Essays on Health Economics:

Trans Fat Policies, Commuting, Physical Activity, and Body Mass Index in the US

by

Ayesha Ali

THESIS

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Economics

in the Department of Economics at

Lancaster University

Lancaster, UK

December 2016

Abstract

This thesis is made up of three empirical essays on obesity and chronic disease risk in the United States. Specifically, it examines health effects of *trans* fat reduction policies, the relationship between commuting and body mass index (BMI), and commuting and physical activity tradeoffs among individuals with differing levels of BMI.

Chapter 2 examines the effects on individual-level measures of health from *trans* fat reduction policies in commercially prepared foods. Difference-in-difference estimation is used to identify changes in blood cholesterol levels prior to and after implementation of a series of related policies among restaurant meal consumers and non-consumers. Individuals with higher levels of consumption were found to have healthier cholesterol levels following implementation of the policies, while non-consumers saw less marked declines in cholesterol. Results remained robust to testing other obesity-related health measures less affected by *trans* fat consumption.

Chapter 3 examines the relationship between active and sedentary commuting and BMI. Contrary to recent literature, this work finds little evidence of a relationship between increased sedentary commuting and higher BMI; instead this work suggests the link between commuting and poor health in the literature may be explained by a strong relationship between active commuting and lower BMI. These findings suggest further work is needed disentangle active and sedentary commuting choices.

Chapter 4 revisits the relationship between sedentary commuting and physical activity to examine two issues. First, to find evidence of causality in this relationship and second, to examine whether this relationship varies by heterogeneity in health status. This work uses a two-part model of time use to examine both physical activity participation and duration decisions. Among healthy-weight and overweight individuals, commuting is associated with a decrease in participation, but not with duration of physical activities. However, among the obese, no significant relationship is observed. Results for non-obese males are robust to an instrumental variables approach (IV), however the instrument is not predictive of commuting behavior among obese males or among women, suggesting that determinants of commuting may be different for groups with differing health status.

For my parents.

Contents

Abstract ii

Dedication iii

Contents iv

List of Tables vi

List of Figures viii

Acknowledgements 1

Declaration of Authorship 2

Chapter 1 3

1.1. Introduction 3

1.2. Economic Framework 5

Chapter 2 10

2.1. Introduction 10

2.2. Background 11

2.2.1 *trans* fat in the food supply 11

2.2.2 *Trans* Fat Reduction Policies 17

2.3. Data 21

2.4. Empirical Strategy 30

2.5. Results 32

2.5.1 Robustness 36

2.5.2. Heterogeneity in Treatment Effect 48

2.6. Discussion and Conclusions 51

Chapter 3 53

3.1. Introduction 53

3.2. Literature Review 54

3.3. Data and Summary Statistics 61

3.4. Methods 69

3.5. Results 71
3.5.1 Heterogeneity and Robustness of Results 73
3.5.2. Discussion of Yang and French Paper 79
3.6. Discussion and Conclusions 85
Chapter 4 87
4.1. Introduction 87
4.2. Literature Review 88
4.3. Data and Sample Selection 94
4.4. Methods 105
4.5. Results 110
4.6. Discussion and Conclusions 120
Chapter 5 123
5.1 Discussion and Conclusions 123
5.1.1. Summary of Findings 123
5.1.2. Policy Implications 124
5.1.3. Limitations and Future Work 125
References 129
Appendix A 140
Appendix B 143
Appendix C 148

List of Tables

Table 2.1: *Trans* Fat Reduction among Top Twenty Quick Service and Fast Casual Restaurants (US) 19
Table 2.2: LDL Cholesterol Categories 23
Table 2.3: Summary Statistics of Socio-Economic and Health Behavior Covariates 27
Table 2. 4: Summary Statistics of Health Outcome Variables 28
Table 2.5: The Influence of Restaurant Meal Consumption on LDL-Cholesterol among US Adults 33
Table 2.6: Estimated impact of *trans* fat regulation on LDL levels, considering variations across time periods and among different restaurant meal consumption frequencies 41
Table 2.7: Estimated impact of *trans* fat regulation on LDL levels, considering variations in weekly restaurant meal consumption 42
Table 2.8: Alternate specifications of Meals Predictor Variable 44
Table 2.9: Relationship between restaurant meals and other FAFH meal consumption 46
Table 2.10: Influence of Restaurant Meal Consumption on Other Health Indicators among US Adults ... 47
Table 2.11: Influence of Restaurant Meal Consumption on LDL among High-Frequency FAFH Subgroups 49
Table 2.12: Influence of Restaurant Meal Consumption on LDL among High-Percentage Fast Food Subgroups 50

Table 3.1: Sample Characteristics by BMI Category 66
Table 3.2: Average Travel and Exercise in Minutes, by BMI Category 68
Table 3.3: The Relationship between Sedentary Commuting, Active Commuting, and BMI 71
Table 3.4: Relationship of Active and Sedentary Commuting and BMI, for Males and Females 74
Table 3.5: Relationship between Modes of Active Commute, Other Physical Activity, and BMI 76
Table 3.6: Variations in measuring Time Spent in Active and Sedentary Commuting 78
Table 3.7: Replication of Yang and French Results with Variations in Sample Selection 81
Table 3.8: Discussion of Results from Yang and French (2013) 84

Table 4.1: Summary Statistics for Male Respondents by Commute Time Category 96
Table 4.2: Summary Statistics for Female Respondents by Commute Time Category 97
Table 4.3: Summary Statistics by BMI Status 99
Table 4.4: Sedentary Commuting Indicator and Physical Activity Participation for Males 111
Table 4.5: Sedentary Commuting Indicator and Physical Activity Participation for Females 112
Table 4.6: Relationship between Sedentary Commuting Participation and Physical Activity Participation by BMI Group 114
Table 4.7: Relationship between Sedentary Commuting Time Spent and Physical Activity Participation by BMI Group 115
Table 4.8: Time Spent in Sedentary Commuting and Time Spent in Physical Activity by BMI Group 116
Table 4.9: Relationship between Sedentary Commuting Participation and Physical Activity Participation for Males, by BMI Category; IV Approach 118

Table A1: Changes in wording of question associated with *Meals* variable... 140
Table A2: Average Number of Meals Out over Time 141

Table B1: Demographic and Socioeconomic Characteristics by BMI in Yang and French Sample 143
Table B2: Average Travel and Exercise in Minutes, by BMI Category in Yang and French Sample 145
Table B3: Comparison of Samples using Yang and French estimation for All Travel............................... 146
Table B4: Comparison of Models for All Travel 147

Table C1: Estimate of 1970 House Price Instrument on Sedentary Commuting Indicator, Probit........... 148

List of Figures

Figure 2.1: Top 25 Food Categories Contributing to Dietary TFA Intake 16
Figure 2.2: Distribution of LDL by Restaurant Meal Consumption, 1999-2010 29
Figure 2.3: Age-Adjusted Trends in Prevalence of High LDL Cholesterol Levels, Use of Cholesterol-Lowering Medications, and Low Saturated-Fat Intake among US Adults Age 40-74. 30
Figure 2.4: Mean LDL by Restaurant Meal Consumption, 1999-2009 37
Figure 2.5: Mean LDL by Restaurant Meal Consumption Among Individuals Not Prescribed Cholesterol-lowering Medications, 1999-2009 38
Figure 2.6: Estimated Yearly Change in LDL as Restaurant Meal Frequency Group Changes 39

Figure 3.1: Unweighted Distribution of Body Mass Index 63
Figure 3.2: Unweighted Proportion of Individuals with BMI ≥ 30 by State 63
Figure 3.3: Unweighted mean sedentary commute time by State.. 64
Figure 3.4: Unweighted Proportion of Active Commuters by State 65

Figure 4.1: Unweighted Distribution of Physical Activity Time among PA Participants 101
Figure 4.2: Unweighted Proportion of Physical Activity Participants by State 102
Figure 4.3: Unweighted Mean Physical Activity Time Among PA Participants, by State......................... 103
Figure 4.4: Distribution of Time Spent Commuting among Commuters 104
Figure 4.5: Unweighted Mean Sedentary Commute Time (in Minutes) by State 105

Figure A1: Density Plot of Meals Out per Week 141

Acknowledgements

This project would not have been possible without the support of many people. First, I would like to express my gratitude to my supervisors, Prof. Colin Green and Prof. Bruce Hollingsworth. Both have read numerous revisions and have offered their support and candid advice from day one, well before there were drafts to revise. I'd also like to thank my viva examiners, Prof. Jennifer Roberts and Prof. Maria Navarro Paniagua, for their insightful comments and encouragement. I would also like to thank the Department of Economics at Lancaster University for the Departmental Scholarship which provided me with the financial means to complete this project, and the departmental faculty and staff for the friendly and collegial atmosphere. I'd like to thank Prof. Ian Walker for his unofficial support and guidance – from his first-year research skills course to writing my letters of recommendation.

I thank my fellow classmates from Economics and Health Economics and my officemates from Accounting for their friendship and support. There are a great many individuals whom I cannot list individually; faculty, fellow classmates, and friends, who have made the process of completing my PhD not only possible, but also memorable.

Last but not the least; I would like to thank my parents, my close friends, and family members who have supported me throughout writing this dissertation and throughout my life in general. Especially my Mother and Mother-in-law, both of whom have encouraged me and are very dearly missed. I'd like to thank my husband as well, for his support as an academic colleague, a friend, and partner. And of course, I'd like to thank my cat, Kishmish, for waiting patiently for me to return home.

Declaration of Authorship

I hereby declare that this dissertation and the work presented in it is entirely my own and I have clearly documented all sources and materials used.

This work has not been submitted in any form for the award of a degree at this university or any other institution, nor has this work been published.

Ayesha Ali

December 2016

Chapter 1

1.1. Introduction

Obesity rates have been on the rise in the United States since the 1960's, when roughly 13% of adults over the age of twenty were clinically obese. Recent data shows that 34.9% or 78.6 million US adults are currently considered to be clinically obese (Burkhauser et al., 2009; Flegal et al., 2010; Ogden and Carroll, 2010; Ogden CL et al., 2014). Increases in obesity have been widespread among the US adult population, with increases occurring in both sexes, all ages, all races, all education levels, and all smoking levels (Mokdad et al., 2003). The World Health Organization (WHO) and the US Centers for Disease Control and Prevention (CDC) have defined obesity as having a body mass index (BMI) greater than or equal to 30, in adults (USDHHS, 2010; WHO, 2000). Obesity is commonly used as an indicator of increased risk of chronic disease. Major diseases strongly associated with obesity are type II diabetes, high blood pressure, high cholesterol, stroke, hypertension, myocardial infarction, cancer, osteoarthritis, asthma, and depression; however this list is not exhaustive (Must et al., 1999; Wang et al., 2011). Estimated annual medical costs of obesity in the US were $147 billion in 2008 dollars and annual medical costs for obese individuals were estimated to be $1,429 higher than for normal weight individuals (Finkelstein et al., 2009).

Aside from genetic factors, lifestyle choices and health behaviors are key determinants of obesity and chronic disease risk (Mokdad et al., 2003). Specifically, sedentary behavior, physical activity, diet, and nutrition, smoking, and alcohol consumption play a role in determining obesity and chronic disease risk. Because of the high costs and increased health risk and because obesity is largely preventable through modification of behavior and lifestyle, obesity prevention and reduction has been the focus of a large body of research and numerous public health policies put into place by governments, school districts, institutions, and companies (Cawley, 2015).

This dissertation contributes to the obesity and chronic disease literature by applying economic theory and methods to understand two of these behaviors: nutrition and physical activity. Specifically, Chapter 2 contributes to the literature on *trans* fat reduction policies and other types of obesity- and lifestyle-related public health policies and regulations such as sugar-sweetened beverage bans, salt reduction regulations, by providing support that these policies are in fact effective in improving health among consumers of otherwise unhealthy foods. Additionally, this work uses clinically measured intermediate health outcomes which are directly affected by *trans* fat consumption; this is less common in health economics research, where disease diagnosis, mortality, or self-reported health status are more common outcome variables. Chapters 3 and 4 contribute to the literature on commuting and health. Findings from Chapter 3 suggest that care is needed in identifying commuting behavior; specifically, active modes of commuting and sedentary modes of commuting should be treated separately when considering their effects on health and health behaviors. Public health policies and urban planning policies based on research which does not account for different types of commuting behaviors may be misguided. Chapter 4 contributes to the literature by re-examining the relationship between sedentary commuting and physical activity using a two-part model of time-use and by examining heterogeneity in this relationship by health status. This work finds some evidence of causality among non-obese males, however the key contribution of this work provides evidence that policies targeted at particular at-risk groups should consider research which focuses on these groups specifically.

The remainder of this chapter provides an introduction of subsequent chapters and presentation of an economic framework within which those chapters fit.

Chapter 2 uses measured health indicators from the 1999 through 2010 National Health and Nutrition Examination Survey (NHANES) dataset, which provides 2-year waves of repeated cross-sectional data. Using a difference-in-difference methodology this work measures the health effects of the implementation of a series of closely-related public health policies aimed at reducing *trans* fat consumption in commercially prepared foods. *Trans* fat consumption has been shown to directly affect low-density lipoprotein (LDL) cholesterol levels, so this is the outcome of interest (Mozaffarian et al., 2004). A continuous treatment group of restaurant-meal consumers is identified and grouped into categories based on the frequencies; these individuals are the most likely to be affected by the *trans* fat reduction policies. Individuals who consume fewer than one restaurant meal per week are identified as the baseline. Controlling for a variety of demographic, socioeconomic, and other health behaviors, this analysis finds restaurant-meal consumers had healthier changes cholesterol levels following implementation of the policies, while non-consumers saw less marked declines in cholesterol measures. Specifically, the treated group saw reductions in LDL levels of nearly 3 mg/dL lower than the lesser-exposed groups. Other obesity-related health measures less affected by *trans* fat consumption saw little or no change when the policy was enacted, giving evidence that the *trans* fat policy was driving the change rather than other health-related changes in the population.

Chapter 3 uses time-use diary data from the 2006 through 2008 waves of the American Time Use Survey (ATUS) and uses a linear regression model to examine the relationship between active and sedentary commuting behaviors and BMI among US adults. In the fields of health economics and urban economics, many papers have examined this relationship using different datasets and specifications for commuting and health outcomes; much of this literature suggests sedentary commuting and urban sprawl are associated with obesity (Eid et al., 2008; Ewing et al., 2014; Frank et al., 2004; Hansson et al., 2011; Hoehner et al., 2012; Lopez, 2004; McCormack and Virk, 2014; Zhao and Kaestner, 2010). While very few papers have found contradictory results (Jilcott et al., 2011; Kelly-Schwartz et al., 2004) this chapter, finds little evidence of a relationship between increased sedentary commuting time and higher BMI. After examining methods used in another recently-published work (Yang and French, 2013), this chapter finds that the link between sedentary commuting and obesity described in the literature could possibly be explained instead by the tradeoff between sedentary commuting and active commuting, and by the strong relationship between active commuting and lower BMI. These findings suggest further work is needed in disentangling the behaviors of active and sedentary commuting choices.

Chapter 4 also uses time-use diary data from the 2006 through 2008 ATUS dataset and revisits the relationship between sedentary commuting and physical activity behaviors to examine three issues. First, this work uses a two-part model of time use to separately examine physical activity participation and duration decisions. This type of model has previously been used with self-reported survey data on commuting and physical activity; however this work is the first to explore that relationship by applying this method to time-use data. Second, this work attempts to find evidence of causality in this relationship, by using an instrumental variables approach where commuting behavior is instrumented by historical housing prices. And third, this work examines whether this relationship varies by heterogeneity in health status. This work finds that among healthy-weight and overweight individuals, commuting is associated with a decrease in participation, but not with duration of physical activities. However, among the obese, no significant relationship is observed for either decision. Results for non-obese males are robust to the instrumental variables approach (IV); however the instrument is not predictive of commuting behavior among obese males or among women, supporting previous evidence that determinants of commuting behavior differ by sex and also suggesting that determinants of commuting may be different for groups with differing health status.

1.2. Economic Framework

Cawley (2004) presents a rational choice framework for understanding and addressing why individuals may choose behaviors which lead to obesity and poor health. This framework is rooted in the economics question of understanding how individuals allocate their scarce resources of time and money in order to maximize their lifetime utility or happiness. Health, itself, can contribute to utility, however individuals may choose to make tradeoffs between health and other things with give them utility.

Individuals face financial constraints which can affect their choice of foods. In an industrialized food system like the US, energy-dense foods, which contain larger amounts of fats and sugars, are relatively cheap while foods like fresh fruits and vegetables, which are less energy dense, are more expensive (Drewnowski, 2004; Drewnowski and Darmon, 2005). Considering a two-good model where individuals can purchase food and all other goods, individual consumers face the trade-off of buying cheap high-fat food and more of all other goods or buying expensive healthier foods and less of all other goods. Related specifically to Chapter 2, where the focus is on *trans* fat regulation, the molecular structure of *trans* fatty acids allows certain foods which are made with *trans* fats to be produced more cheaply than if they were produced with butter or other types of oils. This means that manufacturers and sellers of foods are also making tradeoffs between using *trans* fat which can lower food production and storage costs and increase profit margins or use another healthier type of fat which often means increased production and storage costs and lowered profit margins. Similarly, but not within the scope of this work, time constraints also play a role in an individual's food choices; packaged and processed foods, restaurant and carry-out meals can save time over cooking meals from scratch, allowing individuals with scarce amounts of time to have time for other activities such as working, commuting, childcare, and leisure activities.

All individuals face the time constraint of having only 24 hours in a day and must decide how to allocate time between working for pay, working in the household, physically-active commuting, sedentary commuting, doing physically active leisure, and doing sedentary forms of leisure. Choosing to spend time in one activity reduces time available for other activities. Chapters 3 and 4 use this framework to examine the role that time constraints play in an individual's involvement in physical activity outside of paid work. These chapters specifically examine how variations in the time constraint of working adults may arise through time spent in commuting to and from work and how this variation may affect an individual's choice to engage in physical activity. While not the focus of this work, individuals also face financial constraints which can affect the choice of their participation in physical activity behaviors.

Cawley's framework incorporates elements from the model of labor and leisure choice presented by Becker (1965) which assumes that individuals derive utility from the consumption of "basic commodities" such as meals and leisure activities and that they are able to produce these basic commodities through time and market commodities. This model also builds on Grossman's (1972) model of health production where individuals are both producers and consumers of health, which itself is a sort of capital, or a stock which can decline over time in the absence of investments in health. In Grossman's model, health is both a consumption good, which yields direct utility, and also an investment good, which yields utility to consumers indirectly through increased market productivity, fewer sick days, and higher wages. People can combine time (such as to get exercise or cook meals) with market goods they purchase (such as food, a gym membership, healthcare, tobacco, or alcohol) in order to improve, maintain, or damage their health (Grossman, 1972). When making these decisions of whether or not to eat a high-fat fast food meals, be sedentary, use drugs and alcohol, or engage in other risky behaviors, individuals weigh the utility received from enjoyment these behaviors against the losses in health and welfare caused by these behaviors. Similarly, Cawley's framework, also known as the SLOTH model, attempts to understand how individuals allocate their resources of time and money in order to maximize their lifetime utility, subject to three constraints: time, budget, and biology. In this framework, individuals to maximize the following utility function:

$$(1.1) \qquad U\,(S, L, O, T, H, F, W(S, L, O, T, H, F), H(S, L, O, T, H, F, W), Y).$$

Here, utility is a function of *SLOTH*, where *S*, *L*, *O*, *T*, and *H* are vectors of variables that represent the number of hours spent in sleeping, *S*, leisure, *L*, occupation or paid work, *O*, transportation, *T*, and home production or unpaid work, *H*, respectively. Each of these pursuits directly affects a person's utility, however they can also affect utility indirectly, by affecting one's weight, *W*, and one's health, *H*. An example would be that sedentary forms of leisure, transportation, and other discretionary activities, may lead to higher weight and worse health in the future for an individual.

The exact relationship between utility and weight varies widely between individuals, but generally is nonlinear. Individuals are not able to choose their weight directly but they can affect it through physical activity and caloric intake. *F* represents caloric intake, or food, and can have a direct impact on utility through taste and an indirect impact on utility through weight and health. Health, is affected by weight, by food intake and by the allocation of time or physical activity across SLOTH. Physical activity can affect health directly or indirectly through weight.

Finally, *Y* represents a composite or bundle of all goods other than food.

In this framework, individuals face three constraints:

$$(1.2) \qquad Y + FP_F = w * O,$$

$$(1.3) \qquad S + L + O + T + H = 24,$$

$$(1.4) \qquad \Delta W = c(F) - f(S, L, O, T, H, G) - \delta(G)W.$$

The first of these is a budget constraint: money spent on food and all other goods must add up to one's wage earnings. The second constraint simply states that hours engaged in each of the components of SLOTH must add up to exactly 24 for each day. The third is a biological constraint: changes in weight are determined by caloric intake through food, *c(F)*, caloric expenditure through various activities, *f(S, L, O, T, H, G)*, and through the metabolic rate, δ, which is a function of one's genes, *G*, as is the amount of energy expended. Making suitable assumptions about constraints 1.2 and 1.4, which are implied in Cawley (2004), we arrive at the first order condition that in optimality, the marginal utility of any one activity is equal to the marginal utility of other activities.

Within this framework, individuals may rationally decide to accept a higher body weight in exchange for the utility associated with eating or with sedentary forms of leisure. People will exercise when it is the best use of their scarce time (even though public health advocates may encourage them to do it as long as it increases health). People will consume foods within their budgets that provide the highest net benefit. Gross benefits include immediate pleasure of taste plus any current and future health benefit. Gross costs include financial cost, discounted utility of adverse health impacts, and discounted utility of any future weight gain.

In a multi-period version of this framework, individual's decisions about eating and time allocation in each period will reflect both the immediate and the future marginal costs and benefits. Individuals generally assign less importance to outcomes in the distant future than to those in the present. Cawley points out that in a dynamic framework, the length of life could be made a function of health and weight.

To summarize, this framework assumes that individuals divide their monetary spending between a variety of foods and all other goods, and assumes that individuals divide their time between sleep, leisure, unpaid work in the household, paid work in the labor market, and transportation, in order to maximize their utility. While people do not choose their weight directly, their choices of time allocation and consumption affect weight indirectly. This framework does not examine how individual utility may vary with health, weight, food consumption or other activities, but does provide a means of understanding the trade-offs that individuals face.

In terms of using this framework to understand the problems of obesity, it is useful to consider the marginal benefit of participating in any particular activity. For example, if the marginal utility provided by an hour of a sedentary leisure activity, like playing video games, increased and the marginal utility from all other activities remained the same, then the framework predicts that an individual would re-allocate time to play more video games and do less of other activities.

Chapter 2 examines the effects of public health policies of *trans* fat reduction on cholesterol levels among consumers of restaurant meals, focusing on the biological constraint in this model and how public health policy may affect it. Applying this framework to understand how these policies may have affected individual's food choices raises some concerns about the effectiveness of such policies. In particular, if reduction of *trans* fat improves the healthiness of fast food, restaurant meals, and packaged foods, then individuals who derive utility from consuming healthy foods may find that the marginal utility of these foods has increased relative to the marginal utility of other foods, and thus increase consumption of processed foods formerly containing large amounts of *trans* fat and decrease consumption of other healthier and possibly more expensive foods. This chapter addresses these concerns through a unique robustness check examining whether other health measures which are affected by consuming unhealthy foods but not particularly through *trans* fat have changed as well.

Chapters 3 and 4 both rely heavily upon this framework to understand how individual time use relates to health and focus on the second constraint in this model. Chapter 3 focuses on how time spent involved in commuting may be related to BMI, as a measure of health. In this chapter, a sample of individuals has a time constraint of 24 hours, in which at least seven hours are spent working and some number of minutes may be spent in commuting, or getting to and from work. In this framework, the amount of time spent commuting reduces the overall amount of time available for other activities, including physical activity. This chapter explores how individual-level variation in the time constraint due to variation in commuting relates to health. Similarly, Chapter 4 focuses on how variations in an individual's time constraint may affect their participation in physical activity on a given day, and how these relationships may vary across individuals of differing health status.

Chapter 2

2.1. Introduction

Industrially produced *trans* fats are typically found in fast foods, restaurant and bakery foods, and pre-packaged foods. Their consumption has been strongly linked to an increased cardio-metabolic risk, yet Americans have been increasing their consumption of both restaurant meals and pre-packaged foods (Fryar and Ervin, 2013). The World Health Organization, the Institute of Medicine, The American Heart Association, the US Department of Agriculture, and other leading health organizations recommend minimizing the amount of *trans* fat in the diet (Mozaffarian et al., 2006; Teegala et al., 2009). Two key policies have given food manufacturers incentives to reduce or remove artificial *trans* fats; the FDA required *trans* fats to be listed on the mandatory nutrition label of all processed foods beginning in 2005, and the New York City Department of Health and Mental Hygiene passed a regulation banning the use of artificial *trans* fat in all restaurant foods beginning in 2007. During the time that these policies were being implemented, a series of notable class-action lawsuits against food producers Kraft, McDonalds, and KFC in 2003, ’05, and ’06 provided incentives for food producers to remove or reduce *trans* fat fats (Tarrago-Trani et al., 2006; Unnevehr and Jagmanaite, 2008).

This chapter investigates whether *trans* fat reductions in food prepared away from home (FAFH) due to these policies has had any effect on intermediate health measures. It examines whether there is a causal effect of decreased *trans* fat consumption in restaurant meals on cardio-metabolic risk factors, specifically blood cholesterol levels. This study is among the first to isolate the specific effect of *trans* fat regulation on cholesterol levels and cardiovascular health at the population level. The work helps to fill a gap in the literature pointed out by Angell et al. (2012) and these findings lend support to recent work which has found evidence of reduced mortality from cardiovascular events following the NYC *trans* fat ban (Restrepo and Rieger, 2016).

The main outcome variable in this work is serum low density lipoprotein (LDL) cholesterol; this is in line with causal pathways indicated by the medical literature – *trans* fat consumption directly affects LDL levels and causes greater changes in LDL levels, relative to high-density lipoprotein (HDL) and triglyceride levels (Mozaffarian et al., 2006). High levels of LDL are the main source of cholesterol buildup and blockage in the arteries, leading to cause of heart disease (National Heart, Lung, and Blood Institute (NHLBI), 2005). Using population-level, repeated cross-sectional data, this work is able to control for demographic and socioeconomic factors as well as health behaviors. Using information about the frequency of eating out, this work identifies the magnitude of the change in health indicators due to *trans* fat reduction in restaurant meals.

This chapter proceeds as follows: Section 2.2 provides background information on what *trans* fats are, and why they have become a part of the food supply. This section presents literature on the effects of *trans* fats on health outcomes, on *trans* fat consumption levels prior to the policy, and on the effects of policies related to *trans* fat reduction and removal. The data and changes in restaurant foods created by *trans* fat reduction policies are discussed in section 2.3. Section 2.4, explains the methodology and Section 2.5 presents key results and tests for robustness and treatment heterogeneity. This includes examining the effects on other health indicators: HDL cholesterol, total cholesterol, and triglyceride levels, BMI, and waist circumference. These indicators have seen consistent changes over the past decade, due in part to many changes in food consumption patterns and health behaviors, and serve as placebo tests to show the robustness of the main result. Section 2.5 also discusses implications and limitations of the main findings. Finally, Section 2.6 closes with a summary of the main findings and suggestions for further research.

2.2. Background

2.2.1 *trans* fat in the food supply

Trans fat is a shorthand term for *trans* fatty acid*s*, which are named for the *trans* configuration of the double bond in the fat molecule. In contrast, naturally occurring unsaturated fats have a *cis* configuration in the double bond and saturated fats have no double bond at all.

There are two different types of dietary *trans* fatty acids, one which occurs naturally and another which is industrially produced. Naturally occurring *trans* fats, or ruminant produced *trans* fatty acids (RP-TFA), are found in all meat and dairy products from ruminant animals and make up less than 0.5% of total energy intake in the US (Micha and Mozaffarian, 2008). RP-TFA are made up of vaccenic acid and conjugated linoleic acid and have a *trans* fatty acid concentration of about 6%, whereas industrially produced *trans* fats are made up of elaidic acid and can have a concentration of *trans* fatty acids as high as 60% (Stender et al., 2008). The structure of the *trans* fatty acid molecules in ruminant-produced fat is significantly different from industrially-produced *trans* fatty acids, and as a result they have different effects on health (Chardigny et al., 2008; Remig et al., 2010). Specifically, moderate consumption of RP-TFA is found to have no adverse effect on coronary heart disease (CHD) risk and some research has found possible beneficial effects on lipid levels and CHD risk (Bassett et al., 2010; Jakobsen et al., 2008; Stender et al., 2008).

The focus of this chapter then, is strictly on the industrially-produced *trans* fatty acids, which are simply referred to here as either *trans* fat, IP-TFA, or TFA. Industrially produced fats make up the major source of dietary *trans* fats in the US and contribute to roughly 2-3% of total energy intake in the US (Micha and Mozaffarian, 2008). IP-TFA is created through the industrial process of partial hydrogenation of unsaturated vegetable oils. This process uses a metal catalyst, usually nickel, and high temperatures or pressure to hydrogenate, or saturate, the carbon-carbon double bonds which exist in vegetable oils. If hydrogenation is completed, as is the case in fully hydrogenated oils, the resulting product is a saturated fat and does not have the *trans*-double-bond. When the hydrogenated process is not fully completed, as is the case in partially hydrogenated vegetable oils (PHVO), the resulting product is a mixture of *cis*- and *trans*- fatty acids (Mozaffarian et al., 2006). The percentage of *trans* isomers can vary depending on the level to which vegetable oils have been hydrogenated; for example, tub margarine typically has a lower proportion of *trans* fat than stick margarine (Ascherio et al., 1994). The hydrogenation process was developed in the 1890's and used primarily for industrial processes; it was patented and first applied to food production in 1903. The structural differences resulting from partial hydrogenation result in a number of features which make *trans* fats more amenable to industrial food production. *Trans* fats have a more rigid configuration of molecules which results in a melting point somewhere between the melting points of saturated fats and *cis* unsaturated fats. This means that they are solid at room temperature, while unsaturated fats are liquid. This provides desirable characteristics such as texture and mouthfeel. The molecular structure of *trans* fats also makes them more stable when exposed to oxygen, which means that they are less prone to oxidation and becoming rancid than *cis* unsaturated fats. When used to prepare foods, *trans* fats extend product shelf life, increase the lifespan of frying oils, and decrease refrigeration requirements. *Trans* fats are often cheaper to produce than saturated and unsaturated fats because they are often produced using byproducts of soybean meal production and other vegetable oil production processes (Kodali, 2014). In the US, they are typically found in margarines, shortenings, spreads and are also used in bakery items, deep fried foods, frozen foods, and pre-packaged snacks (American Heart Association, 2010; Stender et al., 2007, 2006). When *trans* fats had initially entered the food supply, they were treated as "generally recognized as safe" (GRAS) and were not initially evaluated for health effects in humans (Ascherio and Willett, 1997). Because of their commercial advantages over non-hydrogenated vegetable oils, *trans* fat usage has increased steadily from their development in the early 1900's until the late 1960's. Usage of *trans* fats became more pronounced during the Great

Depression and World War II period primarily because margarine acted as a cheaper substitute during times of butter rationing. Another reason for their increased usage is that medical research from the 1950’s and ‘60’s associated high levels of blood cholesterol and saturated fat intake with increased mortality from heart attacks and medical professionals recommended that Americans decrease their consumption of saturated fats. Following this advice, margarines made from partially hydrogenated vegetable oils were generally regarded as a healthier alternative to saturated fats like butter (American Heart Association, 2010). This increased usage of *trans* fats has contributed to the modern obesogenic food environment and the built environment by contributing to the lowered cost and thus increased the accessibility and availability of pre-packaged and fast foods. These in turn can contribute to overconsumption of food and increased sedentary activity. (Kirk et al., 2010; Lee et al., 2011; Swinburn et al., 2011).

Medical research on *trans* fats has been extensive; early studies on the negative health effects of *trans* fat intake began appearing in the 1970’s however scientific consensus was not reached. By the 1990’s medical research established links between *trans* fatty acid intake and poor heart and circulatory health (Ascherio and Willett, 1997). Early research on dietary *trans* fat from the landmark Nurses’ Health Study followed a subsample of 80,095 female registered nurses from 1980 onwards to assess risk factors for cancer and cardiovascular disease. *Trans* fat intake was calculated from dietary questionnaires from women with no diagnosis of diabetes, stroke, CHD, or hypercholesterolemia at baseline. These subjects were followed for eight years, during which there were 431 new cases of CHD. Controlling for demographic, health, and dietary characteristics, researchers found that intake of *trans* fats were directly associated with risk of CHD, more so for those women whose margarine consumption over the previous ten years had been stable(Willett et al., 1993). Similar results were observed in the Boston Health Study, which uses a case-control design and studied the relationship between *trans* fatty acid intake and first myocardial infarction in a sample of 239 patients admitted to the intensive care units of one of six hospitals in the Boston area between 1982 and 1983, as well as 282 population control subjects. *Trans* fat consumption was estimated using a food frequency questionnaire, where *trans* fat intake estimates were previously assessed against proportion of *trans* fatty acids found in aspirates of adipose tissue of respondents. Again, after adjusting for age, sex, energy intake, CHD risk factors, and diet, researchers found *trans* fat consumption, particularly in the form of margarine, to be directly and significantly related to risk of myocardial infarction (Ascherio et al., 1994).

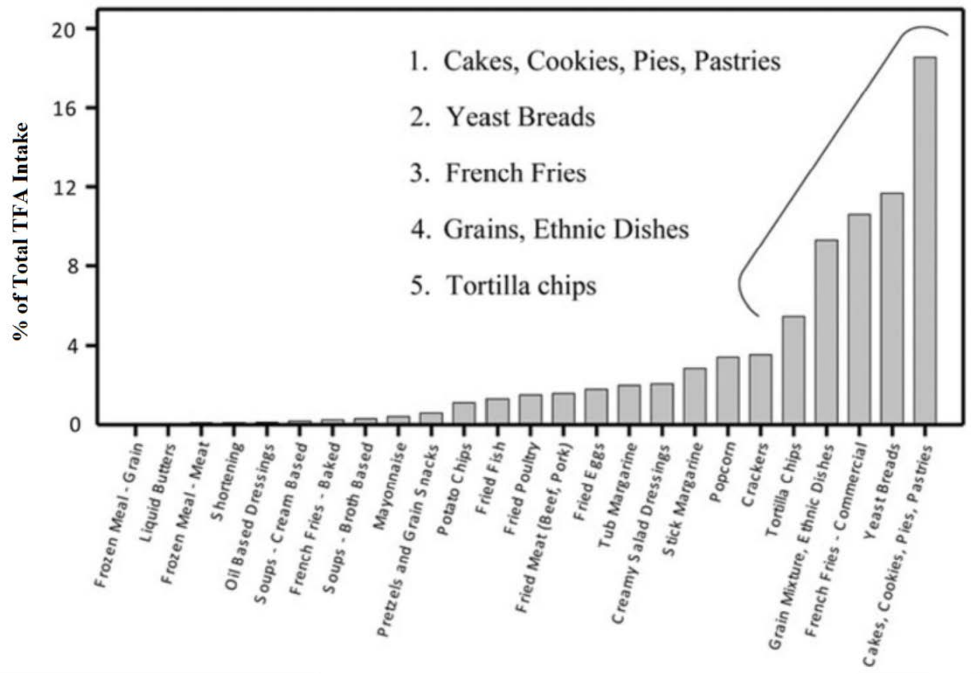
Following early research linking *trans* fat consumption with CHD risk, later research identified numerous causal pathways between consumption and increased CHD risk. Compared to consumption of an equal number of calories from saturated fat or *cis* unsaturated fat, *trans* fatty acid (TFA) consumption was found to raise LDL cholesterol, reduce HDL cholesterol, increase blood levels of triglycerides, and increase systemic markers of inflammation, each of which contribute to an increase in coronary heart disease (CHD) risk (Mozaffarian et al., 2006). Another study estimated *trans* fat intake of about 5 grams per day is associated with a 25% increase in the risk of ischemic heart disease (Oomen et al., 2001). Even at very low levels of consumption, dietary intake of artificial *trans* fatty acids increases the risk of coronary heart disease, more so than consumption of any other macronutrient on a per calorie basis (Ascherio and Willett, 1997; Mozaffarian et al., 2006). The relationship between TFA consumption and incidence of CHD in prospective observational studies is greater than the predicted changes in CHD risk from the changes in blood cholesterol levels alone, which the literature suggests is caused by TFA consumption influencing other CHD risk factors as well (Ascherio et al., 1994). For example, increased *trans* fat consumption has also been associated with increased levels of inflammatory biomarkers and endothelial cell dysfunction (Baer et al., 2004; Lopez-Garcia et al., 2005; Mozaffarian et al., 2006, 2004). *Trans* fat consumption has also been found to be associated with an increased risk of type-II diabetes. A meta-analysis of four prospective cohort studies finds that a 2% increase in total energy intake from TFA, roughly 4.5 grams, is associated with a 23% increase in the incidence of CHD. Further, the effect of replacing *trans* fat varies depending on the type of fatty acid or other macronutrient that it is replaced with – for example exchanging 2.2 grams (1% of total energy intake) of TFA with an equal amount saturated fat results in a reduction of LDL of 0.4 mg/dL and raises HDL by 0.5 mg/dL whereas making the same exchange with 2.2 grams of a *cis* polyunsaturated fat results in a reduction of LDL of 2.3 mg/dL and a rise in HDL of 0.3 mg/dL (Mozaffarian et al., 2006). So, the true effect of a reduction in *trans* fat consumption on CHD outcomes should then be somewhere between the minimum estimate based on the effect of TFA consumption on LDL and the maximum estimate that seen in observational studies (Katan, 2006). These findings suggest that the use of *trans* fats as a healthier alternative to saturated fats is not justified.

Also relevant to the research question presented here, the medical literature addresses how long dietary changes involving *trans* fat intake can take to affect health outcomes. Findings from studies where subjects were given different diets over three to four weeks suggest that effects of *trans* fat consumption on cholesterol levels can be seen relatively quickly. In one study, subjects were fed diets for three weeks at a time, in which the nutrient content was similar save for ten percent of calories which came from three different sources of fat: oleic acid (unsaturated fat), *trans* isomers of oleic acid (*trans* fat), and saturated fatty acids. Effects on LDL levels were identified within the three-week period that subjects were on a particular diet (Mensink and Katan, 1990). In another study, subjects were fed six different diets for periods of 35 days each; effects on cholesterol levels from variations in fatty acid composition of the diets were seen within 28 days from the start of each diet (Lichtenstein et al., 1999).

Another aspect of *trans* fat consumption relevant to this research is to understand how Americans are consuming *trans* fats. The Dietary Guideline for Americans recommends keeping total daily fat consumption under 70 grams, saturated fat consumption under 20 grams, and *trans* fat as low as possible (US Department of Agriculture, 2010). Using data from food intake surveys and a 1995 USDA database of fatty acid content of selected foods, researchers estimated mean TFA consumption from 1989-1991 was approximately 5.3 grams per day (Allison et al., 1999). Between 1999 and 2002, using the same USDA database, the average American consumed 2.5% of total calories per day from *trans* fats – this is roughly 5.5 grams of *trans* fats per day, based upon the average recommended 2,000 calorie diet (Kris-Etherton et al., 2012). Considering that average daily caloric intake for US adults was estimated to be 2,200 calories in 2003, 2.5% of total calories from *trans* fat would be about 6.1 grams of *trans* fats per day (Ng et al., 2014). While this estimate of daily caloric intake does not appear to be particularly excessive, it is worth noting that survey respondents, particularly those who are overweight or obese, often underreport dietary intake; this would suggest that this calculation of *trans* fat consumption is an underestimate (Lichtman et al., 1992; Macdiarmid and Blundell, 1998). The authors note that while consumption in the 1999-2002 period was similar to that in the 1989-1991 period, the striking difference between the two periods was that TFA intake increased among the 90th percentile. This group was made up of males, aged 12-19 and their *trans* fat consumption was estimated to be alarmingly between 11.8 and 92.4 grams per day. The authors emphasize that the range of *trans* fat consumption varies greatly; individuals in the lowest quintile of *trans* fat intake consume between 1.3 to 2.0 grams per day while individuals in the highest quintile average 9.6 to 16.5 grams per day (Kris-Etherton et al., 2012). Since the 1980's roughly 80% of caloric intake from *trans* fat has come from partially hydrogenated vegetable oils. The majority of these are found in commercial baked goods, yeast breads, and commercial French fries, as shown in Figure 2.1, which has been reproduced from the original paper. With the exception of stick margarine, shortening, mayonnaise, and liquid butters, all of these items can be considered to be within the category of "food away from home" or FAFH, which encompasses table-service restaurants, fast-food restaurants, take-out meals, delivery meals, and any other commercially prepared foods, meals and snacks (Lin and

Guthrie, 2012). Most *trans* fats are found in the categories of pre-packaged or commercially prepared foods, restaurant meals, and fast food.

Figure 2.1: Top 25 Food Categories Contributing to Dietary TFA Intake



[1]

Another other major source of *trans* fat consumption comes in the form of fast foods. While average fast food consumption in US adults has gone down slightly in recent years, certain groups are still consuming a large percentage of daily calories from fast foods; 2007-2010 NHANES data show that non-Hispanic black males in the 20-39 age group are consuming over 20% of their daily calories from fast food, obese individuals in the 20-39 age group consume 18% of daily calories from fast food, and lower income individuals in the 20-39 age group consume 16.1% of daily calories from fast food (Fryar and Ervin, 2013; Kris-Etherton et al., 2012). High TFA consumption is also associated with higher consumption of total fat, indicating that food choices and dietary patterns differ for the higher-TFA consumption groups. For example, one study points out that it is easily possible to consume 25 grams of *trans* fats in a single fast food meal consisting of a large serving of chicken nuggets with French fries and 100 grams of cookies (Stender et al., 2008).

[1] Reprinted from "*Trans* Fatty Acid Intakes and Food Sources in the U.S. Population: NHANES 1999-2002" by P.M. Kris-Etherton, M. Lefevre, R.P. Mensink, B. Petersen, J. Fleming, and B.D. Flickinger, 2012, *Lipids*, Volume 47 (10), pp 931-940. Copyright (2012) by Springer International Publishing.

2.2.2 *Trans* Fat Reduction Policies

Following on the research of Ascherio & Willet (1997) and with pressure from consumer interest groups, in 1999 the US government proposed a law requiring manufacturers to list *trans* fat amounts on food nutrition labels, which have been mandatory on all packaged foods since 1994, but did not previously include *trans* fat content. Initially, this law failed to pass, however a series of public lawsuits over *trans* fats and growing attention to *trans* fats as a public health issue created incentives for manufacturers to reduce *trans* fat content in food (Unnevehr and Jagmanaite, 2008). In 2003, the US Food and Drug Administration (FDA) passed a similar law, and manufacturers were given three years, until January 1, 2006, to comply with the *trans* fat labeling law, which allowed products that contain less than 0.5 grams of *trans* fat per serving to be labeled as "zero grams *trans* fat". In response, many large-scale food manufacturers voluntarily removed or reduced *trans* fats in their products so that front-of-packaging labeling could make the advertising claim of "zero grams *trans* fat" in time for the rollout of the law on January 1, 2006 (American Heart Association, 2010). A paper documenting the various approaches food manufacturers use to reduce *trans* fats from food products notes that by 2005, manufacturers of spreads and margarines significantly reduced or eliminated *trans* fats and that some restaurant and food service organizations had already begun using non-hydrogenated oils for frying (Hunter, 2005). Another paper documents numerous changes in product formulation in 2006 which include low *trans* fat shortenings and margarines supplied to the commercial food industry (Tarrago-Trani et al., 2006). The authors note that many of the technologies implemented to create replacements for *trans* fats were actually developed and have been available to the food industry since as early as the mid-1990's; however, prior to the mandatory labelling policy, firms had no regulatory incentive to alter their high *trans* fat formulations.

Advocates of further legislation point out that while nutrition labeling can allow some consumers to make choices to avoid *trans* fats in pre-packaged foods, roughly one third of daily calories come from restaurant foods where *trans* fat information was not generally disclosed at that time (Angell et al., 2012; Fryar and Ervin, 2013; Guthrie et al., 2002; Powell et al., 2012). The process of legislating the reduction of *trans* fats in restaurant foods began in 2005; the New York City Department of Health and Mental Hygiene added assessment of the presence of artificial *trans* fats to the routine of inspecting and regulating restaurants and retail food outlets. Inspectors surveyed a random sample of restaurants to evaluate the use of *trans* fat in oils, shortenings, and spreads used for frying, baking, or cooking and found that where it could be measured, half of restaurants used artificial *trans* fats in food preparation (Angell 2009, NYC DHMH 2008). Following these findings and subsequent intermediate policies focused on public education and voluntary removal of *trans* fats in restaurants, in 2006 New York City became the first US city to pass a regulation limiting *trans* fats in restaurants. All licensed food establishments, including restaurants, bakeries, caterers, street-food vendors, school cafeterias, and senior centers, could only serve foods with less than 0.5 grams of *trans* fat per serving – this applied only to fried foods in 2007 and was extended to baked goods and all other foods by 2008. One key element of the *trans* fat reduction policies that Angell and colleagues (2012) illustrate is the gradual nature of compliance to the regulations. Specifically, they show the percentage of food establishments randomly surveyed in New York City that use artificial *trans* fats and how this percentage increases gradually across time as more stringent policies are implemented.

