Essays in Health Economics: Empirical Studies on Determinants of Health

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

By

Kenneth J. Lee
Executive Masters in Business Administration
University of Pittsburgh, 1995
Bachelor of Science, Chemistry
Carleton College, 1977

Director: Robin Hanson, Professor
Department of Economics

Spring Semester 2011
George Mason University
Fairfax, VA

Copyright: 2011 Kenneth J. Lee
All Rights Reserved

Dedication

This is dedicated to my loving and generous wife, Janice. Thank you for sharing this little trek and our life-long journey.

Acknowledgements

I would like to thank the members of my dissertation committee, Dr. David Levy, Dr. Jack Hadley, and, in particular, Dr. Robin Hanson, for their wise guidance and assistance throughout my studies and the development of this dissertation. Dr. Levy made time for innumerable discussions and offered critical observations on the breadth and depth of econometrics and economics in general. Dr. Hadley provided continuing guidance during my dissertation research and key insights into practical empirical approaches that helped improve the quality of the empirical studies. Dr. Hanson provided leadership, deep insights on, and thoughtful consideration into the vagaries and intricacies of health economics research. Without his help and wise guidance I might still be navigating blindly the ocean of opportunities available for relevant and interesting research.

With great love, I want to thank my wife, Janice. She has endured as much as I have, lived lonely while I toiled in the libraries, and rarely complained. I would have been unable to start and finish this endeavor without her. And I must mention our dogs, Maggie, Zoey, Annie, Jack, and baby Gracie. Often a joy, occasionally a trial, always ready to welcome me home unconditionally. Sometimes you just need to pet a dog.

Table of Contents

Page

List of Tables

List of Figures

Abstract

ESSAYS IN HEALTH ECONOMICS: EMPIRICAL STUDIES ON DETERMINANTS OF HEALTH

Kenneth J. Lee, PhD

George Mason University, 2011

Dissertation Director: Dr. Robin D. Hanson

This dissertation describes results of empirical studies addressing important issues in the field of health economics, one of the fastest-growing fields within economics. The investigated problems include two major topic areas: aggregate health determinant effects on health and individual health determinants effects on health.

For the aggregate study, this dissertation extends current research by including detailed health expenditure data from the Centers for Medicare & Medicaid Services (CMS) at the Department of Health and Human Services; using instrumental variables techniques to reduce the likelihood of cross correlation between expenditure and health outcome variables; and defining a set of state-level factor variables that provide an incisive look into differing state characteristics. The empirical results indicate a consistent negative impact of aggregate health expenditure on all-cause mortality. Income elasticity results indicate that health is not a luxury good

The focus of the individual study involves relationships between geography and health, occupation and health, and the interaction effects between geography and occupation on health. This study uses data defined within the survey of choice, the National Longitudinal Mortality Study (NLMS), for location of birth and standard occupations; and uses occupation variables and state-level characteristic variables, which were both defined through factor analyses. In particular, the race data show consistently worse health for black men and women relative to whites. Being female is always more healthy than being male. Living in rural areas (and suburban areas) is better for health than living in urban areas. Health improves as the amount of education and income rise.

In addition, this study considers the impact of occupation category groupings on health and uses the results of an occupation factor analysis to define job characteristics. Traits related to "job IQ," for example, creativity and cognitive ability, show consistent, significant, and positive impacts on health even with a variety of confounding variables, suggesting that job IQ is fundamental to explaining the impact of occupations on health.

1. Introduction

This dissertation describes results of empirical studies addressing important issues in the field of health economics, one of the fastest-growing fields within economics. The investigated problems include two major topic areas: aggregate health determinant effects on health and individual health determinants effects on health. For the aggregate study, this dissertation extends current research by including detailed health expenditure data from the Centers for Medicare & Medicaid Services (CMS) at the Department of Health and Human Services; using instrumental variables techniques to reduce the likelihood of cross correlation between expenditure and health outcome variables; and defining a set of state-level factor variables that provide an incisive look into differing state characteristics. The focus of the individual study involves relationships between geography and health, occupation and health, and the interaction effects between geography and occupation on health. This dissertation uses data defined within the survey of choice, the National Longitudinal Mortality Study (NLMS), for location of birth and standard occupations; and uses occupation factor variables and state-level factor variables, which were both defined through factor analyses. All analyses in this dissertation extend the literature on the relationship between key determinants and health outcomes, and should be highly relevant to health researchers as well as policy makers, and health care providers.

Chapter 2 reports results of the relationship between health care determinants, including aggregate health care expenditures, and health outcomes based on annual data for the 50 U.S. states (and the District of Columbia) covering 28 years, from 1980–2007. The analysis of the relationships and outcomes consider expenditure data at multiple levels of detail, namely, national health care expenditures based on the location of the provider, national health care expenditures based on the location of the patient, and pharmaceutical and non-drug-related expenditures. Other studies relating health expenditures to health outcomes are affected by the heterogeneity of cross-country data, or the use of analytical techniques that do not account for simultaneous equation bias and endogeneity, omitted variable bias, and the lag between expenditures and outcomes. These issues are addressed in this dissertation using instrumental variables, a wide variety of relevant dependent variables, fixed effects, and panel data.

Chapter 3 reports results of individual health care determinants on a range of health outcomes using data from the NLMS. The study explores the combination of (a) multiple socioeconomic variables on health outcomes through interaction effects, and (b) the use of geographic location variables at multiple levels of detail (Census Region, Census Division, and State). By incorporating the identification of state-level characteristics through a factor analysis of state demographic and ranking variables, this study provides an alternative geographic context for analysis in the manner of Weiss (Weiss 2000).

Chapter 4 explores a deeper investigation of the NLMS data by adding the impact of occupation on health. Occupations are defined at multiple levels of detail, including: detailed occupation (total of 807 distinct occupations); gender-specific recoded occupation groups (total of 88 occupations for men, and 59 occupations for women); a group of 18 major occupation category groupings; and the British Registry General (BRG) groupings, which represent a set of four gender-specific high-level groups. In addition, 225 occupation characteristics were collected from the Occupational Information Network (O*NET) database for each of the 807 detailed occupations, and factor analyses were performed to determine reduced sets of factors representative of occupations. These factors were then combined with the multiple geographic variables, and the state-level factor variables from chapter 2 to investigate the interaction effects on health outcomes. The application of occupation factors that describe the innate characteristics of job abilities, knowledge, skills, work styles, and so on, is unique in the investigation of determinant impacts on health. The use of the state-level factors provides groupings of states that are related through a diverse set of demographic, health, and cultural characteristics, providing a richer alternative to standard geographic groupings.

Chapter 5 summarizes the major empirical findings and briefly discusses the conclusions.

2. U.S. State Aggregate Health Care Determinants and Health Outcomes

2.1 Introduction

This chapter examines the relationship between health care determinants, including health care expenditures, and health outcomes based on aggregated annual data for the 50 U.S. states (and the District of Columbia) covering 28 years, from 1980–2007. The approach generally follows that used in previous studies on the Canadian provinces (Cremieux, Meilleur, et al. 2005; Cremieux, Ouellette, and Pilon 1999) and English program data (Martin, Rice, and PC Smith 2008), including the use of Instrumental Variable (IV) in two-stage least squares (2SLS) analyses to account for potential correlation between expenditures and outcomes. The analyses consider the relationship of expenditures to outcomes controlling for other economic, socio-demographic, and lifestyle factors that may have an impact on health.

The results are generated using U.S. state total health care expenditures and a detailed breakout of state health care expenditures as defined by the U.S. Centers for Medicare & Medicaid Services (CMS). Using this data, I demonstrate a generally negative relationship between higher health spending and better health outcomes. Simulations using bootstrap and jackknife techniques validate the choice of instruments used in the 2SLS analyses, and the negative impact of health expenditures on outcomes.

The CMS detailed category of drug spending has a generally positive impact while non-drug spending has a generally negative impact.

The chapter is organized as follows. First, the background section introduces the literature on health care determinants, including Grossman's theory of health care production. Next, research questions and hypotheses are discussed. The data and the methodologies used in the empirical analyses are then introduced. This is followed by a detailed discussion of the analytical results. Finally, concluding remarks are presented, together with a brief discussion on possible directions for future research.

2.2 Background

There is a vast literature that relates health care determinants to health outcomes within and between countries. The papers referenced below are some key representative papers related to this dissertation.

Studies on the relationship between aggregate health care spending, potential determinants, and health outcomes have provided varied results. Cross-country studies using Organization for Economic Cooperation and Development (OECD) countries have addressed health care spending and the impact of national Gross Domestic Product (GDP) on health outcomes (Barros 1998; L DiMatteo and R DiMatteo 1998; Bac and Pen 2002; Huber and Orosz 2003; Ariste and Carr 2003; Sen 2005) by relating health system characteristics (e.g., population aging, type of health system, and existence of gatekeepers) considered significant to health outcomes (Barros 1998; Or 2000). These studies consider pooled OECD country data and individual country data, and are focused on investigating the GDP (or income) elasticity with respect to health care expenditures.

The use of confounding variables is minimal in the majority of these studies; the focus is instead on the income-health expenditure relationship. This approach, however, may lead to omitted variable bias.

Other studies have considered the impact of aggregate-level health care determinants, including income, on cross-country health outcomes (Gravelle and Backhouse 1987; Pritchett and Summers 1996; Or 2000; Macinko, Starfield, and Shi 2003; Connelly and Doessel 2004; Shaw, Horrace, and Vogel 2005; Arah et al. 2005; Gerdtham and Ruhm 2005; Rajkumar and Swaroop 2008; Biggs et al. 2010). Gravelle and Backhouse used regression analyses to analyze the impact of cross-sectional international data on mortality rates and described the key statistical issues of concern in these analyses. Pritchitt and Summers used time series data on health (infant mortality and life expectancy) and income and determined the income elasticity of infant mortality lies between -0.2 and -0.4. Or, Macinko et al.; Shaw et al.; Arah et al.; and Gerdtham and Ruhm used OECD data with a variety of explanatory variables, including medical system variables, environmental factors, primary care system definitions, pharmaceutical consumption, lifestyle variables, macroeconomic conditions on mortality when labor markets strengthen, and poverty data. Each of these studies uses a variant of ordinary least squares (OLS) on a pooled data sample, with fixed location effects that control for factors that differ across locations but are time invariant. None, however, control for possible endogeneity using techniques like 2-stage least squares (2SLS). Connelly and Doessel use Australian Census data and detect a strong and statistically significant positive impact of medical expenditure on health status. Rajkumar and Swaroop use

World Bank development data for 91 countries, including public health spending, two indices of governance, and the Gini index to determine that, in the presence of good governance, increased health expenditure results in improved infant mortality. Biggs et al. find generally better health results with increases in GDP.

Several papers have been published on determinants in Canadian health outcomes based on provincial data (Cremieux, Ouellette, and Pilon 1999; Cremieux, Jarvinen, et al. 2005; Cremieux, Meilleur, et al. 2005), and at least one focusing on English Primary Care Trust (PCT) geographic areas (Martin, Rice, and PC Smith 2008). The papers show positive health impacts with higher spending, but the results have methodological and data limitations.

Various approaches have been used to consider United States health outcomes and health care expenditures, including focusing on aggregate mortality rates (Auster, Leveson, and Sarachek 1969); regional variation (Fisher, D Wennberg, Stukel, Gottlieb, Lucas, and Pinder 2003a; Fisher, D Wennberg, Stukel, Gottlieb, Lucas, and Pinder 2003b; Dartmouth Team 2010); health system characteristics (Bodenheimer 2005a; Bodenheimer 2005b; Bodenheimer 2005c; Bodenheimer and Fernandez 2005); health survey data (Berk and Monheit 2001); individual mortality data (Sorlie, Backlund, and Keller 1995), race-based data (Murray et al. 2006), and county-level results ((Hadley 1982a). Although the results are mixed, they generally do not show that health spending has significant impacts on outcomes.

Cross-country studies suffer from inherent heterogeneity. Health care measures have definitional and methodological differences. For example, measurements of health care spending between countries are likely to have different bases for data collection, collation, or reporting (Gravelle and Backhouse 1987; Torgerson and Maynard 1998). A variety of characteristics, such as individual characteristics; environmental, cultural, or geographic characteristics; and lifestyle characteristics, all may complicate these analyses. The health care institutional systems in countries differ significantly as well. The nature of these systems, including access to physicians, availability of services, and government intervention, impacts health care spending. Cremieux addressed many of the limitations of cross-country data analyses by focusing on Canadian provincial data. Martin did the same for English PCT area data, and Hadley focused on U.S. counties. The analyses in this chapter use U.S. state data to overcome many of the same limitations due to the similarity of approaches, data definitions, and reporting requirements across the U.S. states.

Methodological difficulties associated with empirical investigations of the determinants of mortality abound, including simultaneous equation bias and endogeneity, omitted variable bias, and the lag between expenditures and outcomes. The Methodology description in subsection 2.5 of this chapter addresses this dissertation's approach to these issues.

In the United States, the health care financial burden is shared between public expenditures (e.g., states and federal government agencies) and private expenditures. There has been a regular increase in annual average, real, total health care spending per capita of about 8.1% between 1975 and 2005. In contrast, real GDP growth over the same period has averaged about 3.2%. Life expectancy at birth (for all races and genders) over the same period rose from about 72.6 to about 77.8 years of age – an average improvement of about two months per year. A positive relationship between health spending and health outcomes could simply reflect greater per capita health spending in healthier states, or the growing parallel trends in health and health spending.

2.2.1 Theoretical Framework

Grossman proposed the first formal model of the determinants of health (Grossman 1972a; Grossman 1972b; Grossman 1999; Grossman 2000). Grossman defined health as a durable capital stock that is inherited and depreciates over time. Health is an endogenous variable that people can improve through investment in medical care, diet, and exercise. Besides the production of health, the model also supports the depreciation of health capital, i.e., individuals age and may choose to invest in products with negative marginal health benefits. Individuals are assumed to invest in health production until the marginal cost of health production equals the marginal benefit of improved health status. The Grossman model provides an economic framework for the relationship between inputs, such as education, income, nutrition, health care, and other environmental or socioeconomic variables, which influence the production of health that can be measured in terms of health status.

A number of different formulations of Grossman's original model equations for the demand for health and the demand for medical care have been published. One production function relevant to this study is (Grossman 1972a):

(1) $$lnH_i = \alpha lnM_i + r_H E - \widetilde{\delta_i} i - ln\delta_0$$

where H_i is the stock of health for individual i, M_i is medical care, E is education, and the last two terms represent depreciation rate terms. The above equation should not be fitted by ordinary least squares (OLS) since $ln\ M_i$ and $ln\ \delta_0$ are likely correlated. Two-stage least squares can be used by first fitting the demand curve for medical care (equation 4-7' from (Grossman 1972a)):

(2) $$lnM_i = B_{WM} lnW + B_{EM} E + B_{iM} i + U_2$$

then using the predicted values of $ln\ M_i$ to estimate the production function. In eqn (2), W is the wage rate, U_2 is a disturbance term, and the model predicts $B_{WM} > 0$, $B_{EM} < 0$, and $B_{iM} > 0$.

Although the basic Grossman functions may seem rather meager in the number of variables driving medical care or health, Grossman interpreted the variables shown to represent a wide array of market goods and factors—often driven by the data available in a particular data set or by reasonable proxies for the named variables. For example, Grossman says in footnote 3 (Grossman 1972b), "…medical care is not the only market good in the gross investment function, for inputs such as housing, diet, recreation, cigarette smoking, and alcohol consumption influence one's level of health." Education is a specifically named variable and Grossman's work emphasizes the importance of education in health production (Grossman 1973; Grossman 2000; Grossman 2005). The stock-of-health-dependent variable has most often been represented by self-reported health, age-adjusted death rates, life expectancy, or infant mortality. Age is a key depreciation factor; in Grossman's model, health capital depreciates with age. Wages may be known for individuals, but more often income (or family income) is used as a proxy for wages. In aggregate studies, income per capita, GDP per capita, or Gross State Product (GSP) per capita are common substitutes for wages.

2.2.2 Empirical Literature Review

Grossman's model was defined at the microeconometric level. Numerous empirical studies have used individual data (see chapter 3 of this dissertation). Other studies have used aggregate data at the U.S. county level, U.S. state level, Canadian province level, or English Primary Care Trust (PCT) level (Grossman 1972a; Corman, Joyce, and Grossman 1987; Hadley 1982a; Hadley 1988; Thornton 2002; Martin, Rice, and PC Smith 2008; Dartmouth Team 2010; Cremieux, Ouellette, and Pilon 1999; Auster, Leveson, and Sarachek 1969).

Auster et al. reported empirical results using 2SLS on cross-sectional data for 1960 and found evidence that medical care reduced age-adjusted state-level death rates (Auster, Leveson, and Sarachek 1969) while controlling for income, education, Standard Metropolitan Statistical Area (SMSA) percentage, manufacturing percentage, alcohol consumption, cigarette consumption, race, and presence of a medical school.

Grossman's early empirical results (Grossman 1972a) use restricted activity days, work loss days, and self-reported health for stock of health proxies, and personal medical outlay is used as the dependent variable in the demand for medical care. The independent variables are age, education, gender, weekly wage rate, family income, and family size. In 2SLS analyses, the elasticity of health stock with respect to medical care outlays is positive and about 0.2, but is significant with only one of the dependent variables—self-reported health.

In one study, Hadley investigated aggregate impacts using county-level Medicare expenditure data (Hadley 1982a) and age-gender-race specific categories of 45-plus year olds. For all-cause mortality rates, Hadley shows that, for all categories, increased medical care expenditures reduce mortality. In another study, Hadley (Hadley 1988) found that greater county-level Medicare spending per beneficiary resulted in significantly lower all-cause mortality rates for all age groups, races, and both genders. In a recent communication, Hadley et al. (Hadley et al. 2011) finds that greater medical spending is associated with better health status of Medicare beneficiaries. Cremieux used panel data for Canadian provinces for 1978–1992 and found that higher health care spending improved outcomes (Cremieux, Ouellette, and Pilon 1999) while controlling for gender, race, physicians per capita, income, education, population density, poverty percentage, alcohol and tobacco consumption, and nutritional intake. The Cremieux study used ordinary least squares (OLS), however, which does not account for the potential endogeneity of health spending.

Thornton used cross-sectional state-level data for 1990 with the age-adjusted death rate as the dependent variable (Thornton 2002). Using 2SLS, the estimated coefficient on medical care expenditures was negative and not significant, while controlling for income, education, alcohol and tobacco consumption, urbanization, marital status, crime rates, and degree of manufacturing. Thornton claims that the marginal contribution of medical care utilization in lowering mortality is quite small. Martin et al. use cross-sectional data for FY2004 from PCT areas—geographic local health areas within England (Martin, Rice, and PC Smith 2008). By focusing on health spending for two programs of care—cancer and circulatory problems—and using 2SLS, Martin et al. find a strong positive impact of health care expenditures on outcomes. Although their theoretical model discussion refers to clinical and environmental factors relevant to the analysis, they only use a minimal set of variables presumably due to lack of available data. Rothberg et al. (Rothberg et al. 2010) find little correlation between reduced mortality for certain conditions and increased spending on patients with those conditions. In particular, chronic obstructive pulmonary disease and sepsis are two conditions for which increases in spending have not translated into improvements in outcomes.

2.3 Research Questions

The major goal of this chapter is to investigate determinants of health outcomes, with an emphasis on particular health outcomes at the U.S. state level using detailed health expenditure data from the Centers for Medicaid & Medicaid Services. Other determinants are considered in the empirical analyses, including education, income, poverty levels, gender, race, and public choice variables representing the makeup of state legislatures and the extent of citizen voting. Table 1 shows the major research questions and the corresponding predicted responses investigated in this chapter.

2.4 Data

The 50 U.S. states are the geographic units for the analysis in this chapter. There is less detail using state-level data than with a smaller defined geographic region, but as with many studies, data availability for both the specific variables of interest and for the span of years of interest was the key driving factor in the choice of geographic unit.

Aggregating to the state level likely masks some interesting detail about Census areas, counties, zip code areas, neighborhoods, and individuals. Hadley et al. (Hadley et al. 2006b) claim that analyses using individual level data should be consistent with area-level analyses to validate the latter. If they are not consistent, and if the individual level analyses are done rigorously, then the individual analyses should be preferred.

Table 1: Major Research Questions and Predicted Responses Investigated in Chapter 2

Research Question	Predicted Response
1. What is the impact of endogenous health expenditure data on U.S. state-level health using a Grossman-type model analysis approach?	Health expenditures have a positive and significant effect on health outcomes
2. What is the impact of detailed versus aggregate health expenditures on health?	Detailed expenditure impacts are a breakout of the aggregate impacts; some being significant some not
3. What are the impacts of socioeconomic status characteristics on health?	Socioeconomic status (SES) factors will impact health, e.g., greater amounts of income and education will have positive impacts
4. What is the influence of demographic characteristics on health?	Demographic factors will impact health, e.g., alcohol and cigarette consumption should have negative health impacts
5. What are the impacts of geographic location on health?	Geographic variation is expected to have an impact on health, for example, rural living has been shown to be healthier than urban living. Impacts are likely to vary by state.
6. What is the income elasticity with respect to health care expenditure? Is health care a luxury good or not?	Studies that show the income elasticity > 1 are likely affected by omitted variable bias. More complete specifications generally show the elasticity < 1.

Fisher et al. (Fisher, D Wennberg, Stukel, Gottlieb, Lucas, and Pinder 2003a; Fisher, D Wennberg, Stukel, Gottlieb, Lucas, and Pinder 2003b) and Chandra et al. (Chandra, Fisher, and Skinner 2007) use regional aggregate data and individual data to show nearly identical results. Interstate heterogeneity is moderated somewhat by the use of geographic dummy variables that account for invariant characteristics of the states.

2.4.1 Data Sources

The empirical analyses in this dissertation are conducted using a constructed composite database of health care data representing the 50 U.S. states and the District of Columbia, collected from multiple sources for the years 1980–2007. After preliminary analysis, the District of Columbia was determined to have significant outliers in many of the data, e.g., physicians per capita, population density, and infant mortality. These outliers were significant enough to have a direct impact on the economic significance of these variables. One approach to resolving this data issue would be the use of a robust regression technique that performs a weighted least squares analysis. In these analyses, the data with the largest residuals receive a lower weighting factor and contribute less to the estimates. Stata has such an ordinary least squares alternative, but it is not applicable to panel data. As a result, for a more consistent reporting of results, the District of Columbia data was combined with the data for Maryland and Virginia to create a "new" state called DMV. As there is significant mobility in the greater metropolitan area of Washington, DC, northern Virginia, and mid-state Maryland for work, social interaction, health care, and education, the combining of this data was determined to be reasonable. The data for the new DMV state replaces that for the District of Columbia, Maryland, and Virginia, resulting in a total of 49 U.S. states used in most of the analyses in this chapter.

Appendix 1 contains a list of the data variables available and the sources of the data. The multiple data sources include the Centers for Medicare & Medicaid Services, Centers for Disease Control and Prevention (CDC), U.S. Department of Agriculture, Bureau of Labor Statistics (BLS), U.S. Census Bureau, non-governmental organizations such as The Tax Foundation, and individual state health organization web sites.

2.4.2 Sample Construction

The analyses presented in this chapter are based on state-level aggregate data from numerous sources. Each data set was collected by year and by state. National collections, like the CDC's National Center for Health Statistics (NCHS), have collected state data for years and were used as the primary baseline.

These data sets were augmented as necessary by data located at each state's Office of Public Health web site. In the case of the U.S. Census Bureau, extensive data are available for each Decennial Census. More limited data are available for the intervening years and for the 2000s; many of the Census Bureau data are projections rather than measured data. For the few data that are missing, linear extrapolation is used to insert the missing values. Overall, imputed data makes up less than 0.3% of all the data used.

In this chapter, most analyses use log-transformed variables. The transformation achieves two things. First, most variables have a right-skewed distribution with a long tail and some have exhibited non-linear relationships with the dependent variables. Transformation captures the non-linearity and also produces data with more normalized distributions. Second, the transformation results in elasticities directly, allowing comparison to previous studies.

2.4.3 Dependent Variables

The primary dependent variables are all-cause age-adjusted state mortality rates and cause-specific age-adjusted state mortality rates. Life expectancy at birth by gender and infant mortality are also available. This chapter focuses on mortality rates as such data are more consistently defined, measured, and accepted in the health economics literature. Lack of data for race-based infant mortality, race-based life expectancy at birth, and life expectancy at 65 years of age, for the entire time span and for all states, prevented the use of these health measures in the analyses.

2.4.4 Explanatory Variables

Braveman et al. (Braveman et al. 2005) consider socioeconomic status variables used in health studies and recommend an outcome- and social group-specific approach to SES measurement and data use. Their recommendations include:

- Education and income are not interchangeable, so both should be used for outcomes research. This dissertation uses both education and income throughout.
- Occupation categories in the U.S. do not appear to be meaningful measures of SES. In Western Europe, occupational categories are based on prestige, skills, social influence, and/or power. Studies have found strong relationships with these classified occupations and health outcomes. Chapter 4 uses occupation characteristics and factor analyses to overcome these limitations; in addition, occupations are ranked by prestige score.

- Neighborhood characteristics can influence health; few studies measure both SES measures and neighborhood features. This dissertation uses geographic features at a number of different levels of detail.
- A given SES measure may have different meanings in different social groups, including varying meanings across age, gender, race, and urban/rural locations. This dissertation uses these dimensions consistently throughout.

The main explanatory variables used in this chapter include:

- **Health expenditure per capita data:** gathered from CMS at the Department of Health and Human Services, including expenditures by providing state and by resident state
- **SES data:** income per capita, high school graduation percentage, college graduation percentage
- **Demographic data:** poverty rate, population density, gender, race, and age
- **Lifestyle data:** smoking consumption, alcohol consumption

A variety of other data are collected and available for sensitivity analyses. For example, the data set contains (see subsection 2.6 for characteristics of these data):

- **Health system variables:** physicians per capita, hospital beds per capita, and hospitals per capita
- **Population data:** male/female percentages; black/white percentages; percentage with private insurance; percentage on Medicare and Medicaid; percentage uninsured

- **Economic data:** Gross State Product (GSP) per capita; percentage of health care expenditure (HCE) considered public expenditures; percentage of HCE considered private expenditures; percentage of HCE spent on Medicare; percentage of HCE spent on Medicaid; gross state debt as a percentage of GSP; gross state debt per capita; Coincident Index (a measure created by the Philadelphia Federal Reserve to indicate state-level economic conditions), and Gini index (a measure of income inequality)
- **Demographic data:** unemployment rate and urban percentage
- **Public Choice data:** registered voters by gender and race; percentage of registered voters voting in previous nationwide election by gender and race; presence of women governors; red state indicator (whether the state voted Republican in the last federal election); percentage of black legislators in State Senate and House; and percentage of women legislators in State Senate and House

Other potential data – for example, nutritional health measures such as per capita expenditures on meat and fat products, or the amount of exercise per person – have not been reported on a per-state basis in the U.S. consistently over the last 30 years. As a consequence, although diet and exercise have a large impact on health there is insufficient data available to allow analyses of these possible determinants. The state level factor analysis, described in subsection 2.5.3, contains some of these variables captured as rankings. The derived factors take these considerations into account.

2.5 Methodology

This section introduces the statistical approaches used and the analytical models estimated. The Stata statistical analysis package (Stata, Version 11.1 2010) was used with all data sets. For instrumental variable analysis, the user-supplied package XTIVREG2 is employed (Schaffer 2007). For formatting the regression tables, the user-supplied package OUTREG2 is used (Wada 2010). The Stata data files (*.dta) and analysis processing files (*.do) are available by request from the author.

2.5.1 Panel-Corrected Standard Error Analysis

The initial approach used is a pooled least squares method correcting for both autocorrelation and heteroskedasticity in the panel data. The basic specification estimated is:

$$lnH_{it} = \alpha_0 + \beta_i lnM_{it} + \gamma_i lnX_{it} + S_i + Y_t + \varepsilon_{it}$$

where H_{it} is the health outcome proxy in state i in year t; M_{it} is the medical care expenditure in state i in year t; X_{it} is a vector of economic, socio-demographic, and lifestyle factors; α_0 is the intercept; S_i is vector of state fixed effects; Y_t is a vector of year fixed effects; and ε_{it} is a disturbance term. The state fixed-effect variables account for potential systematic time invariant differences between states that are not captured by the included control variables; the year fixed-effect variables account for potential state invariant differences that vary over time. Including both fixed effects result in attenuated and less significant outcomes in the analyses in subsection 2.7.

An alternative to using year fixed effects is the use of a time trend. Sensitivity analyses using a time trend resulted in similar coefficient values but consistently smaller standard errors. A trend variable allows for consistent linear data movement either up or down over time; fixed effects allow for both rises and falls over the span of years. To avoid any bias caused by the limitations of a time trend, year fixed effects are used throughout.

Previous analyses have used a variety of econometric approaches. Cremieux, for example, consistently uses generalized least squares (GLS) approaches to health determinant analyses. Beck has shown, however, that this approach produces standard errors with a high level of overconfidence (Beck and Katz 1995; Beck 2001) and recommends using OLS with panel corrected standard errors (PCSE). For the studies conducted in this chapter, the OLS reported results for the U.S. data sets were determined using PCSE to create a baseline for comparison to prior studies and to the 2SLS analyses[1].

2.5.2 Two-Stage Least Squares Analysis

As noted earlier in Section 2.3, there is likely correlation between medical care expenditure variables and health outcome variables requiring an analytical approach that accounts for this bias. For these analyses, I employ an instrumental variable (IV) approach using 2SLS. Valid instruments need to satisfy three conditions. First, they must be correlated with medical care expenditures. Second, they must be uncorrelated with the error term in the primary regression equation. Third, the instruments should not

---

[1] Cremieux's data was obtained from the author. Replication analyses verified the validity of using OLS with PCSE versus GLS.

be correlated with the dependent health outcome variables, except through the expenditure variables.

There are a number of potential instruments available, including economic variables, medical cost variables, population variables, and public choice variables. The choice of instruments for the analyses is described in subsection 2.7. The 2SLS are run with (1) the *robust* option that provides standard errors (SE) robust to the presence of arbitrary heteroskedasticity, and (2) the *cluster(State)* option, providing clustered SEs and statistics robust to intra-group autocorrelation.

2.5.3 State Factor Analysis

State-level rankings for a variety of cultural variables were identified and combined into a dataset representing 36 state characteristic variables. These characteristics, listed in Appendix F, include such rankings as Healthiest State; Teen Birth Rate; Fruit portions eaten per capita per day; Pollution levels; Wal-Mart stores per capita; Starbucks per capita; Manufacturing Output as a percentage of state GSP; Happiness Index; Freedom Index; overall Tax Burden; etc. In each case, the ordering of the rankings was adjusted so that the "better" score has the lower rank. For example, the least Obese state is ranked one (1) and the most Obese state is ranked fifty (50); the state with the highest rate of Church Attendance is ranked one (1), and the state with the lowest Church Attendance is ranked fifty (50); and so on. These 36 state characteristic variables are combined with 19 state demographic variables, including latitude and longitude, density (population per square mile), percentage of white population, and percentage of black population.

The factor analysis approach follows the same methodology as described in Appendix D for the occupation factor analysis used in Chapter 4. The 36 state characteristics were analyzed to gain a better understanding of the relationships among the variables. The 19 demographic variables were analyzed to determine their relationships. The overall set of 55 variables was analyzed and these results are used in subsequent analyses in this chapter, as well as in chapters 3 and 4. This overall factor analysis results in a set of four (4) factors. The four factors explain 70% of the total variance, with Factor 1 accounting for 32%, Factor 2 accounting for 17%, Factor 3 accounting for 13%, and Factor 4 accounting for 8%. Appendix F has more detail on the four factors extracted from the complete data set. Full details of this factor analysis and sample results are available upon request from the author.

2.6 Study Sample Characteristics

This study focuses on U.S. states over a 28-year period (1980–2007). The analyses of the relationship between health outcomes and health care expenditures are conducted for age-adjusted mortality rates and cause-specific mortality rates. State population data serve as empirical weights to account for the relative size of the various states. With certain variables, limited data availability restricts the time periods to less than the default of 28 years.

To identify the relationship between health care spending and health outcomes, it is necessary to account for economic, social-demographic, lifestyle, and geographical data heterogeneity across reporting regions. Previous studies, primarily based on OECD data, focus on countries as the reporting regions (Hansen and A King 1996; G Anderson et al. 2000; Or 2000; G Anderson et al. 2003; Huber and Orosz 2003; Gerdtham and Ruhm 2005). Cremieux focused on Canadian provincial data (Cremieux, Ouellette, and Pilon 1999; Cremieux, Jarvinen, et al. 2005). Following Cremieux, the use of U.S. state data eliminates the inherent heterogeneity found in these cross-country studies.

Studies that focus on the United States have considered regions (Fisher, D Wennberg, Stukel, Gottlieb, Lucas, and Pinder 2003a; Fisher, D Wennberg, Stukel, Gottlieb, Lucas, and Pinder 2003b); states (Auster, Leveson, and Sarachek 1969); and counties (Hadley 1982a). For consistency, the variables as well as the outcome measures must be homogenously determined and reported. The populations observed also must be similar across boundaries. Definitions and methodology differ across countries. Cremieux recognized cultural and geographic variations even across Canadian provinces, but claims that there is homogeneity within the population relative to OECD cross-country analyses; and that federally gathered statistics are based on similar methodologies. U.S. states are also a homogenous group relative to OECD countries. The wider variety of data sources necessary to populate this study for U.S. states might impact the consistency of the data; however, data definitions are typically consistent state-by-state and each category of data is drawn from similar sources, e.g., federal sources for health expenditures, poverty, and unemployment. Also, most national-level reporting (e.g., within CDC and CMS databases) are based on data reported by the states to the agencies following standardized reporting methods defined for the nation as a whole.

2.6.1 Health Care and Economic Variables

Health care expenditures may be an indicator of citizen health. Although several studies indicate that the level of expenditures has minimal impact on overall health, those studies generally compare health in one region and health in another region. In separate studies, Fisher and Wennberg found that quality of care in higher-spending regions was no better, access to care was no better, and there was no difference in patient satisfaction. Rather, regional differences are largely explained by practice variation across regions (J Wennberg, Fisher, and Skinner 2002; Fisher, D Wennberg, Stukel, Gottlieb, Lucas, and Pinder 2003a; Fisher, D Wennberg, Stukel, Gottlieb, Lucas, and Pinder 2003b; J Wennberg 2004). Similarly, Skinner and Wennberg found no difference in mortality rates following acute myocardial infarction, stroke, or gastrointestinal bleeding events in Miami and Minneapolis despite twice the level of Medicare expenditure per capita in Miami and nearly four times the number of Intensive Care Unit (ICU) days (Skinner and J Wennberg 1998). Fuchs coined the term “flat-of-the-curve” medicine to describe this situation where costs can rise without a corresponding increase in health benefit (Fisher, E., et al. 2003a, Fuchs, V. 2004).

National health expenditure (NHE) data are available from CMS and account for both expenditures by state of residence and by state of provider. All spending data are expressed in 2005 dollars.

Other CMS data includes the percentage of state populations on Medicare and Medicaid, the percentage of public and private expenditures (including out-of-pocket expenses), and the percentage of Medicare and Medicaid expenditures. The U.S. Census

Bureau provides data on the percentage of the state population that has private insurance and the percentage that is uninsured. Each of these data can be used in conjunction with the NHE data or as an alternative measure.

Access to care can be a determinant of health. One measure of access is the number of physicians per capita. Presumably, more physicians per capita will lead to a higher availability to health care. It may also lead to higher level of visits per capita due to competition among physicians for the patient population. The number of physician visits is not part of the data set. Higher availability or higher usage both might be expected to increase health outcomes. Other supply-side factors, such as the number of hospital beds per capita, provide additional indicators of access and usage. Hospital Referral Region (HRR) evidence suggests that greater supply of services (e.g., physicians, particularly specialists, and beds) leads directly to higher medical care use (Dartmouth Team 2010).

Per capita income (represented by GDP per capita in many country studies) has been shown to be correlated with health expenditures, but correlation with improved health outcomes is still an open issue. Higher financial resources may result in higher health service availability or indicate a greater ability to pay for services not covered by insurance. Per capita state income and GSP per capita provide two indications of financial resource availability.

Three measures of state-level economic conditions are defined. The Coincident Index is generated monthly by the Philadelphia Federal Reserve Bank and combines four state-level indicators that summarize current economic conditions: nonfarm payroll employment, average hours worked in manufacturing, the unemployment rate, and wage and salary disbursements deflated by the Consumer Price Index (Philadelphia Fed 2010). The Gini Index is measure of state income inequality ranging from a value of zero where all individuals have equal income to a value of one where a single individual has all the income. The overall tax burden is the combined state-local tax burden shouldered by the residents of each state. The data available here are the state rankings (from 1–50) with the value 1 representing the state with the highest tax burden.

2.6.2 Social and Demographic Variables

Geographic and socio-demographic characteristics of populations have an impact on health. Grossman's model assumes age is a key depreciating factor for health, i.e., the stock of health capital decreases continually with age. The data set for this study contains percentages of state population in three age categories gathered from the U.S. Census Bureau web pages. Population variables are available for the following age categories: less than 24 years, 25–64 years, and 65-plus years. Each category is a percentage of the total state population. In addition, the percentages of whites, blacks, and non-whites are available. Finally, the percentages of women and men in the state populations are available.

Higher population density may positively affect health by supporting greater access to health care and lower costs. Some data indicate that greater distance from health care providers is a factor in reducing overall health. Higher density may also negatively affect health, as indicated by a higher life expectancy in rural as opposed to urban areas (Hayward and Gorman 2004); or people in poor health may be less likely to live in rural areas because they need to be closer to medical providers. The overall impact, then, of population density on health outcomes is an empirical issue. Density is determined as the population per square mile using Census Bureau data. States also have differing levels of urban versus rural areas. The percentage of urban area per state is from the U.S. Census Bureau.

The knowledge and awareness of health care and its consequences are a determinant in health. Higher education levels generally correlate with better health (Sorlie, Backlund, and Keller 1995; Lantz et al. 1998; Pincus et al. 1998; Lantz et al. 2001; Hayward and Gorman 2004; Lleras-Muney 2005; Grossman 2000). The level of degree attainment—both in high school and in college—may be an indicator of health. Cremieux limited education analyses to college graduate levels due to differences in high school characteristics across Canadian provinces (Cremieux, Ouellette, and Pilon 1999). In the U.S., common federal-level high school standards make the high school graduation rate a viable additional measure for this study.

Studies also indicate that socioeconomic conditions are indicators of health (M. Marmot et al. 1991; Ennett and Bauman 1993; Sacker et al. 2000). For example, higher employment attainment and social status both affect health positively beyond just financial considerations. Both poverty rates and unemployment levels per capita may be valid proxies for socioeconomic conditions. Poverty rate data were gathered from the Census Bureau and are based on total household income. If the household income is below a federal threshold, all family members are considered to be in poverty. Unemployment level data was gathered from the Bureau of Labor Statistics. Although data is reported monthly, December data was used as a consistent measure of the annual level of unemployment.

2.6.3 Lifestyle Variables

Behavioral characteristics are associated with health. Alcohol and tobacco consumption are two that have received extensive consideration. Some studies have found beneficial impacts to moderate alcohol use (Hummer et al. 1999; Hayward and Gorman 2004). A recent analysis has cast doubt on these earlier studies (Fillmore et al. 2006). In either case, the data used in this study does not delineate between levels of usage. The values are the total annual volume in gallons per capita (including beer, wine, and other spirits) for ages 14 and older.

Tobacco consumption is based on annual CDC surveys of adults by state who respond as current smokers.

2.6.4 Public Choice Variables

Participation in the political process may have an impact on health care spending. Some studies (Rehavi 2007; Lockhart, Green, and Giles-Sims 2010) indicate that a higher percentage of women or minorities in state legislatures has a direct relationship on the level of social spending in the state, including health care spending. Data on participation by women and minorities as state legislators were gathered from the National Conference of State Legislators web site. In addition, percentages of registered voters (white, black, men, and women) and eligible voters voting in each nationwide congressional election were gathered from the U.S. Census Bureau.

2.6.5 Nutritional Variables

Cremieux used household spending on meat products and fat products as a measure of nutritional health. In Canada, provinces have established programs to gather such data to improve nutrition. In the United States, these data are not yet gathered on a consistent basis state-by-state or at the federal level. There are some data on household expenditures and per capita consumption of various products (meat, fat, dairy, etc.); however, most of the data, as reported by the U.S. Department of Agriculture, is at the national level, not the state level. State-level organizations, such as state health departments and agriculture departments, also do not report on this data consistently. Thus, there are no results reported for nutritional variables in this dissertation.

2.7 Results and Discussion

2.7.1 Health Care Expenditure Results[2]

The CDC's National Center for Health Statistics (NCHS) maintains mortality statistics for states and counties across the United States. For use in this research as alternative dependent variables, these age-adjusted mortality statistics (causes of death) by state were downloaded and grouped into an overall All-Cause category and four (4) sub-categories consisting of Tumor-related, Cardiovascular-related, Injury-related, and

---

[2] All results shown in Chapter 2 are for Provider-based Health Care Expenditures. These state data represent health care expenditures based on the location of the provider. This means that patients who cross state boundaries for health care are counted, not in their own state expenditures, but in the state expenditures based on the location of the provider. CMS also provides health care expenditures based on the resident location of the patient. Equivalent analyses were performed with the resident-based data and the results in all cases are equivalent to those shown here, i.e., there are no significant differences in the results.

Other-cause related deaths. Table 2 shows the 2SLS analyses using these dependent variables and the baseline set of explanatory variables.

The instrument set for these 2SLS analyses consists of three variables: physicians per capita, hospital beds per capita, and CMS Dental Services expenditure per capita. The first two are medical care resource variables and should affect total health care expenditures through the volume of care used. In the first stage regressions, both variables have a significant and positive coefficient on health expenditures. The Dental Services expenditure instrument consists of services provided by dentists and dental technicians. Although dental services likely affect the quality of life, such services are unlikely to directly impact the health outcomes used in this dissertation. In the first stage regressions, this instrument has a significant and positive coefficient on health expenditures.

All instruments were subjected to tests of validity and weakness. The first-stage F-test results exceed the recommended minimum value of 10 indicating that the instruments are individually and jointly statistically significant. Hausman tests indicate that the results of the OLS/PCSE estimations are not equivalent to the instrumental variable (IV) 2SLS estimates.

Weak instrumentation arises when the instruments are only weakly correlated with the endogenous regressors. Stata reports the Cragg-Donald Wald F statistic for which Stock and Yogo (Stock and Yogo 2005) published critical values for the statistic for IV estimators. For the instruments used here, the null hypothesis that the instruments are weakly identified is strongly rejected.

The Hanson J-test is a test of overidentifying restrictions. The joint null hypothesis is that the instruments are valid instruments, i.e., uncorrelated with the error term and that the excluded instruments are correctly excluded from the estimation equation (StataVersion 11.1 2010). For the analyses in this chapter this instrument set fails to reject the null supporting the validity of the instruments.

