The Effect of Income and Policy Factors on Physical Activity

by

Saira Khan

Undergraduate Economics Honors Program

Thesis Advisor: Bingxiao Wu

Table of Contents

Introduction 3
Hypothesis 5
Literature Review 6
Methodology 12
Data Description 14
Descriptive Statistics 19
Results 25
Analysis of the Effect of Income on Exercise 25
Controls Analysis 29
Probit and Logit Analysis 35
Policy Factor Analysis 37
Probit and Logit Analysis of Policy Factors 49
Conclusion 51
References 54
Appendix 56

Introduction

A lot of research has been done to examine how socioeconomic class affects health, and the results overwhelmingly show that people of a lower socioeconomic class are more likely to have poorer health. Lantz et al (1998) used data from the American Changing Lives longitudinal survey to observe the impacts of income, education and health behaviors on the risk of dying in the next 7.5 years. They found that income was significant predictor on whether or not someone would die in the next 7.5 years, even more so than education. "Both men and women in the lowest-income category were more than 3 times as likely to die during the follow-up period of the study than those in the highest group, controlling for age, and other sociodemographic variables (Lantz et al., 1998)." In a review, Adler (1993), examined scholarly articles to see if socioeconomic class relates to health in other parts of the world. He found that socioeconomic class was related to mortality in both the aggregate and individual level. People who live in neighborhoods with lower socioeconomic class, and those who are individually poorer were more likely to have a higher mortality risk. Another interesting finding was that the effect of socioeconomic class on health does not show a threshold-type of model. There is no income level that a person must exceed so that their socioeconomic class no longer effects their health, which indicated that the model is more linear than was expected. Socioeconomic class continues to effect health as income increases into the higher categories. This shows that socioeconomic class is an important variable to study in its relationship with health.

Adler (1993) found that people in developed countries were more likely to die of degenerate diseases that are related with life-style behaviors. Lantz et al (1998) found that people in the lowest-income class were more likely to die due to tumor and cardio-vascular

deaths compared to people in the higher-income class. People in the middle-class were more likely to die due to tumor deaths. Lack of physical activity is a risk factor for some of the most popular degenerate diseases in developed countries such as heart disease, cancer, and diabetes. There is also evidence that people of a lower-socioeconomic class exercise less than their higher socioeconomic counterparts. This discrepancy in health behavior could explain why people of a lower-socioeconomic class are less healthy. There is a need to observe how physical activity relates to socioeconomic class, and that is the need this paper hoped to full-fill. This research study explores the relationship between income and exercise. This analysis is followed by a study on how policy factors such as access to physical activity resources affects exercise. Physical activity resources include fitness centers, nature parks, ski areas, bowling arenas, golf clubs, and marinas. In addition, the effect of unemployment rate, property crime, and violent crime is also observed.

This study uses data from five consecutive years so that the time effects of the policy variables can be observed. We hoped that this allows us to naturally observe the increase and decrease of economic variables and see how they effect a person's ability to exercise. If physical fitness centers in an area increased over the five year period, we can see what effect this had on the individuals of that area in terms of whether or not their increased their level of exercise.

Studying these economic variables is important for policy decisions. If our results conclude that people of a lower socioeconomic class do increase their exercise when there are more fitness centers available then maybe cities with a lack of these resources should invest in them. If crimes effects physical activity negatively, it gives policy makers an even bigger incentive to fight crime. This research question is important to study so that we can identify

which policy indicators are important in helping people of a lower socioeconomic class exercise more and improve their health.

Hypothesis

This research study first attempts to see how socioeconomic class affects physical activity. Our hypothesis is that socioeconomic class is positively correlated with physical activity; wealthier people are more likely to engage in physical activity. Secondly, this research study attempts to see what prevents people of a lower socioeconomic class from engaging in physical activity. This research study examines if people of a lower socioeconomic class have lower access to fitness centers, nature parks, ski areas, bowling arenas, golf clubs, and marinas. Our hypothesis is that the more physical activity resources are available in the county that one lives in, the more likely it is that one will exercise. In addition, this research study examines the effects of both violent and property crimes on physical activity because these factors could affect someone's ability to feel safe in their neighborhood and exercise. We expect a negative correlation between crime and exercise. There is evidence that when the market goes down, people become healthier and engage in better health behaviors. This is due to people having more time to exercise when they are unemployed. However, there are also studies that show that when the market is up, people are more likely to exercise. This is because they are happier and richer so they are able to devote more time to exercising. So lastly, this study examines area unemployment rate's effect on physical activity.

Literature Review

Many studies have shown that people of a lower socioeconomic class are less healthy than their higher socioeconomic class counterparts. Because physical exercise is one factor that has been shown to improve health, it is important to study how socioeconomic class affects physical exercise.

Many studies from around the world have shown that people of a lower-socioeconomic class exercise less. Droomers et al. (1998), used a cross-sectional dataset from a Dutch study on socioeconomic heath differences in 1991 to examine whether psychosocial factors (coping resources, personality, and stressors) and material factors (financial situation, employment status, and living conditions) affected physical activity. They used highest attained level of education as their variable for socioeconomic class because they believed it measured socioeconomic class well in the Netherlands. Using a logistic regression model, they found that physical inactivity was more prevalent in people with lower education levels, and that equivalent income and financial problems were statistically significant correlates of physical activity, which shows that the less income one has, the more physically inactive they are. They also found that employment status was a statistically significant correlate of physical activity; those who are employed are more physically active compared to those who are unemployed and those who cannot work due to a disability.

Similarly, Lynch, Kaplan and Salonen (1997), examined why poorer people were more likely to engage in unhealthy behaviors using the 1984-1989 Kuopio Ischaemic Heart Disease Risk Factor Study which asked Finish men about risk factors that affect Cardiovascular Health which included Physical Activity. They found that people of advantaged backgrounds were

significantly less likely to be in the lowest quartile of physical activity, and that those in the least educated group were more likely to report no activity. Both farmers and blue-collar men were more likely to report lower levels of physical activity than people in white-collar jobs.

Similar evidence can be seen in studies conducted in the United States, the population of interest for this study . Lantz et al. (1998) used the 1986 American's Changing Lives longitudinal survey conducted by the University of Michigan. This survey sampled non-institutionalized people 25 years of age or older in the United States and oversampled those 60 years or older to see the relation between different health behaviors and mortality 7.5 years later They found that those with the least amount of education and lowest income were more likely to be in the lowest quartile of physical activity Fahlman, Hall and Lock (2006) conducted a study using a cross-sectional sample of high school females in a Midwestern city in the United States which included a 15-question survey and a FITNESSGRAM fitness test which measured aerobic capacity, BMI, and body fat to observe differences in physical activity. The results showed that high school females of a lower SES did feel more barriers to exercise, had higher BMIs and more body fat

To delve deeper into why people of a lower socioeconomic class exercise less, many researchers have looked into conceptual factors that might affect lower-income individuals' ability to exercise. The Droomers et al. (1998) study that was mentioned earlier found that external locus of control, parochialism (closed attitude), neuroticism (nervous disposition), a lack of active problem focusing, and a lack of emotional social support were more often observed in lower educational groups. Similarly, Murray, Rodgers and Fraser (2011) examined the relationship between control beliefs and socioeconomic status using a one time, cross

sectional mail out survey in a large urban Canadian city. They found that people with higher socioeconomic status exercised more and reported stronger intentions to exercise, which is consistent with the other studies listed above (Murray, Rodgers and Fraser, 2011). Murray, Rodgers and Fraser (2011) wanted to examine if distal (i.e., personal constraints) and proximal control beliefs (i.e., scheduling self-efficacy) mediated the association between socioeconomic status and exercise. They found that both types of beliefs partially explained the variance between people of different socioeconomic class

Baruth et al. (2014) conducted four focus groups between April and May 2008 with overweight or obese women predominantly African American to explore their perceptions and experiences with barriers to exercise. Barriers of exercise included being discouraged due to not seeing quick results, not having fun with exercise, people make cruel comments about their body, and being physically unable to do certain exercises. They also mentioned the competing demands on their time and lack of energy due to having jobs and families to take care of. They also talked about how their neighborhood was a barrier to exercise due to safety issues, violence, and unsupervised children making rude and disrespectful comments, traffic, and lack of access to recreational facilities.

Many of these women talked about the factors present in the built environment that hindered their ability to exercise. There is evidence in the literature to say that the built environment affects a person's ability and willingness to exercise. Rutt and Coleman (2005) conducted a phone survey in El Paso Texas to find out if increasing land mix increased the exercise done by overweight and obese people. Their sample contained many low-income and Hispanic residents and observed each neighborhood's number of GYMs, parks, physical activity

facilities, sidewalks, biking/walking paths. They found that those who lived in high land use areas such as cities were more likely to exercise, but only after controlling for an individual's socioeconomic class which allows them to hypothesis that socioeconomic class may be a factor on whether or not the build environment affects physical exercise.

Powell et al (2006), established the idea that areas with lower average income contained less physical activity resources. He used a business list developed by Dun and Bradstreet to obtain information on physical fitness facility, membership recreation clubs, dance studios, school, and public golf courses and linked it to 2000 Census data on neighborhood income and racial make-up to observe how physical activity resources are linked to these variables. Using multivariate and probit analysis, they found that higher median household income in a neighborhood was associated with higher numbers of all types of physical activity resources, with the highest association between membership recreation clubs and golf courses by increasing median neighborhood income from 25,000 to 75,000, the number of physical fitness facilities increased 17%, membership recreation clubs increased 38%, dance studios and classes 30%, and golf clubs 54% .

Lee, Cubbin and Winkleby (2007), found that neighborhoods with average lower socioeconomic class exercise less when they were researching the literature and wanted to conduct their own study to see if access to physical activity resources mediated that disparity. They used the 1979–90 Stanford Heart Disease Prevention Program (SHDPP), a survey of five cities in a 6 year field trial about community organization and health education in relation with cardiovascular disease. Using multilevel regression models, they found that after controlling for individual income class, people in neighborhoods with average lower socioeconomic class have

more energy expenditure overall but partook in less vigorous physical exercise than neighborhoods with higher average socioeconomic status.

