A Production Function of Chronic Disease

Zach Pitkowsky

Abstract: Chronic ailments like coronary heart disease and cancers are responsible for the majority of the United States' large and inefficient level of health spending. These diseases, for the most part, are preventable through correct lifestyle decisions. Using data from the CHSI dataset, the Food Environment Atlas, and the County Health Rankings dataset at the US county level, this paper uses a US state-dummy-variable model to determine underlying factors that may affect the prevalence of coronary heart disease (CHD), lung cancer, and colon cancer on various causal factors. This study finds that fruit and vegetable consumption and an adult physical activity index are negatively associated with CHD, lung cancer, and colon cancer prevalence. Fat consumption is positively associated with CHD and colon cancer and meat consumption is positively associated with CHD, lung cancer, and colon cancer prevalence. Poverty is positively associated with prevalence of all three disease regressions while metropolitan status of a given county is negatively associated with prevalence of all three chronic conditions. An index for farm workers in each county highlights the strong positive relationship between farm work and disease prevalence across all three regressions. Such results and variation in disease prevalence suggest macro-level trends in health decisions as well as socioeconomic factors across the United States are at the root of the chronic disease epidemic crippling this nation's healthcare system. {limit the last few sentences as well}

JEL Codes: I18, I12, I14

I. Introduction:

The US healthcare system is the most expensive in the world, with average health care spending of $8995 per capita (World Bank 2012). Despite such high spending levels, the US ranks 38th in the World Health Organization's ranking of national healthcare systems (WHO 2000). Moreover, 7 out of the top 10 causes of death in the US are chronic diseases such as heart disease, cancer, asthma and diabetes. These diseases account for 75% of deaths and two of them (heart disease and cancer), account for half of all deaths each year (CDC 2014). These preventable diseases impose a huge burden on the health care system and have substantial economic consequences. Much of this health burden is preventable through nutritional choices, increased physical activity, early detection of risk factors and improvements in the delivery of health care.

While much of the discussion of healthcare in the United States centers around the insurance market, end-of-life costs, the government's involvement in health, and overall expense levels, it ignores the most important underlying issue, the burden of chronic disease. Ailments like coronary heart disease, type II diabetes, stroke, and certain types of cancer are preventable with the right lifestyle choices. Thus, it is important to understand the relationship between the causes of chronic diseases and the rates of sickness in order to achieve better health outcomes at a lower cost. This paper will examine the variability in the burden of chronic disease across US counties and link it to underlying factors that may explain the incidence of such diseases.

There are many different levels at which one can analyze the incidence of disease in the United States but this study will focus on the county level. County level data provides the ability to see aggregate-scale patterns throughout the United States. Additionally, there is a sufficient amount of data available at the county level to run interesting and meaningful regressions. This paper will use four data sets, the Community Health Status Indicators dataset from the Department of Health and Human Services, the County Health Rankings dataset from the

University of Wisconsin, the Food Environment Atlas from the USDA, and the USDA Census of Agriculture. Within these datasets are county-level variables on everything from food intake to rates of coronary heart disease to indicators of physical activity. This paper will set up multiple models, analyzing the regressions of different chronic diseases on factors like diet, exercise, and smoking while controlling for education, poverty, region, and state. Analysis of the coefficients will allow us to understand the impact each of these factors has on the incidence of chronic diseases throughout the United States.

The necessity of this type of investigation is emphasized by the massive cost burdens of disease and potential for lives to be saved and healthcare costs to be diminished if prevention is taken more seriously. It is well understood that chronic diseases are preventable, however it is not well understood how each of the causal factors play into the disease rates observed throughout the United States at an aggregate level. If we can better understand the decisions groups of people are making, we can better target those decisions in far more effective and impactful policy. The solution to many of our healthcare issues is not complex in nature but instead complex in implementation and behavior. The benefits of changing these behaviors, however, has the potential to drastically change the health of a nation.

II. Literature: Medicine Literature and Chronic Disease:

The fact that chronic disease is a function of personal production is not a new or novel concept (Grossman 1972). Empirically, the link between diet, exercise and chronic diseases is well spelled out in the medical science literature. Discussions of diet usually center around obesity, which has been steadily increasing in the United States for the last 50 years (NIH). Preventable obesity is a significant risk factor for multiple types of cancer, type II diabetes, and coronary heart disease, which has been studied using population data in the context of underlying clinical assumptions (Pi-Sunyer 1993). For coronary heart disease, the number one cause of death in the United States, rates of saturated fat and cholesterol consumption are

linked to disease prevalence (Blair et al. 1996). More specifically, red and processed meat consumption has been linked to coronary heart disease, diabetes, stroke, and cancer (McCullough et al. 2002),(Micha et al. 2010). In addition to diet, physical activity plays an important role in overall health and disease prevention (Bouchard et al. 1994),(McCullough et al. 1997). The previously mentioned studies used datasets following personal health decisions and physiological outcomes. Additionally, both diet and physical activity are risk factors for most cancers (World Cancer Research Fund and American Institute for Cancer Research 1997).

On a larger scale, the CDC estimates that eliminating three risk factors (poor diet, physical inactivity, and smoking) would prevent 80% of heart disease and stroke, 80% of type II diabetes, and 40% of cancers worldwide (WHO 2005). Additionally, while smoking has in the past been seen as a major risk factor for many chronic diseases, obesity is now a more significant risk factor (Jia and Lubetkin 2010). This study was done using the Behavioral Risk Factor Surveillance System to determine quality life years and the National Health Interview Survey's Linked Mortality Files to determine causes of death from specific diseases.

III. Literature: Economics of Chronic Disease

The economic implications of chronic diseases are staggering, with 84% of all healthcare spending in 2006 attributed to the 50% of the population with one or more chronic conditions (Robert Wood Johnson Foundation 2010). This data came from the household component of the MEPS (Medical Expenditure Panel Survey) sponsored by the Agency for Healthcare Research and Quality, which measures various health indicators and spending on non-institutionalized members of the US population. The total cost of heart disease and stroke alone in the US in 2012 was $315.4 billion (American Heart Association 2014). It is safe to say that the majority of healthcare spending in the United States goes to the treatment of chronic (preventable) diseases. Beyond the basic care costs, the economic implications of chronic diseases are staggering. 75.3% of working age people in the US have at least one chronic

disease and at least 50% have 2 or more (Nessens et al. 2011). These diseases have a negative impact on worker output, efficiency, and long-term viability (Blair et al. 1996).