Table 2: Results with Causes of Death

	2SLS	2SLS	2SLS	2SLS	2SLS
OUTCOME	All Cause	Tumor	Cardiovascular	Injury	Other
VARIABLES					
Health expenditure	0.0848	0.177***	-0.121	0.145	0.109
	(0.0555)	(0.0470)	(0.0823)	(0.157)	(0.0751)
Income per capita	-0.0248	-0.0774	-0.0622	0.126	0.0977
	(0.0410)	(0.0517)	(0.0499)	(0.128)	(0.0698)
College percent	-0.0231	-0.0309	-0.0564	0.0651	0.0402
	(0.0281)	(0.0222)	(0.0452)	(0.0811)	(0.0418)
High School percent	0.199**	0.111*	0.392***	-0.165	0.223**
	(0.0829)	(0.0656)	(0.124)	(0.241)	(0.108)
Population density	0.0724	0.0503	0.0253	-0.134	0.0500
	(0.0516)	(0.0349)	(0.0522)	(0.152)	(0.0822)
Smoking use	0.0340**	0.0564***	0.0403*	0.0121	0.0219
	(0.0159)	(0.0139)	(0.0227)	(0.0308)	(0.0277)
Alcohol use	0.157***	0.125***	0.102**	0.305**	0.238*
	(0.0523)	(0.0485)	(0.0405)	(0.128)	(0.126)
Percentage female	-0.998**	-1.402***	0.263	0.440	-2.358**
	(0.471)	(0.418)	(0.783)	(1.211)	(0.925)
Percentage white	-0.00841	0.0305	-0.00367	-0.275***	0.0660
	(0.0346)	(0.0249)	(0.0340)	(0.0880)	(0.0908)
Percentage black	-0.0304**	-0.0136	-0.0899**	0.0824**	-0.0121
	(0.0144)	(0.0197)	(0.0373)	(0.0380)	(0.0262)
First Stage F-Test	177.0	177.0	177.0	177.0	177.0
Observations	871	871	871	871	871
R-squared	0.877	0.775	0.964	0.367	0.784

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

Expenditures have a negative health impact on All Cause, Tumor, Injury, and Other mortality rates; the Tumor rate impact is significant at the 1% level. Income per capita does not have a significant impact on health; but, higher income per capita is generally better for health (except for the Injury and Other cause categories).

Health expenditures have a positive and significant health impact on cardiovascular mortality rates. Cardiovascular disease and mortality rates have been studied extensively (M. Marmot et al. 1991; McClellan, McNeil, and Newhouse 1994; Pincus et al. 1998; Davey Smith et al. 1998; Muntaner et al. 2001; Fisher, D Wennberg, Stukel, Gottlieb, Lucas, and Pinder 2003a; Fisher, D Wennberg, Stukel, Gottlieb, Lucas, and Pinder 2003b; Sorlie et al. 2004; Skinner et al. 2005; Fillmore et al. 2006; Skinner, Staiger, and Fisher 2006; MacDonald et al. 2009). Hadley (Hadley 1982a) found a highly significant negative effect of higher spending on cardiovascular mortality. Few studies indicate health expenditures produce a positive impact on cardiovascular mortality; Fisher, for example, demonstrates that higher spending regions in the U.S. provide more care, but these expenditures do not improve outcomes or patient satisfaction.

In this data sample, education is generally not a significant contributor to health. Higher high school graduation rates have a negative impact on health. Results from several empirical tests of Grossman's health care model suggest that education is the most important factor affecting health (Grossman 2000); and in this data the largest coefficients are on the IM and ACM education results. The expected results from equation 2 in Section 2.3, the first stage analysis in the 2SLS, have education variables showing negative relationship with medical care expenditures; the expected results from equation 1 have the education variables showing positive relationship with health outcomes. The high school graduation results shown here (in both regression stages—the first stage is not shown) suggest the opposite, i.e., that a higher high school graduation rate negatively affects health. College graduation rates show mixed and insignificant results.

Smoking use has a negative health impact in all cases and is particularly significant for Tumors; likely demonstrating the impact of lung, throat, and other smoking related morbidities. Alcohol use shows a consistently negative health impact—it is significant for all categories.

Having a higher percentage of women is generally healthier; white percentage shows a similar result. A higher black percentage is generally healthier except for the Injury-related category that shows a significantly negative impact on mortality.

Does the use of instrumental variables in a 2SLS analysis allow a concrete statement about causality? The instruments themselves must have validity, both statistically and theoretically. In this case, all of the statistical tests indicate that the choice of instruments is reasonable, and the instruments are valid, are not weak, and pass the overidentification test. From a theoretical point of view, it is reasonable to assess that the instruments have an impact on health expenditures through the volume of care used or by the availability of income for health-related expenditures, but not health outcomes. As the analyses in subsection 2.7.2 show, there are a number of valid potential instruments in the data set besides these two variables. Analyses using other instruments show similar qualitative results. In addition, the use of panel data, fixed effects for the states, and year fixed effects, reduces the impact of possible omitted variables.

Levy and Peart (Levy and Peart 2008) discuss the intriguing issue of transparency in econometrics and propose a standard rule to induce transparency: take each of the contending models and bootstrap them; the winning model has the smaller bootstrap variance. Freedman (Freedman 1984) has shown that bootstrapping 2SLS analyses is a valid estimation approach. Several models were bootstrapped with a variety of potential instruments. In all cases, the bootstrap standard errors, under-identification tests, and over-identification tests were reviewed, and the instrument set in the model with the smallest bootstrap variance was selected as the instrument set (physicians per capita, beds per capita, and dental expenditures per capita) to use throughout the chapter. Results from using the Stata bootstrap and jackknife commands are shown in Table 3. These modeling results are consistent with those from the 2SLS IV analyses shown in Table 2. Although the standard errors vary slightly in each analysis, the implication is clear: there is a negative, causal effect of health care expenditures on all-cause mortality. The impact of health expenditures is significant and affects mortality rates negatively. Population density negatively affects health, which may mean that urban areas are less healthy than more rural areas or perhaps sicker people live nearer to city-based medical facilities. Smoking and alcohol use both negatively impact health. In all analyses, the bootstrapped standard errors in Table 3 are slightly smaller than those in Table 2.

Table 3: Bootstrap and Jackknife Estimation Results

	Bootstrap 2SLS	Bootstrap 2SLS	Bootstrap 2SLS	Jackknife 2SLS
OUTCOME	**All Cause**	**All Cause**	**All Cause**	**All Cause**
Number of Iterations	50	500	5000	871
VARIABLES				
Health expenditure	0.0849***	0.0849***	0.0849***	0.0849**
	(0.0298)	(0.0311)	(0.0321)	(0.0347)
Income per capita	-0.0249	-0.0249	-0.0249	-0.0249
	(0.0268)	(0.0277)	(0.0280)	(0.0292)
College percent	-0.0231	-0.0231	-0.0231	-0.0231
	(0.0218)	(0.0190)	(0.0196)	(0.0206)
High School percent	0.198***	0.198***	0.198***	0.198***
	(0.0430)	(0.0466)	(0.0452)	(0.0474)
Population density	0.0723***	0.0723***	0.0723***	0.0723***
	(0.0193)	(0.0202)	(0.0208)	(0.0214)
Smoking use	0.0340***	0.0340***	0.0340***	0.0340**
	(0.0119)	(0.0129)	(0.0127)	(0.0133)
Alcohol use	0.157***	0.157***	0.157***	0.157***
	(0.0297)	(0.0277)	(0.0291)	(0.0307)
Percentage female	-1.002***	-1.002***	-1.002***	-1.002***
	(0.285)	(0.326)	(0.315)	(0.352)
Percentage white	-0.00843	-0.00843	-0.00843	-0.00843
	(0.0153)	(0.0193)	(0.0234)	(0.0181)
Percentage black	-0.0304***	-0.0304***	-0.0304***	-0.0304***
	(0.00863)	(0.00881)	(0.00916)	(0.00950)
Observations	871	871	871	871
R-squared	0.968	0.968	0.968	0.968

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

Finally, sensitivity analyses were performed using Public health care expenditure per capita data, Private health care expenditure per capita, or percentage of private expenditure data in place of total health care expenditures used in Table 2. Adjusting the mix of expenditures by using these variables did not change the results reported in Table 2. That is, the other baseline variables show qualitatively the same results, and the expenditure variables demonstrate negative impacts to health. Adding the percentage of public expenditure to the original baseline analyses from Table 2, however, achieves different results. That is, the percentage of public expenditure positively impacts health. This result is consistent with an international panel study by Berger and Messer (Berger and Messer 2002); they claim that increased public financing of health in a variety of OECD countries reduces mortality – measured as the overall mortality rate.

2.7.2 First Stage of the 2SLS Results and Income Elasticity

The first column in Table 4 shows the results of first stage of the 2SLS regressions for the previous baseline analyses. The 2SLS analyses are performed using the *xtivreg2* user-defined Stata routine with the *first* option, which displays the results of the first-stage regression analysis. These analyses are equivalent to fitting the demand curve for medical care (equation 2 in subsection 2.2.1) from Grossman's early work. Also, since these are log-log analyses, the results show elasticities directly, and are used to determine the income elasticity with respect to health expenditures.

The remaining columns in Table 4 show possible instruments and their impact on Health Care Expenditures (HCE) per capita. In general, the choice of instrument must satisfy the following standard characteristics: they have a significant effect on medical expenditure and they are uncorrelated with the error term in the health outcome function, i.e., they have no plausible effect on health outcomes. The choice of instrument must also be rationalized from a theoretical point of view and must pass basic statistical validation tests from a mathematical point of view. For this study, the model described in subsection 2.5.2 requires instruments that correlate with HCE and not with the health outcome used. The first-stage analyses performed below explore possible instruments that satisfy all the criteria. The baseline variables are used in the regressions and are shown along with the instruments used in the analyses in this chapter. Physicans per capita, beds per capita, and dental expenditures per capita demonstrate significant and positive impacts on health expenditures in all analyses.

Model 1 adds economic variables. None of the potential instruments are significant at the 10% level. In Model 2, the demographic variables are not significant, except for Poverty percentage, but it is also highly correlated with the income variable. In Model 3, the percentage of women legislators and the presence of a woman governor are significant. In Model 4, the public choice variables relating to percentage of black registered voters voting is significant. In Model 5, only the percentage of people on Medicaid is significant. As demonstrated by Hadley (Hadley 1988), variables representing insurance coverage are endogenous with health outcomes, particularly mortality. The variables in Model 5 then are not valid instruments for these analyses.

Table 4: Results of the First Stage Analysis

	Baseline	Model 1	Model 2	Model 3	Model 4	Model 5
OUTCOME	Health Exp per capita	Health Exp per capita	Health Exp per capita	Health Exp per capita	Health Exp per capita	Health Exp per capita
VARIABLES						
Income per capita	0.0816	0.1237	0.0738	0.1105	0.2300***	0.0529
	(0.059)	(0.092)	(0.072)	(0.077)	(0.086)	(0.071)
College percent	0.0414	0.0643**	0.0530	0.0571*	0.0440	0.0692*

	Baseline	Model 1	Model 2	Model 3	Model 4	Model 5
OUTCOME	Health Exp per capita	Health Exp per capita	Health Exp per capita	Health Exp per capita	Health Exp per capita	Health Exp per capita
	(0.041)	(0.032)	(0.096)	(0.034)	(0.034)	(0.041)
High School percent	0.1031	-0.1714	0.3288*	-0.0954	0.0385	0.0543
	(0.158)	(0.160)	(0.199)	(0.142)	(0.177)	(0.148)
Population density	-0.0583	-0.0819	-0.0091	-0.1190*	-0.0788	-0.0926
	(0.080)	(0.104)	(0.100)	(0.067)	(0.107)	(0.074)
Smoking use	-0.0247	-0.0294	-0.0203	-0.0318	0.0027	-0.0377**
	(0.017)	(0.018)	(0.017)	(0.020)	(0.015)	(0.018)
Alcohol use	0.1171**	0.2098***	0.0536	0.2141***	0.1363**	0.1495***
	(0.059)	(0.056)	(0.058)	(0.055)	(0.055)	(0.047)
Percentage female	0.1500	0.7180	0.2616	0.8451	-0.4683	-0.0298
	(0.895)	(0.773)	(0.920)	(0.836)	(0.813)	(0.828)
Percentage white	0.0173	-0.0245	0.0784	-0.0157	-0.0454*	0.0158
	(0.086)	(0.065)	(0.104)	(0.065)	(0.026)	(0.074)
Percentage black	0.0165	0.0745**	0.0097	0.0664**	-0.0232	0.0334
	(0.031)	(0.029)	(0.030)	(0.026)	(0.069)	(0.028)
Physicians per capita	0.6892***	0.6481***	0.6887***	0.5919***	0.5217***	0.6735***
	(0.086)	(0.081)	(0.124)	(0.079)	(0.105)	(0.078)
Dental expenditures	0.1532***	0.0937**	0.1454***	0.0874*	0.2184***	0.1260***
	(0.054)	(0.047)	(0.054)	(0.047)	(0.047)	(0.053)
Beds per capita	0.1209**	0.1224***	0.1089**	0.1212***	0.2373***	0.1099**
	(0.050)	(0.042)	(0.045)	(0.040)	(0.055)	(0.044)
Coincident Index		-0.0470				
		(0.081)				
Gini Index		-0.0535				
		(0.319)				
Tax Burden		-0.0003				
		(0.0004)				
Debt as a Percent of GSP		-0.0246				
		(0.056)				
Debt per capita		0.0333				
		(0.059)				
Urban percentage			0.1242			
			(0.165)			
Median Age			-0.0094			
			(0.008)			
Poverty percentage			0.0152**			
			(0.008)			
Hospitals per capita			0.0880			
			(0.055)			

	Baseline	Model 1	Model 2	Model 3	Model 4	Model 5
OUTCOME	Health Exp per capita	Health Exp per capita	Health Exp per capita	Health Exp per capita	Health Exp per capita	Health Exp per capita
Percentage Black Legislators				0.0660		
				(0.170)		
Percentage Women Legislators				-0.1605**		
				(0.072)		
Percentage Black Senators				-0.0734		
				(0.051)		
Percentage Women Senators				-0.0110		
				(0.126)		
Woman Governor				-0.0089*		
				(0.005)		
Red State				-0.0012		
				(0.007)		
Percentage Registered White Voters					0.0043	
					(0.003)	
Percentage Registered Black Voters					0.00002	
					(0.0004)	
Percentage Registered White Voters Voting					0.0017	
					(0.002)	
Percentage Registered Black Voters Voting					0.0012***	
					(0.0005)	
Percentage Registered Male Voters					-0.0041*	
					(0.002)	
Percentage Registered Female Voters					0.0004	
					(0.002)	
Percentage Registered Male Voters Voting					-0.0257	
					(0.017)	
Percentage Registered Female Voters Voting					-0.0290	

	Baseline	Model 1	Model 2	Model 3	Model 4	Model 5
OUTCOME	Health Exp per capita	Health Exp per capita	Health Exp per capita	Health Exp per capita	Health Exp per capita	Health Exp per capita
					(0.018)	
Percentage Registered Voters Voting					0.0522	
					(0.035)	
Percentage Uninsured						-0.0159
						(0.021)
Percentage on Private Insurance						-0.1087
						(0.082)
Percentage on Medicare						0.0106
						(0.012)
Percentage on Medicaid						0.0382***
						(0.014)
Observations	871	619	676	619	497	825
R-squared	0.995	0.990	0.994	0.642	0.736	0.994

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

The potential instrument variables are at best very weak instruments for the CMS detailed expenditure categories or are correlated with baseline variables. The goal of these analyses were to judge the validity and strength (and weakness) of potential instruments. The final set, Physicians per capita, Dental Expenditures per capita, and Beds per capita, were those that best balanced theoretical validity and strength as determined in these first stage analyses.

The income elasticity values are in the range ~0.05–0.23, which indicate that health is not a luxury good. Published determinations of this elasticity have suffered from rather severe omitted variable bias. Sen (Sen 2005) demonstrates clearly in his work, and in previously published papers that he references, that regression results, with only a few explanatory variables, show an elasticity consistently greater than 1.0. Upon adding fixed-effects variables, demand-side variables (e.g., percentage of population over 65 years of age), and supply-side variables (e.g., number of physicians per capita) the elasticity drops to about 0.5. My results are similar. In an initial sensitivity analysis performed with income per capita as the only explanatory variable, the results show an elasticity of ~1.3. Adding the baseline variables, state fixed-effects, year fixed-effects, and a variety of other explanatory variables reduces income elasticities to between 0.05 and 0.67 (results not shown). The results are also insensitive to the analysis method: using PCSE/OLS generates similar results with the range of elasticity values between 0.38 and 0.78.

2.7.3 Drug Expenditure Results

Nixon and Ullman (Nixon and Ulmann 2006) claim that "all studies that included pharmaceutical expenditure (Cremieux et al. 2001; Miller and Frech 2002; Cremieux, Jarvinen, et al. 2005; Cremieux, Meilleur, et al. 2005) found this aspect of health expenditure to be significant and positive for health outcomes." Nixon and Ulmann found similar results for European Union countries. None of these studies, however, used 2SLS or other methodological approaches to control for the potential bias between expenditures and outcomes. A more recent paper (Guindon and Contoyannis 2008) updates Cremieux's results and finds no relationship between spending on pharmaceutical products and infant mortality or life expectancy at 65 years of age in Canada.

Table 5: Drug and Non-Drug Expenditure Analyses, Part 1

	OLS/ PCSE	2SLS	OLS/ PCSE	2SLS	OLS/ PCSE	2SLS
OUTCOME	All Cause	All Cause	Tumor	Tumor	Cardio	Cardio
VARIABLES						
Drug expenditure	-0.0835***	-0.104	-0.0743***	-0.243	-0.132***	-0.336
	(0.0197)	(0.234)	(0.0214)	(0.202)	(0.0268)	(0.362)
Non-drug expenditure	0.0426**	0.0984*	0.0993***	0.210***	-0.0562*	-0.0506
	(0.0202)	(0.0506)	(0.0205)	(0.0479)	(0.0341)	(0.0972)
Income per capita	0.0358	-0.0168	-0.0536	-0.0586	0.0109	-0.0333
	(0.0444)	(0.0433)	(0.0488)	(0.0474)	(0.0522)	(0.0638)
College percent	-0.0151	-0.00680	-0.0223	0.00686	-0.0273	-0.00663
	(0.0182)	(0.0354)	(0.0191)	(0.0392)	(0.0243)	(0.0618)
High School percent	0.309***	0.279	0.230***	0.295**	0.445***	0.593**
	(0.0509)	(0.175)	(0.0552)	(0.150)	(0.0716)	(0.263)
Population density	0.0212	0.0469	0.00293	-0.00871	0.00310	-0.0519
	(0.0177)	(0.0845)	(0.0186)	(0.0553)	(0.0247)	(0.108)
Smoking use	0.0161*	0.0206	0.0390***	0.0252	0.0175	-0.000576
	(0.00963)	(0.0358)	(0.0119)	(0.0326)	(0.0123)	(0.0495)
Alcohol use	0.0986***	0.152***	0.139***	0.114***	0.0327	0.0869
	(0.0280)	(0.0459)	(0.0305)	(0.0384)	(0.0344)	(0.0599)
Percentage female	-0.471	-0.828	-0.891*	-1.005*	0.561	0.829
	(0.430)	(0.528)	(0.520)	(0.604)	(0.611)	(0.816)
Percentage white	-0.00921	-0.00815	0.0242	0.0311*	-0.0176	-0.00345
	(0.0171)	(0.0341)	(0.0213)	(0.0174)	(0.0260)	(0.0304)
Percentage black	-0.0128	-0.0179	0.00152	0.0153	-0.0693***	-0.0527
	(0.00947)	(0.0289)	(0.0113)	(0.0326)	(0.0159)	(0.0558)
Observations	886	871	886	871	886	871
R-squared	0.999	0.884	0.995	0.752	0.998	0.962

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

Table 6: Drug and Non-Drug Expenditure Analyses, Part 2

	OLS/ PCSE	2SLS	OLS/ PCSE	2SLS
OUTCOME	Injury	Injury	Other	Other
VARIABLES				
Drug expenditure	-0.104*	0.00540	0.0384	0.186
	(0.0541)	(0.677)	(0.0332)	(0.496)
Non-drug expenditure	0.117**	0.134	0.0157	0.0674
	(0.0571)	(0.150)	(0.0397)	(0.112)
Income per capita	0.133	0.125	0.149**	0.0814
	(0.107)	(0.132)	(0.0692)	(0.0974)
College percent	0.0462	0.0653	0.000421	0.0129
	(0.0390)	(0.109)	(0.0336)	(0.0740)
High School percent	0.0730	-0.147	0.314***	0.118
	(0.128)	(0.530)	(0.0868)	(0.341)
Population density	-0.191***	-0.134	-0.0136	0.0923
	(0.0446)	(0.241)	(0.0317)	(0.185)
Smoking use	0.00627	0.0119	0.0143	0.0444
	(0.0293)	(0.0879)	(0.0176)	(0.0701)
Alcohol use	0.187***	0.305**	0.145***	0.246
	(0.0567)	(0.141)	(0.0432)	(0.155)
Percentage female	0.557	0.425	-1.867***	-2.674***
	(0.708)	(1.587)	(0.562)	(0.813)
Percentage white	-0.250**	-0.275***	0.0850	0.0660
	(0.105)	(0.0882)	(0.0536)	(0.0865)
Percentage black	0.106***	0.0829	-0.00743	-0.0324
	(0.0364)	(0.0811)	(0.0140)	(0.0590)
Observations	886	871	886	871
R-squared	0.985	0.367	0.995	0.773

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

Table 5 and Table 6 show the results of OLS/PCSE and 2SLS regressions with CMS expenditure categories divided into non-drug expenditures and drug expenditures.

Drug expenditures are positive and significant for OLS analyses, and not significant for 2SLS analyses, for all-cause mortality, tumor-related mortality, and cardiovascular mortality.   For Injury-related and Other-related mortality, the effects negatively impact health but remain non-significant. Non-drug expenditures negatively impact health in the 2SLS analyses for all-cause and tumor-related mortalities. It is clear from these results that the OLS/PCSE results have standard errors that are too conservative. Hausman tests indicate that the results of the OLS/PCSE estimations are not equivalent to the instrumental variable 2SLS estimates. In all cases, the standard errors are less than those with 2SLS.

2.7.4 State-Level Factor Analyses Results

The state-level factor analysis resulted in four (4) factors that represent state characteristic and demographic variables. Using the state factors as geographic-location variables gives the results shown in Table 7. The remaining detailed baseline results do not change dramatically in magnitude or significance and are not shown.

Table 7: State Factors and All-Cause Mortality

	2SLS
OUTCOME	All-Cause
VARIABLES	
1. Big Fish, Small Pond	-0.0982**
	(0.0460)
2.Up-and-Comers	-0.00854
	(0.0414)
3.Heartlanders	-0.0774*
	(0.0415)
4.Empty Nesters	-0.00231
	(0.0336)

	2SLS
OUTCOME	All-Cause
Observations	855
R-squared	0.802

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Big Fish, Small Pond states have high rankings in education, health, and low crime rates. Up-and-Comer states have high ranks in income, graduation percentage, the happiness index, and the freedom index. Heartlanders states have high manufacturing and farm output, and a high percentage of church attendance. Empty Nester states have a high percentage of people over 45 years of age and a low births per capita rank.

Two of the four factors demonstrate a significant result. Factor 1 and Factor 3 have a positive impact on all-cause mortality. Factor 1 Big Fish, Small Pond characteristics are higher in education (IQ Rank, Percentage of Graduates, and Smartest), higher in health (Healthiest, Exercise Frequency, and Percentage with No Insurance), and lower in crime rates (Crime Rate and Violent Crime Rate) rankings. Living in the upper Midwest, Utah, and the New England states rank highly on Factor 1 (see Appendix F) and has a positive impact on all-cause mortality through the benefits of more education, better overall health, more exercise, and lower crime rates.

The characteristics of Factor 2 Up-and-Comers include a higher percentage of people in the 25-44 age group, higher income, high college graduation rate, and higher urbanization. This factor has a wide enough variety of characteristics that population-dense states such as New Jersey, New York, Connecticut, and Maryland rank highly, but also South Carolina, Louisiana, Tennessee, and Oklahoma are ranked near the top due to characteristics such as high Generosity, high Happiness index, and low in Starbucks per capita. The impact on health is not significant.

Factor 3 Heartlanders represents characteristics that include regular church attendance, a high regard for religion, worse overall state economic health, high manufacturing employment, and high farming output. The states ranking high in Factor 3 are the heart land states of the central Midwest from Ohio and Pennsylvania south to Arkansas and Louisiana. The key characteristics for this factor have slightly positive and significant impact on health.

Factor 4 Empty Nesters is constructed primarily from demographic characteristics: high in percentage of 45-64 age group and the 65-plus age group, low in the percentage of 0-17 age group and the 18-24 age group, smaller in square mile area, higher preference for western longitudes, and a higher percentage of females. The impact on health is not significant.

2.7.5 Analyses for a Possible Structural Break

Two additional regressions were performed dividing the years of study into two sub-ranges. The results are shown in Table 8. The first column is for the years 1985-1991 and the key results are the first two rows that show mean health expenditures and mean income per capita. The results show a positive impact of health expenditures on health and a negative impact of income on health. For the years 1992-2004 (second column), the results show a negative impact of health expenditure on health and a positive impact of income on health. A Chow test rejects the null hypothesis of no structural break. Future work should focus on disentangling a possible structural break in the early 1990s that would account for this variation in grouped time frames.

Table 8: Analyses by Year Groupings

	2SLS	2SLS
	1985-1991	1992-2004
OUTCOME	All Cause	All Cause
VARIABLES		
Non-drug expenditure	-0.177*	0.157***
	(0.106)	(0.0407)
Income per capita	0.119*	-0.0347
	(0.0646)	(0.0319)
College percent	-0.201	-0.0170
	(0.150)	(0.0218)
High School percent	0.541***	0.251***
	(0.210)	(0.0564)
Population density	-0.0539	0.0896***
	(0.0531)	(0.0281)
Smoking use	-0.0140	0.0360**
	(0.0158)	(0.0144)
Alcohol use	0.0937**	0.0492
	(0.0456)	(0.0345)
Percentage female	1.202	-1.738***
	(1.469)	(0.342)
Percentage white	-0.431**	0.00180
	(0.194)	(0.0138)
Percentage black	-0.0545*	-0.0401***
	(0.0296)	(0.0125)
Observations	236	632
R-squared	0.815	0.819

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

Population density, smoking results, percentage female, and percentage white are significant and differ as well.

2.8 Conclusions

This chapter presents an empirical analysis of the relationship between health care determinants, including health care expenditures, and health outcomes using aggregated annual data for the 50 U.S. states covering 28 years from 1980–2007. This study confirms and extends previous research by creating a comprehensive data set from national sources that includes recent data on the latest trends in health expenditures, outcomes, and demographics; by using instrumental variable techniques to examine the realtionship of health expenditures, including detailed CMS expenditures in specific health-related categories and the dependent variables of all-cause mortality and cause-specific mortality; re-examining the income elasticity of demand for health care expenditures; and considering geographic impacts on health outcomes using a unique state-level factor analysis of key state-wide characteristics.

The empirical results indicate a consistent negative impact of aggregate health expenditure on all-cause mortality, and a variety of cause-specific mortalities, except for cardiovascular mortality. The effects were moderated by the use of instrumental variable techniques across a variety of relevant and tested instruments. Simulations performed via bootstrap and jackknife techniques validated the negative impacts on health. Alcohol use and smoking have a negative impact on all-cause mortality, and other specific cause-related mortalities. The impact of income is generally beneficial for health outcomes and is generally larger than educational attainment. The percentage of college graduates had no impact on health outcomes, and the percentage of high school graduates correlates with worse overall health outcomes. Sensitivity analyses that add additional explanatory variables do not affect the results for high school graduation percentages. Educational attainment has consistently been reported to improve health outcomes. Many of these studies are cross-sectional rather than longitudinal, and do not use instrumental variables. The implication of this study is that some education may be beneficial and some may not be. Also, there may be an unidentified third variable that affects both education and health and is not accounted for. Another factor may be that education, as used in this study and most others, is considered to be a “black box,” i.e., there is very little insight into the nature of education – for example, the quality of the schools, the diversity of the population, courses taken, or the degree of education funding – and that may impact longitudinal analyses of education.

For the detailed breakout of health care expenditures drug expenditures have a generally positive effect but are not significant as demonstrated by Guindon and Contoyannis (Guindon and Contoyannis 2008). Non-drug expenditures demonstrate a negative impact on health.

Income elasticity results from the first stage of the 2SLS analyses are consistently in the range of 0.05–0.4 indicating that health is not a luxury good. Other analyses (Sen 2005) have pointed out the omitted variable bias inherent in previous studies, and those results are replicated with this data when sufficient explanatory baseline variables are included in the analyses.

A factor analysis of 55 state characteristic and state demographic variables resulted in four factors that rank states into groups with common features. Appendix F shows the results of the rankings with images of the high ranking states associated with the four factors. These factors are used as alternative geographic definitions. Table 8 shows the 2SLS results for all-cause mortality. Each group shows positive impacts on health with Factor 1 and Factor showing significant results. Without the mortality related state characteristics in the factor analysis, the most important remaining characteristics that affect health are education attainment, crime rates, general overall health (such as a ranking of the healthiest states), and the extent of the population that is elderly.

A possible structural break in the impact of health expenditures and income on health is identified as occurring in the early 1990s. A Chow test rejects the null hypothesis of no structural break. Future research should work to disentangle this possible effect.

Follow-on research should fill out the data tables with a more complete set of recent data. Many key variables are missing data in the most recent years and the demands of this research have prevented any additional efforts to mine more data from the variety of sources. A thorough investigation using a broader set of instrumental variables is warranted. Using fixed-effects eliminates impacts on the health estimates by state-specific time invariant variables and by state-invariant adjustments that may change over time. As pointed out by Pritchett and Summers (Pritchett and Summers 1996), however, using a set of plausibly exogenous instruments – with low correlation between the instruments – would provide evidence of a stable, causal relationship between expenditures and health outcomes. In addition, it is reasonable to expect other variables, e.g., education and health industry supply-side variables, to be endogenous. Therefore, instruments for these variables should be examined as well.

3. Individual Health Care Determinants and Health Outcomes

3.1 Introduction

This chapter examines the relationship between health care determinants and health outcomes based on individual data for several hundred thousand people from the National Longitudinal Mortality Study (NLMS). The NLMS is sponsored by the National Cancer Institute; the National Heart, Lung, and Blood Institute; the National Institute on Aging; the National Center for Health Statistics; and the U.S. Census Bureau for the purpose of studying the effects of demographic and socioeconomic determinants on U.S. mortality rates (U.S. Census Bureau 2010). The results include analyses of the contributions of individual level health care determinants to health outcomes, including the likelihood of death, and cause-specific mortality conditions. The results also show a significant relationship between geography and the risk of mortality. Non city-central areas (reasonably called suburbia) and rural areas have significantly better health than urban areas. Geographically, in the United States, northern and western states have better health than southern states.

The NLMS data has variables that indicate whether the individual lives in urban or rural environments, the individual's State of Birth, and the individual's State of Residence. Constructed geographic variables include Census Region and Census

Division dummy variables, and the use of the state-level factor analysis variables described in chapter 2.

This chapter complements chapter 2 in using individual-level data to examine similar relationships between economic, socio-demographic, and lifestyle factors on health outcomes, including a focus on cause-specific mortality data, geography data, and factor analyses. This chapter also provides background information, which will support the occupational studies addressed in chapter 4.

The chapter is organized as follows. First, the background section discusses the literature on individual-level health care determinants. Next, research questions and hypotheses are discussed. The data and the methodologies used in the empirical analyses are introduced and followed by a detailed discussion of the analytical results. Finally, concluding remarks are presented, together with a brief discussion on possible directions for future research.

3.2 Background

Studies on the relationships between individual health care spending and other determinants on health outcomes are limited by the data sources available. Many of the individual studies use survey data with categories of spending provided by the respondents or national data collection studies, such as the Medicare Current Beneficiary Survey (MCBS) in the United States, which are limited to a specific population (e.g., in the MCBS, the population consists of persons enrolled in the Medicare program).

There are few cross-country studies of individuals; most studies are localized to a specific country or to intra-country regions. For example, these studies focus mainly on cross-sectional subsets of country data, such as Danish welfare recipients (Wagstaff 1986a; Wagstaff 1993); German data (Erbsland, Ried, and Ulrich 1995); Swedish data (Gerdtham and Johannesson 1999); Estonian data (Vork 2000); Swiss data (Nocera and Zweifel 1998); British data (M. Marmot et al. 1991; Michael Marmot et al. 1997; Salas 2002; Clark and Royer 2010); and Finnish data (Kiiskinen 2003).

In the U.S., representative examples include: National Opinion Research Center (NORC) data (Grossman 1972a; Grossman 1999); the Health Examination Survey (Newhouse and Friedlander 1980); the National Longitudinal Study of Young Men (Berger and Leigh 1989); the National Health Interview Survey (Bednarek, Pecchenino, and Stearns 2003); the Panel Study of Income Dynamics (Haveman et al. 1994); NYC Vital Statistics (Joyce 1994); the Retirement History Survey (Sickles and Yazbeck 1998); the MCBS (Shang and Goldman 2007); the National Health and Nutrition Examination Survey (NHANES) (Wilper et al. 2009); and the National Longitudinal Mortality Study (NLMS) (Sorlie et al. 1994; Sorlie, Backlund, and Keller 1995; Elo and Preston 1996; Backlund, Sorlie, and Johnson 1996; Gregorio, Walsh, and Paturzo 1997; Richard Anderson et al. 1997; Johnson, Sorlie, and Backlund 1999; Backlund, Sorlie, and Johnson 1999; G Howard et al. 2000; House et al. 2000; Muntaner et al. 2001; Kposowa 2001)

3.2.1 Theory

Because medical spending is often not available in the survey data, most studies focus on the health production function. An alternative formulation to equation (1) in section 2.2.1 is:

(3) $$lnH_i = \alpha + \beta lnW_i - \beta_1 lnP + \beta_2 lnE_i + \beta_3 T_i + \beta_4 lnX_i + u$$

where H_i is the stock of health for individual i, W_i is the individual wage rate, P is the price of medical care, E_i is educational attainment, T_i is age, and X_i is a vector of other relevant explanatory variables. The options used for the stock of health often include self-reported health, physiological measurements, death, number of doctor visits, number of hours spent on physical activity, and so on. Wage rates are often unavailable and are proxied by individual or household income. Similarly, the price of medical care is often unavailable, and most studies, including Grossman's original papers(Grossman 1972b; Grossman 1972a), assumed that the price did not vary "across the relevant units of observation" and thus was dropped from the analyses. There is considerable variation in the "other relevant explanatory variables" used in the studies mentioned above. As in many studies, available data drives what gets included; or a specifically defined purpose for the study dictates both the choice of the survey to use and the use of particular data within the survey. This dissertation focuses on the relationships between occupations and health and the NLMS is a survey that contains occupation, health, and demographic variables.

3.2.2 Empirical Literature Review

Grossman's original empirical analyses (Grossman 1972a) used NORC data, with dependent variables representing positive health. Grossman found the regression coefficient on age to be negative, which is expected if health depreciation rises with age. The education coefficient was positive and significant indicating better health with more education. Wage rates were positively related to health. Family size was positively related to health; and being female was healthier than being male. Wagstaff (Wagstaff 1986b; Wagstaff 1993) used Danish Health Study data, with a slight reformulation of Grossman's equations, and found that the coefficients on the demand for health equation had the expected signs. Erbsland et al. (Erbsland, Ried, and Ulrich 1995) used German Socio-economic Panel data and found significant results with the expected signs, including a variable for private insurance that had a positive effect on health. Gerdtham and Johannesson, using Swedish micro data(Gerdtham and Johannesson 1999), showed that health increases with income and education and decreases with age, being male, living in big cities, and being single.

Vork used self-assessed health (Vork 2000) and demonstrated that a demand for health model supported Grossman's model – income and education improve health and age reduces health. Nocera and Zweifel (Nocera and Zweifel 1998) used time series data which, by and large, confirmed the predictions of the Grossman model. This analysis failed to account for endogeneity, however. Marmot's Whitehall study (M. Marmot et al. 1991) showed a steep inverse association between social class and mortality, including impacts on physical and mental morbidity, and on psychological well-being.

Newhouse and Friedlander (Newhouse and Friedlander 1980) investigated the relationship between medical resources and physiological measures of individual health status. Although they found that additional education and income were associated with fewer physiological measures performed, the overall conclusion was that the impact of additional medical resources was minimal. Berger and Leigh (Berger and Leigh 1989) examined the positive correlation between schooling and good health in detail. They conclude that the observed correlation is due to the direct effect of schooling on the efficiency of producing health.

Howard et al. (G Howard et al. 2000) used the National Longitudinal Mortality Study (NLMS) to investigate the impact of socio-economic status (SES) on racial differences in mortality. Using income and education as SES measures, but not using interaction terms between race and SES, they found that SES plays a substantial role in excess black mortality in ischemic heart disease, lung cancer, and diabetes. Access to health care and lifestyle choices may mitigate these results, but are not available in the data. Backlund et al. (Backlund, Sorlie, and Johnson 1996) examined differences in the inverse gradient between income and mortality at different income levels and age groupings in the NLMS. The income gradient is shown to be much smaller at high income levels than at low to moderate income levels; in addition, the income gradient was much smaller in the elderly than in the working age population. House et al. (House et al. 2000) used the NLMS and reported that city residents have a significant prospective excess mortality risk and this risk is not attributable to differences in terms of age, race, gender, education, income, or marital status.

3.3 Research Questions

The major goal of this study is to investigate determinants of health outcomes, using detailed individual health data from the National Longitudinal Mortality Study. The determinants considered in the empirical analyses include education, income, poverty levels, gender, race, and geographic variables. Consequently, the major research questions and the corresponding predicted responses investigated in this study are shown in Table 9.

Table 9: Major Research Questions and Predicted Responses Investigated in Chapters 3 and 4

Research Question (RQ)	Predicted Response (PR)
1. What is the impact of education, income, and race data on individual health outcomes using the NLMS survey data? Is there an added impact if the approach uses multi-level analysis by making use of the U.S. state-level data?	Individual education, income, race, and gender variables are expected to have significant impacts on outcomes. To the extent that a second-level, e.g., states, impacts the total variance, there will likely be variability in geographic regional results.
2. What is the impact of geography on individual health outcomes? Do state-level factors interact with geographic variables?	Geographic variation ought to have impacts on outcomes. For example, rural living has been shown to be healthier than urban living. Impacts will likely vary by state or other regional grouping.
3. What is the impact of occupation on the individual health outcomes? Are there geography X occupation interaction impacts?	Outcomes will likely vary with occupation, even controlling for education and income. More manual and labor-intensive occupations typically have poorer health outcomes. Interaction impacts are likely to be observed. See chapter 4.
4. What are the key factors from an occupational factor analysis, and how do these factors impact health outcomes?	The impact of occupational factors will likely vary from physical characteristics to more intellectual characteristics. The correlation between factors and outcomes will likely be similar to that between the corresponding occupations. See chapter 4.
5. Do the occupation factors provide insight into psychosocial behavioral aspects of occupations? Do subjective measures impact more or less than object measures?	The literature on psychosocial measures and other subjective measures is mixed. It is an empirical issue with the data available. See chapter 4.
6. Is there an impact of geography combined with key occupation factors? Do state-level factors interact with occupation factors?	To the extent that the factors affect outcomes, there should be an impact by geography. See chapter 4.

3.4 Data

The data used are from the U.S. National Longitudinal Mortality Study (NLMS).

3.4.1 Data Sources

The NLMS is a prospective study of mortality occurring in combined samples of the non-institutionalized U.S. population. It consists of samples taken from selected Current Population Surveys (CPS) conducted by the U.S. Bureau of the Census. Each CPS is a complex, national probability sample of households surveyed monthly to obtain demographic, economic, and social information about the U.S. population, with particular emphasis on employment, unemployment, and other labor force characteristics. The surveys, which are conducted by personal and telephone interviews, have a response rate of close to 96%. The CPS, sponsored by the U.S. Bureau of Labor Statistics, is used, in part, to prepare monthly estimates of the national unemployment rate. CPS surveys are redesigned every 10 years, and households are sampled only once during that period (Johnson, Sorlie, and Backlund 1999). The version of the NLMS used in this chapter is the Public Use Release 3 file, dated June 1, 2008 (obtained from U.S. Census Bureau in November 2008), containing a total of 988,396 individual records (U.S. Census Bureau 2010).

Mortality follow-up information for the NLMS is collected by computer matching its records to the National Death Index (NDI) over an 11 year period. The NDI is a national file containing information collected from death certificates and maintained by the National Center for Health Statistics. The matching of records to the NDI has been shown to be an effective and accurate means of ascertaining mortality information using personal identifiers such as Social Security Number, name, date of birth, sex, race, marital status, state of birth, and state of residence. Mortality rates determined from the NLMS are consistent with estimated rates for the non-institutionalized population of the United States from other sources.

During the CPS household interview, a detailed series of questions elicit information about occupations. If the responses to these questions indicate that the person is in the labor force or has held a job within the last five years, the interviewer asks specific questions relevant to the job description or business. These responses are later coded to a basic three-digit occupation and three-digit industry code, as documented by the U.S. Bureau of the Census. Chapter 4 discusses the use of occupation as a determinant of health outcomes.

3.4.2 Sample Construction

In this dataset, common economic factors, socio-demographic factors – including occupation and industry codes – and lifestyle factors are selected or constructed from available data (see Appendix B for a statistical overview of the NLMS variables).

3.4.3 Dependent Variables

The primary dependent variable is a Death Indicator (=1 if the respondent was matched to an NDI record), which is renamed All-Cause Mortality in chapters 3 and 4. Overall, about 9.1% of respondents died during the follow-up period for this data set. The primary cause of death is coded in Cause1 using the International Classification of Diseases, ninth revision (ICD-9) codes. I recoded these values into a General Cause of

Death variable that represents the same four general categories used in the analysis in chapter 2 (Tumor, Cardiovascular, Injury, and Other causes of death).

3.4.4 Explanatory Variables

Following previous studies and recommendations for SES analyses (Braveman et al. 2005), the baseline set of explanatory variables includes age, race, gender, income, marital status, education, and geography, and is shown in Table 10. This baseline set of variables reflects the named variables in equation (3).

Table 10: Baseline NLMS Variables Used

Baseline Explanatory Variables	Definition/Base Value
Age Dummies	One dummy variable for each age in the analyses
Race	Race/base is Race=0 for White
Female	Gender/base is Female=0 for Male
Rural	Geography/base is Rural=0 for Urban
Married	Marital Status/base is Married=0 for Other Than Married
Education Completed	Education level/base is Education = 8, completed high school
Adjusted Income	Family income/base is Income=8, $25,000-$29,999

For age, an age dummy for each age-year of the study was used. As an alternative age specification, a mean-centered age variable was constructed to aid in the interpretation of regression results; and an Age2 variable was constructed to account for any non-linear impacts of age. Since several of the race categories have a small number of respondents, a Race variable is constructed with three categories: White, Black, and Other. The gender variable and the urban/rural variable are defined within NLMS. The marital status variable in the NLMS was transformed to a dummy variable in which the base value represents "other than married" status and the value one represents married. The NLMS defines income with 14 categories; and education completed with 14 categories. Both are shown in Table 11 with the base categories identified.

