2016

# Income and Physical Activity Choices: A Comparison Between United States and China

Xiaowen Hu
*University of Kentucky*, hu.xiaowen.c@gmail.com
Author ORCID Identifier:
http://orcid.org/0000-0002-7808-3745
Digital Object Identifier: https://doi.org/10.13023/ETD.2016.471

**STUDENT AGREEMENT:**

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

**REVIEW, APPROVAL AND ACCEPTANCE**

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Xiaowen Hu, Student

Dr. C. Jill Stowe, Major Professor

Dr. Carl Dillon, Director of Graduate Studies

# INCOME AND PHYSICAL ACTIVITY CHOICES: A COMPARISON BETWEEN UNITED STATES AND CHINA

---

DISSERTATION

---

A dissertation submitted in partial
fulfillment of the requirements for
the degree of Doctor of Philosophy in
the College of Agriculture, Food and
Environment at the University of
Kentucky

By
Xiaowen Hu
Lexington, Kentucky

Director: Dr. C. Jill Stowe, Associate Professor of Agricultural Economics
Lexington, Kentucky 2016

ABSTRACT OF DISSERTATION

INCOME AND PHYSICAL ACTIVITY CHOICES: A COMPARISON BETWEEN UNITED STATES AND CHINA

The low income population often appears to make poor health choices, including physical activity deficiency. Since outdoor physical activity does not have to be monetarily costly, one explanation for this phenomenon is related to the idea of time preference. Briefly, the benefit of future good health appears to be valued less by those with low income, and they face a choice between consumption today and better health in the future. The objective of this study is to further investigate the determinants of participation in physical activity with an emphasis on the effect of annual household income. This dissertation consists of three empirical studies. The first one utilizes data from the National Health and Nutrition Examination Survey with a two step Heckman selection model. The second study conducts a Kentucky Exercise & Health Survey and applies a two part Cragg's hurdle model. The third study employs panel data models on longitudinal data from China Health and Nutrition Survey. By utilizing the concept of time preferences, the results of empirical analysis show that income is positively related to the probability to participate in physical activity in both countries, while the prediction of the relationship between income and average time spent on physical activity remains ambiguous.

KEYWORDS: Health Choices, Physical Activity Participation, Income, Time Preference.

Author's signature: Xiaowen Hu

Date: December 1, 2016

# INCOME AND PHYSICAL ACTIVITY CHOICES: A COMPARISON BETWEEN UNITED STATES AND CHINA

By
Xiaowen Hu

Director of Dissertation: C. Jill Stowe

Director of Graduate Studies: Carl Dillon

Date: December 1, 2016

To Rulan.

# ACKNOWLEDGMENTS

I would like to express my great gratitude and sincere appreciation to my advisor, Dr. Jill Stowe. Dr. Jill Stowe led me to this research field, poured in her knowledge and encouragement to me without reservation, and provided the most insightful direction to my research. Her guidance, patience and support made this dissertation possible. Being her student is really a wonderful journey of my Ph.D. studies.

I would like to thank my dissertation committee members, Dr. Blomquist, Dr. Davis, Dr. Minier and Dr. Kusunose, who have enhanced the dissertation through their valuable input.

I would like to thank the AEC staff, especially Rita and Chuck, for their support and help in coordinating the Kentucky Exercise & Health Survey. I also want to thank my friends and colleagues, who were entrenched with me in this long, and sometimes treacherous Ph.D. journey. Lastly, thanks to my husband, Dr. Zhiyuan Shen, my mother, Ms. Rongna Lyu, and my most adorable daughter, Rulan, for their unwavering love and unconditional support.

Table of Contents

List of Tables

## Chapter One: Introduction

Health choice is an important area of study for agricultural economists. A vast amount of literature suggests that the low income population often appears to make poor health choices. While one of the significant factors in the lower consumption of healthy foods like fruits and vegetables is price (Cutler and Lleras-Muney, 2010; Park et al., 1996; Stewart et al., 2003), the choice to participate in physical activity should be a different story. Participating in physical activity can be low cost; there are numerous free or low-cost alternatives to paying for a gym membership or buying equipment for sports. Consequently, choosing to live an active lifestyle should not be restricted by the budget constraint. One explanation for this phenomenon is related to the idea of time preference. Briefly, the benefit of future good health appears to be valued less by those with low income, and consumers face a choice between pleasure today and better health in the future (Becker & Murphy, 1988).

The objective of this dissertation is to further investigate the determinants of participation in physical activity. In particular, I focus on the influence of household income. By utilizing the concept of time preference, this dissertation analyzes whether income predicts the probability of exercising and the time spent exercising. If the benefit of future good health is valued more by those with high income, it is expected to find that household income level should be positively correlated with either the probability of participation in physical activity or the average time spent on it.

### 1.1 Motivations and Objective

Today, physical activity participation is a major public health concern as it affects both physical health and psychological wellbeing, and physical inactivity is rated among the top ten leading causes of death in high income countries (World Health Organization, 2002). Physical activity includes a wide range of activities such

as occupational activity, housework, and activities pursued during leisure time, such as sports and exercise, which is the focus of this study. Economists contribute to the explanation of physical inactivity by accounting for budget constraints (i.e. income and time constraints) and perceived benefits, among other factors.

One important determinant of health related choices is household income: a low income household might not be able to afford healthy goods, which are often more expensive. Previous studies have found that low income populations appear more likely to choose unhealthy diets, such as lower consumption of fruits and vegetables, have higher smoking rates, engage in less physical activity, and partake in excessive alcohol consumption. For example, Stephens et al. (1985) find that high income persons are more active than low incomes persons in their leisure time physical activity. Evidence presented by Stewart et al. (2003) suggests that poor households spend less on fruits and vegetables than other households; however, as income rises, they are unlikely to increase the consumption of fruits or vegetables. This latter result suggests that there are factors other than budget constraint affecting their choices.

Low income individuals are less likely to participate in physical activity (Anokye et al., 2012; Breuer and Wicker, 2008; Farrell and Shields, 2002; Garcia et al., 2011; Humphreys and Ruseski, 2006; Meltzer and Jena, 2010), which is generally considered to have numerous health benefits and has the potential to be non costly, which means that the budget constraint should not be binding. One explanation for this finding is related to the idea of time preference. Briefly, participating in physical activity can be seen as one kind of health investment. People sacrifice exercising time for non exercising leisure activities which provide more pleasure in the current period; however, insufficient physical activity has harmful effects on human health in the

next period and undermines future utility. The time preference model implies that when faced with a trade off between long term health and an immediate pleasure that might be detrimental to their future health, consumers who discount the future more heavily tend to select immediate pleasure, even though rationality suggests that individuals are planning to maximize utility over their entire life span. Higher income increases the future cost of being physically inactive, since its negative effects on health would cause a greater loss in future income and thus lifetime income. Therefore, lower income individuals have a higher time preference rate, discount future utility more heavily, and are more likely to choose insufficient levels of physical activity.

Although previous research has provided different perspectives on economic determinants of participation and time spent on physical activity, one obvious extension of this study is to link the effect of income with the concept of time preference in the theoretical model. This allows to interpret the influence of income in a different way: it functions as a budget constraint as well as implies a rational expectation for future health, which may affect the physical activity participation decision.

The influence of income on physical activity participation has been explored in Europe and other developed countries, but little research has focused on the United States population. In Asia, previous studies in China on physical activity have demonstrated that income has a positive relationship either with the likelihood to participate in leisure time physical activity, or with the total time spent on it. However, it remains unclear whether the positive effect of income occurs simultaneously on the frequency and the duration time on leisure time physical activity participation. The research in this dissertation focuses on an empirical investigation of the income-

physical activity connection among the United States population and the China population. The comparison between a developed country and a developing country furthers the understanding of the role economic factors play in determining health choices.

The objective of this study is to further investigate the determinants of individual consumption choice as it relates to leisure time physical activity participation, with a focus on the influence of income. By utilizing the concept of time preference, this study focuses on leisure time physical activity choices among the U.S. population from the National Health and Nutrition Examination Survey (NHANES) data, the China population based on the China Health and Nutrition Survey (CHNS) data, and Kentucky residents with an independently conducted Kentucky Exercise and Health Survey (KEHS). It aims to provide a descriptive comparison along with explanatory models of physical activity participation among the population according to annual household income, age, educational background, and marital status, with the emphasis on the first factor. If the benefit of future good health is valued less by those with low income, it is expected that physical activity participation is positively correlated with income level. The latter part of the empirical analysis yields results consistent with theoretical predictions. Ambiguities from previous research are resolved to some extent.

**1.2 Structure**

This dissertation is structured as follows. Chapter 2 reviews the previous research on income and physical activity participation behavior from five aspects. The first section discusses the development and extension of time preference theory and its application to health choices, which is the foundation of this dissertation. The next section provides empirical evidence on the relationship between income and health

behavior from three typical health consumption decisions and one health outcome: cigarette consumption, drinking behavior, obesity rate and physical activity participation, and underscores the necessity of this study. The third and fourth sections summarize physical activity studies in the United States and in China. The final section briefly discusses some widely utilized empirical methods in relevant studies, including logistic models, two step models and panel data models.

Chapter 3 presents a theoretical model that can be used to generate specific hypotheses. A lifetime aggregate utility function is composed based on neoclassical theory, and income is considered as a budget constraint of the current period as well as a component of future utility. By comparing utility differences under different consumption levels, the model predicts that lower income individuals are less likely to invest in healthy consumption choices and hence are less likely to engage in leisure-time physical activity.

The theoretical predictions of the model are tested empirically in Chapters 4 – 6. Chapter 4 investigates the determinants of physical activity participation in the United States. A two-step Heckman selection model is estimated utilizing data from the National Health and Nutrition Examination Survey (NHANES). The results show that the likelihood of participating in physical activity increases with income, but the relationship between average time spent on physical activity and income is ambiguous.

Chapter 5 investigates physical activity participation determinants among Kentucky residents based on an independently-conducted Kentucky Exercise & Health Survey (KEHS). A Cragg's two-step hurdle model is utilized with two dependent variables: the likelihood of the decision to participate in physical activity and the average time spent on physical activity on a typical day. The empirical findings suggest that income indirectly influences the choice of physical activity

participation through investing in physical activity related expenses such as gym membership fees, league fees, personal training fees or equipment, the presence of workplace exercise facilities, educational attainment, and having physically active friends or family.

Chapter 6 investigates the determinants of physical activity participation in China. Longitudinal data from the China Health and Nutrition Survey (CHNS) are analyzed using a random effects method, and three physical activity behaviors, the probability of physical activity participation, the time spent on physical activity in a typical day during the week (Monday through Friday), and the time spent on physical activity in a typical day on weekends (Saturday and Sunday), are estimated. The study finds that the three sets of estimates all increase with income.

Chapter 7 provides a summary. In general, the theoretical and empirical findings support the hypothesis that low income individuals are more likely to make poor choices with regard to future health, since they discount future utility relatively heavily. The second part of this chapter provides a comparison between physical activity participation determinants between developed countries and developing countries based on the findings from Chapters 4 and 6. Finally, the third part of this chapter includes policy remarks and research limitations and extensions.

## Chapter Two: Literature Review

This chapter reviews the previous research on income and physical activity participation behavior. Section 2.1 discusses the development and extension of time preference theory and its application to health behaviors, which is the theoretical foundation of this dissertation. Section 2.2 provides empirical evidence on the relationship between income and health choices or health outcomes from four aspects: cigarette consumption, consumption of alcoholic beverages, obesity rate, and physical activity participation, and addresses the necessity of this study. Sections 2.3 and 2.4 summarize physical activity research in the United States and China, respectively, providing a comprehensive background for this research. Section 2.5 discusses some widely utilized empirical methods in relevant studies, including logistic models, two step models and panel data models, which provides the empirical foundation for this dissertation.

### 2.1 Time Preference and Consumer Behavior

In the view of neoclassical economics, rational consumers choose between consumption and leisure to maximize their utility function subject to budget and time constraints. Becker (1965) develops a model assuming that households combine time and market goods to produce commodities that increase their utility. Subsequently, Becker and Murphy (1988) state that the demand for addictive goods can also be analyzed by the same demand theory as for any other commodity. They present a consumer choice model of rational addiction and time preference, proposing that people weigh the total costs and benefits of an addictive behavior and make consumption decisions to maximize their lifetime aggregate utility. According to their theory, people determine their optimal consumption of goods based on many factors: their utility function; current prices and income; expected prices and income; initial

stock of consumption capital; depreciation rate of consumption capital; and their time preference rate. The full cost of current consumption includes possible negative consequences occurring at later period of life, but, with heavy discounting, these negative consequences are valued less. Thus, people with high preferences for the present are more likely to consume addictive goods. The authors assume that time preferences are stable across time and that a stronger time preference, which means discounting future utility more heavily, is a contributing factor to addictive consumption.

Becker and Mulligan (1997) further discuss the term "time preference". Unlike previous studies which mostly assume the rate of time preference is exogenous, their study investigates the endogenous determination of time preference. They assume that people are not equally patient, but many of the differences are explainable; they argue that the rate of time preference is associated with income, education and other personal information.

Their study provides three main results. First, the rate of time preference has a complementarity with future utility. With a higher future utility, the benefits of low discounts on the future become greater, because discount rates are weighted by larger utilities. Future utility affects the marginal benefit of discounts according to their distance and the level of the discount factor. Therefore, anything that increases the future utility but does not raise the marginal utility of current consumption will lower the equilibrium discount rate on the future. Second, wealthier individuals would be more patient than individuals with lower income because they have both a lower marginal utility of wealth (which implies a lower cost of investing in time preference) and a higher future utility. Wealth implies patience. Meanwhile, consumption grows more rapidly for more highly educated individuals. Finally, they conclude that people

are not equally patient, and the endogeneity in discount rates will exacerbate the inequality in many different areas, including savings and wealth.

Grossman (1972) constructs a model of the demand for the commodity "good health". It assumes that individuals inherit an initial stock of health that depreciates over time and can be increased by investment such as engaging in healthy behaviors. This is extended by the research of Becker, Grossman, and Murphy (1991), which demonstrates that the framework of time preference can explain both rational addiction and health choices. Labor economics also introduces the definition of time preference (Heckman, 1974, 1976 & 2007) utilizing a lifecycle investment framework and predicts that individuals with higher wage rates have a stronger willingness to invest time in their health.

According to the theoretical analyses above, physical activity, for example, can be seen as one kind of health investment. People sacrifice exercising time for non-exercising leisure activities which provide more pleasure at current period; however, physical activity insufficiency brings harmful effects on human health in the next period and undermines future utility. The time preference model implies that when faced with a trade-off between long-term health and an immediate pleasure that might be detrimental to their health, consumers who discount the future more tend to select immediate pleasure, even though rationality suggests that individuals are planning to maximize utility over their entire life span. Higher income increases the future cost of being physically inactive, since its negative effects on health would cause a greater loss in future income and thus lifetime income. Therefore, lower income individuals have a stronger time preference rate, discount future utility more heavily, and are more likely to exhibit physical activity insufficiency.

Becker and Murphy's theory has wielded great influence on time preference in

the past decades. Various model extensions are developed on the basis of the theory (Bretteville-Jensen, 1999; Cawley, 2004; Downward, 2007), and many studies empirically support its prediction (Binkley, 2010; Slade, 2012; Hu & Stowe, 2016). In accordance with the findings of Becker and Mulligan (1997), Bretteville-Jensen (1999) finds that the discount rate among consumers differs significantly. The SLOTH model proposed by Cawley (2004) further emphasizes the time budget. The SLOTH model assumes that utility depends on a person's weight, health, food, other goods and time spent sleeping (S), at leisure (L), at occupation (O), in transportation (T), and in home production (H), and physical activity and sports are considered as leisure activities. It provides an economic framework to explain physical activity participation. Downward (2007) develops a physical activity participation model to distinguish economic variables, individual and social characteristics, as well as sport and exercise characteristics, incorporating the effects of income, time, and social behavior, which links the neoclassical theory to a post-Keynesian approach.

**2.2 Income and Health Choices**

The overall health of the population depends on many health choices; a few of these include smoking, alcohol consumption, diet and physical activity. The field of agricultural economics has examined these specific areas, investigating how consumers make choices related to their health. One important determinant of health related choices that remains consistent across these studies is household income. This section of the literature review summarizes the relationship between income and cigarette consumption, drinking behavior, obesity rate and physical activity participation.

### 2.2.1 Income and Cigarette Consumption

Several studies have investigated the relationship between household income and individual consumption choices on cigarettes. Binkley (2010) examines the hypothesis that cigarette consumption is positively correlated with income since low income individuals are less willing to sacrifice present pleasure for future health. A logistic regression approach utilizing data from the Behavioral Risk Factor Surveillance System (BRFSS) shows that at annual household income levels lower than $20,000, the probability of starting smoking has a positive relationship with income; but with annual income higher than $25,000, the relationship becomes negative. Conversely, the probability of quitting smoking decreases as income increases at low income levels but increases with income at higher income levels. In general, the probability of starting smoking has a negative relationship with income, while the probability of quitting smoking positively correlates with income. The author asserts that higher income is related to a low rate of time discounting, and individuals with a low rate of time discounting have a stronger incentive to make positive health choices. The correlation between income and cigarette consumption is attributed to the concept of time preference.

Scharff and Viscusi (2011) compose an interesting application illustrating the effect of income on cigarette consumption through the lens of time preference. They utilize a rational model of individual choice to estimate the implicit rates of time preference of smokers and nonsmokers by examining fatality risk-wage decisions in the labor market. Their study follows a multi-period theoretical model of occupational risk and a two-stage empirical model developed by Viscusi and Moore (1989), and obtain the tobacco use data from the Current Population Survey (CPS), the fatality data from the National Traumatic and Occupational Fatality (NTOF), as well as other

data from the Bureau of Economic Analysis on state-specific per capita income (PCI) and the National Center for Health Statistics (NCHS). They find that smokers have a higher rate of time preference compared to nonsmokers, which means that smokers value present utility more while discounting future utility more heavily.