Given the success of these policies in New York City and growing consumer awareness and activism, health departments in many other jurisdictions such as cities, counties, and even states in the US have passed similar laws limiting or removing *trans* fats from menu items in restaurants, bakeries and institutions. In response, more than 50 major national chain restaurants announced that they would voluntarily remove *trans* fats from their menu items nationwide, and by 2007 most had already done so (Associated Press, 2007; Dorfman et al, 2008). Many other hotel groups, food manufacturers, grocery retailers, and other firms have also announced the removal of *trans* fats. One paper in particular highlights the spillover effects onto restaurants for food industry reformulations to reduce *trans* fat content (Unnevehr and Jagmanaite, 2008). These incentives are compliance with the nutrition labelling law, high-profile product liability charges and lawsuits of major food companies such as Kraft and McDonald's, and compliance with the New York City *trans* fat ban. In response, major food service companies replaced *trans* fat-containing frying oils with *trans* fat free alternatives, makers of packaged foods reformulated products, and farm-level suppliers developed oil crops which provided substitute options for *trans* fats. Mozaffarian et al. use FDA food-composition databases and news articles to identify 1993 through 2006 as the pre-policy period, and 2008 through 2009 as the post-policy period for *trans* fat-reduction policies (2010).

As an assessment of whether *trans* fat reductions were occurring at the national level in chain restaurants, I collected news articles and press releases for the top twenty firms using the 2011 annual Quick Service Restaurant industry rankings for quick-serve and fast-casual restaurants in the US (Oches, 2011). I have put together a dataset of the top 50 quick service and fast casual chain restaurants and the dates of each restaurant's implementation of *trans* fat reduction policies. Table 2.1 shows data from the top 20 firms. I found that 17 restaurant chains of the QSR Top 20, had removed *trans* fat from menu items at the national level during the period between 2006 and 2008. A majority (74%) of these restaurants made *trans* fat reductions in 2007. To illustrate the extent of these changes at the national level, Table 2.1 shows the top twenty restaurants in the QSR rankings list, representing roughly 85% of the quick service and fast casual market share in the US, and whether or not they have removed *trans* fats from menu items at the nationwide level.

Table 2.1: *Trans* Fat Reduction among Top Twenty Quick Service and Fast Casual Restaurants (US)

QSR Rank	Company	2011 US Sales (Millions)	% of QSR 50 by Sales	Reformulated Menu Items	Year Enacted
1	McDonald’s	$34,172.00	22.86%	All fried menu items (2007); all baked goods (2008)	2007
2	Subway	$11,400.00	7.63%	All items on the "core menu"	2007
3	Starbucks	$9,750.00	6.52%	All food and beverage items	2007
4	Wendy’s	$8,500.00	5.69%	Zero-*trans* fat oil	2006
5	Burger King	$8,400.00	5.62%	Zero-*trans* fat oil	2007
6	Taco Bell	$7,000.00	4.68%	All menu items	2007
7	Dunkin’ Donuts	$6,500.00	4.35%	All menu items	2007
8	Pizza Hut	$5,500.00	3.68%	Each ingredient <0.5 g per SVG	2007
9	KFC	$4,500.00	3.01%	All menu items, excl. biscuits	2007
10	Chick-fil-A	$4,051.00	2.71%	All menu items	2007
11	Sonic Drive-In	$3,692.80	2.47%	No removal of *trans* fat	N/A
12	Domino’s Pizza	$3,437.90	2.30%	Each ingredient <0.5 g per SVG	N/I
13	Panera Bread	$3,400.00	2.27%	All menu items	2006
14	Arby’s	$3,021.90	2.02%	Zero-*trans* fat oil	2007
15	Jack in the Box	$2,946.30	1.97%	Zero-*trans* fat oil	2008
16	Dairy Queen	$2,660.00	1.78%	Zero-*trans* fat oil	2008
17	Chipotle Mexican Grill	$2,270.00	1.52%	All menu items	N/I
18	Papa John’s	$2,213.60	1.48%	N/I	N/I
19	Hardee’s	$2,100.00	1.40%	Zero-*trans* fat oil	2008
20	Popeye’s Louisiana Kitchen	$1,720.00	1.15%	*Trans* fat-free biscuits only	2006

Source: QSR Magazine Ranking, 2012

There are two primary concerns about reducing *trans* fats in prepared foods. The first is that foods formerly containing *trans* fats are among those that should have limited intake, however if these foods are being advertised as “*trans* fat free,” it may give consumers the illusion of making a healthy choice, or a halo effect, similar to the halo effect of earlier “low fat” claims. In response, consumers may either increase consumption of these foods or decrease consumption of other healthy foods. Research examining food purchasing patterns conducted by the USDA using Nielsen Homescan data show that percentage of food spending on fruits and vegetables fell while packaged and processed foods rose from 1998 to 2006 (Okrent, 2012). Restaurant meals and meals away from home increased between the 1970’s and the 1990’s and between the 1990’s to the 2000’s, suggesting a general upward trend (Guthrie et al., 2002). Whether or not *trans* fat reduction has exacerbated over-consumption of unhealthy meals and meals away from home has yet to be seen and may be another area of future research. The second concern is that *trans* fats may be replaced by some other unhealthy ingredient, perhaps by saturated fats or by some new oil formulations for which the effects on health are still unknown. Much of the research on product reformulations shows efforts from industry, based on economic incentives, to reduce *trans* fat without increasing saturated fat content (Mozaffarian et al., 2010; Stender et al., 2009; Unnevehr and Jagmanaite, 2008; Van Camp et al., 2012). Instead of saturated fat, many new processes and oils have been developed (Eckel et al., 2007; Ratnayake et al., 2008; Tarrago-Trani et al., 2006; Wang et al., 2016). Some early evidence suggests that substitutions of palm oil and interesterified fat may contribute to obesity in animal studies, however little is known about specific effects on cardiovascular health in humans (Magri et al., 2015).

Very recent research on the health effects of *trans* fat reductions finds evidence of an effect on CVD mortality rates one year after bans were implemented from the New York City *trans* fat ban (Restrepo and Rieger, 2016). Medical literature suggests that effects from dietary changes can be seen in cholesterol levels within weeks, while changes in mortality rates would not be seen for about a year following changes in diet. The authors use a synthetic cohort approach to examine pre- and post-policy aggregate CVD and stroke mortality rates. Synthetic control counties were selected from metropolitan statistical areas (MSAs) with populations of one million or more throughout the US. For the counties implementing the ban, the authors find CVD mortality rates decline relative to the synthetic control after the ban is implemented. The authors do check for the possibility that changes in CVD rates may be driven by the financial crisis, which coincided with the ban, or by changes in other CVD risk factors. However, the authors do not test if CVD rate decline started before the ban was implemented. This chapter, on the other hand, addresses *trans* fat reduction policies at the national level, which includes changes in *trans* fat content of both restaurant and pre-packaged foods and examine whether they played a role in improving intermediate health measures prior to the NYC *trans* fat ban.

Understanding the health effects of *trans* fat consumption and the policy change has motivated the research question on whether or not *trans* fat reduction policies in restaurant and packaged foods have affected consumers in a way that improves measures of intermediate health. Information on pre-packaged food consumption spanning the years of the policy are not available in this data, however I can show that the consumption of pre-packaged foods is highly correlated with restaurant meal consumption and that both contribute to FAFH, thus making average weekly restaurant meal consumption an appealing proxy for consumption of all commercially prepared foods affected by these policies. Comparing a measured health outcome such as LDL in individuals exposed to higher levels of FAFH to those who are consume less FAFH both before and after the *trans* fat reduction policies provides a natural experiment. In this work, I use restaurant meal consumption frequency as a proxy for all FAFH consumption. An experiment like this allows us to look at whether and to what extent these policies have had the desired effect of improving health amongst consumers of affected foods. To do this, data which includes measured health outcomes, measures of FAFH consumption, as well socio-economic and health behavior data are needed. This would allow for use of a difference-in-difference method to compare how LDL levels have changed amongst the treated, consumers of FAFH, with the untreated, those who consume less or no FAFH. This research is the first to show that at the population-level, there is a positive health effect of these policies on the most important intermediate risk factor relating *trans* fat consumption and CVD, LDL cholesterol levels.

2.3. Data

This work uses data from the 1999 through 2009 Continuous National Health and Nutrition Examination Survey (NHANES), which is a program of the National Center for Health Statistics (NCHS). NHANES is a series of repeated cross-sectional surveys designed to assess the health and nutritional status of adults and children in the United States. NHANES uses a complex multistage probability sampling design to select participants representative of the civilian, non-institutionalized US population. Certain population subgroups are oversampled in order to increase the reliability and precision of estimates of health measures for these groups. Appropriate sampling weights and sample design variables are used in all analyses. The Continuous NHANES has been collected continuously since 1999. Data is collected in two-year waves or cycles, so the 1999 wave represents data collected both in 1999 as well as in 2000. The dataset is unique within a US context in that it combines household interviews with physical examinations by trained medical personnel. The survey consists of separate questionnaires covering demographic, socioeconomic, dietary, and health-related questions, conducted in respondent's homes by NHANES staff and is followed by standardized medical examinations conducted in NHANES Mobile Examination Centers (MEC) by physicians and health technicians. All survey participants visit the physician at a MEC facility to receive physical examinations which include dietary interviews and body measurements. This includes measures of height and weight from which BMI is calculated. No changes have been made to the NHANES body measurement protocol since the Continuous NHANES began in 1999. All participants over the age of one have a blood sample taken; measures of total cholesterol and high-density lipoprotein cholesterol (HDL) are taken from this sample. One half of all participants were randomly assigned to attend morning examination sessions and the other half attend afternoon examination sessions. Of participants aged 12 and over, half of those scheduled for morning sessions were randomly assigned to fast for nine hours prior to their scheduled appointment time to allow for measures of fasting plasma glucose, two-hour glucose, serum insulin, triglycerides, to be taken. Because the LDL and triglycerides measures are only available for the fasting subsample of participants, sampling weights specific to this group are applied.

The period from 1999/2001 through 2009/2010 starts with a sample size of 62,160 observations. This work limits the sample to individuals aged 20 years or older; interpretation of cholesterol levels and other health indicators are different for youths than for adults. This leaves 32,464 observations. The sample is further reduced to 30,020 after removing individuals who do not provide an annual household income in the questionnaire. Individuals who do not answer the question about highest education level achieved are also removed, decreasing the sample to 29,971. Individuals who lack information on country of birth are excluded, leaving 29,966 observations. Individuals for whom smoking behavior is not available are excluded, reducing the sample to 29,945, and individuals for whom drinking behavior information is not available are excluded, reducing the sample further to 26,117. Exclusion of individuals for whom no information on alcohol consumption reduces the sample size by nearly 14%. While existing literature finds the response rate for the alcohol consumption component of NHANES to be acceptable in terms of representativeness of the US population, there is a possibility that losing so many observations introduces bias (Guenther et al., 2010; Taylor et al., 2016). One concern is that if individuals with unhealthy levels of alcohol consumption are not be providing information on their drinking behavior, this could bias results, depending on whether or not individuals also have patterns of unhealthy eating because alcohol consumption does affect cholesterol levels. Compared to 2014 estimates of heavy drinking and binge drinking prevalence among adults aged 18 and up, 7% and 25% respectively, these data show only a slightly higher percentage of heavy drinkers (10%), and a much lower percentage of binge drinking behavior (13%) (National Institute on Alcohol Abuse and Alcoholism (NIAAA), 2016). Two possible reasons for the difference in binge drinking prevalence may be because binge drinking is more common among younger drinkers and this paper looks at a sample where age is 20+ while the NIAA estimates look at individual aged 18 and up; a second reason for the discrepancy is that binge drinking behavior has been increasing over time (Dwyer-Lindgren et al., 2015). Finally, individuals for whom there is no information on whether or not they eat meals prepared outside of the home are excluded, leaving 26,112 observations. Of these, 11,569 observations have valid measures of LDL, which is the main outcome of interest. From here, 564 observations have survey sampling weights equal to zero, so this results in a final sample of 11,005. Later on, this work estimates the preferred model using other health measures such as triglycerides, total cholesterol, HDL cholesterol, C-reactive protein (CRP), BMI, and waist circumference as outcome variables; each of which has a slightly different number of observations; appropriate sampling weights are used.

The main outcome variable of interest is LDL cholesterol. Values for LDL cholesterol in NHANES data are calculated from measured values of triglycerides, total cholesterol, and HDL cholesterol according to the Friedewald calculation where:

(2.1) $LDL = Total\ Cholesterol - HDL - Triglycerides/5$

This calculation is valid for triglyceride levels less than or equal to 400 mg/dL, so observations with triglycerides above 400 mg/dL are missing LDL cholesterol data (Centers for Disease Control and Prevention (CDC), 2002). For reference, cholesterol levels are measured in milligrams of cholesterol per deciliter of blood, or mg/dL. High cholesterol is a major risk factor for heart disease. Total cholesterol levels below 200 mg/dL are considered desirable and contribute a low risk for heart disease. Total cholesterol is considered borderline high for levels between 200 mg/dL and 239 mg/dL, and high for levels at 240 mg/dL and above (National Heart, Lung, and Blood Institute (NHLBI), 2005). Similarly high LDL cholesterol cut-points are shown in Table 2.2:

Table 2.2: LDL Cholesterol Categories

LDL Cholesterol Level	LDL-Cholesterol Category
Less than 100 mg/dL	Optimal
100-129 mg/dL	Near optimal/above optimal
130-159 mg/dL	Borderline high
160-189 mg/dL	High
190 mg/dL and above	Very high

Source: National Heart, Lung, and Blood Institute, 2005

High levels of triglycerides can also raise heart disease risk; levels of 150 mg/dL to 199 mg/dL are considered borderline high and levels of 200 mg/dL and above are considered to be high. Conversely, HDL cholesterol has a protective effect against heart disease, so higher levels of HDL are preferred. HDL levels below 40 mg/dL are low and are considered to be a risk factor for heart disease. HDL levels of 60 mg/dL or higher lower the risk for heart disease (National Heart, Lung, and Blood Institute (NHLBI), 2005).

Sociodemographic data from the study include age, gender, race, immigrant status, household size, household income, and education. NHANES age data is top-coded at 80. Data includes a dummy variable which equals zero if female, one if male. Dummy variables for five racial/ethnic groups are included: white, black, Mexican-American, other Hispanic, and all other races which includes Asians, Pacific Islanders, Native Americans and Alaska natives. A dummy variable for immigrant status is included, which equals zero if the individual was born within the US and one otherwise. Literature suggests that FAFH eating habits of immigrants vary greatly from the eating habits native-born citizens (Antecol and Bedard, 2006). Another dummy variable for individuals who live in large households, defined as having four or more household members, is included. A dummy variable for education status is included, which equals zero for individuals who have not completed high school and equals one if the individual has a high school degree. Dummy variables for income quintile were created using a categorical variable for household income. NHANES reports household income not as a continuous variable, but as 15 different categories or bands; some literature suggests inaccuracy exists in categorical household income variables, as compared to other measures of income (Micklewright and Schnepf, 2010). Dummy variables for income quintile are created by grouping these bands into quintiles. The lowest quintile (n = 6,075) includes incomes from $0 to $19,999, the second (n = 6,309) includes incomes from $20,000 to $34,999, the third (n = 6,595) includes incomes from $35,000 to $64,999, the fourth (n =1,301) includes incomes from $65,000 to $74,999, and the fifth (n = 5,526) includes incomes of $75,000 and above. The median income in the sample is around $45,000; so an alternate specification for income is a dummy variable which equals zero if household income is below $45,000 and equals one otherwise. In the CAPI interview, individuals who refuse or don't know their income are prompted to answer whether their income is below or above $20,000 and are coded in a separate category. The "below $20K" incomes are included in the first quintile and the "$20K+" incomes in the second quintile; as such both categories are coded as a zero in the $45K dummy variable. The inclusion of income as a control does not greatly affect these estimates, nor does the choice of coding for these individuals.

Indicator variables are also created to control for an individual's health characteristics and behaviors: pregnancy status, smoking status, alcohol consumption status, exercise behaviors, lipid-lowering medication use, and anti-hypertensive medication use. NHANES provides a variable for pregnancy status at the time of interview; this equals zero for all non-pregnant individuals, including males, and one for pregnant individuals. To identify smokers, two questions from the NHANES home interview about smoking behavior are used: "Have you smoked at least 100 cigarettes in your entire life?" and "Do you now smoke cigarettes?" Individuals who answered negatively to both questions are coded as nonsmokers while those who answer yes to either question are coded as smokers. This method of classifying smokers could misclassify infrequent, former, and very recent smokers as non-smokers; however a more precise measure of smoking/non-smoking status is not available with this data. Similarly, dummy variables are created for alcohol consumption. NHANES asks adults the question "In any one year, have you had at least 12 drinks of any type of alcoholic beverage?" Individuals who respond negatively to this question receive the follow-up question of "In your entire life, have you had at least 12 drinks of any type of alcoholic beverage?" Individuals who respond negatively to this question are coded as never-drinkers. Individuals who answered positively to either question then receive the follow-up questions of "In the past 12 months, how often did you drink any type of alcoholic beverage?", "In the past 12 months, on those days that you drank alcoholic beverages, on the average, how many drinks did you have?", and "On how many days in the past 12 months did you have 5 or more drinks of any alcoholic beverage in a single day?" These responses are used to calculate an individual's number of drinks per week and to categorize individuals as moderate and heavy drinkers, as to identify well as binge-drinking behavior, as defined by the NIH (National Institute on Alcohol Abuse and Alcoholism (NIAAA), n.d.). Two indicator variables are created for identifying an individual's participation in physical activity at home and on the job, based on a series of questions about physical activity. I categorize individuals who engage in exercise as those who answer yes to questions asking if they walked or bicycled, did vigorous activity, moderate activity, or muscle strengthening activities over the past 30 days. Individuals are categorized as having a physically-demanding job or doing manual work if they indicate that they do light lifting often, climb stairs or hills often, or do heavy work or carry heavy loads either as a part of their usual daily activities or part of their work activities. During the NHANES home interview, respondents are asked if they have used any prescription drug within the past month; those who answer affirmatively were asked for the name, duration and main reason each product was used. In roughly 84% of cases, the interviewer recorded exact product names from the label on the medication container; otherwise they used the information that was verbally reported by the respondent. To identify individuals using cholesterol-lowering and blood pressure-lowering medications, this prescription drug data is used to create indicator variables for each individual that equals one if the respondent is taking any drug grouped under "statins" or "anti-hyperlipidemic agents" class (for cholesterol-lowering medications) or the "metabolic drugs" class (for blood pressure-lowering medications) in the National Center for Health Statistics Lexicon Plus database of Therapeutic Drug Categories and equals zero otherwise (Tattersall et al., 2011).

A key set of variables in this work has to do with where people eat meals. In all years used in this work, the Diet Behavior and Nutrition component of NHANES asks individuals how many meals they eat in restaurants or outside of the home – this is the variable I use to identify FAFH consumption. The way in which this question was asked has changed slightly over time. In the years 1999, 2001, and 2003, the question was worded as how many times per week, on average, the respondent eats meals that were prepared in a restaurant, which includes eat-in restaurants, carry-out restaurants, and restaurants that deliver food to one's house. In 2005, 2007, and 2009, the Diet Behavior and Nutrition questionnaire asks respondents how many meals per week are eaten that were not prepared at home. In this question, away from home is defined to include meals from dine-in and carry-out restaurants, restaurants that deliver food to the home, cafeterias, fast food places, food courts, food stands, meals prepared at a grocery store, and meals from vending machines. In 2007 and 2009, an additional sentence is added to remind respondents not to include meals that were eaten in a school or community meal program, as these were reported in previous questions. Although the change in the wording of this question appears to be associated with an increase in the mean number of meals consumed outside of the home, I do not find any appreciable difference in the mode or in the distribution of the number of these meals, which is discussed further in Appendix A. This matches research done by the USDA Economic Research Service, which shows that money spent on food away from home had been increasing sharply between 2004 and 2005, relative to slower growth in surrounding years (Todd and Mentzer Morrison, 2014). A categorical variable is created for meals out per week; this equals zero for individuals who eat out less than once per week, equals one for those who eat out one to three times per week, equals two for those who eat out four to six times per week, and equals three for those who eat out seven or more times per week. This grouping is intuitively appealing; data from 1999-2000 NHANES show US adults consumed 2.8 commercially prepared meals per week on average, so one to three meals per week captures the average consumer (Kant and Graubard, 2004). The category of four to six meals per week groups consumers who eat more frequently than average but still less than an average of once per day. The seven meals or more grouping identifies consumers who eat out at least one meal per day or more.

Table 2.3: Summary Statistics of Socio-Economic and Health Behavior Covariates

	(1)	(2)	(3)	(4)	(5)
	All Groups	<1 Meals/ Week	1-3 Meals/ Week	4-6 Meals/ Week	7+ Meals/ Week
Socio-Economic Variables:					
Male	0.48	0.46	0.45	0.53	0.63
Age 20-29	0.17	0.11	0.17	0.24	0.24
Age 30-39	0.17	0.11	0.18	0.19	0.22
Age 40-49	0.17	0.14	0.17	0.19	0.17
Age 50-59	0.14	0.13	0.14	0.16	0.16
Age 60-69	0.16	0.21	0.16	0.11	0.11
Age 70+	0.19	0.30	0.18	0.11	0.10
White	0.52	0.43	0.53	0.56	0.57
Black	0.18	0.21	0.17	0.18	0.17
Mexican Am.	0.20	0.23	0.20	0.17	0.15
Other Hispanic	0.06	0.09	0.06	0.05	0.06
Other Race	0.04	0.04	0.03	0.04	0.04
Immigrant	0.22	0.32	0.21	0.16	0.15
Large Household	0.34	0.33	0.35	0.34	0.29
High School Grad	0.71	0.56	0.73	0.81	0.83
HH Income 45K+	0.43	0.26	0.45	0.54	0.57
Health Behavior Variables:					
Exercise	0.64	0.55	0.65	0.69	0.68
Manual Work	0.30	0.26	0.30	0.33	0.38
Cholesterol Meds	0.16	0.18	0.15	0.13	0.15
Blood Pressure Meds	0.30	0.39	0.30	0.24	0.22
Pregnant	0.04	0.03	0.05	0.04	0.02
Smoker	0.21	0.23	0.20	0.21	0.22
Moderate Drinker	0.76	0.73	0.77	0.77	0.81
Heavy Drinker	0.10	0.09	0.09	0.12	0.12
Binge Drinks	0.13	0.11	0.12	0.16	0.18
No. of Meals Out	2.84	0.20	1.82	4.76	10.15
SD	(0.031)	(0.005)	(0.011)	(0.016)	(0.104)
Observations	11,569	2,853	5,492	1,926	1,298
% Meals Out Fast Food	0.48	-	0.50	0.48	0.43
Observations	7,803	-	4,438	1,765	1,600

Data: NHANES 1999-2010.

Note: Unweighted means for indicator variables for all years are shown, except for % Meals Out Fast Food, for which data is only available for 2007 and 2009 waves of NHANES. Statistics are grouped into categories based on the number of meals out per week. The mean and standard deviation for the Number of Meals Out per week are also reported.

Table 2.3 shows unweighted summary statistics for socio-economic and health behavior covariates for the full sample and by restaurant meal frequency groups. These raw data show males, whites, younger individuals, individuals with higher income, with a high school education, physically demanding jobs, individuals who exercise, and binge drinking habits eat out more often. Conversely, older individuals, Mexican Americans, immigrants, individuals in large households, cholesterol and blood pressure medications, and moderate drinking behavior are correlated with eating meals out less frequently. In 2007 and 2009, an additional question was asked about frequency of consumption of meals from fast food restaurants specifically per week; estimates of the percentage of meals out from fast food are presented at the bottom of Table 2.3. Across all eating groups, roughly half of meals out come from fast food, however, as the number of meals out increases, the percentage of meals from fast food establishments declines.

Table 2. 4: Summary Statistics of Health Outcome Variables

	(1)		(2)	(3)	(4)	(5)
	All Groups		<1 Meals/ Week	1-3 Meals/ Week	4-6 Meals/ Week	7+ Meals/ Week
	N	Mean	Mean	Mean	Mean	Mean
Health Measures:						
LDL	11,569	117.92	118.86	118.22	117.54	115.18
SD		(0.334)	(0.700)	(0.479)	(0.812)	(0.950)
Triglycerides	12,126	144.96	149.11	145.00	142.37	139.46
SD		(1.142)	(2.307)	(1.608)	(3.119)	(3.175)
Total Cholesterol	24,909	200.15	202.38	200.53	198.70	195.82
SD		(0.272)	(0.568)	(0.390)	(0.663)	(0.774)
HDL	24,908	52.88	53.37	53.26	52.24	51.23
SD		(0.103)	(0.211)	(0.150)	(0.248)	(0.288)
CRP	25,010	0.47	0.53	0.46	0.44	0.39
SD		(0.006)	(0.014)	(0.007)	(0.013)	(0.015)
BMI	25,657	28.77	28.50	28.86	28.85	28.86
SD		(0.041)	(0.081)	(0.060)	(0.102)	(0.121)
Waist Circumference	25,357	98.42	98.28	98.39	98.43	98.79
SD		(0.098)	(0.192)	(0.142)	(0.246)	(0.303)

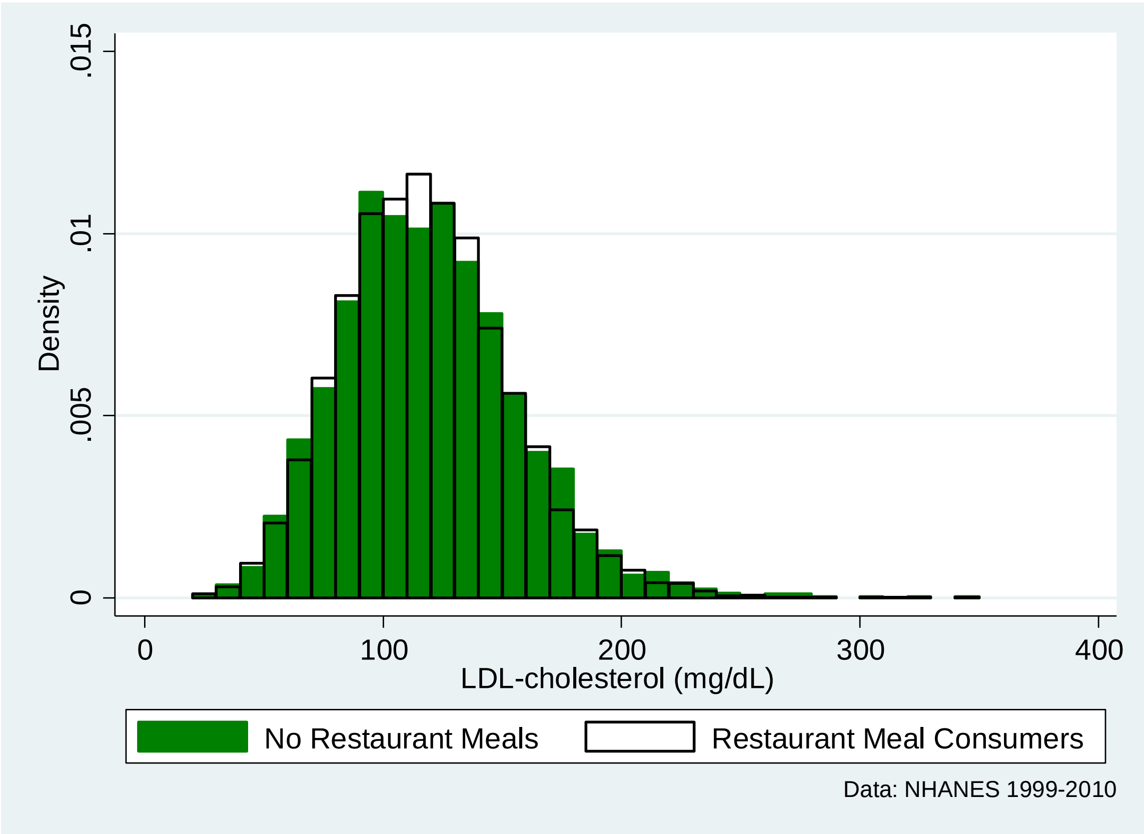
Data: NHANES 1999-2010.

Note: Unweighted means and standard deviations are shown for all health measure variables. Statistics are grouped into categories based on the number of meals out per week.

Table 2.4 presents unweighted means of different health outcome variables for the full sample, as well as by restaurant meal frequency groups. LDL, triglycerides, total cholesterol, and CRP levels are lower among the higher-frequency meals prepared away from home group than the lower-frequency groups. This can be explained by the positive correlation between eating meals prepared away from home and income combined with the negative correlation between income and cholesterol levels. Another factor that explains this is that the number of meals of away from home increases over time while cholesterol levels decline over time. These trends are not present for BMI and waist circumference.

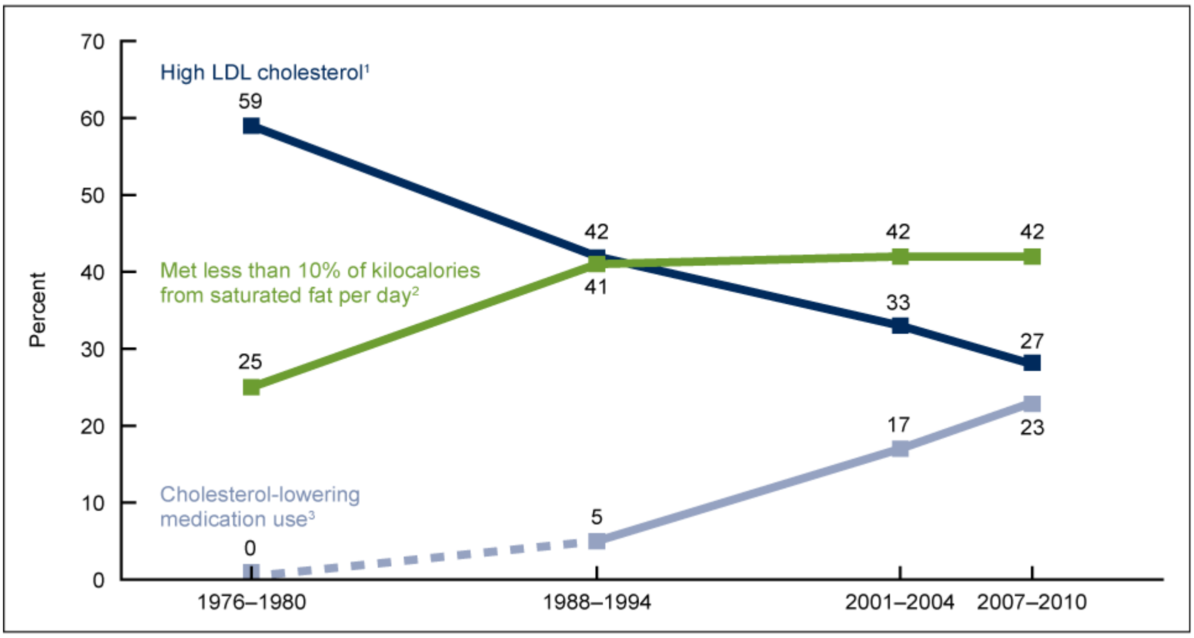
Figure 2.2 shows an unweighted distribution of LDL values across the sample. LDL exhibits some characteristics of a normal distribution, but has a very long right tail and is truncated at zero. The distribution is very similar across restaurant meal frequency. To account for the long tail on the right-hand side, I consider dropping observations with LDL values of 300 or above and also values of 200 or above. In both cases, the conclusions were not affected by leaving out the high-LDL observations.

Figure 2.2: Distribution of LDL by Restaurant Meal Consumption, 1999-2010



During the period of analysis, LDL-cholesterol and total cholesterol levels declined in the US, due in part to a decrease in TFA consumption, an increase in the percentage of adults taking lipid-lowering medications, and an increase in other healthy lifestyle changes (Carroll, et al. 2012). A negative trend is also seen in those adults not taking lipid-lowering medications. Figure 2.3 presents these trends in the US population from 1976 through 2010.

Figure 2.3: Age-Adjusted Trends in Prevalence of High LDL Cholesterol Levels, Use of Cholesterol-Lowering Medications, and Low Saturated-Fat Intake among US Adults Age 40-74.



[1]Significant decreasing linear trends from 1976–1980 to 2007–2010 ($p < 0.05$).
[2]Significant increase from 1976–1980 to 1988–1994 ($p < 0.05$); no significant change from 1988–1994 to 2007–2010.
[3]Significant increasing trend from 1988–1994 to 2007–2010.
NOTES: Data on the use of cholesterol-lowering medication was not collected in 1976–1980. High low-density lipoprotein (LDL) cholesterol is measured levels above the treatment goals established by the National Cholesterol Education Program's Adult Treatment Panel III guidelines. Estimates are age adjusted by the direct method to the 2000 U.S. census population using age groups 40–64 and 65–74.
SOURCE: CDC/NCHS, National Health and Nutrition Examination Survey.

[2]

Given this, part of the empirical challenge in this chapter is to identify the policy effect of lowering *trans* fat consumption while a general decreasing trend in health indicators is already present. The data also show that on average, more frequent consumers of FAFH actually have lower cholesterol levels. Wealthier individuals having lower cholesterol levels is consistent with the literature (Adler and Ostrove, 1999; Kraus et al., 1980; Luepker et al., 1993).

2.4. Empirical Strategy

This work seeks to identify the effect on LDL cholesterol levels from *trans* fat reduction policies in commercially prepared foods. To do this, difference-in-difference specifications are estimated where

[2] Reprinted from “Trends in high LDL cholesterol, cholesterol-lowering medication use, and dietary saturated-fat intake: United States, 1976–2010” by Kuklina EV, Carroll MD, Shaw KM, Hirsch R. NCHS Data Brief No. 117. Hyattsville, MD: National Center for Health Statistics. 2013. < http://www.cdc.gov/nchs/products/databriefs/db117.htm >

the difference between LDL levels in periods before and after the *trans* fat legislation is contrasted between high-frequency consumers of restaurant meals and lower-frequency consumers of restaurant meals. I use a reduced form model based on a general economic framework for understanding health. This reduced form is based on structural models relating health behaviors and utility derived from good health (Cawley, 2004; Cawley and Ruhm, 2011; Grossman, 1999). The main estimating equation is the following difference-in-difference model which explores the variation in cholesterol levels across both time and restaurant meal frequency:

$$(2.2) \qquad LDL_{it} = \beta_0 + \beta_1 PolicyXMeals_{it} + \beta_2 Meals_{it} + \beta_3 X_{it} + \gamma_t + \varepsilon_{it}$$

In this model, low-density lipoprotein *(LDL*), measured in mg/dL for each individual, is the outcome variable. The coefficient β_1 provides the difference-in-difference estimate of the effect of the policy. *PolicyXMeals* is an interaction between the dummy variable *Policy*, which equals one for the year 2005 and beyond and a zero for previous years, and the categorical variable *Meals* which equals zero for individuals who eat out less than once per week, equals one for those who eat out one to three times per week, equals two for those who eat out four to six times per week, and equals three for those who eat out seven or more times per week. Individual level socioeconomic characteristics and health behaviors are included in the matrix *X*. These include: age, age squared, race, immigrant status, household size, high school graduate, income quintile, exercise during leisure time, physical activity on the job, eating out at restaurants, smoking, drinking behavior, and use of lipid-lowering and blood pressure medications. The model includes two-year period fixed effects, γ_t, to control for factors that may affect the outcome variable across all meals out groups equally, such as changes in the process or methods of collecting and measuring the outcome variable and secular decline in LDL cholesterol levels over time. An error term is included and standard errors are clustered to allow for random correlation of observations within the strata and primary sampling unit (PSU) level.

A difference-in-difference model, such as this, relies on the same standard assumptions of the OLS model and additionally assumes a parallel trend assumption. This assumption is violated when something other than the treatment changes in one group but not in the other at the same time as the treatment. Section 2.5.1 presents pre- and post-policy trends and tests for violations of these assumptions.

Both the policy examined and the dataset used lend themselves to a number of identification issues. One of which is the identification of the pre- and post-policy periods. As outlined in Section 2.2, a few different policies were implemented between approximately 2003 and 2008; these include the labelling requirement on processed foods in 2003, the New York City restaurant *trans* fat ban of 2006, high profile class-action lawsuits against Kraft, McDonalds, and KFC in 2003, 2005, and 2006, as well as a number of smaller legislative actions in different localities in the US. This work, and much of the literature on *trans* fat content in processed foods, proposes that the overall effect of these policies could be seen as early as 2005, with effects of additional reductions in *trans* fats occurring later. The robustness section addresses this issue in more detail by looking at the effect of the policy on the treatment groups in individual years, by examining pre-2005 and post-2005 trends of the restaurant meal consumer groups, and also by examining the relationship between restaurant meal consumption and consumption of other types of commercially prepared foods which were also affected by these policies.

Another identification issue is identification of clear treatment and control groups. Consumption of meals away from home identifies those individuals who are affected by the *trans* fat reduction treatment. To consider different specifications of identifying the control and treatment groups, Table 2.6 in the robustness section compares models using different specifications of the *Meals* variable. Another concern is endogenous self-selection in the control group; that individuals who identify as eating less than one restaurant meal per week may be significantly different from the treatment groups, those who do eat at least one meal per week away from home. Particularly, they may have healthier behaviors than the restaurant meal consumers. The robustness section addresses this concern by examining trends in other health outcomes in the pre- and post-policy periods, focusing on those outcomes which are not shown to be as heavily influenced by TFA consumption as LDL levels.

2.5. Results

Table 2.5 presents the main result, the estimation of Equation (2.2), which suggests that the *trans* fat reduction policies led to a statistically significant improvement in the intermediate health measure of LDL cholesterol level among those individuals most affected by the treatment. First, Column 1 estimates a basic version of Equation (2.2) that controls for two-year period fixed effects and no other controls. The coefficient on *PolicyXMeals* in Column 1 shows that *trans* fat reduction policies reduce the effect of eating meals away from home on LDL by roughly 2.2 mg/dL. This estimate can be used to estimate how much TFA was removed from diets by this policy. From the literature, exchanging 2.2 (1% of total energy intake) grams of TFA with an equal amount saturated fat results in an estimated reduction of LDL of 0.4 mg/dL while making the same exchange with 2.2 grams of a *cis* polyunsaturated fat results in an estimated reduction of LDL of 2.3 mg/dL (Mozaffarian et al., 2006), thus this estimate is consistent with a reduction of daily TFA consumption of between 2.2 grams and 13.2 grams. This represents a significant reduction in TFA consumption, as the pre-policy median TFA consumption averaged 6 grams per day.

Table 2.5: The Influence of Restaurant Meal Consumption on LDL-Cholesterol among US Adults

	(1)	(2)	(3)	(4)	(5)
	basic model	demographic controls	health behavior controls	education controls	income controls
Policy X Meals	-2.172**	-2.065**	-2.466**	-2.487**	-2.545**
	(0.994)	(0.997)	(0.977)	(0.973)	(0.980)
Meals	0.290	0.535	0.980	1.034	1.280
	(0.810)	(0.832)	(0.792)	(0.790)	(0.799)
2001 FE	-5.061**	-4.843**	-3.891*	-3.838*	-3.792*
	(1.930)	(2.136)	(2.087)	(2.090)	(2.059)
2003 FE	-8.431***	-8.958***	-7.430***	-7.375***	-7.366***
	(1.739)	(1.876)	(1.898)	(1.887)	(1.832)
2005 FE	-7.688***	-8.104***	-6.179***	-6.094***	-5.871**
	(2.137)	(2.332)	(2.297)	(2.266)	(2.238)
2007 FE	-6.914***	-7.417***	-5.099**	-5.019**	-4.757**
	(1.927)	(2.167)	(2.116)	(2.103)	(2.089)
2009 FE	-6.106***	-6.612***	-4.092*	-3.999*	-3.716
	(2.076)	(2.265)	(2.270)	(2.256)	(2.246)
Age		1.639***	1.795***	1.799***	1.848***
		(0.0866)	(0.0882)	(0.0883)	(0.0877)
Age Squared		-0.0229***	-0.0211***	-0.0213***	-0.0221***
		(0.00129)	(0.00124)	(0.00124)	(0.00124)
Male		2.913***	3.823***	3.793***	3.881***
		(0.748)	(0.780)	(0.782)	(0.778)
Mexican American		-0.185	-0.671	-1.090	-1.591
		(1.325)	(1.375)	(1.401)	(1.362)
Other Hispanic		-2.959	-3.378*	-3.639*	-4.254**
		(1.969)	(1.985)	(1.996)	(1.992)
Black		-1.627	-1.635	-1.803*	-2.209**
		(1.041)	(1.046)	(1.054)	(1.070)
Other Race		-3.835*	-2.900	-2.851	-3.153
		(2.113)	(2.014)	(2.006)	(1.999)
Immigrant		2.801**	2.473*	2.349*	2.475*
		(1.268)	(1.338)	(1.331)	(1.337)
Large Household		-0.900	-1.532*	-1.591*	-1.178
		(0.906)	(0.891)	(0.887)	(0.918)
Smoker			1.227	1.068	0.645
			(0.925)	(0.909)	(0.915)

Table 2.5, Continued

Moderate Drinker			0.396	0.492	0.752
			(1.249)	(1.257)	(1.265)
Heavy Drinker			-2.216	-2.137	-1.813
			(2.159)	(2.165)	(2.182)
Binge Drinks			0.000108	-0.0410	-0.0392
			(1.898)	(1.901)	(1.893)
Exercise			-0.525	-0.379	-0.175
			(0.815)	(0.839)	(0.829)
Manual Work			2.261**	2.259**	2.203**
			(0.936)	(0.936)	(0.937)
Cholesterol Meds			-21.34***	-21.33***	-21.23***
			(1.415)	(1.419)	(1.427)
Blood Pressure Meds			-6.375***	-6.397***	-6.547***
			(0.955)	(0.956)	(0.961)
Pregnant			8.435***	8.406***	8.481***
			(2.532)	(2.526)	(2.529)
High School Graduate				-1.320	-0.698
				(1.028)	(1.085)
Income Quintile 1					2.671**
					(1.259)
Income Quintile 2					0.152
					(1.075)
Income Quintile 3					1.743
					(1.556)
Income Quintile 4					-1.514
					(1.163)
Constant	125.2***	102.4***	97.75***	98.70***	96.82***
	(1.586)	(2.142)	(2.382)	(2.536)	(2.587)
Observations	11,005	11,005	11,005	11,005	11,005
R-squared	0.012	0.060	0.110	0.111	0.112

Data: NHANES 1999-2010

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Note: Results from OLS regression estimating Equation (2.2). LDL-cholesterol measured in mg/dL is the outcome variable in each column. *Meals* is a categorical variable, equal to 0 if the consumer eats at restaurants less than one time per week on average, equal to one if 1-3 meals are eaten per week, equal to 2 if 4-6 meals are eaten per week, and equal to 3 if 7 or more restaurant meals are eaten per week. *PolicyXMeals* is an interaction term between the variable *Policy*, which equals zero in years before 2005 and equals one in 2005 and after. Column 1 includes two-year period fixed effects and no other controls. Column 2 adds controls for the following demographic characteristics: age, age squared, sex, race categories (Mexican American, other Hispanic, black, and other non-white races), immigrant/foreign-born status, and large household size. Column 3 adds controls for health characteristics: smoking status, drinking behavior (moderate drinking, heavy drinking, and binge drinking), exercise participation, physically demanding job, use of cholesterol medications, use of blood pressure medications, and pregnancy status. Column 4 adds an indicator for high school graduate status as a control for education. Column 5 adds controls for income quintiles. Each column includes the controls used in the previous column.

The coefficient on *PolicyXMeals* is greater in magnitude than the coefficient on *Meals.* This indicates that after the policy goes into effect, eating at restaurants is no longer found to be correlated with higher LDL-cholesterol. However, this does not mean that restaurant meals are healthier on average than meals eaten at home, nor that eating at a restaurant has a particularly protective effect on consumers' overall health. Indeed, later on in Table 2.10, I estimate the influence of restaurant meal frequency on BMI and find the both the coefficient on *Meals* and on *PolicyXMeals* to be positive, indicating poorer health. Further, the 95% confidence interval for total effect of *Meals* after 2005 estimated by the sum of *Meals* and *PolicyXMeals* includes zero.