Ross (2000) found similar results; people in poor neighborhoods walked more but participated in less vigorous exercise. Ross used multi-level data from a representative sample of Illinois residents and linked it to census-tract information about poverty and education to observe the effects of these variables on exercise. He found that people in poor neighborhoods walked more than those in more affluent places even though those in poorer neighborhoods had more concerns about safety. However, they found that the more college educated neighborhoods were more likely to walk. This is interesting because these two measures of income (average area income and education) show opposite effects Ross also examined whether the build environment affected vigorous exercises, and found no significant results Although, they did find that individual income affects strenuous exercise. People with individual higher levels of income and education, were more likely to exercise, which is consistent with the studies listed above

Hino et al. (2011), conducted a phone survey in Curitiba, Brazil asking about leisure time walking activity, and moderate to vigorous physical activity and linked it to build environment characteristics 500 m around the participant's home Using multivariate logistic regression, they found that areas with higher income level, higher number of GYMs were more likely to participate in both leisure time walking activity and moderate to vigorous physical activity. This is different from the previous studies listed above in that this study says people living in more affluent areas walk more than people living in less affluent areas, while other studies have said the opposite.

Many of these articles compare average income in a neighborhood to whether one exercises but very little research has been done on comparing individual income, specific build environment, and exercise. A working paper released by the National Bureau of Economic Research used the 2000-2005 Behavioral Risk Surveillance System data and linked it to the Business Patterns dataset to examine physical activity compared to many demographic factors and availability of resources (McInnes and Shinogle, 2009). The research study that I am conducting is similar to this working paper and I have worked on it to improve it's results while also using a more recent dataset.

Similar to the studies discussed, McInnes and Shinogle, 2009, found that greater education and income increases the likelihood of exercise. However, this study contrasted with the Droomers et al., study because said it found that people who were unemployed were more likely to exercise. This research paper also found that the number of gyms per capita did increase physical activity which is consistent with the sources listed above (McInnes and Shinogle, 2009). They found that gas prices were negatively correlated with physical exercise while bus fare prices were positively correlated, which is interesting because these two modes of transportation have opposite effects When stratified for income, this study found that none of these build environment variables significantly affected exercise.

More research is needed to examine the link between built environment variables such as number of parks, GYMs, price of gasoline, price of public transportation, crime and other variables and if these variables are more important for people of a lower socioeconomic class in their willingness to participate in physical activity. This research paper attempts to use a more

recent data set to learn more about specific build environment variables and their relation to socioeconomic class and physical activity.

Methodology

The first economic model in this research study attempts to see the relationship between income and exercise. In order to do this, we used the annual Behavioral Risk Surveillance System dataset that asks a nationally representative survey about health and health behaviors. One of the core questions in the Behavioral Risk Surveillance System is ”During the past month, other than your regular job, did you participate in any physical activities or exercises such as running, calisthenics, golf, gardening, or walking for exercise?” This binary variable is the dependent variable in this research study, and is used to represent exercise. The independent variable is socioeconomic class, which was asked to the individual using the question, “Is your annual household income from all sources:” and then lists the different income categories.

The second part of our analysis focuses on economic variables such as number of fitness centers or GYMs and how they affect exercise. In this analysis, the dependent variable stays as exercise while independent variables corresponding to the economic variables are added to the model. These added variables come from many different sources which include the County Business Patterns Dataset, Uniform Crime Reporting Program Data, and the Local Area Unemployment Statistics. The Behavioral Risk Surveillance System dataset was combined with the economic variable datasets using the FIPS State and County codes. The State and County

FIPS (Federal Information Processing Standards) code is used in many datasets created by the government so it is easy to find datasets that include these codes.

In addition to the variables discussed above, control variables are added which include age, education level, employment status, respondent's sex, metropolitan status code, race, BMI, and number of children. These control variables are added because they might affect the amount of exercise one does. Age is included because younger people are generally more active. Education level is included because educated people are more knowledgeable about the benefits of exercise so they are more likely to exercise. Employment status is included because unemployed people might have more time to exercise. Sex and race are included because both people of different sex and race might have different beliefs about exercise. Metropolitan status code (which measure whether one lives in a city, suburb, or rural area) is included because each environment has its own facilities for exercise. Body Mass Index is included because overweight people might have difficulty exercising. Number of children is included because people with children might have to take care of them for many hours and so would have less time to exercise.

All the datasets explained above were matched and merged with each other. The base dataset is the Behavioral Risk Surveillance System Dataset which contains values for each individual who participated in the survey. Each person's FIPS state and county code was then matched with the information in the County Business Patterns Dataset, the Uniform Crime Reporting Dataset, and the Local Area Unemployment Statistics which is possible because all three of these datasets are county level. Every dataset mentioned above is conducted annually, so there is data available for all five years this research study examines. All five years of data

were matched to their respective information. Then a time variable was generated indicating which year the data belongs to. Then finally, all five years of data was appended together to create a large dataset.

In this economic model there are two sources of variation: 1) Variation between counties, and 2) Variation between years. This study uses county fixed effects to see how change in the independent policy variables effects exercise as time changes. We also used prohibit, and logit models to see the effect.

One challenge of this study endogeneity of the dependent variable, exercise, and the number of physical fitness resources. It is possible that people who are just more active could choose to live in places where there are a higher number of physical fitness resources. This endogeneity issue is hard to solve but it will be kept in mind going forward.

Another challenge for this study is that the binary exercise variable is a weak variable. It only asks the respondent if he or she has exercised at all in the last 30 days. It doesn't ask how much or for how long. It gives very little information about the type of exercise that person does. The variable, however, is advantageous to use because it is included in a dataset that includes state and county data that is vital to our analysis.

Data Description

This research study uses the Behavioral Risk Surveillance System which is an annual survey administered by the Center of Disease Control (CDC) in order to observe the health and health behaviors of people in the United States population. The Behavioral Risk Surveillance System surveys people in all 50 states, the District of Columbia, Puerto Rico, and Guam, with a

limited amount of data from American Samoa, Federated States of Micronesia, and Palau. Only people who are 18 years or older and non-institutionalized can participate. Some risk factors that are assessed are, “tobacco use, HIV/AIDS knowledge, inadequate sleep, hypertension awareness, cholesterol awareness, chronic health conditions, alcohol consumption, fruits and vegetables consumption, arthritis burden, and seatbelt use.” They also ask about demographic information such as income and age. Starting from 2011, the Behavioral Risk Surveillance System uses both landlines and cellphones in order to collect data. Each year there is a set of questions that everyone must be asked called the “core” questions. These are the basic questions such as age or Tabaco use. In addition to these questions, there are “optional modules” relating to a specific topic such as Mental Health or Arthritis. Optional modules of a certain type are given in different years showing no pattern. The states individually carry out the sampling and surveying the population so the Behavioral Risk Surveillance System has strict guidelines on how the survey is conducted. The states have to be able to justify that the sample is representative of all households in the state which have telephones. The survey is conducted through Computer-Assisted Telephone Interview (CATI) systems and the core questions lasts 18 minutes while an optional module would add 5 to 10 minutes to the time.

One of the biggest sources that is used for the economic variables analysis is the County Business Patterns Dataset which is an annual dataset that collects economic data by industry. They collect information about, “number of establishments, employment during the week of March 12, first quarter payroll, and annual payroll.” The County Business Patterns Dataset collects information by using the Business Register database which contains a record for every known establishment with paid employees in all 50 states and many territories. “An

establishment is a single physical location at which business is conducted or services or industrial operations are performed." This means the database doesn't just keep track of companies; it keeps track of every location for each company. This definition is useful for our analysis, because the physical location of the establishments is important. The County Business Patterns Dataset groups establishments by industry code.

The industry category that is most relevant to this research study is 71: "Arts, Entertainment, and Recreation." Each industry has a 6-digit NAICS code. The County Business Patterns dataset is used to measure the number of fitness centers, parks, golf courses, ski places, marinas, and bowling alleys. A list of the variables that is used in this study with their corresponding NAICS code is listed below.

NAICS Code	Name	Description
713940	Fitness and recreational sports centers	This industry comprises establishments primarily engaged in operating fitness and recreational sports facilities featuring exercise and other active physical fitness conditioning or recreational sports activities, such as swimming, skating, or racquet sports.
712190	Nature parks and other similar institutions	This industry comprises establishments primarily engaged in the preservation and exhibition of natural areas or settings.
713910	Golf courses and country clubs	This industry comprises (1) establishments primarily engaged in operating golf courses (except miniature) and (2) establishments primarily engaged in operating golf courses, along with dining facilities and other recreational facilities that are known as country clubs. These establishments often provide food and beverage services, equipment rental services, and golf instruction services.
713920	Skiing facilities	This industry comprises establishments engaged in (1) operating downhill, cross-country, or related skiing areas and/or (2) operating equipment, such as ski lifts and tows. These establishments often provide food and beverage services, equipment rental services, and ski instruction services. Four-season resorts without accommodations are included in this industry.
713930	Marinas	This industry comprises establishments, commonly known as marinas, engaged in operating docking and/or storage facilities for pleasure craft owners, with or without one or more related activities, such as retailing fuel and marine supplies; and repairing, maintaining, or renting pleasure boats.
713950	Bowling centers	This industry comprises establishments engaged in operating bowling centers. These establishments often provide food and beverage services.
713990	All other amusement and recreation industries	This industry comprises establishments (except amusement parks and arcades; gambling industries; golf courses and country clubs; skiing facilities; marinas; fitness and recreational sports centers; and bowling centers) primarily engaged in providing recreational and amusement services.

Figure 1: Descriptions of physical activity resources used in study.

One of the economic variables discussed earlier that is used in this research study is Violent and Property Crimes. The data that includes this variables is downloaded from The Interuniversity Consortium for Political and Social Research (ICPSR), which is an index that contains many different datasets. The name of the dataset is the "Uniform Crime Reporting Program Data: County-Level Detailed Arrest and Offense Data," which is created by the Federal Bureau of Investigation (FBI) using reports submitted by agencies and states. This dataset contains "county-level counts of arrests and offenses for Part I offenses (murder, rape, robbery, aggravated assault, burglary, larceny, auto theft, and arson) and counts of arrests for Part II offenses (forgery, fraud, embezzlement, vandalism, weapons violations, sex offenses, drug and alcohol abuse violations, gambling, vagrancy, curfew violations, and runaways)." This research study uses the "Crimes Reported" file of the dataset which contains data points for "violent crime" and "property crime." Property crime is defined as "the offenses of burglary, larceny-theft, motor vehicle theft, and arson," while Violent crime is defined as, "murder and nonnegligent manslaughter, forcible rape, robbery, and aggravated assault."