### 2.2.2 Income and Excessive Alcohol Consumption

Alcohol consumption is another addictive good often used in empirical studies to investigate the relationship between income and health choices. Ettner (1996) finds that individuals with higher income significantly exhibit greater alcohol intake; however, increasing disposable income does not reduce alcohol consumption and behaviors. Hamilton et al. (1997) utilize data from the General Social Survey (GSS) and find that alcoholic abusers, defined as those who drink at least once a week in the previous 12 months and drink 8 or more drinks on one or more days in the previous week, have lower incomes than moderate drinkers. One drink is defined as one and a half ounces of liquor; or more specifically, one small glass of wine or one pint bottle of beer. Similarly, Cerdá et al. (2011) find that people with lower lifetime income are more likely to possess lower drinking levels. They investigate the data from the Panel Study of Income Dynamics (PSID) with 6,729 respondents aged 18-59 in 1996. However, in their study the next year with 3,111 respondents aged 30-44, the authors find the opposite result: higher income is associated with higher odds of moderate drinking relative to abstinence or alcoholism. Coincidentally, a study of alcoholic behavior (Kenkel et al., 1994) finds that increased heavy drinking is associated with increased income, which contradicts their other findings on the income effects of drinking. Taken together, these studies suggest that the relationship between income and alcohol consumption is inconclusive.

Hu and Stowe (2016) further investigate the relationship between household income level and individual alcohol consumption behavior through the lens of time preference. The model considers income as a budget constraint in the current period as well as a component of future utility, and those with lower income discount future utility more heavily. Utilizing data from the BRFSS and a multinomial Logit empirical model, two drinking behaviors (the frequency of alcohol consumption and the frequency of excessive alcohol consumption), are examined by measuring the influence from affordability and the effect of time preference, respectively. The results show that the frequency of alcohol consumption is positively correlated with income, but excessive alcohol use mostly occurs among the lower income population. Since increasing income is claimed to indicate a weaker time preference, or weighing future utility more heavily, this outcome is consistent with the discussion from Keough et al. (1999). They find that a present time perspective is positively related to self-reported alcohol use, whereas a future time perspective implies less frequent alcohol consumption.

Aside from the role of income in determining individual alcohol consumption, researchers have examined various other factors that might contribute to the decision to use alcohol. For example, Droomers et al. (1999) find that excessive alcohol consumption is more common among less educated individuals. Wilsnack et al. (2000) find gender differences in alcohol use; the rate of heavy drinking is higher in men than in women. Herd (1990) finds evidence of ethnicity effects. For example, despite the similar proportions of abstainers and abusers for blacks and for whites, the determinants of drinking choices are not the same. The effect of income varies by race. Frequent heavier drinking is positively associated with high income as well as the poorest among whites. In contrast, rates of heavier drinking of black men is highest

are those with only modest incomes. Both low income blacks and those with the highest incomes appear to drink less heavily.

### 2.2.3 Income and Obesity

Economics studies on obesity mostly concentrate on the consumption of healthy foods. Previous studies suggest that low income populations appear to choose more unhealthy diets, such as lower consumption of fruits and vegetables. Park et al. (1996) find that the demand of low income households for fruits and vegetables is more responsive to a marginal change in income than that of other households. When income increases, low income households increase their consumption on fruits and vegetables to a greater extent than high income households. Evidence presented by Stewart et al. (2003) suggests that poor households spend less on fruits and vegetables than other households; however, as income rises, they are unlikely to increase the consumption of fruits or vegetables. Cawley (1999) acknowledges that calories can be seen as an addictive good and consumers can form a rational addiction for calories.

Generally, energy-dense foods, mainly include fast food and have a high concentration of calories per bite, should be inferior because health is a normal good. Therefore, people consume less energy-dense foods when income increases (Philipson and Posner, 1999; Philipson, 2001). To some extent, the demand for calorie-dense foods reveals the demand for health. Philipson and Posner (1999) hypothesize that in developed countries, an increased income will result in weight loss because of an increased demand for health. Additionally, Lakdawalla and Philipson (2002) hypothesize that the source of income affects weight through different effects. Earned income has both an income effect and an effect on labor or effort, while unearned income only has an income effect. Cawley and Ruhm (2011) conclude that higher income individuals have a relatively low rate of obesity.

Individuals with low socioeconomic status (as proxied by educational attainment or income level) are more likely to be obese since they tend to have less access to affordable healthy food items (Bao and Chaloupka, 2007; Cawley and Ruhm, 2011; Cutler and Lleras-Muney, 2010). The obesity rate is higher among children who have a working mother, probably because working mothers spend less time on certain activities related to health and nutrition, such as cooking and grocery shopping as well as supervising or eating with children (Cawley and Liu, 2007; Fertig et al., 2009).

### 2.2.4 Income and Physical Activity Participation

While cigarette and alcohol consumption have received much attention in health choice research, increasingly, physical inactivity has become a major public health concern as it affects both physical health and psychological well being. In fact, physical inactivity is rated among the top ten leading causes of death in high-income countries (World Health Organization, 2002). Physical activity includes a wide range of activities such as occupational activity, housework, and sports and exercise, the latter of which is the focus of this dissertation. Economists study physical inactivity by accounting for budget constraints (i.e. income and time constraint) and perceived benefits, among other factors.

A comprehensive study of economic determinants on physical activity participation is attributed to Humphreys and Ruseski (2006). Their study employs recreation and leisure demand models to investigate the relationship between physical activity and economic factors like income. The model generates the prediction that the effect of income on the participation decision and the effect on the amount of time spent on physical activity may work in opposite directions. Their results from empirical analysis of BRFSS data support the prediction that the likelihood of

participation increases with income, but time spent on physical activity declines with income.

In accordance with Humphreys and Ruseski, a statistical investigation on residents of England utilizing random effects probit models indicates that household income has a positive and significant effect on physical activity participation (Farrell and Shields, 2002). Another study in Spain (Garcia et al., 2011) leads to a result that a gender difference exists in the determinants of physical activity participation. Nonlabor income is positively related to the probability of females in physical activity participation; nonetheless, this effect is not significant for males. Overall, most research supports the result that physical activity participation is positively related to household income (Anokye et al., 2012; Breuer and Wicker, 2008; Farrell and Shields, 2002; Garcia et al., 2011; Meltzer and Jena, 2010).

Studies on other socioeconomic characteristics such as education and employment, which are usually positively related to income, represent another area of interest. A higher level of education may imply better understanding of the health benefits from an adequate amount of physical activity. Therefore, a positive relationship between education and physical activity participation is exhibited (Downward and Riordan, 2007; Humphreys and Ruseski, 2007). However, being employed seems to be negatively associated with the amount of time that individual spends on physical activity (Farrell and Shields, 2002), likely because the occupational time limits the leisure time to spend on sports and exercise.

Other than the effects of income, education and employment, the influences of age and gender on physical activity participation are widely discussed in the economics literature. Most studies show that the likelihood of participating in physical activity decreases with age (Breuer and Wicker, 2008; Downward, 2007;

Downward and Riordan, 2007; Farrell and Shields, 2002; Humphreys and Ruseski, 2006; Meltzer and Jena, 2010). Males are more likely to participate in exercise or sports and spend more time on it than females (Farrell and Shields, 2002; Humphreys and Ruseski, 2006, 2007; Meltzer and Jena, 2010). Other health related choices might also affect the choice of physical activity participation. Interestingly, alcohol consumption positively relates to physical activity participation, while cigarette consumption shows a negative relationship with it (Farrell and Shields, 2002). People consistently choose their lifestyle to be healthy - drink moderately, exercise more - or unhealthy - smoke a lot and exercise little. Finally, race, household profile (household size, presence of children, etc.), marital status, and physical and mental health status are other factors that have been widely considered; however, the results are mixed.

Although previous research has provided different perspectives on economic determinants of participation and time spent on physical activity, there is still considerable space for improvement. One obvious extension is to link the effect of income with the concept of time preference in the theoretical model. This enables to interpret the income influence in a different way: in addition to its function as a budget constraint, income implies a rational expectation for future life, which may affect the physical activity decision. Secondly, while studies on physical activity have been conducted in Europe and in developed countries, this research focuses on an empirical investigation on the United States population and the China population. The comparison between a developed country and a developing country furthers the understanding of the role economic factors play in determining health choices.

### 2.3 Physical Activity Study in United States

Most of the empirical studies in United States that investigate the determinants of physical activity participation are based on the data from three large scale

nationwide surveys: the National Health and Nutrition Examination Survey (NHANES), the National Health Interview Survey (NHIS) and the Behavioral Risk Factor Surveillance System (BRFSS). In addition to the three nationwide datasets, some researchers conducted independent regional health surveys designed for their specific research purposes. In this section, the three regional surveys which benefit this study are briefly discussed: the San Diego Health & Exercise Survey in west coast area, the Yale Physical Activity Survey in east coast area, and the Minnesota Leisure Time Physical Activity Questionnaire representing the Midwest.

### 2.3.1 Nationwide Surveys

Physical inactivity is one of the major health concerns in United States that drives significant research interests. The Behavioral Risk Factor Surveillance System the (BRFSS), the National Health and Nutrition Examination Survey (NHANES) and National Health Interview Survey (NHIS) are three most well established and widely acknowledged national surveys that investigate the determinants of physical activity and other health related issues from which most of the economic studies on health problems are derived. They are cross-sectional surveys with different emphases. Both NHANES and NHIS are face-to-face household surveys, while BRFSS is a telephone-based survey.

NHANES is a program of the National Center for Health Statistics (NCHS) under the Centers for Disease Control and Prevention (CDC) that started in the early 1960s and has a nationally representative sample of about 5,000 persons each year. It is a unique survey that combines interviews – which consist of demographic, socioeconomic, dietary, and health-related questions, and physical examinations, which include medical, dental, physiological measurements and laboratory tests to study the health and nutritional status of adults and children in United States. This

survey collects information on health choices like smoking, alcohol consumption, physical fitness and activity, and dietary intake, as well as risk factors that may increase the chance of addiction in one's health choice. The sample for the survey is selected to be representative of the entire U.S. population. To produce reliable statistics, NHANES oversamples persons 60 and older, African Americans, and Hispanics.

The National Health Interview Survey (NHIS) is a nationally representative survey of the civilian non-institutionalized population of the United States. The survey interviews approximately 43,000 households consisting of approximately 106,000 individuals every year. This study collects data on a broad range of health topics as well as possible demographic and socioeconomic determinants of these health characteristics. The questionnaire consists of core questions and supplementary questions. The core questionnaire focuses on basic health and demographic information, while the supplementary questions address health topics responding to new public health data needs and thus may be changed year to year. Physical activity is investigated in the core section of the questionnaire. NHIS is another major program conducted by NCHS. Similar to NHANES, it oversamples Black and Hispanic respondents to provide reliable statistics.

BRFSS is a telephone survey that has tracked health conditions and risk behaviors in the United States yearly since 1984. It is the youngest among the three surveillance systems but has the biggest sample size. In contrast to NHANES and NHIS, this survey comprises only respondents 18 and older. It is based on probability sampling from state health departments and thus is able to provide state-based estimates on health measures such as physical activity. Its questionnaire includes a core component with questions on demographic characteristics and health behaviors,

and emerging core questions on current areas of interest as NHIS does; additionally, BRFSS questionnaires permit state optional modules, which allow for state added questions. It provides the most flexibility and diversity on health studies.

Most of the studies in United States that investigate the determinants of physical activity participation are based on the data from the above three surveys. For example, the study of Humphreys and Ruseski (2006) using BRFSS data suggests that the likelihood of physical activity participation increases with income, but time spent on physical activity declines with income. They build a model of participation based on the SLOTH framework proposed by Cawley (2004) and utilize a two-step Heckman procedure for empirical analysis. The study of Hu and Stowe (2016) with a binary Logit method agrees with this finding that the higher income population is more likely to participate in physical activity. Most research utilizing NHANES data supports the result that physical activity participation is positively related to household income (Lopez and Hynes, 2006; Meltzer and Jena, 2010), and this result is supported by the empirical studies on NHIS data (Caspersen et al., 1986). In general, evidence from previous research agrees that household annual income and the likelihood of participating in physical activity are positively correlated.

### 2.3.2 San Diego Health & Exercise Survey

The San Diego Health & Exercise Survey is a mail survey conducted in 1986 by James Sallis among a randomly selected sample of 6,000 residents in San Diego yielding 2,053 respondents. Their questionnaire has a similar framework to NHANES that measures different types of physical activities, investigates demographic, socioeconomic, dietary, health self perception questions, and also includes questions on risk behaviors such as cigarette and alcohol consumption. A follow-up survey was implemented two years later and yielded 1,701 valid responses. It assessed physical

activity patterns and those potential determinants that draw most theoretical interest or are cross-sectionally associated with physical activity in the baseline survey, and a selected subset of determinant variables that are likely to change during the follow-up interval.

This survey has generated numerous research papers. However, the effect of income on physical activity choice was not a focus of these studies, although income variable is included in the survey. This leaves room for future research.

### 2.3.3 Yale Physical Activity Survey

Loretta DiPietro conducted the Yale Physical Activity Survey in 1988 under Yale University and CDC. This project measures physical activity of older adults (age 60 years and over) in Connecticut representing various socioeconomic backgrounds. The survey is composed of two sections. The first section asks the individual to estimate time spent on work, yard work, exercise, and recreational activities in a typical week during the last month. The second section consists of questions with categorical responses to assess current participation in specific type of physical activity, such as high intensity activity or leisurely walking.

This study does not include income information. Nevertheless, it is a prominent study that provides methods and examples to measure physical activity from a medical perspective, and benefits economists for conducting physical activity related studies based on its survey design and measurement.

### 2.3.4 Minnesota Leisure Time Physical Activity Questionnaire

Another popular physical activity questionnaire used in health related research is the Minnesota Leisure Time Physical Activity Questionnaire. It contains questions pertaining to leisure time activities of nine general categories: walking and miscellaneous, conditioning exercise, water activities, winter activities, sports, lawn

and garden activities, home repair activities, fishing and hunting, and household activities. Individuals are asked to complete the survey with the frequency and the average duration of time spent on each activity in the past four weeks. This questionnaire has been widely applied in a lot of studies on physical activity (Richardson et al., 1994; Elosua et al., 2000). Together with the Yale Physical Activity Survey, the Minnesota Leisure Time Physical Activity Questionnaire provides a foundation for a comprehensive design of physical activity analysis for future economic research on income related issues.

In conclusion, in the Unites States, in addition to the large national health surveys, regional surveys for resident physical activity participation have been carried out by Yale in the east-coast area, in San Diego in west-coast area and Minnesota in the Midwest. However, the influence of income on physical activity participation, which has been studied by European researchers, has received little attention among the United States population. In the following chapters, I am interested in the determinants of consumer choice on physical activity with an emphasis on the effect of income, taking the examples of U.S. population from NHANES data, the Chinese population based on the China Health and Nutrition Survey (CHNS) data (discussed next), and Kentucky residents with an independently conducted health and exercise survey.

**2.4 Physical Activity Study in Other Countries**

The previous section briefly summarizes the physical activity studies focused on residents in the United States. The first part of this section further discusses studies on the state of physical activity participation, data sources and socioeconomic determinants in China; the second part summarizes previous research on physical

activity in more developed and developing countries, aiming at providing a comparison between countries in different stages of development.

### 2.4.1 Physical Activity Study in China

Most of the research published in English studying physical activity in China utilizes data from the China Health and Nutrition Survey (CHNS). CHNS is an international collaborative project led by the Carolina Population Center at the University of North Carolina at Chapel Hill and investigates nutrition and health behaviors in China. It is a longitudinal study first launched in 1989 and repeated in 1991, 1993, 1997, 2000, 2004, 2006, 2009 and 2011. This project consists of different surveys such as the household survey, child survey, adult survey and community survey; the questions regarding physical activity was first introduced in the Household Survey in 1997, became part of the Physical Examination in 2000, and was included in adult surveys in 2004 and the following years thereafter. Similar to NHANES, it collects demographic characteristics such as income, educational attainment, age and gender, and health behaviors such as cigarette and alcohol use, sugary drink consumption and physical activity participation. For physical activity, it includes an activity sheet measuring the frequency and time spent on different types of active or sedentary activities; it also classifies the intensity of physical activity as vigorous or moderate.

Attard et al. (2015) published the most recent study using CHNS data. Their study utilizes the data from 1991 to 2009 of adults aged 18-75 years and uses zero-inflated negative binomial regression models to investigate the associations between urbanicity and income with physical activity in urbanizing China. Urbanicity is the degree to which a given geographical area exhibits urban characteristics, such as population size, population density, type of occupations and percent employed in

agriculture, number of markets, diversity of markets, infrastructure, higher average education and income and greater diversity in education and income (Vlahov and Galea, 2002; Jones-Smith and Popkin, 2010). The results of Attard et al. show that in low urbanicity areas (which means a relatively lower household annual income), income negatively relates to total time on physical activity in 1991, but they are positively correlated in 2000. Meanwhile, in high urbanicity areas (which implies a relatively higher household annual income), the relationship between income and total physical activity is positive at all time points and is statistically significant at most time points after 1997. They also find that leisure physical activity is the only type of physical activity that increases over time; however, a shortcoming of this study is that the relationship between income and leisure physical activity is not investigated.

Similarly, the study of Bauman et al. (2011) finds a positive association between income and physical activity in China. Higher income individuals in China are twice as likely to participate in leisure time physical activity compared with less affluent individuals, while they are less likely to be physically active at work or engaging in active commuting such as walking or riding a bike to work.

In addition to CHNS and its relevant studies, regional studies on physical activity in some particular areas are also available. Shu et al. (2007) conduct a Shanghai Men's Health Study to evaluate physical activity patterns and their association with socioeconomic status. Their study has a sample of 61,582 Chinese men ages 40-74 living in eight communities of urban Shanghai, China. They find that high income people are significantly more likely to participate in leisure time physical activity such as exercise or sports, whereas the physical activity of low income people is associated with transportation and daily living activities. This result is consistent with previous findings (Hu et al., 2002) that income level is positively associated with

exercise and housework but inversely associated with commuting-related physical activities, which is based on a 1996 cross-sectional population survey in urban areas of Tianjin. This study, which yields a sample of 3,976 respondents, is funded by the World Bank and investigates physical activity during leisure time and commuting as well as associated demographics and health behaviors.