Table 11: Listing of NLMS Income and Education Variables

Category	Income Value	Education Value
1	$0–$4,999	None to less than one year of education (E1)
2	$5,000–$7,499	Completed 1-4 years of education (E1, E2, E3, E4)
3	$7,500–$9,999	Completed 5-6 years of education (E5, E6)
4	$10,000–$12,499	Completed 7-8 years of education (E7, E8)
5	$12,500–$14,999	Completed 1 year of high school (H1)
6	$15,000–$19,999	Completed 2 years of high school (H2)
7	$20,000–$24,999	Completed 3 years of high school (H3)
8 (base value)	$25,000–$29,999	High School Graduate (H4)
9	$30,000–$34,999	Completed 1 year of college (C1)
10	$35,000–$39,999	Completed 2 years of college (C2)
11	$40,000–$49,999	Completed 3 years of college (C3)
12	$50,000–$59,999	College Graduate (C4)
13	$60,000–$74,999	Completed 5 years of college (C5)
14	$75,000-plus	Completed 6+ years of college (C6)

The geographic categories provided by NLMS include an urban/rural dichotomous variable; Standard Metropolitan Statistical Area (SMSA) status (in three categories: SMSA in Central City, SMSA not in Central City, and not in SMSA – these generally represent urban, suburban, and rural areas); State of Birth; and State of Residence. Using the State variables and the U.S. Census Bureau Region and Division definitions, four additional geographic variables were constructed: Division of Birth, Region of Birth, Division of Residence, and Region of Residence.

Other constructed variables were created to simplify the choices among sets of enumerated values – typically because in the original data set there are too few values in particular enumerated codes to be meaningful. Thus, Non-White is constructed from Race such that White=0 and all other races=1; and Working is constructed from Employment Status such that Working=1 for employed, Working=2 for not employed, but looking for work, Working=3 for unemployed, and Working=4 for all other groupings, including disabled, retired, student, and homemaker.

3.5 Methodology

The STATA statistical analysis package, v.11, is used for all analyses (StataVersion 11.1 2010). For formatting the regression tables, the user-supplied package OUTREG2 is used (Wada 2010). The Stata data files (*.dta) and analysis processing files (*.do) are available by request from the author.

3.5.1 Logit Analyses

One approach used is a logistic approach with interaction effects. The basic specification estimated is:

$$\text{logit}(\pi) = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ + \beta_i Y_i + \varepsilon$$

where $logit(\pi)$ is the mortality proxy; X is a socio-demographic factor; Z is a geographic factor; XZ is the interaction effect between X and Z; Y_i is a vector of the remaining economic, socio-demographic, or lifestyle factors; β_0 is a the intercept; and ε is a disturbance term. Some regressions use geographical dummy variables to represent any unaccounted for invariant characteristics of geographic locations. Finally, interaction effects between socio-demographic variables and geographic variables are used to determine if there is an impact of geography on the relationship of the socio-demographic variable and health.

Logistic analyses determine coefficients that are the log odds of the relationship between the explanatory and dependent variables. As an alternative reporting approach, odds ratios are frequently used. Odds ratios are simply the exponentiated value of the log odds, and can be generated automatically by Stata. The interpretation varies depending on the nature of the explanatory variable. For example, the odds ratio of a dichotomous explanatory variable is the ratio of the odds of one value relative to the odds of the other value. Either logit coefficients or odds ratios are reported in the following analyses.

3.5.2 Cox Proportional Hazard Analyses

Another approach uses Cox proportional hazards (CPH) regression to determine relative mortality differences after adjustment for the socio-demographic determinants. This is a standard approach used in prior studies (Johnson, Sorlie, and Backlund 1999; Sorlie, Backlund, and Keller 1995). CPH is used to analyze survival data. The procedure regresses survival times (or more specifically, the hazard function) on the explanatory variables. The hazard function is the probability that an individual will experience an event (in the NLMS data, the event is death) within a time interval. It can be interpreted as the risk of dying at time t. The hazard function is:

$$h_i(t) = h_0(t)exp(Z_i)$$

$$Z_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_n X_{in}$$

where $h_0(t)$ is the baseline hazard function; i represents $1,\ldots,n$ individuals; t is a time variable; and X_{ij} are the $1,\ldots j$ explanatory variables for individual i. The baseline hazard function corresponds to the probability of dying when all explanatory variables are zero. "Proportional hazard" refers to the assumption of a constant relationship between the dependent variable and the explanatory variables, i.e., that the hazard functions for any two individuals at any point in time are proportional. Thus, if an individual has a risk of death at some point that is twice as high as another individual, then at all other times the risk of death remains twice as high. This assumption of proportionality needs to be tested. The Kaplan-Meier test plots predicted survival curves against observed survival curves. The closer the observed values are to the predicted, the less likely the proportional hazards assumption is violated. Figure 1 shows a Kaplan-Meier plot for the NLMS data and there is clear commonality between the predicted and observed curves. A second test is a complementary log-log plot, which plots the negative log of the estimated survival function against the log of the survival time. If the curves are parallel, demonstrating proportionality, then the assumption is not violated. Figure 2 shows a plot for the NLMS data which demonstrates sufficient parallelism.

The resulting CPH analyses provide hazard risk ratios directly. A hazard ratio greater than one means that the hazard (risk of death) is higher, and thus, the predicted outcome is worse for individuals with higher values of that variable. Conversely, a hazard ratio less than one means that the hazard (risk of death) is lower and the predicted outcome is better.

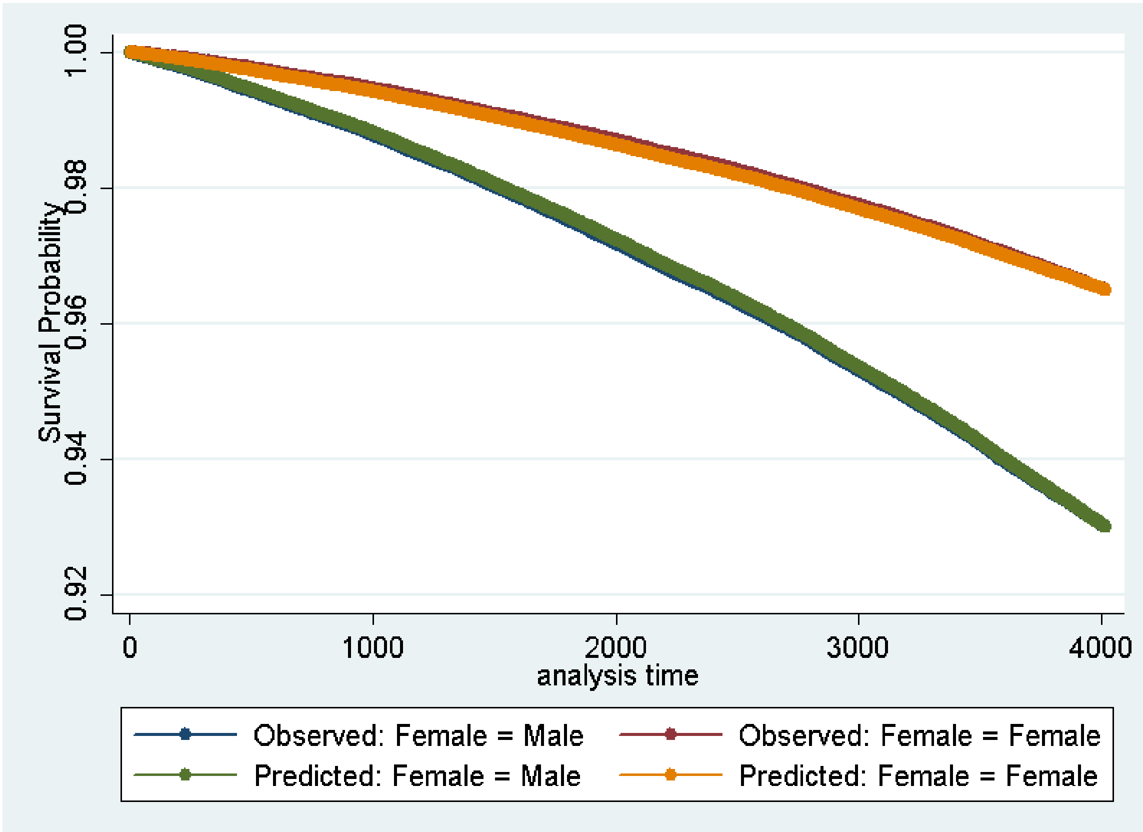

Figure 1: Kaplan-Meier Plot for NLMS Data

3.5.3 State Factor Analysis

The State Factor Analysis approach described in chapter 2 is used in this chapter as well. The analyses in this chapter employ the same four factors determined from the overall analysis of the 55 State variables.

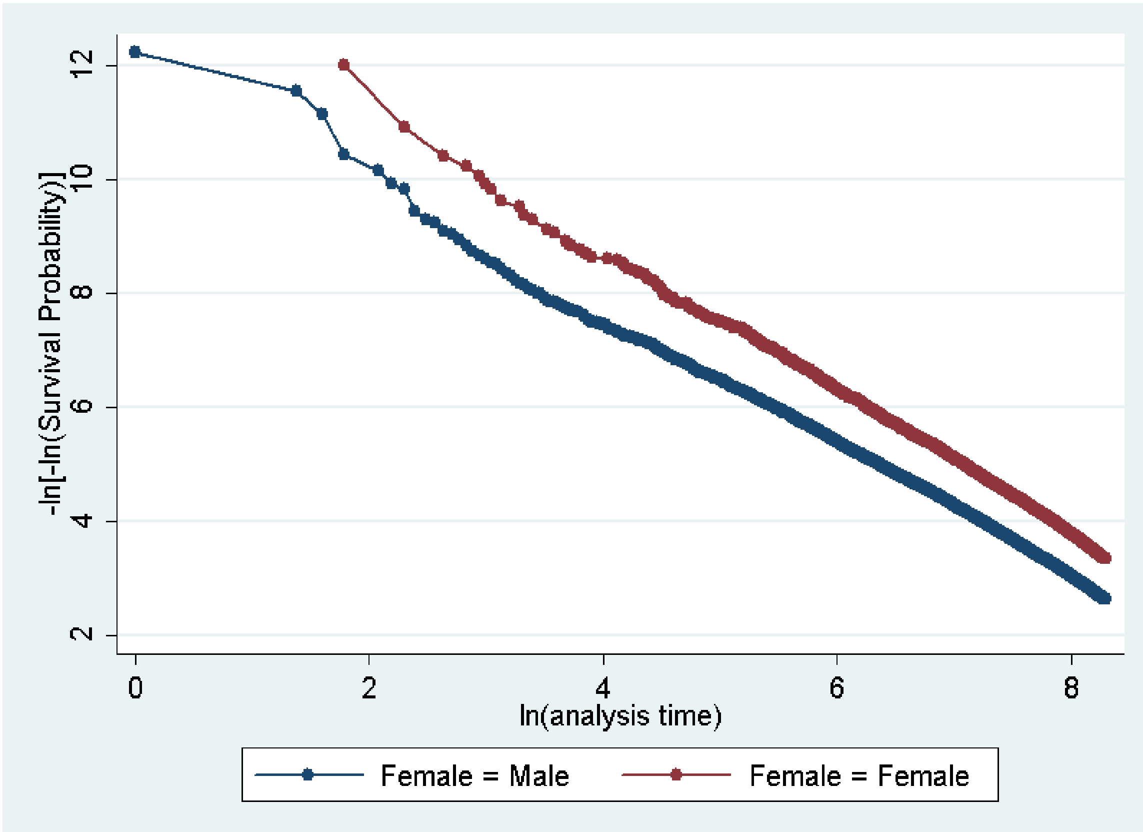

Figure 2: Complementary Log-Log Plot of NLMS Data

3.6 Study Sample Characteristics

This study focuses on individual health status and data from the NLMS. The analyses of the relationship between health outcomes and a variety of socioeconomic status (SES) and demographic variables are conducted using the cause of death to provide specific insight into mortality. There are geographic variables that define large-scale regions of individuals (States and Census Regions/Divisions). There is also an occupation variable that identifies the primary occupational field of the primary survey responder (generally, the head of the household). Chapter 4 focuses on adding the occupation data and the corresponding impacts on health outcomes.

3.6.1 Socio-Demographic Variables

Socio-demographic characteristics of individuals have impacts on health. Standard demographic variables in this data include:

- **Age.** Health worsens with age; age likely has the biggest impact on mortality and health. Using a set of age dummy variables, the coefficients indicate the association between health and age – either positive or negative. For age as a continuous variable, health should get progressively worse with age, but like a time trend will only demonstrate a linear response. For age as a set of age categories, each category of older adults should show worse health. For these analyses the age range of individuals is limited to 25–65, with the assumption that this range represents the working age range of the U.S. population.
- **Gender.** Health results are consistently better for females than for males. Similar results are expected.
- **Race.** Health results vary with race. For this analysis, the constructed Race variable has categories for White, Black, and Other. Other includes principally Hispanic races and Asian races. Results for Blacks are expected to be worse than for Whites. The results for Other is an empirical issue.
- **Education Completed.** This is a measure for the head of household. The knowledge of health care and its consequences are a determinant in health. Higher education levels generally correlate with better health (Sorlie,

Backlund, and Keller 1995; Lantz et al. 1998; Pincus et al. 1998; Lantz et al. 2001; Hayward and Gorman 2004; Lleras-Muney 2005; Grossman 2000).

- **Income.** Individual income has been shown to be correlated with improved health outcomes. Higher financial resources may result in higher health service availability or indicate a greater ability to pay for services not covered by insurance.
- **Marital Status.** Health is consistently reported as better for married people. Similar results are expected.
- **Number in Household.** Some studies show improved health with larger families; in low-income households, more mouths to feed may result in a higher likelihood of living in poverty. The expected results are then an empirical issue.
- **Poverty Level.** Health is expected to decline for those in poverty.
- **Housing Tenure.** This is an indicator of whether the household owns or rents. Owning is likely tied to social status and income, and will likely follow similar trends.
- **Insurance Status.** This is a dummy variable indicating whether the head of household is covered by insurance or not. Having insurance is expected to provide health benefits.
- **Insurance Type.** This variable identifies the type of insurance held by the head of household, e.g., employer-provided, Medicare, Medicaid, TRICARE, etc. Which type provides better health benefits is an empirical issue.

- **Veteran Status.** This is a dummy variable indicating that the head of household did or did not serve in the U.S. military.
- **Employment Status.** Health is generally better for working people than for those out of work. This variable provides multiple status indicators of the employment status of the head of household.
- **Class of Worker.** This variable identifies whether the head of household works in private industry, Government, or is self-employed.

3.6.2 Geographic Variables

Geographic characteristics of populations have an impact on health. For these analyses only individuals born in the United States are included. There can be differences in health and life course outcomes based on country of origin, so that complicating factor is eliminated. The geographic variables in the NLMS include:

- **Urban/Rural.** This dummy variable identifies whether the household lives in a primarily urban or primarily rural area. Rural living is often associated with better health (Hayward and Gorman 2004).
- **Standard Metropolitan Statistical Area (SMSA).** The *SMSA in central city* category will likely track with the urban living. The *not in SMSA* category will likely track with rural living. The *SMSA not in central city* category represents suburban living close to cities and will likely provide results somewhere between the other two categories.
- **State of Residence.** This variable supports the construction of Census Regions and Divisions. Aggregating data to the State/Region/Division level

allows the use of chapter 2 data for multi-level analyses. These analyses face the same issues as in chapter 2 with large-scale aggregation areas. That is, the Census regions and divisions are not representative of specific cultural characteristics of the populations; they are merely geo-political boundaries. That having been said, there is some interest in the geographic differences between north and south, and east and west. Even gross groupings of midwestern states into Census Divisions have different inherent qualities than groupings of southern or northeastern states.

3.7 Results and Discussion

3.7.1 Baseline Analyses and Sensitivity Analyses

The baseline results and sensitivity analyses for this study are shown in Table 12. Sensitivity model 1 adds the number of people living in the household, veteran status, and whether the home is owned or rented. Each of these is significant and may add insight into the individual/health relationship. The addition of the additional variables, insurance status and work status, in the Sensitivity model 2 analysis produces some significant relationships and reduces the number of eligible respondents. The primary purpose of the sensitivity analyses is to determine if other available variables are significant, impact the baseline variables, and are of sufficient interest to include in subsequent analyses. From the initial sensitivity regressions, the variable Married was determined to be a valid addition to the list of baseline variables. The additional variables in the Sensitivity models do not affect the magnitude or significance of the baseline variables dramatically and they are not included further[3].

Since occupation will be a key variable to examine in chapter 4, the age group is limited to that of standard working age people: 25 to 65 years of age. In addition, to minimize other possible external drivers of health, only respondents who are born in the U.S. are considered. These restrictions reduce the pool of respondents to 367,101 individuals.

Table 12: Initial Results from NLMS Analysis

Odds Ratios	Baseline	Sensitivity 1	Sensitivity 2
OUTCOME	All-Cause	All-Cause	All-Cause
VARIABLES/base			
Race:Black/white	1.376***	1.385***	1.478***
Race:Other/white	0.883**	0.908	0.767***
Female/male	0.411***	0.435***	0.455***
Rural/urban	0.895***	0.924***	0.966
Married/other than married	0.686***	0.736***	0.724***
Educ: None/high school H4	0.749**	0.811	0.706*
Educ: E1-E4/high school H4	1.026	1.050	1.031
Educ: E5-E6/high school H4	1.115**	1.134**	1.007
Educ: E7-E8/high school H4	1.160***	1.180***	1.122***
Educ: H1/high school H4	1.147***	1.147***	1.147**
Educ: H2/high school H4	1.175***	1.178***	1.173***
Educ: H3/high school H4	1.186***	1.188***	1.186***
Educ: H4	------	------	------
Educ: C1/high school H4	0.917**	0.914**	0.928
Educ: C2/high school H4	0.962	0.953	0.950
Educ: C3/high school H4	0.924	0.911*	0.891

---

[3] Subsequent analyses in chapter 3 and chapter 4 were performed with additional variables from the sensitivity analyses. There were no qualitative changes in the selected baseline variable results in these studies.

Odds Ratios	Baseline	Sensitivity 1	Sensitivity 2
OUTCOME	All-Cause	All-Cause	All-Cause
Educ: C4/high school H4	0.722***	0.713***	0.710***
Educ: C5/high school H4	0.633***	0.623***	0.582***
Educ: C6/high school H4	0.621***	0.615***	0.600***
Income:0-5K/25K-30K	1.420***	1.360***	1.190**
Income:5K-7.5K/25K-30K	1.501***	1.446***	1.133
Income:7.5K-10K /25K-30K	1.191***	1.157***	1.099
Income:10K-12.5K /25K-30K	1.279***	1.244***	1.107
Income:12.5K-15K /25K-30K	1.124***	1.102**	1.085
Income:15K-20K /25K-30K	1.091**	1.072*	1.016
Income20K-25K/25K-30K	1.061*	1.057*	1.036
Income25K-30K	------	------	------
Income:30K-35K/25K-30K	0.887***	0.879***	0.975
Income:35K-40K/25K-30K	0.861***	0.875***	0.863**
Income:40K-50K/25K-30K	0.758***	0.764***	0.859***
Income:50K-60K/25K-30K	0.870***	0.891***	0.979
Income:60K-75K/25K-30K	0.747***	0.772***	0.899*
Income:75K+/25K-30K	0.766***	0.798***	0.889**
Number In Household		0.973***	0.950***
Veteran/not veteran		1.088***	1.093***
House:Rent/owned		0.815***	0.810***
No Insurance/employer			1.110***
Insurance:Medicare/employer			1.439***
Insurance:Medicaid/employer			1.251**
Insurance:Other Govt Health/ employer			1.086
Insurance: Employer			------
Insurance:Private/employer			0.928*
Work:Unemployed, and looking/ employed			1.448***
Work:Unemployed/employed			1.361***
Work:Other/employed			1.498***
Constant	0.759***	0.861**	0.529***
Observations	355,449	351,099	178,510

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

The odds ratio results in Table 12 show that blacks have a higher mortality than whites, while other races are slightly better off than whites. Females have a lower mortality risk than males; and urban living has a higher mortality risk than rural living. Being married is better than not being married, i.e., divorced, widowed, or never married. There is a clear relationship with both education and income. For Education, having less than a high school diploma is generally worse, and usually significantly worse, for mortality, while having more education is better. Similarly, for income, having less income than the baseline of $25–30K per year is significantly worse for health, while having more income is consistently and significantly better. Based on the Number in Household variable, having your family around you is better for your health. Being a veteran is significantly worse for health, perhaps due to effects of physical injuries suffered or due to long-term stress effects on health from participating in military actions.

Owning your home is better than renting. This result may be another reflection of income or education, i.e., a socioeconomic result that has a basis in several related causes. Including Housing does not affect the significance of the income variables but each variable value is attenuated. Having health insurance is better than not having health insurance. Employer-based insurance is better than Government-supplied health care through Medicare or Medicaid, or being a Government employee. Private health insurance provided individually and not through an employer appears to be better, but is not a significant result. Hadley (Hadley 1988) has shown that insurance is endogenous, however; so, these results, although consistent with expectations, would require additional methodological analysis to address the endogeneity. Finally, working for a living is better for mortality than being unemployed or retired. Once again, this could be a confounding socioeconomic variable with education and/or income or could be a explanatory variable with the insurance variables as many people who work have employer-based health insurance. Both the income and education variables are attenuated and are not significant with the addition of the insurance and working variables. This could reflect confounding or the results could be due to selection effects noted by the large reduction in the number of observations when these variables are added to the model. The baseline model represents a reasonable set of socioeconomic explanatory variables for the following analyses.

3.7.2 Impact of Age Groups

Health gets worse with age, and the inclusion of age in these analyses is crucial. Alternatives to the age dummies include using a continuous age variable with an age-squared variable, and using a variety of categorical age groupings. Table 13 shows the results of using these different age groupings. The remaining detailed baseline results do not change dramatically in magnitude or significance and, except for a selected smaller set, are not shown. The age results, using continuous Age and Age2 variables, are in the first column. Although the age-squared variable is significant, the odds ratio is not much different than one, which does not provide insight into whether health is increasingly worse with age, or demonstrates a decreasing impact over time[4]. The second column has

[4] The coefficients in the analyses are small, but positive, indicating an increasing impact of age on health over time.

two categories, using dummy variables with ages from 25–44 and ages from 45–65. Each value shows the results of that age range relative to all other ages. So, respondents with ages 25–44 die less often than all other ages, while older respondents die more often.

Table 13: Impact of Age

Odds Ratios	Age & Age^2	2 Age Groups	4 Age Groups	Multi-valued Age
OUTCOME	All-Cause	All-Cause	All-Cause	All-Cause
VARIABLES/base				
Age, mean-centered	1.098***			
Age^2	1.000***			
Age 25_44		0.399***		
Age 45_65		2.501***		
Age 25_34			0.263***	
Age 35_44			0.594***	
Age 45_54			1.588***	
Age 55_65			3.881***	
Ages 35-44/Ages 25-34				2.409***
Ages 45-54/Ages 25-34				6.589***
Ages 55-65/Ages 25-34				17.36***
Race:Black/white	1.375***	1.204***	1.247***	1.353***
Race:Other/white	0.882**	0.779***	0.802***	0.873**
Female/male	0.411***	0.453***	0.448***	0.416***
Rural/urban	0.894***	0.847***	0.860***	0.887***
Married/other than married	0.685***	1.306***	1.290***	0.698***
Constant	0.0695***	0.0588***	0.0596***	0.0259***
Observations	355,449	480,494	480,494	348,989

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

Column 3 contains four categories, and once again the results are relative to all other ages. There is a clear, and expected, increase in mortality with age. The younger categories have a lower odds of dying and the older categories have a higher odds. Finally, the last column uses a single multi-valued variable where the baseline category is ages 25–34, and includes only ages from 25 to 65. Relative to this baseline age category containing working age adults with the lowest odds of mortality, all other age categories have an increased and significant odds, with the odds increasing with age.

The few remaining baseline variables shown are typical of the impact on the baseline variables when using detailed age categories. There is some attenuation of the results for the approaches in the second and third columns in which all ages are used. The only dramatic change is in the married results in column 2 and column 3. Being married becomes less healthy when all ages are considered.

The results in column 1 and column 4, which limit the age range to 25 to 65 years of age, are nearly indistinguishable from the results in the original baseline analysis using age dummies. These age analysis approaches were used for many subsequent analyses, and the overall results compared to analyses using the age dummies. The differences were small and are not reported further in this dissertation.

3.7.3 Causes of Death as the Dependent Variable

Table 14 and Table 15 show the relationships between the baseline explanatory variables (age, race, gender, rural/urban, marital status, education, and income) and the age- and race-adjusted causes of death. These dependent variables were constructed to match the cause of death variables used in chapter 2. The most reported on conditions in the published literature are Tumor-related and Cardiovascular-related mortalities.

Table 14: Baseline Analyses Using Causes of Death

Odds Ratios	Baseline	Baseline	Baseline	Baseline
OUTCOME	Tumors	Cardiovascular	Injury	Other
VARIABLES/base				
Race:Black/white	1.230***	1.279***	1.201**	1.558***
Race:Other/white	0.735***	0.812*	1.227	1.064
Female/male	0.705***	0.292***	0.302***	0.412***
Rural/urban	0.915***	0.925***	1.164***	0.759***
Married/other than married	0.859***	0.743***	0.657***	0.547***
Educ: None/high school-H4	0.881	0.628*	1.388	0.646
Educ: E1-E4/high school-H4	0.988	1.075	0.829	1.010
Educ: E5-E6/high school-H4	1.059	1.034	1.244	1.158*
Educ: E7-E8/high school-H4	1.132***	1.173***	1.281***	1.053
Educ: H1/high school-H4	1.128**	1.127*	0.955	1.208***
Educ: H2/high school-H4	1.173***	1.226***	1.110	1.061
Educ: H3/high school-H4	1.179***	1.156**	1.186	1.146*
Educ: C1/high school-H4	0.953	0.908	0.931	0.895
Educ: C2/high school-H4	0.992	0.961	0.879	0.986
Educ: C3/high school-H4	0.980	0.844*	0.841	1.016
Educ: C4/high school-H4	0.805***	0.708***	0.596***	0.776***
Educ: C5/high school-H4	0.703***	0.558***	0.403***	0.849
Educ: C6/high school-H4	0.723***	0.591***	0.534***	0.670***
Income:0-5K/25K-30K	1.144	1.266***	1.801***	1.538***
Income:5K-7.5K/25K-30K	1.300***	1.353***	1.651***	1.566***
Income:7.5K-10K /25K-30K	0.999	1.155	1.156	1.436***
Income:10K-12.5K /25K-30K	1.143*	1.277***	1.732***	1.160*
Income:12.5K-15K /25K-30K	0.986	1.164**	1.142	1.217***
Income:15K-20K /25K-30K	1.117*	1.024	0.994	1.151**
Income20K-25K/25K-30K	1.026	1.012	1.024	1.170**
Income:30K-35K/25K-30K	0.929	0.780***	0.855	1.022
Income:35K-40K/25K-30K	0.930	0.824***	0.824**	0.872**
Income:40K-50K/25K-30K	0.821***	0.720***	0.769**	0.785***
Income:50K-60K/25K-30K	0.988	0.852***	0.856*	0.755***
Income:60K-75K/25K-30K	0.829***	0.714***	0.837	0.692***
Income:75K+/25K-30K	0.874**	0.698***	0.846	0.720***
Constant	0.135***	0.248***	0.0121***	0.167***
Observations	355,449	355,449	355,449	355,449

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

Table 15: Age Impacts Using Causes of Death

Odds Ratios	Age/ Age^2	Ages	Age/ Age^2	Ages	Age/ Age^2	Ages	Age/ Age^2	Ages
OUTCOME	Tumors	Tumors	Cardio.	Cardio.	Injury	Injury	Other	Other
VARIABLES/base								
Age, mean-centered	1.135***		1.149***		1.002		1.084***	
Age^2	0.999***		0.999***		1.000		1.000**	
Ages 35-44/Ages 25-34		3.766***		5.477***		1.012		2.208***
Ages 45-54/Ages 25-34		12.21***		17.73***		1.056		5.193***
Ages 55-65/Ages 25-34		28.70***		47.98***		1.139*		12.66***

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

Race effects track to prior results in this dissertation: worse health for Blacks in all categories, and generally better health for Other races relative to Whites. Females also are healthier overall. Rural living is better for all results except Injury-related deaths. Education results generally follow those for the baseline analysis, with the Tumor-related and Cardiovascular-related categories following most closely. That is, there is a higher risk of death with less than a high school education, and lower risk with more than a high school education. The Other category follows the trend, but is not often significant. The only consistently significant values are for College Graduates (Educ:C4) and higher level graduates (Educ:C6). Income results follow the baseline trend in all categories. Those households making less than $25,000/year have worse health and those making more than $25,000/year have better health relative to the $25,000/year group. Finally, the Age results, in Table 15, all follow expectations, demonstrating higher risk of death in older people relative to the 25-34 age group. The trend in Injury-related deaths is the same, but the results are not significant, and are much smaller in magnitude.

3.7.4 Impact of Geographic Location

Table 16 shows how geography impacts the outcome variable. The State of Birth and related constructed variables – BirthDivision and BirthRegion – had no appreciable association with mortality and are not used in any further analyses. The focus is on urban/rural, SMSA status, and the set of variables related to State of Residence. The remaining detailed baseline results do not change dramatically in magnitude or significance and, except for a selected smaller set, are not shown.

Table 16: Impact of Geographic Variables on Mortality in the NLMS

Odds Ratios	Rural/Urban	SMSA	Residence State	Residence Region	Residence Division
OUTCOME	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause
VARIABLES/base					
Race:Black/white	1.376***	1.362***	1.304***	1.372***	1.356***
Race:Other/white	0.883**	0.880**	0.887*	0.903*	0.896*
Female/male	0.411***	0.412***	0.411***	0.412***	0.412***
Married/other than married	0.686***	0.684***	0.683***	0.670***	0.674***
Rural/urban	0.895***				
SMSA: not City/central city		0.930***			
SMSA: not SMSA/central city		0.890***			
Alabama/New York			1.055		
Alaska/New York			1.321***		
Arizona/New York			0.912		
Arkansas/New York			0.875		
California/New York			1.072		
Colorado/New York			1.047		
Connecticut/New York			1.042		
Delaware/New York			1.189**		
District of Columbia/New York			1.492***		
Florida/New York			1.121**		
Georgia/New York			1.144**		
Hawaii/New York			1.041		
Idaho/New York			1.078		

Odds Ratios	Rural/Urban	SMSA	Residence State	Residence Region	Residence Division
OUTCOME	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause
Illinois/New York			1.101*		
Indiana/New York			1.012		
Iowa/New York			0.859**		
Kansas/New York			0.970		
Kentucky/New York			1.195**		
Louisiana/New York			1.257***		
Maine/New York			0.765***		
Maryland/New York			1.175***		
Massachusetts/New York			1.018		
Michigan/New York			1.160***		
Minnesota/New York			0.935		
Mississippi/New York			1.161**		
Missouri/New York			1.083		
Montana/New York			0.919		
Nebraska/New York			1.032		
Nevada/New York			1.306***		
New Hampshire/New York			1.139		
New Jersey/New York			1.105*		
New Mexico/New York			0.907		
New York/New York			------		
North Carolina/New York			1.196***		
North Dakota/New York			0.869*		
Ohio/New York			1.061		
Oklahoma/New York			1.180**		
Oregon/New York			0.833**		
Pennsylvania/New York			1.044		
Rhode Island/New York			1.128		
South Carolina/New York			1.175**		
South Dakota/New York			0.817***		
Tennessee/New York			1.129*		
Texas/New York			1.125**		
Utah/New York			0.844**		
Vermont/New York			0.911		
Virginia/New York			1.108		
Washington/New York			0.994		
West Virginia/New York			0.885		
Wisconsin/New York			1.031		

Odds Ratios	Rural/Urban	SMSA	Residence State	Residence Region	Residence Division
OUTCOME	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause
Wyoming/New York			0.957		
Region: Northeast/ Northeast				------	
Region: Midwest/ Northeast				1.001	
Region: South/ Northeast				1.111***	
Region: West/ Northeast				1.005	
Division: New England/ Middle Atlantic					0.966
Division: Middle Atlantic/ Middle Atlantic					------
Division: East North Central/ Middle Atlantic					1.043
Division: West North Central/ Middle Atlantic					0.916***
Division: South Atlantic/ Middle Atlantic					1.108***
Division: East South Central/ Middle Atlantic					1.095**
Division: West South Central/ Middle Atlantic					1.079**
Division: Mountain/ Middle Atlantic					0.966
Division: Pacific/ Middle Atlantic					1.014
Constant	0.759***	0.763***	0.686***	0.726***	0.736***
Observations	355,449	355,449	355,449	355,449	355,449

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

The results for the baseline variables are nearly identical in all the analyses. There is minor attenuation in some variables, which would be expected with the addition of more explanatory variables to the analysis. The significance of the baseline variables does not change. In the rural/urban analysis, there is a lower risk when living in rural areas. Using the SMSA status multi-valued variable, the non-urban values (living in the not-city-central areas and in the not-SMSA areas) are both significant and demonstrate positive health results relative to urban living. The Not-SMSA category has a larger impact and is similar to the Rural result in the previous analysis as would be expected, since it represents the most rural of the three variable values. The State analysis uses New York as the baseline state. Any state can be selected as the baseline, and if one chooses the state with the lowest association with the mortality indicator (e.g., Utah or South Dakota) or the state with the highest mortality (i.e., Washington DC), then the logit results can be driven to be significant for every other state (either positive or negative, respectively).

For this analysis, I selected a state (New York) as more representative of a state with average health effects and a state whose health impacts should be interesting. For this Stata analysis, there are 24 significant results at the 10% significance level or better. As New York is an average state health wise, the significant state results reflect those that are significantly worse (e.g., Louisiana, Michigan, Illinois, and New Jersey) and those that are significantly better (e.g., Iowa, Maine, North Dakota, South Dakota, and Utah).

Relative to the Northeast Region (since it contains New York), the only region of significance is the South and the positive coefficient implies worse health. The other two regions reflect better health, but are not significant. With the Middle Atlantic Division (which includes New York) as the baseline, New England, West North Central, and Mountain Divisions have a lower mortality risk. This makes sense as these divisions are made up of the states with better overall health. The South Atlantic, East South Central, and West South Central demonstrate significantly worse mortality. Once again, this makes sense since the Southern states are worse health wise.

3.7.5 Results with Interaction Terms

Interaction terms are used to investigate the indirect effect of a variable Z on a variable X that has a direct effect on the dependent variable Y. Interaction terms are typically constructed by creating the product of variables X and Z, and including the main variables and the product in the analysis. Several analyses were performed, focusing on the interactions between the rural percentage variable and the geographic variables, and between several of the baseline socio-demographic variables and the geographic variables. If the interaction term is significant, the results can be interpreted as, for example, the impact of Geography (e.g., Census Region) on health outcomes differs depending on the urban or rural nature of the Census Region.

Without interactions, the current model assumes that the contextual effect of geography is the same for all genders, all education groups, all income categories, all ages, and all races. Adding interaction terms adjusts this assumption to allow the impact of geography on the rate of dying to depend on race, gender, education, etc. All of the interaction analyses in this section, and in subsection 3.7.7, display odds ratios as these allow straightforward interpretation of the interaction effects.

Table 17 shows interaction effects between Census Residence Divisions or Census Residence Regions and the Urban/Rural or SMSA status variables. The remaining detailed baseline results do not change dramatically in magnitude or significance and are not shown.

Table 17: Interaction Analysis between Region/Division and Rural/SMSA

Odds Ratios	Rural	Rural	SMSA	SMSA
	X Region	X Division	X Region	X Division
OUTCOME	All-Cause	All-Cause	All-Cause	All-Cause
VARIABLES/base				
Rural/urban	0.889***	0.937		
SMSA: not City/central city			0.986	1.003
SMSA: not SMSA/central city			0.939	1.087
Region: South/ Northeast	1.023		1.058	
Region: Midwest/ Northeast	1.118***		1.188***	
Region: West/ Northeast	1.005		1.072	
Division: New England/ Middle Atlantic		1.011		1.086
Division: East North Central/ Middle Atlantic		1.058*		1.130**
Division: West North Central/ Middle Atlantic		0.974		1.007
Division: South Atlantic/ Middle Atlantic		1.137***		1.293***
Division: East South Central/ Middle Atlantic		1.127**		1.175**
Division: West South Central/ Middle Atlantic		1.097**		1.151**
Division: Mountain/ Middle Atlantic		0.999		1.118*
Division: Pacific/ Middle Atlantic		1.014		1.085
Region: South X Rural	0.957			
Region: Midwest X Rural	1.020			
Region: West X Rural	1.003			
Division: New England X Rural		0.897		
Division: East North Central X Rural		0.961		
Division: West North Central X Rural		0.887*		
Division: South Atlantic X Rural		0.964		
Division: East South Central X Rural		0.969		
Division: West South Central X Rural		0.956		
Division: Mountain X Rural		0.924		
Division: Pacific X Rural		0.997		
Region: South X SMSA: not City			0.963	
Region: South X SMSA: not SMSA			0.909	
Region: Midwest X SMSA: not City			0.906*	
Region: Midwest X SMSA: not SMSA			0.937	
Region: West X SMSA: not City			0.914	
Region: West X SMSA: not SMSA			0.925	

Odds Ratios	Rural	Rural	SMSA	SMSA
	X Region	X Division	X Region	X Division
OUTCOME	All-Cause	All-Cause	All-Cause	All-Cause
Division: New England X SMSA: not City				0.942
Division: New England X SMSA: not SMSA				0.749***
Division: East North Central X SMSA: not City				0.909
Division: East North Central X SMSA: not SMSA				0.841**
Division: West North Central X SMSA: not City				1.015
Division: West North Central X SMSA: not SMSA				0.793***
Division: South Atlantic X SMSA: not City				0.811***
Division: South Atlantic X SMSA: not SMSA				0.786***
Division: East South Central X SMSA: not City				0.956
Division: East South Central X SMSA: not SMSA				0.849
Division: West South Central X SMSA: not City				1.006
Division: West South Central X SMSA: not SMSA				0.800**
Division: Mountain X SMSA: not City				0.863*
Division: Mountain X SMSA: not SMSA				0.757***
Division: Pacific X SMSA: not City				0.918
Division: Pacific X SMSA: not SMSA				0.865
Observations	355,449	355,449	355,449	355,449

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

For both Regions and Divisions, living in Rural areas is healthier (odds ratio < 1). Similarly, relative to SMSA Central City, suburban living and non-SMSA living are healthier. The Northeast Region was selected as the baseline region for these analyses since it is the healthiest region and results in negative health impacts for the other three regions in both the Rural analysis and the SMSA analysis. For divisions, the Middle Atlantic Division is one of the healthiest. Using the Middle Atlantic Division as the baseline results in odds ratios larger than one – indicating worse health – for the majority of the remaining divisions, and significant results in several of the divisions.

When interpreting interactions in logistic regression, using odds ratios is generally more straightforward than using the log odds coefficients. The interaction effects between Rural and Regions are not significant. Let me discuss two of the significant interactions – each of the remaining interactions is interpreted similarly. For the Division/Rural analysis, the odds ratio for Rural (0.937) is the odds of Rural people dying relative to Urban people in the Middle Atlantic Division (Division=0). There is a benefit to rural living in the Middle Atlantic states (odds ratio <1). The odds ratio for the West North Central Division (0.974) is the odds of Urban people dying in the West North Central (Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, and South Dakota) relative to Urban people in the Middle Atlantic (Rural=0). It is more healthy to live in urban areas in the upper midwest, but it is not significant. The significant interaction for West North Central (0.887) means that the impact of living in rural areas is 0.89 times the impact on urban living, i.e., the odds of dying in Rural areas for the West North Central is (0.974 * 0.887 =) 0.86 times that of dying in Urban areas in the West North Central. Alternatively, the odds of dying in Rural areas in the West North Central relative to Rural areas in Middle Atlantic is (0.937 * 0.887=) 0.83.

Similarly, for the Division/SMSA analysis, the odds ratio for SMSA not City Central (SnCC) (1.003) is the odds of SnCC people dying relative to SMSA City Central (SCC) people in the Middle Atlantic Division (Division=0), but it is not significant. The odds ratio of not SMSA (nS) (1.087) is the odds of nS people dying relative to SCC people in the Middle Atlantic Division (Division=0) and is also not significant. The significant odds ratio for the South Atlantic Division (1.293) is the odds of SCC people dying in the South Atlantic (Delaware, Washington DC, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, and West Virginia) relative to SCC people in the Middle Atlantic Division (SMSA=0). So, it is less healthy to live in SMSA City Central areas in the South Atlantic than in Middle Atlantic. The significant interaction for South Atlantic (0.811) means that the impact of living in SnCC areas is 0.81 times the impact on SCC living, i.e., the odds of dying in SnCC areas is (1.293 * 0.811 =) 1.05 times as high as that of dying in SCC areas in the South Atlantic. Alternatively, the odds of dying in SnCC areas in the South Atlantic relative to SnCC areas in Middle Atlantic is (1.003 * 0.81 =) 0.81.

Other interaction analyses were performed, including interacting the Urban/Rural variable with Gender, Marital Status, Race, a High School Graduation dummy, a College Graduation dummy, a five-category income variable, and Veteran status. These analyses indicate whether it is healthier, for example, to be married in urban or rural settings or to be a female in urban or rural settings.

Table 18 shows the results. The remaining detailed baseline results do not change dramatically in magnitude or significance and, except for the relevant interaction terms, are not shown.

There are few significant interaction terms which implies there are not many differences in Urban living versus Rural living with respect to the impact of these demographic variables (Gender, Race, Marital status, etc.) on mortality. Two significant results are interpreted as follows. For the high school education result, the odds ratio for Rural (0.945) is the odds of dying for Rural non-graduates divided by the odds of dying for Urban non-graduates (Educ_hs=0) – so there is some benefit (although not significant) to living in rural areas (odds ratio < 1). The odds ratio for graduating high school is 0.811, which means that the odds of dying is less for urban people with a high school degree than for those without a degree (Rural=0). The significant interaction effect means that the impact of living in rural areas is 0.942 times the impact of having a degree on urban people, i.e., the odds of dying in Rural areas for graduates is (0.811 * 0.942 =) 0.76 times that of non-graduates in Rural areas, and living in Rural areas is a benefit. Alternatively, the odds of dying for Rural graduates relative to Urban graduates is (0.945*0.942 =) 0.89.