Column 2 of Table 2.5 adds demographic controls to this model; these include age, age squared, sex, race, immigrant status, and a dummy variable for large household. Marital status is not used as a control variable because the missingness in this variable was correlated with time; the missingness increased in the post-policy years. Column 3 controls for a number of health and behavioral variables. Columns 4 and 5 add additional indicator variables for high school education and income quintiles respectively. Because education and income are highly correlated with each other, controlling for both often results in a smaller estimated effect than controlling for each separately.

While the focus of this chapter is not to estimate all determinants of LDL cholesterol, it is worth noting from this table, that some control variables appear to have a more important relationship with LDL than others. The effect of age and age squared on LDL are statistically significant; this suggests that increases in age are associated with increases in LDL, but at a declining rate. Being male, black, other Hispanic, and foreign-born are all associated significantly with higher LDL. Working in a physically demanding job is associated with higher levels of LDL, and may be capturing an additional socio-economic effect. Pregnancy is also significantly associated with higher levels of LDL; this is in line with medical literature. Taking either blood pressure or cholesterol-reducing medications is also very highly and significantly associated with lower levels of LDL. Of the income and education controls, being in the lowest income quintile, income lower than $20,000, is significantly associated with higher levels of LDL.

Overall, the main result remains consistent as additional controls are added, suggesting that while the control variables are important in understanding the underlying level of blood cholesterol levels, they do not detract from the effect found in the basic model. This is strong evidence in support of the hypothesis that *trans* fat reduction has had a significant effect on the influence of restaurant meals on cholesterol levels.

2.5.1 Robustness

This section explores whether there is evidence that the effect found in the main result, Table 2.5, is due to the *trans* fat reduction policies implemented around the beginning of 2005. First, Figures 2.4 and 2.5 address the issue of parallel trends. Figure 2.6 presents an estimate of the effect of LDL on *Meals* for each year as evidence that the change occurred in 2005, when the *trans* fat reduction policies were beginning to go into effect. The preferred specification in the main estimation uses a categorical specification that I've defined for the frequency of eating meals out. Table 2.6 considers alternative specifications for average weekly restaurant meal consumption and shows robustness to these different specifications. In this work, consumption of restaurant meals serves as somewhat of a proxy for consumption of other types of commercially prepared foods containing *trans* fats and affected by the policies discussed in this work. Data on consumption of all types of FAFH does not exist in this dataset across both pre- and post-policy years, precluding replication of Table 2.5 using all FAFH consumption in place of restaurant meals. To justify use of the *Meals* variable as a proxy for all FAFH consumption, Table 2.7 shows evidence that people who eat restaurant meals also consume more ready-to-eat and frozen meals which are more likely to contain *trans* fat and would be affected by the *trans* fat policies considered in this work. To consider whether *Meals* is also indicative of other health behaviors that may somehow be affected by the *trans* fat policies, Table 2.8 estimates the main model for outcomes other than LDL. The effect of *trans* fat on LDL cholesterol is clearly indicated in medical literature as being different from the effect of *trans* fat on other health indicators. These results match what one would expect if the effect of the policy is in fact through a change in *trans* fat consumption rather than through a change in some other health behavior.

First, I consider the parallel trends, or common trends, assumption in the difference-in-difference model. This assumption is violated when something other than the treatment changes in one group but not in the other at the same time as the treatment. One assumption to ascertain the accuracy of the DID estimate is that the composition of individuals in the two groups remains unchanged over time.

To show common trends in the pre-policy period, Figure 2.4 shows mean LDL levels for the treated group, individuals who consume at least one meal away from home, and the control group, those who report eating less than one meal away from home per week. Although the trend is somewhat noisy and is not precisely parallel, the figure shows that both groups have similarly declining levels of LDL levels prior to the policy. A possible explanation for the similar, but not parallel, trend is that because the data are repeated cross-sections, there is a greater degree of individual-level variation or noise than is typically found in panel data. Another source of variation may be explained by changes in the instrument in 1999. An explanation for the overall declining secular trend is the increased use of statins during this period (Gu et al., 2014; Mann et al., 2008), which is not accounted for in this figure, but is controlled for in the regression estimates. This figure clearly shows that the post policy period LDL increases for both groups, but less so for the food away from home consumers, suggesting that some difference between these two groups has an effect on their LDL levels in the post-policy period. Changes in measurement caused by changes in the survey instrument over time should affect all groups equally and should not contaminate the main result. I attribute the difference in the two groups in the post-policy period to the policy itself.

Figure 2.4: Mean LDL by Restaurant Meal Consumption, 1999-2009

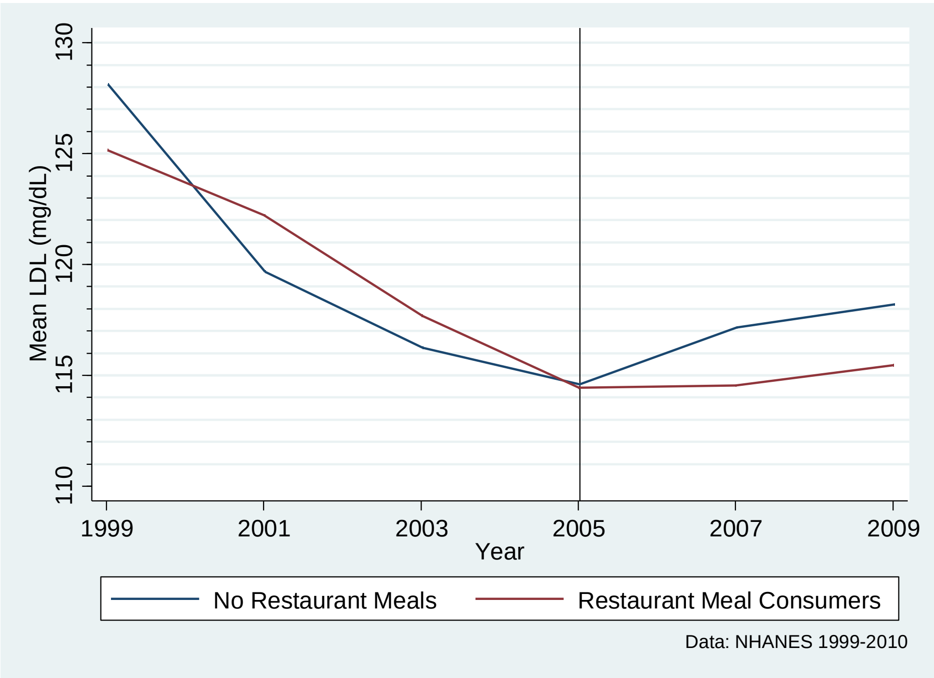
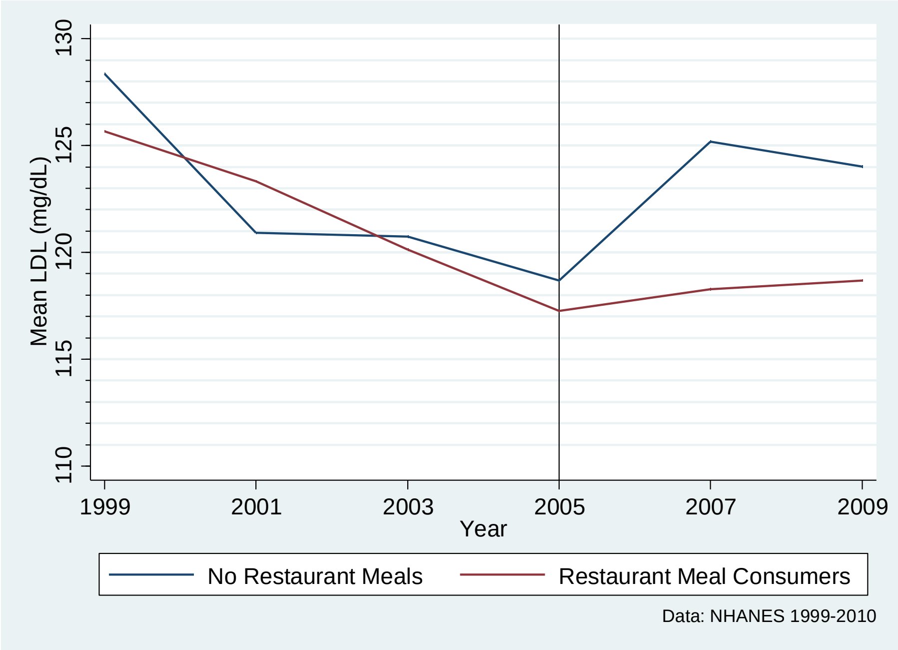


Figure 2.5 shows mean LDL levels for the treated group and the control group among only those individuals who are not taking cholesterol-lowering medications which include statins. This graph shows a larger divergence between restaurant meal consumers and non-consumers among individuals not using cholesterol-lowering medications following the policy period. The pattern present in this figure reinforces the hypothesis that a difference-in-difference approach is appropriate to estimate the effect of *trans* fat regulation over this period.

Figure 2.5: Mean LDL by Restaurant Meal Consumption Among Individuals Not Prescribed Cholesterol-lowering Medications, 1999-2009



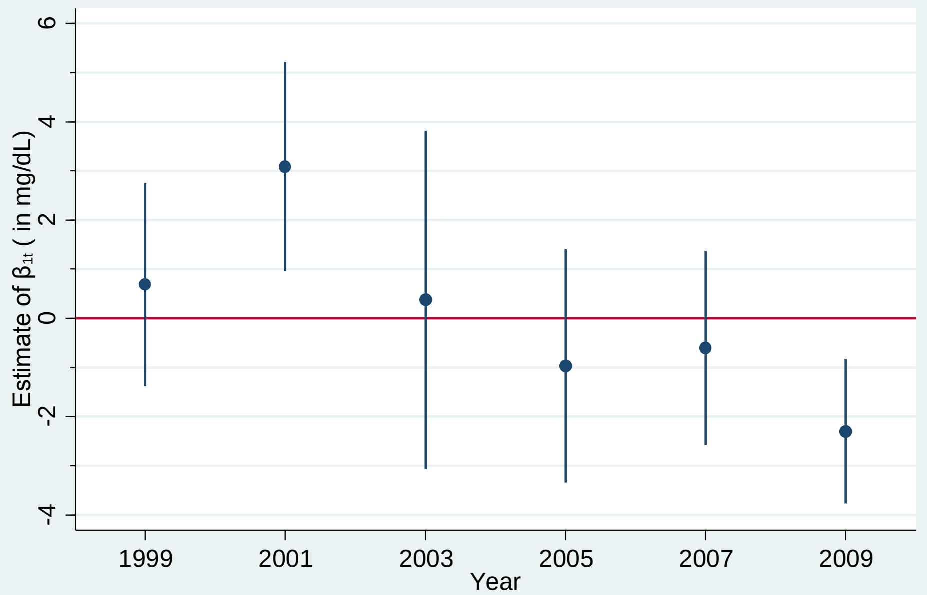
One possible cause for a rising trend in LDL in the post-policy period would be if health behaviors or eating habits changed during this time. The Great Recession occurred in 2007 and 2009 in the US, during the post-policy period, and may provide an explanation for the rise in LDL. A number of papers examine FAFH and restaurant meal consumption during this period, with somewhat mixed results (Smith et al., 2014; Todd and Mentzer Morrison, 2014). If more individuals who were previously eating restaurant meals replace restaurant meals with meals prepared at home, then it could change the composition of the group of individuals who consume restaurant meals. If unhealthier individuals decrease restaurant meal consumption, that could explain why LDL increases so much more for the No Restaurant Meals group. On the other hand, if healthier individuals decrease restaurant meal consumption, we would expect to see much more of a drop in LDL among non-consumers of restaurant meals and much higher LDL levels among the restaurant meal consumers. Similarly, the recession could have affected the quality of food; both food manufacturers and individuals preparing meals at home may have replaced more expensive, healthy, ingredients with cheaper, unhealthier, alternatives this could explain why LDL levels increase during the recession period. As an attempt to control for relationships between macroeconomic variables and LDL levels, statin usage, and eating behaviors, I include year fixed effects in all regressions.

As another check to support the use of 2005 as the policy year, I consider the timing of the policies whose effects I estimate. In order to understand the timing, the following equation is estimated:

(2.3) $LDL_{it} = \beta_{1_t} YearXMeals_{it} + \beta_2 X_{it} + \gamma_t + \varepsilon_{it}$.

This is similar to Equation (2.2), but the interaction of the policy dummy and the meals frequency variable, *PolicyXMeals* has been replaced with an interaction of a dummy for each two-year period in the data with the meals frequency variable, $YearXMeals$ and the intercept has been dropped. This specification shows an estimate of the effect on LDL of a change in *Meals* for each year. Figure 2.6 plots the estimated coefficients of the interaction terms, with 95% confidence intervals of the estimate for each year. This plot suggests that 2005 serves as a transition between a negative effect of *Meals* before the policy change and a non-negative or positive effect of *Meals* after the policy was fully implemented. In this figure, I see that only estimates from 2001 and 2009 are significantly different from zero and that the other three estimates may not be significantly different from each other. However, it does seem consistent with the data presented in Figure 2.4 to pool 2001 and 2003 as pre-policy years and 2005, 2007, and 2009 as post policy years. This confirms the choice of years to use for the *Policy* variable which is used in the main result in Table 2.5. Pooling years reduces the uncertainty around the estimate of *PolicyXMeals* in Table 2.5 compared to the error bars in the coefficient plot in Figure 2.6.

Figure 2.6: Estimated Yearly Change in LDL as Restaurant Meal Frequency Group Changes



Note: Coefficient plot of point estimates and 95% confidence interval for β_{1t} from Equation 2.3.
Data: NHANES 1999-2010

Another concern with properly identifying the policy period has to do with the data collection and examination dates. The dataset does not include date of interview or examination; instead, a variable indicating which six-month period in which the respondent's examination was conducted is available. There are two data-collection periods in each two-year cycle: November 1 of the first year through April 30 of the second year – which I call winter, and May 1 of the second year through October 31 of the second year – which I refer to as summer (Johnson et al., 2013). About 45% of examinations are conducted in winter periods and the remaining 55% occur in the summer. Focusing on the policy year, or the 2005 data cycle, only two months' data are from 2005, and the remaining ten months data is collected in 2006. Because of this, I use data from the 2005 cycle as the beginning of the post-policy period for the *trans* fat labelling policy change that occurred January 1, 2006.

Another potential limitation present in this data is the issue of seasonality in nutrition and eating meals away from home. In order to keep MEC operations running smoothly, certain geographic areas are avoided during winter months (Johnson et al., 2013). This would potentially decrease statistical efficiency for variables affected by seasonal and regional variation, such as cholesterol measures, diet, nutrition, and other health behaviors (Gordon et al., 1987; Joshi et al., 2014; Matthews et al., 2001; Ockene IS et al., 2004). While it is possible to only look at summer exams to avoid geographic heterogeneity in the data due to seasonality, this would reduce the sample size by nearly half and no longer be representative of the population. Instead, I add a dummy variable for winter examinations to account for regional seasonal variation in the preferred specification. Controlling for winter examinations, the estimates on all coefficients did not change significantly. The estimate of the coefficient of interest, *PolicyXMeals*, declined by 0.05 and statistical significance did not change.

One concern we may have is whether there is heterogeneity in the effect on cholesterol levels for individuals with different exposures to the *trans* fat regulations, via their frequency of restaurant meal consumption. One way to test this to use an empirical strategy similar to one that Autor (2003) presents for a difference-in-difference analysis where treatment varies across the treatment category and across time. In his specification, he includes dummy variables for the different treatment categories and additional interaction terms between time periods and each treatment category. Similarly, in this work, it is quite possible that treatment may vary across the different treatment categories (*Meals*). To consider this possibility, and use a model like that suggested in Autor (2003) in Table 2.6, I model the following equation, which is similar to Equation (2.2) but also includes four additional interaction terms, (*<1Meals)XYear, (1-3Meals)XYear, (4-6Meals)XYear*, and (*7+Meals)XYear*. In this equation, I also replace the *Meals* categorical variable in the model with dummy variables for each of the four *Meals* categories. When estimating Equation (2.4), the *(<1Meals)XYear*, *(<1Meals),* and 1999 are omitted as the reference groups.

(2.4) $LDL_{it} = \beta_0 + \beta_1 PolicyXMeals + \beta_2(< 1Meals)XYear_{it} + \beta_3(1 - 3Meals)XYear_{it} + \beta_4(4 - 6Meals)XYear_{it} + \beta_5(7 + Meals)XYear_{it} + \beta_6(< 1Meals) + \beta_7(1 - 3Meals) + \beta_8(4 - 6Meals) + \beta_9(7 + Meals) + \beta_{10}X_{it} + \gamma_t + \varepsilon_{it}$

Table 2.6: Estimated impact of *trans* fat regulation on LDL levels, considering variations across time periods and among different restaurant meal consumption frequencies

	(1) Preferred Model	(2) Meals Dummy Variables	(3) Meals X Time Trend
Policy X Meals	-2.545**	-2.263**	-0.782
	(0.980)	(1.008)	(2.389)
1-3 Meals X Year			0.987
			(1.841)
4-6 Meals X Year			0.293
			(1.177)
7+ Meals X Year			-0.921
			(0.908)
Constant	96.82***	96.05***	94.82***
	(2.587)	(2.582)	(2.893)
Year dummies	*Yes*	*Yes*	*Yes*
Meal dummies	*No*	*Yes*	*Yes*
Observations	12,299	12,299	12,299
R-squared	0.112	0.113	0.113

Data: NHANES 1999-2010

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Note: Results from OLS regression estimating Equation (2.4). Outcome variable is LDL cholesterol, in mg/dL. Column 1 shows preferred model from Table 2.5. Column 2 replaces the *Meals* categorical variable with dummy variables for each of the four *Meals* categories. Column 3 includes the same *Meals* dummy variables as the previous column and also includes interaction terms between *Year* and each of the *Meals* categories, following Autor (2003). All columns include the full set of control variables used in Column 5 of Table 2.5.

In this table, Column 1 is repeated from Column 5 of Table 2.5 for comparison as the preferred model. Column 2 replaces the *Meals* categorical variable with four separate binary variables for each *Meals* category: less than one meal per week, one to three meals per week, four to six meals per week, and seven or more meals per week. This change results in very little difference from the previous column. Column 3 uses the *Meals* specification from the previous column and additionally includes four interaction terms between *Year* and each of the four binary *Meals* categories, as shown in Equation (2.4) and following Autor (2003). Inclusion of these four interaction terms results in a loss in statistical significance and a sizeable reduction in the coefficient of interest. One possible explanation is that this model is over-specified.

Because the 2005 data cycle is assumed to be the change in policy for all meal groups, this paper uses an indicator for the policy period, so I suggest replacing the four interaction terms between the treatment intensity and the year trend, *<1MealsXYear, 1-3MealsXYear*, *4-6MealsXYear*, and *7+MealsXYear* with four interaction terms between each of the binary *Meals* categories and the policy year, 2005. This will capture the effect of the 2005 policy on each of the treatment categories separately. In Table 2.7, I model the following equation:

(2.5) $$LDL_{it} = \beta_0 + \beta_1(<1Meals)XPolicy_{it} + \beta_2(1-3Meals)XPolicy_{it} + \beta_3(4-6Meals)XPolicy_{it} + \beta_4(7+Meals)XPolicy_{it} + \beta_5(<1Meals)_{it} + \beta_6(1-3Meals)_{it} + \beta_7(4-6Meals)_{it} + \beta_8(7+Meals)_{it} + \beta_9 X_{it} + \gamma_t + \varepsilon_{it}$$

Table 2.7: Estimated impact of *trans* fat regulation on LDL levels, considering variations in weekly restaurant meal consumption

	(1) Preferred Model	(2) Policy X Meal Trends
Policy X Meals	-2.545**	-
	(0.980)	-
< 1 Meal X Policy	-	-3.060
	-	(2.638)
1-3 Meals X Policy	-	-6.225***
	-	(2.032)
4-6 Meals X Policy	-	-12.03***
	-	(2.459)
7+ Meals X Policy	-	-7.304**
	-	(2.844)
Constant	96.82***	95.48***
	(2.587)	(2.623)
Observations	12,299	12,299
R-squared	0.112	0.113

Data: NHANES 1999-2010

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Note: Results from OLS regression estimating Equation (2.5). Outcome variable is LDL cholesterol, measured in mg/dL. Column 1 shows the preferred model from Table 2.5. Column 2 replaces the *Meals* categorical variable with dummy variables for each of the four *Meals* categories. Column 2 replaces the *Policy X Meals* interaction term with four separate interaction terms between *Policy* and each of the *Meals* categories, following Autor (2003). All columns include the full set of control variables used in Column 5 of Table 2.5.

For comparison, results from the preferred model are presented in Column 1 of Table 2.7. Column 2 shows results from estimating Equation (2.5); the effects of the policy appear to have no significant effect on LDL for the lowest-frequency restaurant meal consumers, and larger and significant effects among individuals in the other three meal consumption categories. The greatest affect among consumers who eat 4-6 meals per week. That this group has a larger coefficient than the fewer than one meal per week group and the 1-3 meals per week group is in line with the idea that the more an individual eats out, the more they are exposed to the policy, so holding all else constant, we would expect to see a greater effect among this group. Following this, we would expect to see an even larger effect amongst the largest frequency consumers, however we do not find this. Some reasons we don't see the greatest effect among the highest-frequency consumers may be because these individuals are engaging in other unhealthy behaviors not captured in the model, or simply that at a certain amount of restaurant meals, the other unhealthy aspects of these meals outweighs the benefits of *trans* fat reduction in these foods.

Next, I consider the treatment of the estimate of the *Meals* variable, in order to test whether results are driven by the choice of identification of the treatment group. The choice of grouping the variable *Meals*, which measures average consumption of restaurant meals and meals prepared away from home per week, is intuitive but is not the only possible grouping. Table 2.8 presents three possible alternate specifications, which are to use the count of meals directly (Column 2), to use the square root of the count of meals to allow for decreasing marginal effects of additional meals (Column 3), or to think of the data in a binary fashion (Column 4). As a binary variable, zero identifies low frequency consumers, those who consume less than one meal away from home per week; a value of one identifies consumers who eat one or more meals per week. The coefficient on *PolicyXMeals* that is estimated in each of these models will have a different interpretation and the magnitude of the coefficients is not directly comparable because in each model, a one unit change in the value of *Meals* represents a different change in the number of meals eaten in restaurants per week. However, in each case the estimate is consistently negative and meaningful.

Table 2.8: Alternate specifications of Meals Predictor Variable

	(1) Meals: Grouping	(2) Meals: Count	(3) Meals: Sq. Root	(4) Meals: Binary
Policy X Meals	-2.545**	-0.519*	-2.640**	-4.911**
	(0.980)	(0.289)	(1.126)	(2.235)
Meals	1.280	0.210	1.250	3.420**
	(0.799)	(0.259)	(1.000)	(1.609)
Constant	96.82***	97.85***	96.56***	95.53***
	(2.587)	(2.588)	(2.904)	(2.630)
Observations	11,005	11,005	11,005	11,005
R-squared	0.112	0.111	0.112	0.112

Data: NHANES 1999-2010

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Note: Results from OLS regression estimating Equation (2.2) with different specifications of the *Meals* variable. LDL-cholesterol measured in mg/dL is the outcome variable in each column. In Column 1 *Meals* is a categorical variable, equal to 0 if the consumer eats at restaurants less than one time per week on average, equal to one if 1-3 meals are eaten per week, equal to 2 if 4-6 meals are eaten per week, and equal to 3 if 7 or more restaurant meals are eaten per week. In Column 2, *Meals* is a count of the number of restaurant meals per week. *Meals* in Column 3 uses the square root of the number of restaurant meals per week. And in Column 4, *Meals* is a binary variable which equals 0 if the consumer eats restaurant meals fewer than once per week and equal to 1 otherwise. *PolicyXMeals* is an interaction term between the variables *Meals* and *Policy*, which equals zero in years before 2005 and equals one in 2005 and after. All Columns include the full set of control variables used in Column 5 of Table 2.5.

An important aspect of this comparison is that it explores non-linearity in the effect of eating out. The literature suggests that even moderate levels of *trans* fat consumption can have a large adverse effect on health, and an alternative is to estimate a specification which assumes a diminishing marginal health cost to eating out. This can explain why the binary, grouping, and square root estimates show more significance than the estimates where *Meals* is specified to represent the untransformed count of meals eaten in restaurants.

Table 2.9 presents estimates of the relationship between the variable *Meals* and consumption of two other types of FAFH, specifically ready-to-eat prepared meals and frozen prepared meals. This estimation deals with the issue of identification, that the variables *Meals* and *PolicyXMeals* correctly identify those individuals who would be most affected by TFA reductions in commercially prepared foods. One assumption that I test is whether individuals who eat more restaurant meals are more likely to be affected by *trans* fat regulation in other parts of the food system, particularly in pre-processed and pre-packaged foods. FAFH typically includes not only restaurant meals, but other foods that have been commercially prepared, such as packaged and processed foods, frozen meals. For the full pre- and post-policy periods, data are only available on restaurant meal consumption, *Meals*, but not on these other components of FAFH. However, in the 2007 and 2009 data cycles, NHANES added questions on monthly pre-packaged and prepared foods consumption, on monthly frozen meals and frozen pizza consumption, and on weekly fast food and pizza meals out. So, while I cannot directly test the effect of the policy on these two groups, I can show that FAFH is positively correlated with the consumption of other foods affected by the policy. Data from these two years are examined using the following equation:

(2.6) $$Other\ FAFH\ Consumption_{it} = \beta_0 + \beta_1 Meals_{it} + \beta_2 X_{it} + \varepsilon_{it}$$

Two different specifications for *Other FAFH Consumption* are tested: a count of pre-packaged and prepared foods consumed in the past 30 days and a count of frozen meals and frozen pizza consumed in the past 30 days. Because these outcome variables are count data and the conditional distribution of the variable is over-dispersed, I perform a negative binomial regression using prepared foods consumption and frozen meals consumption as the outcome variables of interest. *Meals* are regressed against the number of ready-to-eat meals consumed per month; in Column 1 *Meals* is a count variable, and in Column 2, the *Meals* grouping variable is used. Similarly, in Columns 3 and 4, *Meals*, as a binary variable and as a categorical variable respectively, is regressed against the number of frozen meals or frozen pizza consumed per month. In both of these regressions the full set of controls from the preferred model are included. The estimate of the natural log of the dispersion parameter, α, is reported for each regression. When α is significantly greater than zero, then the data are over dispersed and are better estimated using a negative binomial model than a Poisson model. Estimation of these models confirm the significant positive association between eating meals in restaurants and eating pre-packaged and frozen meals. This relationship implies that using 2005 is correctly identified as the policy year when *trans* fat content of foods consumed by people who eat frequently at restaurants declined.

Table 2.9: Relationship between restaurant meals and other FAFH meal consumption

	(1) Ready-to-Eat Meals in 30 days	(2) Ready-to-Eat Meals in 30 days	(3) Frozen Meals in 30 days	(4) Frozen Meals in 30 days
Meals Category		0.448***		0.142***
		(0.0379)		(0.0325)
Meals per Week	0.103***		0.0310***	
	(0.00955)		(0.00819)	
Constant	-3.529***	-3.809***	-1.806***	-1.893***
	(0.177)	(0.184)	(0.149)	(0.149)
ln α	1.681***	1.668***	1.271***	1.269***
	(0.0497)	(0.0496)	(0.0391)	(0.0393)
Observations	10,097	10,097	10,100	10,100

Data: NHANES 1999-2010

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Note: Results from OLS regression estimating Equation (2.6). Outcome variable in Columns 1 and 2 are the number of ready-to-eat meals in 30 days; in Columns 3 and 4 are the number of frozen meals consumed in 30 days. Independent variable of interest in Columns 1 and 3 are the count of restaurant meals per week, and in Columns 2 and 4 categorical *Meals* variable which is equal to 0 if the consumer eats at restaurants less than one time per week on average, equal to one if 1-3 meals are eaten per week, equal to 2 if 4-6 meals are eaten per week, and equal to 3 if 7 or more restaurant meals are eaten per week. Columns include the full set of control variables used in Column 5of Table 2.5.

Interpretation of the *Meals* coefficient in Columns 1 and 3 is that for a change of one restaurant meal per week, the difference in the logs of expected counts of the number of ready-to-eat for frozen meals per month is expected to change by the value of the coefficient. Similarly, the coefficient on *Meals* in Columns 2 and 4 estimates the effect of moving from one restaurant frequency category to another. Consumption of food from restaurants is strongly and significantly correlated with consumption of other types of pre-packaged, ready-to-eat, and frozen meals. This supports the idea that the *trans* fat reduction in all commercially processed foods, not just in restaurants, plays a role in the improvements in health that are estimated when modeling the effect of eating meals away from home on health.

Another concern with using *Meals* to identify treatment and control groups is endogenous self-selection into the control group; that individuals who identify as eating less than one restaurant meal per week may be significantly different from the treatment groups, those who do eat at least one meal per week away from home. One approach to testing this is to examine whether the policy has effects on other health indicators, particularly those that are not as strongly linked with *trans* fat consumption, but are linked to other healthy behaviors like diet and exercise. Table 2.10 estimates the preferred model, using other indicators of health as the outcome variable of interest. Total cholesterol, HDL cholesterol, and triglycerides, in Columns 1, 2 and 3, respectively, are additional indicators of cardiovascular disease; C-reactive protein, Column 4, is an indicator of metabolic risk and inflammation. To varying degrees, these indicators are affected by TFA consumption, most particularly Total Cholesterol. The remaining columns show measures of body mass and body fat composition and are indicators of obesity and metabolic risk: BMI, in Column 5, and waist circumference, in Column 6.

Table 2.10: Influence of Restaurant Meal Consumption on Other Health Indicators among US Adults

	(1)	(2)	(3)	(4)	(5)	(6)
	Total Cholesterol	HDL	Triglycerides	CRP	BMI	Waist Circumference
Policy X Meals	-2.590***	-0.384*	2.804	0.0210	0.114	0.283
	(0.837)	(0.221)	(3.263)	(0.0133)	(0.102)	(0.224)
Meals	1.669**	-0.201	-0.718	-0.0122	0.217***	0.468***
	(0.731)	(0.180)	(2.927)	(0.00943)	(0.0780)	(0.171)
Constant	164.5***	48.43***	95.20***	0.455***	26.84***	86.43***
	(2.164)	(0.824)	(9.634)	(0.0378)	(0.360)	(0.935)
Observations	24,909	24,908	11,278	25,010	25,657	25,357
R-squared	0.118	0.206	0.056	0.037	0.120	0.190

Data: NHANES 1999-2010

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Note: Results from OLS regression estimating Equation (2.2) with different outcome variables. Outcome variables for Columns 1 through 4 are: total cholesterol, HDL cholesterol, triglycerides, and C-reactive protein levels, measured in mg/dL. BMI is the outcome variable in Column 5 and waist circumference, measured in centimeters, is the outcome variable in Column 6. *Meals* is a categorical variable, equal to 0 if the consumer eats at restaurants less than one time per week on average, equal to one if 1-3 meals are eaten per week, equal to 2 if 4-6 meals are eaten per week, and equal to 3 if 7 or more restaurant meals are eaten per week. *PolicyXMeals* is an interaction term between the variable *Policy*, which equals zero in years before 2005 and equals one in 2005 and after. All columns include the full set of control variables, used in Column 5 in Table 2.5.

This analysis isolates *trans* fat policy as the likely cause of the health effects shown in the main result. As outlined in the medical literature, *trans* fat consumption shows the largest and most immediate effects on LDL and total cholesterol, and diminished effects on HDL, triglycerides, and both diminished and delayed effects on abdominal adiposity (measured here as waist circumference), and BMI. Effects on blood lipids show up more quickly than on measures of body fat composition. So, if *trans* fats are reduced and no other health-improving changes are made, one would expect to see improvements in LDL and total cholesterol, possibly smaller improvements in HDL and triglycerides, depending on the amount of *trans* fat reduced, and likely no improvement in BMI or waist circumference. This is in line with estimates shown in Table 2.10. In fact, this table shows that across all indicators other than total cholesterol, no improvements in health are related to eating at restaurants after the intervention.

One concern with these findings is that HDL, Column 2, and triglycerides, Column 3, should have shown some improvements in health, even if small, however the lack of improvement could be due in part to documented changes in measurement of HDL, or it could be that *trans* fat reductions were not significantly large enough to outweigh other trends in unhealthy eating, or even that newer *trans* fat substitutes may not be as healthy as the *cis* unsaturated fats and saturated fats which were used as comparisons in clinical trials. The results for BMI and waist circumference Columns 5 and 6 are in line with medical literature, as any effects of from reduction in *trans* fat consumption on body fat composition are smaller and show up more slowly than changes in blood lipid levels.

These results provide some evidence that improved LDL and Total Cholesterol after the policy are not due to some other overall increase in good health. Similarly, these results provide some evidence that the change in the wording of the FAFH question in the survey is not driving the result. This supports the assertion that the changes found in both LDL and total cholesterol levels are, in fact, due to the reduction of *trans* fat in the diet and that these changes are more pronounced in those who frequent restaurants than in those who do not.

2.5.2. Heterogeneity in Treatment Effect

To explore heterogeneity in the treatment effect, I test whether similar or even stronger effects are found among certain groups known to have higher consumption of FAFH, which includes restaurant meals, fast food, pre-packaged, prepared, and frozen foods.

If a particular demographic is strongly affected by this policy because they frequently eat fast-food and pre-packaged foods affected by the regulations, then one would expect to find greater health benefits in these groups. This would indicate that prior to the policy this group has higher cholesterol levels due to the larger proportion of FAFH in their diet and that after the policy, the reduction of *trans* fat in FAFH leads to a lowering of cholesterol levels, controlling for demographics, income, education, and health behaviors. However it could be the case that because secular rates of FAFH consumption are increasing, even among the higher-frequency consumers, expected improvements in health from decreased *trans* fats in these foods may be diminished by increases in FAFH consumption over consumption of home-cooked meals.

The literature on FAFH consumption asks two similar questions about the demographic groups who frequently consume commercially prepared foods and fast foods and finds two different answers. Research on commercially prepared foods finds that Americans aged 45 and under, non-Hispanic whites, individuals with a high school education, and with higher income are groups that report a higher mean frequency of consumption of commercially prepared foods (Kant and Graubard, 2004). To examine the effects of the policy among higher-frequency consumers, Table 2.11 estimates the preferred model for those particular subgroups of the population who eat more FAFH: younger adults, non-Hispanic whites, and those with higher income and higher education. While the magnitudes of the estimate of the effect vary somewhat, that difference is not statistically significant; I do not find a heterogeneity of the effect for these particular subgroups.

Table 2.11: Influence of Restaurant Meal Consumption on LDL among High-Frequency FAFH Subgroups

	(1) White	(2) Under 40	(3) Income 45K+	(4) High School Grad
Policy X Meals	-2.324*	-3.728***	-3.232**	-2.759**
	(1.187)	(1.177)	(1.551)	(1.073)
Meals	0.568	1.438	0.998	1.433
	(0.899)	(0.964)	(1.421)	(0.890)
Constant	97.63***	91.41***	96.14***	96.66***
	(3.353)	(4.357)	(4.542)	(2.657)
Observations	5,696	3,785	4,790	7,870
R-squared	0.123	0.087	0.115	0.118

Data: NHANES 1999-2010

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Note: Results from OLS regression estimating Equation (2.2) for different subgroups. Outcome variable is LDL cholesterol, measured in mg/dL. *Meals* is a categorical variable equal to 0 if the consumer eats at restaurants less than one time per week on average, equal to one if 1-3 meals are eaten per week, equal to 2 if 4-6 meals are eaten per week, and equal to 3 if 7 or more restaurant meals are eaten per week. *PolicyXMeals* is an interaction term between the binary variable *Policy*, which equals 0 for years before 2005 and equals 1 for 2005 and later, and the variable *Meals*. All columns include the full set of control variables used in Column 5of Table 2.5. Column 1 samples only individuals whose race is white, Column 2only those who are under age 40, Column 3only those whose income is 45,000 or above, and Column 4only those who have at least a high school degree.

Similarly, the literature identifies subgroups which consume a higher proportion of calories from fast food in particular. Specifically, Non-Hispanic black adults consume a higher percentage of total meals out from fast food establishments, when compared to non-Hispanic white and Hispanic adults, with the disparity being greater in the 20-39 age group. Additionally, when comparing by weight status, obese adults consume the highest percentage of total calories from fast food (Fryar and Ervin, 2013). For these groups, I expect that the type of restaurant may be more likely to be fast food than in other groups. Thus, I expect the effect on LDL of increasing the number of meals eaten out to be higher in these groups – and consequently the effect of the policy in mitigating the effect of these meals to be higher. Table 2.12 presents estimates for those subgroups who consume a greater portion of their FAFH calories from fast food in particular: males, non-Hispanic blacks, and the obese. Looking at subgroups reduces the sample size and increasing standard errors. In this particular case, the anticipated increase in the effect is not observed among the subgroups.

Table 2.12: Influence of Restaurant Meal Consumption on LDL among High-Percentage Fast Food Subgroups

	(1)	(2)	(3)	(4)
	Male	Black	BMI 25+	Abdominal Obesity
Policy X Meals	-1.746	-3.565	-2.152	-1.144
	(1.602)	(2.376)	(1.339)	(1.477)
Meals	0.0734	3.523*	1.699	1.210
	(1.327)	(1.984)	(1.159)	(1.238)
Constant	103.5***	90.15***	104.9***	105.3***
	(4.097)	(4.762)	(3.311)	(4.003)
Observations	5,308	1,974	7,704	5,959
R-squared	0.131	0.070	0.108	0.115

Data: NHANES 1999-2010

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Note: Results from OLS regression estimating Equation (2.2) for different subgroups. Outcome variable is LDL cholesterol, measured in mg/dL. *Meals* is a categorical variable equal to 0 if the consumer eats at restaurants less than one time per week on average, equal to one if 1-3 meals are eaten per week, equal to 2 if 4-6 meals are eaten per week, and equal to 3 if 7 or more restaurant meals are eaten per week. *PolicyXMeals* is an interaction term between the binary variable *Policy*, which equals 0 for years before 2005 and equals 1 for 2005 and later, and the variable *Meals*. All columns include the full set of control variables used in Column 5of Table 2.5. Column 1 samples only males, Column 2only those whose race is Black, Column 3only those whose BMI is 25 or above, and Column 4only those classified as having abdominal obesity.

An explanation for these findings could be that the groups with a larger percentage of FAFH coming from fast food may be making more unhealthy choices, which counteract or outweigh the improvements in LDL that one would expect to see following the policy change.

These findings suggest an area for potential further work to properly identify consumers of FAFH and compare them to the average population. Potentially, these are the groups which have the most to gain from *trans* fat reducing policies. Yet, if effects from the policy are driven by an income effect and those with the lowest income are not seeing improvements in health, then these could be troubling findings. Further research could explore whether the benefits found here are distributed evenly across groups of higher risk and lower risk and explore whether there are disparities in the policy effect across socioeconomic status.

2.6. Discussion and Conclusions

This research examines the health effects of *trans* fat reduction policies at the population level. This work seeks to identify whether the suite of *trans* fat nutrition labelling laws, city- and state-specific *trans* fat bans, and legal actions against major food manufacturers has improved levels of serum LDL, an intermediate measure of health which is directly affected by *trans* fat consumption, among consumers of food prepared away from home at the population level. Unlike recent research examining the health effects from only New York City *trans* fat ban, this work suggests that due to the extent of reformulations throughout all levels of commercially prepared foods, disentangling the effects of one particular ban may not be feasible.

This work finds an average decrease in LDL by 2.75 mg/dL by consumption of meals away from home after the policy was introduced. This work also found a declining marginal ill-health effect from eating meals away from home. Consistent with the literature on the relationship between *trans* fat consumption and LDL, this work does not find significant effects of the policy on other health outcome variables such as triglycerides, HDL cholesterol, BMI, and waist circumference. A positive relationship between eating meals out and eating frozen or pre-packaged meals is found, suggesting that restaurant meal consumers also consume other types of FAFH which were affected by the *trans* fat reduction policies. These findings also support the idea that the NYC *trans* fat ban in 2006 did have an effect on *trans* fat levels and health measures at the national level, but that these effects did not occur in isolation, but rather were built on top of effects from a number of policies.

Since the start of this research, on June 16, 2015, the US FDA has proposed removal of the GRAS status of partially hydrogenated oils, which will in effect, result in an outright banning of *trans* fats in all restaurant and processed foods within the US (US Food and Drug Administration, 2015). Supporters of this proposal expect that these changes will result in an increase in general health in the population, including but not limited to, reduction in CHD risk and in the number of CHD cases. Even more recently, following implementation of the Affordable Care Act (ACA), the FDA has issued a ruling requiring all restaurants with at least 20 locations to provide a clearly visible calorie count on all menu items by December of 2016 (Goldman, 2015). While this regulation may not directly affect *trans* fat content in restaurant meals, it may be difficult to disentangle the health effects from these closely-timed policies. This research provides a first look at the population-level effects of nation-wide *trans* fat reduction in restaurant and commercially prepared foods and suggests these policies can play a significant role in reducing CHD risk. This work finds both a general trend in the population of decreasing cholesterol levels (LDL and total cholesterol) and finds that individuals who consume restaurant meals regularly show significant and larger decreases in LDL and total cholesterol as compared to individuals who rarely or never eat restaurant meals. This effect becomes more concentrated for higher-frequency consumers of restaurant meals and is robust to demographic, socioeconomic, and health controls.

While this work attempts to control for various health behaviors, further research may be needed in order to disentangle the specific effects of *trans* fat removal from fast food, other restaurant foods, and from pre-packaged foods. Most importantly this work suggests that more care is needed to target the effects in those groups whose consumption of fast food and other FAFH is highest. Three particular concerns, going forward, are identifying the health effects of *trans* fat substitutes, identifying the health effects among the most at-risk groups, and identifying whether improvements in health from these policies have outweighed any negative compensatory health behaviors. Nonetheless, these findings suggest that a voluntary reduction of *trans* fats from restaurant and other commercially prepared foods has had a significant effect on population health and that any proposed policies to further reduce *trans* fats from the food supply could be effective in further lowering CHD risk.

Chapter 3

3.1. Introduction

Commuting to work is a part of people's daily routine and makes up the largest share of annual vehicle miles traveled per household in the US (Santos et al., 2011). Average commuting times have been increasing from just under 22 minutes each way since 1980, when the US Census started collecting data on travel times, and after plateauing in the 2000's are at their highest now, with an average commute time of about 26 minutes each way (Ingraham, 2016; Mckenzie and Rapino, 2011). The majority of this commuting, over 90%, is done through sedentary means, such as driving, being a passenger in a vehicle, or using some form of public transportation.

Sedentary lifestyles have been linked with obesity, increased chronic disease risk, and type II diabetes (Allison DB et al., 1999; Lakdawalla and Philipson, 2009; Manson JE et al., 2004; Rössner, 2002). In the SLOTH framework presented in Chapter 1, individuals can spend their time in sleeping, leisure, occupation, transportation, and home production. With the exception of sleep, these activities can be sedentary or active in nature. Most sedentary behavior occurs during work and leisure time, however commuting by motor vehicle, as a form of transportation is also a contributing factor to sedentary behavior (Cawley, 2004). Sedentary commuting has been linked to measures of poorer subjective health and well-being in a variety of contexts (Roberts et al. 2011; van Ommeren and Gutiérrez-i-Puigarnau 2011; Hansson et al. 2011; Dickerson et al. 2014; Künn-Nelen 2015).

This chapter examines associations between commuting and health, specifically between time that individuals spend engaged in active and sedentary modes of commuting and body mass index (BMI), as an indicator of health status. The following questions are asked: what is the relationship between time spent commuting and BMI? Is more time spent in sedentary commuting associated with an increase in BMI? Is more time spent in active commuting associated with lower BMI? These questions are important because the role of commuting on health is studied in the literature and commuting is often associated with poor health. However, BMI depends on many factors, primarily healthy eating and exercise. This work finds evidence contrary to the commonly found relationship between increased commuting and poor health indicators. Instead, this work finds that more time spent in sedentary commuting is not necessarily associated with increased BMI, leaving room in the lives of sedentary commuters to eat well, exercise, or use whatever means at their disposal to maintain a healthy weight. On the other hand, this work does find that more time in active commuting is related to lower BMI. This may be due not only to the exercise active commuting provides, but also to the fact that people who active commute may be more committed to maintaining a healthy lifestyle in other ways. A key component of this work focuses on reconciling these findings against those presented in a recently published paper which uses similar methods and data but finds a negative relationship between sedentary commuting and BMI. These results suggest that previously found associations between ill health and commuting may be largely explained by the positive health effects of active commuting.

This chapter is structured as follows: Section 3.2 provides a review of literature, Section 3.3 presents data and sample characteristics, Section 3.4 proposes a model of commuting behavior and BMI and Section 3.5 presents estimation results from this model and includes sections examining heterogeneity and robustness of the main model and as well as a section replicating results from a similar model presented recently in the literature and comparing against these results. Section 3.6 concludes with a discussion of these findings, their limitations and implications. Appendix B provides additional comparisons of the Yang and French paper which is discussed in Section 3.5.