Taken together, previous studies in China on physical activity demonstrate that that the positive relationship between income and either with the likelihood to participate in leisure time physical activity or with the total time spent on it is robust. However, no previous research has investigated the relationship of income with both; it remains unclear whether the positive effect of income occurs simultaneously on the frequency and duration time on leisure time physical activity participation. This will be addressed in the current study in later chapters.

### 2.4.2 Comparison Across Countries

A statistical investigation on England's population utilizing random effects probit models indicates that household income has a positive and significant effect on physical activity participation (Farrell & Shields, 2002). In addition, Øvrum's study (2011) on the Norwegian population shows that individuals with higher income and socioeconomic status are more likely to choose a healthier life style including higher consumption of fruits and vegetables as well as more physical activity participation. However, the study of Kuvaja-Köllner et al. (2013) in Eastern Finland suggests that the impact of income on the time spent on physical exercise is not significant for older adults aged 55–74 years, although income has a significant impact on health outcomes.

A study in Spain (Garcia et al., 2011) identifies a gender difference in the determinants of physical activity participation. For females, non-labor income is negatively related to participation in physical activity; nonetheless, this effect is not

significant for males. The gender difference in the determination of income on physical activity participation is also discovered by Bauman et al. (2011). They find that in Philippines, women in the highest income group are more likely to be physically active during leisure time. Hallal et al. (2005), which focuses on middle income countries, find that leisure time physical activity (such as walking) is more frequent among high income people in Brazil.

Synthesizing the results from this literature, leisure time physical activity such as sports and exercise are more common among high income people, regardless of whether they reside in developed countries or developing countries. Nevertheless, the influence of income on the time spent on leisure time physical activity is inconclusive. This leaves room for future research.

**2.5 Empirical Methodology Discussion**

Physical activity participation decision modeling mainly includes logistic models and two-step Heckman models as well as multiple classification analysis (Anokye et al., 2011; Downward, 2007; Downward & Riordan, 2007; Farrell & Shields, 2002; Humphreys & Ruseski, 2006, 2007; Shu et al., 2007). Relatively few studies consider the time spent on physical activity participation or the frequency of such participation (Downward & Riordan, 2007; Humphreys & Ruseski, 2006, 2007).

Logistic regression is widely applied in physical activity participation studies because the participation decision is often a binary outcome. It estimates the probability that the discrete dependent variable equals 1 as a function of the determinants (Shu et al., 2007). Alternatively, Probit models are used when assuming the dependent variable follows a standard normal distribution instead of a logistic distribution (Anokye et al., 2011). Other generalized linear methods include multinomial Logit models when the dependent variable is a categorical response, such

as the frequency of physical activity participation; and count data models such as a Poisson model (Øvrum, 2011), the negative binomial model (Øvrum, 2011), and two-part hurdle models.

When modeling participation in physical activity that includes both a participation decision and a time decision, the selectivity may play an important role in the process. The two-step Heckman procedure accounts for selectivity in the empirical analysis of participation in physical activity. The first stage of the Heckman procedure models the participation decision and requires estimating a model with a discrete dependent variable that is equal to 1 if the individual participates in some physical activity and is equal to 0 if the individual does not participate. The second step of the Heckman approach requires individuals make a choice about how much time to spend on physical activity conditional on the decision to participate (Downward & Riordan, 2007; Humphreys & Ruseski, 2006, 2007). The Heckman selection model can also be extended to a Heckman probit model when the dependent variable of the second stage is a binary response, or a Heckman ordered probit model when the frequency of physical activity participation is estimated in the second stage.

Longitudinal data are sometimes available for physical activity participation studies, requiring a different empirical approach. A panel data model is applied when the data in a regression are repeated observations of a respondent over time. It controls for not only observed explanatory variables over time but also unobserved aspects from the repeated behaviors that affect individual decisions. The unmeasured aspect of each individual is called heterogeneity. Fixed effects models and random effects models are commonly utilized in panel data analysis to address heterogeneity. Random effect logit models and random effects probit models (Farrell & Shields, 2002) are utilized to estimate the probability of the decision to participate in physical

activity. In nonlinear models, another important reason to use random effects models is that fixed effects models will result in the loss of all observations that do not vary, e.g. individuals that always participate in physical activity or never participate, and remove many relevant explanatory variables. By utilizing random effects models, the time-invariant variables such as demographic and survey variables are still able to provide information in the regression.

## Chapter Three: Theoretical Foundation

In the view of neoclassical economics, rational consumers choose between consumption and leisure to maximize their utility function subject to budget and time constraints. Becker (1965) develops a model assuming that households combine time and market goods to produce commodities that increase their utility. The full cost of current consumption includes possible negative consequences occurring at a later period of life; however, with heavy discounting of future utility, these negative consequences are valued less. Thus, people with strong preferences for the present, or those that more highly discount their future utility, are more likely to consume goods which are detrimental to one's future health.

These "goods" may be addictive (such as cigarettes or alcohol), of low nutritional value (such as sugary or other highly processed foods), or otherwise affecting future health, such as physical inactivity, which is the focus of this research. People sacrifice physical activity time today for other inactive leisure activities that provide more utility in the current period; however, physical inactivity can be assumed to have harmful effects on human health in the next period and undermine future utility. Having a higher income increases the future cost of being physically inactive, since its negative effects on health would cause a greater loss in future income and thus lifetime income. Consequently, it may be that lower income individuals, who have lower expected lifetime income, may more heavily discount future utility and thereby participate in lower levels of physical activity.

### 3.1 Lifetime Aggregate Utility Function

According to Becker's theory, individuals see their lifetime utility as the aggregation of two periods: the utility of the current period and the expected utility of

the future period. They make consumption decisions to maximize their lifetime aggregate utility as in equation (3.1).

$$U = U_0(c_0, s_0) + \beta_0(I_0) f(s_0) U_1\big(g(I_0)\big) \qquad (3.1)$$

The first part of the equation represents the utility of the current period, period 0. The current utility $U_0$ is determined by the consumption of ordinary goods $c_0$, which provides utility without having any potential health effects, as well as consumption of health goods $s_0$ that have positive or negative future health consequences, such as cigarettes, alcohol and physical activity. Utility $U_0$ increases with the consumption of ordinary goods, so ${U_{0,c}}' > 0$. The sign of ${U_{0,s}}'$ depends on the type of good $s_0$, which will be discussed later in sections 3.2.1 and 3.2.2.

The second component of equation (3.1) represents the current value of expected future utility in period 1. $U_1$ is the future utility function. The term $\beta_0$ is the discount factor in the period that the consumption decision is made; it is assumed to increase with income level $I_0$. In other words, a smaller $\beta_0$ suggests a heavier discounting of future utility, and consequently a stronger current time preference; a larger $\beta_0$ suggests less discounting of future utility, which can be interpreted as having a weaker preference for the current time period.

Assume that the consumption of $s_0$ culminates in a long-run health effect, so that the utility in the future period will be further affected by an additional term $f(s_0)$, which measures how the chosen level of $s_0$ in period 0 improves or reduces health and utility in period 1, with $f(0) = 1$ and $f'(s_0) > 0$ if $s_0$ has a positive health effect, or $f'(s_0) < 0$ if $s_0$ has a negative health effect. This dissertation investigates how consumer decisions are made in the current period; how consumer decisions are made in future period is of no interest. Therefore, the utility in the future period could be modeled as depending solely on expected income, which is assumed to be a function

of the income in period 0, or $U_1 = U_1(g(I_0))$. $I_0$ is income in period 0, and $g'(I_0) > 0$.

### 3.2 Utility Difference Under Different Consumption Levels

This study focuses on the decision in the first period whether to consume $s_0$ or not. Incomes, prices, and preferences are assumed to be fixed within that period. It follows the model in Binkley (2010) to determine the optimal consumption level of $s_0$ by comparing the utilities with different quantities of consumption. Given two levels of consumption, $s_0$ and $s_0^l$, where $s_0^l < s_0$, the difference $D$ in lifetime utilities under the two different consumption levels can be expressed as in equation (3.2).

$$\begin{aligned} D &= \left[U_0(c_0, s_0) + \beta_0(I_0) f(s_0) U_1\left(g(I_0)\right)\right] - \left[U_0\left(c_0, s_0^l\right) + \beta_0(I_0) f\left(s_0^l\right) U_1\left(g(I_0)\right)\right] \\ &= \left[U_0(c_0, s_0) - U_0\left(c_0, s_0^l\right)\right] + \beta_0(I_0) U_1\left(g(I_0)\right)\left[f(s_0) - f\left(s_0^l\right)\right] \\ &= D_0 + D_1 \qquad (3.2) \end{aligned}$$

Equation (3.2) rearranges the utility difference so that it is expressed in terms of current utility variation $D_0$ and expected future utility variation $D_1$.

#### 3.2.1 Consumption of Costly Unhealthy Goods

To illustrate this representation, suppose that $s_0$ is a widely recognized addictive good like cigarettes or excessive alcohol that contributes to current pleasure but has potential negative health effects and diminishes future utility. In equation (3.2), $D_0$ is the utility gained in the current period with a higher consumption of $s_0$ (so that $D_0$ has a positive sign), and $D_1$ represents the future utility loss due to higher $s_0$ consumption in the first period, which suggests that $D_1$ has a negative sign. Obviously, if the absolute value of $D_1$ exceeds $D_0$, $D < 0$. In other words, the utility improvement today is mitigated by the health effect in the future, and rational individuals will choose less and even no consumption of $s_0$. The comparison of $D_0$ and $D_1$ is a key focus.

Now, consider the relationship between income and the values of $D_0$ and $D_1$. To begin with, it is straightforward that the absolute value of $D_1$ is increased with $U_1\big(g(I_0)\big)$, since $f(s_0) - f\big(s_0^l\big)$ is negative and $\beta_0$ is positive. Hence, a higher $I_0$ implies both a higher $\beta$ and a higher $g(I_0)$, and consequently, a larger $U_1$. So, in absolute value, $D_1$ is increasing in $I_0$. Next, consider how $D_0$ relates to income. Standard consumer theory demonstrates that an individual maximizes utility by choosing consumption levels where the marginal utility of a good is greater than or equal to its price times the marginal utility of income. Income has a decreasing marginal utility, and the price is assumed to be fixed. As a result, an individual will accept a good for a smaller marginal utility and thus a larger quantity when income increases. This suggests that $D_0$ is also increasing in $I_0$.

Because $D_0$ and $D_1$ (in absolute value) are both increasing in $I_0$, whether the increase in current utility compensates for the future utility loss due to the negative health effect determines whether higher levels of $s_0$ will be consumed. The relative strengths of the two components of the net effect vary over the income distribution. As mentioned previously, the influence of $D_1$ tends to be minor at low incomes since future utility is heavily discounted. Moreover, $D_0$ plays a dominant role among low income households since a small increase in income enhances their ability to consume more $s_0$. Therefore, individuals with low income have $D > 0$ and are more likely to consume $s_0$. Conversely, high income individuals have $D < 0$ and are less willing to consume a good that diminishes future health. The justification for this is two-fold. First, high income households are relatively less budget constrained, so the effect of $D_0$ is lessened. Second, $D_1$ is increasing in income, since the high income population is more concerned about future utility.

Taken together, this model predicts that the consumption of potentially unhealthy goods is decreasing in income. More generally, the consumption of unhealthy goods increases with income if affordability is main focus and shrinks when future health becomes the major consideration.

### 3.2.2 Consumption of Non-costly Healthy Goods

Now consider a diametrically-opposed situation where $s_0$ represents physical activity; this type of good is beneficial to future health and utility but causes current disutility, so ${U_{0,s}}' < 0$. This assumption is intuitive because if it benefits both current utility and future health at no monetary cost, a rational person will unambiguously choose to consume and there is no need to model this choice.

Specifically, the consumption of $c_0$ is subject to income level $I_0$, where ${c_0}'(I_0) > 0$. It is assumed that physical activity does not involve monetary cost, so the choice of $s_0$ is subject only to the time constraint $t_0$, where ${s_0}'(t_0) > 0$. This is an application of the SLOTH model by Cawley (2004), which assumes that an individual's utility depends on his or her weight, health, food, other goods and time spent sleeping (S), at leisure (L), at occupation (O), in transportation (T), and in home production (H). Physical activity and sports are considered leisure activities. Therefore, given $t_0^l < t_0$, (3.2) could be rewritten as

$$
\begin{aligned}
D = &\left[U_0(c_0(I_0), s_0(t_0)) + \beta_0(I_0) f(s_0(t_0)) U_1\big(g(I_0)\big)\right] \\
&- \left[U_0\big(c_0(I_0), s_0(t_0^l)\big) + \beta_0(I_0) f\big(s_0(t_0^l)\big) U_1\big(g(I_0)\big)\right] \\
&= \left[U_0(c_0(I_0), s_0(t_0)) - U_0\big(c_0(I_0), s_0(t_0^l)\big)\right] \\
&+ \beta_0(I_0) U_1\big(g(I_0)\big)\left[f(s_0(t_0)) - f\big(s_0(t_0^l)\big)\right] \\
&= D_0 + D_1 \qquad (3.3)
\end{aligned}
$$

Here $D_1$ is the expected utility gained in period 1 attributed to a higher level of consumption of $s_0$ in the current period. The term $\left[f(s_0(t_0)) - f\big(s_0(t_0^l)\big)\right]$ has a

positive sign because $f'(s_0) > 0$ and $s_0'(t_0) > 0$. A higher $I_0$ implies both a higher $\beta_0(I_0)$ and a higher $g(I_0)$ and consequently a larger $U_1$, which means $D_1$ is increasing in income $I_0$. Hence, the utility gained from more consumption of $s_0$ is larger for those with higher income. Now, the analysis of $D_0$ and the comparison between $D_0$ and $D_1$ become a key focus.

The difference in utility in period 0 consists of two components: the difference in utility due to the change in consumption of ordinary goods $c_0$, and the difference in utility from the change of consumption on $s_0$. Since it is assumed that $s_0$ has no monetary cost, the different levels of $s_0(t_0)$ and $s_0(t_0^l)$ should not affect the consumption of ordinary $c_0(I_0)$. This suggests that the difference in utility due to consumption of ordinary goods is zero.

Next, consider the change in utility in the first period, $D_0$. Recall the assumption that $s_0'(t_0) > 0$ and $U_{0,s}' < 0$. Therefore, for $t_0^l < t_0$, $s_0(t_0^l) < s_0(t_0)$ and $D_0$ is negative. A higher consumption of $s_0$ results in lower utility in period 0.

Taken together, with a higher level of consumption $s_0(t_0)$ that has a positive health consequence, $D_1$ is a positive term and increases with income while $D_0$ is negative and does not change with income. Hence, with a higher income, the expected future utility benefit is more likely to compensate for the utility loss in the current period, and the lifetime aggregate utility increases.

Therefore, this model predicts that the consumption of healthy and monetarily non-costly goods increases with income, because higher income individuals value future health more due to its influence on expected future utility.

In sections 3.2.1 and 3.2.2, the theoretical model predicts that lower income individuals are less likely to invest in healthy consumption choices, because they have a stronger time preference and discount future utility more heavily. Specifically, this

model suggests that lower income is negatively associated with physical activity participation and positively related to cigarette or excessive alcohol consumption. The empirical examples analyzed in this dissertation are the participation in physical activity in United States, in Kentucky and in China, and are in support of the theoretical finding presented in 3.2.2.

## Chapter Four: The Relationship between Household Income and Physical Activity Participation in United States: A Heckman Selection Model on National Health and Nutrition Examination Survey

This chapter presents the empirical study on physical activity participation in the United States. A two-step Heckman selection model is estimated utilizing data from the National Health and Nutrition Examination Survey (NHANES). The results show that the probability of physical activity participation increases with income, but the relationship between average time spent on physical activity and income is ambiguous.

### 4.1 Data

The data for this study comes from the 2001-2006 National Health and Nutrition Examination Survey (NHANES). This is a unique survey that combines interviews and physical examinations to study the health and nutritional status of adults and children in the United States that started in the early 1960s. It is conducted by the National Center for Health Statistics (NCHS) under the Centers for Disease Control and Prevention (CDC) and has a nationally representative sample of about 5,000 persons each year. Data from 2001 to 2006 are used since questions concerning physical activity were optional before 2001 and changed after 2006. This dissertation only utilizes data for respondents 18 and older.

A two-step Heckman selection model is estimated; the model is described in the next section. The dependent variable for the first step represents whether the respondent participated in physical activity in the past 30 days or not. The original responses come from two questions from the survey questionnaire, measuring

whether the respondent participated in vigorous physical activity[1] and whether the respondent participated in moderate physical activity[2] in the past 30 days. From this, a new binary variable is created as the dependent variable to assess the participation choice based on the two variables; the new variable takes the value 1 if the response to either of these questions is positive. In other words, participation in vigorous activity or participation in moderate activity or both are all counted as participation in physical activity. Table 4.1 reveals the relationship between the two survey variables and the newly created dependent variable.

In the second step of Heckman model, the dependent variable is the natural logarithm of average time spent on physical activity per day in past 30 days. The original survey includes an individual activity sheet recording detailed information of leisure time (i.e., not work-related) physical activity for each respondent such as numbers of times participating and the average duration each time. According to the activity sheet, the total time spent on physical activity per month is available by multiplying the number of times that an individual participated in physical activity and the average duration of activity each time; furthermore, the average time spent on physical activity is obtained by dividing total time by 30. This study takes the natural logarithm of the average time to have the dependent variable more closely follow a normal distribution. The Heckman model has 11,249 observations in the first step, with 64.9% of the respondents reported having participated in physical activity in past 30 days. The second step of the model has 7,298 observations, since respondents who did not participate in physical activity are truncated in this step.

[1] Vigorous activities are defined as exercise, sports and physical active hobbies that last for at least 10 minutes and cause heavy sweating or large increases in breathing or heart rate, e.g. running, lap swimming, aerobics classes, fast bicycling.

[2] Moderate activities are defined as exercise, sports and physical active hobbies that last for at least 10 minutes and cause only light sweating or a slight to moderate increase in breathing or heart rate, e.g. brisk walking, bicycling for pleasure, golf, dancing.