Table 18: Urban/Rural Interaction with Demographic Variables

Odds Ratios							
Interact with:	Race3	Female	High School Education	College Education	Income	Veteran Status	Married
OUTCOME	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause
VARIABLES/base							
Rural/urban	0.883***	0.896***	0.945*	0.906***	0.880**	0.902***	0.852***

Odds Ratios							
Interact with:	Race3	Female	High School Education	College Education	Income	Veteran Status	Married
OUTCOME	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause
Race:Black/white	1.360***	1.380***	1.396***	1.414***	1.384***	1.380***	1.378***
Race:Other/white	0.748***	0.891*	0.881**	0.891*	0.892*	0.891*	0.892*
Rural X Race3:Black	1.084						
Rural X Race3:Other	1.660***						
Female/male	0.431***	0.431***	0.448***	0.424***	0.432***	0.432***	0.431***
Rural X Female		1.002					
High School Grad			0.811***				
Rural X High School			0.942*				
College Grad				0.665***			
Rural X College Grad				0.980			
Income:10K-20K/0-10K					0.831***		
Income:20K-35K/0-10K					0.738***		
Income:35K-50K/0-10K					0.600***		
Income: 50K+/ 0-10K					0.581***		
Rural X Income10K-20K					1.004		
Rural X Income20K-35K					0.969		
Rural X Income35K-50K					0.994		
Rural X Income50K+					1.127*		
Veteran/not a veteran	1.083***	1.084***	1.102***	1.068***	1.086***	1.090***	1.084***
Rural X Veteran						0.983	
Married/not married	0.686***	0.685***	0.701***	0.692***	0.682***	0.685***	0.674***
Rural X Married							1.067
Observations	355,405	355,405	355,485	355,485	355,405	355,405	355,405

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

For the Other Race result, the significant odds ratio for Rural (0.883) is the odds of Rural white folks dying relative to the odds of Urban white folks (Race3=0). For whites, living in rural areas is a plus. The significant odds ratio for Blacks (1.36) means the odds of dying is 1.36 times higher for Urban blacks as for Urban whites (Rural=0).

The significant odds ratio for Other Race (0.748) means that the odds of dying is less for other races in Urban areas than it is for Urban whites (Rural=0). The significant interaction effect for Other Race means that the impact of living in rural areas is 1.66 times the impact of race on urban people, i.e., the odds of dying in Rural areas for Other Races is (0.748 * 1.66 =) 1.24 times higher than that of Whites in Rural areas. In this case, for Other Races, rural living is not a plus. Alternatively, the odds of dying for Other Races in Rural areas relative to the odds of dying for Other Races in Urban areas is (0.883*1.66=) 1.47.

3.7.6 Multi-level Analysis Results

A multi-level analysis was performed to determine the relative contributions from multiple levels of data. In this case, individuals are considered the lowest level of data, and States are considered groupings of individuals, the 2nd-level data. Since the data contains the State of Residence, the individuals can be grouped by State; and the State-level variables from chapter 2 provide demographic data at the State level. The intent is to determine the relative contributions to the overall variance by individuals (within-state variance) and States (between-state variance).

A null baseline analysis (a random effects binary model using the Stata command *xtmelogit*) indicates that the intra-class correlation is about 0.0055, which means that only about 0.55% of the total variance is contributed by the between-state variance. Two tests help interpret the significance of this correlation. First a Wald test gives a value of 17.0, which is compared with a chi-squared distribution on one degree of freedom, giving a p-value < 0.001. Second, the likelihood ratio statistic is reported as 162.0 with a corresponding p-value < 0.00005. Both tests provide strong evidence that the between-state variance is non-zero, i.e., there is significant variation between the states in the risk of death. This is illustrated in the "caterpillar" plot in Figure 3 showing the estimated state residuals shown in rank order. A state whose confidence interval does not overlap the zero line differs significantly from the average at the 5% level.

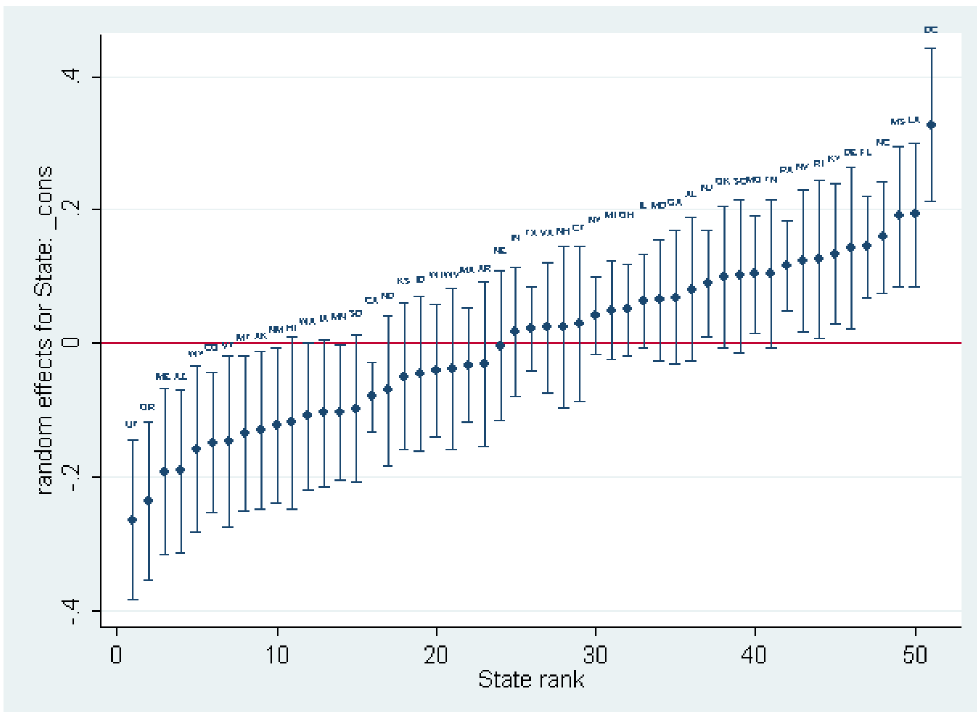

Figure 3: State Residuals with 95% Confidence Intervals

There are 21 states whose confidence interval does not overlap the zero line. The states with the lowest probability of dying (at the bottom left of Figure 3) are Utah, Oregon, Maine, and Arizona; the states with the highest probability of dying are Mississippi, Louisiana, and the District of Columbia.

The next step is to add the baseline variables (age, race, gender, rural, marital status, education, and income. Focusing on income, I now allow the intercept and slope to vary randomly across states. The regression results show a negative income coefficient meaning that increased individual income reduces the risk of death; and shows a negative intercept-slope covariance estimate, which implies that those states with above average mortality risk have below average effects of income. This is illustrated in Figure 4, which shows the estimated slope and intercept residuals for the relationship between the log-odds of dying and income. The states, like the District of Columbia, at the lower right have a higher risk of death after controlling for income, and a weaker relationship between the risk of death and income.

Now, I add a second-level variable: health expenditures per capita as defined in chapter 2. This variable is endogenous with health outcomes, but Stata lacks a straight forward approach for managing endogeneity in binary multi-level regressions. As described by Terza et al. (Terza, A Basu, and Rathouz 2008) the two-stage residual inclusion approach can be used when the dependent variable is binary. Future study should make use of this technique to validate these results. Table 19 shows the results of this analysis. The health expenditure results should be interpreted with caution, as chapter 2 demonstrated that controlling for endogeneity changed the sign of the relationship with all-cause mortality.

The results indicate that states with higher health expenditures are more likely than states with lower health expenditures to have high death rates. The between-state variance dropped substantially over the null model, suggesting that the distribution of one or more of the explanatory variables varies across the states. This is reasonable since some states will have higher proportions of educated citizens and households with relatively more income than others.

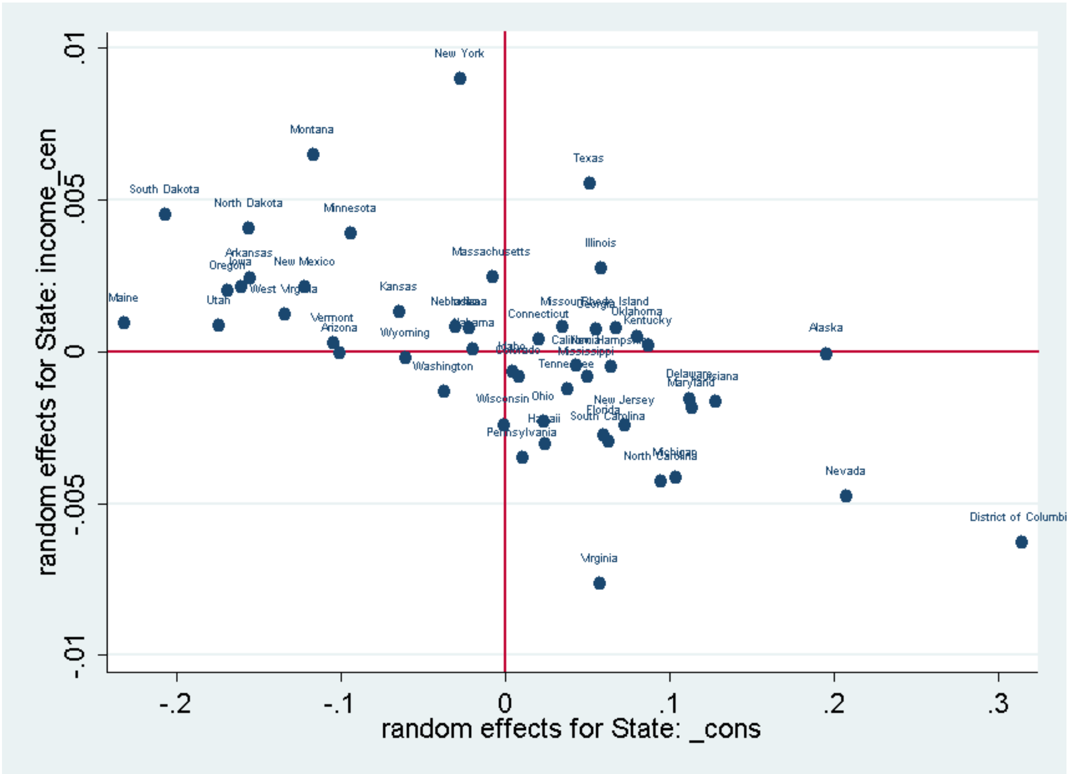

Figure 4: Estimated Slope and Intercept Residuals for the Risk of Death and Income

Table 19: Multi-level Analysis Results

OUTCOME	All-Cause	All-Cause
VARIABLES/base	**Coefficients**	**Odds Ratios**
Race:Black/white	0.257***	1.293***
Race:Other/white	-0.220***	0.803***
Female/male	-0.773***	0.462***
High School Grad/not high school grad	-0.115***	0.891***
College Grad/not college grad	-0.623***	0.536***
Income quintile 2/income 1	-0.170***	0.844***
Income quintile 3/income 1	-0.426***	0.653***
Income quintile 4/income 1	-0.638***	0.528***
Income quintile 5/income 1	-0.562***	0.570***
Health Expenditure quintile 1/expenditure 5	-0.176***	0.839***
Health Expenditure quintile 2/expenditure 5	-0.176***	0.839***
Health Expenditure quintile 3/expenditure 5	-0.0686	0.934
Health Expenditure quintile 4/expenditure 5	-0.529	0.589
Observations	365,673	365,673

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

The final multi-level analysis adds cross-level interaction effects. In this case, the interaction is between mean state income and individual income. The results are summarized in Figure 5.

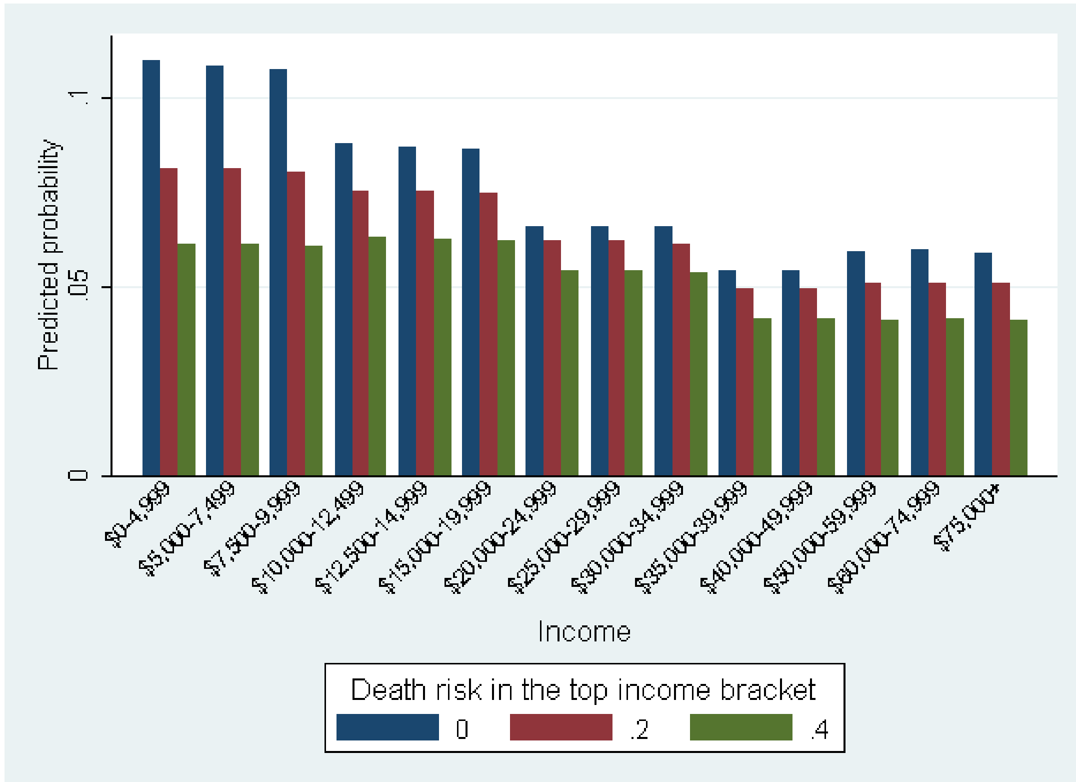

Figure 5: Predicted Probabilities by Individual Income and State Income

The x-axis shows the baseline income categories; each category shows three state income groups (0, 0.2, and 0.4, representing low, medium, and high average state income values). The effect of state income, i.e., the difference in predicted probabilities for the three state averages, is weaker in the middle income categories and strongest in the lowest income categories. Living in a more deprived state appears riskier for poorer individuals than living in a better-off state. In addition, the effect of individual income is stronger in poorer states, i.e., there is a greater impact in the blue bars than in the green bars across Figure 5 from left to right.

3.7.7 Factor Analyses Results

The State-level factors described in chapter 2 are used in this chapter as well. In Table 20, the remaining detailed baseline results do not change dramatically in magnitude or significance and, except for a selected smaller set, are not shown.

Table 20: State Factor Interaction with Rural

Odds Ratios	Baseline		Rural	Rural
	Rural	Factors	+ Factors	X Factors
OUTCOME	All-Cause	All-Cause	All-Cause	All-Cause
VARIABLES/base				
Rural/urban	0.895***		0.904***	1.095
1. Big Fish, Small Pond		0.847***	0.856***	0.906**
2.Up-and-Comers		1.178***	1.151***	1.121**
3.Heartlanders		0.871*	0.878*	0.913
4. Empty Nesters		1.065*	1.075**	1.100**
Rural X Factor 1				0.848**
Rural X Factor 2				1.080
Rural X Factor 3				0.893
Rural X Factor 4				0.912
Observations	355,449	352,004	352,004	352,004

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

Big Fish, Small Pond states have high rankings in education, health, and low crime rates. Up-and-Comer states have high ranks in income, graduation percentage, the happiness index, and the freedom index. Heartlanders states have high manufacturing and farm output, and a high percentage of church attendance. Empty Nester states have a high percentage of people over 45 years of age and a low births per capita rank.

When the State-level factors substitute for Rural as the geography variables (Column 2) each is significant at the 10% level or better. Factor 1 and Factor 3 have a positive impact on health, and Factor 2 and Factor 4 have a negative impact on health. Adding Rural in Column 3 does not change the results. One significant interaction term occurs with Rural, indicating that rural living appears to have an additional effect on the

Big Fish, Small Pond states. Factor 3 Heartlanders shows a positive health impact in this analysis. Adding the interaction effects causes the rural coefficient to switch signs and indicate that more rural areas appear less healthy. The Rural odds ratio (1.095) means that the odds of dying for Rural people in states at the bottom of the Factor 1 scale (Factor 1=0) have a higher risk of dying than Urban people in states at the bottom of the Factor 1 scale. The odds ratio for Factor 1 (0.906) indicates that the odds of dying is less for Urban people in states near the top of the Factor 1 scale versus Urban people at the bottom of the Factor 1 scale (Rural=0).

The significant interaction effect (0.848) means that the impact of living in Rural areas of states at the top of the Factor 1 scale is 0.85 times the impact of living in Rural areas for people near the bottom of the Factor 1 scale (0.85 * 0.91 = 0.77), i.e., people in Rural areas at the top of the scale have a much lower risk of dying. Also, the odds of dying for Rural people at the top of the scale versus Urban people at the top of the scale is (1.095 * .85 =) 0.93.

Another set of interaction analyses, in Table 21, were performed to determine if Race, Gender or Marital Status interact with the State factors. The remaining detailed baseline results do not change dramatically in magnitude or significance and, except for those used in the interaction analyses, are not shown.

Table 21: State Factor Interaction with Demographic Variables

Odds Ratios	Race	Female	Married
	X Factors	X Factors	X Factors
OUTCOME	All-Cause	All-Cause	All-Cause
VARIABLES/base			
Female/male	0.411***	0.519***	0.412***
Race:Black/white	1.102	1.317***	1.317***
Race:Other/white	1.027	0.924	0.927
Married/other than married	0.690***	0.688***	0.638***
Factor 1: Big Fish, Small Pond	0.844***	0.825***	0.802***
Factor 2: Up-and-Comers	1.164***	1.205***	1.006
Factor 3: Heartlanders	0.895	0.931	0.983
Factor 4: Empty Nesters	1.078**	1.133***	0.931
Race:Black X Factor 1	0.944		
Race:Other X Factor 1	4.336***		
Race:Black X Factor 2	0.920		
Race:Other X Factor 2	0.660		
Race:Black X Factor 3	1.291		
Race:Other X Factor 3	0.322***		
Race:Black X Factor 4	1.031		
Race:Other X Factor 4	1.139		
Female X Factor 1		1.129	
Female X Factor 2		0.859*	
Female X Factor 3		0.820	
Female X Factor 4		0.837**	
Married X Factor 1			1.100
Married X Factor 2			1.208**
Married X Factor 3			0.852
Married X Factor 4			1.222***
Observations	352,004	352,004	352,004

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

Big Fish, Small Pond states have high rankings in education, health, and low crime rates. Up-and-Comer states have high ranks in income, graduation percentage, the happiness index, and the freedom index. Heartlanders states have high manufacturing and farm output, and a high percentage of church attendance. Empty Nester states have a high percentage of people over 45 years of age and a low births per capita rank.

For the most part, the baseline variables shown do not vary much when including the interaction terms. The Race variables are not significant when the Race X Factor interactions are present; Female and Married remain significant but the values change. For the significant state factors, the magnitudes of the factors do not change dramatically in the presence of the interaction terms.

The Race interaction terms suggest that for other races it is better to live in the Heartland states and worse to live in the Big Fish, Small Pond states. Females appear even better off in the Up-and-Comer states and in the Empty Nester states; while married people are worse off in both of these factor groups.

3.8 Conclusions

This chapter presented an empirical analysis of the relationship between health care determinants and health outcomes using individual data, for several hundred thousand people in the National Longitudinal Mortality Study (NLMS). This study focuses on the economic, socio-demographic, and lifestyle factor effects on health outcomes. This study extends previous work by using the most recent, and comprehensive, version of the NLMS; considering the impact of age groupings on health outcomes; and examining the impact of geography, including interaction analyses with the key baseline variables, and the state-level factors.

The individual-level data in this chapter allows finer grained analyses of income, education, gender, race, and age than the analyses on aggregate data in chapter 2. In particular, the race data show consistently worse health for black men and women relative to whites, and generally better health for other non-white individuals relative to whites. Being female is always more healthy than being male. Living in rural areas (and suburban areas) is better for health than living in urban areas. Using high school graduation level education as the base value, those with less education have worse health, and those with more education have better health. There is also improved health for those with education beyond a 4-year college degree. Using $25-30K as the base value of household education (and not considering the number of household members), those with less household income have worse health, and those with more household income have better health. The gradients for both education and income move consistently as the education and income categories change from low levels to high levels, and are maintained even while controlling for a variety of other confounding variables. Employer-based insurance was always healthier than using Medicare, Medicaid, or TRICARE (Veteran's Affairs/military health coverage), although these data are subject to endogeneity bias. Being married is healthier than not being married; and being in the work force is healthier than being unemployed, being a student, or being retired. Two-stage least squares analyses, where the second stage is a logit/probit or Cox Proportional Hazard analysis, are not supported easily in Stata. Future work should consider how to incorporate an instrumental variable approach into these individual data analyses.

Using Census geographic regions and divisions, similar results are found as in chapter 2. That is, southern and south central states have worse health, while northern, north-eastern, and western states have better health. The state-level factors, resulting from the factor analysis described in chapter 2, represent different groupings of geography based on the state characteristics and demographic variables. All of the four state-level factors are significant at better than 10% ($p<0.10$) and the results are maintained when co-regressing the factors with the urban/rural variable or the SMSA variable. Factor 1 and Factor 3 represent better health, and Factor 2 and Factor 4 represent worse health.

Interaction effects are important to determine the relative impact of one explanatory variable on another which affects health. A number of interaction analyses were carried out using combinations of the variety of geographic variables and the variety of economic, demographic, and lifestyle variables available in the NLMS. Tables 17, 18, 19, 20, and 21 show relevant and representative results. For the most part, there are few significant interaction effects in any of the analyses. With the analyses reflecting any significant results at the 10% ($p<0.10$) level or better, just by chance I would expect to see about 10% of the results exhibiting significance. Most analyses show no more than, and sometimes less than, 10% significance. In Table 21 there are more interaction effects between the state-level factors and demographic variables. Odds ratios can be calculated from the displayed coefficients which allow a rapid determination of the interaction effects by multiplying the results. Once again, the interaction effects are primarily not significant.

Can we all be this similar across geographic definitions and demographics? There are numerous published reports of neighborhood effects on health. Aggregating individual data to the level of states and groups of states (as is done in the state factor analysis) is probably generalizing too far. That is, if the data were sufficient to identify Census areas, counties, zip code areas, or smaller neighborhood geographies, then there is a much better likelihood of seeing stronger geographic results and interaction effects. See (Weiss 2000) for an exposition on very granular geographic definitions and an ability to identify distinguishing characteristics with implications for focused marketing. The multi-level analyses indicate that state-level distinctions, at least in income, are sufficient to observe differences mortality risk, even when the between-state variance accounts for a small proportion of the overall variance.

The publicly-released version of the NLMS, used in this study, contains limited geographic individual identifying information; a restricted version of the data contains a few more location data, but still may not be sufficient. Future research could make use of the entire NLMS data set, or identify another data set that contains detailed location data, to determine and analyze smaller, geographically more interesting areas and the interactions with a state factor analysis – perhaps combining them with Weiss-like data to further refine the state-level factor analysis.

4. Impact of Occupation on Health Outcomes

4.1 Introduction

This chapter investigates, in more detail than previous studies, the concurrent impact of occupation and geographic factors on mortality and health in the United States. Social inequalities, including social position, social status, or social class, have long been recognized as socioeconomic contributors to mortality and morbidity. The data used to construct these potential determinants are multidimensional and include education, income, power, occupation, occupational prestige, poverty level, access to and knowledge of healthcare, income inequality (e.g., Gini coefficients), employment status, and the like.

Many papers in the United Kingdom and other OECD countries focus on occupation as a key socioeconomic indicator. Many studies in the U.S. use income and/or education, or an index of social status, such as the Duncan Socioeconomic Index measure. This chapter builds on the few papers that have used occupation as the social status measure in the United States, and includes detailed data generated from a factor analysis of occupation characteristics.

The results for the impact of occupations on health generally replicate previous results indicating that non-manual occupations promote better health than manual occupations, and that more prestigious occupations exhibit better health than less prestigious occupations. The 234 occupation characteristics allow for a factor analysis that provides more insight into psychosocial job characteristics, cognitive job characteristics, and physical and environmental job characteristics than any previously reported results. The impact of these job characteristics on health outcomes clarifies how occupations may actually affect health, and provides better definitions of terms than some previously used in health regressions. One key implication is that job IQ, that is, where the nature of the job is best defined by cognitive ability, originality, and reasoning ability may be the most consistent driver of the impact of occupations on health.

The chapter is organized as follows. First, the background section discusses the literature on occupational effects on health outcomes. Next, research questions and hypotheses are discussed. The occupation data and the methodologies used in the empirical analyses are then introduced. This is followed by a detailed discussion of the analytical results. Finally, concluding remarks are presented, together with a brief discussion on possible directions for future research.

4.2 Background

Investigators have repeatedly demonstrated that occupations, and status more generally, are strong factors in predicting health (Michael Marmot et al. 1997). Occupational prestige, social influence, and power are other ways of portraying status. As shown by Marmot and others, more prestigious occupations tend to have lower mortality and morbidity relative to less prestigious occupations. One explanation is the variation in the psychosocial characteristics of occupations, for example the control over one's job situation. International studies have used occupation or occupation status more often as a key socioeconomic status (SES) indicator, while U.S. studies have tended to focus more on income and education as the key SES indicators. The use of a particular indicator often depends on the data availability, the resilience of the data definitions, and the approach to capturing the data. In addition, considering the life course of individuals, there have been many approaches to using data that includes mother's and father's occupation relative to child health, initial occupation of the subjects, longest held occupation of the subjects, last occupation of the subjects, spouse's occupation, and so on. Braveman (Braveman et al. 2005) emphasized the importance for researchers to (1) include a variety of SES measures, (2) not to assume one measure can be interchanged for another, and (3) justify why a study includes a certain set of measures and not others.

The standard occupation definitions in the U.S. are in the Standard Occupational Classification (SOC) System or in the Occupational Information Network (O*NET) database defined by the Bureau of Labor Statistics (Bureau of Labor Statistics 2010a; Bureau of Labor Statistics 2010b). As with any classification system, there are limitations in the scope and level of detail possible. The most detailed occupation categories number well over 800 which pose a computational issue for most empirical analyses. SOC classifies the categories into smaller groupings that are more manageable from an analytical perspective, but which blur the lines of distinction between occupations. For example, the Major Occupation classification in the NLMS (described in subsection 4.3.2) puts Chief Executives, Education Administrators, Coroners, Personnel Recruiters, Tax Examiners and Auditors, and Building Inspectors in the Executive group. Perhaps one can argue, from a health perspective that the job requirements in these positions are similarly demanding and produce a similar impact on health; however, there are few similarities between them when considering the occupation prestige rankings of these Executive occupations. In addition, from a statistical perspective, groupings that are too general in their nature result in collinearity among the explanatory variables. To mitigate these issues, I determine a set of occupation characteristic factors, using factor analysis, taken from the detailed definitions of abilities, knowledge, skills, work activities, etc., defined for each of the 800-plus occupations in the O*NET database. These factors provide a new perspective on occupation relevance and the interpretation relative to health.

Health effects based on geographic locations and levels, e.g., cities, suburbs, rural areas, counties, states; as well as households, neighborhoods, census tracts, regions, and clusters (Diez Roux et al. 2001; Ellen, Mijanovich, and Dillman 2001; Subramanian, Kawachi, and Kennedy 2001; Oakes 2004) are frequently reported. Weiss (Weiss 2000) and others have created cluster models that are available to categorize you and me into geographic marketing groups that characterize our lives. These "neighborhood" effects affect health through physical characteristics, social characteristics, cultural characteristics, or other commonly associated characteristics of households. Many studies only identify large geographic areas, such as state of birth in the NLMS, in attempts to provide some data while de-identifying the survey participants. Finer-grained identifiers may not be captured or may only be allowed in data sets with tightly controlled distribution to protect individual privacy. For this study, the State of Birth is used, but this identifier may aggregate data to such an extent that results have little meaning. There may be more interesting detail about Census areas, counties, or zip code areas; ideally, data collection efforts in the future will capture more complete neighborhood data and characteristics. To mitigate these issues, I determine a set of state-level characteristic factors, using factor analysis, taken from a wide variety of such state rankings as: Smartest state, Healthiest state, state with the highest rate of citizens that exercise, etc. These factors provide a new perspective on geographic "neighborhoods" and the interpretation relative to health.

4.3 Data

The data used in this chapter is the National Longitudinal Mortality Study (NLMS) survey data used in chapter 3, supplemented by the NLMS occupation category data, the occupation factor analyses results, and the state-level factor analysis results described in chapter 2.

4.3.1 Data Sources

A version of the Department of Labor O*NET occupation database is used in factor analyses to identify underlying clusters of common characteristics about occupations.

The O*NET data represents the latest effort by the Department of Labor to create occupation definitions and define occupational characteristics. O*NET was developed to replace the Dictionary of Occupational Titles (DOT) which had been the public standard description of occupations since the late 1930s. The O*NET data, version: v.13[5] (Department of Labor 2009), are used in factor analyses to create a subset of the occupational characteristics sufficient to represent the key factors defining occupations. Two approaches are used to create factors. In the first approach, each of seven key sub-domains (Ability, Education/Experience, Knowledge, Skills, Work Activities, Work Context, and Work Styles) was analyzed separately. This effort created a reduced set of factors relevant to each sub-domain. For example, 52 Ability variables were reduced to four factors: Gross Motor Skills, Cognitive Ability, Fine Motor Abilities, and Auditory/Visual Processing, which accounted for 75% of the total variance. This analysis resulted in a total of 22 factors across the seven sub-domains. In the second factor analysis approach, the entire set of 234 variables (the total from across the seven sub-domains plus nine demographic variables) was analyzed together. This resulted in a more global set of four factors: Reasoning & Complexity, Physical Demands, People vs. Things, and Attention to Detail accounting for 58% of the total variance. See Appendix D for a more complete description of the factor analyses performed and how the resulting factors were determined.

4.3.2 Sample Construction

In order to have sufficient detail on occupations this dissertation uses four occupation classification categories. The most detailed category, called simply Occupation, is the three-digit occupation classification code, based on the 1990 Census

---

[5] Version 13 of the O*NET data was accessed through the Department of Labor O*NET website on February 8, 2009. This version contains complete data on 807 occupations for the occupation characteristics selected for use in this dissertation.

Index of Industries and Occupations, provided directly in the O*NET database and containing 807 occupations. The next grouping is a gender-specific grouping, called Occupation-Recode, based on a BLS Standard Occupation Classification (SOC) code system provided in the Release 2 version of the NLMS, but missing in Release 3. I reconstructed the groupings for males (a total of 88 occupations) and females (a total of 59 occupations) by mapping the 1990 occupation codes in Release 3 back to the 1980 occupation codes used in Release 2. The third grouping is the Major Occupation category (also based on the SOC codes and containing 18 occupations), provided directly in the NLMS. The most general grouping is modeled on the British Registrar General (BRG) definition of social status containing four categories: Professional, Clerical, Skilled Crafts, and Labor occupations. As with the Occupation-Recode categories, I constructed the BRG groupings by assigning the three-digit Occupation codes to the four BRG occupation definitions by gender. No analytical effort was made to study true compatibility of this classification with the British Registrar General's definition.

4.3.3 Dependent Variables

The primary dependent variable is a Death Indicator (=1 if the respondent was matched to an NDI record) which is renamed All-Cause Mortality in chapters 3 and 4. Overall, about 9.1% of respondents died during the follow-up period for this data set. The primary cause of death is coded in Cause1 using the International Classification of Diseases, ninth revision (ICD-9) codes. I recoded these values into a General Cause of Death variable that represents the same four general categories used in the analyses in chapter 2 and chapter 3: Tumor, Cardiovascular, Injury, and Other causes of death.

4.3.4 Explanatory Variables

The baseline explanatory variables are the same NLMS-based variables discussed in subsection 3.4.4 in chapter 3. In addition, I add the Occupation classification variables listed in Appendix E. These include the occupation categories described in section 4.3.2 along with the specific occupation descriptions shown in Appendix C, and the occupation factor analyses variables determined by the approach described in Appendix D and listed in Appendix E. An occupation prestige rank variable and a Duncan Socio-Economic Index variable are included as described in Appendix E. Standard industry codes are also supplied with the NLMS. These data identify the industries associated with the occupations that employ the responders. Sensitivity analyses were performed with the Industry variables, but no results are reported. Finally, I add the state-level factors described in Appendix F.

4.4 Methodology

The STATA statistical analysis package, v.11, is used for all analyses(StataVersion 11.1 2010). For formatting the regression tables, the user-supplied package OUTREG2 is used (Wada 2010). The Stata data files (*.dta) and analysis processing files (*.do) are available by request from the author.

4.4.1 Factor Analyses

Factor analyses were carried out to define a set of occupation-specific factors that represent key characteristics of occupations. See Appendix D for a detailed description and example for how the occupation factor analyses were carried out resulting in multiple Stata datasets used in the regression analyses described in the following Sections. As the occupation categories get more general, i.e., as the factor analyses move from using the Occupation category to the using the BRG categories, the resulting factors tend to become more collinear. This results in Stata dropping many, if not all, of the factors from the analyses for the more general occupation categories. As a result, all of the results reported in this chapter use the Occupation category factors, which are based on the full set of 807 occupations.

This chapter employs the State-level factor analysis described in chapter.

4.4.2 Logit Analyses

One approach used is a logistic approach with interaction effects. The basic specification estimated is:

$$\text{logit}(\pi) = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ + \beta_i Y_i + \varepsilon$$

where $logit(\pi)$ is the mortality proxy; X is an occupational factor; Z is a geographic factor; XZ is the interaction effect between X and Z; Y_i is a vector of the remaining economic, socio-demographic, or lifestyle factors; β_0 is a the intercept; and ε is a disturbance term. Some regressions use occupational dummy variables to represent any unaccounted for invariant characteristics of occupations. Other regressions use the occupation factor analysis results to determine the impact of the key factors of occupations on health. Finally, interaction effects between occupations and geography are used to determine if there is an impact of geography on the relationship of occupation to health. The state-level factor analysis results are used as another set of geographic variables and are interacted with occupations and occupation factors in the analyses below.

4.4.3 Cox Proportional Hazard Analyses

Another approach uses Cox proportional hazards regression to determine relative mortality differences among occupational groups after adjustment for the socio-demographic determinants. This is a standard approach used in prior studies (Johnson, Sorlie, and Backlund 1999; Sorlie, Backlund, and Keller 1995). As described in subsection 3.5.2 and the following subsection, several analyses report hazard ratios.

4.5 Study Sample Characteristics

This study analyzes the concurrent relationship between occupation as a health care determinant and geographic location while controlling for other socioeconomic and demographic conditions. Geographic definitions, such as an urban/rural designation, Standard Metropolitan Statistical Area (SMSA) status, and the U.S. State of Residence, are available in the NLMS data. Empirical results consistently point to urban residents (also SMSA city center residents) as having higher mortality and morbidity rates than rural residents (Hayward and Gorman 2004). These analyses are expected to demonstrate similar results.

When used as a proxy for socioeconomic status, occupation consistently demonstrates an impact on mortality. International studies (M. Marmot et al. 1991; Michael Marmot et al. 1997; Davey Smith et al. 1998; Volkers 2005; MacLeod et al. 2005) demonstrate a sharp inverse relationship between social class, as measured by grade of employment, and mortality for a wide range of diseases. Davey Smith et al. claim to perform one of the few analyses where occupational social class and education are used a co-determinants. For working age men, adjustment for occupational class greatly reduced the association of all-cause mortality with education, leading Davey Smith et al. to state that occupation is a better discriminator of SES differences.

In the United States, the following studies have used the few data sets that contain occupation and other SES variables: the NLMS (Sorlie, Backlund, and Keller 1995; Gregorio, Walsh, and Paturzo 1997; Johnson, Sorlie, and Backlund 1999; Muntaner et al. 2001); the Wisconsin Longitudinal Study (WLS) (Miech and Hauser 2001; Warren and Kuo 2003); the Panel Study of Income Dynamics (PSID) (Duncan et al. 2002; Sindelar et al. 2007; Fletcher, Sindelar, and Yamaguchi 2008; Fletcher and Sindelar 2009); and the Health and Retirement Study (Gueorguieva et al. 2009). The NLMS studies use broad general categories of occupation that tend to be inadequate for use as measures of occupational exposure, and determine that considerable reduction in the relative risks for occupations occurs when income, education, and other explanatory variables are added to the analyses. Sorlie et al. suggest this means that these occupational groups are a less satisfactory measure of social class in the United States. Johnson et al. conclude that the BRG groupings and their 11-category occupational grouping do not represent adequate measures on socioeconomic status. They suggest that occupational differences should include measures on specific job characteristics like control, stress, decision latitude, and complexity. The WLS studies conclude that what people do for a living matters for health above and beyond the impacts of education. But to appreciate the full nature of the effects, job characteristics and job requirements should be measured, not just occupations. In the PSID studies, some job characteristics are identified and used, e.g., physical demands and environmental conditions, and jobs are characterized as first occupation or early occupational choice. Job exposures have little association with health, but increased physical demand reduces health. Also, first occupation with the lowest educational attainment has the worst overall health, and there are large impacts of early blue-collar employment on health. In general, higher health risk is observed as analyses move from more highly skilled occupations to less-skilled and more manual occupations. I expect these datasets to demonstrate similar results.

The remaining socio-demographic variables (gender, race, marital status, education, and so on) are expected to replicate the standard results seen in previously reported studies and in chapter 3.

Table 22 shows the characteristics of the survey population by the major occupation category. The occupations are arranged by occupation prestige ranking with higher prestige rankings to the left and lower rankings to the right. On average, the rate of mortality is higher for lower prestige occupations with 16.Transportation and 6.Private Household workers having the highest mortality. Tumor-related deaths are high for 4.Sales and 5.Administrative support workers. Cardiovascular-related deaths are most highly associated with 9.Farmers and 10.Agricultural employees. Not surprising, perhaps, is that injury-related deaths occur most often in 12.Construction and 13.Extractive (e.g., mining) occupations. The average age for this survey population is about the same across the occupations, with 9.Farmers and 16.Private Household workers having the highest average age, and 13.Extractive workers and 3.Technicians having the lowest.

Blacks are more heavily involved in 8.Service jobs, 6.Private Household jobs, and 17.Manual labor jobs. Whites are more heavily involved in 9.Farming, 4.Sales, and 1.Executive positions. Males dominate the 12.Construction, 11.Mechanical labor, and 13.Extractive occupations while women dominate the 5.Administrative support and 6.Private Household positions. 9.Farming is obviously more rural, but so is 13.Extractive services since strip mining or deep mining generally occur in more rural areas. Those with the most education are the 2.Professionals, which includes teachers; those with the least are 6.Private Household workers. Income follows a similar pattern. 9.Farmers, 14.Precision production workers, and 11.Mechanics are married more often, while 6.Private Household workers are not. Veterans go into 7.Protective Service and 11.Mechanic positions most frequently.

Table 22: Variable Means by Occupation and for ALL Occupations

	ALL	1. Exec	2. Prof	3. Tech	7. Prot Serv	9. Farm	4. Sale	14. Prod	12. Const	5. Admin	11. Mech	13. Extr	8. Serv	16. Trans	15. Mach	10. Agri	6. Priv	17. Labor
Death,%	.054	.053	.033	.038	.065	.079	.053	.069	.068	.041	.064	.055	.066	.080	.059	.068	.088	.075
Tumor,%	.353	.366	.379	.328	.291	.328	.388	.339	.336	.408	.340	.172	.339	.316	.341	.292	.360	.311
Cardio,%	.304	.306	.287	.283	.349	.389	.287	.328	.302	.255	.329	.293	.300	.337	.317	.351	.300	.306
Infect,%	.024	.031	.037	.038	.011	.001	.029	.012	.018	.027	.001	0.00	.031	.017	.015	.020	.001	.021
Injury,%	.102	.079	.100	.138	.110	.106	.092	.102	.154	.077	.120	.310	.085	.128	.103	.144	.057	.143
Other%	.217	.218	.197	.215	.239	.174	.204	.219	.191	.233	.203	.224	.249	.203	.225	.193	.280	.219
Age, yrs	40.3	41.4	39.4	37.4	39.6	45.4	40.9	41.6	39.1	39.9	39.7	37.0	40.9	40.3	40.3	40.5	45.1	38.9
Black,%	.086	.042	.056	.077	.115	.006	.040	.066	.061	.088	.060	.040	.173	.119	.134	.099	.341	.170
White,%	.893	.939	.924	.895	.865	.981	.945	.914	.917	.892	.922	.940	.800	.866	.846	.865	.640	.807
Male,%	.554	.688	.501	.531	.891	.868	.526	.827	.980	.181	.971	.975	.272	.918	.553	.699	.014	.751
Rural,%	.331	.273	.273	.280	.258	.947	.300	.351	.408	.273	.388	.620	.309	.418	.370	.681	.328	.350
SMSA1,%	.259	.263	.278	.276	.313	.008	.249	.235	.214	.289	.213	.065	.308	.226	.257	.106	.321	.283
SMSA2,%	.380	.447	.411	.421	.393	.115	.415	.400	.366	.413	.395	.143	.310	.335	.341	.221	.259	.312
SMSA3,%	.361	.290	.310	.303	.295	.878	.335	.364	.420	.299	.392	.792	.382	.439	.402	.673	.420	.406
Educ, yrs	12.9	14.3	16.0	13.7	12.9	11.9	13.1	11.9	11.7	12.8	11.8	11.3	11.4	11.3	11.1	10.8	10.7	11.0
Income, $1,000s	30.9	38.2	37.2	33.8	31.1	22.2	32.4	32.6	27.7	32.2	30.8	30.1	21.6	28.2	26.6	18.7	14.8	23.3
Married,%	.738	.779	.732	.702	.765	.870	.766	.814	.785	.704	.820	.829	.636	.786	.720	.749	.544	.671
HH Size	3.28	3.16	3.11	3.12	3.33	3.49	3.22	3.42	3.44	3.16	3.46	3.61	3.38	3.49	3.45	3.66	3.39	3.46
Veteran, %	.249	.341	.202	.258	.506	.327	.240	.405	.424	.095	.481	.397	.108	.409	.235	.200	.004	.280
Insurance ,%	.879	.925	.938	.946	.919	.720	.873	.915	.768	.932	.888	.893	.782	.850	.888	.623	.587	.804
Own House,%	.733	.788	.753	.712	.732	.871	.770	.778	.726	.740	.766	.707	.624	.719	.697	.620	.614	.641
N	367578	45119	51691	10609	5574	6849	36775	15179	15621	59374	15369	1048	33543	16993	31711	6518	2987	12538

4.6 Results and Discussion

4.6.1 Analysis of Occupation Categories

Table 23 shows the estimates for the baseline NLMS analysis and for the analyses obtained by adding the occupation categories. Table 23 does not show the occupation category Occupation (with 807 defined occupations); the Stata analysis with this category did not complete after running for eight days. The baseline variable results for the analyses with the occupation categories for males (and also MajorOcc which includes males and females) are qualitatively similar to the results in the baseline analysis. The significance of the male results, the signs of the results, and the magnitudes of the results are all similar. For females, the results demonstrate the same sign and magnitude, but are occasionally not significant. Although the number of women in the sample is less than the number of men (~158,000 to ~197,000), the number of observations is still large enough to provide adequate results. Also, reviewing simple tabular distributions of income and education by gender does not reveal any combinations with a very small number of observations. Others have suggested that there is an occupational disadvantage to women in many of these studies; that occupational status indicators are probably more accurate for white men, rather than for women or non-whites, since the indicators were largely developed on the basis of a white male labor force; and that women are poorly reflected in measures of occupational position or prestige-based indicators (Gregorio, Walsh, and Paturzo 1997). If these suggestions are accurate, the female-only columns may be affected by such bias.