3.2. Literature Review

The relationship between commuting and obesity is closely related to the relationship between urban structure, particularly urban sprawl, and obesity. Some papers find that urban sprawl is linked with lower commuting times, as such, the relationship between sprawl and obesity is worth considering when examining the relationship between commuting time and obesity, as a potential mechanism through which sprawl is associated with obesity (Crane and Chatman, 2003; Reid Ewing et al., 2003; Zolnik, 2011). Issues such as endogeneity and reverse causation arise in the sprawl-obesity literature as well as in the commuting-obesity literature, and similar approaches are used in both contexts. For these reasons, before examining the literature on the relationship between commuting and obesity, I present a number of papers examining the relationship between urban structure and obesity.

In particular, Ewing et al. (2003, 2014) look at the relationship between an index of urban sprawl and BMI as well as other indicators of health and health behaviors among adults at the national level. The authors develop an index of sprawl based on four measures of population density and two measures of street block size. They find that the probability of being overweight or obese, and also to a lesser extent being physically active, is significantly associated with the overall urban form of an individual's county of residence. Due to data limitations and missing data, they did not control for income; this may bias their results because average income is associated both with obesity and the built environment. A larger limitation of the work is that they use county-level characteristics of the built environment; other papers have pointed out that walkability of the built environment can vary considerably from one street to the next in a given neighborhood, so a more localized measure of urban structure may better explain the relationship between urban form and individual health. Lopez (2004) develops a measure of urban sprawl from the 2000 US Census data and individual-level data and uses data from the national-level cross-sectional 2000 Behavioral Risk Factor Surveillance System (BRFSS) dataset. The data includes BMI calculated from self-reported height and weight. Controlling for socioeconomic and demographic characteristics, Lopez uses logistic regression to examine the relationship between urban sprawl and likelihood of being overweight or obese and finds an association between increased risk of overweight and increase in urban sprawl.

Plantinga and Bernell (2007) point out that policies based on these earlier studies assume that urban form is exogenous to an individual's weight, and that urban form plays a role in affecting BMI. They consider instead the possibility that BMI affects an individual's choice of residential location and find that causality runs in both directions. They use longitudinal data on a nationally representative sample of 12,686 men and women aged 14 to 22 years from the National Longitudinal Survey of Youth 1979 (NLSY). Participants were interviewed annually through 1994 and biennially since then. The authors use data from the 1996, 1998, and 2000 surveys, when respondents were between the ages of 31 and 43 and they use a county sprawl index developed by McCann and Ewing (2003). They examine a sample of individuals who move residences to test whether an individual's BMI prior to a move affects their choice of a low- or high-sprawl county. In doing this, they assume that the individual's new residential location cannot influence their BMI prior to moving there. From this, they find BMI is a significant factor in determining whether residence choice is in a dense or a sprawling county, with higher BMI individuals choosing less dense locations. They also test a second sample of movers to test whether changes in locational characteristics influence later changes in BMI, as this is a case where one's weight change after moving to a new residence cannot affect the change in residence. From this, they find that individuals who move to denser counties decrease their BMI; additionally, the greater the change in density, the greater the change in BMI. Their work suggests that more work is needed in determining the direction of causation in the relationship between urban form and health. Limitations of their work are that they only consider those individuals who move; there could be endogenous selection in reasons why an individual moves which may also affect other health behaviors that in turn affect BMI; moving often results in or results from these financial or economic changes, for example moving to a new job can mean a change in salary and change in the cost of living – this could affect both choice of residential location and health behaviors affecting BMI. Second, as a result of this focus, they did not observe how urban form might affect BMI of individuals who do not move to a new residence. They also do not control for any sort of regional variation in location which may play a role in individual health behaviors. Eid et al. (2008) also use longitudinal data from the NLSY to estimate models similar to those presented by Ewing et al. (2003). One limitation to the generalizability of their findings is the rather narrow age range in the selected sample; individuals aged 23-36. They suggest that the previous findings failed to properly control for the influence of an individual's health status in their choice of residential location. Their results show no statistically significant associations between urban sprawl and BMI.

Zhao and Kaestner (2010) use changes in population density as a measure of urban sprawl and an instrumental variables approach as an attempt to establish a causal relationship between urban sprawl and obesity through the use of predicted population densities derived from historical Interstate Highway System plans as an instrument for changes in population density. They use individual-level demographic, socioeconomic, self-reported height and weight data on from the National Health Interview Survey (NHIS) from 1976 to 2001. Data on population, highway infrastructure and MSA-level characteristics were obtained from the Neighborhood Changing Database, the General Location of National System of Interstate Highways, and the Current Population Survey (CPS), respectively. First they predict population density from 1947 planned highway rays, showing that more highway plans are significantly associated with a decrease in population density over time, and that planned highway rays are uncorrelated with observed MSA-level time-varying characteristics. They find no effect of population density on BMI, but do find a significant association between population density and obesity.

Testing multiple hypotheses, Wojan and Hamrick (2015) build on this literature by suggesting that the choice of active commuting is not explained by unobserved characteristics that may be the source of a lower BMI. Their findings do not support previous literature which finds associations between sprawl and higher BMI. Instead, they find that compact or dense urban structure is not associated with higher levels of physical activity than occurs in more sprawling urban areas. These findings suggest that urban structure does not indicate preference for physical activity; specifically residents of compact cities are not more likely than residents of sprawling cities to engage in physical activity. They use data from the 2006-08 American Time Use Survey (ATUS) and its Eating and Health Module and sample adults who report working at their workplace on the diary day, 12,405 observations. The authors use the metropolitan statistical area (MSA) murder rate, MSA-level college enrollment rates of 18 to 24 year olds, number of historical sites in an MSA, adverse weather conditions on the diary day, and whether major cities in an MSA had received a Bicycle Friendly Community certification from the League of American Bicyclists to predict whether the active commuting or other unobservable characteristics were responsible for reductions in BMI. The two key assumptions of an instrumental variable model are that the instrument must be correlated with the endogenous explanatory variable, commuting behavior in this case, and that the instrument must not be correlated with the error term in the explanatory model. While these variables are shown to be correlated with active commuting behavior, the authors do not provide a clear discussion of whether these variables would affect other health behaviors related to BMI. In particular, characteristics such as the murder rate, college enrollment rates, number of historical sites, and Bicycle Friendly certification are functions of the economic conditions in a city and thus are likely to be related to health behaviors of its inhabitants. Weather conditions on diary day may be indicative of overall climate conditions, which are regional characteristics and may be capturing some level of regional variation in BMI.

Martin et al. (2014) provide a review of the literature on the relationship between the built environment and obesity in the US. In this review, they examine whether results from papers using more advanced analytic techniques, such as matching, regression adjustment, propensity scores, difference in differences, instrumental variables, and regression discontinuity find weaker associations than presented by studies using single-equation techniques. The authors suggest that this would be expected if reverse causation or endogeneity explained the associations found between obesity and commuting in single-equation studies using cross-sectional data. They find the use of more advanced methods of analysis does not appear to undermine observed strength of association between the urban built environment and obesity.

When considering the relationships between urban form and BMI or other health outcomes, the primary concern is often with a particular causal pathway; perhaps urban sprawl represents a lack of walkability or perhaps urban sprawl allows for shorter sedentary commutes. There might be other pathways between low urban density and high BMI. On the other hand, urban sprawl and high BMI may be co-determined by other unobservable factors. This chapter focuses on one particular pathway, the relationship between individual-level commuting behavior and individual-level health. This is a different approach from the literature focusing on urban sprawl as it explicitly includes time associated with commuting and relies on a time-use framework for understanding tradeoffs between commuting time and time available for other health-producing activities.

Two recent papers examine the relationship between commuting behavior and BMI within the UK context. Flint et al. (2014) use cross-sectional data from the UK Household Longitudinal Study and find people who use either active modes of commuting or public transportation have lower BMI than people using private transport; in fact they find that active commute and public transportation commutes have roughly equal effects, and suggest that these findings may be driven by a socio-economic effect for which commuting mode serves as a proxy. Martin et al. (2015) use panel data from three waves of the British Household Panel Survey (BHPS) to examine how an individual's changes in mode of travel relate to changes in BMI. Controlling for health-related and socioeconomic factors, they find that changing from commuting by car to either an active mode of commute or public transportation was associated with significant reductions in BMI, when compared against no change from car commuting; effect sizes were larger among active commuters and those with longer commutes. Similarly, they find that changing from active modes of commuting to either public transportation or car commuting was associated with increases in BMI. Use of panel data allows the authors to follow individuals over time and possibly reduces the risk of bias which may exist in between-individual comparisons, however other key determinants of BMI, such as other physical activity and dietary behavior were unobserved in this data. Most recently, Künn-Nelen (2015) argues that using an individual fixed effects model with panel data from the BHPS allows her to estimate a causal relationship between long sedentary commute times and lower levels of subjective health. Contrary to evidence found using cross-sectional data, this work finds objective health and health behaviors remain relatively unaffected by longer sedentary commutes. These results are heterogeneous across gender and commuting mode and findings persisted when only considering a subsample of observations with no change in residential location, job or mode of commute. Most notably, women who face longer sedentary commutes are found to have a higher BMI than those with shorter commutes. Car commuting is associated with more negative health effects while commuting via public transportation has no association between commuting time and the health measures tested in the study. This could be due the fact that users of public transportation often engage in some amount of active commuting, walking or biking, in order to reach train or bus stations, while those who commute by car are able to do so with almost no active commuting. While these findings help in understanding the relationship between commuting and health, commuting behavior is different between the UK and the US, most notably, distances tend to be shorter in the UK, options for public transportation fewer and less commonly used in the US, and a smaller percentage of workers commute by biking or walking in the US (Giuliano and Dargay, 2006; Giuliano and Narayan, 2003).

Similarly, US research looks at the relationship between commuting and BMI. Frank et al. (2004) use data from a cross-sectional travel survey of 10,878 adults in the Atlanta, Georgia region to evaluate the relationship between the built environment, self-reported travel patterns (walking and time spent in a car), BMI and obesity. Their analysis is able to control for many individual-level covariates that were unaccounted for in previous research in this area and they consider heterogeneity by ethnicity and sex. They find that each additional hour spent in a car per day is associated with a 6% greater likelihood of obesity and conversely each additional kilometer walked in a day is associated with a 4.8% reduced likelihood of obesity; additionally, they find that higher levels of mixed land use (which in a broad sense, is when a combination of residential, commercial, cultural, institutional, or industrial uses are blended together in one area and these functions are physically and functionally integrated) are associated with a 12.2% reduction in the likelihood of obesity across gender and ethnicity. A key limitation to these findings is that they are restricted to the city of Atlanta and its suburbs; due to regional variation and a limited range of urban forms, their results may not be generalizable at the national level. Nonetheless, their findings suggest that the relationship between urban form and obesity varies across heterogeneous groups in the population and that this may extend to the relationship between commuting behavior and obesity as well.

Lopez-Zetina et al. (2006) provide a preliminary analysis on obesity and vehicle travel. They use data from the California Health Interview Survey, the California Department of Transportation and the US Census to looking at average county-level vehicle miles of travel (VMT) and commuting times and how they relate to county-level obesity status for adults over the age of 18. They find obesity to be related to indicators of increased automobile usage. Similarly, Jacobson et al. (2011) examine the relationship between trends in VMT per licensed driver and adult obesity rates six years later at the national level. They justify the six-year lag by suggesting that it would take time for an effect to emerge at the national level. While these studies each come with a number of limitations, they do address a need for further understanding of the relationship between commuting behaviors and health, particularly overweight and obesity.

Gordon-Larson et al. (2009) use cross-sectional data from 2,364 participants in the Coronary Artery Risk Development in Young Adults study to associations between active commuting and a variety of measured health indicators among adults who worked outside of the home. The strengths of their study include measured health indicators and biomarkers, detailed active commuting data, and additional controls for adiposity and leisure-time physical activity. Limitations of their study are similar to those occurring in this work; data is cross-sectional, which makes it difficult to examine causality, and second self-selection of active commuting means that individuals who are more inclined to choose healthy lifestyles may be more likely to use active forms of transportation. Other limitations of their study are possible misreporting of commuting, physical activity, and other lifestyle factors. 16.7% of their sample self-reported some amount of active commuting; they found associations between improved health indicators among both males and females who active commute. Their findings suggest more work is needed to understand the amount of active commuting needed for positive health benefits – work presented in this chapter attempts to address this issue.

Using data from the 2006 ATUS, Dunton et al. (2009) examine the interaction between time spent in different types of physical activity and sedentary behaviors on BMI. General transportation and commuting are not separately identified and the authors classify sedentary travel into three categories (fewer than 30 minutes, 30 to 79 minutes, and 80 minutes or more), and active transportation into two categories, (zero minutes and one or more minutes). The results are somewhat exploratory and the authors find that only sedentary transportation of 80 minutes or more is associated with higher BMI, and no effect is found from active transportation. Further, they find that spending less time in sedentary transport is associated with lower BMI for specifically those individuals who engage in some active transportation; this leads the authors to conclude that more work is needed to identify the interaction between time spent in active transportation with sedentary behaviors.

Hoehner et al. (2012) find longer sedentary commuting distance to be associated with measured indicators of poor health, including higher BMI, waist circumference, blood pressure, and lower levels of physical activity and cardio-respiratory fitness. They use cross-sectional data from 4,297 adults in metropolitan counties in the state of Texas, from 2000 through 2007; this data includes measured health outcomes and geocoded home and work addresses. Commuting distance is measured as the shortest distance along the road network between individuals' home and work. Control variables include sociodemographic characteristics, smoking, alcohol intake, family history of diabetes and high cholesterol, BMI, and MET-minutes of self-reported physical activity. One limitation of this study is in the measurement of commuting; information about time spent commuting and the validity of the network route to actual distance traveled was not available to the authors. The study population is limited to predominantly white, mostly male, well-educated, healthier adults of middle-to-upper socioeconomic status mostly in the Dallas-Fort Worth-Arlington Metropolitan Statistical Area, a region ranked among the top five most congested metropolitan areas in the US; while this homogeneity may improve internal validity of the study, these findings may not be generalizable at the population level.

Yang and French (2013) add to the literature by framing the commuting-BMI relationship within the context of time-use and energy expenditure; they examine how individual travel and commuting relate to BMI among individuals who report involvement in sedentary travel or commuting, using data from 2006 and 2007 waves of the ATUS and its Eating and Health Module. They find increased levels of commuting to be associated with higher BMI but increased levels of general travel had no association with BMI; this can be explained by the relatively inflexible nature of commuting, in which individuals choice in origin, destination, and time of day of travel are more limited than in other types of travel. This lends support to work showing that commuting can have negative effects on stress and other subjective measures of health and wellbeing. A key limitation of their work is their choice of model, which relies on the percentage of travel (or commute) time spent in a vehicle as a predictor of interest. Interpretation of their results is somewhat unclear because one would expect that percentage of total travel time spent in a vehicle would have different interpretations depending on the amount, in minutes, of total travel. Section 3.5.2 discusses in detail the shortcomings of this model and proposes a simpler and easier-to-interpret variation of their model.

3.3. Data and Summary Statistics

This chapter uses data from the American Time Use Survey (ATUS). ATUS samples individuals who have completed the eighth (final) interview for the Current Population Survey (CPS), which is a nationally-representative monthly household labor force survey administered by the US Census Bureau. Households are selected for the ATUS to ensure that estimates will be nationally representative. From each household, a “designated person” aged 15 or over is randomly chosen for a telephone interview. The designated person is pre-assigned a day of the week about which to report. In the interview, survey respondents are asked to report activities sequentially from 4 a.m. on the day prior to the interview until 4 a.m. on the day of the interview. Respondents are also asked how long each activity lasted. Other than personal care activities, respondents also report how where and with whom they did the activity. The ATUS also includes updated household composition questions from the last CPS interview and employment status of the respondent and the respondent’s spouse or unmarried partner. All primary activities are given a single six-digit code from the ATUS Coding Lexicon (Bureau of Labor Statistics, U.S. Census Bureau, 2015).

Because time-use, or time-diary, data ask respondents to account for all activities in a given time period rather than asking specific questions about a few select activities, data are relatively free of social desirability bias compared with other types of household surveys (Ploeg et al., 2010). Another feature of the ATUS data is that the recall period is restricted to the previous day; this results in reports that are often more accurate than surveys with longer recall periods. On the other hand, a drawback of this short reference period can lead to questions about the interpretation of results, and whether or not an activity in the short run can impact long-run outcomes (Frazis and Stewart, 2012). In particular, can one infer that an activity done on one particular day will be able to impact an individual's BMI? Generally, whether or not a single activity was performed or for how long would not lead to a change in a long-run outcome such as BMI. However, commuting behavior occurs not just on one day but is repeated throughout the weeks and months while an individual is employed. Other than when a person moves the location of their job or residence or changes their work schedule, commuting behavior is generally the same on a day-to-day basis, with little variation, particularly in terms of mode of transportation (Wojan and Hamrick, 2015). In this case, even though time use on an individual day has little effect on BMI, it can be considered a proxy for the long-run average time use for an individual (Frazis and Stewart, 2012). So in effect, this work measures the relationship between what could be considered an individual's average commuting behavior with their BMI. A key limitation of this work to note is that no causal inference can be made in this relationship.

The ATUS specifically asks questions about respondent's height and weight in the Eating and Health Modules, which were administered in 2006, 2007, and 2008. Self-reported height and weight responses were used to calculate BMI, where weight in kilograms is divided by the square of height in meters. BMI measures the mass of the body and is often used to classify individuals into different weight statuses based on commonly accepted cut points. BMI between 18.5 and 25 is classified as normal weight, between 25 and 30 is classified as overweight, and BMI of 30 and above is classified as obese (National Institutes of Health 1998). Figure 3.1 presents a distribution of BMI values, which exhibits some characteristics of a normal distribution and is truncated at 18.5, which is the lower cut point for normal-range BMI.

Figure 3.1: Unweighted Distribution of Body Mass Index

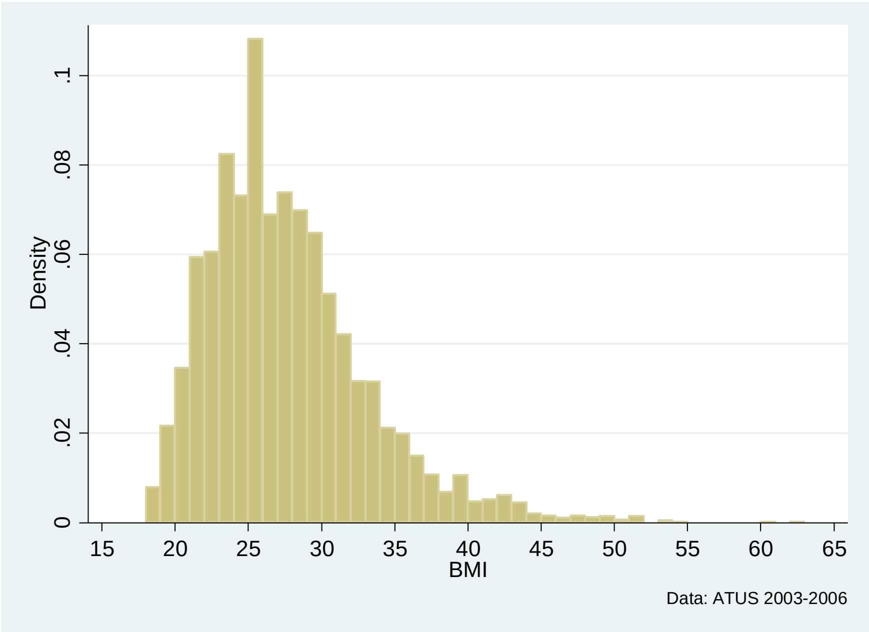
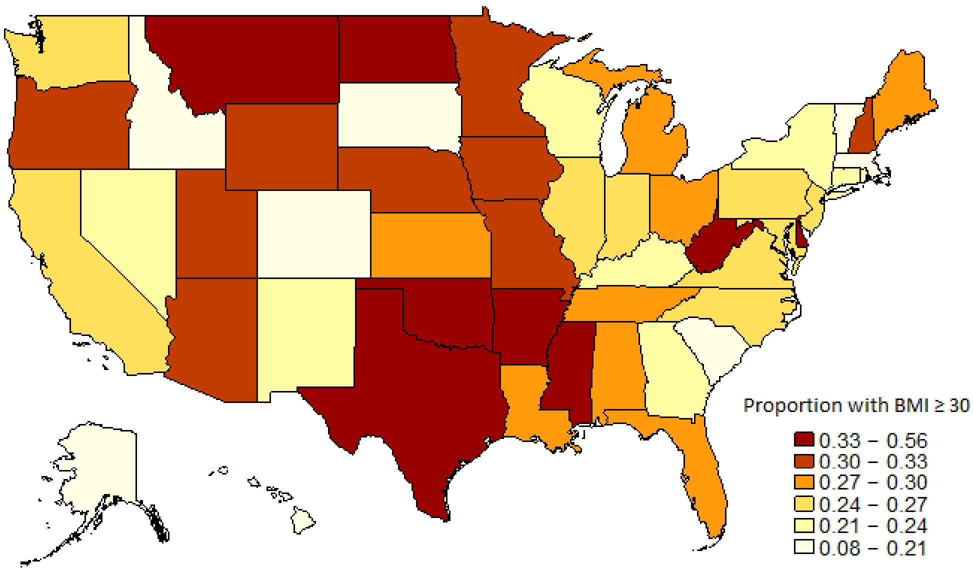


Figure 3.2 presents a map of the unweighted proportion of obese individuals in the selected sample, by state. It is worth noting that this map presents unweighted data so is not representative at the population level, but instead it represents the data available in the selected sample. All analyses are performed with appropriate sample weights to ensure representativeness at the national level.

Figure 3.2: Unweighted Proportion of Individuals with BMI ≥ 30 by State

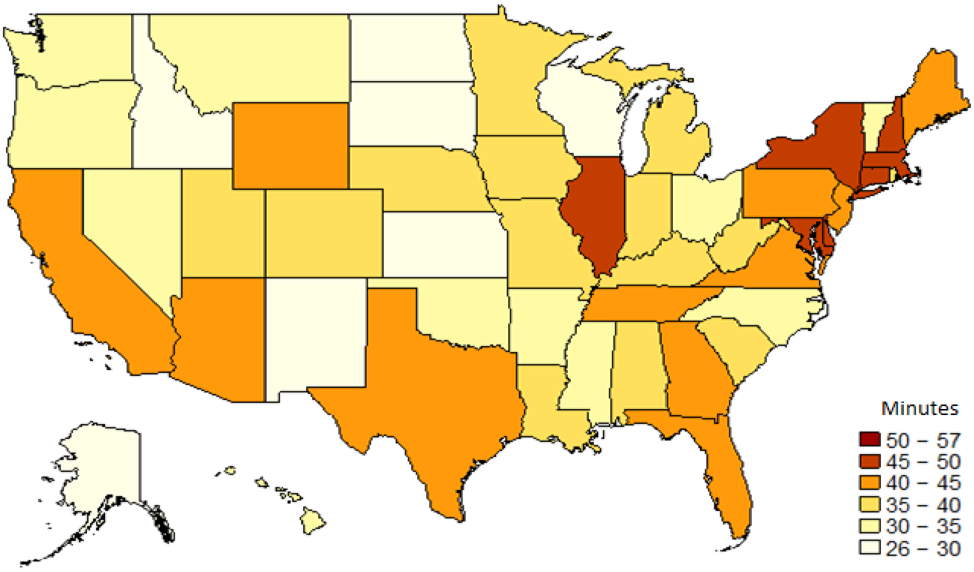


Data: ATUS 2006-2008

Unweighted proportion of individuals with BMI greater than or equal to 30 in the selected sample. Sample is restricted to employed adults age 21 through 65 residing in a metropolitan area, who worked at least seven hours on the diary day, for whom earnings and BMI data are not missing. These unweighted data are not representative of the population at the state or national level.

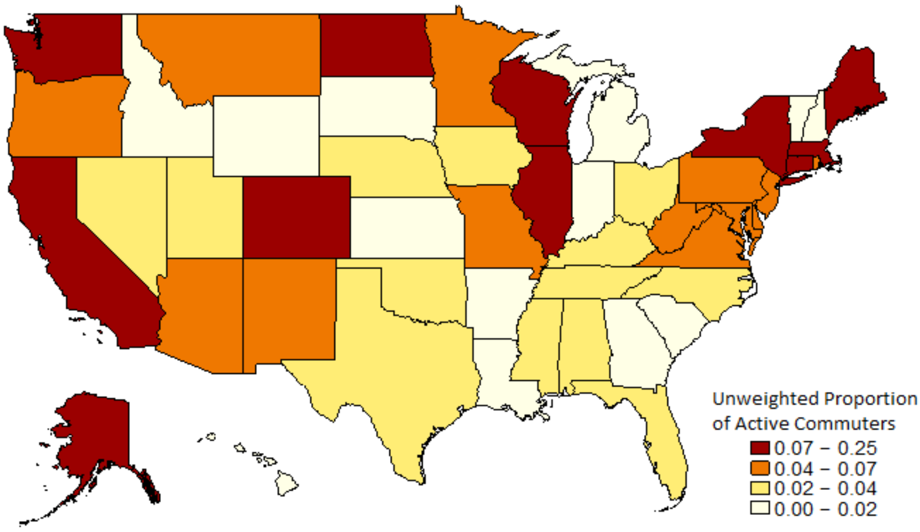
ATUS provides an aggregate of commuting time for each individual, coded as "travel related to work," and defined as strictly any travel occurring immediately before work and any travel occurring immediately after work, provided that the next activity takes place at one's home. This particular method of calculating commute time can underestimate the amount of commuting in situations where an individual does not go home directly after work, but instead makes other trips between work and home. Use of this calculated commute time variable is expected to decrease the size of the relationship found between BMI to a small degree for individuals with those commuting habits, so estimates may in fact be conservative. However, because the same commute time variable is used, results should be consistent with those presented in Yang and French (2013) for comparison. Commuting is categorized as either active, which includes walking or cycling, or sedentary, which is defined as via any other mode of transportation. Figure 3.3 presents unweighted mean commute times among by state, among the selected sample and Figure 3.4 presents an unweighted proportion of individuals who do any active commuting by state.

Figure 3.3: Unweighted mean sedentary commute time by State



Data: ATUS 2006-2008
Unweighted mean sedentary commuting time, in minutes, of selected sample are presented. Sample is restricted to employed adults age 21 through 65 residing in a metropolitan area, who worked at least seven hours on the diary day, for whom earnings and BMI data are not missing. These unweighted data are not representative of the population at the state or national level.

Figure 3.4: Unweighted Proportion of Active Commuters by State



Data: ATUS 2006-2008
Unweighted proportion of individuals in the selected sample who do some active commuting. Sample is restricted to employed adults age 21 through 65 residing in a metropolitan area, who worked at least seven hours on the diary day, for whom earnings and BMI data are not missing. These unweighted data are not representative of the population at the state or national level.

The ATUS has 148,345 respondents in the entire 2003-2013 sample; however this work uses the 2006 through 2008 Eating and Health Module, which has a total of 37,914 observations. Because the focus is on commuting behavior, the sample is restricted to working-age adults, aged 21 through 65 (n = 28,710), who live in urban labor markets, specifically who live in a metropolitan area as defined by the 2000 Census, (n = 23,811). In order to only consider individuals who face a strong time constraint, the sample is limited to individuals who are employed (n = 17,602) and have worked at least seven hours (420 minutes) on the diary day; this brings the sample size down to 7,216 observations. Only individuals who report weekly earnings are included (n = 6,549). For the main analyses, only those individuals with a valid BMI are included (n = 6,190). Because BMI of pregnant individuals have different health implications, pregnant individuals are excluded from the sample. Also excluded are underweight individuals, whose BMI is less than 18.5, as the literature suggests that being underweight might be caused by factors other than time-use, such as genetics and age (Yang and French, 2013). The final sample then is 6,121 observations. Summary statistics for the sample are presented in Table 3.1.

Table 3.1: Sample Characteristics by BMI Category

	(1) All	(2) Normal	(3) Overweight	(4) Obese
Commute	0.92	0.93	0.93	0.92
Sedentary Commute	0.90	0.90	0.91	0.90
Active Commute	0.07	0.09	0.07	0.05
Bike Commute	0.01	0.01	0.01	0.00
Walk Commute	0.07	0.08	0.06	0.05
Any Travel	0.99	0.99	0.99	0.98
Any Sedentary Travel	0.97	0.97	0.98	0.98
Any Active Travel	0.12	0.14	0.12	0.10
Any Bike Travel	0.01	0.01	0.01	0.00
Any Walk Travel	0.12	0.14	0.11	0.10
BMI	27.64	22.47	27.21	34.67
	(0.07)	(0.04)	(0.03)	(0.11)
Age <=25	0.07	0.10	0.06	0.05
Age 26-35	0.23	0.26	0.22	0.23
Age 36-45	0.31	0.31	0.31	0.32
Age 46-55	0.29	0.25	0.30	0.31
Age 56+	0.13	0.11	0.14	0.13
Male	0.55	0.41	0.65	0.57
Spouse or Unmarried Partner in Household	0.59	0.56	0.61	0.59
Spouse is Employed	0.45	0.44	0.46	0.44
Has a Child in Household	0.53	0.53	0.53	0.53
Child Under Age 2 in Household	0.13	0.13	0.13	0.12
White	0.65	0.69	0.65	0.60
Black	0.14	0.09	0.14	0.19
Hispanic	0.16	0.13	0.16	0.17
Asian	0.04	0.08	0.03	0.02
Other Race	0.02	0.01	0.02	0.02
First-generation Immigrant	0.18	0.19	0.19	0.14
No High School	0.07	0.05	0.08	0.09
High School Graduate	0.24	0.20	0.24	0.29
Some College	0.18	0.16	0.18	0.21
College Graduate	0.36	0.40	0.36	0.32
Advanced Degree	0.14	0.19	0.14	0.09
Weekly Income <$400	0.14	0.15	0.12	0.14
Weekly Income $400 - $700	0.28	0.26	0.26	0.32
Weekly Income $700 - $1250	0.33	0.33	0.33	0.34
Occupation with Physical Activity	0.21	0.19	0.22	0.22
Region: Northeast	0.17	0.18	0.17	0.15
Region: Midwest	0.24	0.24	0.24	0.25
Region: South	0.36	0.33	0.37	0.39
Region: West	0.23	0.24	0.23	0.22
Weekend or Holiday	0.17	0.16	0.17	0.19

Table 3.1, Continued

Winter	0.35	0.35	0.34	0.37
Spring	0.17	0.17	0.16	0.17
Summer	0.24	0.21	0.26	0.23
Autumn	0.25	0.26	0.24	0.23
Observations	6,121	2,082	2,361	1,678

Data: ATUS 2006-2008

Note: Unweighted proportions for indicator variables are shown; except for BMI, for which unweighted mean and standard deviation in parenthesis are reported. Statistics are grouped by BMI category.

Table 3.1 presents unweighted summary statistics of the selected sample. This table shows a slightly greater percentage of whites and Asians in the normal-range BMI group than in other BMI groups and a slightly greater percentage of blacks, Hispanics, and other races in the overweight and obese groups. More education, specifically having a college or advanced degree, is observed to be associated with belonging to a lower BMI group. However there is a fairly even dispersion of income levels across the different BMI categories. These data exhibit regional variation in BMI; the Midwest and South have slightly greater proportion of overweight and obese individuals while the Northeast and the West have a relatively larger proportion of normal weight individuals. For the remaining demographic characteristics, there is a fairly even distribution across BMI categories.

Table 3.2 shows the mean of minutes spent in various types of commuting and physical activity, for only those individuals reporting that they took part in each particular activity on the diary day.

Table 3.2: Average Travel and Exercise in Minutes, by BMI Category

	N	(1) All	(2) Normal	(3) Overweight	(4) Obese
All Active Commuting	434	17.56	17.69	17.35	17.68
		(0.97)	(1.33)	(1.78)	(2.20)
Commuting by Bike	32	37.84	30.00	40.87	59.33
		(7.03)	(7.95)	(11.85)	(30.78)
Commuting by Walking	404	15.87	16.61	14.87	16.02
		(0.84)	(1.26)	(1.39)	(1.88)
Sedentary Commuting	5,539	45.11	44.47	45.66	45.12
		(0.49)	(0.82)	(0.80)	(0.93)
Total Commuting	5,654	45.54	45.01	46.03	45.51
		(0.49)	(0.84)	(0.80)	(0.95)
Active Travel	761	19.68	22.18	18.32	17.58
		(0.81)	(1.29)	(1.30)	(1.72)
Sedentary Travel	5,960	76.10	75.03	77.13	75.97
		(0.62)	(1.03)	(1.03)	(1.17)
Total Travel	6,040	77.57	76.82	78.50	77.20
		(0.63)	(1.05)	(1.04)	(1.18)
Leisure Physical Activity	2,228	69.57	68.99	71.04	68.14
		(1.25)	(2.12)	(2.01)	(2.45)

Data: ATUS 2006-2008
Note: Unweighted mean of minutes spent in each activity and standard deviations in parentheses are shown for only those individuals who indicated engaging in the activity on the diary day are shown. Of the 6,121 observations in the sample, N indicates the number of observations which reported participation in each activity on the diary day. Statistics are grouped by BMI category.

This table shows average travel or commute times are much longer in duration in sedentary modes than in active modes. Because travel speed is generally faster in sedentary modes than active, this would suggest that active travel may be used for shorter distance trips than are typically possible with sedentary modes. This may restrict effectiveness of policies to promote active commuting for individuals who reside geographically far from the workplace. These data also show that other than Commuting by Bike, which has a very small sample of individuals who participated in the activity, minutes spent in each of these activities shows homogeneity across BMI categories. This suggests that associations with BMI, or health, may be more a function of participation in a particular activity than the time spent in that activity. The robustness section considers alternative specifications of the preferred model to further explore this issue.

3.4. Methods

This work seeks to examine the relationship between sedentary commuting, active commuting and BMI as an indicator of health. Cawley (2004) presents an economic framework for understanding the contribution of physical activity and other health behaviors to obesity, based on Becker's (1965) model of on choice of allocation of time in non-work activities and Grossman's (1972) extension of this model to investment in health capital.

The implications of this constrained maximization model are that within the framework, individuals may rationally decide to accept a higher body weight in exchange for the utility associated with eating, leisure activities, or time saved in sedentary travel modes. People will exercise when it is the best use of their scarce time (even though public health advocates may encourage them to do it as long as it increases health). People will consume foods within their budgets that provide the highest net benefit. Applying this framework to the context of sedentary and active commuting, gross benefits include immediate pleasure of getting from home to work (or vice versa) quickly plus any current and future health benefit from possibly having more time available for other income-generating or health-producing activities. Gross costs include financial cost, discounted utility of adverse health impacts, and discounted utility of any future weight gain.

One way to understand the choice of active or sedentary commuting is to assume that the different modes of commuting, active and sedentary commuting, are substitute goods, and that marginal benefits and marginal costs of each affect the individual's choice between these two modes. Sallis et al. (1985) find accessibility or proximity to physical activity facilities affects frequency of physical activity; similarly, it is expected that accessibility to active modes of commuting plays role in determining whether an individual would choose active commuting over sedentary. Accessibility to active commuting could be determined by a number of factors including distance between home and workplace, availability of bicycle paths or lanes, walking paths, or sidewalks, perception of safety of neighborhoods between home and workplace, showering facilities and bicycle storage at or near the workplace. On the other hand, for many people, there may be greater accessibility to sedentary commuting, particularly commuting by car. The dispersed urban forms of most American cities and existing road and highway infrastructure encourage automobile dependency (Newman and Kenworthy, 1996). Considering the time cost of active commuting versus sedentary commuting, again due to the built environment sedentary commuting allows individuals to travel at greater speeds than active commuting, particularly for greater distances between home and work, reducing the time cost of sedentary commuting. Similarly, monetary costs may play a role in the commuting mode decision.

Gasoline prices, vehicle maintenance and licensing, and vehicle purchase costs can be significantly greater than costs associated with bicycling and walking. Considering that for the same distance, time costs can be greater while monetary costs are less for active commuting than for sedentary commuting, one could suggest that individuals with a lower income or lower cost of their time may engage in more active commuting than individuals who place a greater value on their time. However, because active commuting essentially combines two actives, commuting and physical activity, it could be the other case that individuals who place a high value on their time may be more drawn to active commuting. This suggests the possibility of the relationship between active commuting and lower BMI being driven by unobservable relative preferences of time vs. money and not by active commuting directly. This also supports controlling for income in analyses of this relationship.

Keeping this framework in mind, this work estimates the following model:

$$(3.1) \quad BMI_{ijt} = \alpha + \beta_1 Sedentary\ Commute_{ijt} + \beta_2 Active\ Commute_{ijt} + \beta_3 X_{ijt} + \gamma_j + \theta_t + \varepsilon_{ijt},$$

where *Sedentary Commute* and *Active Commute* are measured in minutes and an individual may spend time in either, both, or neither activity. The matrix of individual control variables *X* includes age, sex, race, household composition, education, and income. State-level fixed effects, γ_j, are included to capture time-invariant state-level variation in factors such as culture, economic conditions, climate, the built environment, and unobservable factors, all of which may affect BMI. Year-Season fixed effects, θ_t, are also included to account for variation across seasons in factors such as diet and nutrition, exercise behavior, commuting behavior and other factors which may affect BMI but vary by season and by year. Inclusion of these fixed effects accounts for both across-state variation and across-time variation that otherwise may have resulted in omitted variable bias. This allows the results to capture within-state and within-season variation.

Additional analysis in Section 3.5.1 explores issues of heterogeneity by gender and robustness of this model to controlling for other types of physical activity, such as having a physically demanding job and time spent on leisure-time physical activity. Another check of robustness is to examine the relationship with BMI of commuting biking and walking separately, as some research suggests that biking may be more associated with measures of good health than walking.

3.5. Results

Table 3.3 presents the estimation of Equation (3.1), starting with a basic model using no control variables in Column 1. In Column 2, demographic controls are added; these include age, sex, race, presence of children and spouse or unmarried partner in the household, and spouse's employment status. A control for individual's income is added in Column 3, and level of education in Column 4. In Column 5, results from the full model are presented, which includes all controls in the previous model as well as state fixed effects and year-season fixed effects.

Table 3.3: The Relationship between Sedentary Commuting, Active Commuting, and BMI

	(1) Basic Model	(2) Demographic Controls	(3) Income Controls	(4) Education Controls	(5) Preferred Model
Active Commute (minutes)	-0.0374***	-0.0370***	-0.0383***	-0.0332***	-0.0313***
	(0.0106)	(0.0105)	(0.0106)	(0.00996)	(0.00993)
Sedentary Commute (minutes)	0.00345	0.00172	0.00330	0.00324	0.00409
	(0.00307)	(0.00311)	(0.00314)	(0.00311)	(0.00303)
Age <=25		-1.340***	-1.684***	-1.716***	-1.781***
		(0.398)	(0.394)	(0.394)	(0.387)
Age 26-35		-0.648**	-0.722**	-0.661**	-0.692**
		(0.291)	(0.290)	(0.289)	(0.282)
Age 36-45		0.306	0.346	0.388	0.372
		(0.283)	-0.28	(0.277)	(0.274)
Age 46-55		0.246	0.261	0.232	0.190
		(0.259)	(0.257)	(0.254)	(0.251)
Male		1.075***	1.289***	1.105***	1.111***
		(0.186)	(0.189)	(0.193)	(0.187)
Black		2.151***	1.960***	1.938***	2.091***
		(0.293)	(0.292)	(0.289)	(0.291)
Hispanic		1.681***	1.430***	1.189***	1.263***
		(0.303)	(0.308)	(0.305)	(0.317)
Asian		-0.874*	-0.796*	-0.594	-0.540
		(0.468)	(0.463)	(0.465)	(0.464)
Other Race		1.199*	1.209*	1.047	0.935
		(0.659)	(0.638)	(0.655)	(0.634)
First-generation Immigrant		-1.380***	-1.513***	-1.454***	-1.387***
		(0.274)	(0.278)	(0.281)	(0.283)
Spouse/Partner in Household		0.293	0.417	0.393	0.307
		(0.286)	(0.284)	(0.282)	(0.281)
Spouse is Employed		-0.282	-0.283	-0.211	-0.177
		(0.245)	(0.244)	(0.241)	(0.241)
Has a Child in Household		-0.0277	-0.0430	-0.121	-0.114
		(0.203)	(0.203)	(0.201)	(0.202)

Table 3.3, Continued

Child Under Age 2 in HH		0.354	0.358	0.418	0.389
		(0.275)	(0.274)	(0.273)	(0.272)
Log Weekly Earnings			-0.806***	-0.356**	-0.292*
			(0.148)	(0.161)	(0.162)
High School Graduate				-0.186	-0.274
				(0.384)	(0.383)
Some College				-0.182	-0.229
				(0.419)	(0.416)
College Graduate				-1.137***	-1.180***
				(0.386)	(0.384)
Advanced Degree				-2.003***	-1.968***
				(0.420)	(0.416)
Constant	27.53***	26.81***	32.03***	29.89***	29.77***
	(0.148)	(0.290)	(1.011)	(1.077)	(1.293)
State FE					x
Year-Season FE					x
Observations	6,121	6,121	6,121	6,121	6,121
R-squared	0.003	0.050	0.058	0.069	0.092

Data: ATUS 2006-2008

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Note: Results of OLS regression estimating Equation (3.1). Self-reported BMI, measured in kg/m^2, is the outcome variable in each column. *Active Commuting* and *Sedentary Commuting* are reported in minutes spent on the diary day. Column 1 includes no control variables. Column 2 adds demographic control variables, which include age grouping, sex, race categories (Black, Hispanic, Asian, and other non-white races), immigrant/foreign-born status, whether a spouse or partner lives in the household, whether the spouse/partner is employed, whether there is a child in the household, and whether there is a child under age 2 in the household. Column 3 additionally controls for the log of reported weekly earnings. Column 4 additionally controls for education, whether the respondent is a high school graduate, has some college, is a college graduate, or has an advanced degree. Column 5 includes all of these controls as well as state fixed effects and year-season fixed effects. Each column includes controls in previous column.

The coefficient on *Active Commute* in Column 5 can be interpreted as, holding all else constant, a ten-minute increase in active commuting time is associated with 0.3 point lower BMI and is statistically significant. Similarly, the coefficient on *Sedentary Commute* can be interpreted as a ten-minute increase in sedentary commuting time is associated with a 0.04 point higher BMI, but is not statistically significant. Estimated coefficients on the two predictors of interest, minutes spent in active and sedentary commuting, remain robust to addition of control variables. These results indicate that a number of socio-economic and demographic variables play important roles in determining BMI. In particular, age less than 35, being a first-generation immigrant, having higher earnings, and having a college or advanced degree are all significantly associated with lower levels of BMI. Conversely, higher levels of BMI are associated with being male, black, and Hispanic. These coefficients on individual control variables have the expected signs and are consistent with findings in previous literature.[3]

One issue of concern with these findings is that they may be biased upward if the relationship between commuting and BMI is determined by some unobserved factors which are unaccounted for in the model, such as preferences or motivation for healthy behaviors. Considering sedentary commuting in particular, if high-BMI individuals choose residential locations with longer sedentary commutes because these individuals do not have any particular preference or motivation for extra time for exercising or for living close enough for active commuting, then more individuals with higher BMI will have longer sedentary commute times. This would result in an overestimate of β_1, the coefficient on sedentary commute time. How unobserved factors would affect the estimate of β_2, the coefficient on active commuting time, is less clear. If the same assumption that healthy, low-BMI, individuals who choose residential location based on their preference for having a short commute to allow for active commuting or to allow more time for other healthy behaviors, then then the predicted value of β_2 would be an overestimate of the relationship between active commuting and BMI. On the other hand, if only very health-conscious, and possibly low-BMI, individuals are involved in longer active commutes, and more high-BMI individuals who active commute have shorter active commute times, then the estimate of β_2 could be underestimated. It is likely that the relationship between sedentary commute and BMI may be overestimated, but it is unclear in which direction the relationship between active commute and BMI is biased. More research is needed in disentangling the effects and drivers of active commuting.

3.5.1 Heterogeneity and Robustness of Results

Both BMI and commuting behavior exhibit heterogeneity by sex; to account for this, Equation (3.1) is estimated in Table 3.4 for males and females separately. Females are found to have a higher BMI on average. Between males and females, the sign on the point estimates on the sedentary commute variable are different. However, using a Wald chi-squared test to test the difference between coefficients in these two estimates finds that these differences are not statistically significant and they are very small, particularly when compared to the point estimates on active commute. Results indicate

[3] Data includes observations from 263 counties. This result is robust to using county-level fixed effects along with Year-Season fixed effects. In this case, the coefficient on Active Commute is -0.0268 with a standard error of 0.0130. The coefficient on Sedentary Commute is 0.00312 with a standard error of 0.00473.

no statistically significant difference in the relationship between sedentary and active commuting variables and BMI between the two groups.