The analysis includes a broad set of independent variables. First, annual household income is included as a potential factor in determining physical activity behavior. The NHANES data measures household income with eleven intervals. These are combined and re-categorized into eight income variables. The dummy variable indicates whether an individual's income lies in this interval or not. Other demographic variables are also considered as explanatory variables, including age, race, gender, educational attainment, marital status, and household size. There are five race variables, three education variables, and three marital status variables. Other than income and demographic variables, factors included in this study measure whether the respondent has the highest income in the household, as well as self-reported physical health and mental health conditions. These factors are assumed to influence individual risk behaviors and consequently are included in this physical activity analysis. A full set of categorical variable definitions is given in Table 4.2.

Descriptive statistics for the full sample are presented in Table 4.3. The mean of physical activity participation choice is 0.649, which means approximately 64.9% of the sample did participate in physical activity in the time period under analysis. About half of the observations are males, and about half of the observation are whites. Physical and mental health conditions are measured by the self-reported number of unhealthy days in one month; thus, a higher value implies a worse physical or mental health.

Table 4.4 provides the independent variable means for the dependent variable conditional on the response. Individuals who participated in physical activity appear to have a lower average age, as well as better self-reported health conditions both physically and mentally. Among the lower income groups with annual household income up to $34,999, the percentage of respondents who participated in physical

activity is lower than the percentage of observations that did not participate in physical activity. Similarly, for the higher income groups with annual household income more than $55,000, the percentage of individuals who participated in physical activity is greater. The same situation occurs for educational attainment; the percentage of respondents who participated in physical activity is greater in the more highly educated categories. The right column of the table can also be interpreted as the sample mean for the second step of the model. For example, the average age of people who report their physical activity time is 42.79 years, which is slightly younger than the average age of the whole sample as 45.48; the percentage of males in the second step of estimation is 50.7%, slightly higher than the 48.8% in the first step.

**4.2 Empirical methodology**

According to the standard neoclassical theory of consumer utility maximization, it is assumed that an individual has preferences over his health and physical activity participation, which affects his health condition and the length of life. To determine the optimal participation frequency and the time spent on physical activity, the individual maximizes utility taking a set of factors into account, including income level, current health situation, demographic profile and so forth. This study is particularly interested in whether income level is significantly related to physical activity participation. Based on the theoretical discussion before, a reasonable expectation is that high income would contribute to physical activity participation, whether through an increasing likelihood of participation in physical activity or a greater amount of time spent on physical activity in a typical day, or both.

A two-step Heckman selection model is employed to investigate this relationship empirically. According to the theoretical framework and the attributes of

NHANES data, the sample of individuals reporting average time of physical activity is contingent upon the initial choice to participate in the sport. In other words, the sample of average physical activity time could be a non-random sample. Heckman selection model provides a means to correct the bias coming from non-randomly selected samples by implementing the inverse Mills ratio.

The two-step Heckman selection model is a sequential procedure. The first procedure is the selection procedure fitting a Probit model for the likelihood to participate in physical activity as given in equation (4.1).

$$Prob(z = 1) = \Phi(w'\gamma) \qquad (4.1)$$

In equation (4.1), $z$ is the event to participate in physical activity, where $z = 1$ stands for participation, and $z = 0$ means not participated. The term $\Phi$ is the cumulative distribution function of the standard normal distribution, $w$ is a matrix of independent variables which includes annual household income, age, race, gender, educational attainment, employment, marital status, household size, and self-reported physical health and mental health situations, and $\gamma$ is a matrix of unknown parameters.

The inverse Mills ratio is computed as given in equation (4.2), using the Probit estimates obtained from a maximum likelihood method. The term $\phi$ is the probability density function of the standard normal distribution, and $\hat{\gamma}$ are the estimates of regressor coefficients $\gamma$.

$$\hat{\lambda}(w'\gamma) = \frac{\phi(w'\hat{\gamma})}{\Phi(w'\hat{\gamma})} \qquad (4.2)$$

The second step is a regression model for the selected sample. It estimates a log-linear time equation indicated by equation (4.3) using Ordinary Least Squares. The dependent variable $y$ is the logarithm of average time spent on physical activity in a typical day, $x$ is a matrix of independent variables, and $\beta$ is the vector of unknown coefficients. The Heckman model may or may not have the same regressors

for the selection equation and regression. In this model, the independent variables for the first step and the second step are the same, which means $w$ and $x$ have exactly the same variables but only differ in the quantity of observations. Sample selection assumes that the discrete decision $z$ and the continuous decision $y$ have a bivariate distribution with correlation $\rho$. Note that equation (4.4) denotes an underlying logarithm of average physical activity time $y^*$, which is not observed if the respondent did not participate in physical activity; $\sigma$ is the standard deviation of random error term $u$. Finally, the inverse Mills ratio $\hat{\lambda}(w'\gamma)$ estimated in the first stage is included to solve sample selection problems.

$$E(y|z = 1) = x'\beta + \rho\sigma\hat{\lambda}(w'\gamma) \qquad (4.3)$$

$$y^* = x'\beta + u \qquad (4.4)$$

**4.3 Empirical Results and Discussion**

This section reports two sets of estimates: the estimation coefficients for physical activity participation choice, and the average time spent on physical activity. The estimation results for the Heckman selection model are presented in Table 4.5. Individuals with household annual income less than $75,000 have a lower probability of participating in physical activity than those with income higher than $75,000. Household size is negatively associated with physical activity participation choice. Reporting a worse physical or mental health condition are both negatively associated with the choice to participate in physical activity. Males are more likely to exercise than females, while older individuals are less likely to participate in physical activity. Whites have a higher probability of participating in physical activity than all other races except Hispanics. Educational attainment exhibits a positive relationship with physical activity participation choice. Those who have never been married are more

likely to participate in physical activity than those who are married, while all other marital status categories show no significant difference with those who are married.

Results on the average time spent on physical activity for those who do exercise are somewhat different. Income is negatively related to the time spent on physical activity at some points. Some lower income groups ($15,000-$19,999 & $25,000 to $34,999) spend more time on physical activity in a typical day than people with annual household income higher than $75,000. Similarly, more highly educated people spend less time on physical activity. Males spend more time on physical activity, and age is also positively related to the time spent on physical activity. Reporting a worse mental health condition has a negative effect on the time participating in physical activity. Blacks spend more time on physical activity than all other races. Finally, those who have never been married, as well as those do not have a partner, tend to spend more time on physical activity than the married ones.

To summarize, the probability of participating in physical activity increases with income, but the relationship between average time spent on physical activity and income is ambiguous. The first result is consistent with the hypothesis that low income individuals discount expected future utility and hence diminish the cost of reduced longevity. Therefore, the consumption of physical activity, a good with health benefits, for a low income person tends to be smaller. The theoretical foundation of the second finding requires further investigation. The time spent on physical activity is affected by the willingness to take part in physical activity as well as a time constraint. When the former factor plays a key role, people spend more time on physical activity; however, if the latter one dominates, people spend less time on physical activity. While higher income implies a stronger willingness to participate in physical activity, people with higher income may work longer hours and have work-

related social obligations which occupy time, leaving them with less time than the lower income groups to spend on physical activity. Also, an increase in time spent on physical activity does not necessarily imply a health benefit or utility gain. Since the marginal benefits of increased activity beyond a certain point may be diminishing, "more time" may not be optimal. This will be further discussed in Chapter 7.

Educational attainment affects physical activity in a similar way to income. Higher education increases the probability of choosing physical activity, but decreases the average time spent on it. This can be easily interpreted by the same mechanism that education enhances future utility because it is an investment in human capital (Becker et al, 1977). Individuals who do not have a spouse are more likely to take part in physical activity and spend more time on it. An increase in household size leads to a lower probability of physical activity participation, but has no effect on the time spent exercising. Older people are less likely to participate in physical activity, but they spend more time on it when they participate. Being the household reference person (the one with highest income in the household) is not observed to influence the physical activity behavior.

This study does find evidence of gender effects. Males are more likely to participate in physical activity and spend more time on it. Race also plays a role. Whites are more likely to take part in physical activity; however, among those who exercise, blacks spend the most time exercising compared with other races. Better self-reported physical health condition and better mental health condition both contribute to a higher probability of physical activity participation; in addition, better self-reported mental health condition increases the average time spent on physical activity as well. Being healthy means confidence for the future and makes people

place more value on future health, and thus invest more on healthy goods such as physical activity.

Though the results of the empirical estimation generally support the theoretical prediction, limitations leave room for future exploration. First, the estimation assumes that the rates of time preference are affected only by income. Becker et al. (1997) argue that the time preference rate is endogenous and is associated with income, education and other individual demographics. Accordingly, the determinants of time preference rates can be diverse. Additionally, under the consideration of lifelong aggregate utility, the application of one single year's income is less representative than an index that can represent the income trend over several years. The result of this study is reasonable based on the assumption that most of individuals will have a stable income variation during the lifetime; consequently, the income of one particular year is able to partly imply their expectation. Nevertheless, future studies focusing on a larger range of years of income would be beneficial in obtaining more precise outcomes. Finally, further discussion of the interactive function of time preference and "time constraint of income", which implies an opportunity cost of time in earning for an income, would provide insights into the time decision model.

## Chapter Five: The Relationship between Household Income and Physical Activity Participation: A Cragg's Hurdle Model on Kentucky Exercise & Health Survey

This chapter presents the second empirical study that investigates physical activity participation among Kentucky residents based on an independently conducted Kentucky Exercise & Health Survey (KEHS). It aims at incorporating possible physical activity determinants that are not included in previous national surveys. A Cragg's two-step hurdle model is estimated with two dependent variables, which are the likelihood of the decision to participate in physical activity and the average time spent on physical activity on a typical day, respectively. The empirical findings suggest that income indirectly influences the choice of physical activity participation through paying for physical activity, the presence of workplace exercise facilities, educational attainment, and having physically active friends or family members.

### 5.1 Data

The data for this study comes from an independent survey conducted November 2013 – September 2014. The mail survey was first distributed in November 2013 among 1,000 Kentucky residents, weighted by population across all counties on residents 18 and older. A second round was supplemented in September 2014 among 500 Kentucky residents, again weighted by population across all counties, but with the age limited between 18 and 45 to remedy the underrepresented responses of this age group from the first round survey. For both rounds, the same survey questionnaire was utilized (see the Appendix) and followed the Dillman method with postcard reminders. A total of 200 valid responses were received from both rounds, yielding a response rate of 13.33%.

This survey is an independent regional health and exercise survey for Kentucky residents. It investigates possible exercise determinants not included in previous national surveys such as the monetary cost of physical activity, other health related choices (such as cigarette and alcohol consumption), other potential physical activity determinants (such as work time physical activity participation and commuting physical activity and time), and household profile. Using the data collected, a two-step Cragg's hurdle model is estimated; the model is described in the next section. The dependent variable for the first step represents whether the respondent participates in physical activity or not. The original response is the frequency of leisure time physical activity participation identified by six categories: none, less than once a month, 1-2 times a month, 1-2 times a week, 3-4 times a week, and almost daily. From this, a new binary variable is created as the dependent variable to assess the participation choice, where the answer "none" is treated as not having participated in physical activity, and the latter five answers are treated as having participated in physical activity.

In the second step of Cragg's model, the dependent variable is the natural logarithm of the average time spent on physical activity in a typical day. The survey investigates the average time spent on sports or exercise in a typical day. This model takes the natural logarithm of the average time so that the dependent variable more closely follows a normal distribution. The model has 200 observations in the first step, with 77.5% of them reporting as having participated in physical activity. In the second step, there are 155 observations, since respondents who do not participate in physical activity are truncated in this step.

This analysis includes a broad set of independent variables. First, annual household income is included as a potential factor in determining physical activity

behavior. The survey measures household income using eight intervals. In the empirical analysis, these intervals are categorized into three income variables; the dummy variable indicates whether an individual's income is contained in this interval or not. Other demographic variables are also considered as explanatory variables, including age, gender, educational attainment and marital status. Similar to income, the model has three age variables, two education variables and two marital status variables. Other than income and demographic variables, this study includes other factors that may influence physical activity participation behavior but have not been investigated in previous studies. These new variables measure whether an individual pays for physical activity or not, has a workout facility at his workplace or not, considers his residence neighborhood as safe to work out or not, and whether he has close friends or family members who participate in physical activity or not. Table 5.1 provides a full definition of the categorical variables.

Descriptive statistics for the full sample are presented in Table 5.2. The mean of physical activity participation choice is 0.775, which suggests approximately 77.5% of the sample do participate in physical activity in the time under investigation; the average exercise time on a typical day for them is 36 minutes. Only 33.5% of the observations have paid for physical activity, such as gym membership fees, league fees, personal training fees and equipment. This supports the assumption that physical activity can be non-costly from a monetary perspective. About 86.4% of the sample considers their neighborhood as safe to walk or jog, which supports the idea that free physical activity participation is readily available to a large percentage of the population; 61.7% of the respondents are married, while 38.3% are never married, divorced, separated or widowed.

There are a few caveats to remember. First, there are approximately 24% more female respondents than male respondents in our sample. This is normal in survey data since females are more likely to respond to mail surveys than males (Binkley, 2011); however, the sample is not fully randomly representative of the entire Kentucky population. The Kentucky Census (United States Census Bureau, 2016) shows that the percentage of female persons is 50.8% as of July 1, 2015. Second, 42.2% of the observations are reported to be older than 65, which is overrepresented relative to the 15.2% of the whole population according to Kentucky Census Data. Being highly educated is also overrepresented in our sample. 56.8% of the respondents holding a graduate degree, while in the population only 21.8% of persons age 25 years and above have a Bachelor's degree or higher. More highly educated individuals may be more passionate for research projects and the older population may have more time to respond to surveys, which make both groups more likely to return the survey; however, the non-representative sample may create biased regression results.

Table 5.3 provides the independent variable means for the dependent variable conditional on the response. Among the lower income groups with annual household income up to $39,999, the percentage of respondents who participate in physical activity is lower than the percentage of observations that do not participate in physical activity. Similarly, for the higher income groups with annual household income more than $40,000, the percentage of individuals who participate in physical activity is greater. The same situation occurs for educational attainment; the percentage of respondents who participate in physical activity is greater in more highly educated categories. A major portion of people who do not participate in physical activity are those age 65 and above, while nearly half of the people who participate in physical

activity are whose age between 18 and 45. The right column of the table can be interpreted as the sample mean for the second step of our model. For example, the percentage of individuals paying for physical activity in the second step of estimation is 40.4%, higher than the 33.5% in the first step; the percentage of respondents who have work out facilities at their workplace is 22.5% for the truncated sample, which is slight higher than the percentage of the whole sample as 19.5%.

**5.2 Empirical Methodology**

According to the standard neoclassical theory of consumer utility maximization, it is assumed that an individual has preferences over his health and physical activity participation, which affects his health condition and the length of life. To determine the optimal participation frequency and the time spent on physical activity, the individual maximizes utility taking a set of factors into account, including income level, cost of physical activity, accessibility to physical activity facilities, demographic profile and so forth. This study is particularly interested in whether income level is significantly related to physical activity participation. Based on the theoretical discussion before, a reasonable expectation is that high income would contribute to physical activity participation, whether through an increasing likelihood of participation in physical activity or a greater amount of time spent on physical activity in a typical day, or both.

This study investigates this relation empirically by employing a Cragg's two-step hurdle model. The first step is usually called a participation decision: it measures whether an individual participates in an activity or consumes a good. The second step is a consumption decision investigating the amount of consumption conditional on the participation decision made in the first step. Unlike the Heckman selection model assuming that the error terms of the participation and consumption equations are

correlated and the participation decision dominates the consumption decision, the Cragg model assumes that the participation decision and consumption decision are independent. In addition, in the Cragg's two-step model, the coefficients in the two steps can be different ($\gamma$ and $\beta$); it can also have different sets of variables ($x$ and $z$) in the first and second step of the model. However, in this empirical analysis, the set of explanatory variables are the same for both steps, since generally it is difficult to rationalize why one variable should affect participation but not consumption or vice-versa.

The two-step Cragg's hurdle model is a sequential procedure. In the empirical analysis, the first step attempts to associate one's socioeconomic characteristics with the decision whether or not to participate in physical activity. This relationship is obtained through a Probit model as equation (5.1).

$$Prob(y^* > 0) = \Phi(x'\gamma) \qquad (5.1)$$

In equation (5.1), $y$ is the natural logarithm of average time spent on physical activity in a typical day, and $y^*$ implies an underlying variable that is not observed if the average time is not greater than zero. The term $\Phi$ is the cumulative distribution function of the standard normal distribution, $x$ is a matrix of independent variables which includes annual household income, age, gender, educational attainment, marital status, the price paid for physical activity, the physical activity level of close friends and family members, and the accessibility of physical activity facilities in neighborhood and at workplace, and $\gamma$ is a matrix of unknown parameters.

The second step is a truncated regression model; individuals who have average physical activity time equal or smaller than zero are dropped from the sample. Again, as shown in equation (5.2), the dependent variable $y$ is the positive level of average time spent on physical activity in a typical day only observed if $y^* > 0$; $x$ is a matrix

of independent variables that are the same as in the first step and $\beta$ is a matrix of unknown coefficients. The term $\sigma$ is the standard deviation of the random error of $y^*$; $\lambda$ is the inverse Mills ratio function in the format of equation (5.3), where $\Phi$ is the cumulative distribution function of the standard normal distribution and $\phi$ is the probability density function of the standard normal distribution.

$$E(y|y^* > 0) = x'\beta + \sigma\lambda\left(\frac{x'\beta}{\sigma}\right) \qquad (5.2)$$

$$\lambda\left(\frac{x'\beta}{\sigma}\right) = \frac{\phi(\frac{x'\beta}{\sigma})}{\Phi(\frac{x'\beta}{\sigma})} \qquad (5.3)$$

**5.3 Empirical Results and Discussion**

Table 5.4 reports the coefficient estimates for participation in physical activity and the average time spent on physical activity for the Kentucky Health and Exercise Survey data. The younger age group (18-44) appears to be more likely to participate in physical activity than the middle age group (45-64); this finding is consistent with results using NHANES data that age is negatively related to the likelihood of physical activity participation. Educational attainment is positively correlated with the choice of physical activity participation; people who have only accomplished college and below are less likely to participate in physical activity than those with graduate and above degrees. The positive effect of education is identical over the three studies, and generally agrees with all previous studies regarding the choice toward health beneficial consumption.