Looking deeper at the economic relationships that underlie these statistics, a study using US Health and Retirement survey data found that education gaps are associated with the risk of dying from a chronic condition, with a less educated person at a higher risk of death given the same condition as a more educated person (Monteverde et al. 2010). In this context, it is necessary to also consider the impact that socio-economic and societal factors have on chronic disease rates when compared to genetic factors. A paper using the same County Health Rankings dataset as used in this paper, found that the relationship between household income and mortality is stronger at low income levels than high income levels (Cheng and Kindig 2010). This study also found an association between college education rates and mortality amongst US counties. Further, chronic exposure to social stress suggests specific neighborhood impacts on health outcomes (Pickett and Pearl 2001),(Baum et al. 1999). Harris et al. (2014) analyzed county health data in Kentucky, finding socioeconomic factors to outweigh education and other variables in relation to morbidity and mortality. A World Health Organization study authored by Marmot and Wilkinson (1999) (not limited to the US) states that differences in health between population groups are due to characteristics of society and not the actual care people get. Further, it stipulates that when people change social and cultural environments, their disease risks change, suggesting the high importance of surroundings in health outcomes. A survey paper by Smith (2011) asked the question: Is the current industrial food system a market failure and the reason for the United States' obesity epidemic? The paper concludes that the food system's inherent asymmetric information problems have resulted in a "lemon-style" breakdown in the market for processed foods, with consumers believing that their detriments and benefits are different from what they really are. Smith recommends that a new, independent food standards agency be put into place to oversee such issues.

In terms of race and chronic disease prevalence, the CDC Health Disparities and

Inequalities Report (2013) found that death from coronary heart disease was more prevalent amongst non-hispanic blacks than any other racial group. When considering childhood obesity, a significant risk factor for the diseases considered in this study, the report found Mexican-Americans to be at a higher risk of becoming obese than any other racial/ethnic category (CDC 2013). However, for adult obesity non-hispanic blacks are at highest risk of becoming obese (CDC 2013). Lastly, in terms of insurance coverage, non-hispanic blacks and Hispanics are disproportionately uninsured compared to all other racial/ethnic groups (CDC 2013).

Much of the economics literature also focuses on the costs and benefits of interventions aiming to improve underlying health conditions (Wong et al. 2011). Studies on fruit and vegetable pricing suggest that coupons are slightly more effective than price reductions but both strategies effectively increase fruit and vegetable consumption (Dong and Lin 2009) (Dong and Leibtag 2010). Studies have also found that significant taxes (20%) on sugar-rich beverages and processed foods can significantly reduce intake of unhealthy foods (Mozaffarian 2014)(Lin and Smith 2010). Additionally, when looking at the costs of current treatments for chronic diseases, research shows much potential to save money with simple changes like utilizing generic drugs (Shrank et al. 2011).

IV. Data

This paper relies on 4 datasets that contain observations at the United States county level. The datasets have various health and lifestyle variables at the aggregate of each county, which we plan to use to estimate large-scale trends in chronic conditions and lifestyle decisions across the United States. There are 3,144 county and county equivalents in the United States, which is the maximum number of observations for each variable. FIPS codes organize the data and make it possible to run various regression models utilizing dummy variables and intercepts for states as well as divisions, as defined by the US Census Bureau.

The first source is the Community Health Status Indicators (CHSI) dataset, which is a CDC dataset that synthesizes data from various government agencies as indicators of health and well- being in each county. Some of the most relevant variables to this paper are: the percentage of adults exercising, average visits to a primary care physician, prevalence of lung cancer, prevalence of colon cancer, prevalence of coronary heart disease, and a toxic chemical index. Disease rates are measured in deaths per 100,000 people. In searching the literature, one study used CHSI dataset to look at a spatial analysis of health and poverty (Holt 2007). The study made various maps of the United States with the county level data. The only shortcoming of this dataset is that some of the relevant variables are missing observations for several-hundred counties, losing degrees of freedom in the regression analysis.

The second data source is the Food Environment Atlas, which pulls data from various government and congressional reports on food access and consumption. The Food Environment Atlas data is at the county level, allowing it to be easily matched to the CHSI dataset using FIPS codes. Some of the most relevant variables are: obesity rates, fat consumption per household (lbs per year), fruit and vegetable consumption per household (lbs per year), and physical activity measured as percentage of adults who meet recommended weekly physical activity requirements.

The third dataset is the County Health Rankings Dataset, which is from the Robert Wood Johnson Foundation and the University of Wisconsin Population Health Institute. The CHR dataset ranks counties within their respective states in terms of various health outcome indicators. It also contains aggregate average variables on smoking rates in each county, the percentage of binge drinkers in each county, and healthy food store availability (zip codes per county with healthy food stores). This dataset is also linkable to the others by FIPS county codes.

The fourth dataset is the USDA Census of Agriculture, which includes farm output and agricultural labor characteristics. This paper uses county-level totals (linked with FIPS codes) of farm workers to control for the effects farm labor practices have on health outcomes.

The explanatory variables FRUVEG, FATS, and MEAT measure average household consumption (in lbs) of fruits and vegetables, fats, and meats respectfully. Physical Acitivty is an explanatory variable for physical activity, measuring the percent of adults in a given county meeting minimum activity guidelines (150 minutes of moderate activity or 75 minutes of vigorous activity weekly). Smokers measures the percentage of adults who report currently smoking at the time of the survey. Binge drinking measures the percentage of adults that self report binge drinking. Healthy food is a variable included to control for access to healthy foods. It measures the percentage of zip codes within each county that have a market selling healthy foods (fresh fruits and vegetables). Poverty is simply a measure of the percentage of people in each county below the poverty threshold. Toxic_Chem measures the total amount (in lbs) of chemical released for the county in the year measured. Metro is a binary (1 or 0) classification as a county as metropolitan or not, as defined by the Office of Management and Budget. The explanatory variable unemployed is the unemployment rate in each county. College measures the percentage of the population age 25 or older with a 4-year college degree or higher. The Black and Hispanic variables measure the percentage of the population in each county that identify as part of one of the respective classifications. Farm is a transformed variable and measures the fraction of farm workers over total population of each county.