The third dataset that is used is the Local Area Unemployment Statistics (LAUS) which is created by the Bureau of Labor statistics and contains labor force data. This dataset is used to get the economic variable "unemployment rate." The dataset uses information from the Current Employment Statistics (CES) program, State unemployment insurance (UI) systems, decennial census, and annual population estimates to produce its values.

One advantage of the Behavioral Risk Factor Surveillance System is that it is a very large and nationally representative survey. This means that the selection issues are probably very small. The dataset also covers a lot of different topics, so it is unlikely people who are very passionate about exercise said yes to taking this survey more so than would be deemed okay. The fact that the Behavioral Risk Factor Surveillance System has strict standards for how states carry out their sampling is also an advantage.

One advantage of the County Business Patterns dataset is that is covers a large part of the United States Economy, and most of the business units are included in the dataset. However, some establishments are excluded such as: self-employed individuals, employees of private households, railroad employees, agricultural production employees, and most government employees. This is harmful to our analysis because this means government parks are not included in the dataset, and more parks are run by the government. We tried to find a dataset which listed government parks but have not found anything yet. One way to tackle this issue is to find a different way of measuring parks. The amount of money a county spends on employees for their parks could be an indicators on how many parks are in the county. More analysis is needed to solve this issue.

Descriptive Statistics

The first dataset that this research paper includes used data from years 2008-2012. Our datasets included many variables, but the only variables that were included in this research paper are listen in Figure 2.

Description of Variables (Not all included yet).,

Variable Name	Description	Variable Choices	Type	Means
exerany2	Participated in exercise during last month"	0: No 1: Yes	Binary	0: 25.60% 1: 74.40%
income2	Annual household income from all sources:	1: Less than $10,00 2: Less than $15,000 ($10,000 to less than $15,000) 3: Less than $20,000 ($15,000 to less than $20,000) 4: Less than $25,000 ($20,000 to less than $25,000) 5: Less than $35,000 ($25,000 to less than $35,000) 6: Less than $50,000 ($35,000 to less than $50,000) 7: Less than $75,000 ($50,000 to less than $75,000) 8: $75,000 or more	Categorical	1: 5.36% 2: 6.06% 3: 7.82% 4: 9.66% 5: 11.82% 6: 15.01% 7: 16.21% 8: 28.06%
state-county	State and county code	Example “1-13,” 1 being state code, 13 being county code	String	
year	Year of Survey	Years from 2008-2012	Continuous	2008: 18.91% 2009: 19.59% 2010: 20.37% 2011: 21.65% 2012:

				19.48%
marital	Marital Status	1: Married 2: Divorced 3: Widowed 4: Separated 5 Never Married 6: A member of an unmarried couple	Categorical	1: 55.44% 2: 14.77% 3: 12.73% 4: 2.14% 5: 12.53% 6: 2.39%
educa	Highest level of education completed	1: Never attended school or only Kindergarten 2: Grades 1 through 8 (Elementary) 3: Grades 9 through 11 (Some High School) 4: Grade 12 or GED (High School Graduate) 5: College 1 year to 3 years (Some college or technical school) 6: College 4 years or more.	Categorical	1: 0.10% 2: 2.51% 3: 5.58% 4: 28.35% 5: 27.08% 6: 36.38%
employ	Employment Status	1: Employed for wages 2: Self-employed 3: Out of work for more than 1 years 4: Out of work for less than 1 year 5: A homemaker 6: A student 7: Retired 8: Unable to work	Categorical	1: 44.40% 2: 8.30% 3: 2.79% 4: 2.81% 5: 6.56% 6: 1.71% 7: 26.60% 8: 6.83%
sex	Sex	0: Male 1: Female	Binary	0: 39.85 1: 60.15
pregnant	Are you pregnant?	0: No 1: Yes	Binary	0: 96.19 1: 3.81
qlactlm2	Physical Activity Limitation	0: No 1: Yes	Binary	0: 73.81 1: 26.19
_bmi	Body Mass Index	Continuous numbers but with 2 implied decimal places.	Continuous	2913.328
mscode	Metropolitan Status Code	1: In the center city of an MSA 2: Outside the center city of an MSA but inside the	Categorical	1: 33.99% 2: 23.16% 3: 13.75% 4: 0.65%

		county containing the center city 3: Inside a suburban county of the MSA 4: In an MSA that has no center city 5: Not in an MSA		5: 28.46%
_mrace	Non-Hispanic race categories	1: White only 2: Black or African American only 3: Asian Only 4: Native Hawaiian or other Pacific Islander only 5: American Indian or Alaskan Native only 6: Other race only 7: Multiracial	Categorical	1: 83.65 2: 8.75 3: 1.92 4: 0.22 5: 1.30 6: 2.14 7: 2.02
_ageg5yr	Five year age categories	1: Age 18 to 24 2: 25 to 29 3: 30 to 34 4: 35 to 39 5: 40 to 44 6: 45 to 49 7: 50 to 54 8: 55 to 59 9: 60 to 64 10: 65 to 69 11: 70 to 74 12: 75 to 79 13: 80 or older	Categorical	1: 3.19 2: 3.93 3: 5.60 4: 6.75 5: 7.80 6: 9.26 7: 10.98 8: 11.51 9: 11.35 10: 9.45% 11: 7.37% 12: 5.74% 13: 7.06%
_chldcnt	Number of Children	1: No children household 2: One child in household 3: Two children in household 4: Three children in household 5: Four children in household 6: Five or more children in household	Categorical	1: 70.69% 2: 11.66% 3: 11.00% 4: 4.50% 5: 1.49% 6: 0.66%
bpd_fitcenters	**Policy Factor:** Number of	Whole Numbers	Continuous	52.56663

	fitness and recreaitonal sports centers			
bpd_parks	**Policy Factor:** Number of Nature parks and other similar institutions	Whole Numbers	Continuous	.7215446
bpd_golf	**Policy Factor:** Number of Golf clubs and country clubs	Whole Numbers	Continuous	13.28568
bpd_ski	**Policy Factor:** Number of skiing facilities	Whole Numbers	Continuous	.3860345
bdp_bowling	**Policy Factor:** Number of Bowling Centers	Whole Numbers	Continuous	5.090686
bdp_marinas	**Policy Factor:** Number of Marinas	Whole Numbers	Continuous	5.027548
bpd_other	**Policy Factor:** Number of all other amusement and recreaional centers	Whole Numbers	Continuous	22.94331
unemployment_rate	**Policy Factor:** Unemployment Rate	Percentages (example 7%)	Continuous	7.991561
p_crime	**Policy Factor:** Number of property crimes	Whole Numbers	Continuous	15931.82
v_crime	**Policy Factor:** Numer of Violent crimes	Whole Numbers	Continuous	2476.835
landarea	Area in square miles of county	Area in square miles	Continuous	1258.195
land_fitcenters	Bpd_fitcenters divided by area		Continuous	.136975

	of county in square miles			
land_parks	bpd_parks divided by area of county in square miles		Continuous	.0017753
land_golf	bpd_golf divided by area of county in square miles		Continuous	.020926
land_ski	bpd_ski divided by area of county in square miles		Continuous	.0004202
land_bowling	bpd_bowling divided by area of county in square miles		Continuous	.0095934
land_marinas	bpd_marinas divided by area of county in square miles		Continuous	.0114143
land_other	bpd_other divided by area of county in square miles		Continuous	.0461056
land_pcrime	bpd_pcrime divided by area of county in square miles		Continuous	34.80869
land_vcrime	v_crime divided by area of county in square miles		Continuous	6.964426
population	Number of people per county	Whole numbers	Continuous	510622.9
pop_fitcenters	Bpd_fitcenters divided by population of county in number of people			.0001034

pop_parks	bpd_parks divided by population of county in number of people			2.52e-06
pop_golf	bpd_golf divided by population of county in number of people			.0000459
pop_ski	bpd_ski divided by population of county in number of people			2.82e-06
pop_bowling	bpd_bowling divided by population of county in number of people			.000017
pop_marinas	bpd_marinas divided by population of county in number of people			.0000162
pop_other	bpd_other divided by population of county in number of people			.0000629
pop_pcrime	bpd_pcrime divided by population of county in number of people			.0397188
pop_vcrime	v_crime divided by population			.0053804

	of county in number of people			
income3	Broader categories of income compared to income2	1: Less than $15,00 2: Less than $25,000 ($15,000 to less than $25,000) 3: Less than $35,000 ($25,000 to less than $35,000) 4: Less than $50,000 ($35,000 to less than $50,000) 5: Less than $75,000 ($50,000 to less than $75,000) 6: $75,000 or more	Categorical	1: 11.42% 2: 17.48% 3: 11.82% 4: 15.01% 5: 16.21% 6: 28.06%
educa2		2: Grade 8 or less 3: "Grades 9 through 11 (Some High School)" 4: "Grade 12 or GED (High School Graduate)" 5: "College 1 year to 3 years (Some college or technical school)" 6: "College 4 years or more"	Categorical	1: 2.61% 2: 5.58% 3: 28.35% 4: 27.08% 5: 36.38%

Figure 2: Descriptions of all variables and their means.

Results

Analysis of the Effect of Income on Exercise

	(1)	(2)	(3)
	Participated in exercise during last month	Participated in exercise during last month	Participated in exercise during last month
Less than $10,00	0	0	0

	(.)	(.)	(.)
Less than $15,000 ($10,000 to less than $15,000)	0.00190	0.00215	-0.00800***
	(0.00193)	(0.00192)	(0.00204)
Less than $20,000 ($15,000 to less than $20,000)	0.0327***	0.0323***	-0.00166
	(0.00183)	(0.00182)	(0.00196)
Less than $25,000 ($20,000 to less than $25,000)	0.0702***	0.0670***	0.00879***
	(0.00175)	(0.00175)	(0.00192)
Less than $35,000 ($25,000 to less than $35,000)	0.115***	0.109***	0.0305***
	(0.00170)	(0.00170)	(0.00190)
Less than $50,000 ($35,000 to less than $50,000)	0.168***	0.160***	0.0548***
	(0.00164)	(0.00164)	(0.00189)
Less than $75,000 ($50,000 to less than $75,000)	0.217***	0.207***	0.0790***
	(0.00162)	(0.00163)	(0.00193)
$75,000 or more	0.281***	0.267***	0.112***
	(0.00154)	(0.00155)	(0.00195)

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure 3: Three regressions analyzing the effect of income on exercise. (1) Refers to a simple OLS regression with no controls (2) Refers to a Fixed Effects Regression controlling for county and year. (3) Refers to a Fixed Effects Regression controlling for county, year, and many demographic variables. The education control uses narrow education and income categories.

