yogurts with less sugar and calcium and more protein on average than the ones who rarely or never exercise. For the model of obesity, I employ BMI ≥ 30 as the identification strategy for obesity. I find that obese individuals (i.e. BMI ≥ 30) are more price sensitive and prefer yogurts with more sugar and protein and less fat on average than the ones whose BMI < 30. However, when we employ the self-reported obesity to identify obesity, we obtain a regression result that is contradictory to the one using BMI for obesity identification.

Essay 3 employs migration as an identification strategy, meant to capture the endogeneity of food environments. We estimate the association between the change in diet quality and the change in food environment for households who moved across counties or states over the year. Essay 3 finds that even there is occasionally a significant link between the improvement in diet quality and the change in food environment, this link is not consistently appeared in every case.

Based on these three essays of this dissertation, we do find that consumers' preference on a particular product, for instance, yogurts, is correlated with their demographic characteristics such as education, employment, gender, income and race as well as health and healthy behaviors. Consumers with a healthy lifestyle, for instance, doing exercise regularly, distribute more share of their budget on healthy foods such as fruits and vegetables and distribute less on unhealthy foods such as snacks and chips than the ones who are physically inactive. This is also true when I employ an alternative measure for dietary choice. As shown in Essay 2, physically active consumers prefer yogurts with lower level of total fat and sugar and higher level of protein than the ones who do not exercise in most days in a week. Furthermore, the incidence of obesity or type 2 diabetes also associates with consumers' dietary choice.

Even it is difficult for us to identify causality with our models, we do find the existence of this association between diet, health and healthy behaviors. However, while we do count the variation of households with household demographics as control in the models, unobservable factors might lead to an endogeneity issue. This is the issue that is investigated in Essay 3. Essay 3 as well as several recent studies such as Oster (2018) and Allcott et al. (2019) show that the majority of individuals do not significant change their overall diet quality after migration or being diagnosed with type 2 diabetes in short run. But for chronical diseases such as obesity that potentially affect people in a fair long period of time, further study needs to be done.

References

Allcott, H., Diamond, R., Dube, J-P., Handbury, J., Rahkovsky, I., & Schnell, M. (2019). Food Deserts and the Causes of Nutritional Inequality. *Quarterly Journal of Economics*, 134(4).

Oster, E. (2018). Diabetes and diet: purchasing behavior change in response to health information. *American Economic Journal: Applied Economics*, 10(4), 308-48.

VITA

Zefeng (Billy) Dong

Email: mintbilly@gmail.com

Current Employment

Quantitative Associate Sept. 2019
Wells Fargo & Company, San Francisco, CA

Education

Ph.D. in Agricultural, Environmental and Regional Economics 2019
The Pennsylvania State University, University Park, PA

M.S. in Food and Resource Economics 2015
University of Florida, Gainesville, FL

B.A. in Economics 2013
Tianjin University of Science & Technology, Tianjin, China

Publication

Dong, Z., Gao, Z., & McFadden, B. (2017). Impact of food choice on sodium intake patterns from multiple NHANES surveys. *Appetite*, 109, 144-153.

Conference Presentations

Dong, Z., Jaenicke, E., Wrenn, D., & Kuhns, A. (2018). "*The effect of exogenous change in food environment on households' healthfulness of food purchases.*" Selected paper, Annual Meeting, Agricultural & Applied Economics Association, Washington D. C., August.

Dong, Z. Gao, Z. & Lee, J. (2015). "*Food choice and sodium intake in the American diet.*" Selected paper, Annual Meeting, Southern Agricultural Economics Association, Atlanta, GA, February.

Dong, Z. & Guan, Z. (2015). "*The efficiency analysis of greenhouse production by two-stage DEA: a case of petunia growers in Michigan.*" Selected paper, Annual Meeting, Southern Agricultural Economics Association, Atlanta, GA, February.

Research & Teaching Experience

Research Assistant, Penn State, PA 8/2015-5/2019
Supervisor: Edward Jaenicke

Teaching Assistant, Penn State, PA 8/2017-12/2017
Supervisor: Edward Jaenicke

Updated on October 27, 2019