Summary statistics and a correlation for the explanatory variables are available in the data appendix (Table A1)(Table A2). The three independent variables (prevalence of coronary heart disease, lung cancer, and colon cancer) are all skewed in their distributions. To adjust for this, the natural log of the three variables is used in the regression models. Further, to make the coefficient of the Toxic_Chem variable more manageable the variable was scaled by dividing each observation by 10^9. The explanatory variables range in their levels of deviation

amongst the samples, with meat consumption and healthy-food access highly variable and adult physical activity and fat consumption less variable. The Black and Hispanic variables both range from 0 to 85-90% across US counties, which suggests high levels of segregation. The Farm variable, measuring farm workers, ranges from 0 to 2, which suggests that in some counties there are more farm workers than people in the population measure. This is not surprising, as in some counties with large agricultural operations workers may be bussed in from outside counties. The binge drinking variable is the limiting factor in terms of degrees of freedom, with only 2,415 observations.

Some notable correlations are the high; for instance it is 0.5124 between fruit and vegetable and meat consumption, which is surprising because the two variables have opposite sign coefficients in the disease regressions (Table A2). The correlation between meat and fat consumption is high and negative (-0.4703), which is also surprising, as meats are a major source of fat for many individuals. A less surprising relationship is the high negative correlation (-0.3546) between smoking and physical activity. One of the highest correlations was between Black and Poverty (0.4921), which suggests the systemic nature of the socioeconomic variables. Additionally, Black is negatively associated with fat consumption (-0.4002), which is surprising when considering the literature on obesity amongst Black identifying individuals.

V. Empirical Model:

Three different sets of empirical models were created to analyze the data. For each set there is a regression for each of the three disease independent variables (coronary heart disease, colon cancer, lung cancer). The simplest set of models regressed disease prevalences on the explanatory variables explained above. A more controlled model added regional (US Census Bureau divisions) intercepts to each model to control for region-specific factors. The third model and the one whose results are used in this paper, includes US state dummy variables, omitting Alabama.

$$lDiseasePrevalence = RegionalConstants + \beta_1 FruVeg + \beta_2 Fats + \beta_3 Meat + \beta_4 AdultPA + \beta_5 Smokers + \beta_6 BingeDrinking + \beta_7 HealthyFood + \beta_8 Poverty + \beta_9 Metro + \beta_{10} Unemployment + \beta_{11} ToxicChem + \beta_{12} College + \beta_{13} Farm + \beta_{14} Black + \beta_{15} Hispanic$$

In the above equation **β**1, **β** 4, **β** 7, **β** 13, **β** 14, and **β** 15 are expected to be negative, implying that the associated variables decrease the incidence of disease. **β** 2, **β** 3, **β** 5, and **β** 6 are expected to be positive, as they are associated with variables that are understood on the individual level to increase disease risk. The variables associated with **β** 1 to **β** 6 capture the behavioral effects of a county on health outcomes. The remaining variables in the equation capture the socioeconomic aspects of county disease incidence.

VI. Results:

Table 1: State Dummy Regressions for CHD, Lung Cancer, and Colon Cancer

	(1)	(2)	(3)
VARIABLES	lCHD	lLung_Can	lCol_Can
FRUVEG Consumption	-0.00256***	-0.00322***	-0.00222***
	(0.000722)	(0.000637)	(0.000711)
FAT Consumption	0.00897*	-0.00469	0.0110**
	(0.00539)	(0.00476)	(0.00531)
MEAT Consumption	0.00427***	0.00489***	0.00379***
	(0.00101)	(0.000890)	(0.000994)
Physical Activity	-0.0188***	-0.0138***	0.000347
	(0.00326)	(0.00288)	(0.00321)
Smokers	0.000183	0.00301**	-0.000210
	(0.00158)	(0.00140)	(0.00156)
BingeDrinking	0.00348*	-0.000369	0.00320*
	(0.00186)	(0.00164)	(0.00183)
HealthyFood	-0.000333	0.000153	-0.000359
	(0.000406)	(0.000358)	(0.000400)
Poverty	0.0175***	0.0189***	0.00583***
	(0.00173)	(0.00153)	(0.00171)
METRO	-0.188***	-0.0784***	-0.200***
	(0.0171)	(0.0151)	(0.0169)

Unemployed	-0.00760*	-0.00367	-0.0150***
	(0.00455)	(0.00401)	(0.00448)
Toxic_Chem	-3.219*	-2.178	-4.645**
	(1.914)	(1.688)	(1.885)
College	-0.00265**	-0.000816	-0.00284***
	(0.00106)	(0.000931)	(0.00104)
Farm	0.728***	0.439***	0.795***
	(0.0801)	(0.0707)	(0.0789)
Black	-0.00113	-0.00272***	0.00254***
	(0.000703)	(0.000620)	(0.000692)
Hispanic	-0.00316***	-0.00921***	-0.00472***
	(0.000862)	(0.000761)	(0.000849)
State Dummy Variables			
Constant	6.065***	5.103***	2.961***
	(0.236)	(0.208)	(0.232)
Observations	2,078	2,078	2,078
R-squared	0.398	0.436	0.334

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

The results of the various regressions highlight the different sources of variation in the model. The final regression displayed above used US state dummy variables to pick up variation related to state environmental and societal factors. The regression was run on coronary heart disease prevalence, colon cancer prevalence, and lung cancer prevalence (Table A1). The regression model was also run on the three disease prevalences with regional constants (US Census Divisions) as well as without any regional controls (Table A3)(Table A4). In the most basic form, (no dummy variables) there was more significance to be found in the explanatory variables. The coefficients on fruit and vegetable consumption were negative for all three disease models while the coefficients on meat consumption were positive for all three models. Fat consumption has a positive and significant coefficient in only the coronary heart disease and colon cancer models. Adult physical activity has a negative significant coefficient for all three disease regressions. Smoking, interestingly, has a significant negative coefficient for colon cancer and a significant positive coefficient for lung cancer. The coefficients on binge

drinking are positive for both coronary heart disease and colon cancer. HealthyFood has a negative coefficient for for both coronary heart disease and colon cancer. Poverty has positive and significant coefficients for all three models. The Metro dummy variable has a negative coefficient for all three disease prevalences while unemployed is negative and significant for lung and colon cancer. ToxicChem and College both have negative coefficients for coronary heart disease and colon cancer. The Farm worker variable has a positive coefficient for all disease prevalences. Hispanic has a negative and significant coefficient for all diseases while Black is only negative for coronary heart disease and lung cancer. These coefficients are a mix of what we expect and did not expect. Fruit and vegetable consumption and adult physical activity were expectedly negative, as they are commonly related with lower disease risk. On the other hand, fat consumption, smoking, meat consumption, and binge drinking were all positive as expected. For socioeconomic variables, poverty and HealthyFood have positive coefficients as expected. However, unemployed, Hispanic, Black, College, and ToxicChem all have unexpectedly negative coefficients.