Table 3.4: Relationship of Active and Sedentary Commuting and BMI, for Males and Females

	(1) Preferred Model	(2) Males	(3) Females
Active Commute (minutes)	-0.0313***	-0.0317**	-0.0260*
	(0.00993)	(0.0127)	(0.0147)
Sedentary Commute (minutes)	0.00409	0.00490*	0.00115
	(0.00303)	(0.00290)	(0.00647)
Constant	29.77***	30.01***	31.43***
	(1.293)	(1.805)	(1.923)
Observations	6,121	3,351	2,770
R-squared	0.092	0.073	0.151

Data: ATUS 2006-2008

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Note: Results of OLS regression estimating Equation (3.1). Outcome variable is self-reported BMI, measured in kg/m^2. Column 1 presents the preferred model for both males and females, as originally presented in Column 5 of Table 3.3. Column 2 presents estimates of this model for males only and Column 3for females only. All columns include the full set of control variables used in Column 5 of Table 3.3.

Table 3.5 compares the estimate of the preferred model to an alternate specification where biking and walking to work are considered separately, and to a specification where additional controls are added to account for time spent in leisure physical activity and for having a physically-demanding occupation, as specified in Equation (3.2):

$$(3.2) \quad BMI_{ijt} = \alpha + \beta_1 Sedentary\ Commute_{ijt} + \beta_2 Active\ Commute_{ijt} + \beta_3 Leisure\ Physical\ Activity_{ijt} + \beta_4 Physical\ Occupation_{ijt} + \beta_5 X_{ijt} + \gamma_j + \theta_t + \varepsilon_{ijt}.$$

Because cycling generally burns more calories per minute than walking, and the average cycling commute is longer than the average walking commute, this work also examines the relationship with BMI of commuting biking and walking separately. Using data from the UK, Foley et al. (2015) investigated whether participation in active travel is associated with compensatory decreases in other types of physical activity. Contrary to the idea that active commuting and leisure-time physical activity are substitute goods, they found that commuting by walking in particular was associated with increased levels of total physical activity; however, commuting by cycling was associated with small decreases in leisure-time physical activity. Sahlqvist et al. (2012) report similar findings when looking at the relationship between recreational physical activity and active modes of commuting and non-commuting travel in the UK:

$$(3.3) \quad BMI_i = \alpha + \beta_1 Sedentary\ Commute_i + \beta_2 Bike\ Commute_i + \beta_3 Walk\ Commute + \beta_4 X_i + \gamma_j + \theta_t + \varepsilon_i.$$

Column 2 of Table 3.5 shows associations with lower BMI for both cycling to work and walking to work, suggesting that the relationship between lower BMI and involvement in active commuting is not driven solely by mode of active commuting. These effects remain robust to additional controls for time spent in leisure-time physical activity and for having a physically demanding job, shown in Column 3. While cycling burns more calories per minute than walking, the similarity in coefficients on these two activities may be explained by the findings of Foley et al. (2015) that those who walk also engage in other forms of physical activity.

Table 3.5: Relationship between Modes of Active Commute, Other Physical Activity, and BMI

	(1) Preferred Model	(2) Bike & Walk	(3) Other PA Controls	(4) Sedentary Omitted
Active Commute (minutes)	-0.0313***			-0.0315***
	(0.00993)			(0.00997)
Sedentary Commute (mins.)	0.00409	0.00402	0.00384	
	(0.00303)	(0.00305)	(0.00305)	
Bike Commute (minutes)		-0.0389***	-0.0379***	
		(0.0128)	(0.0125)	
Walk Commute (minutes)		-0.0279**	-0.0285**	
		(0.0128)	(0.0128)	
Time spent in Physical Activity			-0.00389**	
			(0.00161)	
Occupation with Physical Activity			-0.497**	
			(0.229)	
Constant	29.77***	29.78***	30.30***	29.71***
	(1.293)	(1.293)	(1.307)	(1.291)
Observations	6,121	6,121	6,121	6,121
R-squared	0.092	0.092	0.094	0.091

Data: ATUS 2006-2008

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Note: Outcome variable is self-reported BMI, measured in kg/m^2. Column 1 presents the preferred model from Column 5 of Table 3.3. Columns 2 and 3 estimate Equation (3.3), where coefficients on biking and walking commuting are estimated separately. Columns 1, 2, and 4 include the full set of control variables used in Column 5 of Table 3.3 and Column 3 additionally controls for reported minutes spent in non-commuting physical activity and having an occupation with high amounts of physical activity. Column 4 estimates the preferred model, but omits *Sedentary Commuting*.

Time spent in active commuting is expected to have diminishing marginal returns to lowering BMI; walking for five additional minutes should have a greater effect if a person does no walking than if a person has walked for an hour. To test this, a squared term is added to the estimation, as shown in Equation (3.4), which is estimated in Column 2 of Table 3.6:

$$(3.4) \quad BMI_{ijt} = \alpha + \beta_1 Sedentary\ Commute_{ijt} + \beta_2 Sedentary\ Commute_{ijt}^2 + \beta_3 Active\ Commute_{ijt} + \beta_4 Active\ Commute_{ijt}^2 + \beta_5 X_{ijt} + \gamma_j + \theta_t + \varepsilon_{ijt}.$$

Equation (3.4) attempts to identify whether the main result is driven by the decision to do active commuting itself rather than by the time an individual regularly spends in active commuting. One concern is that active commuting is a proxy for other healthy behaviors, that by identifying a correlation between lower BMI and active commuting, what is actually being identified is individuals who are making other healthy choices, one of which is active commuting, which are responsible for lowered BMI. To examine this issue, an indicator variable for whether or not an individual does any sedentary commuting or does any active commuting is used to identify the decision to active commute. I use two separate specifications, in Columns 3 and 4 of Table 3.6. Column 3 estimates:

$$(3.5) \qquad BMI_{ijt} = \alpha + \beta_1 Sedentary\ Indicator_{ijt} + \beta_2 Active\ Indicator_{ijt} + \beta_3 X_{ijt} + \gamma_j + \theta_t + \varepsilon_{ijt}.$$

And Column 4 estimates:

$$(3.6) \qquad BMI_{ijt} = \alpha + \beta_1 Sedentary\ Commute_{ijt} + \beta_2 Sedentary\ Indicator_{ijt} + \beta_3 Active\ Commute_{ijt} + \beta_4 Active\ Indicator_{ijt} + \beta_5 X_{ijt} + \gamma_j + \theta_t + \varepsilon_{ijt}.$$

In both of these specifications, the omitted group is a group that does not report any type of commuting on the diary day. These individuals worked from home or telecommuted and make up roughly 8% of the sample. In Equation (3.6), both the time and indicator variables equal zero when zero minutes are spent in that activity; likewise the indicator equals one when the time variable is non-zero.

Table 3.6: Variations in measuring Time Spent in Active and Sedentary Commuting

	(1)	(2)	(3)	(4)
	Preferred Model	Time Squared Model	Time Indicator Model	Time Indicator + Minutes Model
Sedentary Commute (mins.)	0.00409	-0.00684		0.00571*
	(0.00303)	(0.00686)		(0.00334)
Sedentary Commute squared		0.000073		
		(0.000052)		
Active Commute (minutes)	-0.0313***	-0.0715***		-0.0151
	(0.00993)	(0.0214)		(0.0132)
Active Commute squared		0.000519**		
		(0.000229)		
Sedentary Commute Indicator			-0.206	-0.501
			(0.277)	(0.314)
Active Commute Indicator			-1.091***	-0.853*
			(0.367)	(0.489)
Constant	29.77***	29.91***	29.85***	30.20***
	(1.293)	(1.303)	(1.301)	(1.329)
Observations	6,121	6,121	6,121	6,121
R-squared	0.092	0.094	0.092	0.093

Data: ATUS 2006-2008

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Note: Outcome variable is self-reported BMI, measured in kg/m^2. Column 1 presents the preferred model from Column 5 of Table 3.3. Column 2 estimates the Equation (3.4), Column 3 estimates Equation (3.5) and Column 4 estimates Equation (3.6). Note that the omitted group in Columns 3 and 4 are individuals who worked but did not commute on the diary day. All columns include the full set of control variables used in Column 5 of Table 3.3.

Column 2 of Table 3.6 is consistent with diminishing marginal returns to active commuting, which is not a trivial result. This contradicts a suggestion that a person with a twenty-minute-per-day active commute is only slightly healthier than a person with no active commute, but that a person with a forty-minute-per-day active commute is much healthier than the person with a twenty-minute-per-day active commute. In other words, people who have very long active commutes and very healthy BMI are not driving the main result. Columns 3 and 4 can be interpreted two ways: because the amount of time spent on active commuting is not as strongly related to BMI in the presence of an indicator representing participation in active commuting, this suggests that active commuting may be an indicator of healthy lifestyles more broadly. In this interpretation, the number of minutes spent active commuting is not contributing greatly to lower BMI. However, an indicator variable can also play a role similar to a function illustrating diminishing marginal returns.

3.5.2. Discussion of Yang and French Paper

The relationship between sedentary commuting and BMI found in this work are in contrast to the results of Yang and French, who found that automobile-based travel, particularly commuting in a motorized vehicle, is significantly associated with increased BMI. This section reconciles results from this chapter against those put forth by Yang and French (2013), using the sample selection criteria in their paper. In particular, this section discusses how the preferred model presented here is an improvement on the model presented in Yang and French, and allows for disentangling the association between active commuting and BMI specifically. Results from this work are compared to Yang and French because it is the most recent research on the relationship between travel/commuting and obesity using US time-use data. Additional details on the comparison to sample selection criteria from the Yang and French paper are presented in Appendix B.

There are three fundamental differences between the approach to measuring the relationship between BMI and commuting behavior presented here and that presented by Yang and French. The first is the sample selection used. Yang and French include in their sample only those individuals for whom commute time (or travel time) is not zero and not missing and those individuals for whom BMI is greater than or equal to 25 and is not missing. So, they exclude all but those who commute (or travel) and the overweight and obese. The specification that they use, percentage of commute in a vehicle, restricts their sample to those who do commute, by definition a percentage of zero is undefined. The sample selection presented in this work does include individuals who do not commute but restricts the sample to employed individuals who worked at least seven hours on the diary day, in order to fully capture variation in commuting behavior. Because commuting time in metropolitan and rural areas has a different meaning in terms of commuting distance, the sample is also limited to individuals living in metropolitan areas. The sample is also restricted to those individuals who provide weekly earnings data. Normal weight individuals are included in the sample to fully capture the variation in BMI, however Yang and French exclude this group, on the argument that they find that determinants of BMI in their data are different for the normal weight and underweight groups than for the overweight and obese. Specifically, they claim that for these groups BMI is determined not by health behaviors but more so by factors such as genetics, illness, and age. This is supported in the medical literature for the underweight, but not for those in the normal-, overweight, and obese categories. Yang and French do not exclude any individuals on the basis of age however, because this work examines commuting and health, individuals under the age of 20 and over 65 are excluded. In these groups, age, illness, and genetics play a larger role in determination of BMI than the in-between age group. They use data for only the 2006 and 2007 Eating and Health Modules, while this work uses all three years, 2006-08 for the sample. In Table 3.7, the original results published in Yang and French (2013) are copied to Column 1. This chapter approximates their model and sample selection and replicates their results in Column 2. And in Column 3, their model is estimated using the preferred sample selection in this chapter with the added restriction of including only the overweight and obese; in Column 4, the Yang and French model is estimated using the preferred sample selection identified in this work, which includes normal weight, overweight, and obese individuals.

Table 3.7: Replication of Yang and French Results with Variations in Sample Selection

	(1)	(2)	(3)	(4)
	Y&F original	Y&F replication	Preferred Sample, Overweight	Preferred Sample
% of Commute in Vehicle	1.35***	1.363***	1.012*	2.122***
	(0.43)	(0.441)	(0.544)	(0.509)
Total Commute (minutes)	0.00***	0.00261	0.00740**	0.00238
	(0.00)	(0.00260)	(0.00367)	(0.00332)
Constant	26.95***	28.41***	28.10***	23.71***
	(0.73)	(0.629)	(0.737)	(0.663)
Observations	4,625	4,688	3,725	5,654
R-squared	0.031	0.033	0.040	0.055

Data: ATUS 2006-2008

Column 1: Standard errors from published t-statistics. Columns 2-4: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Note: Outcome variable is self-reported BMI, measured in kg/m^2. All columns control for age, sex, race, spouse or unmarried partner in household, household children, education, income, metropolitan status. Column 1 shows published coefficients from Yang & French (2013). Column 2, presents results from replication of the Yang & French model using their sample selection. Column 3 shows a replication of their model using my preferred sample selection criteria and restricting the sample to only overweight and obese individuals. Column 4 presents a replication of the Yang & French model using my preferred sample selection, which includes normal weight, overweight and obese individuals. We use a Wald chi-squared test to test the difference between on the coefficient on percentage of commuting time in a vehicle between the three columns. Comparing Columns 2 and 3 gives a p-value of 0.24, comparing Columns 2 and 4 gives a p-value of 0.24 and comparing Columns 2 and 4 gives a p-value of 0.007.

From this table, the biggest difference between the preferred sample, and that presented in Yang and French occurs when normal weight individuals are included. The results suggest that increasing the variation in BMI increases the estimated coefficient on percentage of commuting time in a vehicle. A Wald chi-squared test is used to test the difference between the coefficient on percentage of commuting time in a vehicle between the three columns. Comparing this coefficient from Columns 2 and 3 gives a p-value of 0.24, and comparing Columns 2 and 4 give a p-value of 0.21, indicating that these differences are not statistically significant. Comparing this coefficient from Columns 3 and 4 gives a p-value of 0.006, indicating that when the obese and overweight groups are examined separately from the normal weight group the difference in the relationship between percentage of commuting in a vehicle and BMI is statistically significant.

The second, and a more minor, source of difference between the two papers are the choice of control variables included. In their paper, Yang and French use dummy variables to control for age groupings, sex, race, married or living with partner, whether there is a child in the household, educational status, weekly income category, and metropolitan status. In the preferred model presented here, these control variables are used as well, but the following additional controls are added to the model: whether there is a child under age two in the household, whether the spouse/partner is employed, and immigrant status; all of these factors are also associated with variation in BMI. State-level and seasonal fixed effects are also included to account for seasonal and regional variation in both BMI and commuting habits. These differences suggest that to compare results presented here to Yang and French, sample selection cannot be ignored. However, when using a close approximation of their model and sample selection, results are very similar to theirs.

The third source of difference between the two papers is the model specification itself. Yang and French present the following model, where BMI is determined by time spent commuting and the percentage of the commute which is sedentary:

$$(3.7) \qquad BMI_i = \alpha + \beta_1 Total\ Commute_i + \beta_2\ \%\ Sedentary\ Commute_i + \beta_3 X_i + \varepsilon_i.$$

The coefficient β_1 estimates the effect of minutes spent commuting on BMI and β_2 measures the role that sedentary commuting plays in BMI by estimating the effect of the percentage of total commuting time that is spent in any type of vehicle.

There are two concerns about this specification. First, using this specification suggests that the relationship between percentage of commute time in a sedentary mode of transportation and BMI does remains constant regardless of total amount of time an individual is commuting. In other words, this specification suggests that, holding all else constant, β_2 would be the same for an individual who has a 10 minute commute with 80% sedentary as for another individual who has a 100 minute commute where 80% is sedentary. This raises questions about the interpretation of the coefficient β_2. One would expect that β_2 would vary in some way as total commute time varies. A second issue is that percentage sedentary commuting and percentage active commuting are mutually exclusive, so estimating the effect of percentage sedentary commuting in this way creates an identification problem. Time spent in the total commute is made up of time spent in any type of vehicle, which is referred to here as sedentary commuting, and time spent walking or cycling, which is referred to as active commuting. This means that for each individual, the percentage of commuting time in a vehicle is equivalent to one minus the percentage of commute that is spent walking or cycling. Mathematically, the Yang and French model can be re-written as:

$$(3.8) \quad BMI_i = (\alpha + \beta_2) + \beta_1 Total\ Commute_i + (-\beta_2)(1 - \%\ Sedentary\ Commute_i) + \beta_3 X_i + +\varepsilon_i.$$

Which would, in effect be the same as:

$$(3.9) \quad BMI_i = (\alpha + \beta_2) + \beta_1 Total\ Commute_i + (-\beta_2)(\%Active\ Commute_i) + \beta_3 X_i + +\varepsilon_i,$$

where $(\alpha + \beta_2)$ serves as the constant and $(-\beta_2)$ estimates the relationship between percentage active commuting and BMI. This shows that using the Yang and French specification, one is unable to discern between the positive association between percentage sedentary commute and BMI, (β_2) from Equation (3.8), and the negative association between percentage active commute and BMI, $(-\beta_2)$ from Equation (3.9).

In the preferred model presented in this work, both the time spent in active commuting and time spent in sedentary commuting are considered, in order to provide a straightforward interpretation of coefficients of interest and to disentangle the effect of sedentary commuting from active commuting. Appendix B presents alternate specifications that were considered, as ways to improve upon the issues that were found with the Yang and French model.

Table 3.8 shows estimates of the Yang and French preferred model, Equation (3.7), in Column 1; note that these estimates were shown previously in Table 3.7, Column 2. Column 2 of Table 3.8 estimates the Yang and French model using their sample selection and the preferred set of control variables in this chapter. Column 3 estimates the preferred model in this chapter, Equation (3.1), using the Yang and French sample. Finally, Column 4 estimates the preferred model in this chapter, using the preferred set of control variables and sample in Column 4; note that Column 4 estimates were shown previously in Table 3.3, Column 5.

Table 3.8: Discussion of Results from Yang and French (2013)

	(1)	(2)	(3)	(4)
	Model: Y&F Controls: Y&F Sample: Y&F	Model: Y&F Controls: Pref. Sample: Y&F	Model: Pref. Controls: Pref. Sample: Y&F	Model: Pref. Controls: Pref. Sample: Pref.
% of Commute in Vehicle	1.363***	1.512***		
	(0.441)	(0.483)		
Total Commute (minutes)	0.00261	0.00403		
	(0.00260)	(0.00307)		
Active Commute (minutes)			-0.0372***	-0.0313***
			(0.0142)	(0.00993)
Sedentary Commute (minutes)			0.00497	0.00409
			(0.00315)	(0.00303)
Constant	28.41***	30.81***	32.26***	29.77***
	(0.629)	(1.046)	(0.976)	(1.293)
Observations	4,688	4,129	4,129	6,121
R-squared	0.033	0.044	0.044	0.092

Data: ATUS 2006-2008

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Note: Outcome variable is self-reported BMI, measured in kg/m^2. Column 1 presents a replication of the Yang and French (2013) model, using their control variables and sample selection criteria. Column 2 estimates the Yang and French model with the full set of control variables used in Column 5 of Table 3.3 and the Yang and French sample selection. Column 3 estimates Equation (3.1) using the full set of control variables used in Column 5 of Table 3.3 and the Yang and French sample selection. Column 4 estimates the preferred model, Equation (3.1), with the preferred sample and full set of control variables.

These results are shown side by side for the purposes of comparison of the interpretation of the coefficients of interest. Yang and French discuss a model very similar what is estimated in Columns 1 and 2 (which are repeated here from Table 3.7 for clarity). Their interpretation of the significant, positive coefficient on the percentage of minutes of commute in a vehicle is that increasing an individual's sedentary commute increases their BMI. However, as shown above, a coefficient of 1.4 or 1.5 in Equation (3.7) is equivalent to a coefficient of -1.4 or -1.5 in Equation (3.8) and Equation (3.9). Thus, it is not clear if this coefficient implies that an increase in an individual's sedentary commute increases their BMI or if an increase in an individual's active commute decreases their BMI. Columns 3 and 4 of Table 3.8 are based on Equation (3.1); Column 3 using the replication of Yang and French's sample and Column 4 using the preferred sample (Column 4 is repeated from Column 5 of Table 3.3). This illustrates the point that while Yang and French do find commuting has a significant association with BMI, a specification similar to Equation (3.7) makes it unclear what that role is. The specification from Equation (3.1) makes it clear that the role of commuting in BMI is likely dominated by active commuting.

3.6. Discussion and Conclusions

This study presents new estimates on the relationship between both active and sedentary commuting behavior and BMI. Results support earlier research that active commuting is associated with better health, where lower BMI indicates better health. However, unlike previous research, this work does not find associations between sedentary commuting and higher BMI. Instead these findings suggest that previous associations between increased sedentary commuting and high BMI were likely driven by active commuting being strongly associated with lower levels of BMI. Because this work finds that active commuting is, in fact, associated with higher levels of leisure-time physical activity, this would suggest that active commuting itself plays a smaller role in BMI, and instead may be an indicator of greater preference for good health.

A key limitation in this work, like much of the literature in this area, is endogenous selection in sedentary and active commuting behavior and health. First, the issue of reverse causality arises; it is not possible to disentangle whether having a lower BMI pushes individuals to choose active modes of commuting or whether walking or cycling to work itself is resulting in healthier body weight. In fact, BMI is driven by a variety of factors related to both genetics and health behaviors; active commuting can be a part of those health behaviors contributing to healthier BMI, but at the same time having a healthy BMI may make active commuting a more comfortable or accessible option for individuals. I would expect that reverse causality would bias these results toward zero. Second, active commuting and BMI may be co-determined by some unobserved factors that cannot be accounted for with this data, resulting in omitted variable bias. For instance, if individuals have an increased preference for good health or have a high level of motivation, this may affect whether they choose active modes of commuting, but this also affects whether they make other healthy choices related to diet and exercise; this in turn may also affect their health outcomes, so that they may be more likely have lower BMI as a result of their preference for healthy behaviors. While I do control for physical activity in the robustness section, which includes exercise, if other healthy behaviors which are correlated with active commuting and are not captured by this variable, I would expect that omitted variable bias would also bias these results toward zero. As an attempt to explore these issues of endogeneity, an instrumental variables approach is presented in the following chapter.

Further work is needed to disentangle the relationships between using public transportation versus commuting in a private vehicle and health. Because users of public transportation often have a component of active commuting in their journey, and this combination may be more accessible to individuals who live too far from work to make an entirely active commute feasible, more work is needed to understand the role of active commuting for those individuals in particular. Understanding whether a small amount of active commuting, in conjunction with a sedentary mode of commuting, has the same relationships with health outcomes and health behaviors as entirely active commuting journeys may direct policies to encourage the most health-promoting types of commuting.

These results raise questions about the effectiveness of policies targeted at reducing obesity through reducing sedentary commuting. While long sedentary commutes are associated with lower productivity and lower levels of well-being, this work suggests that there may be other factors determining the relationship between long sedentary commutes and increased BMI. This work shows that even small amounts of active commuting are associated with lower BMI, and active commuting is associated with lower BMI even after controlling for other types of physical activity. However, there is some evidence of omitted variable bias, that active commuting may be an indicator of other health choices not captured in this survey, as the decision to active commute may be more important than the amount of time spent active commuting. For this reason, further research in the behaviors of active commuters may uncover other healthy habits that drive these results or whether participation in active commuting itself influences involvement in other health-seeking behaviors which result in lower BMI. Similarly, this work supports further investigation into the determinants of active commuting and identifying the causal relationships between active commuting and good health.

Chapter 4

4.1. Introduction

Physical inactivity and sedentary lifestyles have been identified as the sixth leading cause of disease burden in the US and are associated with an increased risk for chronic diseases (Institute for Health Metrics and Evaluation (IHME), 2013; Lee et al., 2012). Numerous studies in the US, Canada, and the UK find increased usage of healthcare services and increased direct medical costs among physically inactive individuals, as compared to active individuals (Katzmarzyk and Janssen, 2004; Pratt et al., 2000; Sari, 2009). Regular exercise, even small increases in activity for inactive individuals, reduces the risk of chronic disease, for which the obese are at greater risk (Sattelmair et al., 2011; Wen et al., 2011). In an effort to encourage individuals to fit exercise into their schedules, the US Centers for Disease Control and Prevention emphasize that any 10-minute episode of physical activity can count towards reaching the goals outlined in their 2008 Physical Activity Guidelines, which serve as the current recommendations for adult physical activity. Adults between the ages of 18 and 65 are recommended to engage in moderate-intensity activity for 150 minutes per week or vigorous-intensity activity for 75 minutes per week. Nonetheless, fewer than half of US adults meet these guidelines and just over one quarter of adults report spending no time at all on physical activity (USDHHS, 2008). More recent research using 22 years of data from the National Health and Nutrition Examination Survey finds that a lack of exercise and physical activity is more strongly correlated with the rise in obesity in the US population than an increase in caloric intake (Ladabaum et al., 2014). A subset of research has demonstrated both correlational and causal links between urban sprawl and obesity (R. Ewing et al., 2003; Lindström, 2008; Lopez, 2004; Sallis et al., 2012).

This chapter builds upon the existing literature, which finds a negative relationship between physical activity and commuting, by focusing on how this relationship varies across individuals with differing health status and by exploring the causal nature of the relationship. In particular, this work focuses on the variation by body mass index (BMI) categories and asks the question “Is the trade-off between physical activity participation and commuting time different for obese individuals than for individuals in the healthy- and over-weight categories, who have a lower chronic disease risk?” This work examines this relationship on both the extensive and intensive margins; both the likelihood that an individual will exercise and the change in time spent exercising as commuting time varies. To address the issue of endogenous selection of commuting time, this work presents an instrumental variables approach from the urban economics and labor economics literature (Baum-Snow, 2007; Giménez and Molina, 2011).

These results indicate a negative relationship between commuting and physical activity participation, which is significant for both men and women. However when examining this relationship among obese individuals separately, a particular subgroup with greater chronic disease risk, no significant relationship between commuting and physical activity participation is found. Amongst physical activity participants, again no significant relationship at the intensive margin is found. These results are robust to a variety of definitions of both commuting and physical activity. Using an instrumental variables approach, with past housing prices as an instrument, this work provides evidence that the results are not simply correlative but represent a potentially causal path between time saved and exercise participation that is not present in at-risk weight groups. This work finds a similar pattern of a negative relationship between commuting and physical activity participation for non-obese males and no significant relationship for obese males in the sample. Housing prices are not a very good predictor of women's commuting behavior; this might reflect differences in determinants of residential and work locations for men and women. Research into differing determinants includes the "tied mover" and "overeducation" hypotheses, originating in work from Mincer (1978) and Frank (1978) respectively, and more recently reviewed by Leuven and Oosterbeek (2011).

These findings indicate that policies aimed at increasing physical activity for obese individuals who do not already engage in physical activity may need to be different than policies aimed at increasing physical activity time for individuals with lower BMI or for individuals who already participate in some physical activity. In particular, these policies may need to provide some additional incentive rather than solely focusing on increasing access or providing more time for physical activity. Further, these results suggest that even increasing physical activity for those who already participate may not be a simple matter.

4.2. Literature Review

Previous medical research establishes a relationship between time spent commuting and a number of adverse health indicators and outcomes, particularly obesity and weight gain (Frank et al., 2004; Sugiyama et al., 2013). Both the transportation and medical literature find a potential mechanism between urban sprawl and obesity through sedentary behaviors such as commuting and traveling, and through commuting by car specifically (Dunton et al., 2009; Frank et al., 2004; Jacobson et al., 2011; Lopez-Zetina et al., 2006). This work examines the relationship between sedentary modes of commuting and physical activity as a possible pathway to the unhealthy outcomes linked with commuting.

Recent work on commuting and obesity implements an instrumental variables approach to establish a causal relationship between commuting and obesity. This chapter uses these approaches to inform an attempt to understand the relationship between commuting and physical activity behavior. Specifically, Zhao and Kaestner (2010) explore causality in the relationship between urban sprawl and obesity by using predicted population densities derived from historical Interstate Highway System plans in two-step instrumental variables approach. They use changes in population density as a measure of urban sprawl, and use an instrumental variables approach to estimate the relationship between urban sprawl and obesity through the use of predicted population densities derived from historical Interstate Highway System plans as an instrument for changes in population density. They use individual-level demographic, socioeconomic, and self-reported height and weight data from the National Health Interview Survey (NHIS) from 1976 to 2001. Data on population, highway infrastructure and MSA-level characteristics were obtained from the Neighborhood Changing Database, the General Location of National System of Interstate Highways, and the Current Population Survey (CPS), respectively. First they predict population density from 1947 planned highway rays, showing that more highway plans are significantly associated with a decrease in population density over time, and that planned highway rays are uncorrelated with observed MSA-level time-varying characteristics. They find no effect of population density on BMI, but do find a significant association between population density and obesity status.

Another recent paper uses instrumental variables as a means to correct for endogeneity in the relationship between time use and BMI. Using cross-sectional data from the American Time Use Survey (ATUS) years 2006 and 2007, Zick et al. (2011) examine the issue of reverse causality in this relationship. They discuss how observed associations between time use and BMI may either be a result of time use behaviors affecting BMI or vice versa, that individual health or BMI affects which activities individuals engage in and how much time they spend in these activities. In order to address this issue and attempt to understand in which direction the causal pathway lies, they use a model of time use where BMI and time use are simultaneously determined. They describe their model where BMI is determined by time use, biological (i.e. health status, age, sex, race/ethnicity), and socio-economic characteristics (i.e. education, employment status, marital status, and number of children). And time spent in particular activities is determined by household roles (i.e. primary meal preparer or grocery shopper), structural factors (i.e. season of interview, weekend/weekday diary day, rural/urban status, geographic region, age, sex, race/ethnicity, marital status, and number of children), prices (such as wage rate and grocery prices), and income. The authors estimate an instrumental variables (IV) model, using instruments which they suggest are correlated to time-use but unrelated to the error term in the BMI estimation. They categorize time use into seven categories: primary eating time, secondary eating time, secondary drinking time, food preparation time, physical activity time, sleep time, and television/video time. They treat each of these as endogenous in a model with multiple endogenous independent variables. They use eight instruments to predict time-use: self-identification as the primary meal preparer in the household, self-identification as the primary grocery shopper in the household, whether the diary day was a weekend, whether the diary day was in the summer, whether the diary day was in the year 2007, the grocery price index, the respondent's wage rate, and the household's annual non-wage income. They then use these predicted time-use variables as regressors in the BMI model. The authors do not argue that the test they use for the strength of their instruments is appropriate when multiple endogenous variables are modelled, and instead test the strength of the full set of instruments on each of the seven first-stage models individually. Also, they provide minimal discussion of the exclusion restriction; it is entirely plausible that many of these variables can jointly influence both time-use and BMI. Nonetheless this work highlights a need for understanding the direction of causality in the relationship between commuting and health outcomes and behaviors. For instance, the direction of causality would have implications for what types of policies may help individuals improve their health.

Recently, a literature has developed in economics that also examines the relationship of how commuting time correlates with time spent in health-producing activities and how time use relates to obesity. Much of this literature builds upon time allocation and human capital frameworks established by Becker in 1965 and Grossman in 1972. Because total time is fixed, the relationship between commuting and physical activity will generally be negative – more time commuting will lead to less time spent in other activities.

Mullahy and Robert (2010) use this framework to examine whether time spent in physical activity is determined by education and also whether education determines how time is spent in all other activities. They use data on adults aged 25-64 from the 2005 and 2006 waves of ATUS data. They model six mutually exclusive and exhaustive categories of time use: sleep, household and personal activities, care for others, work (labor), non-exercise leisure activities, and physical activity. Their definition of physical activity includes any activity coded within the ATUS-defined category of "Sports and Exercise" with the exclusion of a few sub-codes which are less active, such as billiards, boating, bowling, fishing, hunting, and vehicle touring/racing. They include time spent in travel and commuting under the category of household and personal activities. Rather than comparing time spent in PA to time in all other categories, they use a multivariate fractional regression model to simultaneously compare time spent in PA to time spent in each of the other categories of time-use, focusing on how time spent engaged in these activities varies for individuals with differing levels of human capital (education). This method allows them to implicitly include other forms of time use in their model by including the restriction that the sum of estimated time-use categories for each individual is set equal to the total amount of time in a day. They find that individuals with higher human capital, and therefore a higher opportunity cost of time, exercise more on weekend and holidays, when their opportunity cost of time is lower. These groups also exercise more overall. To compensate for spending more time on exercise, these individuals also report less time spent sleeping. The policy relevance of this result is limited because this paper does not argue for the exogeneity of their predictor of interest, not does it use an instrument.

Hoehner et al. (2012) find that longer distances between work and home are correlated with less frequent participation in physical activity as well as decreased cardiorespiratory fitness, greater body mass index, waist circumference, and higher blood pressure. They use cross-sectional data from adults living and working in metropolitan counties in the state of Texas, from 2000 through 2007. Their work adds to the literature by using measured biomarkers to provide a more accurate measure of individual health status. Using respondent's work and home addresses, they calculate commuting distances as the shortest distance along the road network between these two locations. They control for socioeconomic and demographic characteristics, smoking, alcohol intake, family history of diabetes and high cholesterol, BMI, and MET-minutes of self-reported physical activity. Their work provides support to existing literature that identifies increased commuting with decreases in physical activity participation and health status through the use of measured biomarkers; there are some limitations in their assessment of commuting and physical activity. Time spent in physical activity is self-reported and actual commuting time is unknown. While distance between work and home may provide a reasonable proxy, the measurement could be more an artefact of individual choice of residence location or income than a measure of actual time spent commuting. Another limitation is that the sampled population is fairly homogenous, so population subgroups are not well-represented and results may not be generalizable to other groups. Nonetheless, this work suggests evidence of a relationship between commuting, physical activity, and poor health outcomes; this provides support for further work to disentangle these relationships.

Based on Cawley's SLOTH framework of determining physical activity choices, Humphreys and Ruseski (2011) develop and estimate an economic model examining the effect of income and opportunity cost on physical activity behavior. Their model is novel in that it separately distinguishes between the decision to participate (the extensive margin) and the duration of participation decision (the intensive margin). They find that individuals with higher income are more likely to participate in physical activity, but conditional on participation these individuals spend less time being physically active than those with lower income. They use cross-sectional data from the 1998, 1999, and 2000 Behavioral Risk Factor Surveillance System (BRFSS) survey and sample employed adults between the ages of 25 and 54. Respondents are asked if they participated in physical activities in the past month, which two types of physical activity did they spend the most time doing, and how much time did they spend in these activities. Because of the large number of zeros observed in measures of physical activity in their data, the authors estimate a two-part model of participation and time spent in physical activity. I adapt their approach to this work, and note that it has not previously been applied to understanding physical activity using time-use data. Their work goes further by also estimating an instrumental variables estimator, with county-level unemployment rate in each year used to instrument individual income on physical activity decisions. Some limitations of their work come from the data source: first, income and wage data at the individual level are not available; instead they use county-level economic conditions and education as proxies for individual wage and income. Another limitation of their data is that the measure of physical activity is based on self-reported responses about poorly-defined physical activity over the past month. Such data may be subject to measurement error, physical activity may mean different things to different individuals, and certain individuals may be likely to over-report both participation and duration of physical activities. Finally, while the authors present results showing that their instrument, county-level unemployment rate explains variation in individual income and argue that the instrument is uncorrelated with individual-level factors that affect participation in physical activity, it is unclear that this is the case. Particularly, county-level unemployment rates represent county-level economic conditions which can affect county-level characteristics such as crime, litter, vandalism, number and condition of parks and green space, sidewalks, and bicycle lanes, which do impact physical activity (Giles-Corti and Donovan, 2002; Robert Wood Johnson Foundation, 2008; Sallis et al., 2012).

Most recently and closely related to this chapter, Christian (2012) examines trade-offs between commuting time and health-related activities and finds that individuals who spend more time commuting compensate primarily by reducing time spent sleeping, but also by decreasing time engaged in physical activity. Christian uses data from the 2003 through 2010 ATUS, and samples employed adults aged 21 through 65 who live in urban areas. Physical activity is defined as any activity classified by ATUS as "Sports and Exercise" as well as any other activity which exerts a moderate level of exertion. Unlike other research on commuting behavior using this data, Christian defines commuting as "all travel time for any purpose from the time the respondent leaves home until arrival at work and vice versa," regardless of the length or number of stops an individual makes along the way (p. 748). Other work comparing methods of calculating commute time in the ATUS suggests that this definition likely substantially overestimates commuting time, particularly for individuals who make other stops along their travel from home to work and vice versa. It is unclear in which direction this overestimation would bias the results, because it is not known whether commute time is overestimated for individuals who exercise more or who exercise less. This method also disregards commuting mode, so that commuting by car, public transportation, biking and walking are all counted as part of commuting time. While participation in active commuting makes up a very small share (roughly 7% of respondents) of overall commuting, this treatment could bias results relating to physical activity upwards if individuals treat active commuting as a substitute for other physical activity or it could bias results toward zero if active commutes are generally shorter in duration and active commuters are more motivated to engage in healthy behaviors than non-active commuters. Christian notes that this work likely only establishes an upper-bound on commuting trade-offs.

Christian (2012) uses a seemingly unrelated regression (SUR) model, which is a type of simultaneous equations model consisting of several regression equations, each with its own dependent variable, with potentially different sets of exogenous explanatory variables, and with its own error term which is assumed to be correlated with error terms in the other equations. Each individual model can be estimated equation-by-equation using the standard OLS method. SUR uses the feasible generalized least squares estimation method and provides a more efficient estimation of a system of equations when the error terms of each equation are correlated with error terms in the other equations (Zellner, 1962). As such, predicting time-use behavior would be an appropriate application of SUR because time spent in all activities sum to the total amount of time in the survey period, so error terms across equations are correlated. However, there are two cases where SUR estimation is equivalent to equation-by-equation estimation by OLS: 1) when the error terms are actually uncorrelated between equations and 2) when each equation in the system contains exactly the same set of regressors on the right-hand-side (Amemiya, 1985; Greene, 2011). One issue with estimating a series of models that predict time spent in a variety of different activities is that time-use behaviors are co-determined by the same predictors, namely the demographic and socioeconomic characteristics, season and region fixed effects, as well as time spent in all other activities. There is no clear justification to exclude any one predictor from any individual model when estimating time-use. In fact, Christian uses the same explanatory variables across all models. An additional limitation of this work is that self-selection in commuting time may result in biased estimates; Christian notes this and suggests additional work is needed to disentangle preferences for health and location which may drive the results in this work.

This chapter builds upon existing literature, which establishes a relationship between commuting and physical activity, through using time-use data to separately examine both the participation and duration decisions of physical activity, by examining the heterogeneity of this relationship across groups with different health status, and by applying an instrumental variables approach from the urban economics and labor economics literature to address issues of endogenous selection in the effects of commute time and healthy behaviors such as physical activity(Baum-Snow, 2007; Giménez and Molina, 2011).

4.3. Data and Sample Selection

This work uses data from the American Time Use Survey (ATUS) and the ATUS Eating and Health Modules from 2006, 2007, and 2008. The ATUS provides repeated cross-sectional data from households that have recently completed the eighth and final monthly interview of the US Census Bureau's Current Population Survey. Any eligible household member aged 15 or older is randomly selected to be the ATUS respondent. Households with minorities, households with children, and weekend diary days are oversampled to improve the reliability of estimates for these particular subgroups. Sampling weights are provided by ATUS and used in this work to compensate for oversampling of these groups and maintain representativeness at the national level.

This work also uses county-level median value of single-family homes in 1970 as an instrument for commuting behavior. This data was obtained from a dataset developed by Baum-Snow (2007), where county-level data was obtained from the *County and City Data Books* (CCDB) report of decennial census data and aggregated to counties and cities of at least 25,000 inhabitants(U.S. Bureau of the

Census, n.d.). This data was made available upon request from the author, and further details are available in Baum-Snow (2007).

ATUS respondents provide a chronological diary of all primary activities they engaged in from 4 a.m. on the previous day until 4 a.m. on the day of the interview; these activities are reported in per-minute durations with start and end times recorded and are classified into different numerical activity categories. The entire day is accounted for and activities are mutually exclusive, meaning that multi-tasking is not reported. The sample is split evenly between weekdays and weekends and evenly across the weeks in a year. This results in 10% of the sample that report their activities on each weekday and 25% of the sample that report their activities on each day of the weekend. The 2006 to 2008 Eating and Health Modules include additional questions on perceived health status, self-reported height and weight, food preparation activities, and secondary eating activities.

The ATUS has 148,345 respondents in the entire 2003-2013 sample; however this work uses the 2006 through 2008 Eating and Health Module, which has a total of 37,914 observations. Because the focus is on commuting behavior, the sample is restricted to working-age adults, aged 21 through 65 (n = 28,710), who live in urban labor markets, specifically who live in a metropolitan area as defined by the 2000 Census, (n = 23,811). In order to only consider individuals who face a strong time constraint, the sample is limited to individuals who are employed (n = 17,602) and have worked at least seven hours (420 minutes) on the diary day; this brings the sample size down to 7,216 observations. Only individuals who report weekly earnings are included (n = 6,549). For the main analyses, only those individuals with a valid BMI are included (n = 6,190). Because BMI of pregnant individuals have different health implications, pregnant individuals are excluded from the sample. Also excluded are underweight individuals, whose BMI is less than 18.5, as the literature suggests that being underweight might be caused by factors other than time-use, such as genetics and age (Yang and French, 2013). This leaves 6,121 observations; because this work controls for state fixed effects, individuals residing in states where all males or all females within that particular sate have the same physical activity participation outcome are removed. The final sample then is 6,107 observations, in which analysis for males (n = 3,341) and females (n = 2,766) are performed separately. This is for two reasons: medically, BMI has different health implications and health behaviors have differing effects on BMI for the two groups and secondly, determinants of commuting behavior for males and females differ and health outcomes related to commuting are also found to differ (Hemmingsson and Ekelund, 2006; Li et al., 2006; Madden, 1981; Roberts et al., 2011; White, 1977). Summary statistics for the sample are presented below, split by males and females in Table 4.1 and 4.2, respectively.

Table 4.1: Summary Statistics for Male Respondents by Commute Time Category

	Categories of Commuting Times (in Minutes)				
	Overall	0 mins	1-29 mins	30-59 mins	60 + mins
Sample Size	3,341	325	972	1,038	1,006
Age	41.45	42.87	41.11	41.39	41.37
	(10.51)	(10.59)	(10.72)	(10.25)	(10.53)
White	0.66	0.71	0.67	0.67	0.64
Black	0.11	0.10	0.10	0.10	0.12
Hispanic	0.17	0.14	0.16	0.18	0.18
Asian	0.04	0.04	0.05	0.04	0.05
Other Races	0.02	0.02	0.02	0.01	0.01
Less than HS Graduate	0.08	0.07	0.08	0.08	0.09
High School Graduate	0.25	0.28	0.27	0.24	0.22
Some College	0.18	0.17	0.19	0.18	0.16
College Graduate	0.35	0.32	0.33	0.36	0.38
Advanced Degree	0.14	0.17	0.13	0.13	0.15
Income 0 - 24,999	0.10	0.12	0.12	0.09	0.09
Income 25K - 49,999	0.24	0.27	0.27	0.24	0.19
Income 50K - 99,999	0.34	0.30	0.33	0.36	0.36
Income 100,000 +	0.21	0.17	0.18	0.21	0.25
Household Receives Food stamps	0.02	0.02	0.02	0.02	0.02
Has Child in Household	0.54	0.47	0.50	0.56	0.57
Child Age 0 - 2	0.15	0.11	0.16	0.14	0.16
Household Spouse	0.65	0.58	0.59	0.66	0.73
Spouse Works Full-Time	0.31	0.26	0.29	0.32	0.33
Weekend/holiday	0.18	0.25	0.22	0.15	0.14
Winter	0.34	0.37	0.35	0.32	0.35
Spring	0.17	0.17	0.16	0.16	0.19
Summer	0.24	0.25	0.24	0.26	0.23
Autumn	0.24	0.21	0.25	0.26	0.23
Region: Northeast	0.17	0.19	0.15	0.15	0.21
Region: Midwest	0.24	0.24	0.27	0.24	0.20
Region: South	0.35	0.30	0.33	0.36	0.36
Region: West	0.25	0.27	0.25	0.25	0.23
Resp. is Primary Meal Preparer	0.37	0.47	0.43	0.35	0.30
Meal Preparation is shared in HH	0.15	0.13	0.15	0.16	0.16
Does any Leisure Physical Activity	0.25	0.26	0.26	0.26	0.24
Does any PA (Leisure & Non-leisure)	0.39	0.42	0.41	0.40	0.36
Has Physically-demanding Occupation	0.26	0.28	0.28	0.26	0.23
Normal- and Overweight	0.72	0.74	0.71	0.72	0.71
Obese	0.28	0.26	0.29	0.28	0.29
Body Mass Index	28.04	28.03	28.07	27.95	28.12
	(4.82)	(5.10)	(4.65)	(4.69)	(5.01)

Data: ATUS 2006-2008
Note: Unweighted means for indicator variables are shown; additionally, standard errors in parentheses are reported for Body Mass Index (BMI) and Age variables. Statistics are grouped based on time spent commuting.