Table 23: Regression Results with Various Occupation Categories

Hazard Ratios	Baseline	BRG Female	BRG Male	Major Occ	Recode Female	Recode Male
OUTCOME	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause
VARIABLES/base						
Race:Black/white	1.330***	1.387***	1.279***	1.297***	1.358***	1.242***
Race:Other/white	0.896*	0.899	0.889*	0.890**	0.896	0.881*
Female/male	0.431***			0.419***		
Rural/urban	0.899***	0.882***	0.898***	0.931***	0.894***	0.953**
Married/other than married	0.705***	0.788***	0.652***	0.711***	0.795***	0.661***
Educ: None/high school-H4	0.758**	0.658	0.742**	0.751**	0.678	0.746**
Educ: E1-E4/high school-H4	1.011	0.949	1.002	0.997	0.945	1.000
Educ: E5-E6/high school-H4	1.103**	1.027	1.099*	1.080*	1.034	1.081
Educ: E7-E8/high school-H4	1.138***	1.052	1.144***	1.129***	1.043	1.147***
Educ: H1/high school-H4	1.130***	1.156**	1.107**	1.114***	1.151**	1.097**
Educ: H2/high school-H4	1.160***	1.150**	1.152***	1.143***	1.142**	1.137***
Educ: H3/high school-H4	1.173***	1.087	1.203***	1.160***	1.081	1.189***
Educ: H4	------	------	------	------	------	------
Educ: C1/high school-H4	0.926**	0.910	0.943	0.932**	0.908	0.947
Educ: C2/high school-H4	0.967	1.007	0.961	0.978	1.009	0.967
Educ: C3/high school-H4	0.935	0.861	0.976	0.957	0.879	0.987
Educ: C4/high school-H4	0.738***	0.788***	0.740***	0.764***	0.828***	0.751***
Educ: C5/high school-H4	0.651***	0.709***	0.652***	0.687***	0.781**	0.669***
Educ: C6/high school-H4	0.640***	0.776***	0.640***	0.681***	0.843*	0.664***
Income:0-5K/25K-30K	1.368***	1.379***	1.359***	1.396***	1.360***	1.436***
Income:5K-7.5K/25K-30K	1.445***	1.416***	1.477***	1.461***	1.396***	1.540***
Income:7.5K-10K /25K-30K	1.169***	1.194**	1.166**	1.177***	1.183**	1.207***
Income:10K-12.5K /25K-30K	1.256***	1.130*	1.365***	1.257***	1.131*	1.390***
Income:12.5K-15K /25K-30K	1.120***	1.113	1.139***	1.123***	1.111	1.158***
Income:15K-20K /25K-30K	1.084**	1.081	1.096**	1.086**	1.077	1.106**
Income20K-25K/25K-30K	1.059*	1.085	1.049	1.057*	1.085	1.049
Income25K-30K	------	------	------	------	------	------
Income:30K-35K/25K-30K	0.891***	0.978	0.857***	0.889***	0.978	0.857***
Income:35K-40K/25K-30K	0.869***	0.850***	0.874***	0.872***	0.852***	0.875***
Income:40K-50K/25K-30K	0.767***	0.869**	0.734***	0.769***	0.871**	0.736***
Income:50K-60K/25K-30K	0.873***	0.916	0.864***	0.880***	0.915	0.869***
Income:60K-75K/25K-30K	0.758***	0.756***	0.760***	0.763***	0.756***	0.766***
Income:75K+/25K-30K	0.774***	0.843**	0.759***	0.782***	0.831***	0.771***
1. Professional – female		------				

Hazard Ratios	Baseline	BRG Female	BRG Male	Major Occ	Recode Female	Recode Male
OUTCOME	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause
2. Clerical/professional – female		0.966				
3. Skilled Crafts/professional – female		0.864				
4. Labor/professional – female		1.055				
1. Professional – male			------			
2. Clerical/professional – male			1.085***			
3. Skilled Crafts/professional – male			1.037			
4. Labor/professional – male			1.069**			
1.Executive/professional				1.062*		
2.Professional				------		
3.Technician/professional				1.107*		
4.Sales/professional				1.089**		
5.Clerical/professional				1.129***		
6.Private Household/professional				1.169**		
7.Protective Services/professional				1.132**		
8.Service-not protective/professional				1.238***		
9.Farm Managers/professional				0.777***		
10.Farm Workers/professional				0.981		
11.Mechanics/professional				1.048		
12.Construction/professional				1.072		
13.Extractive/professional				1.103		
14.Precision Production/professional				1.071		
15.Machine Operators/professional				1.079*		
16.Transportation/professional				1.183***		
17.Handlers,Laborers/professional				1.217***		
1.Accountants/teachers					1.026	
2.Computer Specialists/teachers					1.322	
3.Librarians/teachers					0.907	
4.Mathematicians/teachers					2.809**	
5.Life,Physical Scientists/teachers					0.741	
6.Nurses,Therapists/teachers					1.123	
7.Health Technicians/teachers					1.371**	
8.Social Scientists/teachers					1.040	
9.Social Workers/teachers					1.129	
10.Teachers					------	
11.Technicians/teachers					0.956	
12.Writers,Entertainers/teachers					1.388**	

Hazard Ratios	Baseline	BRG Female	BRG Male	Major Occ	Recode Female	Recode Male
OUTCOME	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause
13.Other Professional/teachers					1.482***	
14.Buyers,Sales Managers/teachers					1.300**	
15.Restaurant Managers/teachers					1.232	
16.School Administrators/teachers					1.188	
17.Other Managers/teachers					1.236**	
18.Peddlers/teachers					1.942	
19.Insurance brokers/teachers					1.322*	
20.Sales Clerks/teachers					1.126	
21.Salesmen/teachers					1.050	
22.Other Sales Workers/teachers					1.242	
23.Bank Tellers/teachers					0.790	
24.Bookkeepers/teachers					1.167	
25.Cashiers/teachers					1.196*	
26.Counter Clerks/teachers					1.134	
27.Interviewers/teachers					1.059	
28.File Clerks/teachers					1.077	
29.Office Machine Operators/teachers					1.082	
30.Payroll Clerks/teachers					1.641	
31.Receptionists/teachers					1.090	
32.Secretaries/teachers					1.107	
33.Stenographers/teachers					1.642	
34.Telephone Operators/teachers					1.577***	
35.Typists/teachers					1.204	
36.Other Clerical/teachers					1.214**	
37.Foremen/teachers					1.015	
38.Other Craftsmen/teachers					1.273*	
39.Assemblers/teachers						
40.Bottling Operatives/teachers					0.816	
41.Examiners,Inspectors/teachers					1.365**	
42.Seamstresses/teachers					0.968	
43.Laundry Operatives/teachers					1.353*	
44.Graders,Sorters/teachers					0.791	
45.Packers,Wrappers/teachers					1.330*	
46.Sewers,Stitchers/teachers					0.332	
47.Textile Operatives/teachers					1.009	
48.Other Operatives/teachers					1.218**	

Hazard Ratios	Baseline	BRG Female	BRG Male	Major Occ	Recode Female	Recode Male
OUTCOME	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause
49.Transport Operatives/teachers					1.161	
50.Laborers-not farm/teachers					1.360**	
51.Farmers/teachers					1.168	
52.Farm laborers/teachers					0.995	
53.Cleaning Workers/teachers					1.321***	
54.Cooks/teachers					1.348***	
55.Waitresses/teachers					1.363***	
56.Health Service Workers/teachers					1.387***	
57.Cosmetologists/teachers					0.789	
58.Other Personal Service/teachers					1.370***	
59.Private Household/teachers					1.410***	
1.Accountants/teachers						1.084
2.Architects/teachers						0.517*
3.Computer Specialists/teachers						0.924
4.Engineers/teachers						0.938
5.Lawyers,Judges/teachers						0.821
6.Chemists/teachers						0.741
7.Life,Physical Scientists/teachers						1.108
8.Dentists/teachers						1.256
9.Pharmacists/teachers						1.235
10.Physicians/teachers						0.940
11.Other Health Practitioners/teachers						0.990
12.Heath Technicians/teachers						1.887***
13.Religious Workers/teachers						0.812
14.Social Scientists/teachers						1.016
15.Social Workers/teachers						1.234
16.Teachers						------
17.Engineering Technicians/teachers						0.993
18.Writers,Entertainers/teachers						0.937
19.Other Professionals/teachers						1.124
20.Buyers,Sales Managers/teachers						1.021
21.School Administrators/teachers						1.041
22.Public Administrators/teachers						1.051
23.Other Managers/teachers						0.966
24.Managers,Administrators/teachers						1.019
25.Insurance Brokers/teachers						0.888

Hazard Ratios	Baseline	BRG Female	BRG Male	Major Occ	Recode Female	Recode Male
OUTCOME	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause
26.Real Estate Brokers/teachers						0.861
27.Other Sales Workers/teachers						1.171*
28.Salesmen/teachers						0.982
29. Bank Tellers,Cashiers/teachers						1.038
30.Bookkeepers/teachers						1.446***
31.Postal Clerks/teachers						1.009
32.Other Clerical/teachers						1.242***
33.Upholsterers/teachers						1.102
34.Bakers/teachers						0.675
35.Cabinetmakers/teachers						0.807
36.Carpenters/teachers						0.946
37.Road Machine Operatives/teachers						0.927
38.Electricians/teachers						0.983
39.Masons/teachers						1.155
40.Painters,Paperhangers/teachers						1.158
41.Plasterers/teachers						0.915
42.Plumbers/teachers						1.086
43.Other Construction/teachers						1.098
44.Foremen/teachers						0.996
45.Linemen-Power/teachers						1.018
46.Locomotive Engineers/teachers						1.379**
47.Auto Mechanics/teachers						1.074
48.Other Mechanics/teachers						0.950
49.Machinists/teachers						1.068
50.Sheetmetal Workers/teachers						1.130
51.Tool&Die Makers/teachers						0.953
52.Other Metal Craftsmen/teachers						1.213
53.Printing Craftsmen/teachers						0.867
54.Power Station Operators/teachers						1.223
55.Other Craftsmen/teachers						1.103
56.Assemblers/teachers						0.717
57.Examiners,Inspectors/teachers						1.253**
58.Gas Station Attendants/teachers						1.259
59.Laundry Operatives/teachers						1.291
60.Butchers/teachers						1.219
61.Mine Operators/teachers						1.093

Hazard Ratios	Baseline	BRG Female	BRG Male	Major Occ	Recode Female	Recode Male
OUTCOME	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause
62.Packers,Wrappers/teachers						1.459**
63.Painters/teachers						1.066
64.Precision Machine Operatives/teachers						1.183
65.Sawyers/teachers						0.660*
66.Firemen/teachers						0.847
67.Textile Operatives/teachers						0.756
68.Welders/teachers						1.057
69.Other Metal Operatives/teachers						1.048
70.Other Specified Operatives/teachers						1.124
71.Other Operatives/teachers						0.998
72.Bus Drivers/teachers						0.982
73.Taxicab Drivers/teachers						1.357**
74.Truck Drivers/teachers						1.120
75.Other Transport Operatives/teachers						1.521**
76.Construction Laborers/teachers						1.305***
77.Freight Handlers/teachers						1.126
78.Other Specified Laborers/teachers						1.047
79.Other Laborers/teachers						1.056
80.Farmers/teachers						0.714***
81.Farm Laborers/teachers						0.939
82.Cleaning Service Workers/teachers						1.187**
83.Food Service Workers/teachers						1.422***
84.Health Service Workers						1.881***
85.Personal Service Workers/teachers						1.207
86.Protective Service Workers/teachers						1.139
87.Other Service Workers/teachers						1.237
88.Private Household Workers/teachers						0.249*
Observations	355,449	158,424	197,025	355,449	158,424	196,987

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

For each occupation category, the omitted group is the group containing teachers. This is most clearly seen in the OccupationRecode categories where there are specific groups for teachers (group 10 for females and group 16 for males). In the other categories, the aggregate group that included teachers was selected. In the BRG-Male category, both Clerical and Labor groups demonstrate significantly higher mortality risks relative to the Professional (Teachers) group—these are also the least prestigious groups. For the BRG-Female category, there are no significant results, although the Labor group has the highest mortality relative to Teachers. In the Major Occupation category (which includes both males and females), there are a number of significant results. Several groups have a higher risk (at the 10% level or better) of mortality than Professionals (Teachers), including Executives, Technicians, Sales, Clerical, Private Household workers, Protective Services (firemen and police), Services – not protective, Machine Operators, Transportation workers, and Handlers/Laborers. Most of these are a breakout of the significant categories from BRG-Male: Clerical and Laborer groups. The insignificant category from BRG-Male – Skilled Craftsmen can be tied to the few insignificant Major Occupation categories, e.g., precision production workers and mechanics. The one significant result that demonstrates improved mortality risk relative to Professionals is Farmers. This result persists even with the correction for rural living, where the vast majority of Farmers reside. The magnitude of the Rural variable is reduced from the baseline result for both the Major Occupation category and for the OccupationRecode-Male category both of which have specific (and significant) categories for Farmers.

For the OccupationRecode-Male category, there are 17 significant results (at the 10% level or better) relative to the Teacher base group, out of 88 total groups. The significant results that have a negative association with health include a rather eclectic set of occupation categories, the groups for Health Technicians, Bookkeepers, Other Clerical, Locomotive Engineers, Inspectors, Packers, Taxicab Drivers, Other Transportation Operators, Construction Laborers, Cleaning Service workers, Food Service workers, and Health Service workers. Several, like Locomotive Engineers, Packers, Other Transportation workers, and Construction Laborers face hazardous occupational working conditions. Other groups on the list are some of the lowest prestige occupations on the list. Those groups with positive health relative to teachers include Architects, Farmers, and Sawyers. The latter group has a hazardous job (cutting wood), and could be a statistical anomaly. Since income, education, gender, geography, and race are controlled for, these results likely reflect other factors not specifically included in the analysis, namely job risk, job hazard, working conditions, and/or job prestige. The groups with higher mortality risk have a much lower average prestige ranking than those groups with a lower risk. Using the 1989 National Opinion Research Center (NORC) Prestige Scores, which range from 20 (low prestige) to 77 (high prestige) for the 88 OccupationRecode-male groups, the average prestige score for the higher risk groups is 33.9 and the average risk for the lower risk groups is 52.1. There is a significant reduced risk of mortality when regressing occupation prestige rankings on mortality (see below). It is consistent that the higher risk groups have a lower average prestige score.

For the OccupationRecode-Female category, there are 22 significant results (at the 10% level or better) relative to the Teacher base group out of 58 total groups. All of these significant results negatively impact health and include Health Technicians, Telephone Operators, Cleaning workers, Cooks, Waitresses, Health Service workers, and Household workers. There are a few groups that demonstrate a lower risk of mortality but none have significant coefficients. These non-significant groups include: Bank Tellers, Librarians, Seamstresses, Sewers, and Cosmetologists. A comparison of the average prestige scores for females is much less interesting. The higher-risk groups have an average score of 38.9 and the lower-risk (non-significant) groups have an average score of 37.1. Perhaps this is an indication of occupational disadvantage exhibited in this data set.

4.6.2 Visual Analysis of Occupation Categories

Table 24 shows the results of analyzing the OccupationRecode by gender categories using the Cox Proportional Hazard (CPH) approach. The results shown are only for the occupation groups (the other data do not change qualitatively from the previous analyses). The two analyses for each gender are (1) an "unadjusted" analysis with only Age and Race as explanatory variables and (2) an "adjusted" analysis using Education and Income covariates as well as Age and Race.

Table 24: Cox Proportional Hazard Analysis of Occupation Recode Categories

NOTE: other results not shown

Hazard Ratios	Unadj.	Adjusted		Unadj.	Adjusted
	Cox PH male	Cox PH male		Cox PH female	Cox PH female
VARIABLES/base			Variable/base		
1.Accountants/teachers	1.211*	1.095	1.Accountants/ teachers	1.191	1.073
2.Architects/teachers	0.523*	0.545*	2.Computer Specialists/ teachers	1.450	1.427
3.Computer Specialists/teachers	1.065	0.920	3.Librarians/teachers	0.967	0.934
4.Engineers/teachers	1.024	0.933	4.Mathematicians/ teachers	3.078**	3.189**
5.Lawyers,Judges/ teachers	0.778*	0.825	5.Life,Physical Scientists/teachers	0.782	0.791
6.Chemists/teachers	0.838	0.753	6.Nurses,Therapists/ teachers	1.268**	1.162
7.Life,Physical Scientists/teachers	1.137	1.091	7.Health Technicians/teachers	1.712***	1.416***
8.Dentists/teachers	1.070	1.223	8.Social Scientists/teachers	1.112	1.104
9.Pharmacists/teachers	1.290	1.204	9.Social Workers/teachers	1.330	1.180
10.Physicians/teachers	0.828	0.923	10.Teachers	1.0	1.0
11.Other Health Practitioners/teachers	1.318	1.044	11.Technicians/teachers	1.131	1.014
12.Heath Technicians/teachers	2.503***	1.928***	12.Writers, Entertainers/ teachers	1.630***	1.430**
13.Religious Workers/teachers	0.966	0.783*	13.Other Professional/teachers	1.669***	1.553***
14.Social Scientists/teachers	1.028	1.021	14.Buyers,Sales Managers/teachers	1.560***	1.311**
15.Social Workers/teachers	1.416*	1.241	15.Restaurant Managers/teachers	1.604	1.269
16.Teachers	1.0	1.0	16.School Administrators/teachers	1.207	1.235
17.Engineering Technicians/teachers	1.285***	0.999	17.Other Managers/teachers	1.463***	1.291***
18.Writers, Entertainers/ teachers	1.282**	0.974	18.Peddlers/teachers	2.518	2.082
19.Other Professionals/teachers	1.345***	1.125	19.Insurance brokers/teachers	1.467***	1.377**
20.Buyers,Sales Managers/teachers	1.393***	1.003	20.Sales Clerks/teachers	1.427**	1.157
21.School Administrators/teachers	0.944	1.011	21.Salesmen/teachers	1.323*	1.065
22.Public Administrators/teachers	1.257*	1.038	22.Other Sales Workers/teachers	1.536***	1.267*
23.Other Managers/teachers	1.154	0.951	23.Bank Tellers/teachers	1.019	0.839
24.Managers, Administrators/ teachers	1.288***	1.011	24.Bookkeepers/teachers	1.427***	1.190*
25.Insurance Brokers/teachers	1.093	0.866	25.Cashiers/teachers	1.562***	1.219*
26.Real Estate Brokers/teachers	1.097	0.856	26.Counter Clerks/teachers	1.455*	1.165

Hazard Ratios	Unadj.	Adjusted		Unadj.	Adjusted
	Cox PH male	Cox PH male		Cox PH female	Cox PH female
27.Other Sales Workers/teachers	1.538***	1.169*	27.Interviewers/teachers	1.370**	1.119
28.Salesmen/teachers	1.507**	0.986	28.File Clerks/teachers	1.378*	1.125
29. Bank Tellers, Cashiers/ teachers	1.696***	1.044	29.Office Machine Operators/teachers	1.389*	1.158
30.Bookkeepers/teachers	2.131***	1.504***	30.Payroll Clerks/teachers	1.894	1.763
31.Postal Clerks/teachers	1.452***	1.033	31.Receptionists/teachers	1.343**	1.125
32.Other Clerical/teachers	1.831***	1.274***	32.Secretaries/teachers	1.345***	1.158
33.Upholsterers/teachers	1.812***	1.102	33.Stenographers/teachers	2.217***	1.777*
34.Bakers/teachers	1.091	0.688	34.Telephone Operators/teachers	2.099***	1.697***
35.Cabinetmakers/ teachers	1.353	0.799	35.Typists/teachers	1.526***	1.272*
36.Carpenters/teachers	1.563***	0.925	36.Other Clerical/teachers	1.502***	1.265**
37.Road Machine Operatives/teachers	1.597***	0.916	37.Foremen/teachers	1.254	1.071
38.Electricians/teachers	1.421***	0.979	38.Other Craftsmen/teachers	1.626***	1.303*
39.Masons/teachers	1.879***	1.133	39.Assemblers/teachers	1.067	0.838
40.Painters,Paperhangers/teacher s	2.029***	1.164	40.Bottling Operatives/teachers	1.823***	1.404**
41.Plasterers/teachers	1.441	0.874	41.Examiners, Inspectors/ teachers	1.282	0.971
42.Plumbers/teachers	1.641***	1.073	42.Seamstresses/teachers	2.011***	1.387**
43.Other Construction/teachers	1.857***	1.099	43.Laundry Operatives/teachers	1.096	0.801
44.Foremen/teachers	1.430**	0.989	44.Graders, Sorters/ teachers	1.829***	1.362**
45.Linemen-Power/teachers	1.381**	1.009	45.Packers, Wrappers/ teachers	1.273	0.323
46.Locomotive Engineers/teachers	1.892***	1.382**	46.Sewers, Stitchers/ teachers	1.415***	1.007
47.Auto Mechanics/teachers	1.659***	1.055	47.Textile Operatives/teachers	1.683***	1.253**
48.Other Mechanics/teachers	1.392***	0.935	48.Other Operatives/teachers	1.526**	1.151
49.Machinists/teachers	1.547***	1.075	49.Transport Operatives/teachers	1.832***	1.374**
50.Sheetmetal Workers/teachers	1.770***	1.128	50.Laborers-not farm/teachers	1.362*	1.054
51.Tool&Die Makers/teachers	1.359*	0.964	51.Farmers/teachers	1.230	0.896
52.Other Metal Craftsmen/teachers	1.821***	1.213	52.Farm laborers/teachers	1.903***	1.338***
53.Printing Craftsmen/teachers	1.233	0.886	53.Cleaning Workers/teachers	1.886***	1.343***
54.Power Station Operators/teachers	1.686***	1.208	54.Cooks/teachers	1.928***	1.417***
55.Other Craftsmen/teachers	1.556***	1.094	55.Waitresses/teachers	1.898***	1.419***
56.Assemblers/teachers	0.894	0.679	56.Health Service Workers/teachers	1.113	0.796

Hazard Ratios	Unadj.	Adjusted		Unadj.	Adjusted
	Cox PH male	Cox PH male		Cox PH female	Cox PH female
57.Examiners,Inspectors/ teachers	1.831***	1.265**	57.Cosmetologists/ teachers	1.863***	1.388***
58.Gas Station Attendants/teachers	2.200***	1.245	58.Other Personal Service/teachers	2.154***	1.415***
59.Laundry Operatives/teachers	2.299***	1.357	59.Private Household/teachers		
60.Butchers/teachers	1.971***	1.209			
61.Mine Operators/teachers	1.649***	1.066			
62.Packers,Wrappers/ teachers	2.510***	1.485**			
63.Painters/teachers	1.736***	1.072			
64.Precision Machine Operatives/teachers	1.861***	1.214			
65.Sawyers/teachers	1.142	0.642*			
66.Firemen/teachers	1.124	0.841			
67.Textile Operatives/teachers	1.301	0.765			
68.Welders/teachers	1.595***	1.043			
69.Other Metal Operatives/teachers	1.604***	1.043			
70.Other Specified Operatives/teachers	1.804***	1.137			
71.Other Operatives/teachers	1.599***	1.008			
72.Bus Drivers/teachers	1.499***	0.971			
73.Taxicab Drivers/teachers	2.396***	1.410**			
74.Truck Drivers/teachers	1.821***	1.114			
75.Other Transport Operatives/teachers	2.390***	1.554**			
76.Construction Laborers/teachers	2.165***	1.304***			
77.Freight Handlers/teachers	1.927***	1.157			
78.Other Specified Laborers/teachers	1.734***	1.069			
79.Other Laborers/teachers	1.850***	1.084			
80.Farmers/teachers	1.190**	0.660***			
81.Farm Laborers/teachers	1.735***	0.933			
82.Cleaning Service Workers/teachers	2.037***	1.214**			
83.Food Service Workers/teachers	2.687***	1.519***			
84.Health Service Workers	3.025***	1.951***			
85.Personal Service Workers/teachers	1.989***	1.273*			
86.Protective Service	1.730***	1.146			

Hazard Ratios	Unadj. Cox PH male	Adjusted Cox PH male		Unadj. Cox PH female	Adjusted Cox PH female
Workers/teachers					
87.Other Service Workers/teachers	2.096***	1.250			
88.Private Household Workers/teachers	0.518	0.286*			
Observations	202,817	197,171		162,818	158,565

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

Figures 6 and 7 show plots of the relative risks from the CPH model for men and women, respectively, updating similar plots shown by (Johnson, Sorlie, and Backlund 1999). Each graph is arranged by the BRG groups, and each group is ordered by the age- and race-adjusted relative risk. These views provide a visual assessment of socioeconomic status and the association with income and education. For both men and women the base category is Teachers, which has a relatively low risk. For men, occupations with a low risk relative to teachers include Architects, Lawyers, Physicians, and Chemists. The number of male Private Household workers is too small to provide accurate, measureable results. Occupations with a high risk relative to teachers include Health Technicians, Bookkeepers, Painters, Food Service workers, and Health Service workers. Figure 6 clearly shows that the impact of education and income increases from the Professional BRG category down to the Laborer BRG category as the differences between the curves increases across the categories.

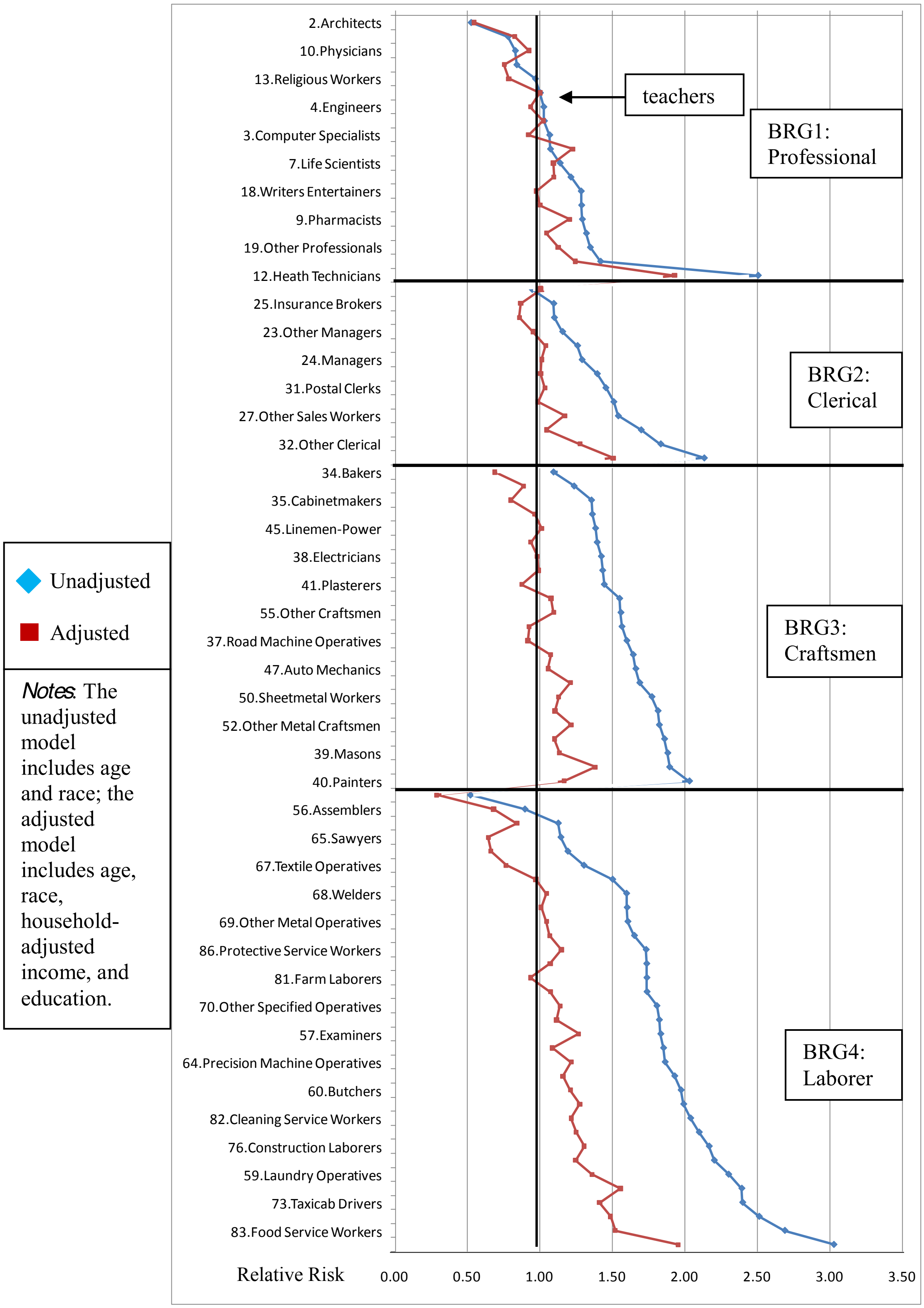

Figure 6: Adjusted and Unadjusted Relative Risks of Mortality among Males Aged 25–65 within Specific Occupations

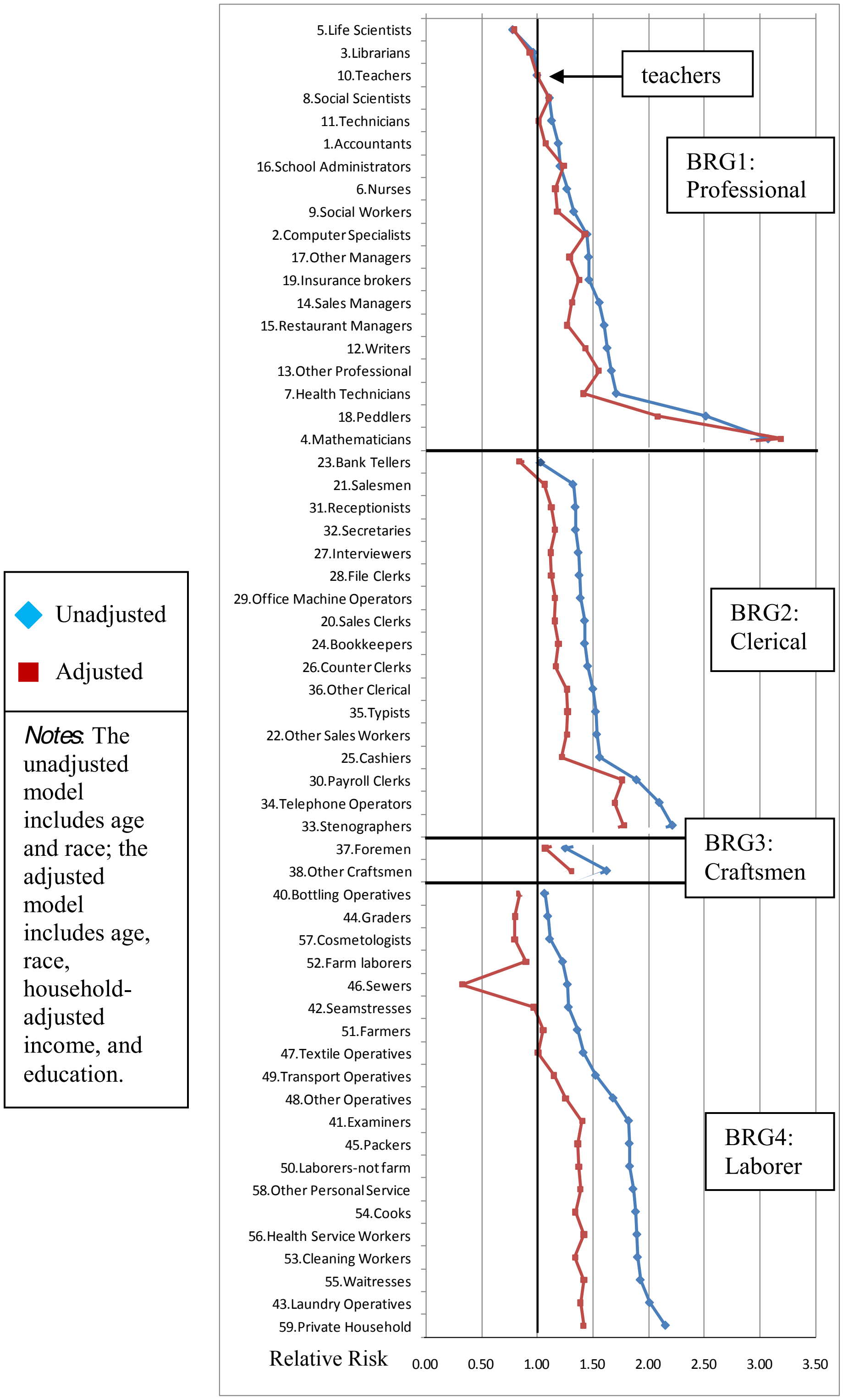

Figure 7: Adjusted and Unadjusted Relative Risks of Mortality among Females Aged 25–65 within Specific Occupations

Notes: The first point is adjusted for age and race; the second for age, race, and household-adjusted income; the third for age, race, and education; and the fourth for age, race, household-adjusted income, and education.

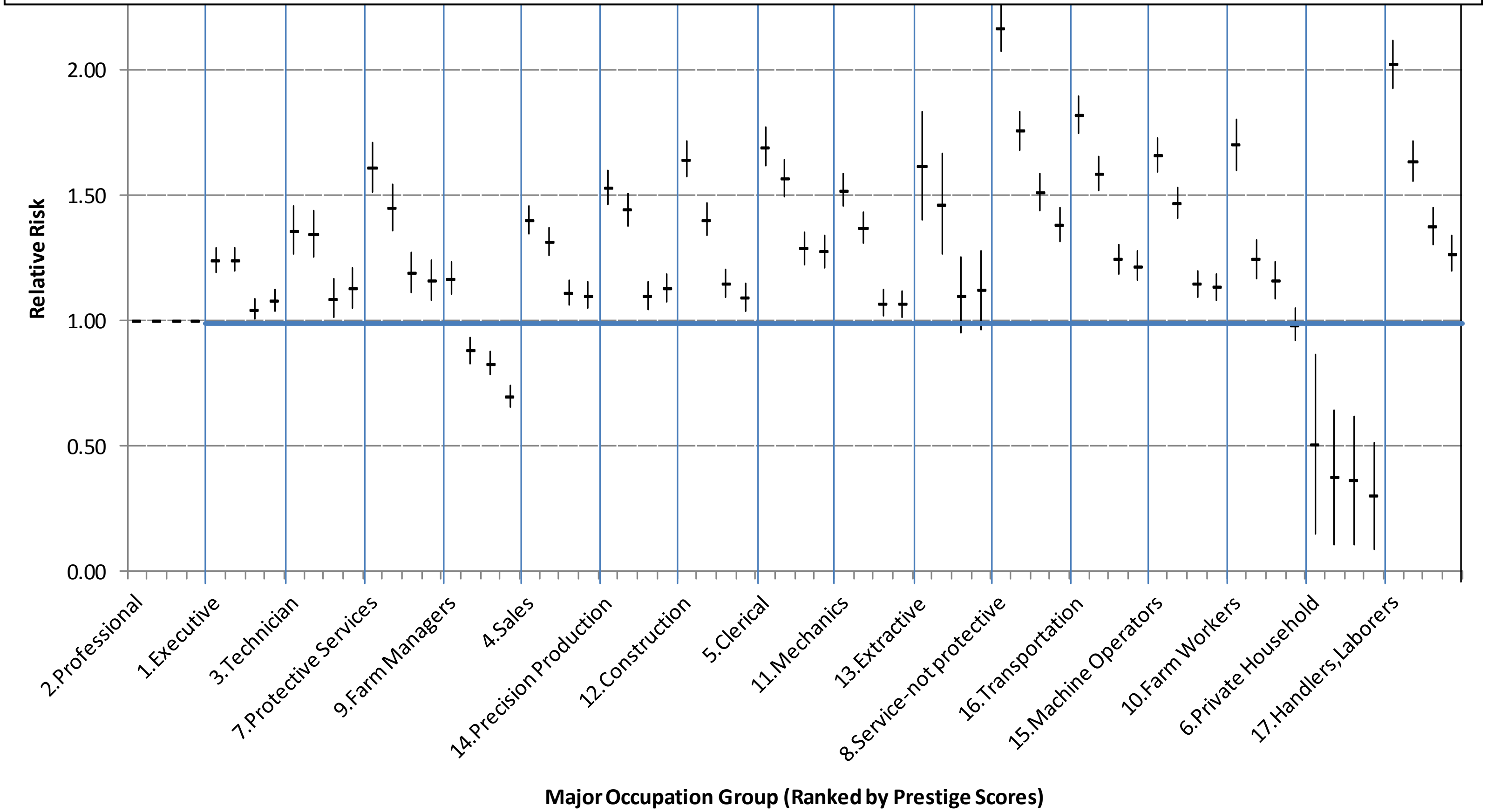

Figure 8: Relative Risks of Mortality among Males Aged 25–65 within Major Occupations, adjusted for Age, Race, Income, and Education

Notes: The first point is adjusted for age and race; the second for age, race, and household-adjusted income; the third for age, race, and education; and the fourth for age, race, household-adjusted income, and education.

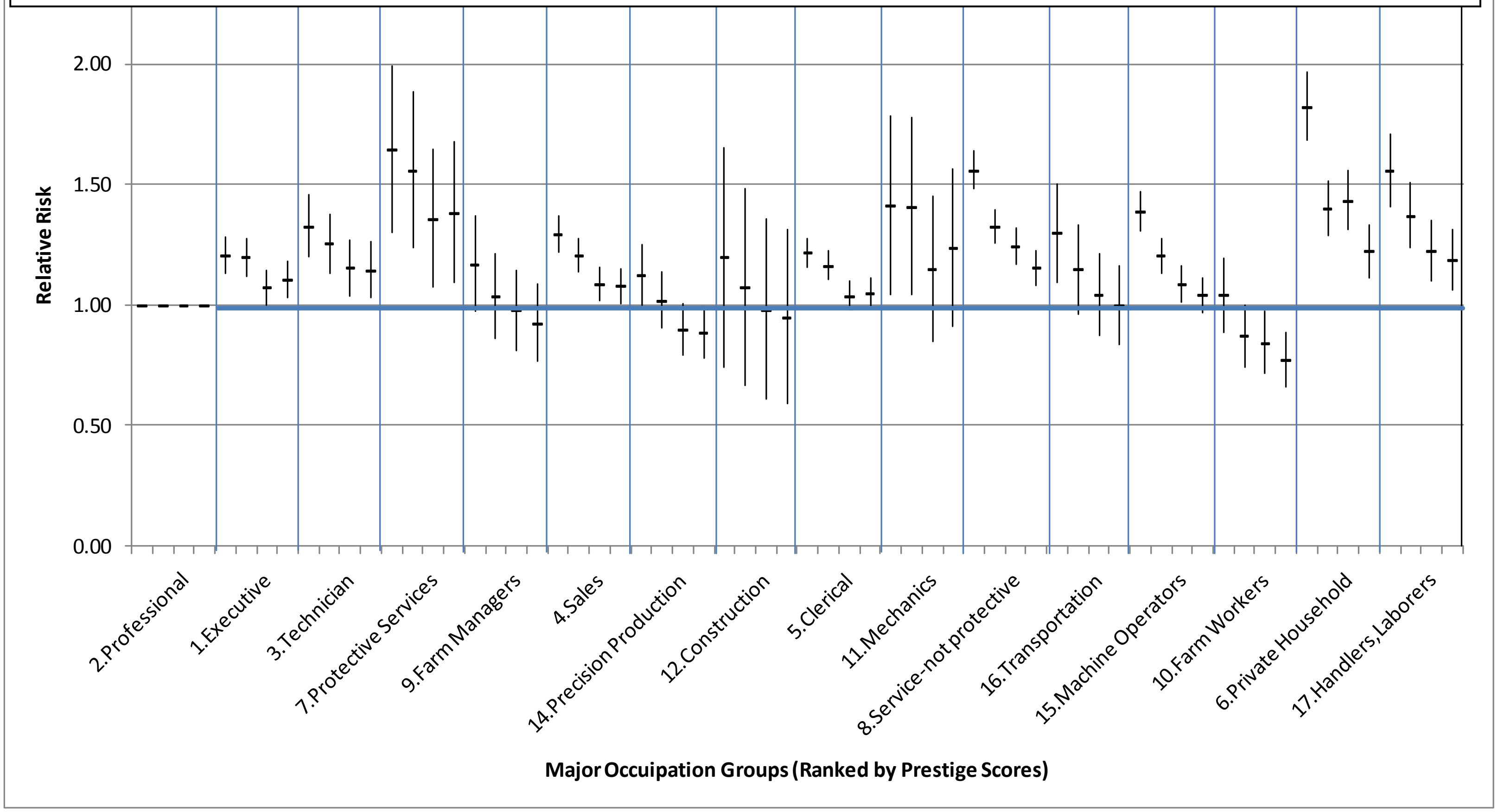

Figure 9: Relative Risks of Mortality among Females Aged 25–65 within Major Occupations, adjusted for Age, Race, Income, and Education

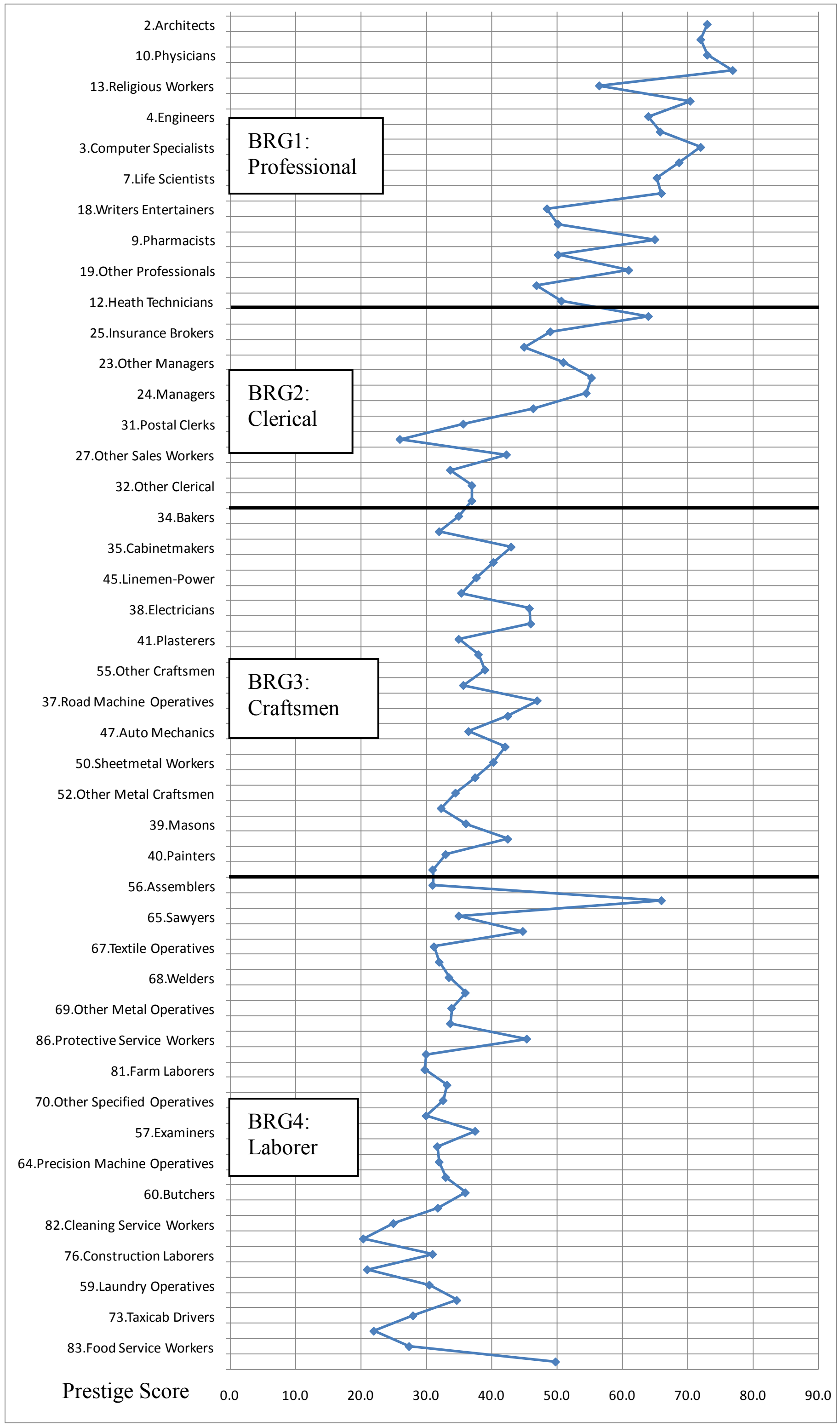

Figure 10: Prestige Scores among Males Aged 25–65 within Specific Occupations

For women, occupations with low risk relative to teachers include Life Scientists, Bank Tellers, and Cosmetologists. Occupations with high risk relative to teachers include Mathematicians, Stenographers, Laundry workers, Waitresses, and Private Household workers. The impact of education and income increases for women, but not nearly as dramatically as for men. In general, even when controlling for income and education, results for women are similar. As pointed out by (Johnson, Sorlie, and Backlund 1999), there are clear differences within the BRG groups although the differences between groups are small. This suggests that the primary responsibility for the differential risks are the specific occupational impacts (e.g., exposure to environment, behaviors, stresses, and level of responsibility of specific occupations), rather than social classes.