The regional model, which uses US Census Bureau divisions as intercept terms for each chronic disease prevalence, has less significance among the variables (Table A4). Fruit and vegetable, fat, and meat consumption coefficients were unchanged in sign and significance from the basic model. The physical activity, smoking, and HealthyFood variables lose their significance in the colon cancer regression. Poverty and the Metro dummy variable maintain their coefficient signs and significance in all models. Unemployed, interestingly, picks up significance in the transition from the basic model to the divisions model and has negative coefficients in all three models. The coefficients on ToxicChem have no change in sign or significance from the basic model and the coefficients on college pickup significance, as they are negative for all three models. The Farm variable's coefficients remain unchanged in the divisions model. Concerning race and ethnicity, contrary to the basic model, the regional model for colon cancer has a positive significant coefficient on Black while the coefficients for coronary

heart disease and lung cancer remain the same as in the basic model. There is also no change in coefficient sign or significance for the Hispanic variable in this model. Overall, the regional model both losses and gains significance in certain areas, suggesting important regional trends with still much variation left to behavioral aspects of health.

The model this paper primarily uses includes state dummy variables, omitting Alabama (Table 1). The state dummy model has various differences from the divisional dummy model. There is no change in the sign or significance of the coefficients on fruit and vegetable consumption, fat consumption, or meat consumption when moving from the regional model to the state model. For coronary heart disease, lung cancer, and colon cancer, a 1 pound change in fruit and vegetable consumption is associated with a -0.256%, -0.322%, and -0.222% decrease in disease prevalence, respectively. For coronary heart disease and colon cancer, a 1 pound increase in household fat consumption is associated with a 0.897% or 1.10% increase in disease prevalence, respectively. In terms of meat consumption, a 1 pound increase in household consumption is associated with a 0.427% increase in coronary heart disease prevalence, a 0.489% increase in lung cancer prevalence, and a 0.379% increase in colon cancer prevalence. Like the regional model, physical activity has negative and significant coefficients for only the coronary heart disease and lung cancer models. A 1% change in county adults who meet physical activity guidelines is associated with a -1.88% decrease in coronary heart disease prevalence and a -1.38% decrease in lung cancer prevalence. Also like the regional model, smoking has a negative coefficient only in the lung cancer regression. A 1% increase in the number of regular smokers in a given county is associated with a 0.301% increase in lung cancer prevalence. In the state model, the healthy food access variable loses significance on the coefficient in the coronary heart disease regression. Interestingly, unlike the regional model but similar to the basic model, coronary heart disease and colon cancer have positive significant coefficients on binge drinking in the state model. A 1% increase in the number binge drinkers in a given county is associate with a 0.348% increase in the prevalence

of coronary heart disease and a 0.320% increase in the prevalence of colon cancer. The state model loses significance in the coefficient on healthy food. From the regional model there is no change in sign or significance for the coefficients on poverty. For coronary heart disease, lung cancer, and colon cancer, a 1% change in the poverty rate is associated with a 1.75% increase in coronary heart disease prevalence, a 1.89% increase in lung cancer prevalence, and a 0.583% increase in colon cancer prevalence. The coefficients on Metro in the state dummy model are unchanged in sign and significance from the regional model. The binary metropolitan classification is associated with a -18.8% decrease in coronary heart disease prevalence, a -7.84% decrease in lung cancer prevalence, and a -20.0% decrease in colon cancer prevalence. Unemployed lost its significance in the lung cancer regression, however, a 1% change in the unemployment rate is associated with a -0.760% decrease in coronary heart disease prevalence and a -1.50% decrease in colon cancer prevalence. ToxicChem maintained a significant coefficient for both the coronary heart disease and the colon cancer regressions in the state dummy model. The coefficients on ToxicChem are large due to scaling, as noted in the previous section. The coefficients on College are unchanged in sign and significance from the regional model except for a loss of significance for the lung cancer regression. A 1% increase in the number of adults with a college education in a given county is associated with -0.265% reduction in coronary heart disease prevalence and a -0.284% reduction in colon cancer prevalence. The coefficients on the Farm variable do not change in sign or significance in the state model. A 1% change in the number of farm workers in a given county is associated with a 72.8% increase in the prevalence of coronary heart disease, a 43.9% increase in the prevalence of lung cancer, and a 79.5% increase in the prevalence of colon cancer. The Black variable's coefficients lose significance for coronary heart disease in the state model but maintain significance for both lung and colon cancer with the same signs as in the regional model. A 1% change in the number of Black identifying individuals in a given county is associated with a -0.272% decrease in lung cancer prevalence and a 0.254% increase in the prevalence of colon

cancer. Lastly, there is no change in sign or significance in the Hispanic variable from the regional to the state model. A 1% change in the number of Hispanic identifying individuals in a county is associate with a -0.316% decrease in the prevalence of coronary heart disease, a -0.921% decrease in the prevalence of lung cancer, and a -0.472% decrease in the prevalence of colon cancer. While several coefficients lost significance in the transition from the regional intercept to the state dummy model, it was a far smaller change than expected.

VII. Conclusion:

Over the past century there has been a seismic shift in disease patterns in the United States, from communicable diseases like tuberculosis burdening the health care system to chronic ailments like diabetes, heart disease, and cancer accounting for the majority of health spending (Lim et al. 2013). While this shift occurred, other trends also took hold, including significant increases in life spans on the positive end and the troubling rise in obesity rates on the negative end (NIH). As the literature review highlighted, chronic diseases impose massive societal burdens in both economic and social terms. They account for the majority of healthcare spending in the United States and inhibit production in the workplace. Fortunately, unlike communicable diseases, chronic ailments are largely preventable by lifestyle decisions. In this light, it is necessary to better understand the epidemiology of these diseases in the United States if we ever hope to effectively combat their proliferation and societal burden. County data gives us an ability to observe aggregate scale trends as they relate to disease burden, lifestyle decisions, and socioeconomic factors.