Table 4.2: Summary Statistics for Female Respondents by Commute Time Category

	Categories of Commuting Times (in Minutes)				
	Overall	0 mins	1-29 mins	30-59 mins	60 + mins
Sample Size	2766	255	1043	885	583
Age	42.34	43.73	41.55	42.21	43.35
	(11.09)	(11.46)	(11.22)	(11.00)	(10.71)
White	0.62	0.61	0.65	0.64	0.55
Black	0.17	0.17	0.16	0.17	0.21
Hispanic	0.14	0.17	0.14	0.13	0.15
Asian	0.04	0.04	0.03	0.04	0.06
Other Races	0.02	0.01	0.02	0.02	0.02
Less than HS Graduate	0.06	0.09	0.06	0.06	0.06
High School Graduate	0.23	0.22	0.24	0.22	0.23
Some College	0.19	0.24	0.20	0.18	0.17
College Graduate	0.37	0.27	0.37	0.41	0.37
Advanced Degree	0.14	0.18	0.13	0.12	0.16
Income 0 - 24,999	0.14	0.12	0.16	0.14	0.09
Income 25K - 49,999	0.26	0.23	0.27	0.27	0.24
Income 50K - 99,999	0.32	0.33	0.30	0.31	0.36
Income 100,000 +	0.16	0.18	0.15	0.16	0.19
Household Receives Food stamps	0.05	0.04	0.06	0.05	0.03
Has Child in Household	0.52	0.48	0.57	0.51	0.49
Child Age 0 - 2	0.11	0.12	0.12	0.09	0.10
Household Spouse	0.51	0.51	0.50	0.51	0.52
Spouse Works Full-Time	0.41	0.41	0.41	0.41	0.40
Weekend/holiday	0.16	0.24	0.17	0.16	0.13
Winter	0.36	0.35	0.35	0.37	0.37
Spring	0.16	0.13	0.17	0.16	0.16
Summer	0.23	0.24	0.23	0.21	0.26
Autumn	0.25	0.29	0.26	0.26	0.22
Region: Northeast	0.16	0.15	0.15	0.18	0.19
Region: Midwest	0.24	0.24	0.25	0.26	0.21
Region: South	0.38	0.39	0.38	0.38	0.38
Region: West	0.21	0.22	0.22	0.19	0.22
Resp. is Primary Meal Preparer	0.76	0.77	0.78	0.75	0.75
Meal Preparation is shared in HH	0.11	0.13	0.11	0.12	0.10
Does any Leisure Physical Activity	0.19	0.21	0.21	0.18	0.18
Does any PA (Leisure & Non-leisure)	0.40	0.50	0.42	0.36	0.38
Has Physically-demanding Occupation	0.15	0.16	0.14	0.16	0.13
Normal- and Overweight	0.74	0.71	0.74	0.74	0.75
Obese	0.26	0.29	0.26	0.26	0.25
Body Mass Index	27.14	27.64	27.12	27.09	27.03
	(6.15)	(6.35)	(5.89)	(6.31)	(6.27)

Data: ATUS 2006-2008

Note: Unweighted means for indicator variables are shown; additionally, standard errors in parentheses are reported for BMI and Age variables. Statistics are grouped based on time spent commuting.

To examine heterogeneity by health status in the preferred model, the sample is split by BMI category and compared across the different groups. BMI measures the mass of the body in relation to height and weight and is used to classify individuals into different weight statuses based on commonly accepted cut points. ATUS data provides a calculation of BMI from respondent's self-reported height and weight. This type of self-reported BMI has been used in previous studies; some researchers have found modest underestimations of overweight and obesity rates while others have found BMI calculated from self-reported height and weight to be both reliable and a valid way to measure BMI for non-elderly adults (Burkhauser and Cawley, 2008; Kuczmarski et al., 2001; Pinkston and Stewart, 2009). BMI between 18.5 and 25 is classified as normal weight, between 25 and 30 is classified as overweight, and BMI of 30 and above is classified as obese. Using BMI classifications, obese individuals are at an elevated risk for chronic disease, especially heart diseases, and are an important target of many health interventions (Must et al., 1999). Although other measures, such as waist circumference, waist to hip ratio, and waist to height ratio, percentage body fat, lipid and lipoprotein levels, have been suggested as superior to BMI in predicting risk for chronic disease, BMI has been traditionally used as a predictor for chronic disease risk and is still often used when studying populations, simply because it can provide an estimate of risk levels in the population and is less invasive and less costly to obtain than other measures (Huxley et al., 2009). Because the medical literature suggests that a BMI of 30 and above is associated with an elevated risk for coronary heart disease and diabetes, Individuals with a BMI less than 30 are compared against those with BMI 30 or more in this work (Harris et al., 1993; Hubert et al., 1983; WHO Expert Committee, 1995). Table 4.3 presents summary statistics of participation rates and time spent in a variety of commuting and physical activities, split by sex and by BMI status.

Table 4.3: Summary Statistics by BMI Status

	Males			Females		
	All BMI % or mean	BMI < 30 % or mean	BMI 30+ % or mean	All BMI % or mean	BMI < 30 % or mean	BMI 30+ % or mean
Does No Commuting	0.08	0.07	0.08	0.08	0.07	0.09
Does Any Sedentary Commuting	0.90	0.90	0.91	0.91	0.91	0.90
Mean Commute Time	49.15	49.11	49.23	40.34	40.55	39.72
SD	(38.22)	(38.21)	(38.28)	(33.45)	(33.70)	(32.75)
Does Any Active Commuting	0.07	0.08	0.05	0.07	0.07	0.06
Mean Commute Time	18.56	18.79	17.52	16.29	15.82	17.84
SD	(21.24)	(21.58)	(19.80)	(18.83)	(17.96)	(21.62)
Physically Demanding Occupation	0.26	0.26	0.25	0.15	0.14	0.17
Does Any Physical Activity	0.41	0.43	0.38	0.42	0.44	0.36
Mean PA Time	70.16	69.76	71.30	56.36	57.77	51.51
SD	(62.73)	(63.03)	(61.93)	(53.89)	(54.59)	(51.23)
Does any Leisure-Only PA	0.25	0.26	0.23	0.20	0.21	0.15
Mean PA Time	76.47	76.41	76.64	61.28	61.29	61.23
SD	(53.88)	(53.11)	(56.23)	(44.83)	(45.20)	(43.49)
Number of Observations	3,341	2,391	950	2,766	2,045	721

Data: ATUS 2006-2008
Note: Unweighted means for indicator variables and time variables are shown; additionally, standard errors for time spent in various activities are shown in parentheses. Statistics are grouped by sex and BMI status.

This table shows that the percentage of individuals involved in different types of commuting behavior is fairly even across BMI categories, as is the time spent in these activities. Among men, distribution of occupations with physical activity is fairly even, but among women, a slightly higher percentage of obese women are in these types of occupations. Physical activity participation is lower among obese males and females as compared to the normal and overweight groups, however average time spent in physical activity among participants is quite similar across BMI status.

The set of explanatory variables in this analysis allow us to explore how the relationship between commuting behavior may affect physical activity while controlling for a number of covariates which may themselves have an effect on physical activity. All analyses control for respondent age and age squared because there may be non-linear relationships between age and physical activity behavior. Controls for socioeconomic and demographic variables such as race, education, log of weekly earnings, food stamp recipient status are also included; previous literature has found that a number of these factors are related to individual physical activity participation (Powell et al., 2004; Trost et al., 2002). Controls are also included for individual household composition characteristics such as whether or not there is a child in the household, whether there is a child under age two in the household, respondent's marital status, spouse's employment status, and participation in household meal preparation. Existing literature has shown that individuals who are married, have children, particularly young children, or are involved in non-market labor, such has household duties, may have more time commitments compared with single respondents, those without children, and those with fewer household responsibilities (Brown and Roberts, 2011; Humphreys and Ruseski, 2009; Popham and Mitchell, 2006).

This work uses a quite broad definition of physical activity with the motivation that any physical activity can affect an individual's health outcomes, including non-leisure physical activity (Arrieta and Russell, 2008). This is done by using metabolic equivalent (MET) values to quantify the intensity of physical activity; a MET is defined as the ratio of energy expenditure in a particular activity to energy expenditure while at rest. Activities with a MET value of 3.00 to 6.00 are defined as moderate-intensity physical activity and those with a MET value greater than 6.00 are classified as vigorous-intensity; both count towards the CDC's Physical Activities Guidelines (Ainsworth et al., 2000; USDHHS, 2008). Following the work of Tudor-Locke et al. (Tudor-Locke et al., 2009) who have linked the ATUS activity lexicon with MET values from the Compendium of Physical Activities, physical activity is defined here as any activity in the ATUS activity lexicon that generates MET value of 3.00 or more. These include activities classified as sports and recreation, lawn and garden work, home maintenance and repair, exterior house cleaning and playing sports with children. In the definition of physical activity in this work, any non-commuting travel that is undertaken by walking or cycling is also included as physical activity. Because specific activities an individual undertakes while at work are not identifiable in this data, activities during an individual's working hours are excluded from the measure of physical activity in this work. Instead, MET values linked to the 2002 Census Occupational Classification System codes are available and broadly identify the physical demands of different occupations (Tudor-Locke et al., 2011). To account for physical activity individuals do on the job all analyses include a control variable for having a physically demanding occupation that equals one if the main activities of an individual's primary occupation are at least three METs, and a zero otherwise.

Figure 4.1: Unweighted Distribution of Physical Activity Time among PA Participants

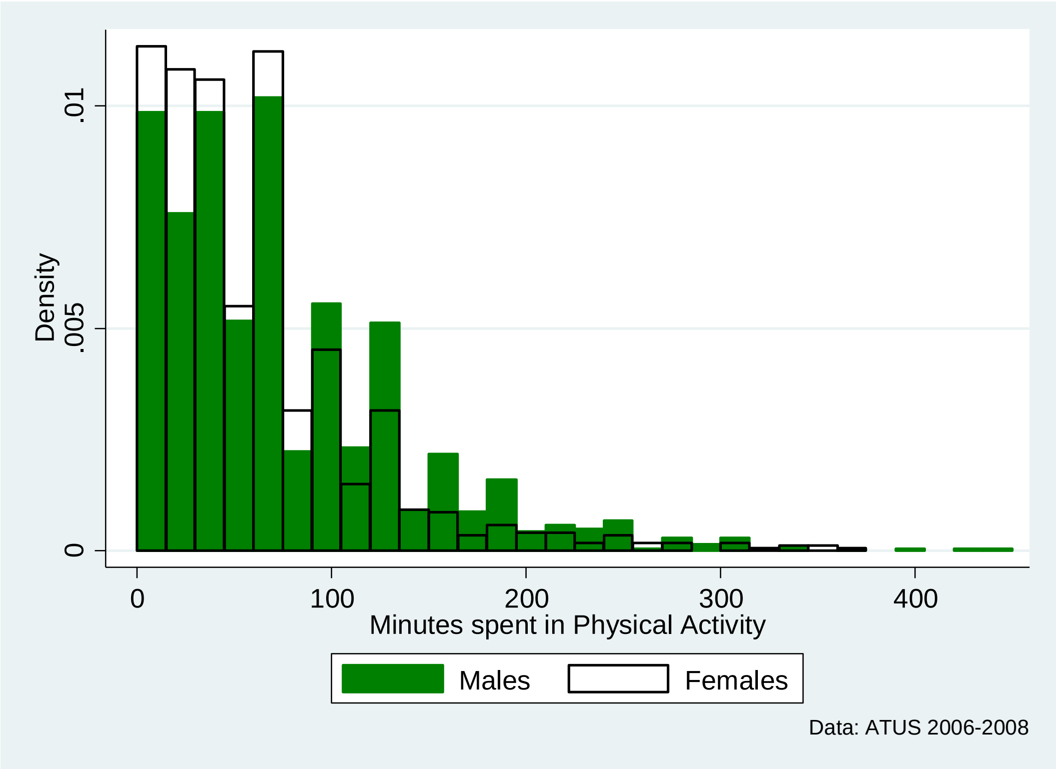
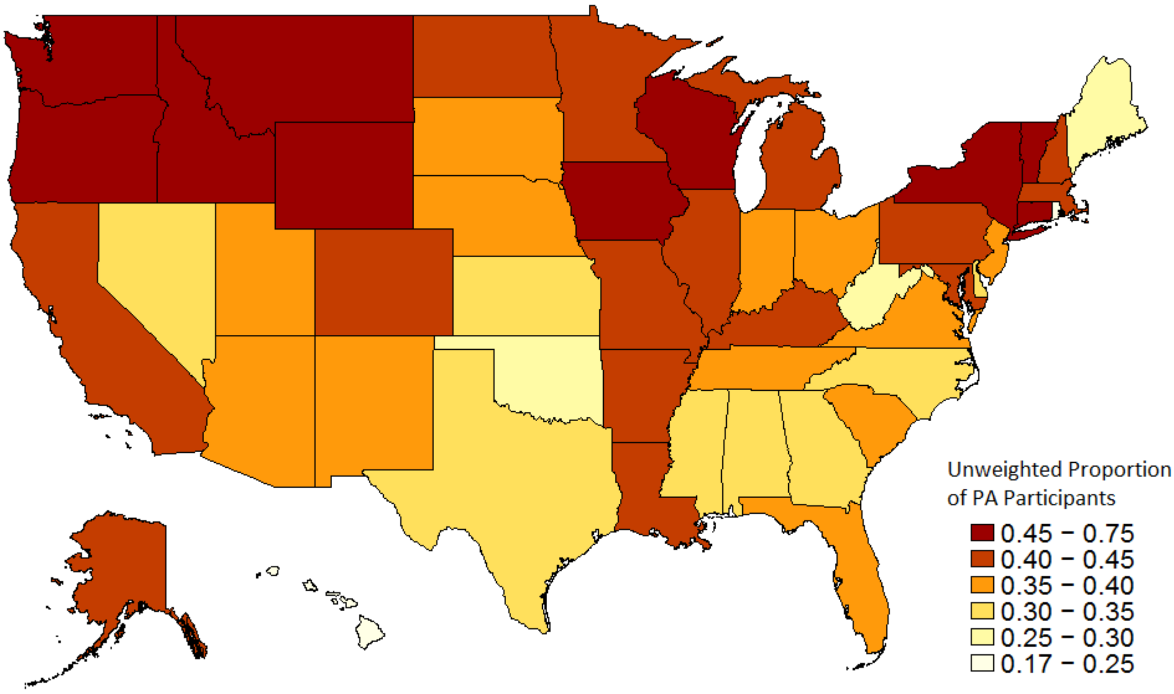


Figure 4.1 presents a histogram of unweighted number of minutes spent in physical activity among those individuals who participate in physical activity on the diary day, males and females shown separately. As expected, PA time is truncated at 0 and has characteristics of an exponential distribution, but there is bunching at five- and ten-minute intervals.

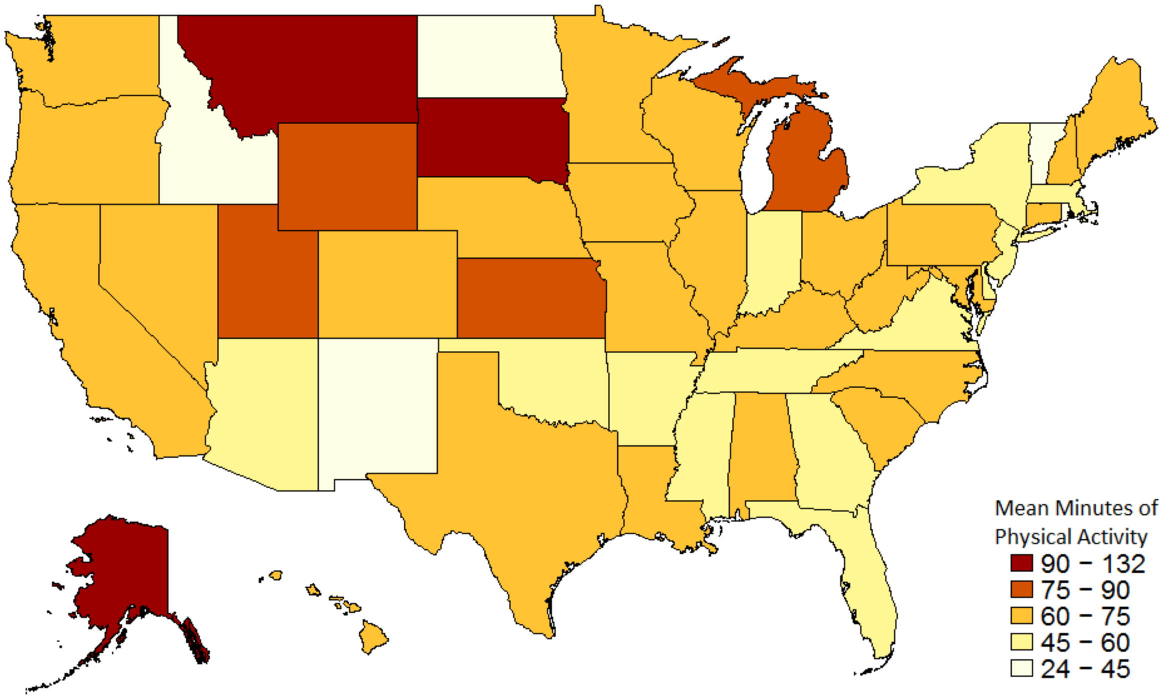
Figure 4.2: Unweighted Proportion of Physical Activity Participants by State



Data: ATUS 2006 - 2008
Unweighted proportion of individuals in the selected sample who do some physical activity (PA) on the diary day. Sample is restricted to employed adults age 21 through 65, residing in a metropolitan area, who worked at least seven hours on the diary day, and for whom BMI is greater than or equal to 18.5. These unweighted data are not representative of the population at the state or national level.

Figure 4.2 presents a map of the unweighted proportion of individuals who participate in physical activity by state, and Figure 4.3 presents mean time spent in physical activity among those individuals who do participate in physical activity on the diary day, by state. Because these figures present unweighted data, they are not representative of the US population, but instead represent the selected sample in the ATUS data; all analyses use appropriate sample weights to insure representativeness at the national level.

Figure 4.3: Unweighted Mean Physical Activity Time Among PA Participants, by State



Data: ATUS 2006 - 2008
Note: Unweighted mean of time spent in physical activity (PA) among those individuals who participate in PA on the diary day. Sample is restricted to employed adults age 21 through 65 residing in a metropolitan area, who worked at least seven hours on the diary day, for whom BMI is at least 18.5, and who participated in physical activity on the diary day. These unweighted data are not representative of the population at the state or national level.

Using time-use data, the literature suggests a variety of methods to define and calculate commuting time. This analysis uses the ATUS definition of "travel related to work," which is defined strictly as any travel occurring immediately before work and any travel occurring immediately after work, provided that the next activity takes place at one's home (Kimbrough, Gray, 2015). This definition can underestimate the amount of commuting in situations where an individual does not go home directly after work but instead makes other trips between work and home. Using this definition of commuting time, I expect this to decrease the size of the effects found to a small degree, so estimates may be conservative. I make a distinction between active commuting, which is commuting by walking or cycling, and sedentary commuting, which is defined as commuting via any other mode of transportation. There is a literature that highlights the differences in health outcomes from active commuting and sedentary commuting (Gordon-Larsen et al., 2009; Hartog et al., 2010). Active commuting is associated with improved measures of health and well-being, whereas increased sedentary commuting is associated with lower levels of health and well-being. Including active commuting as part of physical activity would have the effect of negatively biasing the coefficient on commuting behavior if individuals engaged in active commuting treat it as a substitute for sedentary commuting. On the other hand, including active commuting in the calculation of (sedentary) commuting time would positively bias the coefficient on commuting, especially if individuals who engage in active commuting substitute it for time spent in physical activity. Because of this, active commuting is not being included in the definitions of either commuting or physical activity in this work; going forward all references to commuting will specifically be referring to sedentary commuting and focusing on the relationship between sedentary commuting and physical activity. To consider the effect of whether or not an individual commutes, an indicator variable, *dCommuting* is used. As with the indicator for participation in physical activity, this equals zero when the length of the commute is less than ten minutes and equals one for commutes of at least ten minutes.

Figure 4.4: Distribution of Time Spent Commuting among Commuters

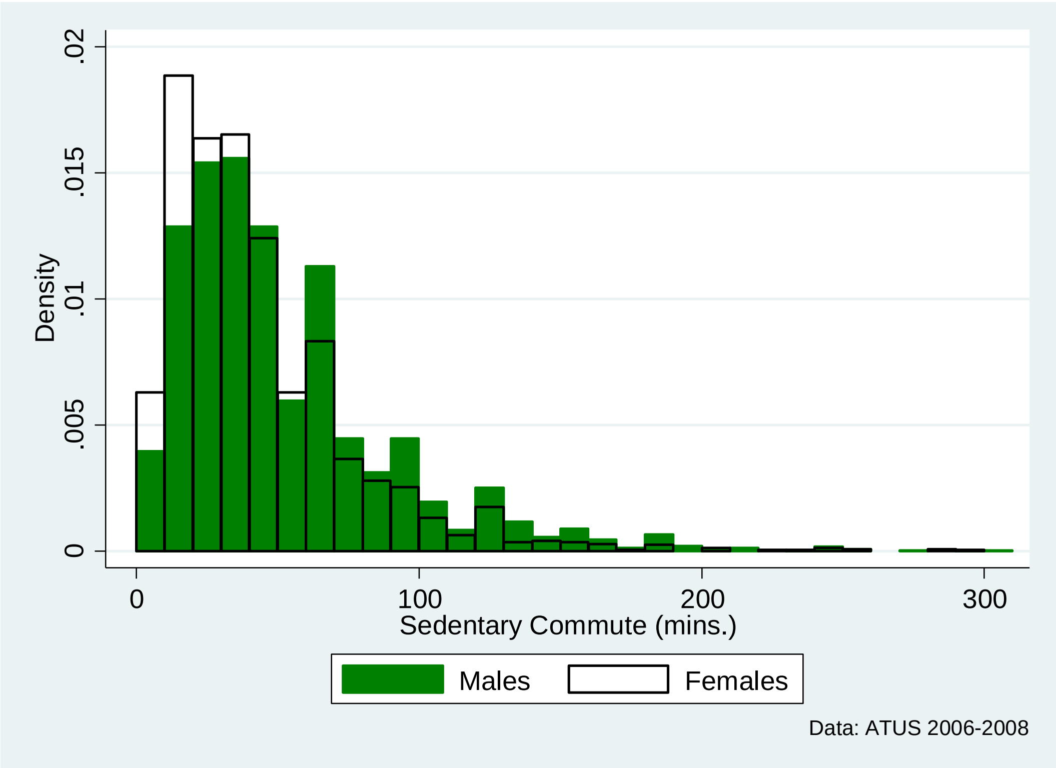
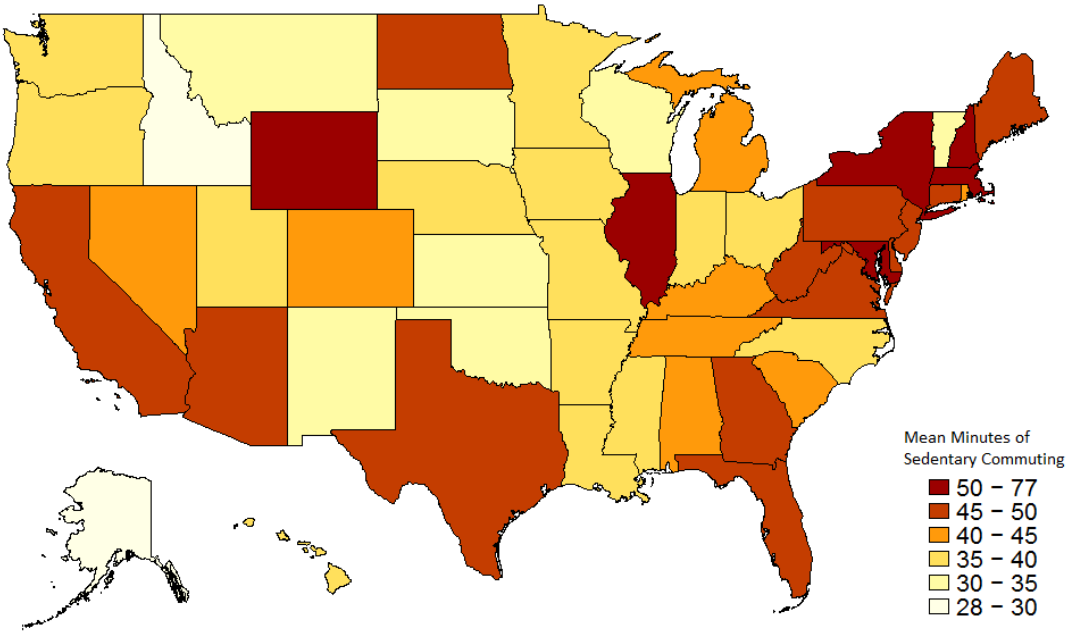


Figure 4.4 presents unweighted time spent in sedentary commuting among those individuals who do commute, with males and females shown separately. The distribution of sedentary commuting times is truncated at zero, with a very long right tail.

Figure 4.5: Unweighted Mean Sedentary Commute Time (in Minutes) by State



Data: ATUS 2006 - 2008
Unweighted mean of Sedentary Commute time in a day. Sample is restricted to employed adults age 21 through 65 residing in a metropolitan area, who worked at least seven hours on the diary day, for whom BMI is at least 18.5, and who participated in sedentary commuting on the diary day. These unweighted data are not representative of the population at the state or national level.

Figure 4.5 presents unweighted mean time spent in sedentary commuting among sedentary commuters by state. Again, because these data are unweighted, they are not representative at the national level, however appropriate sampling weights are used to ensure representativeness in all subsequent analyses. Nonetheless, these commute times by state are fairly close to the other, nationally representative, statistics on mean commute times by state (U.S. Census Bureau, n.d.).

4.4. Methods

Initially, this work models differences in commuting time as an exogenous variation in the amount of time available for other activities such as physical activity, the variable of interest. The motivation is that by looking at variation in commuting time for a set of individuals who otherwise face approximately similar time constraints, one can understand how individuals might respond to variation in the amount of time available in the day by making trade-offs between time engaged in health-producing behaviors, such as physical activity and time spent commuting.

Because individuals face a time constraint of 24 hours and this work examines a sample of individuals who work at least seven hours during that time, increases in commuting time are expected to reduce time available for other discretionary activities such as leisure and physical activity. For this reason a negative and significant relationship between commuting time and physical activity is expected. With time-use data, a large number of zeros exist when individuals do not participate in a particular activity and there are different methods to deal with this (Frazis and Stewart, 2012). When the zeros in time-use data come from a mismatch between the period of interest (such as a long-term health outcome) and the period of the data (such as an individuals' activities in a single day) OLS is suggested as the most appropriate estimation technique; however that approach is not taken here because the period of the outcome of interest, individual engagement in physical activity on a single day, matches the period of the predictor, individual commuting behavior on a single day (Frazis and Stewart, 2012). Instead, this work follows literature that suggests a two-part model is appropriate for first identifying the extensive margin and then the intensive margin for those who do participate (Cragg, 1971; Jones, 2000; Cawley and Liu, 2007; Humphreys and Ruseski, 2011). To examine the extensive margin, or whether or not an individual participates in physical activity, a Probit model is used to estimate the change in probability of physical activity participation dependent upon whether or not the individual commutes. That is, the following reduced-form model is estimated:

$$(4.1) \qquad PR(dPA_i = 1 | Commuting_i, X_i) = \phi(\beta_0 + \beta_1 dCommuting_i + \beta_2 X_i + \varepsilon_i),$$

where *dPA* is an indicator variable which is equal to one if the respondent spends at least ten minutes of physical activity on the diary day, which is the minimum amount of time for any one episode of activity, as suggested by the CDC Guidelines for Physical Activity (USDHHS, 2008). In this model *dCommuting* is a similarly defined indicator variable for commuting[4]. This model gives us an estimate of the likelihood of participating in physical activity dependent on the likelihood an individual spends at least ten minutes of their day commuting. -

A similar specification to this model provides an alternate way to examine the extensive margin. In Equation (4.2) the probability of participation in physical activity as dependent upon the amount of time the individual commutes. This second model estimates the probability an individual would spend

[4] Note: The ten-minute threshold for dummy variables on commuting and physical activity result in conservative estimates. However, results remain robust to using other specifications where these dummy variables equal one for a one-minute duration and five-minute duration of physical activity and sedentary commuting. Coefficients increase slightly as durations decrease. Keeping *dPA* at the ten-minute threshold and reducing *dCommuting* to a five-minute threshold gives a coefficient of -0.0879, and a one-minute threshold results in a coefficient of -0.0946, both statistically significant at the $p<0.05$ level. Similarly, holding *dCommuting* at the 10-minute threshold and reducing *dPA* to a five-minute threshold results in a coefficient of -0.0645, statistically significant at the $p<0.10$ level and reducing *dPA* to a one-minute threshold coefficient of -0.0682, statistically significant at the $p<0.05$ level.

at least ten minutes engaged in physical activity using an equation of the same form but define instead *Commuting* as the total number of minutes the respondent spends commuting on the diary day:

(4.2) $$PR(dPA_i = 1|Commuting_i, X_i) = \phi(\beta_0 + \beta_1 Commuting_i + \beta_2 X_i + \varepsilon_i).$$

I expect the coefficient on commuting behavior to be larger in the estimation of Equation (4.1) than in Equation (4.2) because the effect of whether or not an individual meets some minimum amount of commuting would influence time for physical activity much more than any one-minute increase in the total commuting time would have.

To examine the intensive margin, or how much time the physically active individuals spend in these activities, the sample is limited to only those individuals who actually do any physical activity on the diary day. An OLS model is used to regress the time these individuals spent engaged in physical activity on the time they spend commuting:

(4.3) $$PA_i = \beta_0 + \beta_1 Commuting_i + \beta_2 X_i + \varepsilon_i.$$

In all models, *X* is the set of control variables and all analyses are carried out separately for males and females. Results of non-linear Probit models (4.1) and (4.2) are reported as the average marginal effects for all non-binary variables and can be interpreted as the average change in probability of participating in physical activity as the variable increases by one unit. For the binary commuting indicator, the effect is the change in probability as the indicator changes from zero to one in Equation (4.1). All estimates are made clustering at the county-month-year level so that standard errors are robust to non-independence of observations from the same county-month-year. This is especially necessary when including instrumental variables, as the instruments are constant for all observations within the same county.

The decision to commute is represented by the variable, *dCommuting*. One concern is that this decision may be co-determined with the decision to engage in physical activity. In particular, individuals who are more likely to be physically active may choose residential locations nearer to their jobs or choose a commuting time with less congestion in order to allow for either active commuting or for a shorter commute which then allows them more time for participation in physical activity. In other words, if these individuals who are more likely to be physically active are also more time-sensitive than sedentary individuals, then this could bias the coefficient upward or it result in reverse causality, where involvement in physical activity is causing the individual to find a way to shorten their commute time. If this is the case, the commuting variable is correlated with the error term, ϵ. One way to deal with this is to identify an instrument which is correlated with the explanatory commuting variable but is not correlated with the physical activity outcome variable.

An instrumental variables approach is commonly used to establish causality in situations of endogenous selection like this (Wooldridge, 2010). A valid instrument must satisfy two conditions: first, an instrumental variable must be strongly correlated with the endogenous explanatory variable, conditional on other covariates, and second, the instrument cannot be correlated with the error term, meaning essentially that the instrument cannot have the same issue of endogeneity as the original predicting variable. This second condition is often called the exclusion restriction. Previous research has presented instrumental variables as a means of disentangling the causal effects of urban sprawl or sedentary commuting on health, health behaviors, and labor market outcomes; however no valid instruments have been identified for predicting the effect of active commuting choices at the individual level on BMI (Giménez and Molina, 2011; Hymel, 2009; Schauder and Foley, 2015; Wojan and Hamrick, 2015; Zhao and Kaestner, 2010). The instrumental variables approach proposed in Schauder and Foley (2015) most closely relate to the estimation in this chapter; the authors use season in which the survey was administered and rent as an instrument to predict active commuting participation. Although the authors show their instruments are strongly correlated with active commuting, they may still suffer from endogenous selection. In particular, both seasonal and regional variations occur in BMI as well as in other health behaviors which co-determine BMI (Visscher and Seidell 2004; Pivarnik et al. 2003; Ma et al. 2005; Plasqui and Westerterp 2004; Scott et al. 2009; Ford et al. 2005; Reis et al. 2004; Dutton and McLaren 2011). The authors also point out that renting (versus owning) a home is a choice individuals make that could be co-determined by unobserved factors that affect health outcomes. Similarly, fuel prices are also ruled out as an instrument for commuting time on health outcomes because while changes in fuel prices may affect commuting behavior, when individuals are faced with higher prices, they may substitute spending on cheaper, less nutritious, energy-dense foods (Gicheva et al. 2010), which can then affect their BMI.

In an attempt to address this issue, this work uses lagged housing prices as an instrument for sedentary commuting behavior. Brueckner and Fansler (1983) suggest that lagged real estate prices can meet these criteria in that real estate prices strongly determine urban sprawl, a determinant in the commute time. Similarly, Baum-Snow (2007) and Hymel (2009) use historic highway infrastructure plans as a source of exogenous variation in highway development, to estimate a causal effect of highway infrastructure on suburban development, using the rationale that original highway networks were planned to link faraway places and not to facilitate commuting from central cities to suburbs, so original infrastructure plans would not be co-determined with current patterns in commuting and suburban development. Giménez and Molina (2011) use single-period lags of housing prices as well as future-period housing prices to predict current period labor supply; their identification strategy being that the combination of past-period and future-period housing prices may predict labor market outcomes where current-period housing prices may be co-determined with labor market outcomes. Following similar reasoning as in these papers, this work proposes that housing prices from earlier periods are predictive of urban infrastructure, and thus commuting behavior today, and also would not be endogenous to an individual's current period choice of physical activity participation on a given day. However, there are a number of weaknesses in using past county-level housing prices as an instrument for commuting behavior. First, this work uses mean county-level house prices, this does not take into account the degree of county-level variation in urban structure, so predictive power is limited (Glaeser et al., 2008). Also because the lagged housing price variable is constant at the county level, geographic county-level or MSA-level fixed effects are not included in regressions which use these instruments. Previous literature suggests that males' and females' commuting behavior is very different and decisions are made based on differing criteria, and I find in the first-stage regressions that lagged housing prices do not accurately predict women's commute behavior, despite the fact that they do so for men (Madden, 1981; Roberts et al., 2011; Sermons and Koppelman, 2001; White, 1977). Another limitation is that past housing prices are only predictive of commuting behavior among males, so the instrument variables model is only estimated for males in the sample.

To estimate Equation (4.1) using an instrumental variables approach, I estimate the model using a recursive bivariate Probit method, which is recommended in situations where the outcome and an endogenous predictor of interest are both binary variables (Wooldridge, 2010). This method is presented as:

(4.4) $dPA_i = 1[z_{1i}\delta_1 + \alpha\, dCommuting_i + u_i > 0]$

(4.5) $dCommuting_i = 1[z_i\delta_2 + v_i > 0]$,

where z_1 is the matrix of all control variables, z_2 is the matrix of the instrumental variables, in this case 1970 county-level housing prices, and z is the matrix of all control variables and instrumental variables, (z_1, z_2). This method estimates Equations (4.4) and (4.5) simultaneously, and also estimates another parameter, ρ, which is equal to the correlation of the error terms, (u, v). If ρ is not equal to zero, then the error terms of Equations (4.4) and (4.5) are correlated and estimating Equation (4.4) alone is inconsistent for the coefficient on *dCommuting,* α. Using the joint distribution of $(dPA, dCommuting)$ to estimate both of these equations simultaneously gives a consistent estimate for α which uses the instrument from Equation (4.5) to account for endogeneity. On the other hand, when ρ equals zero, the error terms are not correlated and it would be sufficient to estimate Equation (4.4) alone, as in the non-instrumented in Equation (4.1). For this analysis, weakness of the instrument is tested by using a χ^2-test to compare the first-stage regression to a regression of participation in sedentary commuting on control variables alone.

To interpret the Probit and recursive bivariate Probit results, average marginal effects are calculated for each coefficient. In the case of binary predictors, such as the binary commute variable, average marginal effects are estimated by first estimating from the model a fitted probability of participating in physical activity for each individual assuming the binary predictor for that individual is equal to one. Second, I estimate from the model a fitted probability of participating in physical activity for each individual assuming the binary predictor for that individual is equal to zero. Then I estimate a marginal effect for each individual which is the difference between those two estimates. From this, the average marginal effect for the population can be estimated.

This instrument is weak when used to predict the total time spent commuting as in Equations (4.2) and (4.3). Baum-Snow (2007) also uses a historic measure of highway infrastructure as an instrument for urban sprawl, but this is also weak when used as an instrument with the data used in this chapter.

4.5. Results

Table 4.4 presents the estimates of Equation (4.1) for males, applying different sets of control variables: demographic controls in Column 2, education in Column 3, income in Column 4, and meal preparation in Column 5. The full model, including state fixed effects is presented Column 6.

Table 4.4: Sedentary Commuting Indicator and Physical Activity Participation for Males

	(1) No Controls	(2) Demo-graphic Controls	(3) Education Controls	(4) Income Controls	(5) Meal Prep Controls	(6) Full Model, FE
Sedentary Commute Indicator	-0.0608*	-0.0569*	-0.0608*	-0.0641**	-0.0639**	-0.0500*
	(0.0322)	(0.0316)	(0.0324)	(0.0319)	(0.0320)	(0.0295)
Respondent Age		-0.00449				-0.00978
		(0.00744)				(0.00736)
Respondent Age Squared		0.0000533				0.000116
		(0.0000882)				(0.00009)
Black		-0.0272				-0.0183
		(0.0330)				(0.0319)
Hispanic		-0.0762***				-0.0492
		(0.0290)				(0.0336)
Asian		0.0257				0.0148
		(0.0458)				(0.0461)
Other Race		-0.141*				-0.136*
		(0.0846)				(0.0801)
Spouse/Partner in HH		0.0364				0.0382
		(0.0355)				(0.0394)
Spouse is Employed		0.0377				0.0412
		(0.0270)				(0.0267)
Has a Child in Household		0.0138				0.0263
		(0.0260)				(0.0252)
Child Under Age 2 in HH		0.106***				0.107***
		(0.0299)				(0.0296)
High School Graduate			0.0921**			0.0782*
			(0.0396)			(0.0408)
Some College			0.0873**			0.0706
			(0.0404)			(0.0441)
College Graduate			0.143***			0.101**
			(0.0374)			(0.0441)
Advanced Degree			0.127***			0.0679
			(0.0433)			(0.0528)
Log Weekly Earnings				0.0471***		0.0213
				(0.0161)		(0.0193)
HH Receives Food Stamps				0.0260		-0.0130
				(0.0734)		(0.0712)
Occupation with PA				-0.0347		-0.0174
				(0.0249)		(0.0250)
Resp. Primary Meal Preparer					-0.0215	0.0201
					(0.0219)	(0.0269)

Table 4.4, Continued.

Shared Meal Preparation					0.0114	-0.0150
					(0.0272)	(0.0259)
State Fixed Effects						X
Observations	3,341	3,341	3,341	3,341	3,341	3,341

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Note: Results from Probit regression estimating Equation (4.1). The outcome variable in each column is the binary indicator of whether or not an individual spends at least ten minutes in physical activity on the diary day; Average marginal effects are reported. Column 1 includes no control variables. Column 2 controls for the following demographic characteristics: age, age squared, race categories (Black, Hispanic, Asian, and other non-white races), whether a spouse or partner lives in the household, whether the spouse/partner is employed, whether there is a child in the household, and whether there is a child under age 2 in the household. Column 3 controls for education, whether the respondent is a high school graduate, has some college, is a college graduate, or has an advanced degree. Column 4 controls for the log of reported weekly earnings, whether the household received food stamps, and whether the respondent has an occupation that is physically demanding. Column 5 controls for household food preparation tasks: whether the respondent is the primary meal preparer and whether meal preparation is shared in the household. Column 6 includes all controls and Column 7 includes all controls as well as state fixed effects. In all columns, sampling weights are applied and standard errors are clustered at county-month-year level.

Similarly, Table 4.5 presents estimates of Equation (4.1) for females, building the model by controlling for different sets of variables, column by column.

Table 4.5: Sedentary Commuting Indicator and Physical Activity Participation for Females

	(1)	(2)	(3)	(4)	(5)	(7)
	No Controls	Demo-graphic Controls	Education Controls	Income Controls	Meal Prep Controls	Full Model, FE
Sedentary Commute Indicator	-0.0760**	-0.0660**	-0.0779**	-0.0786**	-0.0745**	-0.0742**
	(0.0308)	(0.0302)	(0.0305)	(0.0307)	(0.0310)	(0.0296)
Respondent Age		0.00978				0.00385
		(0.00861)				(0.00834)
Respondent Age Squared		-0.0000919				-0.0000266
		(0.0000995)				(0.0000967)
Black		-0.0812***				-0.0503
		(0.0301)				(0.0313)
Hispanic		-0.0310				-0.0170
		(0.0321)				(0.0374)
Asian		0.0515				0.0472
		(0.0664)				(0.0654)
Other Race		0.0918				0.0957
		(0.0954)				(0.0905)
Spouse/Partner in HH		-0.0409				-0.0296
		(0.0517)				(0.0538)

Table 4.5, Continued.

Spouse is Employed		0.0279				0.0199
		(0.0499)				(0.0511)
Has a Child in Household		-0.00807				-0.00787
		(0.0289)				(0.0290)
Child Under Age 2 in HH		0.280***				0.274***
		(0.0357)				(0.0349)
High School Graduate			0.0107			0.0201
			(0.0553)			(0.0584)
Some College			0.0440			0.0556
			(0.0541)			(0.0570)
College Graduate			0.0805			0.0674
			(0.0528)			(0.0601)
Advanced Degree			0.102*			0.0672
			(0.0584)			(0.0695)
Log Weekly Earnings				0.0469**		0.0233
				(0.0202)		(0.0217)
HH Receives Food Stamps				0.0812		0.0680
				(0.0600)		(0.0606)
Occupation with PA				-0.0434		-0.0301
				(0.0310)		(0.0322)
Resp. is Primary Meal Preparer					0.0397	0.0335
					(0.0344)	(0.0326)
Shared Meal Preparation					0.0408	0.0202
					(0.0511)	(0.0471)
State Fixed Effects						X
Observations	2,766	2,766	2,766	2,766	2,766	2,766

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Note: Results from Probit regression estimating Equation (4.1). The outcome variable in each column is the binary indicator of whether or not an individual spends at least ten minutes in physical activity on the diary day; Average marginal effects are reported. Column 1 includes no control variables. Column 2 controls for the following demographic characteristics: age, age squared, race categories (Black, Hispanic, Asian, and other non-white races), whether a spouse or partner lives in the household, whether the spouse/partner is employed, whether there is a child in the household, and whether there is a child under age 2 in the household. Column 3 controls for education, whether the respondent is a high school graduate, has some college, is a college graduate, or has an advanced degree. Column 4 controls for the log of reported weekly earnings, whether the household received food stamps, and whether the respondent has an occupation that is physically demanding. Column 5 controls for household food preparation tasks: whether the respondent is the primary meal preparer and whether meal preparation is shared in the household. Column 6 includes all controls and Column 7 includes all controls as well as state fixed effects. In all columns, sampling weights are applied and standard errors are clustered at county-month-year level.

Although the coefficient on the sedentary commuting indicator remains consistently between -0.05 and -0.07 across all models and is equal to -0.0500 for males and -0.0742 for females when all control variables are included, it shows there is some variation for when different controls are used and that variation is not uniform between genders. This can be interpreted as implying that someone who commutes at all in a given day will expect a decrease in their probability of participating in physical activity on that day of 5.0 percentage points for males and 7.4 percentage points for females. The probability of participating in any type of physical activity in the data is about 39% for males and 40% for females, so this represents a decrease of about 12.7% for males and of 18.4% for females.[5]

Table 4.6: Relationship between Sedentary Commuting Participation and Physical Activity Participation by BMI Group

Probit	(1) All BMI	(2) BMI < 30	(3) BMI 30+
Males			
Sedentary Commute Indicator	-0.0500*	-0.102***	0.0620
	(0.0295)	(0.0377)	(0.0462)
Observations	3,341	2,386	939
Females			
Sedentary Commute Indicator	-0.0742**	-0.114***	0.0340
	(0.0296)	(0.0362)	(0.0444)
Observations	2,766	2,044	709

Robust standard errors in parentheses.

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

Note: Results from Probit regression estimating Equation (4.1). The outcome variable in each column is the binary indicator of whether or not an individual spends at least ten minutes in physical activity on the diary day; Average marginal effects are reported. All columns include the full set of controls and state fixed effects from Column (7) of Tables 4.4 and 4.5. Sampling weights are applied and standard errors are clustered at the county-month-year level. Column (1) estimates Equation (4.1) for the full sample, Column (2) for individuals with BMI less than 30, and Column (3) for individuals with BMI greater than or equal to 30, where BMI status is a proxy for health.