In Figure 3, three regressions were completed to see the effect of income on exercise. The first one was a simple OLS regression with no controls. The second was a fixed effects regression using year and county as controls. The third was a fixed effects regression controlling for county, year, and many demographic variables.

In all three regressions, all coefficients referring to an income greater than 20,000 showed that they were statistically significant. Some coefficients referring to an income lower than 20,000 were insignificant, while other were negative which is not the expected result. The insignificance of the coefficients referring to the lower levels of income could be explained by noticing how the categories are narrower compared to the upper levels of income. These narrow categories mean that the amount of people who fit in those categories are small. The income category for less than $10,000 annual income was only 5.36% of the survey while the income category for greater than $75,000 was 28.07% of the population. As the income categories increase, the range of the categories increases while the people assigned to that income category also increases.

To combat this problem, the categories were rearranged in the regressions of Figure 4. Because the Behavioral Risk Surveillance System asked their questions with those specific categories only, the rearranging was attempted by adding the already existing categories. In the regressions in Figure 4, the income category of less than 10,000 was added to the income category of 10,000-15,000. In addition, the income category of 15,000-20,000 was added to 20,000-25,000.

After examining the controls for Regression 3 in Figure 3, it was observed that the education variable was insignificant in most categories. This could be due to the same problem as the income categories. The education categories include “No school or Kindergarten only” and “Grades 1-8” which is a very small percentage of the population. To combat this problem these two categories were added together. More explanation of the changes in categories for income and education can be seen in Appendix Figure 2.

	(1)	(2)	(3)
	Participated in exercise during last month	Participated in exercise during last month	Participated in exercise during last month
Less than $15,00	0	0	0
	(.)	(.)	(.)
Less than $25,000 ($15,000 to less than $25,000)	0.0525***	0.0503***	0.00840***
	(0.00124)	(0.00124)	(0.00134)
Less than $35,000 ($25,000 to less than $35,000)	0.114***	0.108***	0.0347***
	(0.00135)	(0.00135)	(0.00150)
Less than $50,000 ($35,000 to less than $50,000)	0.167***	0.159***	0.0589***
	(0.00128)	(0.00128)	(0.00149)
Less than $75,000 ($50,000 to less than $75,000)	0.216***	0.206***	0.0831***
	(0.00126)	(0.00127)	(0.00154)

$75,000 or more	0.280***	0.265***	0.116***
	(0.00114)	(0.00117)	(0.00156)

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure 4: Three regressions analyzing the effect of income3 on exercise. These regressions differ from Figure 1 in at the income categories have become more broad. (1) Refers to a simple OLS regression with no controls (2) Refers to a Fixed Effects Regression controlling for county and year. (3) Refers to a Fixed Effects Regression controlling for county, year, and many demographic variables. This regression also differs from Figure 1 in that the education categories have also become broader. Refer to Figure 2 in Appendex to see the change. A more detailed table of these regressions can be found in the Appendix Table 1

In Figure 4, all the coefficients are significant, and all coefficients show the trend that we had hoped. As income increases, exercise increases. Compared to Figure 3, the regressions in Figure 4 also show a greater effect.

Controls Analysis

Referring Table 1 in the Appendix, one can see that as the year increases, people exercise more. This can be attributed to society's trend to be more health-conscious by stressing the importance of exercise. All year coefficients are significant.

Most of the marital coefficients are not significant, which shows that marital status might not have an effect on exercise. It makes sense that marital status would not affect exercise. In future regressions, this variable will not be included.

In the initial regression seen in Figure 3, almost all of the education coefficients were insignificant. This insignificance was attributed to the first two education categories being a

very small percentage of the population, and this attribution caused us to change the categories into a broader one as seen in Appendix Figure 2. After this change, all education coefficients are significant, though the “Some High School” coefficient is only weakly significant. This can be attributed to the same reason mentioned above, very few people fall into this category in today’s age of free and easily available education. All coefficients show that as education increases, exercise increases.

	Participated in exercise during last month
Employed for wages	0
	(.)
Self-employed	0.0224***
	(0.00128)
Out of work for more than 1 years	0.0268***
	(0.00219)
Out of work for less than 1 year	0.0301***
	(0.00215)
A homemaker	0.0453***
	(0.00148)
A student	0.0466***

	(0.00311)
Retired	0.0619***
	(0.00117)
Unable to work	-0.0834***
	(0.00162)

Figure 5: Employment Coefficients referring to Regression 3 of Table 1 in Appendix.

It is interesting to note that all employment coefficients are significant. This is surprising because one does not immediately assume that employment has an effect on exercise. It could be explained that those who are self-employed have more autonomy and that is why they are able to exercise more. Those out of work or a homemaker might have more free time to exercise. This can also be said about students. Even though this regression controls for age, the student coefficient shows that they exercise more which can also be attributed to more free time or more access to sports clubs and teams. It was interesting to note that retired people exercise more when most are of older age. Because this regression controls for age, this result was able to be extracted. It could be said that retired people exercise more because they also have free time. Those unable to work exercise less which makes sense because if someone has a disability, they would be unable to work while also be unable to exercise. One weakness of this variable is that it could correlate with the Physical Activity Limitation variable that is also included in this regression.

It was also significant that females were less likely to exercise. This makes sense because woman might not be as enthused about sports due to gender roles. Though it could also be said that females have a greater pressure to look thin than men so they would exercise more so this result is confusing. Body Mass Index shows a significant weak effect. This variable is one of the only continuous variables and each data point includes the 2 decimal digits in it which explains the tiny coefficient.

Almost all the Metropolitan Status Code coefficients are insignificant so whether or not someone lives in a city or a suburb doesn't affect whether or not they exercise. However, the "In an MSA that has no center city" coefficient shows a negative effect and is significant which is what we expected. The people in this category live in more rural counties than the ones in the other coefficients so maybe they do not have as much access to physical activity resources. The insignificance of this effect could be attributed to how we are already controlling for County in this regression. This variable will be taken out in subsequent regressions because it does not have much of a significant effect.

In the center city of an MSA	0
	(.)
Outside the center city of an MSA but inside the county containing	-0.00337***

the center city	
	(0.000908)
Inside a suburban county of the MSA	-0.0173***
	(0.00108)
In an MSA that has no center city	-0.0152***
	(0.00421)
Not in an MSA	-0.0102***
	(0.000865)

Figure 6: Coefficients of Metropolitan Status Code (MSA) variable in a regression that does not control for county.

When the county control was taken out, the Metropolitan Status Code was significant and gave us the negative effect we were hoping for. This shows that these to variables are correlated.

Almost all the race categories were significant and showed that most non-white races exercise less. American Indian or Alaskan Native only was not significant but can be explained due to the small number of people who fit into this category. This is interesting because we are controlling for income as well so race by itself has an effect. Multiracial has a positive effect which is not what we were expecting. All the age categories were also significant and that as age increases, exercise decreases.

No children household			0
			(.)

One child in household			-0.0135***
			(0.00120)
Two children in household			-0.0121***
			(0.00133)
Three children in household			-0.0170***
			(0.00185)
Four children in household			-0.0279***
			(0.00296)
Five or more children in household			-0.0346***
			(0.00432)

Figure 7: Coefficients of Number of Children Variables referring to Regression 3 of Table 1 in Appendix.

All children coefficients are negative and significant which shows that compared to having no children, having any children causes one to exercise less. This can be attributed to lack of time. The negative effect of one child is more than the negative effect of two children which is interesting.

Regression 4 of Table 1 of the Appendix shows the results after taking out the control variables that were deemed unnecessary during this analysis. This includes Metropolitan Status Code and Marital Status. The results do not change much compared to Regression 3 but hopefully taking

these bad controls out gives us more accurate information about the effect of income on exercise.

Probit and Logit Analysis

Probit and Logit models were also considered because the outcome variable was binary.

	(Probit)	(Probit with Controls)	(Logit)	(Logit with Controls)
	Participated in exercise during last month	Participated in exercise during last month	Participated in exercise during last month	Participated in exercise during last month
Participated in exercise during last month				
Less than $15,00	0	0	0	0
	(.)	(.)	(.)	(.)
Less than $25,000 ($15,000 to less than $25,000)	0.137***	0.0162***	0.221***	0.0243***
	(0.00370)	(0.00418)	(0.00597)	(0.00686)
Less than $35,000 ($25,000 to less than $35,000)	0.305***	0.0895***	0.496***	0.143***
	(0.00410)	(0.00472)	(0.00668)	(0.00780)
Less than $50,000	0.465***	0.165***	0.763***	0.269***

($35,000 to less than $50,000)				
	(0.00393)	(0.00468)	(0.00648)	(0.00779)
Less than $75,000 ($50,000 to less than $75,000)	0.630***	0.254***	1.047***	0.423***
	(0.00395)	(0.00484)	(0.00661)	(0.00816)
$75,000 or more	0.885***	0.410***	1.507***	0.708***
	(0.00365)	(0.00482)	(0.00623)	(0.00823)

Figure 8: Probit and Logit models of income on exercise. More information can be seen in Table 3 of the Appendix.

In this regression, mscode was kept as a control. This is because we could not control for county because we could not use the absorb command with probit. The mscode variable controls for location-specific effects.

Both the simple Probit model and the Probit model with controls, showed that as income increases, the z-score of exercising increases. The logit regression shows similar results; as income increases, the log odds of exercising increases.

Policy Factor Analysis

```
. correlate bpd_fitcenters bpd_parks bpd_golf bpd_ski bdp_bowling bdp_marinas bpd_other
(obs=1704104)
```

	bpd_fi~s	bpd_pa~s	bpd_golf	bpd_ski	bdp_bo~g	bdp_ma~s	bpd_ot~r
bpd_fitcen~s	1.0000						
bpd_parks	0.5749	1.0000					
bpd_golf	0.8330	0.5135	1.0000				
bpd_ski	0.2950	0.1692	0.2600	1.0000			
bdp_bowling	0.8311	0.3857	0.7754	0.2848	1.0000		
bdp_marinas	0.5340	0.3806	0.5170	0.1115	0.4493	1.0000	
bpd_other	0.9606	0.5833	0.8502	0.3225	0.8252	0.5748	1.0000

Figure 9: Correlation Matrix for Policy Factors.