The explanatory variable for fruit and vegetable consumption has coefficients that align with our expectations expected signs, as eating healthy should be negatively associated with higher chronic disease rates (Pi-Sunyer 1993),(Blair et al. 1996). More specifically, fruit and vegetable consumption have been found to be associated with modest reduction in risks of major chronic diseases (Hung et al. 2004). Additionally, since the model included variables to

control for poverty and food access, fruit and vegetable consumption is an indicator of food choices made on the aggregate county level. The coefficient on meat consumption's signs align with the literature that spelled out the links between meat consumption and various chronic disease risks (McCullough et al. 2002),(Micha et al. 2010). Interestingly, the significance of meat consumption is still present after adding state dummy variables. It was expected that differing levels of meat consumption could be attributed to different regions but this is not solely the case. Additionally, the fat consumption variable has positive coefficients for both coronary heart disease and colon cancer. While much of the fat most people consume comes from meat, there is clearly enough other sources of fat (possibly non meat animal protein) to have significance in these coefficients. This aligns with the literature that spells out the link between fat consumption and disease prevalence (Blair et al. 1996).

The variable measuring the percentage of adults engaging in physical activity has a significant negative coefficient for both coronary heart disease and lung cancer prevalences (Table 1). Interestingly, significance is lost on the coefficient in the colon cancer regression from the basic to the regional model, which could suggest some regional variability in exercise possibly due to the number of sunny days in different locals (Table A3),(Table A4). This result aligns with much of the literature that suggests physical activity plays a significant role in disease prevalence and prevention (Bouchard et al. 1994),(McCullough et al. 1997),(World Cancer Research Fund and American Institute for Cancer Research 1997). Smokers, unsurprisingly, has a significant and positive coefficient for the lung cancer regression (Table 1).

The coefficient on the poverty variable supports the literature that suggests there is a stronger relationship between health outcomes and income at the low end than the high end (Cheng and Kindig 2010). It also aligns with county-level research in Kentucky that found socioeconomic factors to outweigh education and other variables in terms of overall morbidity and mortality, as the variable measuring college attendance is insignificant in this regression analysis (Harris et al. 2014). This relationship is not surprising, as healthcare and healthy

lifestyles can be expensive. Since the data measures disease prevalence in terms of deaths, the coefficient on poverty may only be saying that poverty is related to risk of death from a disease instead of actual disease incidence. This is logical since treatment for many chronic ailments is far more expensive than the lifestyle changes necessary for prevention. The socioeconomic disease burdens are not just related to income; research shows that chronic exposure to social stress impacts health outcomes (Pickett and Pearl 2001)(Baum et al. 1999). This speaks to the systematic aspects of health, that while hard to measure, are necessary to consider in any discussion of disease reduction.

The coefficient on the binary variable measuring each county's metropolitan status is negative and significant for all three disease regressions (Table 1). This relationship could speak to issues of access when it comes to health and health services. In a city there are far more options in terms of doctors and hospitals, gyms for exercise, and healthy food stores. Another potential reason for the result is that people may live and work in cities and then go to die in less expensive elderly care facilities in outer-metro areas. This would skew disease prevalence (measured by deaths in this study) to areas outside of metropolitan areas. Since the correlation between healthy food access and Metro is low, the metropolitan variable may be explaining a further relationship between cities and chronic disease prevalence (Table A2).

The coefficients on Unemployed are not what we expect, as the literature and other variables link poor socioeconomic conditions to higher disease prevalence (Harris et al. 2014). An explanation for this relationship may be the unemployment rate's limited ability to measure actual labor situations due to factors like labor force participation. College has significant and negative coefficients for the coronary heart disease and colon cancer regressions (Table 1). This relationship is what we expect from the literature, as education is negatively associated with risk of disease death and the data uses death rates to estimate disease prevalence (Cheng and Kindig 2010). The Black variable, perplexingly, has a significant coefficient in both the coronary heart disease and colon cancer regressions, however, the sign is not the same in both.

We expect the coefficient to be positive, as Black populations have been linked with higher disease incidences, and this is the case in the colon cancer regression (CDC 2013). The reasoning for the negative coefficient in the coronary heart disease regression could be the high standard deviation in Black percentage throughout the counties. It may be more likely that modern segregation happens on town and city levels instead of county and state levels. However, this does not explain the conflicting coefficients. The coefficients on the Hispanic variable are also unexpected and in contradiction with the literature (CDC 2013). The explanations for the similar issues with the Black variable may be applicable, however, there is clearly something unaccounted for in the data.

The Farm variable, which measures percentage of farm workers in a given county, has positive significant coefficients in all three state dummy regressions (Table 1). This is not surprising, as several studies have linked the farming occupation to being associated with worse overall health (Rola 1993),(Gerrard 1998). The reasons for this are not just the physical labor aspect of farming. Instead, it is linked to the pesticides and other dangerous chemicals many industrial farmers use in their trade. A limitation of this finding is that since farmers may work in a different county from which they live (and die) there is little assurance of the link between being a farmer and dying of a chronic disease, on an individual level.

The limitations of this research are significant, as county-level aggregate data cannot be substituted for individual level observations. Thus, it is hard to extrapolate these findings to things like individual lifestyle decisions. Further, the fact that disease prevalences are measured by death rates limits the ability to use the data to understand current trends, as diseases were accumulated over many years. Instead, the research speaks to broader trends of food consumption, socioeconomics, and disease.

Despite the limitations, the analysis has implications in terms of how to combat the epidemic of chronic disease. Fruit and vegetable consumption is significant in the regression analysis and several studies show that price reductions (subsidies) and coupons on fruits and

vegetables are useful in increasing consumption (Dong and Lin 2009),(Dong and Leibtag 2010). These strategies merit further investigation and even implementation. With these types of programs, there may also be a need for improvements in access in food deserts. Further, the significance in adult physical activity suggests a potential to change behaviors positively not just in food consumption but also in exercise habits. The significance in the meat and fat variables may speak to market problems with the industrial food system, as consumers lack necessary nutritional information when making purchases (Smith 2011). Implications from the significance in poverty, Metro, College, Black, and Farm variables speak to the larger systemic issues influencing the health behaviors of millions of Americans. Issues of poverty and access are crucial to changing eating and activity behaviors.

The results of this paper are significant in that they highlight population-level trends in chronic disease rates and their associations with the factors science has proven cause such diseases. The lack of individual data and inclusion of state dummy variables showed the amount of regional variability in the factors considered. The implications of this research are that the significant factors in the state-dummy model are most likely the factors that, if targeted, could have the largest impact on disease rates (fruit and vegetable consumption, fat consumption, meat consumption, smoking, poverty, college education and access). Surely more research is done using individual data and data on those living with chronic ailments, not just those who die from them. The importance of this research is obvious, as chronic ailments are responsible for a majority of the deaths and excessive health care spending in the United States.