Table 25 shows the results of analyzing the Major Occupation category using the Cox Proportional Hazard (CPH) approach. The results shown are only for the occupation groups (the other data do not change qualitatively from the previous analyses). The four analysis models for each gender are adjusted (1) for age and race; (2) for age, race, and household-adjusted income; (3) for age, race, and education; and (4) for age, race, household-adjusted income, and education.

Table 25: Cox Proportional Hazard Analyses of Major Occupation Categories

NOTE: remaining results not shown

Hazard Ratios	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
	Male	Male	Male	Male	Female	Female	Female	Female
VARIABLES/base								
1.Executive/ professional	1.243***	1.246***	1.048	1.083*	1.206***	1.198***	1.074	1.107
2.Professional	------	------	------	------	------	------	------	------
3.Technician/ professional	1.363***	1.348***	1.090	1.132*	1.328***	1.256**	1.154	1.146
4.Sales/professional	1.403***	1.319***	1.115**	1.104**	1.293***	1.208***	1.085	1.078
5.Clerical/professional	1.695***	1.571***	1.291***	1.279***	1.218***	1.165***	1.039	1.049

Hazard Ratios	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
	Male	Male	Male	Male	Female	Female	Female	Female
6.Private Household/ professional	0.509	0.376	0.364	0.302*	1.824***	1.401***	1.433***	1.222**
7.Protective Services/professional	1.612***	1.455***	1.193***	1.163**	1.646**	1.561**	1.359	1.384
8.Service-not protective/professional	2.168***	1.759***	1.517***	1.386***	1.561***	1.325***	1.243***	1.154**
9.Farm Managers/professional	1.171***	0.883**	0.832***	0.698***	1.171	1.035	0.978	0.925
10.Farm Workers/professional	1.705***	1.249***	1.163**	0.987	1.040	0.871	0.845	0.773*
11.Mechanics/ professional	1.524***	1.375***	1.072	1.068	1.412	1.410	1.151	1.236
12.Construction/ professional	1.645***	1.406***	1.149***	1.094*	1.197	1.075	0.981	0.951
13.Extractive/ professional	1.620***	1.468***	1.102	1.125				
14.Precision Production/ professional	1.536***	1.445***	1.101*	1.132**	1.126	1.020	0.901	0.887
15.Machine Operators/ professional	1.663***	1.470***	1.148***	1.136***	1.388***	1.205***	1.087	1.041
16.Transportation/ professional	1.825***	1.589***	1.247***	1.220***	1.298*	1.147	1.043	0.999
17.Handlers,Laborers/ professional	2.025***	1.639***	1.378***	1.271***	1.556***	1.372***	1.226**	1.186
Observations	202,855	197,268	202,796	197,209	162,818	158,588	162,795	158,565

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

Figures 8 and 9 show plots of the relative risks from the CPH model for men and women, respectively, updating similar plots shown by (Johnson, Sorlie, and Backlund 1999). The Major Occupation categories are arranged in decreasing occupation prestige order using NORC 1989 prestige scores. For both men and women, the reference group is Professional (which includes Teachers).

For men, there is a tendency for increased risk of mortality with decreasing prestige score. The Farmers group, the only major occupation with a relative risk consistently below Teachers, does not follow the prestige trend. For most of the groups, the impact of Education (the third point) is larger than the impact of Income alone (the second point). Farmers and Laborers are the only groups that do not demonstrate the Education impact—both occupations in which formal education may be less important.

For women this arrangement does not demonstrate an increasing pattern of risk with prestige as would be expected if mortality risks are inversely correlated with prestige. Farm workers are the only group that has relative risks at or below that of Teachers. The impact of Education is generally larger than that of Income. Overall, however, the results of adding Education and Income are less than with men as evidenced in Figures 6 and 7 as well.

Finally, Figure 10 shows the NORC 1989 prestige scores for men and the Occupation Recode and BRG categories. Figure 10 shows a clear inverse relationship within the Professional and Clerical BRG groups, somewhat mirroring the slight trend within Figure 8 that indicated an inverse relationship between prestige and relative risk. In the Craftsmen and Laborer groups, however, there is no clear trend; the relationship is virtually non-existent. The extreme points within categories are Religious workers (low), Entertainers (low), and Social Workers (low) in the Professional group; Real Estate brokers (low) and Salesmen (low) in the Clerical group; and Firemen (high), Protective Service workers (Police) (high), Farmers (high), and Health Service workers (high) in the Laborer group.

4.6.3 Impact of Age Groups

Table 26 shows the impact of using Age variables different than the baseline age dummies. The Income and Education baseline variables do not change dramatically in magnitude or significance and are not shown. The first column contains the results using continuous Age and Age2 variables. The second column has two categories, using dummy variables with ages from 25–44 and ages from 45–65. Each value shows the results of that age range relative to all other ages. Table 26 shows that respondents with ages 25–44 die less often than all other age groups, while older respondents die more often.

Table 26: Age Impacts

Hazard Ratios	Age & Age^2	2 Age Groups	4 Age Groups	Multi-valued Age
OUTCOME	All-Cause	All-Cause	All-Cause	All-Cause
VARIABLES/base				
Age, mean-centered	1.095***			
Age^2	1.000			
Age 25_44		0.397***		
Age 45_65		2.257***		
Age 25_34			0.264***	
Age 35_44			0.582***	
Age 45_54			1.493***	
Age 55_65			3.338***	
Ages 35-44/Ages 25-34				2.376***
Ages 45-54/Ages 25-34				6.322***
Ages 55-65/Ages 25-34				15.50***
Race:Black/white	1.295***	1.207***	1.242***	1.278***
Race:Other/white	0.889**	0.804***	0.826***	0.878**
Female/male	0.418***	0.441***	0.443***	0.419***
Rural/urban	0.929***	0.854***	0.869***	0.920***
Married/other than married	0.709***	1.261***	1.242***	0.720***
1.Executive/ professional	1.061*	1.119***	1.116***	1.059
2.Professional	------	------	------	------
3.Technician/ professional	1.106*	0.863***	0.896**	1.085
4.Sales/ professional	1.087**	0.998	0.993	1.101**
5.Clerical/ professional	1.129***	1.014	1.019	1.134***
6.Private Household/ professional	1.166**	1.157***	1.121**	1.190**
7.Protective Services/ professional	1.130**	1.155***	1.162***	1.105
8.Service-not protective/ professional	1.238***	0.911***	0.915***	1.256***
9.Farm Managers/ professional	0.776***	1.179***	1.118**	0.811***
10.Farm Workers/ professional	0.981	0.663***	0.679***	0.972

Hazard Ratios	Age & Age^2	2 Age Groups	4 Age Groups	Multi-valued Age
OUTCOME	All-Cause	All-Cause	All-Cause	All-Cause
11.Mechanics/ professional	1.048	0.811***	0.841***	1.028
12.Construction/ professional	1.073	0.798***	0.829***	1.061
13.Extractive/ professional	1.106	0.648***	0.688***	1.068
14.Precision Production/ professional	1.071	0.951	0.963	1.067
15.Machine Operators/ professional	1.080*	0.788***	0.811***	1.060
16.Transportation/ professional	1.182***	0.874***	0.909**	1.157***
17.Handlers, Laborers/ professional	1.218***	0.657***	0.681***	1.192***
Observations	355,449	480,494	480,494	348,989

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

Column 3 contains four categories. Once again, the results are relative to all other ages. There is a clear and expected increase in mortality with age. Finally, the last column uses a single multi-valued variable where the baseline category is ages 25–34. Relative to this baseline age category that contains working age adults with the lowest mortality, all other age categories have an increased and significant mortality risk, with the risk increasing with age. As in the results in chapter 3 (see Table 13), the married hazard ratio becomes riskier to health when the entire age range is used (column 2 and column 3).

The occupation variable results in columns 1 and 4 are essentially equivalent as are the results in columns 2 and 3. Since the basis for the regression, with respect to the Age variables, is the same in columns 1 and 4 and in columns 2 and 3, the equivalency of the results is not unexpected. The Age and Age2 variables (in column 1), and the multi-valued age variable (used in column 4) were used in subsequent analyses as a sensitivity check on the age dummies; the results were similar to those analyses and are not shown.

Age impacts and cause-specific mortality conditions are shown in Table 27, Table 28, Table 29, and Table 30. The age variables used are Age and Age2, and the multi-valued age variable in Table 27 and Table 28. Table 29 and Table 30 show analyses results for specific age ranges: 25–44 years of age, 45–64 years of age, and 65–plus years of age.

Table 27 shows the results for Tumor-related deaths and for Cardiovascular-related deaths. Table 28 shows the results for Injury-related deaths and Other-related deaths. In both tables, the baseline variables of Race, Gender, and Rural/Urban generally demonstrate the results expected, i.e., blacks have higher risk of death than whites, while other races have a lower risk; females have a lower risk than males; and living in rural areas is better than living in urban areas. The majority of these results have significant hazard ratios. The results for the Major occupation categories are mixed, with generally all categories showing a higher risk of death than the base Professional category (except for farmers, farm workers, and extractive workers). Tumor-related deaths are most important in Sales workers, Construction workers, Transportation workers, and general Laborers. Cardiovascular deaths are more prevalent in Private Household workers and Service workers.

Table 27: Age Impacts on Tumor-Related and Cardiovascular-Related Deaths

Hazard Ratios	Baseline	Age & Age^2	Ages	Baseline	Age & Age^2	Ages
OUTCOME	Tumors	Tumors	Tumors	Cardio.	Cardio.	Cardio.
VARIABLES/base						
Age, mean-centered		1.135***			1.150***	
Age^2		0.999***			0.999***	
Ages 35-44/Ages 25-34			3.795***			5.514***
Ages 45-54/Ages 25-34			12.42***			18.07***
Ages 55-65/Ages 25-34			29.95***			49.89***
Race:Black/ white	1.235***	1.234***	1.217***	1.226***	1.226***	1.206***
Race:Other/ white	0.746***	0.746***	0.704***	0.808*	0.807*	0.813*
Female/male	0.685***	0.685***	0.690***	0.273***	0.273***	0.269***
Rural/urban	0.937**	0.937**	0.933**	0.953	0.953	0.926**
Married/other than married	0.851***	0.851***	0.858***	0.744***	0.744***	0.761***
1.Executive/ professional	1.091	1.090	1.075	0.999	1.000	1.012
2.Professional	------	------	------	------	------	------
3.Technician/ professional	1.063	1.064	1.045	1.117	1.116	1.102
4.Sales/ professional	1.176***	1.175***	1.163**	0.989	0.990	1.024
5.Clerical/ professional	1.134**	1.134**	1.108*	1.047	1.048	1.089
6.Private Household/ professional	1.013	1.013	0.993	1.420***	1.419***	1.506***
7.Protective Services/ professional	1.111	1.110	1.023	1.145	1.147	1.168
8.Service-not protective/ professional	1.151**	1.151**	1.126*	1.245***	1.248***	1.314***
9.Farm Managers/ professional	0.819**	0.819**	0.827*	0.777***	0.778***	0.834*
10.Farm Workers/ professional	0.975	0.975	0.938	0.980	0.978	0.993
11.Mechanics/ professional	1.196**	1.195**	1.141*	0.982	0.982	0.995
12.Construction/ professional	1.241***	1.242***	1.206**	0.932	0.930	0.933
13.Extractive/ professional	0.614	0.615	0.606	0.932	0.930	0.960
14.Precision Production/ professional	1.124	1.124	1.094	0.999	0.998	1.022
15.Machine Operators/ professional	1.077	1.077	1.036	1.089	1.088	1.088
16.Transportation/ professional	1.277***	1.278***	1.220***	1.122	1.122	1.128
17.Handlers, Laborers/ professional	1.280***	1.281***	1.240***	1.132	1.127	1.138
Educ: None/ high school-H4		0.882	0.877		0.609**	0.666*
Educ: E1-E4/ high school-H4		0.997	1.028		1.042	0.997
Educ: E5-E6/ high school-H4		1.069	1.118		1.017	1.061
Educ: E7-E8/ high school-H4		1.138***	1.196***		1.152***	1.215***
Educ: H1/ high school-H4		1.121*	1.134**		1.101	1.129*

Hazard Ratios	Baseline	Age & Age^2	Ages	Baseline	Age & Age^2	Ages
OUTCOME	Tumors	Tumors	Tumors	Cardio.	Cardio.	Cardio.
Educ: H2/ high school-H4		1.166***	1.179***		1.203***	1.205***
Educ: H3/ high school-H4		1.177***	1.177***		1.147**	1.143**
Educ: H4		------	------		------	------
Educ: C1/ high school-H4		0.962	0.947		0.920	0.887*
Educ: C2/high school-H4		1.007	1.003		0.972	0.949
Educ: C3/high school-H4		1.009	0.979		0.868	0.853
Educ: C4/high school-H4		0.841***	0.815***		0.726***	0.719***
Educ: C5/high school-H4		0.754***	0.719***		0.575***	0.552***
Educ: C6/high school-H4		0.781***	0.757***		0.607***	0.604***
Income:0-5K/25K-30K		1.199**	1.225**		1.303***	1.299***
Income:5K-7.5K/25K-30K		1.355***	1.374***		1.398***	1.413***
Income:7.5K-10K /25K-30K		1.023	1.048		1.155*	1.160*
Income:10K-12.5K /25K-30K		1.164**	1.199**		1.284***	1.318***
Income:12.5K-15K /25K-30K		1.004	1.023		1.173**	1.171**
Income:15K-20K /25K-30K		1.120**	1.134**		1.031	1.046
Income20K-25K/25K-30K		1.029	1.029		1.014	1.014
Income25K-30K		------	------		------	------
Income:30K-35K/25K-30K		0.921	0.903		0.779***	0.755***
Income:35K-40K/25K-30K		0.928	0.912*		0.829***	0.815***
Income:40K-50K/25K-30K		0.814***	0.807***		0.718***	0.694***
Income:50K-60K/25K-30K		0.986	0.970		0.857***	0.845***
Income:60K-75K/25K-30K		0.827***	0.825***		0.718***	0.698***
Income:75K+/ 25K-30K		0.873**	0.875**		0.704***	0.702***
Observations		355,449	348,989		355,449	348,989

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

Table 28: Age Impacts on Injury-Related and Other-Related Deaths

Hazard Ratios	Baseline	Age & Age^2	Ages	Baseline	Age & Age^2	Ages
OUTCOME	Injury	Injury	Injury	Other	Other	Other
VARIABLES/base						
Age, mean-centered		1.006**			1.086***	
Age^2		1.000*			1.000***	

Hazard Ratios	Baseline	Age & Age^2	Ages	Baseline	Age & Age^2	Ages
OUTCOME	Injury	Injury	Injury	Other	Other	Other
Ages 35-44/Ages 25-34			1.041			2.214***
Ages 45-54/Ages 25-34			1.126*			5.306***
Ages 55-65/Ages 25-34			1.305***			13.47***
Race:Black/ white	1.181**	1.179**	1.172**	1.516***	1.515***	1.494***
Race:Other/ white	1.206	1.201	1.193	1.049	1.051	1.056
Female/male	0.315***	0.315***	0.315***	0.359***	0.359***	0.361***
Rural/urban	1.150***	1.149***	1.142***	0.808***	0.809***	0.802***
Married/other than married	0.654***	0.656***	0.657***	0.555***	0.555***	0.564***
1.Executive/ professional	0.941	0.943	0.945	1.170**	1.172**	1.168**
2.Professional	------	------	------	------	------	------
3.Technician/ professional	1.161	1.162	1.177	1.161	1.161	1.104
4.Sales/ professional	1.081	1.080	1.062	1.098	1.098	1.128
5.Clerical/ professional	1.003	1.005	1.002	1.262***	1.263***	1.276***
6.Private Household/ professional	1.313	1.303	1.372	1.211	1.210	1.236
7.Protective Services/ professional	1.160	1.166	1.157	1.125	1.132	1.111
8.Service-not protective/ professional	1.259*	1.261*	1.273*	1.352***	1.355***	1.379***
9.Farm Managers/ professional	1.023	1.022	1.044	0.625***	0.627***	0.675***
10.Farm Workers/ professional	1.275	1.271	1.265	0.835	0.834	0.843
11.Mechanics/ professional	1.095	1.099	1.087	0.931	0.931	0.906
12.Construction/ professional	1.383**	1.382**	1.398***	0.899	0.900	0.890
13.Extractive/ professional	2.416***	2.419***	2.453***	1.104	1.106	0.934
14.Precision Production/ professional	1.147	1.149	1.158	1.072	1.073	1.064
15.Machine Operators/ professional	1.112	1.114	1.112	1.052	1.053	1.036
16.Transportation/ professional	1.372**	1.370**	1.354**	1.051	1.053	1.029
17.Handlers, Laborers/ professional	1.453***	1.452***	1.469***	1.116	1.114	1.060
Educ: None/ high school-H4	1.276	1.274	1.326	0.661	0.659	0.730
Educ: E1-E4/ high school-H4	0.779	0.782	0.692	1.043	1.043	1.080
Educ: E5-E6/ high school-H4	1.182	1.189	1.231	1.190**	1.189**	1.227**
Educ: E7-E8/ high school-H4	1.224**	1.227**	1.220**	1.090	1.089	1.163***
Educ: H1/ high school-H4	0.916	0.915	0.938	1.218***	1.217***	1.246***
Educ: H2/ high school-H4	1.070	1.069	1.076	1.083	1.081	1.090
Educ: H3/ high school-H4	1.152	1.154	1.127	1.168**	1.168**	1.196**
Educ: H4	------	------	------	------	------	------
Educ: C1/ high school-H4	0.961	0.964	0.971	0.890	0.891	0.889

Hazard Ratios	Baseline	Age & Age^2	Ages	Baseline	Age & Age^2	Ages
OUTCOME	Injury	Injury	Injury	Other	Other	Other
Educ: C2/high school-H4	0.920	0.924	0.903	0.979	0.981	0.982
Educ: C3/high school-H4	0.895	0.897	0.889	1.021	1.021	0.989
Educ: C4/high school-H4	0.659***	0.662***	0.656***	0.776***	0.777***	0.797***
Educ: C5/high school-H4	0.452***	0.456***	0.459***	0.869	0.872	0.859
Educ: C6/high school-H4	0.606***	0.610***	0.612***	0.691***	0.693***	0.689***
Income:0-5K/25K-30K	1.780***	1.783***	1.787***	1.599***	1.597***	1.610***
Income:5K-7.5K/25K-30K	1.632***	1.639***	1.584***	1.637***	1.642***	1.674***
Income:7.5K-10K /25K-30K	1.135	1.138	1.164	1.440***	1.446***	1.491***
Income:10K-12.5K /25K-30K	1.707***	1.712***	1.730***	1.186**	1.192**	1.219**
Income:12.5K-15K /25K-30K	1.130	1.129	1.140	1.234***	1.235***	1.264***
Income:15K-20K /25K-30K	0.988	0.988	1.000	1.158**	1.161**	1.183**
Income20K-25K/25K-30K	1.024	1.025	1.031	1.170**	1.170**	1.190***
Income25K-30K	------	------	------	------	------	------
Income:30K-35K/25K-30K	0.852	0.854	0.864	1.006	1.007	0.973
Income:35K-40K/25K-30K	0.829**	0.830**	0.844*	0.868**	0.869**	0.862**
Income:40K-50K/25K-30K	0.769**	0.770**	0.768**	0.773***	0.774***	0.752***
Income:50K-60K/25K-30K	0.874	0.873	0.879	0.747***	0.748***	0.734***
Income:60K-75K/25K-30K	0.852	0.849	0.863	0.681***	0.679***	0.633***
Income:75K+/ 25K-30K	0.868	0.863	0.886	0.701***	0.700***	0.689***
Observations	355,449	355,449	348,989	355,449	355,449	348,989

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

In Table 28, there are two changes to the baseline variable results. First, other races have a higher risk of injury-related deaths and other-related deaths than blacks or whites. Second, there is a higher risk of injury-related deaths in rural areas than in urban areas. Because farmers and extractive workers are much more likely to live in rural areas (as shown in Table 22), both groups have their highest risk of death in the injury category. Laborers and construction workers also show a high level of injury-related deaths. Service workers and clerical workers show a high rate of other-related deaths.

In Table 29 and Table 30, the baseline variables of Race, Gender, and Rural/Urban generally demonstrate the results expected, i.e., blacks have higher risk of death than whites, while other races have a lower risk; females have a lower risk than males; and living in rural areas is better than living in urban areas. The two exceptions are the same as those in describe for Table 28, i.e., in Table 30, other races show a higher risk for Injury-related and Other-related deaths and rural people have a higher risk for injury-related deaths. More details are now evident. For other races, the higher rate is limited to those less than 65 years of age; for those over 65 the rates return to the much lower rate seen in the baseline analysis. For rural injury-related deaths, the rate remains high across all age groups.

Table 29: Age Groups and Causes of Death, Part 1

Hazard Ratios	25-44 years	45-65 years	65-plus years	25-44 years	45-65 years	65-plus years
OUTCOME	Tumors	Tumors	Tumors	Cardio.	Cardio.	Cardio.
VARIABLES/base						
Race:Black/ white	1.187*	1.233***	1.209*	1.474***	1.169***	0.888
Race:Other/ white	0.653*	0.773**	0.962	0.685	0.838	0.560**
Female/male	0.983	0.629***	0.452***	0.274***	0.272***	0.422***
Rural/urban	1.019	0.919***	0.898*	0.969	0.951	0.844***
Married/other than married	0.772***	0.857***	0.895*	0.778***	0.735***	0.900*
1.Executive/ professional	0.896	1.150**	0.902	1.288	0.953	1.123
2.Professional	------	------	------	------	------	------
3.Technician/ professional	0.729	1.204*	1.168	1.601**	1.033	1.108
4.Sales/ professional	1.149	1.205***	0.896	1.130	0.961	1.053
5.Clerical/ professional	1.042	1.165**	1.159	1.373*	0.996	1.195
6.Private Household/ professional	1.263	1.047	0.711*	3.311***	1.296*	1.102
7.Protective Services/ professional	0.592*	1.280**	0.899	0.717	1.223*	1.063

Hazard Ratios	25-44 years	45-65 years	65-plus years	25-44 years	45-65 years	65-plus years
OUTCOME	Tumors	Tumors	Tumors	Cardio.	Cardio.	Cardio.
8.Service-not protective/ professional	1.000	1.211***	0.916	2.096***	1.134	1.022
9.Farm Managers/ professional	0.687	0.856	0.778*	0.873	0.746***	1.018
10.Farm Workers/ professional	0.648	1.064	0.613***	1.512	0.902	1.053
11.Mechanics/ professional	0.949	1.286***	1.256	1.279	0.933	0.943
12.Construction/ professional	1.023	1.320***	0.804	1.171	0.893	0.925
13.Extractive/ professional	0.870	0.545	0.921	2.663**	0.620	3.481**
14.Precision Production/ professional	0.980	1.178**	1.013	1.207	0.962	0.990
15.Machine Operators/ professional	1.056	1.102	1.104	1.494**	1.023	1.420**
16.Transportation/ professional	1.274	1.300***	0.935	1.458*	1.066	1.047
17.Handlers, Laborers/ professional	1.072	1.353***	0.808	1.517**	1.066	1.027
Educ: None/ high school-H4	0.939	0.858	0.965	0.965	0.591**	1.102
Educ: E1-E4/ high school-H4	0.646	1.014	0.941	0.860	1.079	0.878
Educ: E5-E6/ high school-H4	0.666	1.105	0.893	0.932	1.039	0.993
Educ: E7-E8/ high school-H4	1.165	1.138***	1.088	1.190	1.157***	1.051
Educ: H1/ high school-H4	1.448***	1.067	0.903	1.353*	1.070	0.986
Educ: H2/ high school-H4	1.053	1.182***	1.068	1.375**	1.180***	1.080
Educ: H3/ high school-H4	0.992	1.209***	1.183	1.315*	1.120*	1.357***
Educ: H4	------	------	------	------	------	------
Educ: C1/ high school-H4	0.985	0.950	0.928	1.077	0.883	0.908
Educ: C2/high school-H4	0.835	1.067	1.276**	1.008	0.972	0.833
Educ: C3/high school-H4	0.989	1.025	1.088	0.850	0.881	1.077
Educ: C4/high school-H4	0.778**	0.865**	0.890	0.711**	0.738***	0.961
Educ: C5/high school-H4	0.431***	0.885	1.053	0.602*	0.578***	0.868
Educ: C6/high school-H4	0.652***	0.835**	0.733**	0.587***	0.618***	0.853
Income:0-5K/25K-30K	1.276	1.180*	0.906	1.228	1.307***	1.284*
Income:5K-7.5K/25K-30K	1.392*	1.348***	1.081	1.554**	1.357***	1.404***
Income:7.5K-10K /25K-30K	0.922	1.044	1.237	1.285	1.127	1.203
Income:10K-12.5K /25K-30K	1.137	1.171**	1.162	1.478**	1.241***	1.085
Income:12.5K-15K /25K-30K	0.807	1.047	1.333**	1.049	1.186**	1.149
Income:15K-20K /25K-30K	1.360**	1.070	1.114	1.020	1.028	1.127
Income20K-25K/25K-30K	1.254*	0.978	1.073	0.963	1.024	1.047
Income25K-30K	------	------	------	------	------	------
Income:30K-35K/25K-30K	0.897	0.926	0.804	0.855	0.765***	0.829
Income:35K-40K/25K-30K	0.788*	0.957	0.827	0.858	0.824***	1.120

Hazard Ratios	25-44 years	45-65 years	65-plus years	25-44 years	45-65 years	65-plus years
OUTCOME	Tumors	Tumors	Tumors	Cardio.	Cardio.	Cardio.
Income:40K-50K/25K-30K	0.929	0.786***	0.684**	0.705**	0.722***	0.704**
Income:50K-60K/25K-30K	1.121	0.948	0.877	0.942	0.841***	1.043
Income:60K-75K/25K-30K	0.844	0.816***	0.946	0.768	0.708***	0.757*
Income:75K+/ 25K-30K	0.855	0.855**	0.853	0.657**	0.702***	0.786*
Observations	234,117	121,332	11,148	234,117	121,332	11,148

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

Table 30: Age Groups and Causes of Death, Part 2

Hazard Ratios	25-44 years	45-65 years	65-plus years	25-44 years	45-65 years	65-plus years
OUTCOME	Injury	Injury	Injury	Other	Other	Other
VARIABLES/base						
Race:Black/ white	1.188*	1.112	1.257	1.925***	1.334***	0.982
Race:Other/ white	1.197	1.213	0.890	1.094	1.037	0.704
Female/male	0.307***	0.333***	0.349***	0.307***	0.390***	0.471***
Rural/urban	1.136**	1.148*	1.117	0.693***	0.860***	0.935
Married/other than married	0.637***	0.703***	0.729	0.412***	0.635***	0.962
1.Executive/ professional	0.808	1.182	0.385*	1.091	1.207**	1.226
2.Professional	------	------	------	------	------	------
3.Technician/ professional	1.089	1.375		0.986	1.277*	1.290
4.Sales/ professional	1.033	1.176	0.667	1.058	1.118	1.185
5.Clerical/ professional	0.970	1.084	0.969	1.286*	1.254**	1.263*
6.Private Household/ professional	1.162	1.411	0.462	3.233***	1.035	1.260
7.Protective Services/ professional	1.133	1.213	0.937	0.734	1.339**	1.426**
8.Service-not protective/ professional	1.305*	1.189	0.768	1.430***	1.343***	1.117
9.Farm Managers/ professional	0.623	1.387	0.608	0.735	0.607***	1.129
10.Farm Workers/ professional	1.180	1.439	1.368	0.953	0.802	1.093
11.Mechanics/ professional	0.983	1.330	0.656	0.671**	1.060	1.269
12.Construction/ professional	1.339*	1.432	0.491	0.827	0.940	1.220
13.Extractive/ professional	2.380***	2.255	7.136*	1.005	1.196	1.288
14.Precision Production/ professional	0.978	1.490*	0.503	0.961	1.120	0.932
15.Machine Operators/	1.198	0.900	1.205	0.836	1.145	1.176

Hazard Ratios	25-44 years	45-65 years	65-plus years	25-44 years	45-65 years	65-plus years
OUTCOME	Injury	Injury	Injury	Other	Other	Other
professional						
16.Transportation/ professional	1.303*	1.488*	1.028	0.823	1.163	0.722
17.Handlers, Laborers/ professional	1.394**	1.531*	0.870	0.972	1.191	1.061
Educ: None/ high school-H4	0.705	1.689	1.223	0.714	0.675	0.717
Educ: E1-E4/ high school-H4	0.448	0.988	1.794	1.181	1.083	1.041
Educ: E5-E6/ high school-H4	0.926	1.442*	1.301	1.011	1.263***	0.929
Educ: E7-E8/ high school-H4	1.206	1.299**	1.416	1.218	1.080	1.107
Educ: H1/ high school-H4	1.036	0.814	1.290	1.075	1.257***	1.339***
Educ: H2/ high school-H4	1.135	1.026	1.578	1.146	1.063	0.964
Educ: H3/ high school-H4	1.187	1.130	1.614	1.288*	1.120	1.065
Educ: H4	------	------	------	------	------	------
Educ: C1/ high school-H4	0.940	0.996	1.144	0.963	0.850*	1.323**
Educ: C2/high school-H4	0.895	0.972	1.945*	1.116	0.903	1.170
Educ: C3/high school-H4	0.810	1.181	1.297	0.952	1.063	1.201
Educ: C4/high school-H4	0.546***	1.048	2.011*	0.665***	0.851**	1.021
Educ: C5/high school-H4	0.362***	0.772	0.748	0.730	0.972	1.231
Educ: C6/high school-H4	0.430***	1.116	0.740	0.710**	0.687***	0.927
Income:0-5K/25K-30K	1.713***	1.907***	1.352	1.662***	1.562***	1.049
Income:5K-7.5K/25K-30K	1.277	2.346***	1.862	1.569***	1.670***	1.463***
Income:7.5K-10K /25K-30K	1.082	1.252	1.764	1.576***	1.373***	1.438***
Income:10K-12.5K /25K-30K	1.538***	2.083***	1.412	1.109	1.211*	1.286**
Income:12.5K-15K /25K-30K	1.019	1.392*	1.738	1.315**	1.194**	1.180
Income:15K-20K /25K-30K	0.953	1.068	1.962*	1.158	1.146*	1.126
Income20K-25K/25K-30K	1.004	1.075	0.705	1.262**	1.128	0.893
Income25K-30K	------	------	------	------	------	------
Income:30K-35K/25K-30K	0.849	0.861	1.257	0.934	1.039	0.824
Income:35K-40K/25K-30K	0.791**	0.910	0.667	0.812*	0.894	0.802
Income:40K-50K/25K-30K	0.727**	0.853	0.948	0.644***	0.824**	0.864
Income:50K-60K/25K-30K	0.872	0.858	0.789	0.728***	0.755***	0.900
Income:60K-75K/25K-30K	1.017	0.642**	1.058	0.800	0.642***	0.552***
Income:75K+/ 25K-30K	1.089	0.605***	1.617	0.779	0.671***	0.728**
Observations	234,117	121,332	11,148	234,117	121,332	11,148

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

The results for the Major occupation categories vary across the age groups. Nearly every category has an age group that has a lower risk of death than Professionals and an age group with a higher risk of death. For example, Table 27 showed Sales, Construction, Transportation, and general Laborer workers to have the most significant risk of tumor-related deaths. In Table 29, for each category, there is a significant and high risk for the 45–65 year age group and a (non-significant) lower risk in the 65–plus year age group. For the Service workers and Private Household workers that show high cardiovascular risk in Table 27, it is the 25–44 year age group in Table 29 that bears that high risk. The risk drops dramatically in higher ages and becomes non-significant. Although as shown in Table 27, extractive workers and machine operators do not demonstrate a significant risk of cardiovascular risk, in Table 29 these categories have an interesting pattern. Both show a high risk at lower ages (25–44 years of age) and again at higher ages (65–plus years of age); the middle age group has a non-significant risk. One explanation is: inexperienced younger workers in these fields may suffer health incidents perhaps from stress or pressure from the rigors of a risky and demanding job. Middle age workers, by contrast, have a sense of the requirements and demands of the jobs, and their work experience allows them to manage themselves and their health, while elderly workers in these fields are past the point where they should be working under these demanding conditions, and once again, through lack of focus or diminution of physical attributes allow the job stresses to affect their health. In addition, elderly workers in these occupations may suffer more from lingering stress, environmental effects, or physical labor impacts, which then affect their health and cause of death in later years even when retired.

In Table 28, laborers and construction workers had the highest rate of injury-related deaths; service workers and clerical workers had the highest rate of other-related deaths. In Table 30, for injury-related deaths both laborers and construction workers have the most significant rate of death in the youngest age category (25–44 years of age), a less significant (but slightly higher rate) in the middle age group, and a “normal” rate in the 65–plus years of age group. One interpretation for these results may be that experience is a great companion. Younger, less-experienced workers suffer death through injuries at a higher rate. Once into their older years, working in these industries is no more of a life burden than in working in other professions; and after 65 years of age your health is better than if you had spent your career years working as a Professional. For other-related deaths, both clerical and service workers suffer a consistent risk of death throughout their lives, i.e., age and job experience do not seem to adjust the risk of dying by other causes dramatically. Having obtained the results with these broad age categories, future research should create smaller, more focused age groupings to better determine which age ranges are most important with these occupation categories. The NLMS is particularly suited for such a future study since it has nearly one million records.

Additional interesting results are demonstrated with the income and education impacts in Table 29 and Table 30. NLMS data shows that lower than average income is significantly worse for health and higher than average income is significantly better. The impact of education is similar. In Table 29, for both the tumor-related outcomes and the cardiovascular-related outcomes the education results generally follow the same pattern for the 25–44 years and the 45–65 years age groups. In both cases, the 65–plus age group results are mixed and typically not significant. In Table 30, this pattern is repeated for the Other-related outcome. For the Injury-related outcome, the 45–65 year age group shows a mixed and non-significant set of results. Income results in both tables are similar, i.e, the younger age group and the 45–65 year age group tend to follow the "standard" trend while the 65–plus year age group has mixed and non-significant results. These results support prior work (Adams et al. 2003; Kiuila and Mieszkowski 2007) in which the impact of SES variables on mortality are strongest for younger persons and persons in good health, and weakest for older persons and persons in poor health. In both papers, the mortality gradient persists into old age, but the significance is slight or non-existent at the older ages. The interpretation is that health outcomes are largely pre-determined by antecedent health-related factors throughout the life course. In those papers, the income-related impacts are those most affected. This may be due to those over 65–plus years of age receiving Medicare benefits and Social Security benefits, which tend to balance out variations in income. In Table 29 and Table 30, there is a general improvement in health risk as income increases, but many results are not significant. In all cases, the age group with the most consistent and significant results is the 45–65 year age group. This could reflect that the younger age group is not sufficiently established with a stream of discretionary income that can be used for health improvement, and that the older age group, as explained above, has leveled out their income through retirement. The 45–65 year age group, however is in its peak earning years and has accumulated sufficient occupational standing and life experience that increased income is primarily a benefit to improved health.

4.6.4 Occupation and Geographic Interaction Results

Table 31 shows interaction analyses using the Urban/Rural geography variable and the Major Occupation category for different demographic groups. The results are shown for all individuals, and by race and gender (white males, black males, white females, and black females). There are much smaller numbers of black men and black women in the analyses. Among all women (black and white), a few Major Occupation groups have no results due to the small number of women in those groups.

For white men and white women, the impact of education remains similar to the All group: a significantly higher result with less than a High School degree and a significantly lower result with more education than a High School degree. For black men and black women, those with less than a High School degree have a higher risk of death, but those with more than a High School education have mixed results with none being significant. Income results are similar. The only consistent result is for those in the highest income category (Income > $75,000) where the risk of death is significantly reduced.

Table 31: Occupation Results Interacted with Urban/Rural by Race and Gender

Hazard Ratios	All	All X Rural	White M	White M X Rural	Black M	Black M X Rural	White F	White F X Rural	Black F	Black F X Rural
OUTCOME	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause
VARIABLES /base										
Race:Black/ white	1.297***	1.298***								
Race:Other/ white	0.890**	0.890**								
Female/male	0.419***	0.419***								
Married/other than married	0.711***	0.710***	0.652***	0.652***	0.726***	0.725***	0.782***	0.782***	0.836**	0.841**
Educ: None/ high school-H4	0.751**	0.752**	0.761*	0.764*	0.583	0.606	0.595	0.586	1.782	1.756
Educ: E1-E4/ high school-H4	0.997	0.997	1.018	1.021	1.026	1.018	0.799	0.793	1.235	1.173
Educ: E5-E6/ high school-H4	1.080*	1.081*	1.055	1.054	1.209*	1.205*	0.984	0.978	1.021	1.008
Educ: E7-E8/ high school-H4	1.129***	1.130***	1.146***	1.148***	1.163*	1.164*	1.088	1.087	1.012	1.003
Educ: H1/ high school-H4	1.114***	1.115***	1.114**	1.115**	1.034	1.037	1.221***	1.220***	1.050	1.026
Educ: H2/ high school-H4	1.143***	1.145***	1.157***	1.159***	1.091	1.089	1.185***	1.185***	1.011	1.012
Educ: H3/ high school-H4	1.160***	1.161***	1.171***	1.173***	1.372***	1.368***	1.066	1.065	1.108	1.108
Educ: H4	------	------	------	------	------	------	------	------	------	------
Educ: C1/ high school-H4	0.932**	0.932**	0.953	0.953	0.875	0.874	0.911	0.911	0.878	0.870
Educ: C2/high	0.978	0.978	0.962	0.962	1.102	1.098	1.021	1.022	0.962	0.951

Hazard Ratios	All	All X Rural	White M	White M X Rural	Black M	Black M X Rural	White F	White F X Rural	Black F	Black F X Rural
OUTCOME	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause
school-H4										
Educ: C3/high school-H4	0.957	0.956	0.983	0.982	1.084	1.084	0.904	0.906	0.838	0.835
Educ: C4/high school-H4	0.764***	0.764***	0.753***	0.753***	0.923	0.915	0.820***	0.823***	0.827	0.815
Educ: C5/high school-H4	0.687***	0.687***	0.664***	0.664***	1.009	1.001	0.712***	0.713***	1.248	1.245
Educ: C6/high school-H4	0.681***	0.681***	0.658***	0.659***	1.053	1.048	0.770***	0.770***	1.272	1.269
Income:0-5K/25K-30K	1.396***	1.397***	1.344***	1.345***	1.668***	1.660***	1.373***	1.373***	1.309	1.279
Income:5K-7.5K/25K-30K	1.461***	1.460***	1.460***	1.456***	1.905***	1.899***	1.402***	1.403***	1.472**	1.456**
Income:7.5K-10K /25K-30K	1.177***	1.177***	1.213***	1.212***	1.184	1.168	1.255**	1.256**	0.932	0.926
Income:10K-12.5K /25K-30K	1.257***	1.256***	1.395***	1.391***	1.403***	1.393***	1.131	1.128	1.042	1.028
Income:12.5K-15K /25K-30K	1.123***	1.123***	1.154***	1.152***	1.151	1.157	1.109	1.109	1.191	1.184
Income:15K-20K /25K-30K	1.086**	1.086**	1.117**	1.117**	1.031	1.032	1.059	1.059	1.133	1.123
Income20K-25K/25K-30K	1.057*	1.057*	1.050	1.048	1.063	1.059	1.077	1.076	1.168	1.158
Income25K-30K	------	------	------	------	------	------	------	------	------	------
Income:30K-35K/25K-30K	0.889***	0.890***	0.869***	0.870***	0.766*	0.764*	0.968	0.965	1.085	1.073
Income:35K-40K/25K-30K	0.872***	0.872***	0.876***	0.876***	0.892	0.890	0.862**	0.860**	0.845	0.835
Income:40K-50K/25K-30K	0.769***	0.769***	0.733***	0.732***	0.832	0.830	0.861**	0.859**	0.860	0.858

Hazard Ratios	All	All X Rural	White M	White M X Rural	Black M	Black M X Rural	White F	White F X Rural	Black F	Black F X Rural
OUTCOME	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause
Income:50K-60K/25K-30K	0.880***	0.880***	0.866***	0.866***	0.866	0.868	0.926	0.925	0.943	0.918
Income:60K-75K/25K-30K	0.763***	0.763***	0.764***	0.763***	0.726*	0.726*	0.718***	0.717***	1.349	1.348
Income:75K+/ 25K-30K	0.782***	0.781***	0.787***	0.786***	0.452***	0.455***	0.882*	0.882*	0.371***	0.373***
Rural/urban	0.931***	0.942	0.937***	0.953	0.975	1.222	0.877***	0.853	0.905	1.274
1.Executive/ professional	1.062*	1.076*	1.094**	1.109**	1.111	1.135	1.090	1.091	1.008	1.066
2.Professional	------	------	------	------	------	------	------	------	------	------
3.Technician/ professional	1.107*	1.038	1.144*	1.108	0.921	1.073	1.128	0.908	1.280	1.415
4.Sales/ professional	1.089**	1.091**	1.116**	1.101*	1.121	1.140	1.085	1.094	1.039	1.177
5.Clerical/ professional	1.129***	1.127***	1.251***	1.260***	1.227	1.269	1.022	0.997	1.147	1.190
6.Private Household/ professional	1.169**	1.129	0.300	0.510	0.311		1.199	1.200	1.618**	1.578**
7.Protective Services/ professional	1.132**	1.079	1.200***	1.134	0.825	0.905	1.312	1.346	1.347	1.439
8.Service-not protective/ professional	1.238***	1.253***	1.317***	1.327***	1.334*	1.367*	1.147**	1.157*	1.366*	1.429*
9.Farm Managers/ professional	0.777***	1.178	0.759***	1.078	0.916	2.273	1.011	1.238		
10.Farm Workers/ professional	0.981	0.980	0.992	0.889	0.911	1.158	0.825	1.075	1.492	2.886**

Hazard Ratios	All	All X Rural	White M	White M X Rural	Black M	Black M X Rural	White F	White F X Rural	Black F	Black F X Rural
OUTCOME	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause
11.Mechanics/ professional	1.048	1.056	1.103*	1.115*	0.968	1.003	1.186	1.362	1.108	1.289
12.Construction/ professional	1.072	1.045	1.117**	1.097	0.999	0.960	0.736	0.830		
13.Extractive/ professional	1.103	1.217	1.209	1.354	0.842	1.104				
14.Precision Production/ professional	1.071	1.103*	1.168***	1.222***	0.781	0.791	0.910	0.879	0.973	0.833
15.Machine Operators/ professional	1.079*	1.089*	1.135**	1.128**	1.055	1.092	1.057	1.072	1.095	1.260
16.Transportation/ professional	1.183***	1.197***	1.235***	1.297***	1.156	1.108	1.125	1.062	0.560	0.529
17.Handlers, Laborers/ professional	1.217***	1.187***	1.243***	1.174**	1.186	1.249	1.185	1.239	1.322	1.348
Rural X 1.Exec		0.950		0.948		0.832		0.995		0.611
Rural X 2.Prof		------		------		------		------		------
Rural X 3.Tech		1.263*		1.116				1.915***		
Rural X 4.Sales		0.992		1.042		0.893		0.973		0.193
Rural X 5. Clerical		1.007		0.970		0.723		1.104		0.780
Rural X 6.Private		1.114		0				1.007		1.013
Rural X 7.Protect		1.201		1.212		0.239		0.861		
Rural X 8.Service		0.952		0.966		0.806		0.984		0.724
Rural X 9. FarmMgr		0.634***		0.678**		0.296		0.819		
Rural X 10. FarmLbr		0.994		1.172		0.545		0.736		0.181*

Hazard Ratios	All	All X Rural	White M	White M X Rural	Black M	Black M X Rural	White F	White F X Rural	Black F	Black F X Rural
OUTCOME	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause
Rural X 11. Mech		0.975		0.965		0.749		0.534		
Rural X 12. Construction		1.064		1.039		1.053		0.614		
Rural X 13. Extract		0.827		0.811						
Rural X 14. Precise		0.902		0.859		0.888		1.121		1.761
Rural X 15. Machine		0.968		1.012		0.754		0.973		0.395**
Rural X 16. Transport		0.962		0.880		1.076		1.155		0.992
Rural X 17. Laborer		1.074		1.151		0.686		0.888		0.822
Observations	355,449	355,449	178,676	178,676	14,375	14,375	138,487	138,487	16,419	16,419

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

Adding the interaction terms has minimal impact on education and income results, but makes the health benefit of rural living non-existent. For white men, there are several significant occupation categories, the majority of which reflect relatively worse health than that of Professional (Teachers). This result matches the data results shown in Figure 6 and Figure 7. For black men, white women, and black women, the occupations that have results demonstrating better health are Service workers and Private Household workers, but most occupations are not significant.