Having a workout facility at workplace has a positive effect on the choice to take part in physical activity; moreover, paying for physical activity, such as gym membership fees or sports equipment, is strongly associated with the probability of physical activity participation. These two factors have not been investigated in previous studies, and the implication of empirical results of these two unique

variables in this model is straightforward intuitively. The workout facility in the workplace provides more flexibility to take part in physical activity. Additionally, financially investing in physical activity suggests a strong willingness or intention to participate in physical activity, and may also serve as a commitment mechanism. Therefore, both of these factors contribute to an increased probability of physical activity participation.

Paying for physical activity also appears to influence the average time spent on physical activity; those who pay for physical activity spend more time exercising. This is because paying for physical activity can act as a commitment mechanism, just like peer pressure. Another variable correlated with the average exercising time is whether an individual has close friends or family members who participate in physical activity; this variable is a new addition compared to other studies. People who have physically active friends or family tend to spend more time on physical activity, even though it is unknown if they exercise together. This suggests that health behavior is influenced by peer pressure.

In the final model, there is no statistically significant relationship between income and the dependent variables. In a preliminary simple model considering solely the effect of income on physical activity choice, results suggest that income and physical activity participation are positively related. However, in the multivariate model, controlling for other explanatory variables eliminates the effect of income. Further analyses are performed to better understand how income and those explanatory variables are related. To do this, multiple simple logistic regressions are performed, using income as independent variable, and varying the dependent variables among paying for physical activity, workout facility at the workplace, age,

education, and friends and family play sports and exercise, which are all significantly related to physical activity participation choice.

Table 5.5 displays the logistic regressions results. Those in the low income category (annual household income 0 to $39,999) are less likely to pay for physical activity compared with middle income category (annual household income $40,000 to $79,999), while the middle income category and high income category (annual household income $80,000 and above) are not statistically different. Also, those in the low income category are less likely to have physical activity facilities at their work place; middle income category and high income category do not statistically differ. Income does not appear to be significantly correlated with age in this model. As expected, those in the low income category are more likely to be less educated (college and below); conversely, those in the high income category are more likely to be more educated. Finally, those in the high income category are more likely to have close friends or family members participate in physical activity than middle income category, while this relationship is not statistically significant between those in the middle income category and those in the low income category.

Combining the results from Table 5.4 and Table 5.5, the empirical results suspect that income indirectly influences the choice of physical activity participation through paying for physical activity, the presence of workplace exercise facilities, educational attainment, and having physical active friends or family members.

## Chapter Six: The Relationship between Household Income and Physical Activity Participation: Panel Data Analysis on China Health and Nutrition Survey

This chapter presents the last empirical study of the dissertation, which investigates the determinants of physical activity participation in China with a focus on the impact of income and possible implications regarding time preference. Longitudinal data from China Health and Nutrition Survey (CHNS) are used employing a random effects method. The results suggest that both the probability of physical activity participation and the time spent on physical activity are positively associated with income in China.

### 6.1 Data

The data for this study comes from the China Health and Nutrition Survey (CHNS). CHNS is an international collaborative project led by the Carolina Population Center at the University of North Carolina at Chapel Hill investigating nutrition and health behaviors in China. It is a longitudinal study first launched in 1989 and was repeated in 1991, 1993, 1997, 2000, 2004, 2006, 2009 and 2011. This project consists of different surveys such as a household survey, child survey, adult survey and community survey; the questions regarding physical activity was first introduced in the Household Survey in 1997, became part of the Physical Examination in 2000, and was included in adult surveys in 2004 and the following years. This study utilizes the data from 2004 to 2011, which is obtained from the longitudinal dataset released in 2015. The panel is unbalanced because not every individual is observed in every year; the minimum number of observation times for an individual is 1 and the maximum is 4.

A random effects Logit model and two standard random effects models are estimated; the models will be described in the next section. The first model is a

random effects Logit model with the dependent variable representing whether the respondent participated in some level of leisure time physical activity or not in a typical day. The original survey includes an individual activity sheet recording all physical activities taken in leisure time for each respondent. Based on this, a new binary variable is created as the dependent variable to assess the participation choice; the new variable takes the value 1 if the response to either physical activity in the activity sheet is positive. In other words, taking part in any form of physical activity is recognized as leisure time physical activity participation.

In the second step, two random effects models are estimated; the dependent variables are the logarithm of time spent on physical activity in a typical day during the week (Monday through Friday) in the first model and on weekends in the second model. The CHNS investigates the time spent on physical activity on weekdays and weekends separately, which enables it to display the difference of individual physical activity behavior during the week and on weekends. Again, according to the detailed information of leisure time physical activity from the individual activity table, by summing up the time that an individual spent on each physical activity, the total time spent on physical activity during a typical day is obtained; furthermore, this study takes the natural logarithm of the time to have the dependent variable more closely follow a normal distribution. The random effects Logit model has 17,765 observations with 12.0% of the observations reporting as having participated in physical activity. The second step of the model has 2,002 observations for time spent on physical activity in a typical day during the week and 2,006 for that on weekends, since respondents who did not participate in physical activity are truncated.

The analysis includes a broad set of independent variables. First, annual household income is included as a potential factor in determining physical activity

behavior. Unlike the NHANES survey, the CHNS data measures the actual annual household income rather than categorizing income. The study takes the natural logarithm of the income to have the variable more closely follow a normal distribution. Other demographic variables are also considered as explanatory variables, including age, gender, educational attainment, marital status, employment status, and whether the respondent resides in urban or rural area. Other than income and demographic variables, consumption towards cigarette, alcohol and sugary drinks are also considered in the empirical analysis. These consumption behaviors are health related and assumed to influence individual risk behaviors and consequently are included in this physical activity analysis.

Table 6.1 describes and defines the categorical variables used in the study. Since approximately 20% of the respondents have never been in school, illiteracy is set as one category of education attainment. Compulsory education means having completed 9 years of China's compulsory education of primary school and junior middle school. The other two education variables are completing high school, and completing college or graduate school. Any amount of cigarette consumption has a negative effect on health and is considered as a health-related consumption, while only a significant amount of alcohol and sugar intake becomes detrimental to health and therefore a consumption of alcoholic beverage, soft drinks or sugary drinks for more than twice a week is included in the empirical model.

A comprehensive set of summary statistics is provided in Table 6.2, which includes overall, between and within summary statistics. The overall summary statistics are the means, standard deviations, minimums and maximums of the pooled data. Between summary statistics are based on variation between individuals, whereas

within summary statistics are based on individual variation over time from own averages.

The overall mean of physical activity participation choice is 0.12, which means only 12% of the sample chose to participate in physical activity. The overall mean of time spent on physical activity in a typical day during the week is 75 minutes, slightly lower than the 80 minutes devoted to leisure time physical activity on weekends. About half of the observations are males, and about half of the observations have completed 9 years of compulsory education. Nearly one third of the observations reside in urban areas, and nearly one third of the observations are employed for wages. Finally, 85.9% of the respondents are married.

**6.2 Empirical Methodology**

According to the standard neoclassical theory of consumer utility maximization, assume that an individual has preferences over his health and physical activity participation, which affects the health condition and the length of life. To determine the optimal participation frequency and the time spent on physical activity, the individual maximizes utility taking a set of factors into account, including the income level, the demographic profile, the risk preference and so forth. This study is particularly interested in whether income level significantly relates to physical activity participation. The theoretical model presented in Chapter 3 predicts that high income would contribute to physical activity participation, whether through an increasing likelihood of participation in physical activity or a greater amount of time spent on physical activity in a typical day, or both.

This relationship is investigated empirically by utilizing a panel data model. The advantage of using panel data is the ability to account for changes across time and individuals while controlling for unobserved individual effects. The first model is

a random effects Logit model with the binary response measuring whether the individual participated in physical activity or not. The second model and third model are random effects (linear) models investigating the time spent on physical activity in a typical day during the week and on weekends. To start with, the standard random effects model takes the specification given in equation (6.1).

$$y_{it} = x_{it}'\beta + \varepsilon_{it} \qquad (6.1)$$

In equation (6.1), $y_{it}$ is a continuous dependent variable for individual $i$ over time $t$, $x_{it}$ is a matrix of independent variables, and $\beta$ is a matrix of unknown coefficients. The error term $\varepsilon_{it}$ can be decomposed as shown in equation (6.2),

$$\varepsilon_{it} = u_i + e_{it} \qquad (6.2)$$

where $u_i$ are the individual-specific effects which are distributed independently of the regressors and $e_{it}$ is the composite error term.

Note that $Var(\varepsilon_{it}) = \sigma_u^2 + \sigma_e^2$ and $cov(\varepsilon_{it}, \varepsilon_{is}) = \sigma_u^2$, so the intraclass correlation of the error is $\rho_\varepsilon = cor(\varepsilon_{it}, \varepsilon_{is}) = \sigma_u^2/(\sigma_u^2 + \sigma_e^2)$, which is the fraction of the variance in the error due to the individual-specific effects. $\rho_\varepsilon$ will approach 1 if the error changes both over time and across individuals, which is called idiosyncratic error. STATA reports the estimates of $\sigma_u, \sigma_e$ and $\rho_\varepsilon$, which will be presented in next section.

The second model and third model take the form of the random effects regression as described above. In these models, the dependent variable $y_{it}$ is the time spent on physical activity in a typical day during the week and on weekends, and $x_{it}$ is a matrix of independent variables which includes annual household income, age, gender, educational attainment, marital status, employment status, residence area and the consumption level of cigarette, alcohol beverage and sugary drinks.

A random effects Logit model used for this first model, as given as equation (6.3).

$$Prob(y_{it} \neq 0|x_{it}) = Prob(x'_{it}\beta + u_i + e_{it} > 0|x_{it}) = \Phi(x'_{it}\beta + u_i) \qquad (6.3)$$

Here, $y_{it}$ is the dependent variable taking the binary response whether the individual participated in physical activity or not, and $x_{it}$ is a matrix of independent variables used in models 2 and 3. The term Φ is the cumulative distribution function of logistic regression. The $u_i$ are independent and identically distributed following $N(0, \sigma_u^2)$, and the value of $\sigma_u$ will be estimated. Finally, $e_{it}$ are independent and identically distributed following a logistic distribution with mean zero and variance $\sigma_e^2$, where $\sigma_e^2$ is constant and equals to $\frac{\pi^2}{3}$, which is independent of $u_i$.

**6.3 Empirical Results and Discussion**

Table 6.3 presents the results for physical activity choice (column 1), and the time spent on physical activity during the week (column 2) and on weekends (column 3).

The results from model 1, the random effects Logit model on physical activity choice, indicate that annual household income is positively correlated with physical activity participation choice. Married people, and those who had been married but do not have a partner right now, are both less likely to participate in physical activity than people who have never married. Older people are more likely to exercise. Males are more likely to participate in physical activity. Education shows a positive relationship with physical activity participation. People are more likely to participate in physical activity if they live in an urban area. People who are employed for a wage are more likely to take part in physical activity than the self-employed.

The decision to make other unhealthy choices may also be related to decisions regarding physical activity. Cigarette consumption is negatively associated with

physical activity participation, while alcohol consumption and consumption of sugary drinks are both positively related to the probability of physical activity participation. In a developing country such as China, the health detriment of cigarettes has become better understood in recent years; however, the detrimental effects of alcohol and sugary drinks, have not been as widely recognized as cigarettes. In addition, the fact that the CHNS is conducted mainly in rural areas and surrounding suburbs, where the consumption of alcohol and sugary drinks is more a symbol of wealth and therefore less likely among low income groups, this finding is reasonable. In other words, consumption of alcohol and sugary drinks might not be good indicators of time preference in less developed countries.

The determinants on the time spent on physical activity are quite different from the participation model itself. The first model analyzes the time spent on physical activity during the week (Monday to Friday). Household income plays a positive role, while people who are employed for a wage tend to spend less time on physical activity. People who attain higher education with college degree and above spend more time on physical activity than individuals only have completed compulsory education, but no significant difference shows between those who are illiterate, who have completed compulsory education and who have completed high school. Income, employment status and education attainment, these three factors also function the same in determining the time spent on physical activity on the weekend. An additional determinant in the model analyzing weekend exercise time is that urban people are likely to spend more time on physical activity on the weekend, but there is no link with time spent on physical activity during the week. Age or marital status does not exhibit a relationship with average exercise time per day.

To summarize, the probability of participating in physical activity, as well as the time spent on physical activity both during the week and on weekends, increases with income. As discussed previously, low income individuals discount expected future utility and hence diminish the cost of reduced longevity. Therefore, the consumption of beneficial goods for a low income person tends to be smaller. The effect of education functions similar to income. Education is positively related to both the probability of physical activity participation and the time spent on it. This can be easily interpreted by the same mechanism that education raises the possibility of future utility because education is an investment in human capital (Becker et al, 1977). However, employment status affects physical activity in a different way. Those employed for wages have a higher probability of choosing physical activity, but spend less time on it. In China, being employed for a wage implies a relatively higher and more stable income source, which further implies a greater weight on future utility and therefore induces a positive consumption towards beneficial goods. Nevertheless, being employed for a wage might also limit the time people have for leisure activities such as physical activity, and thus lead to a smaller amount of time spent on it.

The determinants for time spent on physical activity during the week and on weekends are almost identical except one factor, residing in an urban area. This variable is positively related to weekend exercising time. One possible explanation is that urban residents might have more leisure time for activities including physical activity during the weekend, because urban people usually work during the week while rural residents, most of whom are self-employed, have more flexibility in time allocation. In addition, those working for a wage are probably more likely to have set hours to work (often during the week), leaving weekends more available for exercise; while many individuals who are self-employed work longer hours and work on the

weekends, leaving them less time to exercise on the weekends. It is also interesting to note that many variables that affect the probability of participating in physical activity have no significant relationship with the time spent on it. This may suggest that the choice to be active represents the attitude towards time preference; however, the amount of consumption on health goods might not be the best descriptor of time preference.

## Chapter Seven: Conclusion and Future Research

This dissertation investigates the determinants of individual consumption choice on a health beneficial and monetarily non-costly good, leisure time physical activity, in developed and developing economics. By utilizing the concept of time preference, this study focuses on leisure time physical activity choice of U.S. population from the National Health and Nutrition Examination Survey (NHANES) data, China population based on the China Health and Nutrition Survey (CHNS) data, and Kentucky residents with an independently-conducted Kentucky Exercise & Health Survey. Analyses of data from these surveys show that income plays a positive role in determining the probability to participate in physical activity in both countries. However, while the time spent on physical activity increases with income in China, the relationship between average time spent on physical activity and income is ambiguous in United States.

### 7.1 Summary

In recent years, health choices as they relate to physical activity have attracted a lot of attention. Physical activity participation is one example of a health-related decision that consumers make. Most physical activities can provide health benefits and are monetary non-costly, while insufficient physical activity can potentially do harm to the human body. This dissertation is interested in the relationship between income and physical activity participation, and what that implies about time preference. It investigates how the household income level may be related to the probability to participate in physical activity and the time spent on it for individuals.

Based on the medical finding that physical activity will benefit health, the assumption is that when making physical activity participation decisions, a rational consumer has to balance his choice between the increased lifetime aggregate utility

from impending health benefits and the current disutility caused by the opportunity cost of physical activity participation. Individuals with stronger time preferences discount the future more heavily and place more weight on current utility; those with weaker time preferences value future utility more and are willing to sacrifice present utility to maintain better health in the future. The cost of utility loss increases as income increases since utility depends, in part, on income. Hence, a higher income should be correlated with a greater likelihood of participating in physical activity, while a lower income may be correlated with insufficient physical activity participation.

This dissertation is comprised of three empirical studies on physical activity participation behavior. The first one estimates a two-step Heckman selection model employing data from the 2001-06 National Health and Nutrition Examination Surveys (NHANES). The empirical results are mostly consistent with theoretical findings. The probability of participating in physical activity increases with income. This is a strong signal that the low income group discounts expected future utility and, thus, diminishes the cost of reduced longevity. Educational attainment affects physical activity in a similar way to income. Higher education increases the probability of choosing physical activity, but decreases the average time spent on it. Males and individuals who do not have a spouse are more likely to take part in physical activity and spend more time on it. An increasing household size leads to a lower probability of physical activity participation. Older people are also less likely to participate in physical activity; however, they spend more time on it when they participate. This study also introduces self-reported physical and mental health condition as explanatory variables. Better self-reported physical health condition and better mental health condition both contribute to a higher probability of physical activity

participation; in addition, better self-reported mental health condition increases the average time spent on physical activity as well.

The second study comes from the independently conducted Kentucky Exercise & Health Survey (See Appendix). An estimation of a two-step Cragg's hurdle model finds that income has an indirect positive influence on the choice of physical activity participation through paying for physical activity, the presence of workplace exercise facilities, educational attainment, and having physically active friends or family. Specifically, low income people are less likely to pay for physical activity and are less likely to have accessible physical activity facilities at their workplace, while having access to a workout facility in the workplace positively correlates with the probability of participation, and paying for physical activity positively correlates with both the probability of participation and the time spent exercising. Income also positively relates to educational attainment, which is positively associated with the probability of participating in physical activity. Finally, high income individuals are more likely to have close friends or family members who participate in physical activity; this new added variable is demonstrated a positive relationship with the average time spent on physical activity.

The third study utilizes longitudinal data from the China Health and Nutrition Survey (CHNS). Results from random effects models suggest that both the probability of participating in physical activity and the time spent on physical activity increase with income. Individuals who are employed for wages are more likely to participate in physical activity, but spend less time on it compared with those of other employment status. Residing in an urban area makes one more likely to participate, but regarding the time spent on physical activity, it only has a positive effect on the time spent exercising on the weekend. A novelty of the third empirical analysis out of

the three studies is that it incorporates other health behaviors as explanatory variables. Cigarette consumption is found to be negatively associated with physical activity participation, while alcohol consumption and consumption of sugary drinks are both positively related to the probability of physical activity participation. They do not relate to the time spent on physical activity.