References

American Diabetes Association. (2013). Economic costs of diabetes in the U.S. in 2012. *Diabetes Care, 36*(4), 1033-1046. doi:10.2337/dc12-2625 [doi]

Arndt, S., Acion, L., Caspers, K., & Blood, P. (2013). How reliable are county and regional health rankings? *Prevention Science, 14*(5), 497-502.

Bartley, M., Ferrie, J., Montgomery, S., MARMOT, M., & WILKINSON, R. (1999). Social determinants of health. *Social Determinants of Health,*

Baum, A., Garofalo, J., & YALI, A. (1999). Socioeconomic status and chronic stress: Does stress account for SES effects on health? *Annals of the New York Academy of Sciences, 896*(1), 131-144.

Blair, S. N., Horton, E., Leon, A. S., Lee, I. M., Drinkwater, B. L., Dishman, R. K., . . . Kienholz, M. L. (1996). Physical activity, nutrition, and chronic disease. *Medicine and Science in Sports and Exercise, 28*(3), 335-349.

Bouchard, C. E., Shephard, R. J., & Stephens, T. E. (1994). Physical activity, fitness, and health: International proceedings and consensus statement. *International Consensus Symposium on Physical Activity, Fitness, and Health, 2nd, may, 1992, Toronto, ON, Canada,*

Centers for Disease Control and Prevention (CDC). (2013). CDC health disparities and inequalities report—United states, 2013. *Morbidity and Mortality Weekly Report,*

Cheng, E. R., & Kindig, D. A. (2012). Disparities in premature mortality between high- and low-income US counties. *Preventing Chronic Disease, 9*, E75. doi:E75 [pii]

Chiuve, S. E., McCullough, M. L., Sacks, F. M., & Rimm, E. B. (2006). Healthy lifestyle factors in the primary prevention of coronary heart disease among men: Benefits among users and nonusers of lipid-lowering and antihypertensive medications. *Circulation, 114*(2), 160-167. doi:CIRCULATIONAHA.106.621417 [pii]

Deo, S., Iravani, S., Jiang, T., Smilowitz, K., & Samuelson, S. (2013). Improving health outcomes through better capacity allocation in a community-based chronic care model. *Operations Research, 61*(6), 1277-1294.

Divajeva, D., Marsh, T., Logstrup, S., Kestens, M., Vemer, P., Kriaucioniene, V., . . . Webber, L. (2014). Economics of chronic diseases protocol: Cost-effectiveness modelling and the future burden of non-communicable disease in europe. *BMC Public Health, 14*(1), 456.

Dong, D. (2010). *Promoting fruit and vegetable consumption: Are coupons more effective than pure price discounts?* DIANE Publishing.

Dong, D., & Lin, B. (2009). *Fruit and vegetable consumption by low-income americans* Citeseer.

Fraser, G. E., & ebrary, I. (2003). *Diet, life expectancy, and chronic disease*. Oxford; New York: Oxford University Press.

Gerrard, C. E. (1998). Farmers' occupational health: Cause for concern, cause for action. *Journal of Advanced Nursing, 28*(1), 155; 155-163; 163.

Gittelsohn, J. (2012). Interventions in small food stores to change the food environment, improve diet, and reduce risk of chronic disease. *Preventing Chronic Disease, 9*

Glade, M. J. (1999). Food, nutrition, and the prevention of cancer: A global perspective. american institute for cancer Research/World cancer research fund, american institute for cancer research, 1997. *Nutrition (Burbank, Los Angeles County, Calif.), 15*(6), 523-526. doi:S0899900799000210 [pii]

Halpin, H. A., Morales-Suarez-Varela, M. M., & Martin-Moreno, J. M. (2010). Chronic disease prevention and the new public health. *Public Health Reviews, 32*(1), 120-154.

Harris, A. L., Scutchfield, F. D., Heise, G., & Ingram, R. C. (2014). The relationship between local public health agency administrative variables and county health status rankings in kentucky. *Journal of Public Health Management and Practice : JPHMP, 20*(4), 378-383. doi:10.1097/PHH.0b013e3182a5c2f8 [doi]

Heath, G. W. (2009). Physical activity transitions and chronic disease. *American Journal of Lifestyle Medicine,*

Herrera, N. (2011). Access to affordable and nutritious food: Measuring and understanding food deserts and their consequences. *Eating Right: The Consumption of Fruits and Vegetables,* , 1-137.

Holt, J. B. (2007). The topography of poverty in the united states: A spatial analysis using county-level data from the community health status indicators project. *Preventing Chronic Disease, 4*(4), A111. doi:A111 [pii]

Howard, D. H., Thorpe, K. E., & Busch, S. H. (2010). Understanding recent increases in chronic disease treatment rates: More disease or more detection? *Health Economics, Policy and Law, 5*(04), 411-435.

Jia, H., & Lubetkin, E. I. (2010). Trends in quality-adjusted life-years lost contributed by smoking and obesity. *American Journal of Preventive Medicine, 38*(2), 138-144.

Lantz, P. M., & Pritchard, A. (2010). Socioeconomic indicators that matter for population health. *Preventing Chronic Disease, 7*(4), A74. doi:A74 [pii]

Lim, S. S., Vos, T., Flaxman, A. D., Danaei, G., Shibuya, K., Adair-Rohani, H., . . . Andrews, K. G. (2013). A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: A systematic analysis for the global burden of disease study 2010. *The Lancet, 380*(9859), 2224-2260.

Lin, B., Smith, T. A., & Lee, J. (2010). The effects of a sugar-sweetened beverage tax: Consumption, calorie intake, obesity, and tax burden by income. *Agricultural & Applied Economics Association Meeting,*

Lorig, K. R., Ritter, P., Stewart, A. L., Sobel, D. S., Brown Jr, B. W., Bandura, A., . . . Holman, H. R. (2001). Chronic disease self-management program: 2-year health status and health care utilization outcomes. *Medical Care, 39*(11), 1217-1223.

Mackay, J., & Mensah, G. A. (2004). The atlas of heart disease and stroke. *The Atlas of Heart Disease and Stroke,*

McCullough, M. L., Feskanich, D., Stampfer, M. J., Giovannucci, E. L., Rimm, E. B., Hu, F. B., . . . Willett, W. C. (2002). Diet quality and major chronic disease risk in men and women: Moving toward improved dietary guidance. *The American Journal of Clinical Nutrition, 76*(6), 1261-1271.