The policy factors that we will be examining include access to physical activity resources such as number of fitness centers and bowling areas. Others include unemployment rate, and violent and property crimes. In the table above, one can see that many of the physical activity resources are correlated so it might not be a good idea to throw all of these similar variables together in a regression. Instead, we will be examining the effect of these variable by themselves. Property and violent crime are also severely correlated so these will also be added separately.

Figure 10 refers to a regression of the broad income variable, year, the broad education category, employment, sex, physical activity limitation, BMI, race, age, children and one policy factor at a time on exercise. The results are according to which policy factor was included.

Name of Variable	bpd_fitcenters	bpd_parks	bpd_golf	bpd_ski	bdp_bowling	bdp_marinas
Estimate	.0000558***	.0057888***	.0002709***	.0117568***	.0004481***	.0002972***
Standard Error	3.71e-06	.0002467	.0000183	.0003637	.0000427	.0000306
Mean	52.56663	.7215446	13.28568	.3860345	5.090686	5.027548

Name of Variable	bpd_other	unemployment_rate	p_crime	v_crime
Estimate	.0002052***	-.0003536	1.27e-07***	3.99e-07*
Standard Error	9.28e-06	.0001372	1.10e-08	5.73e-08
Mean	22.94331	7.991561	15931.82	2476.835

$^{*}\ p < 0.05,\ ^{**}\ p < 0.01,\ ^{***}\ p < 0.001$

Figure 10: Regression of the broad income variable, year, the broad education category, employment, sex, physical activity limitation, BMI, race, age, children and one policy factor on exercise. The results are according to which policy factor was included.

In this regression, the only control that was left out was our County Fixed Effects. All the physical activity resource variables are significant but that might change when County Fixed Effects are added. The unemployment variable was not significant which makes sense because in our literature review there were conflicting findings on this variable. Property and Violent Crimes were significant, but did not give us the coefficient we expected nor did it give us a large effect. It was expected that property and violent crime would have a negative effect on exercise. All the coefficients, regardless of significance, show very small effects when comparing each coefficient to the mean.

Now we will conduct the same one variable at a time analysis, but we will absorb our county variable.

Name of Variable	bpd_fitcenters	bpd_parks	bpd_golf	bpd_ski	bdp_bowling	bdp_marinas	bpd_other
Estimate	-.0001765*	.0012089*	-.0006711*	-.0003475	-.0008024	-2.80e-06	.0001093
Standard Error	.0000817	.0005215	.0002863	.0016866	.000431	.0003881	.0001327
Mean	52.56663	.7215446	13.28568	.3860345	5.090686	5.027548	22.94331

Name of Variable	unemployment_rate	p_crime	v_crime
Estimate	.0008863*	-6.74e-08	-5.15e-07*
Standard Error	.0003047	4.35e-08	2.39e-07
Mean	7.991561	15931.82	2476.835

$^{*}\ p < 0.05,\ ^{**}\ p < 0.01,\ ^{***}\ p < 0.001$

Figure 11: Regression of the broad income variable, year, the broad education category, employment, sex, physical activity limitation, BMI, race, age, children, county and one policy factor on exercise. This regression differs from Figure 10 in that it controls for county. The results are according to which policy factor was included.

After controlling for County fixed effects, most physical activity resource coefficients because insignificant or very weakly significant. Unemployment rate is the only variable that shows a different reaction; the unemployment rate become significant after controlling for county. One

interesting thing to note is that property and violent crime become negative after controlling for county. Property crime is insignificant while violent crime is only weakly significant, but both variables show a negative effect which was expected. This effect is very small so it is not an acceptable result.

According to this analysis, the physical activity coefficients are only significant when we are not controlling for County Fixed Effects, and even in this case the result is very small. For example, referring to Figure 10 for every 1 fitness center added, exercise increases only by .0000558.

Once County Fixed effects are applied, the significance disappears. Unemployment rate, property crime, and violent crime also show insignificant or very small effects.

Because the physical activity resources are correlated to one another and the two types of crimes are correlated with each other, we decided to run a regression with the most significance of these variables. Parks are the most significant and greatest in magnitude coefficient for the physical activity recourses. Violent Crimes are more significant than Property Crimes in the regression with County is absorbed.

We decided to run a regression on exercise with all the controls including county, Parks, Unemployment Rate, and Violent Crimes.

Name of Variable	Bpd_parks	Unemployment Rate	Violent Crimes
Estimate	.0010759*	.0009173***	-5.21e-07*
Standard Error	.0005237	.0003055	2.41e-07
Mean	.7215446	7.991561	2476.835

$^{*}\ p < 0.05,\ ^{**}\ p < 0.01,\ ^{***}\ p < 0.001$

Figure 12: Regression of the broad income variable, year, the broad education category, employment, sex, physical activity limitation, BMI, race, age, children, number of parks, unemployment rate, and number of violent crimes.

It is interesting to note that when Parks, Unemployment Rate, Violent crimes are all thrown in the same regression, all three are significant (though parks and violent crimes are weakly significant). Parks has a positive sign as we would expect, and violent crimes has a negative sign which is what we would expect. All three are very small in magnitude which again, alludes to the idea that these variable do not have an effect on exercise.

In our next analysis, the variables were scaled by both land and population. The only variable that was not scaled was unemployment rate because that variable is not affected by the population or land size. This is also the only variable that is a percentage and not a quantity. The physical activity resources tell us how many of the resources there are. The crimes variables also tell us how many crimes have been committed.

First we will observe the coefficients for these policy factors for land. Here we took the initial variables and divided it by land size. The land size is measured in square miles.

Variable	Obs	Mean	Std. Dev.	Min	Max
land_fitce~s	1704104	.136975	.7889376	0	19.6671
land_parks	1704104	.0017753	.0127095	0	.3504161
land_golf	1704104	.020926	.0290511	0	.5256242
land_ski	1704104	.0004202	.0015041	0	.087604
land_bowling	1704104	.0095934	.0221984	0	.4380201
land_marinas	1704104	.0114143	.0328771	0	.3725166
land_other	1704104	.0461056	.170697	0	4.862024
land_pcrime	1704104	34.80869	102.0429	0	8909.756
land_vcrime	1704104	6.964426	26.93216	0	764.3015

$^{*}\ p < 0.05, {}^{**}\ p < 0.01, {}^{***}\ p < 0.001$

Figure 13: Descriptive Statistics of all land variables.

This table shows the range of values the scaled by land variables show. The range varies a lot. The max of property crime is 8909.756. This could be explained because smaller counties could have more crimes due to overcrowding. This could be possible for the large fitness center variable as well. A more overcrowded area could have more fitness centers to meet the demand.

Now we will conduct a regression for each of these variables. The table below refers to a regression of the broad income variable, year, the broad education variable, employment, sex,

physical activity limitation, bmi, race, age, number of children and one policy factor on exerany2. There are the results according to which policy factor was included.

Name of Variable	land_fitcenters	land_parks	land_golf	land_ski	land_bowling	land_marinas
Estimate	.0014738***	.1645925***	.0259518*	3.806202***	.0013469	0214321*
Standard Error	.0004208	.0262511	.011426	.2190254	.0149375	.0100062
Mean	.136975	.0017753	.020926	.0004202	.0095934	.0114143

Name of Variable	land_other	land_pcrime	land_vcrime
Estimate	.0100187***	.0000125***	.0000308*
Standard Error	.0019525	3.27e-06	.0000124
Mean	.0461056	34.80869	6.964426

$^{*}\ p < 0.05,\ ^{**}\ p < 0.01,\ ^{***}\ p < 0.001$

Figure 14: Regression of the broad income variable, year, the broad education category, employment, sex, physical activity limitation, BMI, race, age, children and one policy factor on exercise. These policy factors are scaled by land. The results are according to which policy factor was included.

All the coefficients are significant except for the bowling variable. All the physical activity resource variables show a positive effect which is what we expected. Property and violent crimes coefficients have positive coefficients which is not what we expected but the results are significant.

Now we will redo the above regression but control for County Fixed Effects by absorbing the county variable to see if changes arise.

Name of Variable	land_fitcenters	land_parks	land_golf	land_ski	land_bowling	land_marinas
Estimate	.0039216	-.0014225	.0306697	-.1362557	-.1360675	-.0412551
Standard Error	.0174922	.0536033	.0735253	.3153195	.1255251	.0800891
Mean	.136975	.0017753	.020926	.0004202	.0095934	.0114143

Name of Variable	land_other	land_pcrime	land_vcrime
Estimate	-.0133935	1.52e-06	.0000107
Standard Error	.0193216	7.28e-06	.0000414
Mean	.0461056	34.80869	6.964426

$^{*}\ p < 0.05,\ ^{**}\ p < 0.01,\ ^{***}\ p < 0.001$

Figure 15: Regression of the broad income variable, year, the broad education category, employment, sex, physical activity limitation, BMI, race, age, children, county and one policy factor on exercise. These policy factors are scaled by land. This regression differs from Figure 14 in that it controls for county. The results are according to which policy factor was included.

After controlling for County Fixed Effects, not a single coefficient is significant.

Next, we try this same analysis but with scaling for population.

Variable	Obs	Mean	Std. Dev.	Min	Max
pop_fitcen~s	1704104	.0001034	.0000503	0	.0007236
pop_parks	1704104	2.52e-06	7.81e-06	0	.0001279
pop_golf	1704104	.0000459	.0000395	0	.0004237
pop_ski	1704104	2.82e-06	.0000122	0	.0002922
pop_bowling	1704104	.000017	.0000192	0	.0002791
pop_marinas	1704104	.0000162	.0000416	0	.000794
pop_other	1704104	.0000629	.0000824	0	.0019589
pop_pcrime	1704104	.0397188	.1472056	0	10.7529
pop_vcrime	1704104	.0053804	.0261218	0	1.816754

$^{*}\ p < 0.05, ^{**}\ p < 0.01, ^{***}\ p < 0.001$

Figure 16: Descriptive Statistics of all land variables.

The coefficients' range is a lot smaller compared to Figure 13 where we looked at the coefficients when scaled for land.