Table 4.6 focuses only on the coefficient on the sedentary commuting indicator as estimated by Equation (4.1) and uses BMI as a proxy for health status, comparing the normal and overweight individuals (those with BMI greater than or equal to 18.5 and less than 30) against the obese (BMI greater than or equal to 30). Making this distinction reduces the number of observations in each

[5] In these results, physical activity and commuting indicators use a ten-minute threshold. This threshold results in conservative estimates. Using a five-minute and a one-minute minimum threshold in either variable increases magnitude of estimated effect sizes and t-statistic of results.

estimation and therefore increases the standard errors. More interestingly, though, these results show a difference in the physical activity behavior of the two BMI groups. The decrease in the probability of participating in physical activity for an individual who does sedentary commuting as compared to someone who does not is larger for normal and overweight individuals than it is for obese individuals. But for obese individuals, that coefficient is no longer significantly different from zero and is, in fact, estimated to be positive. In order to test whether these differences are statistically significant, I estimate these two regressions simultaneously and test the null hypothesis that the coefficients on the commuting indicators are equal across the two groups. The χ2 statistics that the coefficient on normal and overweight males is equal to the coefficient on obese males is 6.38 (p value of 0.0116). The same test comparing normal and overweight females to obese females has χ2 statistic of 5.12 (p value = 0.0236). So, for both males and females the null hypothesis can be rejected and the difference between the normal and overweight group against the obese group is found to be statistically significant.

Table 4.7: Relationship between Sedentary Commuting Time Spent and Physical Activity Participation by BMI Group

	(1)	(2)	(3)
Probit	All BMI Groups	BMI < 30	BMI ≥ 30
Males			
Minutes Spent Sedentary Commute	-0.000976***	-0.000975***	-0.00105**
	(0.000260)	(0.000309)	(0.000422)
Observations	3,341	2,386	939
Females			
Minutes Spent Sedentary Commute	-0.00133***	-0.00158***	-0.000736
	(0.000329)	(0.000386)	(0.000537)
Observations	2,766	2,044	709

Robust standard errors in parentheses.

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

Note: Results from Probit regression estimating Equation (4.2). The outcome variable in each column is the binary indicator of whether or not an individual spends at least ten minutes in physical activity on the diary day; Average marginal effects are reported. All columns include the full set of controls and state fixed effects from Column (7) of Tables 4.4 and 4.5. Sampling weights are applied and standard errors are clustered at the county-month-year level. Column (1) estimates Equation (4.1) for the full sample, Column (2) for individuals with BMI less than 30, and Column (3) for individuals with BMI greater than or equal to 30, where BMI status is a proxy for health.

Table 4.7, presents the time commuting coefficient estimated from Equation (4.2) comparing normal and overweight individuals to obese individuals. The coefficient for normal and overweight males, -.000975, indicates that an increase in commute time of ten minutes decreases the probability of participating in physical activity by nearly 0.10 percentage points or about 0.26% (=0.10/38) at mean activity levels for males. The coefficient is only slightly larger, -0.00105, for obese men and this difference is not significant ($\chi2$ = 0.11, p value = 0.7440). For normal and overweight women the coefficient is similar, -0.00158, implying an increase in commute time of ten minutes decreases the probability of participating in physical activity by 0.16 percentage points or about 0.42% (=0.16/38) at mean activity levels for females. In this case the coefficient is smaller and, as is the case for males, I cannot reject the hypothesis that this coefficient is the same for the two BMI groups ($\chi2$ = 0.98, p value = 0.3221).

Table 4.8: Time Spent in Sedentary Commuting and Time Spent in Physical Activity by BMI Group

	(1)	(2)	(3)
OLS	All BMI Groups	BMI < 30	BMI ≥ 30
Males			
Sedentary Commuting (minutes)	0.0468	0.0265	0.120
	(0.0577)	(0.0655)	(0.147)
Observations	1,379	1,019	360
R-squared	0.108	0.142	0.231
Females			
Sedentary Commuting (minutes)	-0.0335	-0.0885	0.145
	(0.0542)	(0.0622)	(0.107)
Observations	1,164	901	263
R-squared	0.091	0.100	0.396

Robust standard errors in parentheses.

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

Note: Results from Probit regression estimating Equation (4.3). The outcome variable in each column is the number of minutes an individual spends engaged in physical activity on the diary day. All columns include the full set of controls and state fixed effects from Column (7) of Tables 4.4 and 4.5. Sampling weights are applied and standard errors are clustered at the county-month-year level. Column (1) estimates Equation (4.1) for the full sample, Column (2) for individuals with BMI less than 30, and Column (3) for individuals with BMI greater than or equal to 30, where BMI status is a proxy for health.

Table 4.8 examines the time commuting coefficient estimated from Equation (4.3) assuming that the individuals do participate in some physical activity. This greatly reduces the number of observations and as a result, standard errors are relatively high and the coefficients are no longer statistically significant. The relationship that this estimation shows suggests that for those individuals who do commute, there is almost zero trade-off in minutes between different time uses. The coefficients are positive with confidence intervals centered near zero for all males and for obese females. Even though these coefficients are not themselves statistically significant, I do test if the difference between the coefficients for obese individuals against normal and overweight individuals is significant. For males, the $\chi 2$ distributed test statistic for this difference is 1.58, and for females the test statistic is 2.65, which corresponds to a p value of 0.1038, implying that I cannot reject the null hypothesis that these two coefficients are similar across BMI groups for both males and females. To test the robustness of these findings, I have also considered dropping individuals who engage in only active commuting. Doing this results in the signs of the coefficients staying the same, however some statistical significance is lost as the number of observations decreases and standard errors correspondingly increase.

Table 4.9 shows the results from an instrumental variables approach using the recursive bivariate Probit method described above in Equations (4.4) and (4.5). This method estimates a maximum likelihood of a joint distribution, which is more complicated than a standard Probit model, and had convergence issues when the full set of control variables was included. Because the control variables do not play a significant role in determining the coefficient on the commute variable as shown in Tables 4.4 and 4.5, they are not included in this estimation.

I have also estimated Equations (4.4) and (4.5) for women; however housing prices appear to be a very poor predictor of commuting behavior for women in the first stage. This may be due to differences in determinants of residential and job choices between men and women. (Turner and Niemeier, 1997; MacDonald, 1999). Appendix C includes an estimated model of the instrument predicting women's commuting behavior.

Table 4.9: Relationship between Sedentary Commuting Participation and Physical Activity Participation for Males, by BMI Category; IV Approach

	Males		
	(1)	(2)	(3)
	All BMI Groups	BMI < 30	BMI ≥ 30
Equation (4.5): dependent variable: *dCommute*			
1970 House Price/$10,000	-0.0107***	-0.0139***	-0.00575
	(0.00274)	(0.00329)	(0.00632)
Equation (4.4): dependent variable: *dPA*			
Sedentary Commute Indicator	-0.6105***	-0.6156***	0.4035***
	(0.0127)	(0.0147)	(0.0243)
ρ	0.9990***	0.9994***	-0.9891***
P-value of χ2 of ρ= 0	<0.001	<0.001	<0.001
χ2 of instrument	15.40***	18.21***	0.83
P-value	<0.001	<0.001	0.363
Sample Size	1,469	1,070	399

Robust standard errors in parentheses.

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

Note: Results from recursive Bivariate Probit regression estimating Equations (4.4) and (4.5). The outcome variable in each column is the number of minutes an individual spends engaged in physical activity on the diary day. Note: This table shows results of a recursive bivariate Probit model; the first two rows show coefficients and standard errors from estimation of Equation (4.5), and is the "first stage" regression, showing the effect of the instrument (past housing prices) on the endogenous predictor (participation in sedentary commuting). The second two rows show coefficients and standard errors from estimation of Equation (4.4); this is the estimate of the effect of the instrument (past housing prices) on the outcome of interest (physical activity participation). χ2 is a test of instrument strength in the first stage. All columns include the full set of controls and state fixed effects from Column (7) of Tables 4.4 and 4.5. Sampling weights are applied and standard errors are clustered at the county-month-year level. Column (1) estimates Equation (4.1) for the full sample, Column (2) for individuals with BMI less than 30, and Column (3) for individuals with BMI greater than or equal to 30, where BMI status is a proxy for health.

Because these equations are estimated simultaneously, there is no first-stage regression in the recursive bivariate Probit method. However, instrument strength can be estimated from the Equation (4.5) model. Because the outcome variables are binary and a Probit model is used, instead of the standard F-test for instrument strength in the first stage, a likelihood ratio test with a *χ2* distribution is typically used as a test for the predictive power of the instrument on the instrumented variable. In this case, the *χ2* values of 15.40 and 18.21 indicate a strong instrument in the male pooled BMI sample and in the male normal and overweight sample, however, this test indicates the instrument is a weak predictor of commuting for the obese males, as well as for women, as noted above. The second assumption a valid instrument must meet is the exclusion restriction; that the instrument (lagged housing prices) does not predict the outcome (physical activity) through any means other than through the endogenous predictor (commuting behavior). While it is not possible to test this empirically, the effect of lagged housing prices on current period physical activity may possibly be mediated through socioeconomic characteristics or characteristics of the built environment. With other data sources, future work could focus on disentangling these effects.

The average marginal effects (61%) estimated in Columns 1 and 2 of Table 4.9 are six to ten times as large as the more reasonable estimates from the un-instrumented models estimated for males in Table 4.6, Columns 1 and 2. Estimates in Table 4.9 represent in part an estimate of a local average treatment effect (LATE), or an estimate of the effect on only the “compliers”, those individuals who commute when house prices are lower and who do not commute in areas where house prices are higher(Angrist and Pischke, 2009). The large magnitude of these estimates may be due to the large, positive value of the correlation of the error terms in the two models, ρ. The high value of ρ in Columns 1 and 2 shows a high correlation between commuting and physical activity decisions, conditional on 1970 housing prices. The high value of ρ allows for a tighter distribution in fitted estimates of the commute indicator variable than in the observed values of those indicator variables. If a tighter distribution of fitted estimates of *dCommute* affects Equation (4.4), it could inflate the estimated effect of that variable on the physical activity indicator. As a result, it may be best to interpret the coefficients of the commuting indicator in Table 4.9 by considering a scenario where an individual has a 10% increase in the probability of commuting; this would decrease the probability of participating in physical activity by about 6% for normal and overweight men. For obese men, the estimated effect is much noisier and the point estimate is much smaller.

At first glance, this estimation continues to show a difference between normal and overweight group of men and the obese men; obese individuals continue to show less responsiveness to a change in the probability of commuting. However, when this model is estimated using only obese males, the instrument, housing prices from 1970, is no longer a good instrument for commuting behavior. This suggests the difference in the estimates of the effect of commuting on physical activity, as shown in Table 4.9, should not be used as evidence for the hypothesis that obese individuals respond to a change in commute time differently than overweight and normal weight individuals. On the other hand, the comparison of the estimate of the effect of housing prices from 1970 on commuting behavior between obese individuals and individuals who are overweight or normal weight suggests that the determinants of commuting behavior, and not just of physical activity, are different for obese individuals. However, this difference is not statistically significant.

The literature on commuting and physical activity examines men and women separately, but the obese and non-obese are typically pooled together. However, by not pooling them in Table 6, it is evident that lagged housing prices, which do predict commuting behavior for the pooled sample, are not predictive of commuting behavior for the obese group. These results suggest that perhaps the decisions related to commuting behavior, such as residential location, job choice, route and time of day of commute, may be different between not only women and men, but also between obese and non-obese individuals. This provides evidence that BMI, or health status, determines residential location and commuting behaviors, and supports the need for further examination of these relationships.

4.6. Discussion and Conclusions

Physical inactivity and overweight and obesity are important risk factors for chronic disease and, ultimately, for mortality. Fewer than half of US adults meet the US Center for Disease Control and Prevention recommendation of 150 minutes per week of moderate-intensity physical activity or 75 minutes per week vigorous-intensity physical activity (USDHHS, 2008). This chapter focuses on individual time-use to study the role of commuting in physical activity and pays particular attention to the overweight and obese, two groups of high risk.

The paper shows a correlation between sedentary commuting and decreased physical activity participation in non-obese individuals, and that there is a significant difference in this relationship between the non-obese and the obese. This work finds no relationship between sedentary commuting and physical activity participation among the obese. These relationships are robust to a number of different specifications. However, to strengthen the argument for causality in this work, it would be necessary to identify a stronger instrument for variations in commuting time. Possible solutions would be to use a larger lag in the housing prices, following the rationale that a larger lag would be less likely to be endogenous with current period behaviors yet would still be predictive of the built environment which affects commuting time. In future extensions of this work, I would like to use recently developed race-specific BMI cut-points for obesity, which take into account race-specific variation in cardiovascular and metabolic health risks. An additional area of work involves refining and comparing the definitions and calculations of both commuting time and physical activity in the ATUS according to various methods recommended in the literature; this could provide a clearer understanding of these relationships (Brown and Borisova, 2007; Kimbrough, Gray, 2015).

There are some additional limitations with this work, which stem from the nature of the ATUS data. First, it is a repeated cross-section, so actual changes in commuting time or in the amount of free time for a given individual cannot be observed, which would be possible with panel data. Second, the understanding of how joint time investment in home production activities or spouse's own participation in physical activity may affect an individual's physical activity participation is limited as ATUS does not provide spouse's and other family member's time-use data. Third, analyses are based on a single diary day, so it is not possible to address whether or not individuals who do not do physical activity on long days of commuting and working may substitute physical activity on shorter working days or weekend days. If time spent in various activities on a given day deviate from an individual's overall time-use patterns, then this would likely bias the coefficient estimates toward zero. Fourth, this work is unable to account for additional aspects of individual health behavior, such as food consumption behavior, alcohol consumption, and smoking behaviors. Finally, all data is self-reported and subject to some level of bias, although time-use diary data suffers less from this bias than other types of self-reported survey data, it is still possible that individuals with some level of health may under- or over-report various activities. More recent research has taken advantage of availability of accelerometer data for physical activity, mobile device data for transportation and commuting time, and measured health outcomes; however obtaining all of these measures in one dataset may be a challenge.

Previous research has focused on the relationships between urban sprawl or commuting and health outcomes such as obesity, chronic disease risk, and subjective health and well-being. Few papers examine the relationship between commuting and physical activity as a mechanism through which commuting behavior may impact these other health outcomes. This work adds to the literature on commuting and physical activity by examining the extensive and intensive margins of physical activity decisions using more precise measures of commuting time and physical activity through time-use data. This work also provides evidence for a causal relationship; that commuting time affects time spent in physical activity. Further, this work examines how this relationship varies by obesity status.

This prior research also does not address questions of how physical activity behavior in at-risk groups responds to variation in commuting. Policy makers proposing policies aimed at reducing obesity-related diseases might wish to target at-risk groups to thereby increase effectiveness of their policies. Pooling all groups reduces the ability to understand the health outcomes implications of public policy. By estimating a relationship for normal-weight, overweight, and obese individuals separately, this work shows the value of dis-aggregating the role of commuting behavior on physical activity decisions. In doing so, these results suggest that low levels of physical activity among the obese are not due to time constraints imposed by commuting; this raises questions about what is driving low levels of physical activity in this group. Table 4.3 shows the percentage of individuals involved in different types of commuting behavior is fairly even across BMI categories, as is the time spent in these activities, suggesting that the findings of differences in commuting and physical activity elasticities are not driven by differences in commuting patterns among the different BMI groups. This table also shows that the percentage of obese women in physically demanding jobs is slightly higher than the percentage of non-obese women in these types of jobs, suggesting a possibility that obese women in these occupations may be making more trade-offs between on-the-job activity and other types of physical activity. These results and summary statistics suggest that more research is needed to understand the factors leading to lower levels of physical activity among the obese. Existing literature shows that commuting time and health seeking behavior have an inverse relationship and suggest that urban planners include this in the design of transportation systems. This paper's results agree that these two factors are inversely related, but our results caution that the obese and overweight individuals may not respond in a significant way to commute-time-based incentives to increase physical activity. Further research is necessary to understand effective ways to increase physical activity time use in at-risk overweight and obese individuals.

Chapter 5

5.1 Discussion and Conclusions

5.1.1. Summary of Findings

This dissertation contributes to existing knowledge in the health economics of obesity, chronic disease risk, commuting behaviors and health behaviors. Specifically, Chapter 2 examines the effects on individual LDL cholesterol levels from *trans* fat reduction policies and provides a first look at individual-level effects of a nation-wide food regulation policy. This work finds significant improvements in LDL levels among more frequent restaurant meal consumers in the post-policy period. These findings suggest additional policies to further reduce *trans* fat at the national level should result in increased improvements in population cholesterol levels.

Chapters 3 and 4 together give complementary perspectives on the role of commuting behavior in obesity. Chapter 3 shows that increasing sedentary commutes do not lead to increased BMI, but instead that active commuting specifically is associated with lower levels of BMI. These findings contradict previous work which finds strong associations between sedentary commuting and higher BMI. This work reconciles these findings with another recently published piece of work which uses the same data and similar methods, and finds that much of the previously established relationship between commuting and obesity may be driven instead by a strong relationship between active commuting and healthier BMI in a small subset of the population. This work raises questions for health and urban policies which aim to reduce obesity through targeting commuting infrastructure and behaviors.

Chapter 4 examines physical activity participation and duration decisions and their relationship to commuting behavior. This chapter also explores heterogeneity by BMI status in this relationship. This work supports previous literature which shows a negative relationship between these two activities by showing that when facing less commuting, normal weight individuals respond with increasing their level of physical activity. However, this relationship does not persist among the obese; instead, obese individuals maintain low levels of physical activity regardless of commuting behavior. For normal- and overweight men, these findings are shown to be robust to an instrumental variables approach; however the instrument used does not explain commuting behavior for women or for obese men. This finding supports literature on differences in determinants of commuting behavior between sexes, however, more importantly, this finding provides new evidence that commuting determinants may differ by health status, suggesting that health may be determining commuting behaviors.

5.1.2. Policy Implications

Policy implications that can be drawn from Chapter 2 are that *trans* fat reduction policies and regulations have been successful in improving population health among consumers of commercially prepared foods. This work provides evidence that early policies such as information campaigns have slightly improved cholesterol levels, but that gradually implemented more stringent regulations have further improved cholesterol levels. This suggests that the proposed removal of GRAS status of partially hydrogenated vegetable oils should result in further improvements in population cholesterol levels among consumers of commercially prepared foods. These findings support interventions to regulate the food industry in an effort to provide more healthy prepared foods, including interventions such as sugar-sweetened beverage taxes and bans and sodium reduction regulations.

Policy implications suggested by Chapters 3 and 4 are primarily related to improvements in research from which public health policy and urban planning policy may be drawn. A growing area in public policy focuses on changing the urban structure or built environment to promote health. Many of these policies focus on increasing public transportation and active commuting; however findings from this work raise questions as to whether these policies would necessarily result in the desired level of improvements in population health. Specifically, both chapters highlight a need for further research to disentangle effects of different modes of commuting; this work separately considers active and sedentary modes of commuting and finds much of the relationship between poor health and commuting is driven by a strong relationship between active commuting and indicators of good health. While these findings support policies which promote active commuting, the work does not establish a causal relationship and active commuting itself may be an indicator of unobservable factors such as individual preference or motivation. So it is not clear whether promoting active commuting would necessarily then change other health behaviors which may themselves be contributing to the health benefits associated with active commuting. Both chapters highlight the complexity of the relationship between commuting choices and health and through this raise questions on the effectiveness of policies aimed at improving health through changes to the built environment and commuting behaviors.

5.1.3. Limitations and Future Work

A number of limitations are noted here. First, Chapter 2 faces complications with identifying a clear pre- and post-policy period and therefore cannot disentangle effects of any one particular policy, but instead assesses the effects of the combination of *trans* fat reduction policies, all of which contributed to reductions in *trans* fat content of food prepared away from home. Because individuals consume foods from a variety of different sources outside of a laboratory setting, and because of the multitude of overlapping policies and changes in food formulation, this work suggests that research attempting to assess the effects on any one particular policy at the national level may possibly over-estimate its effects. A second limitation of this work is in identifying the treated group; consumers of restaurant meals are considered to be the most exposed to *trans*-fat-containing foods; this work shows these individuals are also more likely to consume other types of FAFH, however a clearer identification of overall food consumption provided by food diary data may provide an improved estimate of effects of *trans* fat reduction policies. Another key limitation of this data is the lack of geographic variables available – because many *trans* fat policies were implemented at the local level, a difference-in-difference approach could be used to single out effectiveness of particular policies relative to changes in the overall food environment. At the national level, controlling for geographic variation might reduce noise in the estimates in this work because geographic variation exists in health and eating behaviors. This work does not attempt to assess cost-effectiveness of these policies, however with the addition of geographic data and food pricing or cost data, more work could be done to asses cost-effectiveness of individual policies.

Future work in this area could build upon this work by identifying health effects of the various *trans* fat substitutes, identifying health effects among the most at-risk groups, and assessing whether the improvements in health from these policies have outweighed any negative compensatory health behaviors.

Limitations in Chapters 3 and 4 stem from the data available. First, general limitations of the data are that it is cross-sectional, so changes in commuting behavior for an individual cannot be observed; panel data would allow for such observations. Second, individual time-use may be co-determined by time-use of other individuals within the household; joint time investments in home production activities or spouse's participation in physical activity may affect an individual's physical participation and other health behaviors. However time diary data for only one individual within a household is collected in the ATUS, so it is not possible to take these important and interesting factors fully into account. Another general limitation of this data is that it is self-reported. It is argued that time-use diary data suffers less from reporting bias and social desirability bias than other types of self-reported survey data, however it is still possible that individuals with differing levels of health may over- or under-report participation in various activities. Future research could examine whether this bias exists in time-use diary data. Two key determinants of BMI are diet and physical activity; while time-use data is arguably less prone to social desirability bias in physical activity measures than surveys which ask specific questions about physical activity participation, the dataset has minimal information from which to determine the types of food an individual consumes as well as and other health habits. Plausibly, healthfulness of foods eaten could be inferred from time spent in food preparation activities, however if a home-cooked meal was prepared by somebody other than the respondent, this information would not be captured. Similarly, data is available on where and with whom the respondent ate meals, however it is entirely likely that individuals could eat commercially prepared foods at home or with others and could plausibly take home-cooked meals elsewhere for consumption, so assumptions made about healthfulness of foods based on location could be biased. In the Eating and Health Module, additional data are available on eating as a secondary activity, or grazing as it is commonly known. While much of this data is useful in understanding eating behaviors, without any information on what is being eaten, this understanding remains limited. Other health behaviors not included in this data are smoking and alcohol consumption behaviors; both of these have impacts on BMI, so may limit findings of this work. There are also limitations in the measurement of commuting itself; while this measure is arguably a more accurate assessment of individual commuting behavior than simply using MSA-level or county-level indicators of sprawl, this measure of commuting introduces bias to the results. Specifically, this measure of calculating commute times in ATUS data fails to identify commuting trips for individuals who make stops along the way home from work. If individuals who make stops along the way represent a subset of individuals who are more time-sensitive or have a preference for reducing time spent running errands in an effort to leave more time for health-producing behaviors, then this would result in biased estimates. Particularly, if these individuals generally have lower BMI, then underestimating their commuting time would provide an overestimate of the relationship with BMI. Similarly, if these individuals are more likely to participate in physical activity, then underestimating their commute time could result in an overestimate of the effect of commuting behavior on physical activity behaviors. This work makes note of this limitation and possible bias. Future work could provide a better understanding of how this bias may affect health or health behaviors by comparing these results using different methods of calculating commuting times within the ATUS data.

Chapter 3 specifically faces the following limitations, first no causality can be inferred in this relationship; commuting may be affecting BMI or factors which affect BMI such as preferences for exercising or cooking and eating healthier foods might in turn be affecting factors related to commuting decisions such as residential location, job location, and time and mode of commuting to and from work. While the highlight of this chapter is identification of a strong relationship between active commuting and decreased BMI, this work is unable to identify whether active commuting itself is a proxy for preferences for healthy behaviors. Although among US commuters cycling and walking make up only 7% of commuting trips, this effect seems to be driving the results of unhealthy outcomes related to sedentary commuting. More work is needed to identify whether it is actually the walking and cycling itself which leads to improvements in health or, more likely, what other factors drive these individuals to choose healthier behaviors. Policies targeted at changing individual behavior, also known as "nudge" policies, are increasing in popularity, however their overall effectiveness, particularly for the most at-risk groups, may lie in improved understanding of what behaviors lead to better health and how to changes those behaviors among individuals who have the most to gain from those changes.

In addition to those outlined above, Chapter 4 faces additional limitations: First, because the analysis is based on a single diary day, it is not possible to address whether individuals who do not do physical activity on long days of commuting and working are in fact substituting physical activity on shorter working days or non-working days; additional data may be needed for understanding such issues. Second, and perhaps of more concern, although this work attempts to identify a causal relationship between commuting behavior and physical activity; the fact that the instrument used to predict commuting behavior is predictive only for men with BMI less than 30 suggests that commuting behavior and health behaviors may be co-determined by some other unobservable factors or that BMI itself may be responsible to some degree for commuting decisions. Especially because of the recent emphasis on using urban design to affect health, more research is needed not only to disentangle these effects and identify direction of causality, but also to understand determinants of commuting and health behaviors among particularly at-risk groups like the obese. Similarly, determinants of commuting behavior are markedly different for men than women; more research is needed to understand determinants of women's commuting behaviors and how they relate to engagement in health-producing behaviors like physical activity.

This dissertation presents research on common health behaviors and choices people face. It has found that there is a role for careful, considered study of heterogeneity in choices and in outcomes. When this heterogeneity is considered, positive outcomes of policy can be easier to see, as in Chapter 2, as can the difficulty of changing behaviors in at risk individuals as in Chapters 3 and 4. These results highlight a number of challenges remaining in using survey data to understand health behaviors related to obesity and chronic disease risk. Policies seeking to alter nutritional quality of food should seek research which targets subgroups with higher risk; similarly policies seeking to shape the built environment or promote healthy behaviors should consider the effects of these policies on specific at-risk groups.

References

Adler, N.E., Ostrove, J.M., 1999. Socioeconomic Status and Health: What We Know and What We Don't. Annals of the New York Academy of Sciences 896, 3–15. doi:10.1111/j.1749-6632.1999.tb08101.x

Ainsworth, B. E. et al., 2000. Compendium of physical activities: an update of activity codes and MET intensities. Medicine and science in sports and exercise 32, S498–S504.

Allison, D.B., Egan, S.K., Barraj, L.M., Caughman, C., Infante, M., Heimbach, J.T., 1999. Estimated intakes of trans fatty and other fatty acids in the US population. J Am Diet Assoc 99, 166-174-176. doi:10.1016/S0002-8223(99)00041-3

Allison DB, Fontaine KR, Manson JE, Stevens J, VanItallie TB, 1999. Annual Deaths Attributable to Obesity in the United States. JAMA 282, 1530–1538. doi:10.1001/jama.282.16.1530

Amemiya, T., 1985. Advanced Econometrics, 1 edition. ed. Harvard University Press, Cambridge, Mass.

American Heart Association, 2010. A History of Trans Fat [WWW Document]. URL http://www.heart.org/HEARTORG/GettingHealthy/FatsAndOils/Fats101/A-History-of-Trans-Fat_UCM_301463_Article.jsp (accessed 3.5.13).

Angell, S.Y., Cobb, L.K., Curtis, C.J., Konty, K.J., Silver, L.D., 2012. Change in Trans Fatty Acid Content of Fast-Food Purchases Associated With New York City's Restaurant RegulationA Pre–Post Study. Ann Intern Med 157, 81–86. doi:10.7326/0003-4819-157-2-201207170-00004

Angrist, J.D., Pischke, J.-S., 2009. Mostly Harmless Econometrics: An Empiricist's Companion, 1 edition. ed. Princeton University Press, Princeton.

Antecol, H., Bedard, K., 2006. Unhealthy assimilation: Why do immigrants converge to American health status levels? Demography 43, 337–360. doi:10.1353/dem.2006.0011

Arrieta, A., Russell, L.B., 2008. Effects of Leisure and Non-Leisure Physical Activity on Mortality in U.S. Adults over Two Decades. Annals of Epidemiology 18, 889–895. doi:10.1016/j.annepidem.2008.09.007

Ascherio, A., Hennekens, C.H., Buring, J.E., Master, C., Stampfer, M.J., Willett, W.C., 1994. Trans-fatty acids intake and risk of myocardial infarction. Circulation 89, 94–101. doi:10.1161/01.CIR.89.1.94

Ascherio, A., Willett, W.C., 1997. Health effects of trans fatty acids. Am J Clin Nutr 66, 1006S–1010S.

Associated Press, 2007. McDonald's finally picks trans-fat-free oil [WWW Document]. msnbc.com. URL http://www.nbcnews.com/id/16873869/ns/health-diet_and_nutrition/t/mcdonalds-finally-picks-trans-fat-free-oil/ (accessed 4.11.13).

Autor, D.H., 2003. Outsourcing at Will: The Contribution of Unjust Dismissal Doctrine to the Growth of Employment Outsourcing. Journal of Labor Economics 21, 1–42. doi:10.1086/344122

Baer, D.J., Judd, J.T., Clevidence, B.A., Tracy, R.P., 2004. Dietary fatty acids affect plasma markers of inflammation in healthy men fed controlled diets: a randomized crossover study. Am J Clin Nutr 79, 969–973.

Bassett, C.M.C., Edel, A.L., Patenaude, A.F., McCullough, R.S., Blackwood, D.P., Chouinard, P.Y., Paquin, P., Lamarche, B., Pierce, G.N., 2010. Dietary Vaccenic Acid Has Antiatherogenic Effects in LDLr−/− Mice. J. Nutr. 140, 18–24. doi:10.3945/jn.109.105163

Baum-Snow, N., 2007. Did Highways Cause Suburbanization? The Quarterly Journal of Economics 122, 775–805.

Becker, G.S., 1965. A Theory of the Allocation of Time. The Economic Journal 75, 493–517. doi:10.2307/2228949

Brown, H., Roberts, J., 2011. Exercising choice: the economic determinants of physical activity behaviour of an employed population. Soc Sci Med 73, 383–390. doi:10.1016/j.socscimed.2011.06.001

Brown, Cheryl, and Tatiana Borisova, 2007. Understanding commuting and grocery shopping using the American Time Use Survey.

Brueckner, J.K., Fansler, D.A., 1983. The Economics of Urban Sprawl: Theory and Evidence on the Spatial Sizes of Cities. The Review of Economics and Statistics 65, 479–482. doi:10.2307/1924193

Bureau of Labor Statistics, U.S. Census Bureau, 2015. American Time Use Survey User's Guide: Understanding ATUS 2003 to 2014.

Burkhauser, R.V., Cawley, J., 2008. Beyond BMI: The value of more accurate measures of fatness and obesity in social science research. Journal of Health Economics 27, 519–529. doi:10.1016/j.jhealeco.2007.05.005

Burkhauser, R.V., Cawley, J., Schmeiser, M.D., 2009. The timing of the rise in U.S. obesity varies with measure of fatness. Economics & Human Biology 7, 307–318. doi:10.1016/j.ehb.2009.07.006

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. doi:10.1016/j.jhealeco.2015.03.001

Cawley, J., 2004. An economic framework for understanding physical activity and eating behaviors. American Journal of Preventive Medicine 27, 117–125. doi:10.1016/j.amepre.2004.06.012

Cawley, J., Liu, F., 2007. Maternal Employment and Childhood Obesity: A Search for Mechanisms in Time Use Data (Working Paper No. 13600). National Bureau of Economic Research.

Cawley, J., Ruhm, C., 2011. The Economics of Risky Health Behaviors (Working Paper No. 17081). National Bureau of Economic Research.

Centers for Disease Control and Prevention (CDC), 2002. NHANES 1999 - 2000: Cholesterol - LDL & Triglycerides Data Documentation, Codebook, and Frequencies.

Chardigny, J.-M., Destaillats, F., Malpuech-Brugère, C., Moulin, J., Bauman, D.E., Lock, A.L., Barbano, D.M., Mensink, R.P., Bezelgues, J.-B., Chaumont, P., Combe, N., Cristiani, I., Joffre, F., German, J.B., Dionisi, F., Boirie, Y., Sébédio, J.-L., 2008. Do trans fatty acids from industrially produced sources and from natural sources have the same effect on cardiovascular disease risk factors in healthy subjects? Results of the trans Fatty Acids Collaboration (TRANSFACT) study. Am J Clin Nutr 87, 558–566.

Cragg, J.G., 1971. Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods. Econometrica 39, 829–844. doi:10.2307/1909582

Crane, R., Chatman, D.G., 2003. As Jobs Sprawl, Whither the Commute? ACCESS Magazine 1.

Dickerson, A., Hole, A.R., Munford, L.A., 2014. The relationship between well-being and commuting revisited: Does the choice of methodology matter? Regional Science and Urban Economics 49, 321–329. doi:10.1016/j.regsciurbeco.2014.09.004

Dishman, R.K., Sallis, J.F., Orenstein, D.R., 1985. The determinants of physical activity and exercise. Public Health Rep 100, 158–171.

Drewnowski, A., 2004. Obesity and the food environment. American Journal of Preventive Medicine 27, 154–162. doi:10.1016/j.amepre.2004.06.011

Drewnowski, A., Darmon, N., 2005. The economics of obesity: dietary energy density and energy cost. Am J Clin Nutr 82, 265S–273S.

Dunton, G.F., Berrigan, D., Ballard-Barbash, R., Graubard, B., Atienza, A.A., 2009. Joint associations of physical activity and sedentary behaviors with body mass index: results from a time use survey of US adults. Int J Obes (Lond) 33, 1427–1436. doi:10.1038/ijo.2009.174

Dutton, D.J., McLaren, L., 2011. Explained and Unexplained Regional Variation in Canadian Obesity Prevalence. Obesity 19, 1460–1468. doi:10.1038/oby.2010.339

Dwyer-Lindgren, L., Flaxman, A.D., Ng, M., Hansen, G.M., Murray, C.J.L., Mokdad, A.H., 2015. Drinking Patterns in US Counties From 2002 to 2012. American Journal of Public Health 105, 1120–1127. doi:10.2105/AJPH.2014.302313

Eckel, R.H., Borra, S., Lichtenstein, A.H., Yin-Piazza, S.Y., 2007. Understanding the Complexity of Trans Fatty Acid Reduction in the American Diet American Heart Association Trans Fat Conference 2006: Report of the Trans Fat Conference Planning Group. Circulation 115, 2231–2246. doi:10.1161/CIRCULATIONAHA.106.181947

Eid, J., Overman, H.G., Puga, D., Turner, M.A., 2008. Fat city: Questioning the relationship between urban sprawl and obesity. Journal of Urban Economics 63, 385–404. doi:10.1016/j.jue.2007.12.002

Ewing, R., Meakins, G., Hamidi, S., Nelson, A.C., 2014. Relationship between urban sprawl and physical activity, obesity, and morbidity – Update and refinement. Health & Place 26, 118–126. doi:10.1016/j.healthplace.2013.12.008

Ewing, R., Pendall, R., Chen, D., 2003. Measuring Sprawl and Its Transportation Impacts. Transportation Research Record: Journal of the Transportation Research Board 1831, 175–183. doi:10.3141/1831-20

Ewing, R., Schmid, T., Killingsworth, R., Zlot, A., Raudenbush, S., 2003. Relationship between urban sprawl and physical activity, obesity, and morbidity. American Journal of Health Promotion 18, 47–57.

Finkelstein, E.A., Trogdon, J.G., Cohen, J.W., Dietz, W., 2009. Annual Medical Spending Attributable To Obesity: Payer-And Service-Specific Estimates. Health Aff 28, w822–w831. doi:10.1377/hlthaff.28.5.w822

Flegal, K.M., Carroll, M.D., Ogden, C.L., Curtin, L.R., 2010. Prevalence and Trends in Obesity Among US Adults, 1999-2008. JAMA: The Journal of the American Medical Association 303, 235. doi:10.1001/jama.2009.2014

Flint, E., Cummins, S., Sacker, A., 2014. Associations between active commuting, body fat, and body mass index: population based, cross sectional study in the United Kingdom. BMJ 349, g4887. doi:10.1136/bmj.g4887

Foley, L., Panter, J., Heinen, E., Prins, R., Ogilvie, D., 2015. Changes in active commuting and changes in physical activity in adults: a cohort study. International Journal of Behavioral Nutrition and Physical Activity 12, 161. doi:10.1186/s12966-015-0323-0

Ford, E.S., Mokdad, A.H., Giles, W.H., Galuska, D.A., Serdula, M.K., 2005. Geographic Variation in the Prevalence of Obesity, Diabetes, and Obesity-Related Behaviors**. Obesity Research 13, 118–122. doi:10.1038/oby.2005.15

Frank, L.D., Andresen, M.A., Schmid, T.L., 2004. Obesity relationships with community design, physical activity, and time spent in cars. American Journal of Preventive Medicine 27, 87–96. doi:10.1016/j.amepre.2004.04.011

Frank, R.H., 1978. Why Women Earn Less: The Theory and Estimation of Differential Overqualification. The American Economic Review 68, 360–373.

Frazis, H., Stewart, J., 2012. How to Think about Time-Use Data: What Inferences Can We Make about Long- and Short-Run Time Use from Time Diaries? Annals of Economics and Statistics 231–245.

Fryar, C., Ervin, R., 2013. Caloric intake from fast food among adults: United States, 2007-2010. 2013 NCHS data brief.

Gicheva, D., Hastings, J., Villas-Boas, S., 2010. Investigating Income Effects in Scanner Data: Do Gasoline Prices Affect Grocery Purchases? American Economic Review 100, 480–484. doi:10.1257/aer.100.2.480

Giles-Corti, B., Donovan, R.J., 2002. The relative influence of individual, social and physical environment determinants of physical activity. Social Science & Medicine 54, 1793–1812. doi:10.1016/S0277-9536(01)00150-2

Giménez, J.I., Molina, J.A., 2011. Commuting Time and Labour Supply: A Causal Effect? (SSRN Scholarly Paper No. ID 1771251). Social Science Research Network, Rochester, NY.

Giuliano, G., Dargay, J., 2006. Car ownership, travel and land use: a comparison of the US and Great Britain. Transportation Research Part A: Policy and Practice 40, 106–124. doi:10.1016/j.tra.2005.03.002

Giuliano, G., Narayan, D., 2003. Another Look at Travel Patterns and Urban Form: The US and Great Britain. Urban Stud 40, 2295–2312. doi:10.1080/0042098032000123303

Glaeser, E.L., Kahn, M.E., Rappaport, J., 2008. Why do the poor live in cities? The role of public transportation. Journal of Urban Economics 63, 1–24. doi:10.1016/j.jue.2006.12.004

Goldman, T., 2015. Health Policy Brief: The FDA's Menu-Labeling Rule. Health Affairs.

Gordon, D.J., Trost, D.C., Hyde, J., Whaley, F.S., Hannan, P.J., Jacobs, D.R., Ekelund, L.G., 1987. Seasonal cholesterol cycles: the Lipid Research Clinics Coronary Primary Prevention Trial placebo group. Circulation 76, 1224–1231. doi:10.1161/01.CIR.76.6.1224

Gordon-Larsen P, Boone-Heinonen J, Sidney S, Sternfeld B, Jacobs DR, Jr, Lewis CE, 2009. Active commuting and cardiovascular disease risk: The cardia study. Arch Intern Med 169, 1216–1223. doi:10.1001/archinternmed.2009.163

Greene, W.H., 2011. Econometric Analysis, 7 edition. ed. Pearson, Boston.

Grossman, M., 1999. The Human Capital Model of the Demand for Health (Working Paper No. 7078). National Bureau of Economic Research.

Grossman, M., 1972. On the Concept of Health Capital and the Demand for Health. Journal of Political Economy 80, 223–55.

Gu, Q., Paulose-Ram, R., Burt, V.L., Kit, B.K., 2014. Prescription Cholesterol-lowering Medication Use in Adults Aged 40 and Over: United States, 2003–2012 (No. No. 177), NCHS Data Brief. National Center for Health Statistics, Hyattsville, MD.

Guenther, P., Bowman, S., Goldman, J., 2010. Alcoholic Beverage Consumption by Adults 21 Years and Over in the United States: Results from the National Health and Nutrition Examination Survey, 2003-2006 (Technical Report). Center for Nutrition Policy and Promotion, and Agricultural Research Service, U.S. Department of Agriculture.

Guthrie, J.F., Lin, B.-H., Frazao, E., 2002. Role of Food Prepared Away from Home in the American Diet, 1977-78 versus 1994-96: Changes and Consequences. Journal of Nutrition Education and Behavior 34, 140–150. doi:10.1016/S1499-4046(06)60083-3

Hansson, E., Mattisson, K., Björk, J., Östergren, P.-O., Jakobsson, K., 2011. Relationship between commuting and health outcomes in a cross-sectional population survey in southern Sweden. BMC Public Health 11, 834. doi:10.1186/1471-2458-11-834

Harris, T.B., Ballard-Barbasch, R., Madans, J., Makuc, D.M., Feldman, J.J., 1993. Overweight, Weight Loss, and Risk of Coronary Heart Disease in Older Women The NHANES I Epidemiologic Follow-up Study. Am. J. Epidemiol. 137, 1318–1327.

Hartog, J.J. de, Boogaard, H., Nijland, H., Hoek, G., 2010. Do the Health Benefits of Cycling Outweigh the Risks? Environmental Health Perspectives 118, 1109–1116.

Hemmingsson, E., Ekelund, U., 2006. Is the association between physical activity and body mass index obesity dependent? Int J Obes 31, 663–668. doi:10.1038/sj.ijo.0803458

Hoehner, C.M., Barlow, C.E., Allen, P., Schootman, M., 2012. Commuting Distance, Cardiorespiratory Fitness, and Metabolic Risk. Am J Prev Med 42, 571–578. doi:10.1016/j.amepre.2012.02.020

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, 968–977. doi:10.1161/01.CIR.67.5.968

Humphreys, B., Ruseski, J., 2009. The Economics of Participation and Time Spent in Physical Activity (Working Paper No. 2009–9). University of Alberta, Department of Economics.

Humphreys, Brad R, Ruseski, Jane E, 2011. An Economic Analysis of Participation and Time Spent in Physical Activity. The B.E. Journal of Economic Analysis & Policy 11, 1–38.

Hunter, J.E., 2005. Dietary levels of trans-fatty acids: basis for health concerns and industry efforts to limit use. Nutrition Research 25, 499–513. doi:10.1016/j.nutres.2005.04.002

Huxley, R., Mendis, S., Zheleznyakov, E., Reddy, S., Chan, J., 2009. Body mass index, waist circumference and waist:hip ratio as predictors of cardiovascular risk—a review of the literature. Eur J Clin Nutr 64, 16–22. doi:10.1038/ejcn.2009.68

Hymel, K., 2009. Does traffic congestion reduce employment growth? Journal of Urban Economics 65, 127–135. doi:10.1016/j.jue.2008.11.002

Ingraham, C., 2016. The astonishing human potential wasted on commutes [WWW Document]. Washington Post. URL https://www.washingtonpost.com/news/wonk/wp/2016/02/25/how-much-of-your-life-youre-wasting-on-your-commute/ (accessed 6.10.16).

Institute for Health Metrics and Evaluation (IHME), 2013. GBD Compare [WWW Document]. URL http://www.healthdata.org/data-visualization/gbd-compare (accessed 5.11.15).

Jacobson, S.H., King, D.M., Yuan, R., 2011. A note on the relationship between obesity and driving. Transport Policy 18, 772–776. doi:10.1016/j.tranpol.2011.03.008

Jakobsen, M.U., Overvad, K., Dyerberg, J., Heitmann, B.L., 2008. Intake of ruminant trans fatty acids and risk of coronary heart disease. Int. J. Epidemiol. 37, 173–182. doi:10.1093/ije/dym243

Jilcott, S.B., Moore, J.B., Wall-Bassett, E.D., Liu, H., Saelens, B.E., 2011. Association between Travel Times and Food Procurement Practices among Female Supplemental Nutrition Assistance Program Participants in Eastern North Carolina. Journal of Nutrition Education and Behavior 43, 385–389. doi:10.1016/j.jneb.2010.11.004

Johnson, C.L., Paulose-Ram, R., Ogden, C.L., Carroll, M.D., Kruszon-Moran, D., Dohrmann, S.M., Curtin, L.R., 2013. National Health and Nutrition Examination Survey: Analytic Guidelines, 1999–2010.

Jones, A.M., 2000. Chapter 6 Health econometrics, in: Newhouse, A.J.C. and J.P. (Ed.), Handbook of Health Economics. Elsevier, pp. 265–344.