The estimated relationship between household income and consumption of physical activity is consistent across the three empirical analyses and supports the predictions of the theoretical model. Household income level plays an important role in determining physical activity participation behavior because it affects not only the consumer's current budget but also his or her future expected budget. The utility of the latter one becomes dominant as income increases. This provides useful information for the purpose of physical activity promotion, which can improve population health. To begin with, economic policy design can aim at providing greater income opportunities for the low income population. The prediction suggests that high income increases the value of one's future and, thus, more incentives to choose a healthier lifestyle today. Education has a similar influence so investment in education for the poor may also increase those incentives.

**7.2 Comparison Across United States and China**

According to the United Nations Statistics Division, the designations "developed" and "developing" are only intended for statistical convenience; there is no established convention for the designation of "developed" and "developing" countries or areas in the United Nations system. However, in common practice, United States is considered "developed" and China is considered "developing" (United Nations Statistics Division, 2016). The World Bank classifies countries into four income groups. High income countries having Gross National Income (GNI) per

capita above US$12,476 are classified as developed, while developing countries include low income countries that had GNI per capita of US$1,025 or less, lower middle income countries that had GNI per capita between US$1,026 and US$4,035, and upper middle income countries that had GNI per capita between US$4,036 and US$12,475 (World Bank, 2016). The International Monetary Fund (IMF) uses a flexible classification system that considers per capita income level, export diversification, and degree of integration into the global financial system (IMF, 2016). According to all three organizations, China is classified as a developing country and United States is classified as a developed country.

Table 7.1 displays the comparison of means of dependent variables between the U.S model from Chapter 4 and the China model from Chapter 6. In the NHANES sample, the overall mean of physical activity participation choice is 0.649, representing a 64.9% of the respondents that participated in physical activity. At the same time, it has a mean of approximate 30 minutes for the average exercise time per day. The overall mean of physical activity participation choice of CHNS sample is 0.12, which means only 12% of the sample chose to participate in physical activity. The overall mean of time spent on physical activity in a typical day during the week is 75 minutes, slightly lower than the 80 minutes devoted to leisure time physical activity on weekends. Obviously, a much greater percentage of the U.S. sample participates in physical activity than in the China sample, which generally coincides with the income assumption from the theoretical prediction in Chapter 2. United States has a higher GNI per capita of $54,960 in 2015, while the GNI per capita of China is $7,820 (World Bank, 2016). Besides, the data from NHANES and CHNS might not be equivalently representative of the population. While NHANES examines a nationally representative sample located in counties across the country, CHNS takes

place in only 15 provinces and municipal cities[3] in mostly middle China as well as a few east coast areas. Since economics development is unbalanced across east, middle and west China, this may generate the divergence in response.

Meanwhile, the average time spent on physical activity for those who participate is more than doubled in China than in the United States. The determinant of this result is worthy of investigation.

Table 7.2 reveals the comparison between United States and China on the determinants that are included in both models on physical activity participation. The variables utilized in both empirical studies include income, education, gender, age and marital status. Income plays a positive role in determining the probability of participating in physical activity both in the United States and in China, which strongly supports the theoretical assumption that low income individuals have a stronger time preference and discount future utility more heavily, and are less likely to make consumption choices beneficial to one's health. Education also increases the probability of physical activity participation in both countries, because it is an investment in human capital so that enhances future utility (Becker et al, 1977). To be clear, the income variable used in the China study is a continuous variable representing the logarithm of household annual income, whereas the U.S. study employs a categorical income variable, and individuals with annual household income less than $75,000 are less likely to participate in physical activity than those with income higher than $75,000. Education variables are categorical variables in both studies, but differ in the way they are categorized. Despite the differences in the specific design of the explanatory variables, the income and education variables show

[3] The provinces and municipal cities are Beijing, Chongqing, Guangxi, Guizhou, Heilongjiang, Henan, Hubei, Hunan, Jiangsu, Liaoning, Shaanxi, Shandong, Shanghai, Yunnan, Zhejiang.

a similar pattern in the U.S. and in China in determining the probability of physical activity participation.

Males are more likely to take part in physical activity both in the U.S. and in China; this is a general consensus in most of the literature investigating physical activity participation across countries (Farrell and Shields, 2002; Humphreys and Ruseski, 2006, 2007; Meltzer and Jena, 2010). Additionally, the "never married" groups in both countries are more likely to participate in physical activity. The only determinant that differs between the two countries is age. Age is positively related to participation in physical activity in China, while it is negatively related in United States. This dichotomy is intriguing, taking into account the differences in culture, retirement policy, social welfare system and the way people capture and accumulate wealth. For example, China has the world's lowest retirement age: currently it is 60 for men, 55 for female civil servants and 50 for female workers (Ministry of Human Resources and Social Security of the People's Republic of China, 2014). Conversely, the full retirement age in United States is 66, while the earliest workers may retire and receive reduced retirement benefits is age 62 (Social Security Administration, 2016). This means while many Chinese are retired and are planning their future life and leisure activity, people of the same age in United States are still at work and might not have adequate time and effort to do so. In addition, neighborhood committees in China organize leisure activities for retirees regularly such as square dancing[4], which is an activity started and developed years ago and now is popular from everywhere in China, both urban and rural. This also supports the evidence from the Kentucky study that close friends and family members participating in physical activity will promote the probability of participation.

[4] https://en.wikipedia.org/wiki/Square_dancing_(China)

As explained before, identifying the determinants of time spent on physical activity is more ambiguous. For all those factors that investigated, education is the only significant factor in both countries, and the effects differ. Education plays a positive role in determining one's time spent on physical activity in China but has a negative effect in United States. Follow the human capital theory, higher education implies more weight on future utility? and consequently a stronger willingness to exercise. But the time spent on physical activity is affected both by the willingness to take part in physical activity and the time constraint. It is obvious that in this example, the former factor has a dominant role in China so that more highly educated people spend more time on physical activity. However, in the United States, as discussed in Chapter 4, the time constraint might be more important. More highly educated people may be involved in many more social or volunteer activities, which occupy a lot of time and leave less time available for leisure time physical activity.

Additionally, the time spent on physical activity can be seen as the amount of consumption of physical activity. Recall the law of diminishing marginal utility in economics: when increasing the consumption of one product while holding all other conditions constant, the marginal utility from each additional unit of that product declines. Thus, an increase on time spent on physical activity does not always imply a health benefit or utility gain. World Health Organization (WTO)[5] recommends at least 150 minutes of moderate intensity aerobic physical activity or at least 75 minutes of vigorous intensity aerobic physical activity throughout a week or an equivalent combination of moderate- and vigorous-intensity activity for adults aged 18–64. For additional health benefits, the time should be doubled. As a result, when people make health decision towards physical activity participation, it does not necessarily mean to

[5] http://www.who.int/dietphysicalactivity/factsheet_adults/en/

maximize the time participated to obtain maximum health benefit. Consequently, the amount of consumption of health goods might not be the best descriptor of time preference, although the choice to be active represents the attitude towards time preference.

### 7.3 Limitations and Extensions

The results of the empirical estimation provide strong support for the theoretical prediction that low income people discount expected future utility and are less likely to make consumption decision towards health. Nonetheless, limitations leave room for future research. First, this analysis assumes that the rates of time preference are affected only by income, but this assumption is likely too simplistic. Becker and Mulligan (1997) have argued that time preference rate is endogenous and is associated with income, education, and other personal information. Accordingly, the determinants of time preference rates can be diverse. A future study identifying the determinants of the rate of time preference could help address this limitation. Second, the theoretical model only includes an income budget constraint; the results from this analysis suggest that also incorporating a time constraint may prove beneficial. Cawley (2004) presents a SLOTH model that provides an economic framework to explain physical activity participation and emphasizes the time budget. His study assumes that utility depends on a person's weight, health, food, other goods and time spent sleeping (S), at leisure (L), at occupation (O), in transportation (T), and in home production (H), and physical activity and sports are considered leisure activities. Thus, further discussion of the interactive function of time preference and the role that time serves as a budget to earn for an income would provide insights into the time decision model.

Additionally, under the consideration of lifelong aggregate utility, the use of one single year's income is less representative than an index that can represent the income trend over several years. The results of this dissertation are most applicable in situations in which individuals will have a relatively stable income during their lifetime; consequently, the income of one particular year is sufficient on which to base their future expectation. Nevertheless, future studies aimed at accounting income information spanning many years would be beneficial in addressing this limitation.

| Table 4.1 The Relationship of Vigorous Activity, Moderate Activity and Physical Activity Choice on NHANES Data | | |
|---|---|---|
| Vigorous Activity | Moderate Activity | Physical Activity Choice |
| Yes | Yes | Yes |
| Yes | No | Yes |
| No | Yes | Yes |
| No | No | No |

| Table 4.2 Definitions of Categorical Variables on NHANES Data | | |
|---|---|---|
| Variable | Value | Definition |
| Participation Choice | 1 | Participated in physical activity during a typical day. |
| Male | 1 | Gender is male; otherwise female. |
| Mexican | 1 | Race is Mexican. |
| Hispanic | 1 | Race is Hispanic. |
| White | 1 | Race is white. |
| Black | 1 | Race is Black. |
| Other Race | 1 | Race is any race other than Mexican, Hispanic, white or black. |
| High School and Below | 1 | Education level is high school and below. |
| Some College or Associate Degree | 1 | Education level is some college or associate degree. |
| College Graduate or Above | 1 | Education level is college graduate and above. |
| Married and Cohabited | 1 | Marital status is never married or cohabited. |
| Widowed/Divorced/Separated | 1 | Marital status is widowed or divorced or separated. |
| Single | 1 | Marital status is never married. |
| Reference Person | 1 | Being the person that has highest income in the household. |
| Income $0 to $9,999 | 1 | Household annual income is between 0 and $9,999. |
| $10,000 to $14,999 | 1 | Household annual income is between $10,000 and $14,999. |
| $15,000 to $19,999 | 1 | Household annual income is between $15,000 and $19,999. |
| $20,000 to $24,999 | 1 | Household annual income is between $20,000 and $24,999. |
| $25,000 to $34,999 | 1 | Household annual income is between $25,000 and $34,999. |
| $35,000 to $54,999 | 1 | Household annual income is between $35,000 and $54,999. |
| $55,000 to $74,999 | 1 | Household annual income is between $55,000 and $74,999. |
| $75,000 and above | 1 | Household annual income is $75,000 and above. |

| Table 4.3 Sample Descriptive Statistics on NHANES Data | | | | | |
|---|---|---|---|---|---|
| Variable | Obs | Mean | SD | Min. | Max. |
| Participation Choice | 11249 | 0.649 | 0.477 | 0 | 1 |
| Average Exercise Time per Day | 7298 | 30.418 | 58.440 | 0 | 932 |
| Log Average Time | 7298 | 3.137 | 1.286 | 0 | 6.837 |
| Male | 11249 | 0.488 | 0.500 | 0 | 1 |
| Age | 11249 | 45.477 | 20.044 | 18 | 85 |
| Mexican | 11249 | 0.209 | 0.406 | 0 | 1 |
| Hispanic | 11249 | 0.033 | 0.180 | 0 | 1 |
| White | 11249 | 0.509 | 0.500 | 0 | 1 |
| Black | 11249 | 0.210 | 0.407 | 0 | 1 |
| Other Race | 11249 | 0.038 | 0.192 | 0 | 1 |
| High School and Below | 11249 | 0.565 | 0.496 | 0 | 1 |
| Some College or Associate Degree | 11249 | 0.247 | 0.431 | 0 | 1 |
| College Graduate or Above | 11249 | 0.188 | 0.390 | 0 | 1 |
| Married and Cohabited | 11249 | 0.580 | 0.494 | 0 | 1 |
| Widowed/Divorced/Separated | 11249 | 0.188 | 0.391 | 0 | 1 |
| Single | 11249 | 0.232 | 0.422 | 0 | 1 |
| Reference Person | 11249 | 0.509 | 0.500 | 0 | 1 |
| Physical Unhealthy Days | 11249 | 3.343 | 7.348 | 0 | 30 |
| Mental Unhealthy Days | 11249 | 3.454 | 7.381 | 0 | 30 |
| Household Size | 11249 | 3.143 | 1.644 | 1 | 7 |
| Income $0 to $9,999 | 11249 | 0.074 | 0.261 | 0 | 1 |
| $10,000 to $14,999 | 11249 | 0.081 | 0.273 | 0 | 1 |
| $15,000 to $19,999 | 11249 | 0.079 | 0.270 | 0 | 1 |
| $20,000 to $24,999 | 11249 | 0.085 | 0.279 | 0 | 1 |
| $25,000 to $34,999 | 11249 | 0.134 | 0.340 | 0 | 1 |
| $35,000 to $54,999 | 11249 | 0.202 | 0.401 | 0 | 1 |
| $55,000 to $74,999 | 11249 | 0.122 | 0.327 | 0 | 1 |
| $75,000 and above | 11249 | 0.224 | 0.417 | 0 | 1 |

| Table 4.4 Explanatory Variable Means for Physical Activity Choice on NHANES Data | | |
|---|---|---|
| Participation Choice | 0 (N=3951) | 1 (N=7298) |
| Average Exercise Time per Day | NA | 30.418 |
| Log Average Time | NA | 3.137 |
| Male | 0.453 | 0.507 |
| Age | 50.438 | 42.791 |
| Mexican | 0.273 | 0.174 |
| Hispanic | 0.033 | 0.034 |
| White | 0.433 | 0.551 |
| Black | 0.228 | 0.200 |
| Other Race | 0.033 | 0.041 |
| High School and Below | 0.696 | 0.494 |
| Some College or Associate Degree | 0.211 | 0.267 |
| College Graduate or Above | 0.093 | 0.239 |
| Married and Cohabited | 0.602 | 0.568 |
| Widowed/Divorced/Separated | 0.245 | 0.157 |
| Single | 0.153 | 0.275 |
| Reference Person | 0.537 | 0.494 |
| Physical Unhealthy Days | 4.561 | 2.684 |
| Mental Unhealthy Days | 3.910 | 3.207 |
| Household Size | 3.194 | 3.116 |
| Income $0 to $9,999 | 0.094 | 0.062 |
| $10,000 to $14,999 | 0.107 | 0.068 |
| $15,000 to $19,999 | 0.100 | 0.067 |
| $20,000 to $24,999 | 0.101 | 0.076 |
| $25,000 to $34,999 | 0.156 | 0.122 |
| $35,000 to $54,999 | 0.202 | 0.202 |
| $55,000 to $74,999 | 0.106 | 0.131 |
| $75,000 and above | 0.134 | 0.273 |

| Table 4.5 Physical Activity Participation Choice and Average Time Spent among United States Population | | |
|---|---|---|
| Determinants | Choice (N=11,249) | Time (N=7,298) |
| Male | 0.145*** | 0.187*** |
| Age | -0.012*** | 0.009*** |
| Mexican | -0.302*** | 0.024 |
| Hispanic | -0.06 | 0.117 |
| Black | -0.247*** | 0.215*** |
| Other Race | -0.131* | 0.113 |
| Some College or Associate Degree | 0.271*** | -0.223*** |
| College Graduate or Above | 0.587*** | -0.336** |
| Widowed/Divorced/Separated | 0.028 | 0.101* |
| Single | 0.333*** | 0.155* |
| Reference Person | -0.015 | -0.065 |
| Physical Unhealthy Days | -0.013*** | 0.005 |
| Mental Unhealthy Days | -0.004** | -0.005* |
| Household Size | -0.064*** | 0.019 |
| Income $0 to $9,999 | -0.491*** | 0.222 |
| $10,000 to $14,999 | -0.394*** | 0.143 |
| $15,000 to $19,999 | -0.355*** | 0.189* |
| $20,000 to $24,999 | -0.289*** | 0.002 |
| $25,000 to $34,999 | -0.331*** | 0.209** |
| $35,000 to $54,999 | -0.241*** | 0.040 |
| $55,000 to $74,999 | -0.197*** | 3.289 |
| $\rho$ | | *-0.886* |
| $\sigma$ | | *1.604* |
| $\lambda$ | | *-1.421* |