The All group has only two significant interactions: for Farmers, the significant value (0.63) implies that farmers in rural areas are much more likely than farmers in urban areas to have a lower death risk. This is not surprising given all the evidence that farmers have a lower risk of death and that the vast majority of farmers live in rural areas. The other group is Technicians with a significant interaction value (1.26) implying that rural living is more risky when one is a technician. White males only have a significant result (0.68) for Farmers. Black men have no significant interactions. White women Technicians have worse risk in rural areas. For black women, there is significant interaction (0.18) for Farm Labor, but that is one of the groups with the smallest number of black women and may be a statistical anomaly. There is also a significant interaction (0.4) for Machine operators. Since these analyses report results that have $p<0.10$ (i.e., at the 10% or better level), merely by chance 10% of the interaction results should be significant, which would be 1–2 interaction effects for each demographic group. It is, thus, possible to say that for these analyses there are no more significant geographyXoccupation interactions than would occur by chance.

Additional analyses were performed using different combinations of geography (e.g., State of Residence, SMSA Status) and occupation (e.g., Occupation Recode categories). There were too few significant interactions (or even significant occupation or geography results) to warrant saying anything interesting about the impact of geography on occupation mortality risk.

4.6.5 Occupation Factors Results

Occupation factors were determined using the factor analysis approach described in Appendix D. The results are two sets of factors: (1) a set of 22 factors across seven domains, and (2) a set of four overall factors determined using the entire set of 234 occupation characteristic variables.

Table 32 shows Cox Proportional Hazard (CPH) analyses using the 22 factors. Each analysis in the first seven columns uses factors from one domain. One of the analyses uses the four overall factors and an additional analysis uses the NORC Prestige scores and the Duncan SEI scores. The baseline variables all show results consistent with those in previous analyses. The signs, magnitudes, and significance are virtually unchanged.

In the Ability factor group, occupations requiring a higher cognitive ability or a high degree of fine motor abilities have a significant and beneficial impact on mortality risk. The gross motor skills ability has a negative impact on health.

For the Education/Experience factor group, those occupations requiring more training and more education and/or experience improve health significantly. In general, this is not surprising due to the education completed variable results in the analyses in chapters 3 and 4 reported so far. Since this analysis also controls for educational attainment, it is indicative of the strong association between education, experience, and health in this data set. Occupations with a high degree of educational focus are healthier.

In the Knowledge factor group, occupations that have a higher business knowledge and higher engineering knowledge have a significant and beneficial impact on health. Occupations with a high degree of social science knowledge or biomedical knowledge do not have an appreciable health impact.

In the Skills factor group, occupations with a requirement for organizational skills or technical skills improve health significantly. Perhaps this is similar to the relationships with business and engineering knowledge or cognitive ability.

In the Work Activities factor group, occupations requiring more analysis and decision making have a positive health impact. Occupations that involve interacting with others and working with things have no significant impact.

In the Work Context factor group, occupations that are physically challenging and occupations that operate in an office context both significantly improve health, while socially challenging occupations have a significantly negative association with health.

Finally, in the Work Style factor group, occupations with a high degree of leadership have a positive health impact and occupations with a high degree of cooperation have a negative health impact.

In the Overall factor group, occupations with a high degree of Reasoning and Complexity have a positive health benefit and those with a high value on People versus Things have a negative health association. These groupings may be similar to the

Cognitive ability, Analyze and Decide, and Leadership factors and to the Socially Challenging and Cooperation factors.

A final analysis shows that jobs with higher occupation prestige rankings have a significantly lower health risk; the Duncan SEI score has no significant impact on health with this data set.

From the review in subsection 4.5, occupational/job characteristics are suggested to be better drivers of health outcomes than occupation categories. A variety of characteristics has been discussed in the literature. Table 33 summarizes some of the representative papers. The most cited papers are the Whitehall and Whitehall II studies by Marmot et al. The results show substantial mortality differences by administrative rank while controlling for income, demographics, and social background. The bottom line is that job position matters. The mechanisms work through relationships with others—both peers and subordinates, differences in job strain or stress, and degree of control in the job. Since these positions have a high degree of stability and insurance coverage is not a concern in Britain, these psychosocial explanations carry more weight. Smith (J Smith 1999) pointed out that the Whitehall samples represent a single employer. More variation in the work environment would be desirable. There was no observation of the study participants prior to their employment, and so self-selection by health status into job grades is possible. Job-related factors are unlikely to be all that matter in this story: family and environmental attributes are missing and spouse's grade of employment was not examined. Nevertheless, as shown in Table 33, a sense of control, influence, and stress are relevant psychosocial indicators expressed by Marmot. The related occupational factors in this study are generally consistent. Skills:Organizational Skills, Work Activities:Analyze and Decide, Work Style:Leadership, and Work Activities:Work with Others have key psychosocial attributes that focus on responsibility, stress tolerance, self-control, guiding and directing subordinates, and coaching others. These match well with the characteristics of the lower mortality positions described by Marmot. The factor Overall: People vs. Things is similar to the factor Work Context: Socially Challenging because it has such attributes as impact of decisions on others, frequency of conflict situations, stress tolerance, and dealing with physically aggressive or angry people. The balancing between potentially healthy psychosocial attributes and unhealthy socially challenging attributes results an increase in mortality risk in this factor.

Table 32: Initial Results with Occupation Factors

Hazard Ratios	Ability Factors	Educ/Exp Factors	Knowledge Factors	Skills Factors	Work Activities Factors	Work Context Factors	Work Style Factors	Prestige, Duncan SEI	Overall Factors
OUTCOME	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause
VARIABLES/base									
Race:Black/white	1.318***	1.322***	1.316***	1.325***	1.332***	1.318***	1.326***	1.324***	1.305***
Race:Other/white	0.890*	0.890*	0.889*	0.890*	0.892*	0.888*	0.893*	0.889*	0.888**
Female/male	0.420***	0.419***	0.417***	0.415***	0.426***	0.410***	0.415***	0.431***	0.410***
Married/other than married	0.708***	0.709***	0.709***	0.709***	0.708***	0.710***	0.709***	0.708***	0.709***
Rural/urban	0.898***	0.899***	0.900***	0.904***	0.894***	0.908***	0.900***	0.894***	0.915***
Educ: None/high school-H4	0.781*	0.785*	0.783*	0.793*	0.791*	0.804	0.808	0.775*	0.741**
Educ: E1-E4/high school-H4	1.010	1.014	1.013	1.022	1.018	1.025	1.034	1.002	0.994
Educ: E5-E6/high school-H4	1.060	1.064	1.062	1.070	1.070	1.072	1.086*	1.056	1.082*
Educ: E7-E8/high school-H4	1.109***	1.114***	1.112***	1.120***	1.115***	1.118***	1.128***	1.102***	1.129***
Educ: H1/high school-H4	1.096**	1.101**	1.102**	1.105***	1.103**	1.100**	1.112***	1.091**	1.118***
Educ: H2/high school-H4	1.139***	1.142***	1.141***	1.145***	1.145***	1.140***	1.153***	1.135***	1.146***
Educ: H3/high school-H4	1.141***	1.142***	1.143***	1.146***	1.145***	1.139***	1.151***	1.135***	1.161***
Educ: H4	------	------	------	------	------	------	------	------	------
Educ: C1/high school-H4	0.924**	0.922**	0.921**	0.920**	0.921**	0.922**	0.916**	0.928*	0.933**
Educ: C2/high school-H4	0.962	0.960	0.959	0.956	0.958	0.958	0.951	0.968	0.982
Educ: C3/high school-H4	0.949	0.945	0.942	0.939	0.947	0.933	0.932	0.961	0.953
Educ: C4/high school-H4	0.754***	0.753***	0.751***	0.755***	0.752***	0.748***	0.741***	0.769***	0.766***
Educ: C5/high school-H4	0.645***	0.652***	0.645***	0.660***	0.652***	0.637***	0.635***	0.674***	0.684***

Hazard Ratios	Ability Factors	Educ/Exp Factors	Knowledge Factors	Skills Factors	Work Activities Factors	Work Context Factors	Work Style Factors	Prestige, Duncan SEI	Overall Factors
OUTCOME	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause
Educ: C6/high school-H4	0.671***	0.678***	0.655***	0.685***	0.675***	0.659***	0.654***	0.709***	0.672***
Income:0-5K/25K-30K	1.347***	1.350***	1.354***	1.357***	1.349***	1.371***	1.357***	1.343***	1.372***
Income:5K-7.5K/25K-30K	1.457***	1.458***	1.462***	1.467***	1.460***	1.475***	1.466***	1.454***	1.445***
Income:7.5K-10K /25K-30K	1.154***	1.156***	1.159***	1.161***	1.158***	1.165***	1.160***	1.152***	1.163***
Income:10K-12.5K /25K-30K	1.235***	1.237***	1.239***	1.241***	1.239***	1.242***	1.241***	1.234***	1.248***
Income:12.5K-15K /25K-30K	1.099**	1.100**	1.102**	1.103**	1.102**	1.106**	1.105**	1.099**	1.114***
Income:15K-20K /25K-30K	1.081**	1.081**	1.081**	1.083**	1.082**	1.085**	1.084**	1.080**	1.083**
Income20K-25K/25K-30K	1.079**	1.079**	1.079**	1.080**	1.079**	1.080**	1.080**	1.078**	1.059*
Income25K-30K	------	------	------	------	------	------	------	------	------
Income:30K-35K/25K-30K	0.891***	0.890***	0.890***	0.891***	0.891***	0.890***	0.890***	0.891***	0.891***
Income:35K-40K/25K-30K	0.886***	0.885***	0.885***	0.884***	0.884***	0.884***	0.883***	0.886***	0.875***
Income:40K-50K/25K-30K	0.770***	0.769***	0.769***	0.768***	0.768***	0.769***	0.767***	0.770***	0.773***
Income:50K-60K/25K-30K	0.895***	0.893***	0.894***	0.891***	0.890***	0.893***	0.888***	0.895***	0.886***
Income:60K-75K/25K-30K	0.758***	0.756***	0.757***	0.754***	0.754***	0.756***	0.752***	0.758***	0.769***
Income:75K+/25K-30K	0.799***	0.795***	0.795***	0.790***	0.790***	0.797***	0.788***	0.797***	0.788***
Ability: Gross Motor Skills, Strength, and Endurance	1.171**								
Ability: Cognitive Ability	0.759***								
Ability: Fine Motor Abilities	0.832***								
Ability: Auditory and Visual Processing	0.951								
Educ/Exp: Training		0.830***							

Hazard Ratios	Ability Factors	Educ/Exp Factors	Knowledge Factors	Skills Factors	Work Activities Factors	Work Context Factors	Work Style Factors	Prestige, Duncan SEI	Overall Factors
OUTCOME	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause
Educ/Exp: Education & Experience		0.824***							
Knowledge: Social Science			1.055						
Knowledge: Business			0.788***						
Knowledge: Bio-Medicine			1.006						
Knowledge: Engineering			0.814***						
Skills: Organizational				0.787***					
Skills: Quantitative				0.995					
Skills: Technical				0.886**					
Work Activities: Analyze & Decide					0.860*				
Work Activities: Interact with Others					0.940				
Work Activities: Work with Things					0.978				
Work Context: Physically Challenging						0.866***			
Work Context: Office						0.719***			
Work Context: Socially Challenging						1.417***			
Work Context: Repetition						1.053			
Work Style: Leadership							0.705***		
Work Style: Cooperation							1.362***		
Occupation Prestige Ranking								0.663***	

Hazard Ratios	Ability Factors	Educ/Exp Factors	Knowledge Factors	Skills Factors	Work Activities Factors	Work Context Factors	Work Style Factors	Prestige, Duncan SEI	Overall Factors
OUTCOME	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause
Ducan SEI Score								1.098	
Overall: Reasoning & Complexity									0.738***
Overall: Physical Demands									0.972
Overall: People versus Things									1.201***
Overall: Attention to Detail									0.977
Observations	302,890	302,890	302,890	302,890	302,890	302,890	302,890	302,890	354,973

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

Table 33: Comparison of Occupation Factors to Literature Results

Occupation Characteristic	Reference(s)	Reported Impact on Health	Corresponding Occupational Factor(s)	Occupation Factor Impact on Health Risk
Conscientiousness (Orderliness, Self-control, Conventionality) Sense of Control (High Pace, Supporting, Subordinates) Salience in Authority Decision Latitude Salience in Influence	(M. Marmot et al. 1991; Michael Marmot et al. 1997; Lantz et al. 1998; Zhou 2005; Roberts et al. 2007)	Decrease in mortality	Skills: Organizational Skills	Decrease, $p<0.01$
			Work Activities: Analyze and Decide	Decrease, $p<0.01$
			Work Style: Leadership	Decrease, $p<0.01$
			Work Activities: Interact with Others	Decrease, not sig.
			Overall: People vs. Things	Increase, $p<0.01$
Extraversion	(Roberts et al. 2007)	Decrease in mortality	Work Activities: Analyze and Decide	Decrease, $p<0.01$
Optimism	(Lantz et al. 1998)	Decrease in mortality	Work Activities: Interact with Others	Decrease, not sig.
Agreeableness	(Roberts et al. 2007)	No clear association	Work Style: Cooperation	Increase, $p<0.01$
Openness to Experience (Creativity) Salience in Creativity Cognitive Abilities	(Zhou 2005; Roberts et al. 2007)	No clear association	Ability: Cognitive Ability	Decrease, $p<0.01$
			Overall: Reasoning & Complexity	Decrease, $p<0.01$
Stress (chronic and acute)	(M. Marmot et al. 1991; Michael Marmot et al. 1997; Lantz et al. 1998; Adler and Newman 2002; MacLeod et al. 2005)	Increase in mortality	Work Style: Cooperation	Increase, $p<0.01$
			Work Context: Socially Challenging	Increase, $p<0.01$
			Overall: Physical Demands	Decrease, not sig.
Neuroticism (Pessimism)	(Roberts et al. 2007)	Increase in mortality	Overall: People vs. Things	Increase, $p<0.01$
Degree of Physical Hazard Environmental Effects (Heat, Cold, Exposure)	(Warren and Kuo 2003; Fletcher, Sindelar, and Yamaguchi 2008)	Increase in mortality	Work Context: Physically Challenging	Decrease, $p<0.01$
			Overall: Physical Demands	Decrease, not sig.
Strength (Standing, Lifting, Pulling, Pushing, Amount of Controls used)	(Warren and Kuo 2003; Fletcher, Sindelar, and Yamaguchi 2008)	Increase in mortality	Work Context: Office Context (negative)	Increase, $p<0.01$
			Work Activities: Work with Things	Decrease, $p<0.01$
			Overall: Physical Demands	Decrease, not sig.

There are four key differences, as shown in Table 33, between the previous literature and the results in this dissertation:

(1) Roberts reports mixed results on the degree of Agreeableness on mortality. The factor Work Style: Cooperation (with attributes such as concern for others, self-control, social orientation, and flexibility) fits best with the concept of agreeableness, and this factor shows a marked increased risk of mortality, both in significance and magnitude. The occupation attribute category Work Style creates just two factors: Leadership and Cooperation. The Leadership factor (with attributes such as achievement, persistence, innovation, initiative, and leadership) demonstrates as strong a positive impact on health as Cooperation does negative. It may be that the attributes for Cooperation reflect a social relationship that adds to a lack of authority or sense of control over others, while the Leadership factor reflects control, independence, and more latitude to focus on effort and achievement. The result is that the Cooperation factor has a consistently negative impact on health. From a more economic perspective, the Leadership factor may be more directly related to more education, a more prestigious position, or a higher income—all of which affect health positively. The Cooperation factor may be more directly related to less education, a less prestigious position, less income, or the stress of trying to work with leaders.

(2) Zhou and Roberts report mixed results for the association of creativity and cognitive abilities with health outcomes. This is explained as a lack of consistent interpretation of the attributes of these social factors. In this data set, the

occupation factors Ability: Cognitive Ability and Overall: Reasoning & Complexity embody the breadth of related attributes, including complex problem solving, critical thinking, thinking creatively, originality, innovation, deductive reasoning, and inductive reasoning. Both occupation factors are consistently significant and positively impact health. This likely reflects a number of job-related effects; for example, these attributes contribute to success and advancement, which then improves social standing, income, and position. Similarly, higher prestige occupations correlate with better health, and these attributes are representative of occupations with higher prestige, e.g., engineers, physicians, and mathematicians. Finally, these two occupation factors, particularly Overall: Reasoning & Complexity, contain a large number of attributes that are often included in intelligence quotient (IQ) values. To the extent that the Overall: Reasoning & Complexity factor is representative of the "job IQ" of the occupation, there is clear relationship between increased job IQ and better health. [Note: as is shown in Table 34, when the Overall factors are co-regressed with the other 22 domain factors, the Overall: Reasoning & Complexity is the only factor that remains significant with a positive impact on health across all regressions. This suggests that job IQ is fundamental to explaining the relationship of occupations and health.] Figure 11 shows the state-by-state geographic distribution of IQ ranking, the Ability: Cognitive Ability factor, and the Overall: Reasoning & Complexity factor (dark blue represents a higher ranking; light green represents the lowest ranking). Although individual IQ rank and job IQ are rather different constructs, there is some commonality between the images, particularly in the northern and northeastern states. As the Overall factor contains most of the attributes of the Ability factor, it is not surprising that these two images are more similar.

(3) and (4) Warren and Kuo, and Fletcher, Sindelar, and Yamaguchi report that the degree of physical hazard, exposure to hazardous occupation environmental effects, and harsh physical demands cause an increase in mortality due to the dangerous nature of the work. In addition, manual work is more often associated with physical hazards than non-manual work, and manual work is subject to more exposure to environmental effects.

There is a confounding effect of prestige, as manual occupations are near the bottom of many prestige ranking ladders. Warren and Kuo looked at a variety physical and environmental job characteristics regressed against a variety of self-reported health issues as dependent variables. They found mixed results with most relating physical work to aching muscles or stiff and swollen joints as health issues. Their conclusion is that what people do for a living does matter for health beyond the benefits of educational attainment.

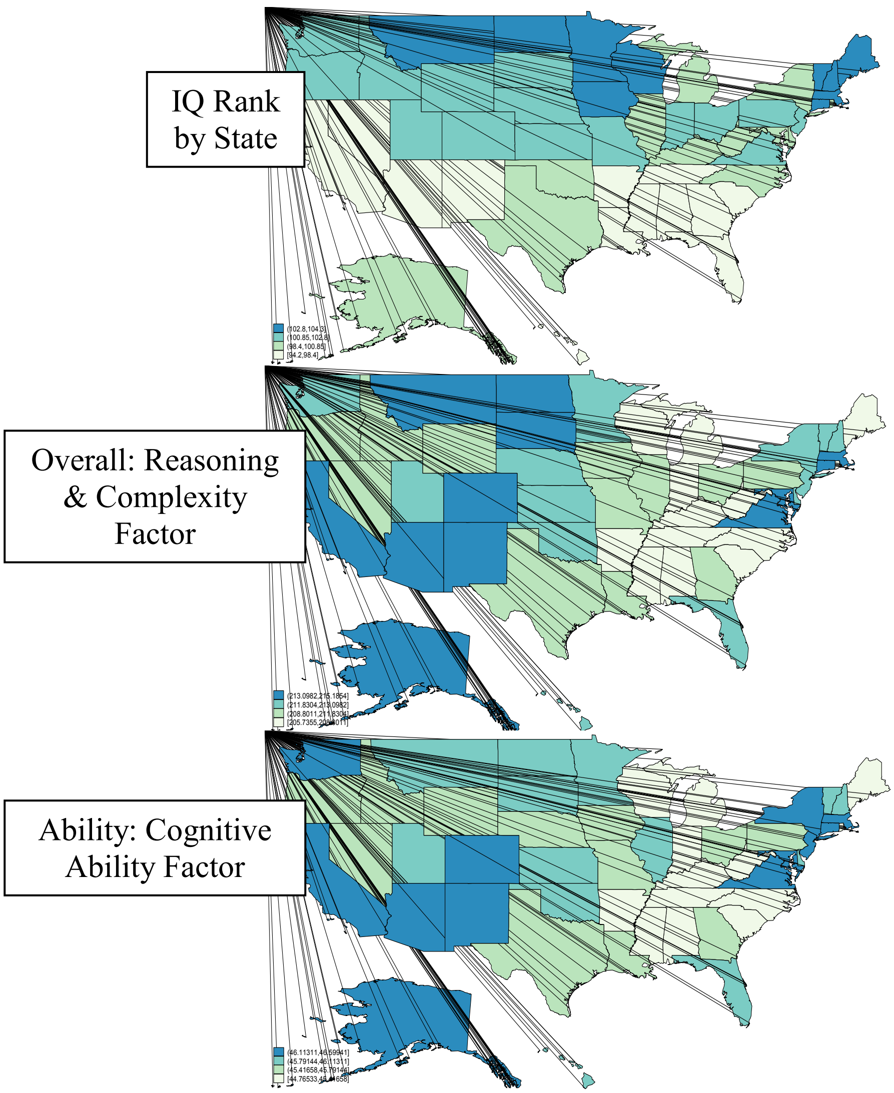

Figure 11: IQ and Job IQ Geographic Distribution

Fletcher and Sindelar found that the impact of job characteristics varies by demographic group. For example, there were very small impacts for white males and for black males, a one standard deviation increase in physical demands decreases health by an amount equivalent to four years of aging. They also found more health impacts with older workers. Including labor income, however, cushions the negative effects of job exposures and these negative impacts on blacks and older workers are no longer significant. In Table 33, the occupation factors related to the degree of physical hazard, Work Context: Physically Challenging and Overall: Physical Demands, decrease the risk of mortality. The factors related to strength, Work Activities: Work with Things and Overall: Physical Demands, decrease the risk of mortality. The factor Work Context: Office Context contains a number of office-related attributes that contribute to a higher factor value, and a number of physical attributes that contribute negatively to the factor value. Thus, any result of this factor for a particular occupation is a balance between positive office terms and negative physical terms. The result in Table 34 shows a positive and significant health impact. This means that a larger office-related context for an occupation positively impacts health (implying that a larger physical-related context would negatively affect health). Numerous other reported results demonstrate the positive health effects of physical exercise (two previously referenced reports are (Lantz et al. 1998; Lantz et al. 2001)). The results in Table 33 that show a decrease in mortality risk with increased degree of physical hazard or increased strength of activity could reflect that regular physical exertion improves health. Although some aspects of physical work may cause injury, with this data set the overall association with physical work apparently works equivalently to an increase in exercise to improve health outcomes.

Table 34: Co-regression with Overall Factors and Occupation Factors

NOTE: remaining results not shown

Hazard Ratios	22 Group Factors	22 Group Factors + Overall Factors	Ability Factors + Overall	Educ/Exp Factors + Overall	Knowledge Factors + Overall	Skills Factors + Overall	Work Activities Factors + Overall	Work Context Factors + Overall	Work Style Factors + Overall
OUTCOME	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause
VARIABLES									
Ability: Gross Motor Skills, Strength, and Endurance	0.956	0.933	1.075						
Ability: Cognitive Ability	0.828	0.966	1.032						
Ability: Fine Motor Abilities	0.820*	0.756**	0.768**						
Ability: Auditory and Visual Processing	0.939	0.869	0.934						
Educ/Exp: Training	0.917	0.931		0.966					
Educ/Exp: Education & Experience	1.079	0.944		1.083					
Knowledge: Social Science	0.962	0.986			1.115				
Knowledge: Business	0.994	1.275**			0.961				
Knowledge: Bio-Medicine	0.857*	1.240*			0.990				
Knowledge: Engineering	0.812*	0.560***			0.985				
Skills: Organizational	1.096	1.411*				1.520***			
Skills: Quantitative	1.191	1.159				1.288***			
Skills: Technical	0.795*	0.674***				0.794*			
Activities: Analyze & Decide	1.158	1.442*					1.628***		
Activities: Interact with Others	0.727***	0.771**					0.847		
Activities: Work with Things	1.469***	1.251					1.057		
Context: Physically Challenging	1.052	0.973						1.055	
Context: Office	0.823	0.931						0.952	
Context: Socially Challenging	1.501***	1.693***						1.511***	
Context: Repetition	0.964	0.905						1.022	

Hazard Ratios	22 Group Factors	22 Group Factors + Overall Factors	Ability Factors + Overall	Educ/Exp Factors + Overall	Knowledge Factors + Overall	Skills Factors + Overall	Work Activities Factors + Overall	Work Context Factors + Overall	Work Style Factors + Overall
OUTCOME	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause
Style: Leadership	1.037	1.035							1.083
Style: Cooperation	1.081	1.285**							1.142
Overall: Reasoning & Complexity		0.513***	0.669***	0.696***	0.501***	0.633***	0.722***	0.659***	0.684***
Overall: Physical Demands		1.699**	1.198	0.984	1.102	0.943	0.883	0.982	1.008
Overall: People versus Things		0.508***	1.170**	1.180*	1.197**	1.211***	0.966	1.097	1.212***
Overall: Attention to Detail		1.008	1.112	1.009	0.951	0.786***	0.946	0.987	0.979
Observations	302,890	302,890	302,890	302,890	302,890	302,890	302,890	302,890	302,890

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

In Table 34, the domain occupation factors are co-regressed with the Overall factors. The first analysis includes all 22 group factors. Many results are attenuated and non-significant. The results that are significant at the 1% level are Activities: Interact with Others, which has a positive impact on health, and Activities: Work with Things and Work Context: Socially Challenging, which have a negative impact on health. The next analysis adds the Overall factors to the 22 group factors. Work Context: Socially Challenging, Work Style: Cooperation, and Knowledge: Business are significant at the 5% or better level with negative health results. Ability: Fine Motor Abilities, Knowledge: Engineering, Skills: Technical, and Activities: Interact with Others are significant at the 5% or better level with positive health impacts. In the Overall category, Reasoning & Complexity and People versus Things both affect health positively and significantly; and Physical Demands impacts health negatively.

The remaining seven analyses use the seven domain groups plus the Overall factors. In general, the individual factors have less significant hazard ratios than the results in the analyses in Table 32. For the overall variables, the Reasoning & Complexity factor remains significant and beneficial to health in all cases. This factor contains a large number of attributes that are often included in intelligence quotient (IQ) values. To the extent that the Overall: Reasoning & Complexity factor is representative of the "job IQ" of the occupation, there is clear relationship between increased job IQ and better health. In the first five analyses Overall:People versus Things is harmful to health and significant. Since the overall factor, Reasoning & Complexity remains significant in all analyses, it is perhaps unsurprising the majority of the domain group factors related to reasoning ability, cognitive ability, or leadership ability are no longer significant. For example, Ability: Cognitive Ability factor is not significant; the Education and Experience factors are not significant; Skills: Organizational and Skills: Quantitative are significant and negatively affect health and Skills: Technical has a positive impact on health; Work Activities: Analyze and Decide is significant but negatively impacts health; and Work Context: Socially Challenging remains significant with a negative result. One can infer from these results that overall innate reasoning and cognitive aptitude are more important drivers of the relationship between occupations and health, than are education or experience or specific management skills.

In the previous results (Table 32), the Overall: People versus Things factor is significant and harmful to health in the Overall factor analysis. In these analyses, this factor is significant and harmful to health, except in the two co-regressions where there are two other social factors, i.e., Context: Socially Challenging and Style: Cooperation. When co-regressed, these two domain factors remain harmful while Overall: People versus Things is no longer significant.

4.6.6 Interaction Results and State Factor Results

Without interactions, the current model assumes that the contextual effect of occupation is the same for all geographic regions. By adding interaction terms, the impact of occupation on the rate of dying can depend on location. For these interaction analyses, I use the standard NLMS geography variables (Rural/Urban and SMSA Status) and the state-level factors to represent location and the Major occupation category and the occupation factors to represent job status.

Table 35 shows the results for the use of Rural/Urban, SMSA Status, and state-level factors regressed with the Major occupations and the overall occupation factors. Table 36 repeats the analyses with the interactions between Rural/Urban, SMSA Status, and the state-level factors with the occupation variables. Table 35 has consistent results for all combinations of geographic and occupation variables. The Rural and SMSA variables reflect better health with more rural conditions.

Table 35: Baseline Results with Occupation Factors and State Factors

NOTE: remaining results not shown

Hazard Ratios	Rural	SMSA Status	State Factors	Rural	SMSA Status	State Factors
	Major Occ	Major Occ	Major Occ	Occ Factors	Occ Factors	Occ Factors
OUTCOME	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause
VARIABLES/base						
Rural/urban	0.931***			0.915***		
SMSA: not City/central city		0.944***			0.942***	
SMSA: not SMSA/central city		0.923***			0.909***	
Factor 1: Big Fish, Small Pond			0.880***			0.868***
Factor 2: Up-and-Comers			1.149***			1.166***
Factor 3: Heartlanders			0.875*			0.878*
Factor 4: Empty Nesters			1.078**			1.078**
1.Executive/ professional	1.062*	1.062*	1.059			
2.Professional	------	------	------			
3.Technician/ professional	1.107*	1.108*	1.102*			
4.Sales/ professional	1.089**	1.090**	1.085**			
5.Clerical/ professional	1.129***	1.128***	1.123***			
6.Private Household/ professional	1.169**	1.171**	1.174**			
7.Protective Services/ professional	1.132**	1.131**	1.129*			
8.Service-not protective/ professional	1.238***	1.238***	1.235***			
9.Farm Managers/ professional	0.777***	0.762***	0.760***			

Hazard Ratios	Rural	SMSA Status	State Factors	Rural	SMSA Status	State Factors
	Major Occ	Major Occ	Major Occ	Occ Factors	Occ Factors	Occ Factors
OUTCOME	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause
10.Farm Workers/ professional	0.981	0.975	0.967			
11.Mechanics/ professional	1.048	1.048	1.039			
12.Construction/ professional	1.072	1.072	1.072			
13.Extractive/ professional	1.103	1.104	1.078			
14.Precision Production/ professional	1.071	1.071	1.068			
15.Machine Operators/ professional	1.079*	1.078*	1.076*			
16.Transportation/ professional	1.183***	1.180***	1.181***			
17.Handlers, Laborers/ professional	1.217***	1.216***	1.225***			
Overall: 1.Reasoning & Complexity				0.738***	0.731***	0.733***
Overall: 2.Physical Demands				0.972	0.967	0.968
Overall: 3.People versus Things				1.201***	1.206***	1.209***
Overall: 4.Attention to Detail				0.977	0.982	0.973
Observations	355,449	355,449	352,004	354,973	354,973	351,530

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

Big Fish, Small Pond states have high rankings in education, health, and low crime rates. Up-and-Comer states have high ranks in income, graduation percentage, the happiness index, and the freedom index. Heartlanders states have high manufacturing and farm output, and a high percentage of church attendance. Empty Nester states have a high percentage of people over 45 years of age and a low births per capita rank.

In Table 36, adding interaction terms attenuates the geography variables and reduces the level of significance. The Major occupation variable results do not change dramatically in the rural and SMSA columns, but are smaller and less significant in the state factors columns. Occupational factors and state factors are not longer significant when interacted with geographic variables.

Table 36: Interaction Effects with Occupation Factors and State Factors

NOTE: remaining results not shown

Hazard Ratios	Rural	SMSA Status	State Factors	Rural	SMSA Status	State Factors
	X Major Occ	X Major Occ	X Major Occ	X Occ Factors	X Occ Factors	X Occ Factors
OUTCOME	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause
VARIABLES/base						
Rural/urban	0.942			0.760***		
SMSA: not City/central city		0.942			0.953	
SMSA: not SMSA/central city		1.004			0.754***	
Factor 1: Big Fish, Small Pond			1.040			1.158
Factor 2: Up-and-Comers			1.169			0.843
Factor 3: Heartlanders			0.670*			1.091
Factor 4: Empty Nesters			1.040			1.164
1.Executive/ professional	1.076*	1.124*	0.961			
2.Professional	------	------	------			
3.Technician/ professional	1.038	1.143	1.385			
4.Sales/ professional	1.091**	1.122*	0.977			
5.Clerical/ professional	1.127***	1.167**	1.083			
6.Private Household/ professional	1.129	0.968	0.489			
7.Protective Services/ professional	1.079	1.084	0.552			
8.Service-not protective/ professional	1.253***	1.282***	0.963			
9.Farm Managers/ professional	1.178	1.380	0.989			
10.Farm Workers/ professional	0.980	0.885	0.563			
11.Mechanics/ professional	1.056	1.114	1.003			
12.Construction/ professional	1.045	1.208**	1.186			
13.Extractive/ professional	1.217	1.487	1.802			
14.Precision Production/ professional	1.103*	1.123	0.643			
15.Machine Operators/ professional	1.089*	1.086	0.949			
16.Transportation/ professional	1.197***	1.140*	1.048			
17.Handlers, Laborers/ professional	1.187***	1.205**	0.437**			
Overall: 1.Reasoning & Complexity				0.813***	0.892	1.052
Overall: 2.Physical Demands				0.969	0.966	1.069
Overall: 3.People versus Things				1.094	1.072	1.585
Overall: 4.Attention to Detail				0.892	0.803*	0.806
1.Exec X Rural	0.950					
2.Prof X Rural	------					
3.Tech X Rural	1.263*					

Hazard Ratios	Rural	SMSA Status	State Factors	Rural	SMSA Status	State Factors
	X Major Occ	X Major Occ	X Major Occ	X Occ Factors	X Occ Factors	X Occ Factors
OUTCOME	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause
4.Sales X Rural	0.992					
5. Clerical X Rural	1.007					
6.Private X Rural	1.114					
7.Protect X Rural	1.201					
8.Service X Rural	0.952					
9. FarmMgr X Rural	0.634***					
10. FarmLbr X Rural	0.994					
11. Mech X Rural	0.975					
12. Construction X Rural	1.064					
13. Extract X Rural	0.827					
14. Precise X Rural	0.902					
15. Machine X Rural	0.968					
16. Transport X Rural	0.962					
17. Laborer X Rural	1.074					
1.Exec X SMSA: not City		0.933				
2.Prof X SMSA: not City		------				
3.Tech X SMSA: not City		0.922				
4.Sales X SMSA: not City		1.016				
5. Clerical X SMSA: not City		0.998				
6.Private X SMSA: not City		1.437**				
7.Protect X SMSA: not City		1.005				
8.Service X SMSA: not City		1.071				
9. FarmMgr X SMSA: not City		0.522				
10. FarmLbr X SMSA: not City		1.120				
11. Mech X SMSA: not City		0.964				
12. Construction X SMSA: not City		0.801**				
13. Extract X SMSA: not City		0.651				
14. Precise X SMSA: not City		0.991				
15. Machine X SMSA: not City		1.055				
16. Transport X SMSA: not City		1.118				
17. Laborer X SMSA: not City		1.042				
1.Exec X SMSA: not SMSA		0.912				
2.Prof X SMSA: not SMSA		------				
3.Tech X SMSA: not SMSA		1.005				

Hazard Ratios	Rural	SMSA Status	State Factors	Rural	SMSA Status	State Factors
	X Major Occ	X Major Occ	X Major Occ	X Occ Factors	X Occ Factors	X Occ Factors
OUTCOME	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause
4.Sales X SMSA: not SMSA		0.893				
5. Clerical X SMSA: not SMSA		0.893				
6.Private X SMSA: not SMSA		1.247				
7.Protect X SMSA: not SMSA		1.138				
8.Service X SMSA: not SMSA		0.844**				
9. FarmMgr X SMSA: not SMSA		0.523				
10. FarmLbr X SMSA: not SMSA		1.063				
11. Mech X SMSA: not SMSA		0.867				
12. Construction X SMSA: not SMSA		0.877				
13. Extract X SMSA: not SMSA		0.695				
14. Precise X SMSA: not SMSA		0.868				
15. Machine X SMSA: not SMSA		0.917				
16. Transport X SMSA: not SMSA		0.972				
17. Laborer X SMSA: not SMSA		0.973				
1.Exec X F1.Big Fish			0.744**			
2.Prof X F1.Big Fish			------			
3.Tech X F1.Big Fish			0.976			
4.Sales X F1.Big Fish			0.757*			
5. Clerical X F1.Big Fish			0.787			
6.Private X F1.Big Fish			0.356***			
7.Protect X F1.Big Fish			0.941			
8.Service X F1.Big Fish			0.976			
9. FarmMgr X F1.Big Fish			0.583**			
10. FarmLbr X F1.Big Fish			0.785			
11. Mech X F1.Big Fish			0.882			
12. Construction X F1.Big Fish			0.563***			
13. Extract X F1.Big Fish			1.474			
14. Precise X F1.Big Fish			0.908			
15. Machine X F1.Big Fish			1.007			
16. Transport X F1.Big Fish			1.155			
17. Laborer X F1.Big Fish			0.988			
1.Exec X F2.Up-and-Comers			0.821			
2.Prof X F2.Up-and-Comers			------			
3.Tech X F2.Up-and-Comers			0.884			

Hazard Ratios	Rural	SMSA Status	State Factors	Rural	SMSA Status	State Factors
	X Major Occ	X Major Occ	X Major Occ	X Occ Factors	X Occ Factors	X Occ Factors
OUTCOME	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause
4.Sales X F2.Up-and-Comers			1.104			
5. Clerical X F2.Up-and-Comers			0.950			
6.Private X F2.Up-and-Comers			0.731			
7.Protect X F2.Up-and-Comers			0.526**			
8.Service X F2.Up-and-Comers			1.070			
9. FarmMgr X F2.Up-and-Comers			1.246			
10. FarmLbr X F2.Up-and-Comers			0.792			
11. Mech X F2.Up-and-Comers			0.872			
12. Construction X F2.Up-and-Comers			1.030			
13. Extract X F2.Up-and-Comers			1.687			
14. Precise X F2.Up-and-Comers			1.025			
15. Machine X F2.Up-and-Comers			0.954			
16. Transport X F2.Up-and-Comers			1.370			
17. Laborer X F2.Up-and-Comers			0.999			
1.Exec X F3.Heartlanders			1.478			
2.Prof X F3.Heartlanders			-------			
3.Tech X F3.Heartlanders			0.835			
4.Sales X F3.Heartlanders			1.167			
5. Clerical X F3.Heartlanders			1.263			
6.Private X F3.Heartlanders			6.570**			
7.Protect X F3.Heartlanders			4.041**			
8.Service X F3.Heartlanders			1.435			
9. FarmMgr X F3.Heartlanders			0.987			
10. FarmLbr X F3.Heartlanders			2.545*			
11. Mech X F3.Heartlanders			1.204			
12. Construction X F3.Heartlanders			1.311			
13. Extract X F3.Heartlanders			0.357			
14. Precise X F3.Heartlanders			1.904			
15. Machine X F3.Heartlanders			1.082			
16. Transport X F3.Heartlanders			0.719			
17. Laborer X F3.Heartlanders			3.357**			
1.Exec X F4.Empty Nest			0.965			
2.Prof X F4.Empty Nest			------			
3.Tech X F4.Empty Nest			0.902			

Hazard Ratios	Rural	SMSA Status	State Factors	Rural	SMSA Status	State Factors
	X Major Occ	X Major Occ	X Major Occ	X Occ Factors	X Occ Factors	X Occ Factors
OUTCOME	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause
4.Sales X F4.Empty Nest			1.168			
5. Clerical X F4.Empty Nest			0.966			
6.Private X F4.Empty Nest			0.855			
7.Protect X F4.Empty Nest			0.771			
8.Service X F4.Empty Nest			0.963			
9. FarmMgr X F4.Empty Nest			0.972			
10. FarmLbr X F4.Empty Nest			1.001			
11. Mech X F4.Empty Nest			0.975			
12. Construction X F4.Empty Nest			0.884			
13. Extract X F4.Empty Nest			0.974			
14. Precise X F4.Empty Nest			1.118			
15. Machine X F4.Empty Nest			1.140			
16. Transport X F4.Empty Nest			1.492**			
17. Laborer X F4.Empty Nest			1.253			
1.Reasoning X Rural				0.799**		
2.Physical X Rural				1.057		
3.People X Rural				1.360***		
4.Attention to Detail X Rural				1.307**		
1.Reasoning X SMSA: not City					0.760**	
2.Physical X SMSA: not City					0.983	
3.People X SMSA: not City					1.089	
4.Attention to Detail X SMSA: not City					1.198	
1.Reasoning X SMSA: not SMSA					0.777**	
2.Physical X SMSA: not SMSA					1.032	
3.People X SMSA: not SMSA					1.272*	
4.Attention to Detail X SMSA: not SMSA					1.463**	
1.Reasoning X F1.Big Fish						0.542***
2.Physical X F1.Big Fish						0.842
3.People X F1.Big Fish						0.758
4.Attention to Detail X F1.Big Fish						1.574*
1.Reasoning X F2.Up-and-Comers						1.359
2.Physical X F2. Up-and-Comers						1.384*
3.People X F2. Up-and-Comers						1.584*
4.Attention to Detail X F2.Up-and-Comers						0.681

Hazard Ratios	Rural	SMSA Status	State Factors	Rural	SMSA Status	State Factors
	X Major Occ	X Major Occ	X Major Occ	X Occ Factors	X Occ Factors	X Occ Factors
OUTCOME	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause
1.Reasoning X F3.Heartlanders						0.905
2.Physical X F3. Heartlanders						0.847
3.People X F3. Heartlanders						0.715
4.Attention to Detail X F3. Heartlanders						1.145
1.Reasoning X F4.Empty Nest						0.866
2.Physical X F4. Empty Nest						1.047
3.People X F4. Empty Nest						0.982
4.Attention to Detail X F4. Empty Nest						0.955
Observations	355,449	355,449	352,004	354,973	354,973	351,530

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

Big Fish, Small Pond states have high rankings in education, health, and low crime rates. Up-and-Comer states have high ranks in income, graduation percentage, the happiness index, and the freedom index. Heartlanders states have high manufacturing and farm output, and a high percentage of church attendance. Empty Nester states have a high percentage of people over 45 years of age and a low births per capita rank.