This simple regression between exercise and education also shows the same conclusion that the previous regression did. As education increases, exercise increases. In many of the articles in my literature review, some scholars used education as a measure of income so it is interesting to see that the same effect is present here.

Name of Variable	pop_fitcenters	pop_parks	pop_golf	pop_ski	pop_bowling	pop_marinas
Estimate	153.1048***	685.6554***	35.70522***	616.6161***	4.33173	58.78882***
Standard Error	6.515305	41.80137	8.236573	26.35481	16.99749	7.741933
Mean	.0001034	2.52e-06	.0000459	2.82e-06	.000017	.0000162

Name of Variable	pop_other	pop_pcrime	pop_vcrime
Estimate	111.1765***	-.0034324	-.0075333
Standard Error	3.925111	.0022492	.0126096
Mean	.0000629	.0397188	.0053804

$^{*}\ p < 0.05,\ ^{**}\ p < 0.01,\ ^{***}\ p < 0.001$

Figure 17: Regression of the broad income variable, year, the broad education category, employment, sex, physical activity limitation, BMI, race, age, children and one policy factor on exercise. These policy factors are scaled by population. The results are according to which policy factor was included.

Almost all the Physical Activity resources are significant except for bowling. They all show a positive correlation which is to be expected. The coefficients are a lot higher than the other regressions. This can be attributed to how dividing each original physical activity resource number by the population makes the variable very, very small. So a 1 unit increase in physical fitness resource/population is unlikely. Property crimes and violent crimes are insignificant.

Name of Variable	pop_fitcenters	pop_parks	pop_golf	pop_ski	pop_bowling	pop_marinas
Estimate	-61.44363	-12.10728	-74.30145	-133.7525	-21.20523	70.22061
Standard Error	21.83076	81.80916	46.46684	148.4145	61.18663	65.63275
Mean	.0001034	2.52e-06	.0000459	2.82e-06	.000017	.0000162

Name of Variable	pop_other	pop_pcrime	pop_vcrime
Estimate	2.313975	-.0011096	.0035046
Standard Error	24.72467	.0025369	.0142214
Mean	.0000629	.0397188	.0053804

Figure 18: Regression of the broad income variable, year, the broad education category, employment, sex, physical activity limitation, BMI, race, age, children, county and one policy factor on exercise. These policy factors are scaled by population. This regression differs from Figure 17 in that it controls for county. The results are according to which policy factor was included.

All the coefficients are insignificant after controlling for County Fixed Effects. This could be attributed to how there is little variation year to year per county. In the following table, the number of physical activity resource in 2012 was subtracted by the number of physical activity resources in 2008 to show the change in the physical activity resource. Observing the table, one can see that the change is very small, and that is why our county effects are not being controlled effectively.

bpd_fitcenters	bpd_parks	bpd_golf	bpd_ski	bdp_bowling	bdp_marinas	bpd_other
-.0778247	-.0037506	-.0646976	-.0014065	.0421941	.0375059	-.0661041

Figure 21: The means of the change of each physical activity coefficient from 2008 to 2012.

Probit and Logit Analysis of Policy Factors

A probit model was used to verify the results bought upon our Fixed Effects Regression in the previous section. The table below shows a batch of regressions where each policy factor variable was inserted separately. The regression included all controls (except mscode and marital) and one policy factor.

bpd_fitcenters	bpd_parks	bpd_golf	bpd_ski	bdp_bowling	bdp_marinas	bpd_other
.0002008***	.0215336***	.0009582**	.0438742***	.0015631**	.0010692**	.0007435***
.0000132	.0008849	.0000645	.0013097	.0001489	.0001083	.0000331

unemployment_rate	p_crime	v_crime
-.0014268***	4.34e-07***	1.30e-06***
.0004681	3.87e-08	2.00e-07

Figure 20: Probit regression of the broad income variable, year, the broad education category, employment, sex, physical activity limitation, BMI, race, age, children, county and one policy factor on exercise. These policy factors are not scaled. Results are according to which policy factor was included.

The following is a logit regression with the same controls as the probit regression. This is a logit regression on exercise with one policy factor and all controls (except mscode and marital) being on the right hand side. The table shows the coefficient of the one policy factor that was included.

bpd_fitcenters	bpd_parks	bpd_golf	bpd_ski	bdp_bowling	bdp_marinas	bpd_other
.0003477***	.037506***	.0016451***	.0756012***	.0026824***	.0018567***	.001293***
.000023	.0015517	.0001122	.0022943	.0002578	.0001895	.0000583

unemployment_rate	p_crime	v_crime
-.0023393***	7.50e-07***	2.27e-06 ***
.0007993	6.72e-08	3.45e-07

Figure 21: Logit regression of the broad income variable, year, the broad education category, employment, sex, physical activity limitation, BMI, race, age, children, county and one policy factor on exercise. These policy factors are not scaled. Results are according to which policy factor was included.

All the coefficients in the probit and logit model are significant. All the physical activity resource coefficients are positive, though the effect is still very small in these regressions as well. Unemployment rate, property crimes, and violent crimes show very small coefficient, but their results are also significant.

Conclusion

During the first part of our analysis where we analyzed the effect of income on exercise, we obtained significant and useful results that show that income does affect exercise. For the $75,000 dollars or up category, the probability of a person exercising in the last 30 days increases by approximately 11% even when controls are added to the regression. This is a pretty large result especially when the mean of our exercise variable is 75%. This result gives policy makers a reason to keep lower-income people in mind when making decisions on how to improve the population's health. Policy makers should think of ways to incentivize lower-income people to go to GYMs, join sports clubs, and get more active.

The controls part of our analysis also showed some interesting results. Education, even when controlling for income, still shows an effect on exercise. Educating people can also have an effect on getting people to be healthier. Another interesting finding is that even when controlling for income, racial minorities still exercise less. Focusing on these minority groups can also help increase the health of the population.

The policy factor section of our study showed that there is a small, significant effect of policy factors on exercise. However when we control for county, that significance goes away. This shows that there is a significant effect between counties, but there it is not significant effect within counties.

One of the limitations of the research study is that our county variables do not have a large amount of variation which could explain why the significance goes away when we control for county. This can be seen in Figure 21 in the results section. This presents a problem of omitted

variable bias because we are not capturing the County Fixed Effects correctly. Future studies should try to lengthen the time period that is observed. A lengthier time period could show more variation over time about the increase and decrease of the policy variables.

Another weakness of this research study is that our results can also be explained using the endogeneity theory. It is possible that people who like to exercise, choose to live in a place that gives them access to a lot of physical activity resources. This weakness can be countered by saying that most people do not consider physical activity resources when choosing a place to live. It is possible that people who are very enthusiastic about exercise might, but a normal person would not. This research study examines over a million people's responses throughout the United States, and so this research study does capture the effect of an average person's thoughts about exercise.

The exercise variable that was used for this analysis measured if a person exercised in the last 30 days. This is a weak variable because it does not give a lot of information about a person's exercise behavior. Future studies should try to find a better dependent variable that could give us more information such as number of times exercised in the last month or minutes spent exercising or the difficulty of the exercise.

One strength of this research study is the large sample size which helps lower selection bias in the survey. The large number of variables available to us also allows us to control for many different variables that might affect out dependent variable. The perfect dataset would allow us to keep these strengths while also allowing us to have a stronger exercise variable. One limitation of finding a better dataset was that many surveys and data collection techniques do

not include the respondent's county because they want their participants to have their anonymity. This makes conducting this type of research study difficult.

This research study was able to add to the literature on how income effects exercise. It also added a lot of information about controls such as education, employment, number of children etc. The policy factor analysis shows that there is a small positive effect of the number of physical fitness resources on exercise. In future studies, this small effect should be studied using more detailed exercise variables.

References

Adler, N. (1993). Socioeconomic Inequalities in Health. JAMA, 269(24), p.3140.

Baruth, M., Sharpe, P., Parra-Medina, D. and Wilcox, S. (2014). Perceived Barriers to Exercise and Healthy Eating Among Women from Disadvantaged Neighborhoods: Results from a Focus Groups Assessment. Women & Health, 54(4), pp.336-353.

Bureau of Labor Statistics. *Local Area Unemployment Statistics.* The United States Department of Labor, [2008-2012].

Centers for Disease Control and Prevention (CDC). *Behavioral Risk Factor Surveillance System Survey Data.* Atlanta, Georgia: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, [2008-2012]

Droomers, M., Schrijvers, C., van de Mheen, H. and Mackenbach, J. (1998). Educational differences in leisure-time physical inactivity: a descriptive and explanatory study. Social Science & Medicine, 47(11), pp.1665-1676.

Fahlman, M., Hall, H. and Lock, R. (2006). Ethnic and Socioeconomic Comparisons of Fitness, Activity Levels, and Barriers to Exercise in High School Females. Journal of School Health, 76(1), pp.12-17.

Hino, A., Reis, R., Sarmiento, O., Parra, D. and Brownson, R. (2011). The built environment and recreational physical activity among adults in Curitiba, Brazil. Preventive Medicine, 52(6), pp.419-422.

Lantz, P., House, J., Lepkowski, J., Williams, D., Mero, R. and Chen, J. (1998). Socioeconomic Factors, Health Behaviors, and Mortality. JAMA, 279(21), p.1703.

Lee, R., Cubbin, C. and Winkleby, M. (2007). Contribution of neighbourhood socioeconomic status and physical activity resources to physical activity among women. Journal of Epidemiology & Community Health, 61(10), pp.882-890.

Lynch, J., Kaplan, G. and Salonen, J. (1997). Why do poor people behave poorly? Variation in adult health behaviours and psychosocial characteristics by stages of the socioeconomic lifecourse.Social Science & Medicine, 44(6), pp.809-819.

McInnes, M. and Shinogle, J. (2009). Physical Activity: Economic and Policy Factors. NBER Working Paper Series. Cambridge, MA: National Bureau of Economic Research.

Murray, T., Rodgers, W. and Fraser, S. (2011). Exploring the relationship between socioeconomic status, control beliefs and exercise behavior: a multiple mediator model. J Behav Med, 35(1), pp.63-73.

Powell, L., Slater, S., Chaloupka, F. and Harper, D. (2006). Availability of Physical Activity–Related Facilities and Neighborhood Demographic and Socioeconomic Characteristics: A National Study.Am J Public Health, 96(9), pp.1676-1680.