McCullough, M. L., Feskanich, D., Stampfer, M. J., Giovannucci, E. L., Rimm, E. B., Hu, F. B., . . . Willett, W. C. (2002). Diet quality and major chronic disease risk in men and women: Moving toward improved dietary guidance. *The American Journal of Clinical Nutrition, 76*(6), 1261-1271.

Monteverde, M., Noronha, K., Palloni, A., & Novak, B. (2010). Obesity and excess mortality among the elderly in the united states and mexico. *Demography, 47*(1), 79-96.

Mozaffarian, D., Rogoff, K. S., & Ludwig, D. S. (2014). The real cost of food: Can taxes and subsidies improve public health? *Jama, 312*(9), 889-890.

Mozaffarian, D., Afshin, A., Benowitz, N. L., Bittner, V., Daniels, S. R., Franch, H. A., . . . American Heart Association Council on Epidemiology and Prevention, Council on Nutrition, Physical Activity and Metabolism, Council on Clinical Cardiology, Council on Cardiovascular Disease in the Young, Council on the Kidney in Cardiovasc. (2012). Population approaches to improve diet, physical activity, and smoking habits: A scientific statement from the american heart association. *Circulation, 126*(12), 1514-1563. doi:CIR.0b013e318260a20b [pii]

Naessens, J. M., Stroebel, R. J., Finnie, D. M., Shah, N. D., Wagie, A. E., Litchy, W. J., . . . Nesse, R. E. (2011). Effect of multiple chronic conditions among working-age adults. *The American Journal of Managed Care, 17*(2), 118-122. doi:47687 [pii]

Pickett, K. E., & Pearl, M. (2001). Multilevel analyses of neighbourhood socioeconomic context and health outcomes: A critical review. *Journal of Epidemiology and Community Health, 55*(2), 111-122.

Pi-Sunyer, F. X. (1993). Medical hazards of obesity. *Annals of Internal Medicine, 119*(7_Part_2), 655-660.

Renata Micha, R., Wallace, S. K., & Mozaffarian, D. (2010). Red and processed meat consumption and risk of incident coronary heart disease, stroke, and diabetes mellitus.

Richards, H., & Barry, R. (1998). US life tables for 1990 by sex, race, and education. *Journal of Forensic Economics, 11*(1), 9-26.

Robert Wood Johnson Foundation. (2010). *County health rankings dataset.* (Dataset No. 1).Robert Wood Johnson Foundation & University of Wisconsin.

Rola, A. C. (1993). *Pesticides, rice productivity, and farmers' health: An economic assessment* IRRI CABI.

Shrank, W. H., Choudhry, N. K., Liberman, J. N., & Brennan, T. A. (2011). The use of generic drugs in prevention of chronic disease is far more cost-effective than thought, and may save money. *Health Affairs (Project Hope), 30*(7), 1351-1357. doi:10.1377/hlthaff.2010.0431 [doi]

Slade, A. N. (2012). Health investment decisions in response to diabetes information in older americans. *Journal of Health Economics, 31*(3), 502-520.

Smith, T. G., Chouinard, H. H., & Wandschneider, P. R. (2011). Waiting for the invisible hand: Novel products and the role of information in the modern market for food. *Food Policy, 36*(2), 239-249.

Stuckler, D. (2008). Population causes and consequences of leading chronic diseases: A comparative analysis of prevailing explanations. *Milbank Quarterly, 86*(2), 273-326.

Tunstall-Pedoe, H. (2006). Preventing chronic diseases. A vital investment: WHO global report. geneva: World health organization, 2005. pp 200. CHF 30.00. ISBN 92 4 1563001. also

published on http://www. who. int/chp/chronic_disease_report/en. *International Journal of Epidemiology, 35*(4), 1107-1107.

Van Remoortel, H., Giavedoni, S., Raste, Y., Burtin, C., Louvaris, Z., Gimeno-Santos, E., . . . Vogiatzis, I. (2012). Validity of activity monitors in health and chronic disease: A systematic review. *Int J Behav Nutr Phys Act, 9*(1), 84.

Wolf, A. M., & Colditz, G. A. (1998). Current estimates of the economic cost of obesity in the united states. *Obesity Research, 6*(2), 97-106.

Wong, J. B., Coates, P. M., Russell, R. M., Dwyer, J. T., Schuttinga, J. A., Bowman, B. A., & Peterson, S. A. (2011). Economic analysis of nutrition interventions for chronic disease prevention: Methods, research, and policy. *Nutrition Reviews, 69*(9), 533-549.

Data Appendix:

Table A1: Summary Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
lCHD	2899	5.381707	0.399184	4.265493	6.469095
lLung_Can	2899	4.278866	0.361081	2.424803	5.580107
lCol_Can	2899	3.224862	0.378351	1.589235	4.32148
FRUVEG Consumption	2899	173.1263	19.43072	143	252
FAT Consumption	2899	18.64057	2.30682	13	24
MEAT Consumption	2899	70.6516	12.85442	31	120
PhysicalAcitivty	2899	64.12839	4.267828	51.8	73.3
Smokers	2264	22.55954	5.794856	2.41	44.9
BingeDrinking	2415	13.49778	5.183783	0	35.3
HealthyFood	2898	38.98281	19.87774	0	100
Poverty	2899	15.26547	5.971958	3.1	46.2
METRO	2899	0.369093	0.482642	0	1
Unemployed	2898	5.862388	2.07967	1.6	22.9
Toxic_Chem	2899	0.001143	0.00509	-2.22E-06	0.189112
College	2898	17.71993	8.455861	4.920534	69.70375
Farm	2777	0.042443	0.102183	4.20E-06	2.01914
Black	2899	9.665919	14.68213	0	86
Hispanic	2899	7.289997	12.35339	0.1	97.2