*(***, **, * denote significance at the 1%, 5%, 10% level, respectively.)*

| Table 5.1 Definitions of Categorical Variables on KEHS Data | | |
|---|---|---|
| Variable | Value | Definition |
| Participation Choice | 1 | Participated in physical activity during a typical day. |
| Pay for Physical Activity | 1 | Have paid for participating in sports or other physical activities. |
| Safe Neighborhood | 1 | Feel neighborhood is safe to walk or jog. |
| Workplace Facility | 1 | Have exercise facilities or programs at workplace. |
| Male | 1 | Gender is male; otherwise female. |
| Income $0 to $39,999 | 1 | Household annual income is between 0 and $39,999. |
| $40,000 to $79,999 | 1 | Household annual income is between $40,000 and $79,999 |
| $80,000 and above | 1 | Household annual income is $80,000 and above. |
| Age 18-45 | 1 | Age is between 18 and 45. |
| 46-54 | 1 | Age is between 46-54. |
| 65 and above | 1 | Age is 65 and above. |
| Married | 1 | Marital status is married. |
| Non-married | 1 | Marital status is widowed/divorced/separated or never married. |
| College and Below | 1 | Education level is college graduate and below. |
| Graduate and Above | 1 | Education level is graduate school and above. |
| Physical Active Family & Friends | 1 | Have family members or close friends play sports or exercise. |

| Table 5.2 Sample Descriptive Statistics on KEHS Data | | | | | |
|---|---|---|---|---|---|
| Variable | Obs | Mean | SD | Min. | Max. |
| Participation Choice | 200 | 0.775 | 0.414 | 0 | 1 |
| Average Exercise Time per Day | 155 | 36.005 | 43.843 | 1.066 | 300 |
| Log Average Time | 155 | 3.033 | 1.169 | 0.064 | 5.707 |
| Pay for Physical Activity | 200 | 0.335 | 0.473 | 0 | 1 |
| Safe Neighborhood | 200 | 0.864 | 0.344 | 0 | 1 |
| Workplace Facility | 200 | 0.195 | 0.397 | 0 | 1 |
| Male | 200 | 0.380 | 0.487 | 0 | 1 |
| Income $0 to $39,999 | 200 | 0.358 | 0.481 | 0 | 1 |
| $40,000 to $79,999 | 200 | 0.338 | 0.474 | 0 | 1 |
| $80,000 and above | 200 | 0.304 | 0.461 | 0 | 1 |
| Age 18-45 | 200 | 0.398 | 0.491 | 0 | 1 |
| 46-54 | 200 | 0.175 | 0.381 | 0 | 1 |
| 65 and above | 200 | 0.422 | 0.495 | 0 | 1 |
| Married | 200 | 0.617 | 0.487 | 0 | 1 |
| Non-married | 200 | 0.383 | 0.487 | 0 | 1 |
| College and Below | 200 | 0.432 | 0.497 | 0 | 1 |
| Graduate and Above | 200 | 0.568 | 0.497 | 0 | 1 |
| Physical Active Family & Friends | 200 | 0.799 | 0.402 | 0 | 1 |

| Table 5.3 Explanatory Variable Means for Physical Activity Choice on KEHS Data | | |
|---|---|---|
| Participation Choice | 0 (N=45) | 1 (N=155) |
| Average Exercise Time per Day | 0 | 36.005 |
| Log Average Time | 0 | 3.033 |
| Pay for Physical Activity | 0.089 | 0.404 |
| Safe Neighborhood | 0.800 | 0.882 |
| Workplace Facility | 0.089 | 0.225 |
| Male | 0.378 | 0.381 |
| Income $0 to $39,999 | 0.489 | 0.321 |
| $40,000 to $79,999 | 0.267 | 0.358 |
| $80,000 and above | 0.244 | 0.321 |
| Age 18-45 | 0.178 | 0.460 |
| 46-54 | 0.222 | 0.161 |
| 65 and above | 0.600 | 0.373 |
| Married | 0.622 | 0.615 |
| Non-married | 0.378 | 0.385 |
| College and Below | 0.644 | 0.373 |
| Graduate and Above | 0.356 | 0.627 |
| Physical Active Family & Friends | 0.622 | 0.849 |

| Table 5.4 Physical Activity Participation Choice and Average Time Spent in Kentucky | | |
|---|---|---|
| Determinants | Choice (N=200) | Time (N=155) |
| Pay for Physical Activity | 0.955*** | 0.804*** |
| Safe Neighborhood | 0.323 | -0.104 |
| Workplace Facility | 0.644* | 0.057 |
| Male | -0.028 | 0.246 |
| Income $0 to $39,999 | -0.224 | -0.042 |
| $80,000 and above | -0.486 | 0.208 |
| Age 18-45 | 0.909*** | -0.338 |
| 65 and above | 0.120 | -0.343 |
| Married | 0.013 | -0.102 |
| College and Below | -0.591** | -0.135 |
| Physical Active Family & Friends | 0.444 | 0.480* |
| Pseudo $R^2$ | 0.221 | |
| $\sigma$ | | 1.006 |

*(***, **, * denote significance at the 1%, 5%, 10% level, respectively.)*

| Table 5.5 The Relationship of Variables with Income on KEHS Data (N=200) | | | | | |
|---|---|---|---|---|---|
| | Pay for Physical Activity | Workplace Facility | Age 18-45 | College and Below | Physical Active Family & Friends |
| Income $0 to $39,999 | -1.123*** | -0.916* | 0.052 | 0.974*** | -0.506 |
| Income $80,000 and above | 0.503 | 0.123 | 0.004 | -1.212*** | 1.717*** |
| *Pseudo $R^2$* | *0.071* | *0.030* | *0.000* | *0.079* | *0.087* |

*(***, **, * denote significance at the 1%, 5%, 10% level, respectively.)*

| Table 6.1 Definitions of Categorical Variables on CHNS Data | | |
|---|---|---|
| Variable | Value | Definition |
| Participation Choice | 1 | Participated in physical activity during a typical day. |
| Widowed/Divorced/Separated | 1 | Marital status is widowed or divorced or separated. |
| Single | 1 | Marital status is never married. |
| Married | 1 | Marital status is married. |
| Illiteracy | 1 | Have not been in school at all. |
| Compulsory Education | 1 | Education level is 9-year compulsory education. |
| High School | 1 | Education level is high school. |
| College and Above | 1 | Education level is college graduate and above. |
| Urban | 1 | Reside in urban area; otherwise in rural area. |
| Smoke | 1 | Consume cigarettes. |
| Alcohol | 1 | Consume alcoholic beverage more than twice a week. |
| Sugary Drinks | 1 | Consume soft drinks or sugary drinks more than twice a week. |
| Male | 1 | Gender is male; otherwise female. |
| Employed | 1 | Primary occupation is paid worker for enterprise or other person. |

| Table 6.2 Sample Descriptive Statistics on CHNS Data | | | | | |
|---|---|---|---|---|---|
| Variable | | Mean | SD | Min. | Max. |
| Participation Choice | Overall | 0.120 | 0.325 | 0 | 1 |
| | Between | | 0.314 | 0 | 1 |
| | Within | | 0.183 | -0.680 | 0.953 |
| Exercise Time on Weekday | Overall | 75.402 | 69.043 | 2 | 1202 |
| | Between | | 68.087 | 2 | 1202 |
| | Within | | 21.755 | -209.599 | 360.402 |
| Log Exercise Time on Weekday | Overall | 4.048 | 0.737 | 0.693 | 7.092 |
| | Between | | 0.719 | 0.693 | 7.092 |
| | Within | | 0.236 | 2.16 | 5.84 |
| Exercise Time on Weekend | Overall | 80.461 | 63.326 | 1 | 600 |
| | Between | | 60.431 | 1 | 600 |
| | Within | | 22.65 | -209.539 | 370.461 |
| Log Exercise Time on Weekend | Overall | 4.139 | 0.725 | 0 | 6.397 |
| | Between | | 0.702 | 0 | 6.397 |
| | Within | | 0.238 | 2.092 | 6.186 |
| Widowed/Divorced/Separated | Overall | 0.064 | 0.245 | 0 | 1 |
| | Between | | 0.236 | 0 | 1 |
| | Within | | 0.099 | -0.736 | 0.897 |
| Single | Overall | 0.077 | 0.266 | 0 | 1 |
| | Between | | 0.302 | 0 | 1 |
| | Within | | 0.103 | -0.673 | 0.910 |
| Married | Overall | 0.859 | 0.348 | 0 | 1 |
| | Between | | 0.365 | 0 | 1 |
| | Within | | 0.138 | 0.026 | 1.659 |
| Age | Overall | 47.166 | 14.489 | 18 | 97.100 |
| | Between | | 15.260 | 18 | 95.150 |
| | Within | | 3.448 | 38.316 | 55.632 |
| Annual Household Income | Overall | 29850.37 | 37840.66 | 0 | 900600 |
| | Between | | 38090.33 | 0 | 900600 |
| | Within | | 20869.91 | -347969.6 | 469527.2 |
| Log Household Income | Overall | 9.793 | 1.052 | 1.775 | 13.711 |
| | Between | | 0.978 | 4.129 | 13.711 |
| | Within | | 0.591 | 4.52 | 14.049 |
| Illiteracy | Overall | 0.202 | 0.401 | 0 | 1 |
| | Between | | 0.363 | 0 | 1 |
| | Within | | 0.153 | -0.631 | 1.035 |
| Compulsory Education | Overall | 0.520 | 0.500 | 0 | 1 |
| | Between | | 0.477 | 0 | 1 |
| | Within | | 0.190 | -0.314 | 1.353 |
| High School | Overall | 0.204 | 0.403 | 0 | 1 |
| | Between | | 0.396 | 0 | 1 |
| | Within | | 0.142 | -0.629 | 1.038 |
| College and Above | Overall | 0.070 | 0.254 | 0 | 1 |
| | Between | | 0.280 | 0 | 1 |
| | Within | | 0.085 | -0.73 | 0.903 |
| Urban | Overall | 0.326 | 0.469 | 0 | 1 |
| | Between | | 0.487 | 0 | 1 |
| | Within | | 0 | 0.326 | 0.326 |
| Smoke | Overall | 0.301 | 0.459 | 0 | 1 |
| | Between | | 0.435 | 0 | 1 |
| | Within | | 0.170 | -0.533 | 1.134 |

| Table 6.2 Sample Descriptive Statistics on CHNS Data (Continued) | | | | | |
|---|---|---|---|---|---|
| Alcohol | Overall | 0.256 | 0.436 | 0 | 1 |
| | Between | | 0.397 | 0 | 1 |
| | Within | | 0.210 | -0.578 | 1.089 |
| Sugar Drinks | Overall | 0.110 | 0.312 | 0 | 1 |
| | Between | | 0.302 | 0 | 1 |
| | Within | | 0.159 | -0.640 | 0.860 |
| Male | Overall | 0.505 | 0.500 | 0 | 1 |
| | Between | | 0.500 | 0 | 1 |
| | Within | | 0 | 0.505 | 0.505 |
| Employed for Wage | Overall | 0.331 | 0.471 | 0 | 1 |
| | Between | | 0.457 | 0 | 1 |
| | Within | | 0.195 | -0.502 | 1.165 |

| Table 6.3 Physical Activity Participation Choice and Time Spent among China Population | | | |
|---|---|---|---|
| Determinants | Choice (N=17,765) | Time on Weekday (N=1,748) | Time on Weekend (N=1,759) |
| Log Household Income | 0.524*** | 0.039* | 0.050** |
| Widowed/Divorced/Separated | -1.297*** | -0.090 | -0.112 |
| Married | -1.171*** | -0.045 | -0.025 |
| Age | 0.025*** | 0.001 | -0.002 |
| Male | 0.348*** | 0.021 | 0.047 |
| Illiteracy | -0.819*** | 0.006 | 0.017 |
| High School | 0.813*** | 0.040 | 0.038 |
| College and Above | 1.104*** | 0.075* | 0.073* |
| Urban | 1.368*** | 0.046 | 0.103*** |
| Employed for Wage | 0.553*** | -0.287*** | -0.209*** |
| Smoke | -0.465*** | -0.056 | -0.054 |
| Alcohol | 0.247*** | -0.014 | -0.012 |
| Sugary Drinks | 0.538*** | 0.027 | 0.066 |
| $R^2$ *Within* | | *0.032* | *0.007* |
| $R^2$ *Between* | | *0.043* | *0.030* |
| $R^2$ *Overall* | | *0.044* | *0.027* |
| $\sigma_u$ | *1.720* | *0.367* | *0.249* |
| $\sigma_e$ | | *0.622* | *0.649* |
| $\rho$ | *0.474* | *0.258* | *0.128* |

*(***, **, * denote significance at the 1%, 5%, 10% level, respectively.)*

| Table 7.1 The Comparison of Dependent Variable Means of U.S. and China Models | | | |
|---|---|---|---|
| Variable | United States | China | |
| Participation Choice | 0.649 | 0.12 | |
| Average Exercise Time per Day | 30.418 | Mon-Fri | Sat-Sun |
| | | 75.402 | 80.461 |

| Table 7.2 The Comparison of Determinants on Physical Activity Participation in U.S. and in China | | | | |
|---|---|---|---|---|
| Determinant | Choice | | Time | |
| | United States | China | United States | China |
| Income | + | + | NS | + |
| Male | + | + | + | NS |
| Age | - | + | + | NS |
| Married | - | - | - | NS |
| Education | + | + | - | + |

**Appendix**

**Kentucky Exercise & Health Survey**

*This survey asks a few questions about your work time and leisure time physical activity choices, cigarette consumption and alcohol consumption. There are no right or wrong answers. Please read each question carefully and answer it to the best of your ability. The information that you give will be kept in strictest confidence and only used for research purpose. Thank you.*

1. On average, how often do you regularly participate in the activities below during your **leisure time** (the time not used for work or sleep)? (Please check one per row.)

| | None | Less than once a month | 1-2 times a month | 1-2 times a week | 3-4 times a week | Almost daily |
|---|---|---|---|---|---|---|
| Sports or exercise | | | | | | |
| Housework or gardening | | | | | | |

2. How much time do you spend on those activities **on a typical day** during your **leisure time**? (Please fill in a number in either column for each row.)

| | Minutes | **OR** | Hours |
|---|---|---|---|
| Sports or exercise | | | |
| Housework or gardening | | | |

3. About how many hours **a week** do you usually work? ____ hours (Please fill in a number.)

4. Do you usually walk or bike to work or school? (Please check one.)

| Yes | | No | |
|---|---|---|---|

If yes, how long does a round trip take? (Please fill in a number in either column.)

| Minutes | **OR** | Hours |
|---|---|---|
| | | |

5. How often do you participate in physical activities (carrying loads, construction work, farming, etc.) as part of your **work**? (Please check one.)

| | None | Less than once a month | 1-2 times a month | 1-2 times a week | 3-4 times a week | Almost daily |
|---|---|---|---|---|---|---|
| Work time physical activity | | | | | | |

6. How much time do you spend doing physical activity **at work** on a typical day? (Please fill in a number in either column.)

| Minutes | **OR** | Hours |
|---|---|---|
| | | |

7. How often do your family members or close friends regularly play sports or exercise? (Please check one.)

| None | Less than once a month | 1-2 times a month | 1-2 times a week | 3-4 times a week | Almost daily |
|---|---|---|---|---|---|
| | | | | | |

8. How often do you exercise or play sports with your family members or close friends? (Please check one.)

| None | Less than once a month | 1-2 times a month | 1-2 times a week | 3-4 times a week | Almost daily |
|---|---|---|---|---|---|
| | | | | | |

9. Do you pay for participating in sports or other physical activities (gym membership fees, league fees, personal training fees, equipment, etc.)? (Please check one.)

| Yes | | No | |
|---|---|---|---|

If yes, about how much do you spend **annually**? $_____(Please fill in a number.)

10. Do you feel it is safe to walk or jog in your neighborhood? (Please check one.)

| Yes | | No | |
|---|---|---|---|

11. Are there any exercise facilities or programs where you work? (Please check one.)

| Yes | | No | |
|---|---|---|---|

If yes, how often do you use them? (Please check one.)

| None | Less than once a month | 1-2 times a month | 1-2 times a week | 3-4 times a week | Almost daily |
|---|---|---|---|---|---|
| | | | | | |

12. Check all the reasons that you play sports or exercise.

| Enjoyment | Meet friends or new people | Control weight or build up muscle | Prevent disease or boost energy | Release stress or improve mood | Sleep better | Other (Explain below) |
|---|---|---|---|---|---|---|
| | | | | | | |

*Other (Please explain here):

13. Check all the reasons that prevent you from playing sports or exercising as much as you would like.

| Have physical activity as part of work | Lack of interest | Lack of money | Lack of skill | Lack of workout partner | Lack of facility | Lack of benefits | Other (Explain below) |
|---|---|---|---|---|---|---|---|
| | | | | | | | |

*Other (Please explain here):

14. How often do you smoke cigarettes? (Please check one.)

| None | Less than once a month | 1-2 times a month | 1-2 times a week | 3-4 times a week | Almost daily |
|---|---|---|---|---|---|
| | | | | | |

15. How often do you smoke more than 20 cigarettes in a single day? (Please check one.)

| None | Less than once a month | 1-2 times a month | 1-2 times a week | 3-4 times a week | Almost daily |
|---|---|---|---|---|---|
| | | | | | |

16. How often do you have at least one drink of any alcoholic beverage, such as beer, wine, a malt beverage or liquor? (Please check one.)

| None | Less than once a month | 1-2 times a month | 1-2 times a week | 3-4 times a week | Almost daily |
|---|---|---|---|---|---|
| | | | | | |

17. One drink is equivalent to a 12-ounce beer, a 5-ounce glass of wine, or a drink with one shot of liquor. If you are **male**, how often do you have at least **5** drinks in a single day? If you are **female**, how often do you have at least **4** drinks in a single day? (Please check one.)

| None | Less than once a month | 1-2 times a month | 1-2 times a week | 3-4 times a week | Almost daily |
|---|---|---|---|---|---|
| | | | | | |

18. How many years have you lived in Kentucky? (Please check one.)

| Less than 1 Year | 1-3 Years | 3-5 Years | 5-10 Years | More than 10 Years |
|---|---|---|---|---|
| | | | | |

19. What is your age? (Please check one.)

| 18-24 | 25-34 | 35-44 | 45-54 | 55-64 | 65 and above |
|---|---|---|---|---|---|
| | | | | | |

20. What is your gender? (Please check one.)

| Male | | Female | |
|---|---|---|---|

21. How many adult males, adult females and children are there in your household? (Please fill in the number in each cell.)

| | Men | Women | Children under 18 |
|---|---|---|---|
| Number | | | |

22. How would you describe your **physical** and **mental** health status? (Please check one per row.)

| | Excellent | Very good | Good | Fair | Poor |
|---|---|---|---|---|---|
| Physical health | | | | | |
| Mental health | | | | | |

23. Which of the following best describes your marital status? (Please check one.)

| Married | Divorced | Widowed | Separated | Never married | Unmarried but living with significant other |
|---|---|---|---|---|---|
| | | | | | |

24. What is the highest grade or year of school you completed? (Please check one.)

| High school or below | Some college | Graduate or above |
|---|---|---|
| | | |

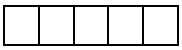
25. Which of the following best describes your employment status? (Please check one.)

| Employed for wages | Self-employed | Out of work for less than 1 year | Out of work for 1 year or more | A Home maker | A Student | Retired | Unable to work |
|---|---|---|---|---|---|---|---|
| | | | | | | | |

26. Which of the following best describes your **annual household** income? (Please check one.)

| | | | |
|---|---|---|---|
| 0-$9,999 | | $30,000-$39,999 | |
| $10,000-$14,999 | | $40,000-$59,999 | |
| $15,000-$19,999 | | $60,000-$79,999 | |
| $20,000-$29,999 | | $80,000 and above | |

27. Please fill in the 5-digit ZIP Code where you live.



*Please fold this survey and mail it back in the postage-paid envelope provided. Please return as soon as possible.*

**THANK YOU VERY MUCH!**

**References**

Anokye, N. K., Pokhrel, S., Buxton, M., and Fox, J. 2012. The Demand for Sports and Exercise: Results from an Illustrative Survey. European Journal of Health Economics, 13:277–287.