Joshi, P., Martin, S., Blaha, M., McEvoy, J., Santos, R., Cannon, C., Blumenthal, R., Jones, S., 2014. SEASONAL VARIATIONS IN LIPID PROFILES FROM 2.8 MILLION US ADULTS: THE VERY LARGE DATABASE OF LIPIDS (VLDL 14). J Am Coll Cardiol 63. doi:10.1016/S0735-1097(14)61458-3

Kant, A.K., Graubard, B.I., 2004. Eating out in America, 1987–2000: trends and nutritional correlates. Preventive Medicine 38, 243–249. doi:10.1016/j.ypmed.2003.10.004

Katan, M.B., 2006. Regulation of trans fats: The gap, the Polder, and McDonald's French fries. Atherosclerosis Supplements, First International Symposium on Trans Fatty Acids and Health Rungstedgaard, Rungsted Kyst, Denmark 7, 63–66. doi:10.1016/j.atherosclerosissup.2006.04.013

Katzmarzyk, P.T., Janssen, I., 2004. The Economic Costs Associated With Physical Inactivity and Obesity in Canada: An Update. Can. J. Appl. Physiol. 29, 90–115. doi:10.1139/h04-008

Kelly-Schwartz, A.C., Stockard, J., Doyle, S., Schlossberg, M., 2004. Is Sprawl Unhealthy? A Multilevel Analysis of the Relationship of Metropolitan Sprawl to the Health of Individuals. Journal of Planning Education and Research 24, 184–196. doi:10.1177/0739456X04267713

Kimbrough, Gray, 2015. Measuring Commuting in the American Time Use Survey.

Kirk, S.F.L., Penney, T.L., McHugh, T.-L.F., 2010. Characterizing the obesogenic environment: the state of the evidence with directions for future research. Obesity Reviews 11, 109–117. doi:10.1111/j.1467-789X.2009.00611.x

Kodali, D.R., 2014. Trans Fats Replacement Solutions. Elsevier.

Kraus, J.F., Borhani, N.O., Franti, C.E., 1980. Socioeconomic Status, Ethnicity, and Risk of Coronary Heart Disease. Am. J. Epidemiol. 111, 407–414.

Kris-Etherton, P.M., Lefevre, M., Mensink, R.P., Petersen, B., Fleming, J., Flickinger, B.D., 2012. Trans fatty acid intakes and food sources in the U.S. population: NHANES 1999-2002. Lipids 47, 931–940. doi:10.1007/s11745-012-3704-z

Kuczmarski, M.F., Kuczmarski, R.J., Najjar, M., 2001. Effects of age on validity of self-reported height, weight, and body mass index: findings from the Third National Health and Nutrition Examination Survey, 1988-1994. J Am Diet Assoc 101, 28-34-36. doi:10.1016/S0002-8223(01)00008-6

Künn-Nelen, A., 2016. Does Commuting Affect Health? Health Econ. 25, 984–1004. doi:10.1002/hec.3199

Ladabaum, U., Mannalithara, A., Myer, P.A., Singh, G., 2014. Obesity, Abdominal Obesity, Physical Activity, and Caloric Intake in US Adults: 1988 to 2010. The American Journal of Medicine 127, 717–727.e12. doi:10.1016/j.amjmed.2014.02.026

Lakdawalla, D., Philipson, T., 2009. The growth of obesity and technological change. Economics & Human Biology 7, 283–293. doi:10.1016/j.ehb.2009.08.001

Lee, I.-M., Shiroma, E.J., Lobelo, F., Puska, P., Blair, S.N., Katzmarzyk, P.T., 2012. Impact of Physical Inactivity on the World's Major Non-Communicable Diseases. Lancet 380, 219–229. doi:10.1016/S0140-6736(12)61031-9

Lee, R.E., McAlexander, K., Banda, J., 2011. Reversing the Obesogenic Environment. Human Kinetics.

Leuven, E., Oosterbeek, H., 2011. Chapter 3 - Overeducation and Mismatch in the Labor Market1, in: Eric A. Hanushek, S.M. and L.W. (Ed.), Handbook of the Economics of Education, Handbook of The Economics of Education. Elsevier, pp. 283–326.

Li, C., Engström, G., Hedblad, B., Calling, S., Berglund, G., Janzon, L., 2006. Sex differences in the relationships between BMI, WHR and incidence of cardiovascular disease: a population-based cohort study. Int J Obes 30, 1775–1781. doi:10.1038/sj.ijo.0803339

Lichtenstein, A.H., Ausman, L.M., Jalbert, S.M., Schaefer, E.J., 1999. Effects of Different Forms of Dietary Hydrogenated Fats on Serum Lipoprotein Cholesterol Levels. New England Journal of Medicine 340, 1933–1940. doi:10.1056/NEJM199906243402501

Lichtman, S.W., Pisarska, K., Berman, E.R., Pestone, M., Dowling, H., Offenbacher, E., Weisel, H., Heshka, S., Matthews, D.E., Heymsfield, S.B., 1992. Discrepancy between Self-Reported and Actual Caloric Intake and Exercise in Obese Subjects. New England Journal of Medicine 327, 1893–1898. doi:10.1056/NEJM199212313272701

Lin, B.-H., Guthrie, J.F., 2012. Nutritional Quality of Food Prepared at Home and Away From Home, 1977-2008 (Economic Information Bulletin No. 142361). United States Department of Agriculture, Economic Research Service.

Lindström, M., 2008. Means of transportation to work and overweight and obesity: A population-based study in southern Sweden. Preventive Medicine 46, 22–28. doi:10.1016/j.ypmed.2007.07.012

Lopez, R., 2004. Urban Sprawl and Risk for Being Overweight or Obese. Am J Public Health 94, 1574–1579. doi:10.2105/AJPH.94.9.1574

Lopez-Garcia, E., Schulze, M.B., Meigs, J.B., Manson, J.E., Rifai, N., Stampfer, M.J., Willett, W.C., Hu, F.B., 2005. Consumption of Trans Fatty Acids Is Related to Plasma Biomarkers of Inflammation and Endothelial Dysfunction. J. Nutr. 135, 562–566.

Lopez-Zetina, J., Lee, H., Friis, R., 2006. The link between obesity and the built environment. Evidence from an ecological analysis of obesity and vehicle miles of travel in California. Health & Place 12, 656–664. doi:10.1016/j.healthplace.2005.09.001

Luepker, R.V., Rosamond, W.D., Murphy, R., Sprafka, J.M., Folsom, A.R., McGovern, P.G., Blackburn, H., 1993. Socioeconomic status and coronary heart disease risk factor trends. The Minnesota Heart Survey. Circulation 88, 2172–2179. doi:10.1161/01.CIR.88.5.2172

Ma, Y., Olendzki, B.C., Li, W., Hafner, A.R., Chiriboga, D., Hebert, J.R., Campbell, M., Sarnie, M., Ockene, I.S., 2005. Seasonal variation in food intake, physical activity, and body weight in a predominantly overweight population. Eur J Clin Nutr 60, 519–528. doi:10.1038/sj.ejcn.1602346

Macdiarmid, J., Blundell, J., 1998. Assessing dietary intake: Who, what and why of under-reporting. Nutrition Research Reviews 11, 231–253. doi:10.1079/NRR19980017

MacDonald, H.I., 1999. Women's Employment and Commuting: Explaining the Links. Journal of Planning Literature 13, 267–283. doi:10.1177/08854129922092397

Madden, J.F., 1981. Why Women Work Closer to Home. Urban Stud 18, 181–194. doi:10.1080/00420988120080341

Magri, T.P.R., Fernandes, F.S., Souza, A.S., Langhi, L.G.P., Barboza, T., Misan, V., Mucci, D.B., Santos, R.M., Nunes, T.F., Souza, S.A.L., de Mello Coelho, V., Tavares do Carmo, M. das G., 2015. Interesterified fat or palm oil as substitutes for partially hydrogenated fat in maternal diet can predispose obesity in adult male offspring. Clinical Nutrition 34, 904–910. doi:10.1016/j.clnu.2014.09.014

Mann, D., Reynolds, K., Smith, D., Muntner, P., 2008. Trends in Statin Use and Low-Density Lipoprotein Cholesterol Levels Among US Adults: Impact of the 2001 National Cholesterol Education Program Guidelines. Ann Pharmacother 42, 1208–1215. doi:10.1345/aph.1L181

Manson JE, Skerrett PJ, Greenland P, VanItallie TB, 2004. The escalating pandemics of obesity and sedentary lifestyle: A call to action for clinicians. Arch Intern Med 164, 249–258. doi:10.1001/archinte.164.3.249

Martin, A., Ogilvie, D., Suhrcke, M., 2014. Evaluating causal relationships between urban built environment characteristics and obesity: a methodological review of observational studies. International Journal of Behavioral Nutrition and Physical Activity 11, 142. doi:10.1186/s12966-014-0142-8

Martin, A., Panter, J., Suhrcke, M., Ogilvie, D., 2015. Impact of changes in mode of travel to work on changes in body mass index: evidence from the British Household Panel Survey. J Epidemiol Community Health jech-2014-205211. doi:10.1136/jech-2014-205211

Matthews, C.E., Freedson, P.S., Hebert, J.R., Stanek, E.J., Merriam, P.A., Rosal, M.C., Ebbeling, C.B., Ockene, I.S., 2001. Seasonal Variation in Household, Occupational, and Leisure Time Physical Activity: Longitudinal Analyses from the Seasonal Variation of Blood Cholesterol Study. Am. J. Epidemiol. 153, 172–183. doi:10.1093/aje/153.2.172

McCann, B.A., Ewing, R., 2003. Measuring the Health Effects of Sprawl: A NationalAnalysis of Physical Activity, Obesity, and Chronic Disease, Surface Transportation Policy Project. Smart Growth America.

McCormack, G.R., Virk, J.S., 2014. Driving towards obesity: A systematized literature review on the association between motor vehicle travel time and distance and weight status in adults. Preventive Medicine 66, 49–55. doi:10.1016/j.ypmed.2014.06.002

Mckenzie, B., Rapino, M., 2011. Commuting in the United States: 2009 (No. ACS-15), American Community Survey Reports. US Department of Commerce, Economics and Statistics Administration, US Census Bureau.

Mensink, R.P., Katan, M.B., 1990. Effect of Dietary trans Fatty Acids on High-Density and Low-Density Lipoprotein Cholesterol Levels in Healthy Subjects. New England Journal of Medicine 323, 439–445. doi:10.1056/NEJM199008163230703

Micha, R., Mozaffarian, D., 2008. Trans Fatty Acids: Effects on Cardiometabolic Health and Implications for Policy. Prostaglandins Leukot Essent Fatty Acids 79, 147–152. doi:10.1016/j.plefa.2008.09.008

Micklewright, J., Schnepf, S.V., 2010. How reliable are income data collected with a single question? Journal of the Royal Statistical Society: Series A (Statistics in Society) 173, 409–429. doi:10.1111/j.1467-985X.2009.00632.x

Mincer, J., 1978. Family Migration Decisions. Journal of Political Economy 86, 749–73.

Mokdad, A., Ford, E., Bowman, B., Dietz, W., Vinicor, F., Bales, V., Marks, J., 2003. Prevalence of obesity, diabetes, and obesity-related health risk factors, 2001. JAMA 289, 76–79. doi:10.1001/jama.289.1.76

Mozaffarian, D., Jacobson, M.F., Greenstein, J.S., 2010. Food Reformulations to Reduce Trans Fatty Acids. New England Journal of Medicine 362, 2037–2039. doi:10.1056/NEJMc1001841

Mozaffarian, D., Katan, M.B., Ascherio, A., Stampfer, M.J., Willett, W.C., 2006. Trans Fatty Acids and Cardiovascular Disease. New England Journal of Medicine 354, 1601–1613. doi:10.1056/NEJMra054035

Mozaffarian, D., Rimm, E.B., King, I.B., Lawler, R.L., McDonald, G.B., Levy, W.C., 2004. Trans Fatty Acids and Systemic Inflammation in Heart Failure. Am J Clin Nutr 80, 1521–1525.

Mullahy, J., Robert, S.A., 2010. No time to lose: time constraints and physical activity in the production of health. Rev Econ Household 8, 409–432. doi:10.1007/s11150-010-9091-4

Must A, Spadano J, Coakley EH, Field AE, Colditz G, Dietz WH, 1999. The disease burden associated with overweight and obesity. JAMA 282, 1523–1529. doi:10.1001/jama.282.16.1523

National Heart, Lung, and Blood Institute (NHLBI), 2005. High Blood Cholesterol: What You Need To Know (NIH Publication No. No. 05-3290). USDHHS, NIH.
National Institute on Alcohol Abuse and Alcoholism (NIAAA), 2016. Alcohol Facts and Statistics [WWW Document]. URL https://www.niaaa.nih.gov/alcohol-health/overview-alcohol-consumption/alcohol-facts-and-statistics (accessed 8.18.16).
National Institute on Alcohol Abuse and Alcoholism (NIAAA), n.d. Drinking Levels Defined [WWW Document]. URL http://www.niaaa.nih.gov/alcohol-health/overview-alcohol-consumption/moderate-binge-drinking (accessed 5.3.16).
National Institutes of Health, National Heart, Lung, and Blood Institute, 1998. Clinical Guidelines on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults:The Evidence Report. National Institutes of Health. Obes. Res. 6 Suppl 2, 51S–209S.
Newman, P.W., Kenworthy, J.R., 1996. The land use—transport connection. Land Use Policy 13, 1–22. doi:10.1016/0264-8377(95)00027-5
Ng, S.W., Slining, M.M., Popkin, B.M., 2014. Turning point for US diets? Recessionary effects or behavioral shifts in foods purchased and consumed. Am J Clin Nutr 99, 609–616. doi:10.3945/ajcn.113.072892
Oches, S., 2011. The 2011 QSR 50 [WWW Document]. QSR magazine. URL https://www.qsrmagazine.com/reports/2011-qsr-50 (accessed 5.3.16).
Ockene IS, Chiriboga DE, Stanek III EJ, et al, 2004. Seasonal variation in serum cholesterol levels: Treatment implications and possible mechanisms. Arch Intern Med 164, 863–870. doi:10.1001/archinte.164.8.863
Ogden, C.L., Carroll, M.D., 2010. Prevalence of overweight, obesity, and extreme obesity among adults: United States, trends 1960–1962 through 2007–2008, NCHS Health E-Stats. National Center for Health Statistics, Hyattsville, MD.
Ogden CL, Carroll MD, Kit BK, Flegal KM, 2014. PRevalence of childhood and adult obesity in the united states, 2011-2012. JAMA 311, 806–814. doi:10.1001/jama.2014.732
Okrent, R.V., Abigail, 2012. USDA Economic Research Service - EIB102 [WWW Document]. URL http://www.ers.usda.gov/publications/eib-economic-information-bulletin/eib102.aspx (accessed 4.21.16).
Oomen, C.M., Ocké, M.C., Feskens, E.J., van Erp-Baart, M.A., Kok, F.J., Kromhout, D., 2001. Association between trans fatty acid intake and 10-year risk of coronary heart disease in the Zutphen Elderly Study: a prospective population-based study. Lancet 357, 746–751.
Pinkston, J.C., Stewart, J., 2009. How does time use affect the probability of becoming obese. Presented at the American Time Use Research Conference, College Park, MD.
Pivarnik, J.M., Reeves, M.J., Rafferty, A.P., 2003. Seasonal variation in adult leisure-time physical activity. Med Sci Sports Exerc 35, 1004–1008. doi:10.1249/01.MSS.0000069747.55950.B1
Plantinga, A.J., Bernell, S., 2007. The Association Between Urban Sprawl and Obesity: Is It a Two-Way Street? Journal of Regional Science 47, 857–879. doi:10.1111/j.1467-9787.2007.00533.x
Plasqui, G., Westerterp, K.R., 2004. Seasonal Variation in Total Energy Expenditure and Physical Activity in Dutch Young Adults. Obesity Research 12, 688–694. doi:10.1038/oby.2004.80
Ploeg, H.P. van der, Merom, D., Chau, J.Y., Bittman, M., Trost, S.G., Bauman, A.E., 2010. Advances in Population Surveillance for Physical Activity and Sedentary Behavior: Reliability and Validity of Time Use Surveys. Am. J. Epidemiol. 172, 1199–1206. doi:10.1093/aje/kwq265
Popham, F., Mitchell, R., 2006. Leisure time exercise and personal circumstances in the working age population: longitudinal analysis of the British household panel survey. J Epidemiol Community Health 60, 270–274. doi:10.1136/jech.2005.041194
Powell, L.M., Nguyen, B.T., Han, E., 2012. Energy Intake from Restaurants: Demographics and Socioeconomics, 2003–2008. American Journal of Preventive Medicine 43, 498–504. doi:10.1016/j.amepre.2012.07.041

Powell, L.M., Slater, S., Chaloupka, F.J., 2004. The relationship between community physical activity settings and race, ethnicity and socioeconomic status. Evidence-Based Preventive Medicine 1, 135–144.

Pratt, M., Macera, C.A., Wang, G., 2000. Higher Direct Medical Costs Associated With Physical Inactivity. The Physician and Sportsmedicine 28, 63–70. doi:10.3810/psm.2000.10.1237

Ratnayake, W.M.N., L'Abbe, M.R., Mozaffarian, D., 2008. Nationwide product reformulations to reduce trans fatty acids in Canada: when trans fat goes out, what goes in? Eur J Clin Nutr 63, 808–811. doi:10.1038/ejcn.2008.39

Reis, J.P., Bowles, H.R., Ainsworth, B.E., Dubose, K.D., Smith, S., Laditka, J.N., 2004. Nonoccupational Physical Activity by Degree of Urbanization and U.S. Geographic Region: Medicine & Science in Sports & Exercise 2093–2098. doi:10.1249/01.MSS.0000147589.98744.85

Remig, V., Franklin, B., Margolis, S., Kostas, G., Nece, T., Street, J.C., 2010. Trans Fats in America: A Review of Their Use, Consumption, Health Implications, and Regulation. Journal of the American Dietetic Association 110, 585–592. doi:10.1016/j.jada.2009.12.024

Restrepo, B.J., Rieger, M., 2016. Trans fat and cardiovascular disease mortality: Evidence from bans in restaurants in New York. Journal of Health Economics 45, 176–196. doi:10.1016/j.jhealeco.2015.09.005

Robert Wood Johnson Foundation, 2008. Where We Live Matters for Our Health: Neighborhoods and Health (Issue Brief No. 3), Neighborhoods and Health. Robert Wood Johnson Foundation.

Roberts, J., Hodgson, R., Dolan, P., 2011. "It's driving her mad": Gender differences in the effects of commuting on psychological health. Journal of Health Economics 30, 1064–1076. doi:10.1016/j.jhealeco.2011.07.006

Rössner, S., 2002. Obesity: the disease of the twenty-first century. Int. J. Obes. Relat. Metab. Disord. 26 Suppl 4, S2-4. doi:10.1038/sj.ijo.0802209

Sahlqvist, S., Song, Y., Ogilvie, D., 2012. Is active travel associated with greater physical activity? The contribution of commuting and non-commuting active travel to total physical activity in adults. Preventive Medicine 55, 206–211. doi:10.1016/j.ypmed.2012.06.028

Sallis, J.F., Floyd, M.F., Rodríguez, D.A., Saelens, B.E., 2012. Role of Built Environments in Physical Activity, Obesity, and Cardiovascular Disease. Circulation 125, 729–737. doi:10.1161/CIRCULATIONAHA.110.969022

Santos, A., McGuckin, N., Nakamoto, H.Y., Gray, D., Liss, S., 2011. Summary of travel trends: 2001 National Household Travel Survey (Trends in travel behavior, 1969-2009 No. FHWA-PL-ll-022). U.S. Department of Transportation, Federal Highway Administration, Washington, DC.

Sari, N., 2009. Physical inactivity and its impact on healthcare utilization. Health Econ. 18, 885–901. doi:10.1002/hec.1408

Sattelmair, J., Pertman, J., Ding, E.L., Kohl, H.W., Haskell, W., Lee, I.-M., 2011. Dose Response Between Physical Activity and Risk of Coronary Heart Disease A Meta-Analysis. Circulation 124, 789–795. doi:10.1161/CIRCULATIONAHA.110.010710

Schauder, S.A., Foley, M.C., 2015. The relationship between active transportation and health. Journal of Transport & Health 2, 343–349. doi:10.1016/j.jth.2015.06.006

Scott, M.M., Dubowitz, T., Cohen, D.A., 2009. Regional differences in walking frequency and BMI: What role does the built environment play for Blacks and Whites? Health & Place 15, 897–902. doi:10.1016/j.healthplace.2009.02.010

Sermons, M.W., Koppelman, F.S., 2001. Representing the differences between female and male commute behavior in residential location choice models. Journal of Transport Geography 9, 101–110. doi:10.1016/S0966-6923(00)00047-8

Smith, L.P., Ng, S.W., Popkin, B.M., 2014. Resistant to the recession: low-income adults' maintenance of cooking and away-from-home eating behaviors during times of economic turbulence. Am J Public Health 104, 840–846. doi:10.2105/AJPH.2013.301677

Stender, S., Astrup, A., Dyerberg, J., 2009. What Went In When Trans Went Out? New England Journal of Medicine 361, 314–316. doi:10.1056/NEJMc0903380
Stender, S., Astrup, A., Dyerberg, J., 2008. Ruminant and industrially produced trans fatty acids - health aspects. Food & Nutrition Research 52. doi:10.3402/fnr.v52i0.1651
Stender, S., Dyerberg, J., Astrup, A., 2007. Fast food: unfriendly and unhealthy. Int J Obes (Lond) 31, 887–890. doi:10.1038/sj.ijo.0803616
Stender, S., Dyerberg, J., Astrup, A., 2006. High Levels of Industrially Produced Trans Fat in Popular Fast Foods. New England Journal of Medicine 354, 1650–1652. doi:10.1056/NEJMc052959
Sugiyama, T., Ding, D., Owen, N., 2013. Commuting by Car. American Journal of Preventive Medicine 44, 169–173. doi:10.1016/j.amepre.2012.09.063
Swinburn, B.A., Sacks, G., Hall, K.D., McPherson, K., Finegood, D.T., Moodie, M.L., Gortmaker, S.L., 2011. The global obesity pandemic: shaped by global drivers and local environments. The Lancet 378, 804–814. doi:10.1016/S0140-6736(11)60813-1
Tarrago-Trani, M.T., Phillips, K.M., Lemar, L.E., Holden, J.M., 2006. New and existing oils and fats used in products with reduced trans-fatty acid content. J Am Diet Assoc 106, 867–880. doi:10.1016/j.jada.2006.03.010
Taylor, A.L., Denniston, M.M., Klevens, R.M., McKnight-Eily, L.R., Jiles, R.B., 2016. Association of Hepatitis C Virus With Alcohol Use Among U.S. Adults: NHANES 2003–2010. American Journal of Preventive Medicine 51, 206–215. doi:10.1016/j.amepre.2016.02.033
Teegala, S.M., Willett, W.C., Mozaffarian, D., 2009. Consumption and Health Effects of Trans Fatty Acids: A Review. Journal of AOAC International 92, 1250–1257.
Todd, J.E., Mentzer Morrison, R., 2014. USDA Economic Research Service - Less Eating Out, Improved Diets, and More Family Meals in the Wake of the Great Recession [WWW Document]. URL http://www.ers.usda.gov/amber-waves/2014-march/less-eating-out,-improved-diets,-and-more-family-meals-in-the-wake-of-the-great-recession.aspx#.Vx422_krLct (accessed 4.25.16).
Trost, S.G., Owen, N., Bauman, A.E., Sallis, J.F., Brown, W., 2002. Correlates of adults' participation in physical activity: review and update. Med Sci Sports Exerc 34, 1996–2001. doi:10.1249/01.MSS.0000038974.76900.92
Tudor-Locke, C., Leonardi, C., Johnson, W.D., Katzmarzyk, P.T., 2011. Time spent in physical activity and sedentary behaviors on the working day: the American time use survey. J. Occup. Environ. Med. 53, 1382–1387. doi:10.1097/JOM.0b013e31823c1402
Tudor-Locke, C., Washington, T.L., Ainsworth, B.E., Troiano, R.P., 2009. Linking the American Time Use Survey (ATUS) and the Compendium of Physical Activities : methods and rationale. Journal of Physical Activity and Health 6, 347–353.
Turner, T., Niemeier, D., 1997. Travel to work and household responsibility: new evidence. Transportation 24, 397–419. doi:10.1023/A:1004945903696
Unnevehr, L.J., Jagmanaite, E., 2008. Getting rid of trans fats in the US diet: Policies, incentives and progress. Food Policy 33, 497–503. doi:10.1016/j.foodpol.2008.05.006
U.S. Bureau of the Census, n.d. County and City Data Books 1947-1994.
U.S. Census Bureau, n.d. Quick Facts: Mean travel time to work (minutes), workers age 16+, 2009-2013 [WWW Document]. URL http://www.census.gov/quickfacts/map/LFE305214/00 (accessed 11.15.16).
US Department of Agriculture, 2010. Report of the Dietary Guidelines Advisory Committe on the Dietary Guidelines for Americans, 2010. US Department of Agriculture.
US Food and Drug Administration, 2015. Press Announcements - The FDA takes step to remove artificial trans fats in processed foods [WWW Document]. URL http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm451237.htm (accessed 5.4.16).
USDHHS, 2010. The Surgeon General's Vision for a Healthy and Fit Nation. U.S. Department of Health and Human Services, Office of the Surgeon General, Rockville, MD.

USDHHS, 2008. 2008 Physical Activity Guidelines for Americans [WWW Document]. URL http://www.cdc.gov/physicalactivity/everyone/guidelines/ (accessed 5.11.15).
Van Camp, D., Hooker, N.H., Lin, C.-T.J., 2012. Changes in fat contents of US snack foods in response to mandatory trans fat labelling. Public Health Nutrition 15, 1130–1137. doi:10.1017/S1368980012000079
van Ommeren, J.N., Gutiérrez-i-Puigarnau, E., 2011. Are workers with a long commute less productive? An empirical analysis of absenteeism. Regional Science and Urban Economics 41, 1–8. doi:10.1016/j.regsciurbeco.2010.07.005
Visscher, T.L.S., Seidell, J.C., 2004. Time trends (1993–1997) and seasonal variation in body mass index and waist circumference in the Netherlands. Int J Obes Relat Metab Disord 28, 1309–1316. doi:10.1038/sj.ijo.0802761
Wang, F.C., Gravelle, A.J., Blake, A.I., Marangoni, A.G., 2016. Novel trans fat replacement strategies. Current Opinion in Food Science, Food chemistry and biochemistry • Food bioprocessing 7, 27–34. doi:10.1016/j.cofs.2015.08.006
Wang, Y.C., McPherson, K., Marsh, T., Gortmaker, S.L., Brown, M., 2011. Health and economic burden of the projected obesity trends in the USA and the UK. The Lancet 378, 815–825. doi:10.1016/S0140-6736(11)60814-3
Wen, C.P., Wai, J.P.M., Tsai, M.K., Yang, Y.C., Cheng, T.Y.D., Lee, M.-C., Chan, H.T., Tsao, C.K., Tsai, S.P., Wu, X., 2011. Minimum amount of physical activity for reduced mortality and extended life expectancy: a prospective cohort study. The Lancet 378, 1244–1253. doi:10.1016/S0140-6736(11)60749-6
White, M.J., 1977. A Model of Residential Location Choice and Commuting by Men and Women Workers. Journal of Regional Science 17, 41–52. doi:10.1111/j.1467-9787.1977.tb00471.x
WHO, 2000. Obesity: preventing and managing the global epidemic. Report of a WHO consultation, World Health Organization Technical Report Series. World Health Organization, Geneva, Switzerland.
WHO Expert Committee, 1995. Physical status : the use of and interpretation of anthropometry. WHO, Geneva.
Willett, W.C., Stampfer, M.J., Manson, J.E., Colditz, G.A., Speizer, F.E., Rosner, B.A., Hennekens, C.H., Hennekens, C.H., Willett, W.C., Stampfer, M.J., Colditz, G.A., Willett, W.C., Sampson, L.A., Rosner, B.A., 1993. Intake of trans fatty acids and risk of coronary heart disease among women. The Lancet 341, 581–585. doi:10.1016/0140-6736(93)90350-P
Wojan, T.R., Hamrick, K.S., 2015. Can Walking or Biking to Work Really Make a Difference? Compact Development, Observed Commuter Choice and Body Mass Index. PLOS ONE 10, e0130903. doi:10.1371/journal.pone.0130903
Wooldridge, J.M., 2010. Econometric Analysis of Cross Section and Panel Data. MIT Press.
Yang, J., French, S., 2013. The travel–obesity connection: discerning the impacts of commuting trips with the perspective of individual energy expenditure and time use. Environment and Planning B: Planning and Design 40, 617 – 629. doi:10.1068/b38076
Zellner, A., 1962. An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. Journal of the American Statistical Association 57, 348–368. doi:10.1080/01621459.1962.10480664
Zhao, Z., Kaestner, R., 2010. Effects of urban sprawl on obesity. Journal of Health Economics 29, 779–787. doi:10.1016/j.jhealeco.2010.07.006
Zick, C.D., Stevens, R.B., Bryant, W.K., 2011. Time use choices and healthy body weight: A multivariate analysis of data from the American Time use Survey. International Journal of Behavioral Nutrition and Physical Activity 8, 84. doi:10.1186/1479-5868-8-84
Zolnik, E.J., 2011. The Effect of Sprawl on Private-Vehicle Commuting Outcomes. Environ Plan A 43, 1875–1893. doi:10.1068/a42466

Appendix A

One concern with the measure of FAFH, or food away from home, in NHANES data is that the wording of the question changes in 2005. Initially the question asked how many meals are eaten in a restaurant, then later how many meals are eaten that were prepared outside of the home. This wording includes meals at restaurants but also includes meals from vending machines, grocery store delis and food counters, food trucks, and other establishments. Table A1 shows exact wording of the question associated with the *Meals* variable.

Table A1: Changes in wording of question associated with *Meals* variable

Year	Question wording
1999 2001 2003	On average, how many times per week do you eat meals that were prepared in a restaurant?
2005	I'm going to ask you about meals. By meal, I mean breakfast, lunch and dinner. On average, how many meals per week do you get that were not prepared at a home? Please include meals from both dine-in and carry out restaurants, restaurants that deliver food to your home, cafeterias, fast-food places, food courts, food stands, meals prepared at a grocery store, and meals from vending machines.
2007 2009	I'm going to ask you about meals. By meal, I mean breakfast, lunch and dinner. During the past 7 days, how many meals did you get that were prepared away from home in places such as restaurants, fast food places, food stands, grocery stores, or from vending machines? {Please do not include meals provided as part of the school lunch or school breakfast./Please do not include meals provided as part of the community programs you reported earlier.}

Source: NHANES 1999-2010 Diet Behavior & Nutrition Data Documentation

The concern here is that this change in wording has changed the meaning of the variable meals out and that this in turn changes the effect that one would expect to see on cholesterol and other health measures. Statistically, the effect of the change in the wording of the question is likely to be very small, because it did not change the mode of responses or the shape of the distribution of the responses, as shown in Figure A1. While the means did change, as shown in Table A2, the change is consistent with what one would expect given the upward trend in eating out over time.

Table A2: Average Number of Meals Out over Time

Mean of *Meals* Variable Over Time

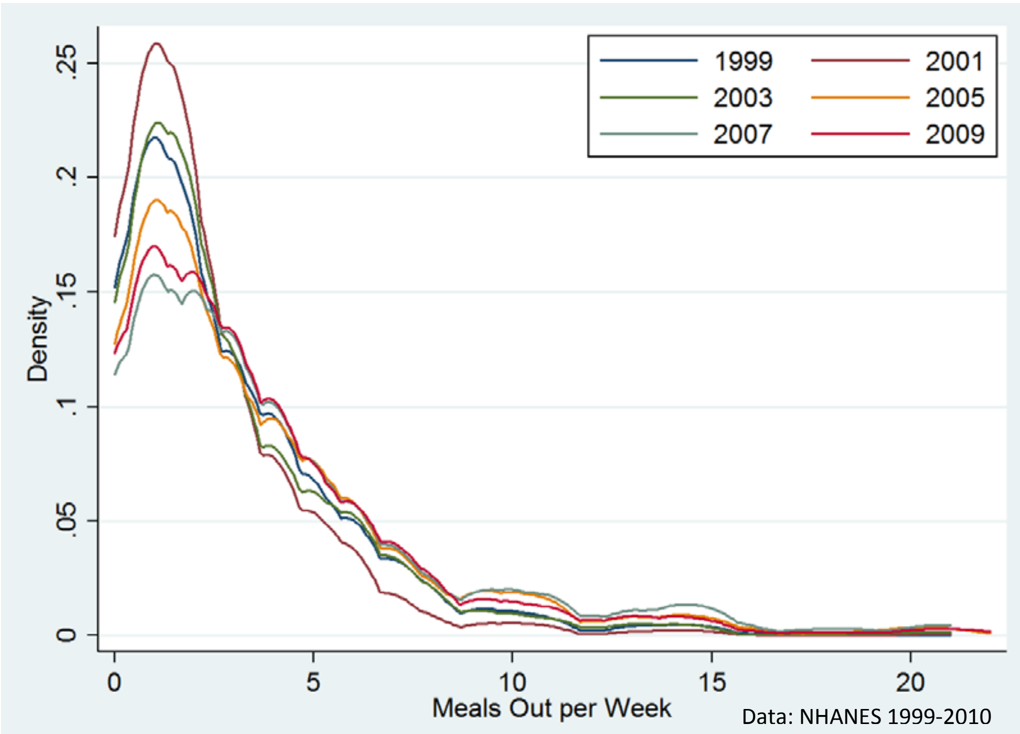
	1999	2001	2003	2005	2007	2009
Mean	2.80	2.25	2.88	3.58	3.92	3.49
Std. err.	0.10	0.07	0.07	0.14	0.16	0.09

Data: NHANES 1999-2010

The mean of meals out increases by 0.7 between the 2003 and 2005 survey cycles, when the survey question changed. However, relative to changes in other years, this change in means is not large.

Figure A1 presents a kernel density plot, which shows how the distribution of *Meals* out changes across years and whether or not the mode of meals out changes. To make this plot, I use a bandwidth of 0.75 to allow for smoothing because *Meals* are reported in discrete quantities. I chose to use a kernel density plot instead of a histogram for ease of comparison across years.

Figure A1: Density Plot of Meals Out per Week



This plot shows the number of people reporting very few meals out is higher in 1999 through 2003 than in the later three periods and the number of people reporting three or more meals out is higher in the later three years than in the earlier three. However, as with the means, the differences across years are small.

Other than the change in the question's wording, are there other causes for the increase in meals out? One might theorize that eating out is associated with economic fluctuations; however these data do not show any such correlation. The peak of the tech bubble was in 2001 and the housing bubble was in 2007. One would expect to see more eating out and less eating at home in the periods leading up to those peaks, in the 1995 and 2005/06 data, however this is not the case. Similarly, the recessions hit their lowest lows at the end of 2001 and April 2009. In the periods prior to these lows, coinciding with the 2001 and 2007/08 data, one would expect to see less eating out and more meals prepared at home and actually do see this trend with the 2001 data.

Appendix B

This appendix provides additional information on the sample selection presented in Yang and French (2013) as well as a comparison of methods between this work and Yang and French, with all travel as the outcome variable of interest. Summary statistics of the commuting sample and the travel sample are presented in Tables B1 and B2, using an approximate replication of sample selection criteria from Yang and French.

Table B1: Demographic and Socioeconomic Characteristics by BMI in Yang and French Sample

	Y&F Travel Sample			Y&F Commuting Sample		
	(1) All BMI	(2) Over-weight	(3) Obese	(4) All BMI	(5) Over-weight	(6) Obese
Age <=25	0.09	0.09	0.08	0.09	0.09	0.08
Age 26-35	0.21	0.20	0.23	0.17	0.17	0.18
Age 36-45	0.28	0.29	0.28	0.24	0.23	0.24
Age 46-55	0.29	0.28	0.29	0.23	0.23	0.24
Age 56+	0.16	0.17	0.15	0.29	0.30	0.28
Male	0.59	0.62	0.55	0.52	0.56	0.48
Spouse or Unmarried Partner in HH	0.60	0.61	0.57	0.57	0.58	0.56
Spouse is Employed	0.44	0.45	0.42	0.40	0.40	0.39
Has a Child in Household	0.53	0.54	0.52	0.49	0.49	0.49
Child Under Age 2 in Household	0.11	0.12	0.10	0.11	0.11	0.11
White	0.66	0.69	0.63	0.67	0.70	0.63
Black	0.15	0.13	0.18	0.14	0.12	0.18
Hispanic	0.15	0.15	0.16	0.14	0.14	0.15
Asian	0.02	0.02	0.02	0.02	0.02	0.01
Other Race	0.02	0.02	0.02	0.02	0.02	0.03
No High School	0.10	0.10	0.10	0.14	0.13	0.15
High School Graduate	0.27	0.26	0.29	0.29	0.27	0.31
Some College	0.20	0.18	0.23	0.19	0.18	0.21
College Graduate	0.32	0.34	0.29	0.28	0.30	0.25
Advanced Degree	0.11	0.12	0.09	0.10	0.11	0.08
Weekly Income <$400	0.20	0.18	0.23	0.15	0.14	0.15
Weekly Income $400 - $700	0.25	0.23	0.27	0.17	0.17	0.19
Weekly Income $700 - $1250	0.26	0.25	0.27	0.19	0.18	0.19
Metropolitan Status	0.82	0.83	0.81	0.81	0.82	0.80
Region: Northeast	0.16	0.16	0.16	0.17	0.17	0.16
Region: Midwest	0.26	0.26	0.27	0.26	0.26	0.26
Region: South	0.37	0.36	0.38	0.37	0.35	0.38
Region: West	0.21	0.22	0.20	0.21	0.22	0.19
Weekend or Holiday	0.25	0.24	0.25	0.49	0.49	0.50

Table B1 continued.

Winter	0.36	0.35	0.36	0.35	0.34	0.35
Spring	0.17	0.17	0.17	0.17	0.18	0.17
Summer	0.23	0.24	0.23	0.24	0.24	0.24
Autumn	0.24	0.24	0.24	0.24	0.24	0.24
Occupation with Physical Activity	0.25	0.25	0.25	0.47	0.46	0.47
Commute	1.00	1.00	1.00	0.38	0.38	0.37
Sedentary Commute	0.98	0.98	0.98	0.37	0.37	0.36
Active Commute	0.07	0.07	0.05	0.02	0.03	0.02
Bike Commute	0.00	0.01	0.00	0.00	0.00	0.00
Walk Commute	0.06	0.07	0.05	0.02	0.03	0.02
Any Travel	1.00	1.00	1.00	1.00	1.00	1.00
Any Sedentary Travel	0.99	0.99	0.99	0.98	0.98	0.98
Any Active Travel	0.12	0.13	0.11	0.13	0.14	0.12
Any Bike Travel	0.01	0.01	0.00	0.00	0.01	0.00
Any Walk Travel	0.12	0.12	0.10	0.13	0.13	0.12
BMI	30.14	27.19	34.38	30.32	27.16	34.69
	(0.07)	(0.03)	(0.10)	(0.04)	(0.02)	(0.07)
Observations	4,688	2,761	1,927	12,388	7,191	5,197

Data: ATUS 2006-2008
Note: Unweighted means for indicator variables are shown; unweighted mean and standard error in parentheses are reported for BMI. Statistics are grouped by BMI status.

Table B2 presents mean values of time spent in active and sedentary travel and commuting and leisure-time physical activity for only those individuals who participated in each of these activities on the diary day, using the Yang and French sample criteria.

Table B2: Average Travel and Exercise in Minutes, by BMI Category in Yang and French Sample

	(1) N	(2) All	(3) Overweight	(4) Obese
All Active Commuting	305	15.77	15.82	15.67
		(1.07)	(1.40)	(1.58)
Commuting by Bike	16	38.44	39.07	34.00
		(10.90)	(12.50)	(4.00)
Commuting by Walking	289	14.52	14.11	15.30
		(0.91)	(1.11)	(1.59)
Sedentary Commuting	4,590	43.17	43.97	42.02
		(0.67)	(0.93)	(0.92)
Total Commuting	4,688	43.29	44.11	42.12
		(0.67)	(0.93)	(0.93)
Active Travel	1,599	19.83	20.40	18.93
		(0.66)	(0.82)	(1.10)
Sedentary Travel	12,089	84.68	85.63	83.38
		(0.72)	(0.94)	(1.11)
Total Travel	12,388	85.20	86.29	83.70
		(0.71)	(0.93)	(1.10)
Leisure Physical Activity	7,712	140.72	144.37	135.24
		(1.51)	(1.97)	(2.34)

Data: ATUS 2006-2008

Note: Unweighted means and robust standard errors in parentheses are reported for participants of each activity. Statistics are grouped by BMI status.

The preferred sample in this chapter includes only working-age adults, aged 21 through 65, while Yang and French include in their commuting analysis all ages from 15 upwards who report commuting on the diary day, and in their travel analysis, all ages from 15 upwards who report traveling (excluding normal- and underweight individuals. In their sample, total commuting makes up approximately half of total travel time.

In an effort to examine whether results from the comparison with the Yang and French work presented in this chapter would be similar if examining total travel time, Tables B3 and B4 are presented here.

Table B3: Comparison of Samples using Yang and French estimation for All Travel

	(1)	(2)	(3)	(4)
	Y&F original	Y&F replication	Preferred Sample, excl. BMI<25	Preferred Sample
% of Travel in Vehicle	0.74**	0.818***	1.353**	2.457***
	(0.185)	(0.313)	(0.557)	(0.502)
Total Travel (minutes)	0.00	0.000348	0.00325	0.00133
	(0.00)	(0.000806)	(0.00248)	(0.00218)
Constant	28.90***	30.27***	28.00***	23.58***
	(0.369)	(0.392)	(0.715)	(0.629)
Observations	12,208	12,388	3,984	6,040
R-squared	0.029	0.026	0.038	0.054

Data: ATUS 2006-2008

Col. 1: Standard errors from previously published t-statistics. Col 2-4: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Note: Outcome variable is self-reported BMI, measured in kg/m^2. All columns control for age, sex, race, spouse or unmarried partner in household, household children, education, income, metropolitan status. Column 1 shows published coefficients from Yang & French (2013). Column 2, presents results from replication of the Yang & French model using their sample selection. Column 3 shows a replication of their model using my preferred sample selection criteria and restricting the sample to only overweight and obese individuals. Column 4 presents a replication of the Yang & French model using my preferred sample selection, which includes normal weight, overweight and obese individuals.

Table B3 is analogous to Table 3.7 in Chapter 3 and compares their main results using all travel as the predictor of interest, using a sample that matches theirs against the preferred sample presented in this chapter.

Table B4 is analogous to Table 3.8 in the paper; using all travel (not just commuting) as the predictor of interest, it compares the Yang and French model, sample selection, and choice of controls against those presented in the preferred model in this work. Column 4 of Table B4 estimates Equation 1 in the paper using all sedentary travel and all active travel as the predictors of interest.

Table B4: Comparison of Models for All Travel

	(1) Model: Y&F Controls: Y&F Sample: Y&F	(2) Model: Y&F Controls: Pref. Sample: Y&F	(3) Model: Pref. Controls: Pref. Sample: Y&F	(4) Model: Pref. Controls: Pref. Sample: Pref.
% of Travel in Vehicle	0.818***	0.876*		
	(0.313)	(0.488)		
Total Travel (minutes)	0.000348	0.000453		
	(0.000806)	(0.00107)		
Sedentary Travel (minutes)			0.000747	0.00172
			(0.00107)	(0.00211)
Active Travel (minutes)			-0.0142**	-0.0274***
			(0.00614)	(0.00795)
Constant	30.27***	30.81***	31.65***	29.68***
	(0.392)	(0.788)	(0.681)	(1.128)
Observations	12,388	7,793	7,793	6,121
R-squared	0.026	0.034	0.034	0.071

Data: ATUS 2006-2008
Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1
Note: Outcome variable is self-reported BMI, measured in kg/m^2. Column 1 presents a replication of the Yang and French (2013) model, using their control variables and sample selection criteria. Column 2 estimates the Yang and French model with the full set of control variables used in Column 5 of Table 3.3 and the Yang and French sample selection. Column 3 estimates Equation (3.1) using the full set of control variables used in Column 5 of Table 3.3 and the Yang and French sample selection. Column 4 estimates the preferred model, Equation (3.1), with the preferred sample and full set of control variables.

Appendix C

Table C1 presents average marginal effects of a Probit estimate of county-level mean single-family housing price from 1970 on the sedentary commuting indicator variable. No control variables are included in the model. To improve readability of the estimates, 1970 price is divided by $10,000. This shows the relationship between historic housing prices and the commuting decision is statistically significant only in normal weight and overweight men.

Table C1: Estimate of 1970 House Price Instrument on Sedentary Commuting Indicator

Probit	(1) All Males	(2) BMI < 30 Males	(3) BMI 30+ Males	(4) All Females	(5) BMI < 30 Females	(6) BMI 30+ Females
1970 Price/$10,000	-0.00962**	-0.0114**	-0.00280	0.00506	0.00201	0.0172*
	(0.00419)	(0.00521)	(0.00815)	(0.00604)	(0.00721)	(0.00932)
Observations	1,475	1,072	403	1,227	906	321

Data: ATUS 2006-2008
Note: Results from Probit estimation of county-level mean single-family house price in 1970 on sedentary commuting indicator. Model is estimated without inclusion of control variables. Standard errors in parentheses. Standard errors are clustered at the county-month-year level.
*** p<0.01, ** p<0.05, * p<0.1