Table A2: Correlation Matrix

	FRUVEG Consumption	FAT Consumption	Meat Consumption	Physical Activity	Smokers	Binge Drinking	Healthy Food	Poverty	METRO
FRUVEG Consumption	1								
FAT Consumption	0.2614	1							
Meat Consumption	0.5124	-0.4073	1						
Physical Activity	0.3257	0.2669	-0.0887	1					
Smokers	-0.1209	-0.0831	0.0589	-0.3546	1				
BingeDrinking	0.1509	0.3498	-0.169	0.3809	-0.1706	1			
HealthyFood	0.0756	-0.0694	0.0779	-0.0068	-0.0843	0.0091	1		
POV_RATE	-0.3425	-0.2643	0.0508	-0.333	0.1956	-0.2877	-0.0059	1	
METRO	0.1893	-0.1547	0.1464	0.0077	-0.0253	-0.0282	0.0525	-0.3265	1
unemployed	0.0722	-0.0828	0.1262	-0.2006	0.2958	-0.2264	0.1169	0.2081	0.0653
Toxic_Chem	0.0354	-0.0161	0.0204	-0.01	0.0237	0	0.0445	-0.0404	0.1241
College	0.1885	0.0459	0.0088	0.2415	-0.4759	0.314	0.1512	-0.2135	0.1078
Farm	-0.0479	0.0005	-0.026	0.0648	-0.04	0.0292	-0.0297	0.0034	-0.1888
Black	-0.1777	-0.4002	0.2371	-0.3054	0.0829	-0.2542	0.0917	0.4921	0.0538
Hispanic	0.0457	-0.2495	0.0579	0.1661	-0.1543	0.0373	0.0381	0.0656	0.0987
	Unemployed	ToxicChem	College	Farm	Black	Hispanic			
Unemployed	1								
Toxic_Chem	0.0384	1							
College	-0.3175	-0.0138	1						
Farm	-0.1096	-0.0447	-0.0287	1					
Black	0.2419	0.0402	-0.0494	-0.111	1				
Hispanic	0.0043	0.0368	0.0917	-0.0409	-0.0763	1			

Table A3: Basic Regressions for CHD, Lung Cancer, and Colon Cancer

	(1)	(2)	(3)
VARIABLES	lCHD	lCol_Can	lLung_Can
FRUVEG Consumption	-0.00327***	-0.00232***	-0.00334***
	(0.000664)	(0.000654)	(0.000584)
FAT Consumption	0.0109**	0.0132***	-0.00481
	(0.00498)	(0.00491)	(0.00438)
MEAT Consumption	0.00495***	0.00359***	0.00548***
	(0.000956)	(0.000942)	(0.000841)
Physical Activity	-0.0176***	-0.00566***	-0.0110***
	(0.00211)	(0.00208)	(0.00186)
Smokers	-3.64e-06	-0.00247*	0.00361***
	(0.00152)	(0.00149)	(0.00133)
BingeDrinking	0.00387**	0.00525***	-0.000294
	(0.00173)	(0.00170)	(0.00152)
HealthyFood	-0.000853**	-0.000986**	6.48e-07
	(0.000405)	(0.000399)	(0.000356)
Poverty	0.0175***	0.00529***	0.0200***
	(0.00167)	(0.00164)	(0.00146)
METRO	-0.187***	-0.207***	-0.0686***
	(0.0176)	(0.0173)	(0.0155)
Unemployed	-0.555	-2.085***	-2.052***
	(0.707)	(0.696)	(0.622)
Toxic_Chem	-3.774*	-3.996**	-1.273
	(1.951)	(1.921)	(1.715)
College	-0.00234**	-0.00324***	-0.00134
	(0.000989)	(0.000974)	(0.000870)
Farm	0.703***	0.761***	0.413***
	(0.0791)	(0.0779)	(0.0695)
Black	-0.00276***	0.000963	-0.00355***
	(0.000630)	(0.000621)	(0.000554)
Hispanic	-0.00306***	-0.00466***	-0.00933***
	(0.000786)	(0.000774)	(0.000691)
Constant	6.368***	3.562***	5.010***
	(0.172)	(0.169)	(0.151)
Observations	2,078	2,078	2,078
R-squared	0.332	0.262	0.379

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

Table A4: Divisional (USCB) Constants Regressions for CHD, Lung Cancer, and Colon Cancer

	(1)	(2)	(3)
VARIABLES	lCHD	lCol_Can	lLung_Can
FRUVEG Consumption	-0.00230***	-0.00175**	-0.00314***
	(0.000699)	(0.000684)	(0.000618)
FAT Consumption	0.00992*	0.00986**	-0.00482
	(0.00508)	(0.00497)	(0.00449)
MEAT Consumption	0.00423***	0.00347***	0.00556***
	(0.000960)	(0.000940)	(0.000848)
Physical Acitivty	-0.0200***	-0.00337	-0.0110***
	(0.00256)	(0.00251)	(0.00226)
Smokers	0.00111	0.000165	0.00417***
	(0.00155)	(0.00151)	(0.00137)
BingeDrinking	0.00244	0.00248	-0.000802
	(0.00177)	(0.00173)	(0.00156)
HealthyFood	-0.000714*	-0.000584	4.13e-05
	(0.000409)	(0.000400)	(0.000361)
Poverty	0.0184***	0.00691***	0.0206***
	(0.00168)	(0.00165)	(0.00149)
METRO	-0.187***	-0.199***	-0.0739***
	(0.0174)	(0.0170)	(0.0154)
Unemployed	-0.0136***	-0.0213***	-0.00762**
	(0.00420)	(0.00411)	(0.00371)
Toxic_Chem	-3.927**	-4.438**	-1.736
	(1.942)	(1.902)	(1.716)
College	-0.00278***	-0.00343***	-0.00169*
	(0.00104)	(0.00102)	(0.000916)
Farm	0.736***	0.823***	0.466***
	(0.0817)	(0.0800)	(0.0722)
Black	-0.00238***	0.00129**	-0.00365***
	(0.000654)	(0.000640)	(0.000578)
Hispanic	-0.00283***	-0.00449***	-0.00938***
	(0.000786)	(0.000769)	(0.000694)
div1	6.414***	3.398***	4.886***
	(0.206)	(0.201)	(0.182)
div2	6.405***	3.378***	4.944***
	(0.208)	(0.204)	(0.184)
div3	6.509***	3.444***	5.013***
	(0.205)	(0.201)	(0.182)
div4	6.495***	3.508***	5.016***

	(0.200)	(0.196)	(0.177)
div5	6.477***	3.404***	5.024***
	(0.205)	(0.201)	(0.181)
div6	6.408***	3.448***	4.994***
	(0.193)	(0.189)	(0.171)
div7	6.518***	3.438***	5.017***
	(0.195)	(0.191)	(0.172)
div8	6.469***	3.436***	4.971***
	(0.208)	(0.203)	(0.184)
div9	6.425***	3.296***	4.925***
	(0.204)	(0.200)	(0.180)
Observations	2,078	2,078	2,078
R-squared	0.996	0.990	0.996

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