Ross, C. (2000). Walking, exercising, and smoking: does neighborhood matter?. Social Science & Medicine, 51(2), pp.265-274.

Rutt, C. and Coleman, K. (2005). Examining the relationships among built environment, physical activity, and body mass index in El Paso, TX. Preventive Medicine, 40(6), pp.831-841.

United States Census Bureau. *County Business Patterns*, United States Census Bureau [2008-2012]

United States Department of Justice. Federal Bureau of Investigation. Uniform Crime Reporting Program Data: County-Level Detailed Arrest and Offense Data, [2008-2012]. ICPSR35019-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2014-06-12.

Appendix

Figure 1:

Appendix (1) Refers to a simple OLS regression with no controls (2) Refers to a Fixed Effects Regression controlling for county and year. (3) Refers to a Fixed Effects Regression controlling for county, year, and many demographic variables. (4) Refers to a Fixed Effects Regression controlling for county, year, and many demographic variables after controls analysis was done. Marital (Marital Status) and Mscode (Metropolitan Status Code) were taken out. Marital was insignificant and Mscode correlated with County control.

All regressions use broader income and education categories.

	(1)	(2)	(3)	(4)
	Participated in exercise during last month	Participated in exercise during last month	Participated in exercise during last month	Participated in exercise during last month
Less than $15,00	0	0	0	0
	(.)	(.)	(.)	(.)
Less than $25,000 ($15,000 to less than $25,000)	0.0525***	0.0503***	0.00840***	0.00764***
	(0.00124)	(0.00124)	(0.00134)	(0.00127)
Less than $35,000 ($25,000 to less than $35,000)	0.114***	0.108***	0.0347***	0.0334***
	(0.00135)	(0.00135)	(0.00150)	(0.00142)
Less than $50,000 ($35,000 to less than $50,000)	0.167***	0.159***	0.0589***	0.0574***
	(0.00128)	(0.00128)	(0.00149)	(0.00139)
Less than $75,000 ($50,000 to less than $75,000)	0.216***	0.206***	0.0831***	0.0816***
	(0.00126)	(0.00127)	(0.00154)	(0.00141)
$75,000 or more	0.280***	0.265***	0.116***	0.115***
	(0.00114)	(0.00117)	(0.00156)	(0.00139)
Year=2008		0	0	0
		(.)	(.)	(.)
Year=2009		0.00522***	0.00301**	0.00332**
		(0.00105)	(0.00104)	(0.00103)
Year=2010		0.00625***	0.00821***	0.00831***
		(0.00105)	(0.00103)	(0.00103)
Year=2011		0.0115***	0.0104***	0.00870***
		(0.00103)	(0.00107)	(0.00103)
Year=2012		0.0311***	0.0238***	0.0248***

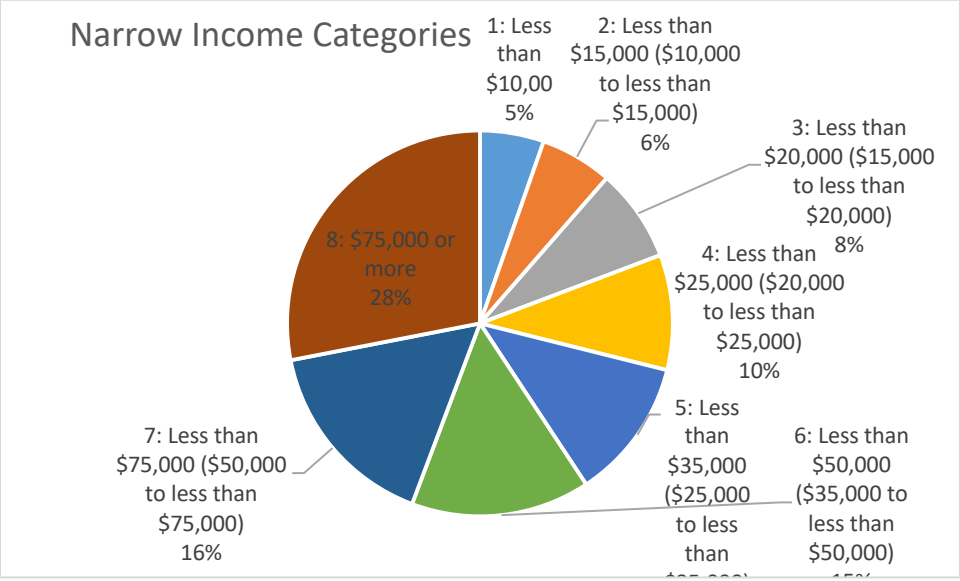
		(0.00106)	(0.00113)	(0.00106)
Married			0	
			(.)	
Divorced			0.00191	
			(0.00107)	
Widowed			0.00128	
			(0.00122)	
Separated			-0.00422	
			(0.00245)	
Never Married			-0.00291*	
			(0.00127)	
A member of an unmarried couple			0.00999***	
			(0.00237)	
Grade 8 or Less			0	0
			(.)	(.)
Grades 9 through 11 (Some High School)			0.00518*	0.00553*
			(0.00257)	(0.00248)
Grade 12 or GED (High School Graduate)			0.0516***	0.0530***
			(0.00226)	(0.00219)
College 1 year to 3 years (Some college or technical school)			0.114***	0.115***
			(0.00230)	(0.00222)
College 4 years or more			0.167***	0.168***
			(0.00233)	(0.00225)
Employed for wages			0	0
			(.)	(.)
Self-employed			0.0224***	0.0214***
			(0.00128)	(0.00123)
Out of work for more than 1 years			0.0268***	0.0254***
			(0.00219)	(0.00207)
Out of work for less than 1 year			0.0301***	0.0296***
			(0.00215)	(0.00202)
A homemaker			0.0453***	0.0445***
			(0.00148)	(0.00143)
A student			0.0466***	0.0446***
			(0.00311)	(0.00272)
Retired			0.0619***	0.0609***
			(0.00117)	(0.00115)
Unable to work			-0.0834***	-0.0849***

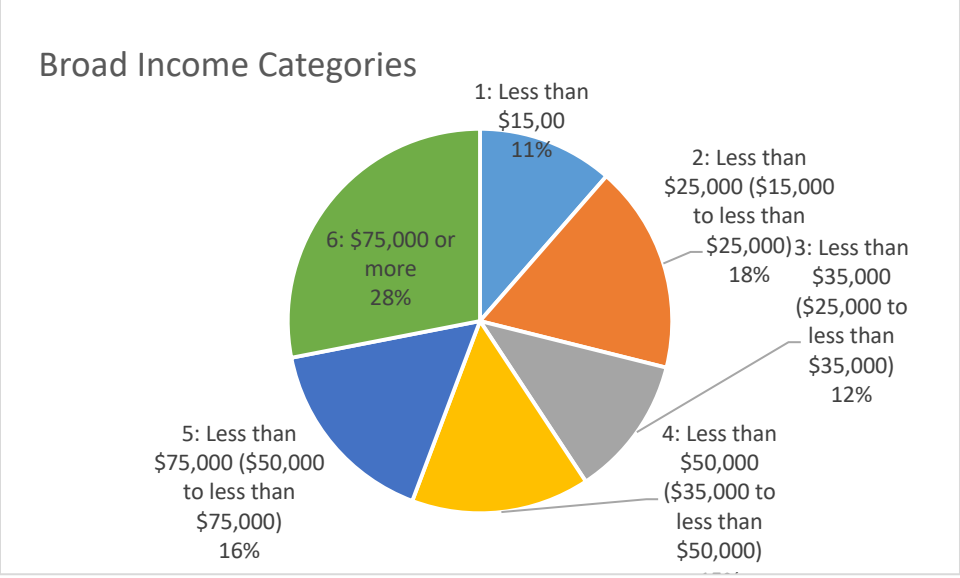
			(0.00162)	(0.00155)
Male			0	0
			(.)	(.)
Female			-0.0173***	-0.0172***
			(0.000721)	(0.000681)
Does NOT have a Physical Activity Limitation			0	0
			(.)	(.)
Physical Activity Limitation			-0.127***	-0.125***
			(0.000840)	(0.000814)
Calculated Body Mass Index			-0.0000259***	-0.0000262***
			(0.000000298)	(0.000000293)
In the center city of an MSA			0	
			(.)	
Outside the center city of an MSA but inside the county containing the center city			-0.00126	
			(0.00106)	
Inside a suburban county of the MSA			0.000794	
			(0.00273)	
In an MSA that has no center city			-0.0647**	
			(0.0202)	
Not in an MSA			-0.00425	
			(0.00311)	
White only			0	0
			(.)	(.)
Black or African American only			-0.0277***	-0.0284***
			(0.00138)	(0.00131)
Asian Only			-0.0705***	-0.0725***
			(0.00272)	(0.00255)
Native Hawaiian or other Pacific Islander only			-0.0339***	-0.0335***
			(0.00761)	(0.00700)
American Indian or Alaskan Native only			-0.00588	-0.00487
			(0.00311)	(0.00295)