Attard, S. M., Howard, A. G., Herring, A. H., Zhang, B., Du, S., Aiello, A. E., Popkin, B.M. and Gordon-Larsen, P. 2015. Differential Associations of Urbanicity and Income with Physical Activity in Adults in Urbanizing China: Findings from the Population-Based China Health and Nutrition Survey 1991-2009. International Journal of Behavioral Nutrition and Physical Activity, 12(1):152.

Bauman, A., Ma, G., Cuevas, F., Omar, Z., Waqanivalu, T., Phongsavan, P., Keke, K. and Bhushan, A. 2011. Cross-National Comparisons of Socioeconomic Differences in the Prevalence of Leisure-Time and Occupational Physical Activity, and Active Commuting in Six Asia-Pacific Countries. Journal of Epidemiology and Community Health, 65(1): 35-43.

Becker, G. S. 1965. A Theory of the Allocation of Time. The Economic Journal, 75: 493-517.

Becker, G. S., and K. M. Murphy. 1988. A Theory of Rational Addiction. Journal of Political Economy, 96: 675-700.

Becker, G. S. 1993. Nobel Lecture: The Economic Way of Looking at Behavior. Journal of Political Economy, 101: 385-409.

Becker, G. S., and C. B. Mulligan. 1997. The Endogenous Determination of Time Preference. The Quarterly Journal of Economics, 112: 729-758.

Becker, G. S., M. Grossman., and K. M. Murphy. 1991. Rational Addiction and the Effect of Price on Consumption. The American Economic Review, 81:237-241.

Binkley, J. 2010. Low Income and Poor Health Choices: The Example of Smoking. American Journal of Agricultural Economics, 92: 972–984.

Bretteville-Jensen, A.L. 1999. Addiction and Discounting. Journal of Health Economics, 18: 393-407.

Breuer, C., and Wicker, P. 2008. Demographic and Economic Factors Concerning the Inclusion in the German Sport System. A Micro-analysis of the Years 1985 to 2005. European Journal for Sport and Society, 5: 35-43.

Brownson, R. C., Boehmer, T. K., and Luke, D. A. 2005. Declining Rates Of Physical Activity in the United States: What Are the Contributors? Annual Review of Public Health, 26:421–443.

Caspersen, C. J., Christenson, G.M. and Pollard, R. A. 1986. Status of the 1990 Physical Fitness and Exercise Objectives-Evidence from NHIS 1985. Public Health Reports, 101: 587–592.

Cawley, J., Moran, J. and Simon K. 2010. The Impact of Income on the Weight of Elderly Americans. Health Economics, 19: 979–993.

Cawley, J. 1999. Obesity and Addiction. PhD Dissertation. University of Chicago.

Cawley, J. 2004. An Economic Framework for Understanding Physical Activity and Eating Behaviours. American Journal of Preventive Medicine, 27: 117-125.

Cawley, J. and Ruhm, C. 2011. The Economics of Risky Health Behaviors. NBER Working Paper 17081.

Cawley, J. and Liu, F., 2007. Maternal Employment and Childhood Obesity: A Search for Mechanisms in Time Use Data. NBER Working Paper 13600.

Cerdá, M. , Johnson-Lawrence, V. D. , and Galea, S. 2011. Lifetime Income Patterns and Alcohol Consumption: Investigating the Association between Long- and Short-term Income Trajectories and Drinking. Social Science and Medicine, 73: 1178–1185.

China Ministry of Human Resources and Social Security. Online. Available at http://www.mohrss.gov.cn/gkml/xxgk/201404/t20140430_129459.htm. [Retrieved March 2016].

Cutler, D., and E. Glaeser. 2005. "What Explains Differences in Smoking, Drinking, and Other Health- Related Behaviors?" American Economic Review, 95: 238–242.

Cutler, D. M., E.L. Glaeser., and J.M. Shapiro. 2003. Why Have Americans Become More Obese? Cambridge MA: National Bureau of Economic Research (Working Paper 9446).

Cutler, D.M. and Lleras-Muney, A. 2010. Understanding Differences in Health Behaviors By Education. Journal of Health Economics, 29:1-28.

DiPietro, L., Ostfeld, A. M. and Nadel, E. R. 1993. A Survey for Assessing Physical Activity Among Older Adults. Medicine and Science in Sports and Exercise, 25:628-642.

Downward, P. 2007. Exploring the Economic Choice to Participate in Sport: Results from the 2002 General Household Survey. International Review of Applied Economics, 21: 633-653.

Downward, P., and Riordan, J. 2007. Social Interactions and the Demand for Sport: an Economic Analysis. Contemporary Economic Policy, 25: 518-537.

Droomers, M. , Schrijvers, C. T. M. , Stronks, K. , Mheen, D. , and Mackenbach, J. P. 1999. Educational Differences in Excessive Alcohol Consumption: The Role of Psychosocial and Material Stressors. Preventive Medicine, 29: 1-10.

Elosua R., Garcia M., Aguilar A., Molina L., Covas M.I. and Marrugat J. 2000. Validation of the Minnesota Leisure Time Physical Activity Questionnaire in Spanish Women. Medicine and Science in Sports and Exercise, 32: 1431-1437.

Ettner, S. J. 1996. New Evidence on the Relation between Income and Health. Journal of Health Economics, 15: 67-85.

Farrell, L., and Shields, M. A. 2002. Investigating the Economic and Demographic Determinants of Sporting Participation in England. Journal of Royal Statistics Society A, 165: 335-348.

Fertig, A., Gerhard, G. and Tchernis, R. 2009. The Connection between Maternal Employment and Child Obesity: Inspecting the Mechanisms. Review of Economics of the Household 73: 227–255.

Garcia, J., Lera, F., and Suarez, M. J. 2011. Estimation of a Structural Model of the Determinants of the Time Spent on Physical Activity and Sport: Evidence for Spain. Journal of Sports Economics, 12:515-537.

Grossman, M. 1972. On the Concept of Health Capital and the Demand for Health. Journal of Political Economy 80(2): 223–255.

Hallal, P. C., M. R. Azevedo, F. F. Reichert, F. V. Siqueira, C. L. P. Araújo, and C. G. Victora. 2005. Who, When, and How Much? Epidemiology of Walking in a Middle-Income Country. American Journal of Preventive Medicine, 28(2): 156–161.

Hamilton, V., and Hamilton B. H. 1997. Alcohol and Earnings: Does Drinking Yield a Wage Premium?. The Canadian Journal of Economics, 30: 135-151.

Heckman, J. J. 1974. Life Cycle Consumption and Labor Supply: An Explanation of the Relationship between Income and Consumption Over the Life Cycle. The American Economic Review, 64: 188-194.

Heckman, J. J. 1976. A Life-Cycle Model of Earnings, Learning, and Consumption. Journal of Political Economy, 84: S11-S44.

Heckman, J. J. 2007. The Economics, Technology, and Neuroscience of Human Capability Formation. Proceedings of the National Academy of Sciences of the United States of America, 104:13250–13255.

Herd, D. 1990. Subgroup Differences in Drinking Patterns among Black and White Men: Results from a National Survey. Journal of Studies on Alcohol and Drugs, 51: 221-232.

Hovell, M.F., Hofstetter, C.R., Sallis, J.F., Rauh, M.J.D., and Barrington, E. 1992. Correlates of Change in Walking for Exercise: an Exploratory Analysis. Research Quarterly for Exercise and Sport, 63:425-434.

Hu, G., H. Pekkarinen, O. Hanninen, Z. Yu, H. Tian, Z. Guo, and A. Nissinen. 2002. Physical Activity During Leisure and Commuting in Tianjin, China. Bulletin of the World Health Organization, 80 (12): 933–938.

Hu, X., and Stowe, C. J. 2016. The Effect of Income on Health Choices: Alcohol Use. Journal of Agribusiness, 36: 193-210.

Humphreys, B. R., and Ruseski, J. E. 2006. Economic Determinants of Participation in Physical Activity and Sport. IASE Working Paper No. 06-13.

Humphreys, B. R., and Ruseski, J. E. 2007. Participation in Physical Activity and Government Spending on Parks and Recreation. Contemporary Economic Policy, 25: 538-552.

International Monetary Fund. World Economic Outlook. Online. Available at http://www.imf.org/external/pubs/ft/weo/faq.htm#q4b. [Retrieved October 2016].

Jones-Smith, J.C. and Popkin, B.M. 2010. Understanding Community Context and Adult Health Changes in China: Development of an Urbanicity Scale. Social Science & Medicine, 71(8): 1436-1446.

Kenkel D.S., Ribar, D.C., Cook, P.J., and Peltzman, S. 1994. Alcohol Consumption and Young Adults' Socioeconomic Status. Brookings Papers on Economic Activity, Microeconomics, 119-175.

Keough, K.A., Zimbardo, P.G., and Boyd, J.N. 1999. Who's Smoking, Drinking, and Using Drugs? Time Perspective as a Predictor of Substance Use. Basic and Applied Social Psychology, 21: 149-164.

Kuvaja-Köllner, V., Valtonen, H., Komulainen, P., Hassinen, M. and Rauramaa, R. 2013. The Impact of Time Cost of Physical Exercise on Health Outcomes by Older Adults: the DR's EXTRA Study. The European Journal of Health Economics, 14(3): 471-479.

Lakdawalla D., Philipson T. 2002. The Growth of Obesity and Technological Change: A Theoretical and Empirical Examination. NBER Working Paper.

Lee S.A., Xu W.H., Zheng W., Li H., Yang G., Xiang Y.B., and Shu X.O. 2007. Physical Activity Patterns and Their Correlates among Chinese Men in Shanghai. Medicine and Science in Sports and Exercise, 39(10): 1700-1707.

Lopez, R.P. and Hynes, H.P. 2006. Obesity, Physical Activity, and the Urban Environment: Public Health Research Needs. Environmental Health, 5:25.

Meltzer, D. O., and Jena, A. B. 2010. The Economics of Intense Exercise. Journal of Health Economics, 29: 347–352.

Park, J., R. Holcomb, K. Raper, and O. Capps. 1996. A Demand System Analysis of Food Commodities by U.S. Households Segmented by Income. American Journal of Agricultural Economics, 78: 290-300.

Philipson T. 2001. The World-Wide Growth in Obesity: an Economic Research Agenda. Health Economics 10: 1-7.

Philipson T., Posner R. A. 1999. The Long-run Growth in Obesity as a Function of Technological Change. NBER Working Paper 7423.

Powell, L. M., Slater, S., Mirtcheva, D., Bao, Y., and Chaloupka, F. J. (2007). Food Store Availability and Neighborhood Characteristics in the United States. Preventive Medicine, 44: 189-195.

Richardson, M.T., Leon, A.S., Jacobs, D.R., Ainsworth, B.E. and Serfass, R. 1994. Comprehensive Evaluation of the Minnesota Leisure Time Physical Activity Questionnaire. Journal of Clinical Epidemiology, 47: 271-281.

Sallis, J.F., Hovell, M.F., Hofstetter, C.R., Faucher, P., Elder, J.P., Blanchard, J., Caspersen, C.J., Powell, K.E., and Christenson, G.M. 1989. A Multivariate Study of Determinants of Vigorous Exercise in a Community Sample. Preventive Medicine, 18:20-34.

Sallis, J.F., Hovell, M.F., Hofstetter, C.R., and Barrington, E. 1992. Explanation of Vigorous Physical Activity During Two Years Using Social Learning Variables. Social Science and Medicine, 34:25-32.

Scharff, R. L., and Viscusi, W. K. 2011. "Heterogeneous Rates of Time Preference and the Decision to Smoke". Economic Inquiry, 49: 959-972.

Schmeiser, M. D. 2009. Expanding Wallets and Waistlines: The Impact of Family Income on the BMI of Women and Men Eligible for the Earned Income Tax Credit. Health Economics, 18: 1277-1294.

Slade, A. N. 2012. Health Investment Decisions in Response to Diabetes Information in Older Americans. Journal of Health Economics, 31: 502-520.

Stephens, T., D. R. Jacobs, and C. C. White. 1985. A Descriptive Epidemiology of Leisure-Time Physical Activity. Public Heath Reports, 100(2): 147–158.

Stewart, H., N. Blisard, and D. Joliffe. 2003. Do Income Constraints Inhibit Spending on Fruits and Vegetables among Low-income Consumers?. Journal of Agricultural and Resource Economics, 28: 465– 480.

United Nations Statistics Division. Standard Country or Area Codes for Statistical Use. Online. Available at http://unstats.un.org/unsd/methods/m49/m49.htm. [Retrieved October 2016].

United Nations Statistics Division. Composition of Macro Geographical (Continental) Regions, Geographical Sub-Regions, and Selected Economic and Other Groupings. Online. Available at http://unstats.un.org/unsd/methods/m49/m49regin.htm#ftnc. [Retrieved October 2016].

United States Census Bureau. QuickFacts Kentucky. Online. Available at http://www.census.gov/quickfacts/table/IPE120214/21. [Retrieved May 2016].

United States Social Security Administration. Retirement Planner: Benefits By Year Of Birth. Online. Available at https://www.ssa.gov/planners/retire/agereduction.html. [Retrieved March 2016].

Viscusi, W. K., and M. J. Moore. 1989. Rates of Time Preference and Valuations of the Duration of Life. Journal of Public Economics, 38: 297–317.

Vlahov, D. and Galea, S. 2002. Urbanization, Urbanicity, and Health. Journal of Urban Health, 79(1): S1-S12.

Wilsnack, R.W., Vogeltanz, N.D., Wilsnack. S.C., and Harris, T.R. 2000. Gender Differences in Alcohol Consumption and Adverse Drinking Consequences: Cross-cultural Patterns. Addiction, 95: 251-265.

World Bank. GNI per Capita Ranking, Atlas Method and PPP Based. Online. Available at http://databank.worldbank.org/data/download/GNIPC.pdf. [Retrieved October 2016].

World Bank. World Bank Country and Lending Groups. Online. Available at https://datahelpdesk.worldbank.org/knowledgebase/topics/19280-country-classification. [Retrieved October 2016].

World Health Organization: World health report. 2002.

Øvrum, A. 2011. Socioeconomic Status and Lifestyle Choices: Evidence from Latent Class Analysis. Health Economics, 20 (8): 971–984.

**Vita**

**Xiaowen Hu**

**EDUCATION**

- M.Sc., Economics, University of Kentucky, Lexington, KY, 2010-2011
- B.Sc., Agricultural Economics, Renmin University of China, Beijing, China, 2006-2010

**RESEARCH INTERESTS**

- Consumer Behavior, Health Economics; Discrete Choice Modeling.

**PUBLICATIONS**

- **Hu, X.**, and C. J. Stowe. 2016. "The Effect of Income on Health Choices: Alcohol Use." Journal of Agribusiness. 36: 193-210.

**PRESENTATIONS (3 Selected Papers; 1 Selected Poster)**

- **Hu, X.** and C. J. Stowe. "Income and Health Choices: Physical Activity Evidence from China." Selected paper at the SAEA Annual Meeting, San Antonio, TX, February 6-9, 2016.
- **Hu, X.**, and C. J. Stowe. "Income and Physical Activity: A Comparison between United States and China." Selected paper at the SAEA Annual Meeting, Dallas, TX, February 1-4, 2014.
- **Hu, X.**, and C. J. Stowe. "The Effect of Income on Health Choices: Physical Activity and Alcohol Use." Selected poster at the AAEA & CAES Joint Annual Meeting, Washington D.C., August 4-6, 2013.
- **Hu, X.**, and C. J. Stowe. "The Effect of Income on Health Choices: Alcohol Use." Selected paper at the SAEA Annual Meeting, Orlando, FL, February 2-5, 2013.

**GRANTS**

- **Hu, X.(PI)**, S. Quan, Y. Jia, and S. Zhang. "Impact of the Functioning of a Rural Enterprise on the Income of Peasants Living in That Area.", ￥40,000, funded by the Ministry of Education of China, National Innovative and Experimental Program, Nov 2007.

**RESEARCH AND WORK EXPERIENCE**

- Aug 2010-present, **Research Assistant**, University of Kentucky
  - **Dissertation Project III** "Income and Physical Activity in United States: An Empirical Analysis on National Health and Nutrition Examination Survey.", Dec 2014-Present.
  - **Dissertation Project II** "Income and Health Choice: Physical Activity Evidence from China.", April 2014-Present.
  - **Dissertation Project I** "Income and Physical Activity: A Kentucky Health Choice Survey.", Sept 2013-Present.
  - **Independent Study Project V** "The Effect of Income on Health Choices: Physical Activity and Alcohol Consumption.", Jan 2013-Aug 2013.
  - **Independent Study Project IV** "Classification of Decision Strategies from EEG Data during Economic Decision Making.", Sept 2012-Dec 2012.

*Collaborated with fellows at Department of Computer Science.

- **Independent Study Project III** "Income and Its Effect on Health Choice: Alcohol Use.", Jan 2012-Sept 2012.
- **Independent Study Project II** "China's Food Safety Law: Did It Make Chinese Eat More Safely?", March 2012-May 2012.
- **Independent Study Project I** "The Relationship between Air Quality and Obesity Rate - Based on the Case Study in United States.", Oct 2011-Dec 2011.

- Sept 2007-July 2010, **Research Assistant**, Renmin University of China
  - **Independent Study Project** "Impact of the Functioning of a Rural Enterprise on the Income of Peasants Living in That Area.", Nov 2007-March 2009.

**TEACHING EXPERIENCE**

- Aug 2014-Dec 2014, **Teaching Assistant**, University of Kentucky
  - Recitation instructor for graduate microeconomic theory class (AEC503); actively assisted primary instructor with course planning on price theory and application/decision making/game theory.
  - Provide feedback and mentoring to students.

**SERVICE**

- Information Officer, Graduate Student Organization in Agricultural Economics, University of Kentucky, 2011-2012.