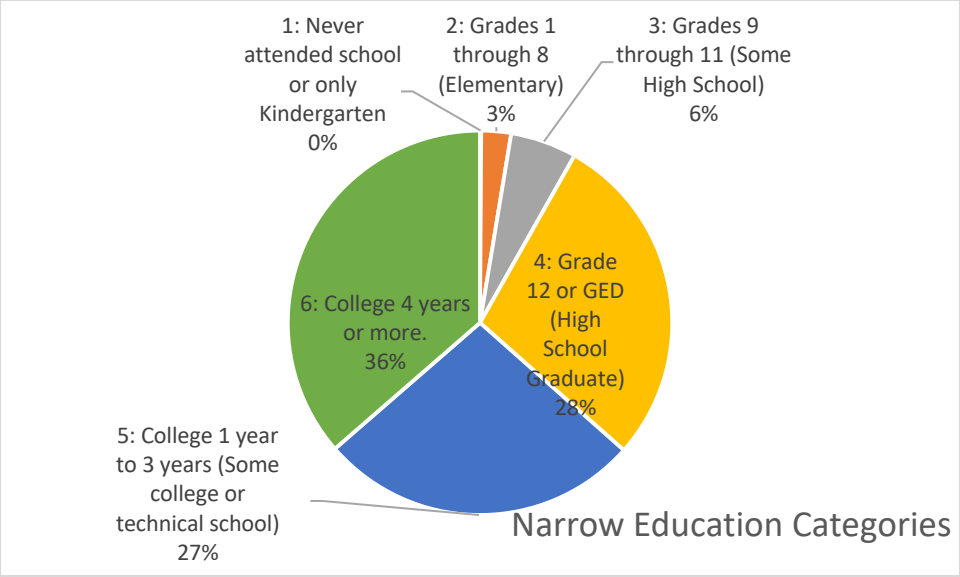
Other race only			-0.0269***	-0.0294***
			(0.00250)	(0.00233)
Multiracial			0.0146***	0.0145***
			(0.00254)	(0.00239)
Age 18 to 24			0	0
			(.)	(.)
25 to 29			-0.0329***	-0.0325***
			(0.00301)	(0.00253)
30 to 34			-0.0514***	-0.0525***
			(0.00285)	(0.00241)
35 to 39			-0.0629***	-0.0646***
			(0.00280)	(0.00237)
40 to 44			-0.0684***	-0.0713***
			(0.00275)	(0.00231)
45 to 49			-0.0706***	-0.0737***
			(0.00270)	(0.00225)
50 to 54			-0.0795***	-0.0825***
			(0.00270)	(0.00222)
55 to 59			-0.0926***	-0.0952***
			(0.00273)	(0.00224)
60 to 64			-0.107***	-0.110***
			(0.00277)	(0.00227)
65 to 69			-0.117***	-0.119***
			(0.00287)	(0.00238)
70 to 74			-0.135***	-0.138***
			(0.00298)	(0.00248)
75 to 79			-0.167***	-0.169***
			(0.00308)	(0.00258)
80 or older			-0.215***	-0.216***
			(0.00308)	(0.00254)
No children household			0	0
			(.)	(.)
One child in household			-0.0135***	-0.0133***
			(0.00120)	(0.00112)
Two children in household			-0.0121***	-0.0125***
			(0.00133)	(0.00123)
Three children in household			-0.0170***	-0.0176***
			(0.00185)	(0.00173)
Four children in household			-0.0279***	-0.0268***
			(0.00296)	(0.00279)
Five or more children in household			-0.0346***	-0.0367***

			(0.00432)	(0.00407)
Intercept	0.583***	0.580***	0.778***	0.780***
	(0.000964)	(0.00119)	(0.00393)	(0.00325)
Observations	1704104	1704104	1522776	1634956

Figure 2: Broad VS Narrow Categories for Income and Education Variable







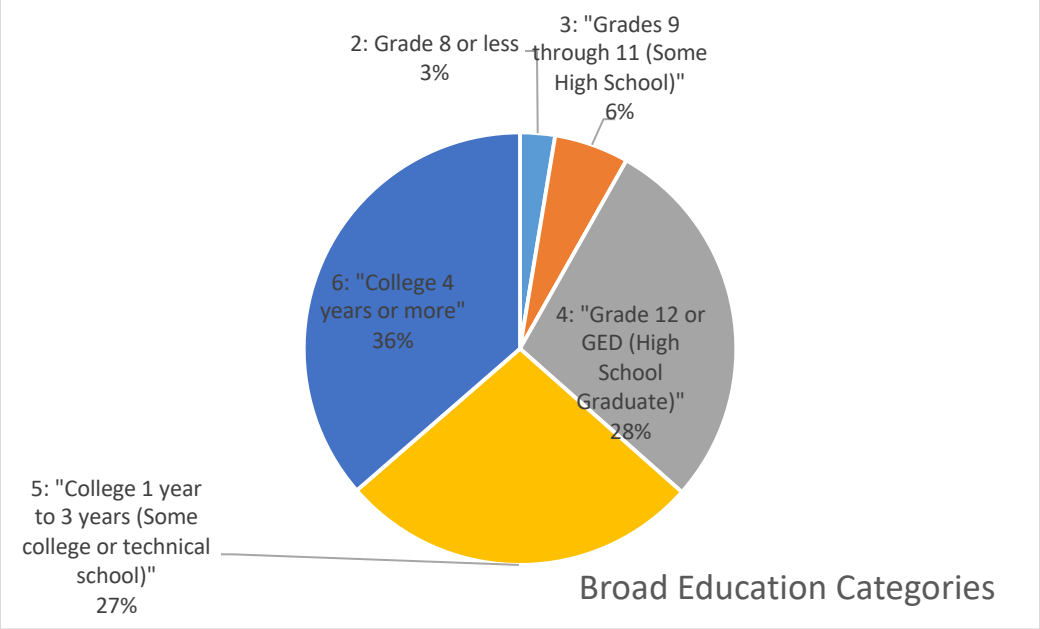


Figure 3: Probit and Logistic Models

	(Probit)	(Probit with Controls)	(Logit)	(Logit with Controls)
	Participated in exercise during last month	Participated in exercise during last month	Participated in exercise during last month	Participated in exercise during last month
Participated in exercise during last month				
Less than $15,00	0	0	0	0
	(.)	(.)	(.)	(.)
Less than $25,000	0.137***	0.0162***	0.221***	0.0243***

($15,000 to less than $25,000)				
	(0.00370)	(0.00418)	(0.00597)	(0.00686)
Less than $35,000 ($25,000 to less than $35,000)	0.305***	0.0895***	0.496***	0.143***
	(0.00410)	(0.00472)	(0.00668)	(0.00780)
Less than $50,000 ($35,000 to less than $50,000)	0.465***	0.165***	0.763***	0.269***
	(0.00393)	(0.00468)	(0.00648)	(0.00779)
Less than $75,000 ($50,000 to less than $75,000)	0.630***	0.254***	1.047***	0.423***
	(0.00395)	(0.00484)	(0.00661)	(0.00816)
$75,000 or more	0.885***	0.410***	1.507***	0.708***
	(0.00365)	(0.00482)	(0.00623)	(0.00823)
Year=2008		0		0
		(.)		(.)
Year=2009		0.00902*		0.0138*
		(0.00353)		(0.00603)
Year=2010		0.0266***		0.0449***
		(0.00349)		(0.00597)
Year=2011		0.0296***		0.0540***
		(0.00362)		(0.00620)
Year=2012		0.0781***		0.127***
		(0.00385)		(0.00660)
Grade 8 or less		0		0
		(.)		(.)
Grades 9 through 11 (Some High School)		0.00443		0.00595
		(0.00793)		(0.0129)
Grade 12 or GED (High School Graduate)		0.131***		0.211***
		(0.00697)		(0.0114)
College 1 year to 3 years (Some college or technical school)		0.333***		0.547***
		(0.00711)		(0.0117)
College 4 years or more		0.558***		0.943***
		(0.00725)		(0.0120)
Employed for wages		0		0
		(.)		(.)
Self-employed		0.0993***		0.173***
		(0.00460)		(0.00808)

Out of work for more than 1 years		0.0949***		0.162***
		(0.00719)		(0.0121)
Out of work for less than 1 year		0.105***		0.178***
		(0.00727)		(0.0124)
A homemaker		0.165***		0.286***
		(0.00511)		(0.00879)
A student		0.186***		0.332***
		(0.0115)		(0.0206)
Retired		0.217***		0.366***
		(0.00400)		(0.00683)
Unable to work		-0.186***		-0.286***
		(0.00508)		(0.00834)
Male		0		0
		(.)		(.)
Female		-0.0550***		-0.0940***
		(0.00245)		(0.00421)
Does NOT have a Physical Activity Limitation		0		0
		(.)		(.)
Physical Activity Limitation		-0.398***		-0.671***
		(0.00274)		(0.00457)
Calculated Body Mass Index		-0.0000819***		-0.000134***
		(0.000000947)		(0.00000157)
In the center city of an MSA		0		0
		(.)		(.)
Outside the center city of an MSA but inside the county containing the center city		-0.0123***		-0.0201***
		(0.00316)		(0.00544)
Inside a suburban county of the MSA		-0.0612***		-0.103***
		(0.00369)		(0.00632)
In an MSA that has no center city		-0.0557***		-0.0919***
		(0.0148)		(0.0256)
Not in an MSA		-0.0365***		-0.0602***
		(0.00294)		(0.00502)
White only		0		0

		(.)		(.)
Black or African American only		-0.144***		-0.237***
		(0.00405)		(0.00681)
Asian Only		-0.201***		-0.322***
		(0.00873)		(0.0153)
Native Hawaiian or other Pacific Islander only		-0.0730**		-0.116**
		(0.0257)		(0.0443)
American Indian or Alaskan Native only		-0.0243*		-0.0410*
		(0.00992)		(0.0166)
Other race only		-0.0961***		-0.160***
		(0.00806)		(0.0136)
Multiracial		0.0850***		0.150***
		(0.00843)		(0.0145)
Age 18 to 24		0		0
		(.)		(.)
25 to 29		-0.125***		-0.211***
		(0.0108)		(0.0191)
30 to 34		-0.188***		-0.318***
		(0.0101)		(0.0179)
35 to 39		-0.227***		-0.389***
		(0.00983)		(0.0174)
40 to 44		-0.244***		-0.425***
		(0.00961)		(0.0170)
45 to 49		-0.251***		-0.438***
		(0.00940)		(0.0166)
50 to 54		-0.281***		-0.491***
		(0.00931)		(0.0165)
55 to 59		-0.323***		-0.561***
		(0.00935)		(0.0165)
60 to 64		-0.370***		-0.640***
		(0.00943)		(0.0167)
65 to 69		-0.400***		-0.692***
		(0.00972)		(0.0171)
70 to 74		-0.463***		-0.796***
		(0.00999)		(0.0175)
75 to 79		-0.560***		-0.958***
		(0.0102)		(0.0179)
80 or older		-0.697***		-1.181***
		(0.0101)		(0.0177)
No children household		0		0
		(.)		(.)

One child in household		-0.0481***		-0.0818***
		(0.00414)		(0.00715)
Two children in household		-0.0323***		-0.0550***
		(0.00463)		(0.00811)
Three children in household		-0.0491***		-0.0855***
		(0.00647)		(0.0113)
Four children in household		-0.0806***		-0.141***
		(0.0103)		(0.0178)
Five or more children in household		-0.103***		-0.179***
		(0.0148)		(0.0253)
Intercept	0.208***	0.861***	0.333***	1.438***
	(0.00286)	(0.0121)	(0.00460)	(0.0208)
Observations	1704104	1525256	1704104	1525256