The first two columns show the results for the Rural/Urban and Major occupation interactions, and the SMSA and Major occupation interactions. In the Rural column, the Farmer occupation result changes from a health beneficial impact (0.78 in Table 35) to a health harmful impact (1.18 in Table 36) when the interaction terms are added. The interaction term indicates that rural living is much better for Farmers' health than urban living and the combination of the occupation term and interaction term (1.18 * 0.63 =) 0.75 is essentially equivalent to the original occupation term (0.78) in Table 35. In the SMSA column, construction work appears healthier outside of urban areas; service workers appear healthier in more rural areas; and private household work appears much less healthy outside of urban areas.

There are few significant interaction results between the state-level factors and the Major occupations except for the 1.Big Fish factor and 3.Heartlanders factor. The state factors shift upward (Factor 1) or downward (Factor 3) slightly from the results in Table 35, while the interaction terms demonstrate the opposite effect. The overall interaction effects for 1.Big Fish are positive health benefits for occupations living in the key Big Fish states (upper Midwest states and New England states, see Figure 15) relative to Professionals (Teachers). For 3.Heartlanders, the majority of the effects in this category are negative.

For the occupation factor results, the interactions with rural demonstrate positive health benefits for Reasoning & Complexity and negative health benefits for People vs. Things (PvT) and Attention to Detail (AtD). This pattern repeats for the SMSA category, *not SMSA*, which is most similar to the rural definition. For the "suburban" SMSA category only the Reasoning & Complexity (R&C) is significant. There are four interaction terms in the last column of Table 36 that are different from 1.0. The first suggests that it is beneficial to be in a Reasoning & Complexity occupation in the Big Fish states. As these states are those with the highest IQ rank and smartest rank, this is consistent. The second suggests that it is not beneficial to hold an Attention to Detail occupation in the Big Fish states. The last two suggest that it is not beneficial to be in Physically Demanding occupations or in people-centric occupations in the Up-and-Comer states.

An overall picture of rural people may be constructed from these results. That is, using these results, rural people are observed to be clever and incisive ((R&C < 1.0), prefer working with machines to dealing with people (PvT > 1.0), and tend to be generalists and not focused on precision, detailed oriented work (AtD > 1.0).

A final analysis was done, shown in Table 37 that co-regressed the occupation factors and the major occupation categories. These results demonstrate that the significance of the occupations themselves persist even with the factor categories included; and that the factor categories, including R&C, are for the most part no longer significant.

Table 37: Occupations and Occupation Factor Co-Regression

Hazard Ratios	Baseline	Ability Factors	Educ/Exp Factors	Knowledge Factors	Skills Factors	Work Activities Factors	Work Context Factors	Work Style Factors	Overall Factors
OUTCOME	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause
VARIABLES/base									
1.Executive/professional	1.062*	1.045	1.048	1.052	1.021	1.064	1.054	1.040	1.068*
2.Professional	------	------	------	------	------	------	------	------	------
3.Technician/professional	1.107*	1.146**	1.113*	1.143**	1.134**	1.093	1.110*	1.120*	1.115*
4.Sales/professional	1.089**	1.077*	1.074*	1.118**	1.092**	1.108**	1.089**	1.087**	1.076*
5.Clerical/professional	1.129***	1.119**	1.107**	1.142***	1.123***	1.128***	1.135***	1.122***	1.124***
6.Private Household/professional	1.169**	1.144	1.150	1.153	1.191	1.183	1.192	1.168	1.146*
7.Protective Services/professional	1.132**	1.095	1.123*	1.131*	1.143**	1.139**	1.120*	1.139**	1.110
8.Service-not protective/professional	1.238***	1.180***	1.218***	1.260***	1.263***	1.257***	1.220***	1.244***	1.196***
9.Farm Managers/professional	0.777***	0.774***	0.784***	0.803***	0.820***	0.774***	0.813***	0.793***	0.780***
10.Farm Workers/professional	0.981	0.967	0.968	1.019	1.024	0.985	0.997	0.997	0.960
11.Mechanics/professional	1.048	1.078	1.077	1.147**	1.129**	1.056	1.066	1.089*	1.036
12.Construction/professional	1.072	1.034	1.083	1.146**	1.113*	1.076	1.070	1.091*	1.048
13.Extractive/professional	1.103	1.164	1.162	1.224	1.221	1.150	1.126	1.178	1.074
14.Precision Production/professional	1.071	1.087	1.081	1.150**	1.121*	1.072	1.087	1.101*	1.065
15.Machine Operators/professional	1.079*	1.090	1.079	1.147***	1.132**	1.069	1.084	1.103**	1.067
16.Transportation/professional	1.183***	1.171**	1.149***	1.197***	1.205***	1.153**	1.165***	1.173***	1.148**
17.Handlers,Laborers/professional	1.217***	1.173**	1.189**	1.266***	1.242***	1.233***	1.210***	1.230***	1.177***
Ability: Gross Motor Skills, Strength, and Endurance		1.240**							
Ability: Cognitive Ability		0.959							
Ability: Fine Motor Abilities		0.862							

Hazard Ratios	Baseline	Ability Factors	Educ/Exp Factors	Knowledge Factors	Skills Factors	Work Activities Factors	Work Context Factors	Work Style Factors	Overall Factors
OUTCOME	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause	All-Cause
Ability: Auditory and Visual Processing		0.948							
Educ/Exp: Training			0.941						
Educ/Exp: Education & Experience			0.919						
Knowledge: Social Science				1.043					
Knowledge: Business				0.960					
Knowledge: Bio-Medicine				1.144*					
Knowledge: Engineering				0.846**					
Skills: Organizational					0.997				
Skills: Quantitative					1.047				
Skills: Technical					0.884				
Activities: Analyze & Decide						1.095			
Activities: Interact with Others						0.882			
Activities: Work with Things						1.036			
Context: Physically Challenging							0.986		
Context: Office							0.885		
Context: Socially Challenging							1.197**		
Context: Repetition							1.014		
Style: Leadership								0.924	
Style: Cooperation								1.078	
Overall: Reasoning & Complexity									0.967
Overall: Physical Demands									1.048
Overall: People versus Things									1.076
Overall: Attention to Detail									0.942
Observations	355,449	302,890	302,890	302,890	302,890	302,890	302,890	302,890	354,973

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

4.7 Conclusions

This chapter investigates, in more detail than previous studies, the concurrent impact of occupation and geographic factors on mortality and health in the United States. This study extends previous work by

- using the most recent, and more comprehensive, version of the NLMS
- considering the relationship of a variety of occupation category groupings and health
- using the results of an occupation factor analysis of a new occupation descriptor data set; and
- examining the coordinated effect of geography and the state-level factors, and occupations and occupation factors on health.

Standard occupational categories may be more accurate for white men than for women or non-whites because they were originally developed on the basis of a largely white and male labor force. In addition, occupation prestige rankings from the 1970s and 1980s may suffer from the same bias. Over time, prestige rankings tend to change only in small ways and reflect cultural shifts in how occupations are judged, e.g., lawyers going down in rank, while firefighters and police go up. The Duncan Socioeconomic Index is a legacy from the 1950s and 1960s as well. Perhaps as a result of this historic bias, the baseline regressions with occupation categories (shown in Table 23) demonstrate that the male occupation groupings (BRG male and Recode male) closely match the overall baseline results. For females, the results are often the same sign, but are smaller in magnitude and less significant. The figures in subsection 4.6.2 show similar results, i.e., the figures for males show larger impacts than the figures for women. Significant interaction effects between occupations and the urban/rural variable would show that the health impact by occupation varies depending on whether the occupation is an urban or rural occupation. Once again, there are few significant interactions. 9.Farmers is the only significant result suggesting better health in rural areas; however, 95% of farmers in the data set live in rural areas.

The occupation factor analysis provides previously unpublished insight into the impact of job characteristics on health outcomes. Table 33 summarizes relevant literature on job characteristics and maps the occupation factors in this study to previously reported characteristics. There are similarities, and interesting differences. As previously reported, creativity and cognitive abilities have no clear association with health. Table 33 and Table 34 show consistent, significant, and positive impact on health from the Ability: Cognitive Ability factor and the Overall: Reasoning & Complexity factor. The previous studies most often focus on one or a very narrow set of characteristics and gather results based on survey data. Each occupation factor in this study, however, is a composite of a much larger number of traits, and the factors emerged and are characterized by the traits that contribute most heavily toward the factor. For example, the Overall: Reasoning & Complexity factor has ~70 traits with a factor loading of 0.6 or higher; of these 70 traits, 31 relate clearly and directly to intelligence, perception, and cognition. Does this factor relate better to actual cognitive ability than the Salience in Creativity reported by Zhou? That correspondence is an interesting and fruitful follow-on opportunity to the work performed so far with these occupation factors. The fact that the Ovcrall: Reasoning & Complexity factor is consistently, significantly, and positively related to health suggests that it is a better measure. High-prestige occupations and occupational activity associated with more rank, power, and control are consistently shown to be healthier than occupations with more manual, less control, and less reasoning-based aspects. The empirical results of this study strongly support that outcome while Zhou's are mixed. If as several authors suggest, job characteristics rather than occupations themselves are better drivers of health, this study is a good starting point for further work. There may also be a relationship between the occupation factors and income or education. Although the coefficients on the income and education variables remain qualitatively the same when adding the occupation factors as additional explanatory variables, investigating the interaction of these baseline variables with the factors may determine whether income and education have independent effects or whether the factors affect health due to correlated impacts.

The interaction effects shown in Table 36 are disappointing. Interaction effects between occupations and state-level factors, and between occupation factors and state-level factors, are often not significant. The implication is that local cultural effects, which ought to be observable, must be attenuated when data is aggregated to the state level. This suggests the need to do future research at lower levels of geographic area granularity.

There are significant opportunities for future research using occupation factors and various economic, demographic, and lifestyle variables. The factors should be used with other data sets that include occupations to investigate a wide variety of relationships.

5. Conclusions

In this dissertation, I report results of empirical studies that address important research areas in health economics: health care determinants of key health outcomes, geographic impacts on health outcomes, and occupational impacts on health outcomes. These empirical studies examined aggregate determinants at the U.S. state level as well as individual determinants, and used factor analysis of occupation characteristics and state characteristics to create unique insights into the drivers of health.

There are many results presented in chapters 2, 3, and 4 (and there were many more analyses performed without the results being reported). This set of conclusions provides an integrated view of the work and addresses the key findings. The common threads through the chapters of this dissertation are the impact of determinants on overall health, the impact of determinants on causes of death, the impact of geography (both rural/urban distinctions and regional distinctions), and the impact of occupation related information. Chapter 2 is limited by the use of data at the state level. Some anomalous results may be due to the lack of detail at the state level and to potentially substantial intra-state heterogeneity in population health and socio-economic characteristics and in sub-state health care markets. Study of county-level detail or other more granular definitions would likely address some of these concerns and would be a valuable check on the state-level results. Further investigation into the possible structural break in the early 1990s may also provide some interesting insight.

The impact of determinants on overall health focuses on how spending affects all-cause mortality and cause-specific mortality. Despite hundreds of studies over the last 40 years, there is still uncertainty about the positive or negative impact of health expenditures on health. Studies have used panel data or cross-section data; some may have controlled for a wide variety of explanatory variables or a very few; some may have used instrumental variables (IV) to disentangle some aspects of causality or not; and some have performed cross-country studies, intra-country studies, and fine-grained studies of limited populations. Studies with limited explanatory variables are subject to omitted variable bias; studies with cross-country analyses struggle with differing health measures and meanings between countries; and studies not using IV, first differences, or fixed effects may have biased coefficients and/or standard errors.

An interesting result in this dissertation is that more health care spending has a negative effect on health. In chapter 2, results consistently show lower overall mortality with more spending. These results are resilient within bootstrap and jackknife simulations. Fisher and his colleagues at the Dartmouth Atlas project have argued that greater use of resources and greater expenditures per capita are inversely proportional to better outcomes or more satisfied patients. Fuchs has described “flat of the curve” medicine as those marginal applications of health care that provide little (or no) benefit with added cost. The data set used in chapter 2 is an aggregate accounting of health expenditures against outcomes, controlling for a variety of demographic, lifestyle, and economic variables. The data cover a broad spectrum of ages and demographics without a focus on a particular population, e.g., Medicare enrollees or those people within the last six months of life. Using a combination of state and year-fixed effects; controlling for age, education, and income; and employing validated instruments in two-stage least squares (2SLS) analyses results in negative impacts to health. The results are robust to sensitivity analyses, the use of other instruments, and a variety of statistical analytical methods. The results are consistently negative even with the trends of increasing income, increasing health expenditures, and improving health outcomes over the years of study.

Each chapter also looks at the impact of determinants on causes of death. The data sets in each chapter provide causes of death using the International Classification of Diseases, ninth revision (ICD-9) codes. These causes can be aggregated a variety of ways to consider different causes of death. In this dissertation, I chose to create four categories representing Tumor-related deaths, Cardiovascular-related deaths, Injury-related deaths, and Other deaths. The results in chapter 2, using aggregate data, indicate that greater medical spending has a beneficial effect on Cardiovascular-related mortality, but not for the remaining causes. Higher income per capita has a generally beneficial association with mortalities. Smoking and alcohol use impact health negatively across the board. The largest effect of alcohol use is on Injury-related deaths.

From the results in chapter 3, lower mortality is associated with being married, female, living in rural communities, and a race other than white or black. The education anomalies for high school education levels in the chapter 2 data may be affected by considering education as a black box, i.e., there is very little insight into the nature of education in this study and most others – for example, the quality of the schools, the diversity of the population, courses taken, or the degree of education funding – and that may impact longitudinal analyses of education. There are likely other omitted variables that affect these education category results and would illuminate the rationale for these results. Psychosocial effects on attitudes of high school students, similar to those discussed in chapter 4 with respect to occupations, would be a fruitful research topic. For example, is there a sense of invincibility, are there any effects of inexperience while driving, and what results can be attributed to the use of drugs might be useful avenues to pursue. By the time many young people graduate from college, attitudes may be reoriented toward career, their future (possibly marriage and family) and thus, other socio-economic status influences overcome the negative effect seen in chapter 2 for high school education levels.

Chapter 4 adds occupation to the mix. Tables 27 and 28 show which causes of death are more likely to be impacted by which occupations. Sales, construction, transportation, and laborers have a stronger association with Tumor-related deaths while private household, service occupations, and farmers are more strongly associated with cardiovascular-related mortality. For farmers, it is a positive impact on health. Deaths from injuries are most strongly related to construction, extractive, and laborer occupations. Other causes of death are impacted by clerical and service occupations. Adding the occupations affects the rural results most dramatically for the Injury mortality. This can be explained by looking back at Table 22, which shows that most of the risky occupations in the Injury category are the occupations with the highest percentage of rural people; for example, farmers and farm workers have the highest rural percentage and the Injury category is the only cause of death category in which farmers and farm workers have a negative impact on health.

Tables 29 and 30 in chapter 4 address causes of death and add age groupings to the analyses. The age groupings provide interesting insight into which age ranges and occupations, taken together, affect health. Consider one occupation as an example of what these results reveal. In Tables 27 and 28, being a farmer is associated with better health than being a Professional (Teacher) for all causes of death except Injury. In Table 29, farmers demonstrate lower Tumor-related deaths at all ages. For cardiovascular-related and other-related mortalities, farmers under the age of 65 are associated with lower mortality, and farmers over the age of 65 have a higher mortality. Perhaps farmers keep working beyond the age of 65 while professionals retire, so this is an indication of the impact of manual work related health effects on older people, or perhaps it is indicative of the cumulative impact of a more manual occupation over a lifetime compared to a professional career. For Injury-related mortalities, younger (presumably more healthy) and older (presumably retired) farmers have a much lower risk, while the middle-aged farmer has a much higher risk of injury-related death than professionals. Further research should consider a smaller, more discrete, range of ages; a more detailed occupation listing; and, perhaps, finer-grained mortality groupings to better elucidate whether the intriguing age effects persist.

The geographic effects in this dissertation are limited by the granularity of the data available. State effects are the primary geographic level. A large literature has emerged over the last 15 years on the impact of "neighborhood" effects on health. Small areas, neighborhoods, and census tracts, have effects even with demographic and lifestyle explanatory variables present. The use of state variables in these studies likely aggregates the data to such an extent that many interesting relationships are hidden. In addition, the use of dummy variables to identify geographic regions does not reveal anything about the nature of the regional differences that may impact health. It is merely an approach to determine if mortality varies across areas with broad geographic similarities (perhaps latitude, longitude, average temperature, or height above sea level) or gross cultural characteristics (perhaps attitudes, rural versus urban, farming versus manufacturing, or seaside versus mountainside). Given those constraints, the results show that people in southern states and regions have consistently worse health than western and northeastern states and regions, and that people in rural areas have better health than urban areas. Interaction effects can best be seen in Table 17 in chapter 3 with the Standard Metropolitan Statistical Areas (SMSA) and Census Divisions. The results are relative to the base of *SMSA City Central* (meaning urban) areas. All results indicate that living anywhere other than urban areas has lower health risk, and living in the *Not SMSA* category (meaning mostly rural) has a consistently significant and positive impact on health. These effects are stronger than interactions with the Rural variable itself. Interactions with demographic variables in Table 18 do not reveal many significant results.

The multi-level analyses in chapter 3 provide a key result. Figure 5 indicates that living in a poorer state has a larger impact on health than a richer state, and that there is a larger difference between being poor in a rich state than being rich in a poor state. The implication is that individual income inequality is not as important as state income inequality.

The most interesting results emerge from chapter 4 when considering the impact of occupations and the occupation factors on health. Occupations are seldom used in U.S. studies of health determinants, the preference going to either income and education or to socio-economic status indices such as the Duncan Socioeconomic Index, Nam-Powers Occupation Status Score, Nam-Powers Socioeconomic Score, or the Hollingshead Index of Social Position.

The references in chapter 4 that do include occupation as an explanatory variable frequently identify the need for better measures for illuminating the impact of occupations on health and point to more detailed insight into the nature of work, the social characteristics of the job, or environmental exposures experienced as likely candidates. The domain occupation factor analysis, performed for this chapter, captures these kinds of insights in 22 factors by considering the abilities, education and experience, knowledge, skills, work activities, work context, and work style traits that characterize occupations. The overall factor analysis groups these traits into four factors that contain recognizably related characteristics. Factor 1 Reasoning & Complexity (R&C) represents cognitive ability, critical thinking, and innovation; Factor 2 Physical Demands (PD) represents the impact of physical demands of work; Factor 3 People vs. Things (PvT) contrasts interpersonal work context and activity with skills and activities related to working with mechanical equipment; and Factor 4 Attention to Detail (AtD) focuses on precision work and eye-hand coordination. In co-regressions, using the 22 domain factors and the four overall factors, Factor 1 consistently, significantly, and positively impacts health. That is, even with the other factors present, Reasoning & Complexity remains significant while the others are typically attenuated and lose significance. The conclusion from these analyses is that the R&C factor has the most persistent relationship with health outcomes, even in the face of a variety of confounding variables. Perhaps it is a better indicator of what characteristic of occupations helps drive health. A final analysis was done, as shown in Table 37 that co-regressed the occupation factors and the major occupation categories. These results demonstrate that the significance of the occupations themselves persist even with the factor categories included, and that the factor categories, including R&C, are for the most part no longer significant. This analysis supports an argument that refines the foregoing conclusion to say that even with a large variety of confounding variables, including occupation factors, there are still persistent characteristics of the occupations themselves that maintain their significant relationship with health. Clearly, other yet unidentified variables are at work here.

There are additional opportunities for exploring these occupation factors based on published work in psychology. The papers listed in Table 33 demonstrate efforts to relate specific job-related social characteristics to health. Most of these papers focus on a single psycho-social characteristic. As noted, some of these papers, such as those of Marmot, conclude that since British civil employees have a high degree of job stability and universal health insurance coverage, the remaining qualities that are in play relate to lack of job control or influence, or the impact of stress (M. Marmot et al. 1991; Michael Marmot et al. 1997; Bosma, Stansfield, and Michael Marmot 1998). Smith pointed out limitations, however, in Marmot's use of this employee cohort (J Smith 1999). Two other recent papers (Crum and Langer 2007; Hsu, Chung, and Langer 2010) investigate the role of perception and mind-set on job characteristics that affect health. Crum and Langer demonstrate that perception of a physical job as good exercise providing benefits toward an active lifestyle positively affects physical health-related characteristics (e.g., blood pressure and body mass index) relative to a control group. One study reported by Hsu et al. related the wearing of work uniforms to mortality. They report that low-income workers that wear uniforms exhibit poorer health than workers that do not wear uniforms. When worker incomes rise above a certain level (~$24,000 per year), the results are reversed—workers wearing uniforms had better health. Hsu et al. relate these results to job control and age-related cues, i.e., low-income workers see wearing uniforms as lack of job control while higher income workers may see uniforms as a buffer for being aware of one's age.

There is a definable relationship between the occupation factors and the psychosocial characteristics shown in Table 33. There is general support for the prior results. In four specific cases discussed in chapter 4, the occupation factors appear to illuminate the impact of social characteristics better than the previously published focused studies, perhaps because the extent of the occupational traits in the O*NET database allows the concurrent analyses of multiple psychosocial characteristics. Future related research should consider how to use the occupation factors identified here for investigation of more direct impacts of social variables on health, or interaction effects with the wearing of a uniform on the job, or the relationship between occupation factors and occupations that require licenses to perform the job. The impact across states that do not require licenses might demonstrate different occupation factor effects than those that do require licenses.

These investigations address interesting elements in the continuing debate over determinants of health; it is my hope that these results will contribute to a deeper understanding of the debate and inspire further research on health determinants.

Appendix A. Descriptions of Chapter 2 Variables

Table 38 provides definitions and source information for the dependent variables used in chapter 2. Other sets of dependent variables were collected from Federal and State sources; for example, Life Expectancy at age 65, Life Expectancy by Race and Gender, and Infant Mortality by Race. These data were not sufficiently complete across the years of this study to allow for use in the analyses.

Table 38: Definitions of the Dependent Variables

Dependent Variable	Definition	Source
Male life expectancy	Life expectancy at birth	U.S. Census Bureau and State Health Departments
Female life expectancy	Life expectancy at birth	U.S. Census Bureau and State Health Departments
Infant mortality	Mortality rates before 1-year-old per thousand live births	National Center for Health Statistics (NCHS), Centers for Disease Control (CDC); and State Health Departments
All cause mortality	Age, race-adjusted all cause death rate	NCHS, CDC
Cause 1	Death rate by tumors	NCHS, CDC
Cause 2	Death rate by cardiovascular disease	NCHS, CDC
Cause 3	Death rate by injury	NCHS, CDC
Cause 4	Death rate by all other causes	NCHS, CDC
Teen Birth Percentage	Teen births (age < 20) as a percentage of total births	NCHS, CDC
Low Birth Weight Percentage	Low birth weight (2500 g or less) as a percentage of total births	NCHS, CDC

Table 39 provides definitions and source information for the explanatory variables used in chapter 2. Other sets of explanatory variables were collected from Federal and State sources; for example, Medicare expenses per capita, Medicaid expenses per capita, and percentage of the population with private insurance. These data were not sufficiently complete across the years of this study to allow for inclusion in the analyses.

Table 40 provided definitions for the CMS detailed health care expenditure variables.

Table 39: Definitions of the Explanatory Variables

Explanatory Variable	Definition	Source
Alcohol use per capita	Per capita consumption of alcoholic beverages	National Institute on Alcohol Abuse and Alcoholism
Beds	Number of hospital beds per 10,000 population	National Center for Health Statistics (NCHS), Centers for Disease Control (CDC)
Black %	Percentage of blacks in the state population	U.S. Census Bureau
Density	State population/geographical area	U.S. Census Bureau
Education Level – College	State-specific proportion of the population with a Bachelors degree or higher	U.S. Census Bureau
Education Level – High School	State-specific proportion of the population with a high school degree	U.S. Census Bureau
Female %	Percentage of females in the state population	U.S. Census Bureau
GSP per capita	Deflated Gross State Product per capita	Centers for Medicare and Medicaid Services (CMS)
Health care expenditures per capita	State-specific sum of all private and public personal health care spending per capita	CMS
Income per capita	Deflated income per capita	The Tax Foundation
Non-White %	Percentage of non-whites in the state population	U.S. Census Bureau
Physicians per capita	State-specific number of civilian physicians per 10,000 population	Vital Statistics of the U.S., 2006, CDC
Population 25-44	Percentage of the state population that is 25-44 years of age	U.S. Census Bureau
Population 45-64	Percentage of the state population that is 45-64 years of age	U.S. Census Bureau
Population 65+	Percentage of the state population that is 65 years or older	U.S. Census Bureau
Poverty	Percentage of households at or below the Federal poverty level	U.S. Census Bureau
Tobacco use	Percentage of individuals that smoke cigarettes	Behavioral Risk Factor Surveillance System (BRFSS)
Unemployment	Percentage of unemployed individuals	U.S. Department of Labor, Bureau of Labor Statistics
Urban %	Percentage of the state considered to be urban	U.S. Census Bureau
White %	Percentage of whites in the state population	U.S. Census Bureau
Coincident index	State summary index of economic conditions	Philadelphia Federal Reserve
Gini index	State measure of income inequality	U.S. Census Bureau
Tax Rank	Rank of state based on tax burden	The Tax Foundation

Explanatory Variable	Definition	Source
	(1==highest)	
Black House	Percentage of State House members that are black	National Conference of State Legislators
Black Legislators	Percentage of State legislative members that are black	National Conference of State Legislators
Black Senators	Percentage of State Senate members that are black	National Conference of State Legislators
Democratic House	State House has a Democrat majority (==1)	National Conference of State Legislators
Democratic House advantage	Democratic vote advantage in the State House	National Conference of State Legislators
Democratic Senate	State Senate has a Democrat majority (==1)	National Conference of State Legislators
Democratic Senate advantage	Democratic vote advantage in the State Senate	National Conference of State Legislators
Red State	State voted Republican in last National election	U.S. Census Bureau
Registered Black	Percentage of eligible blacks registered to vote	U.S. Census Bureau
Registered Men	Percentage of eligible men registered to vote	U.S. Census Bureau
Registered White	Percentage of eligible whites registered to vote	U.S. Census Bureau
Registered Women	Percentage of eligible women registered to vote	U.S. Census Bureau
Voted	Percentage of eligible voters voting	U.S. Census Bureau
Voted Black	Percentage of eligible black voters voting	U.S. Census Bureau
Voted Men	Percentage of eligible male voters voting	U.S. Census Bureau
Voted White	Percentage of eligible white voters voting	U.S. Census Bureau
Voted Women	Percentage of eligible female voters voting	U.S. Census Bureau
Woman Governor	Governor of State is a woman (==1)	National Conference of State Legislators
Women House	Percentage of State House members that are women	National Conference of State Legislators
Women Legislators	Percentage of State legislative members that are women	National Conference of State Legislators
Women Senators	Percentage of State Senate members that are women	National Conference of State Legislators
Medicaid expenditure %	State Medicaid expenditures as a percentage of total HC expenditures	CMS
Medicaid expenditure per capita	State Medicaid expenditures per capita	CMS
Medicaid %	Percentage of the state population registered for Medicaid	CMS

Explanatory Variable	Definition	Source
Medicare expenditure %	State Medicare expenditures as a percentage of total HC expenditures	CMS
Medicare expenditure per capita	State Medicare expenditures per capita	CMS
Medicare %	Percentage of the state population registered for Medicare	CMS
Private	Percentage of the state population that has private health insurance	U.S. Census Bureau
Private expenditures	State Private expenditures per capita	CMS
Private %	State Private expenditures as a percentage of total HC expenditures	CMS
Public expenditures	State Public expenditures per capita	CMS
Public %	State Public expenditures as a percentage of total HC expenditures	CMS
Uninsured %	Percentage of the state population that is uninsured	U.S. Census Bureau

Table 40: CMS Detailed Expenditure Categories

CMS Detailed Expenditure Categories	Definition	Source
Dental Services	Services provided by dentists, dental surgeons, and dental technicians	CMS
Durable Medical Products	Items such as contact lenses, eyeglasses, orthopedic production, bottled oxygen, wheelchairs, and hearing aids	CMS
Home Health Care	Skilled nursing care in the home; companion services, medical social services, medical equipment and supplies, counseling, vocational therapy, dietary services, speech therapy	CMS
Hospital Care	Services provided by hospitals for patients, e.g., room and board, operating room fees, inpatient pharmacy, resident doctor fees	CMS
Nursing Home Care	Nursing and rehabilitative services provided in freestanding nursing homes	CMS
Other Non-Durable Medical Products	Retail items such as rubber medical sundries, heating pads, bandages, non-prescription drugs and analgesics	CMS

CMS Detailed Expenditure Categories	Definition	Source
Other Personal Health Care	Industrial in-plant services; medical care delivered in community centers, senior citizen centers, schools, and military field stations	CMS
Other Professional Services	Services provided by other health practitioners, e.g., chiropractors, podiatrists, optometrists, and physical, occupational, and speech therapists	CMS
Physician and Clinical Services	Services performed by Doctors of Medicine, Doctors of Osteopathy, and medical laboratories	CMS
Prescription Drugs	Retail sales of prescription drugs, in pharmacies, via mail-order, and in mass-merchandising establishments	CMS

Appendix B. Descriptions of Chapter 3 Variables

Table 41 provides definition and source information for the dependent variables used in the NLMS analyses described in chapter 3.

Table 41: Chapter 3 Dependent and Explanatory Variables

Variable	Description	Source	# of Obs.	Mean	Std. Dev	Min	Max
Health Status Variables							
DeathIndictor	Death Indicator	NLMS	988346	0.091	0.288	0	1
Cause1	Cause of Death 1, ICD-9	NLMS	89904	393.06	189.8	3	988
MajCause	Major Cause of Death	constructed	89904	6.436	3.828	1	17
GenCause	General category of cause of death	constructed	89904	2.801	1.613	1	5
Tumors_cause1	GenCause== 1.0000	constructed	89904	0.250	0.433	0	1
Cardio_cause2	GenCause== 2.0000	constructed	89904	0.367	0.482	0	1
Injury_cause3	GenCause== 3.0000	constructed	89904	0.062	0.242	0	1
Other_cause4	GenCause== 4.0000	constructed	89904	0.303	0.460	0	1
Demographic Variables							
AdjIncome	Adjusted Income	NLMS	947480	8.276	3.617	1	14
Age	For those aged 25-65 only	NLMS	476300	41.63	11.8	25	65
Age_cen	Age - mean-centered, for those aged 25-65 only	constructed	367180	0.000	11.035	-15.3	24.7
Age25_44	Age dummy for ages 25-44	constructed	988346	0.305	0.461	0	1
Age45_64	Age dummy for ages 45-64	constructed	988346	0.186	0.389	0	1
Age65_more	Age dummy for ages 65+	constructed	988346	0.095	0.293	0	1
New_age	Enumerated age variable	constructed	579566	1.641	0.744	1	3
EducCompleted	Education Completed	NLMS	769879	7.973	2.802	1	14
Female	Sex-updated	constructed	988346	0.519	0.500	0	1
Race	For those aged 25-65 only	NLMS	474445	1.15	0.48	1	5
Hispanic	Hispanic	NLMS	956401	7.635	1.399	1	8
Housing	Housing Tenure	NLMS	963659	1.318	0.506	1	3
Insurance	Insurance	NLMS	447755	0.814	0.389	0	1
InsuranceType	Insurance Type	NLMS	447755	2.937	1.730	0	5
Married	Marital status-updated	constructed	866865	0.482	0.500	0	1
NonWhite	White/Non-white race	constructed	985282	0.131	0.338	0	1
NumberInHouse	Number in Household	NLMS	988346	3.652	1.786	1	24
Poverty Percent	Poverty Level	constructed	988346	0.887	0.782	0	2

Variable	Description	Source	# of Obs.	Mean	Std. Dev	Min	Max
Race3	Race-updated	constructed	985282	0.161	0.440	0	2
Veteran	Veteran Status	NLMS	739891	0.159	0.366	0	1
WorkerClass	Class of Worker	NLMS	518190	1.433	0.791	1	5
Working	Employment Status-updated	constructed	760608	2.190	1.411	1	4
Geographic Variables							
BirthDivision	Census Birth Division	constructed	452031	4.732	2.4261	1	9
BirthRegion	Census Birth Region	constructed	452031	2.459	1.0134	1	4
PlaceOfBirth	State of Birth	NLMS	482454	28.77	15.66	1	59
ResidenceDivision	Census Residence Division	constructed	988346	5.015	2.5758	1	9
ResidenceRegion	Census Residence Region	constructed	988346	2.583	1.0677	1	4
Rural	Urban/Rural-updated 0=Urb, 1-Rural	NLMS	975299	0.333	0.4714	0	1
SMSAStatus	SMSA	NLMS	975285	2.109	0.7913	1	3
StateResidence	State of Residence	NLMS	988346	53.65	26.11	11	95
Unused NLMS Variables							
Cause2	Cause of Death 2, ICD-9	NLMS	63168	4.753	3.993	0	9
Certified	Certifed by Coroner	NLMS	59191	1.217	0.427	1	3
DayOfWeek	Day of Death	NLMS	89909	4.013	2.007	1	7
EmployStatus	Employment Status	NLMS	760608	2.530	1.855	1	5
FollowUp	Survey Follow-up	NLMS	988346	3843	651.0	1	4017
Hospital	Place of Death - Hospital Related	NLMS	76614	3.139	1.468	1	5
Hour	Hour of Death	NLMS	59191	2.999	1.356	1	5
Interval	Interval Btw Onset and Death	NLMS	33342	1.592	1.380	0	4
MaritalStatus	Marital Status	NLMS	866865	2.634	1.851	1	5
PlaceOfDeath	Place of Death	NLMS	89242	1.847	1.287	1	6
Race	Race	NLMS	985282	1.182	0.541	1	5
RelHHouse	Relationship to Head of Household	NLMS	973673	2.535	1.402	1	6
Sex	Sex	NLMS	988346	1.519	0.500	1	2
SSN	SSN Present	NLMS	988346	0.673	0.469	0	1

Appendix C. Listing of Occupations by Defined Groups

British Registrar General (BRG) occupation categories: 4 categories

1. Professional
2. Clerical
3. Skilled Crafts
4. Labor

Major Occupation categories: 18 categories

1. Executive	10. Farm Workers
2. Professional	11. Mechanics
3. Technician	12. Construction
4. Sales	13. Extractive
5. Clerical	14. Precision Production
6. Private Household	15. Machine Operators
7. Protective Services	16. Transportation
8. Service-not protective	17. Handlers, Laborers
9. Farm Managers	18. Military

Occupation Recode categories for females: 59 categories

1. Accountants	31. Receptionists
2. Computer Specialists	32. Secretaries
3. Librarians	33. Stenographers
4. Mathematicians	34. Telephone Operators
5. Life, Physical Scientists	35. Typists
6. Nurses, Therapists	36. Other Clerical
7. Health Technicians	37. Foremen
8. Social Scientists	38. Other Craftsmen
9. Social Workers	39. Assemblers
10. Teachers	40. Bottling Operatives
11. Technicians	41. Examiners, Inspectors
12. Writers, Entertainers	42. Seamstresses
13. Other Professional	43. Laundry Operatives

14. Buyers, Sales Managers	44. Graders, Sorters
15. Restaurant Managers	45. Packers, Wrappers
16. School Administrators	46. Sewers, Stitchers
17. Other Managers	47. Textile Operatives
18. Peddlers	48. Other Operatives
19. Insurance brokers	49. Transport Operatives
20. Sales Clerks	50. Laborers-not farm
21. Salesmen	51. Farmers
22. Other Sales Workers	52. Farm laborers
23. Bank Tellers	53. Cleaning Workers
24. Bookkeepers	54. Cooks
25. Cashiers	55. Waitresses
26. Counter Clerks	56. Health Service Workers
27. Interviewers	57. Cosmetologists
28. File Clerks	58. Other Personal Service
29. Office Machine Operators	59. Private Household
30. Payroll Clerks	

Occupation Recode categories for males: 88 categories

1.Accountants	45.Linemen-Power
2.Architects	46.Locomotive Engineers
3.Computer Specialists	47.Auto Mechanics
4.Engineers	48.Other Mechanics
5.Lawyers,Judges	49.Machinists
6.Chemists	50.Sheetmetal Workers
7.Life,Physical Scientists	51.Tool&Die Makers
8.Dentists	52.Other Metal Craftsmen
9.Pharmacists	53.Printing Craftsmen
10.Physicians	54.Power Station Operators
11.Other Health Practitioners	55.Other Craftsmen
12.Heath Technicians	56.Assemblers
13.Religious Workers	57.Examiners,Inspectors
14.Social Scientists	58.Gas Station Attendants
15.Social Workers	59.Laundry Operatives
16.Teachers	60.Butchers
17.Engineering Technicians	61.Mine Operators

18.Writers,Entertainers	62.Packers,Wrappers
19.Other Professionals	63.Painters
20.Buyers,Sales Managers	64.Precision Machine Operatives
21.School Administrators	65.Sawyers
22.Public Administrators	66.Firemen
23.Other Managers	67.Textile Operatives
24.Managers,Administrators	68.Welders
25.Insurance Brokers	69.Other Metal Operatives
26.Real Estate Brokers	70.Other Specified Operatives
27.Other Sales Workers	71.Other Operatives
28.Salesmen	72.Bus Drivers
29. Bank Tellers, Cashiers	73.Taxicab Drivers
30.Bookkeepers	74.Truck Drivers
31.Postal Clerks	75.Other Transport Operatives
32.Other Clerical	76.Construction Laborers
33.Upholsterers	77.Freight Handlers
34.Bakers	78.Other Specified Laborers
35.Cabinetmakers	79.Other Laborers
36.Carpenters	80.Farmers
37.Road Machine Operatives	81.Farm Laborers
38.Electricians	82.Cleaning Service Workers
39.Masons	83.Food Service Workers
40.Painters,Paperhangers	84.Health Service Workers
41.Plasterers	85.Personal Service Workers
42.Plumbers	86.Protective Service Workers
43.Other Construction	87.Other Service Workers
44.Foremen	88.Private Household Workers

Occupation categories: total of 807 occupations. This information is available by request from the author.

Appendix D. Factor Analyses of Occupation Characteristics

Background

Factor analysis is a statistical approach used to analyze interrelationships among a large number of variables and to explain these variables in terms of their common underlying dimensions or factors. Variables that are correlated with one another, and which are largely independent of other subsets of variables, are combined into factors. The approach is basically a data reduction technique to condense the information contained in the original variables into a smaller set suitable for use in further analyses.

The purpose of this factor analyses is to explore the structures of the O*NET descriptor domains, specifically Abilities, Education and Experience, Knowledge, Skills, Work Activities, Work Context, and Work Styles. Of the ten possible domains, these seven were determined to best match the analyses in this thesis[6]. Brief descriptions of the domains are shown in Table 42.

Table 42: Descriptions of O*NET Domains Used

Domain	Description
Abilities	Enduring attributes of the individual that influence performance
Education and Experience	Prior educational experience or training required to perform in a job
Knowledge	Organized sets of principles and facts applying in general domains
Skills	Developed capacities that facilitate learning, the more rapid acquisition of knowledge, or performance of activities that occur across jobs
Work Activities	General types of job behaviors occurring on multiple jobs
Work Context	Physical and social factors that influence the nature of work
Work Styles	Personal characteristics that can affect how well someone performs a job

Abilities include 52 different abilities in 4 broad categories (Cognitive, Psychomotor, Physical, and Sensory). Examples are *Oral Comprehension* and *Visualization*.

Education and Experience includes 4 areas: *Related Work Experience, Required Level of Education, On-the-Job-Training*, and *On-Site Training*.

Knowledge includes 33 different knowledge areas. Examples are *Design* and *Building and Construction*.

[6] The domains not used are: Interests, Job Zones, and Work Values.

Skills include 10 basic skills, such as *Reading Comprehension*, and 25 cross-functional skills, such as *Complex Problem Solving*.

Work Activities include 41 activity descriptors. Examples are *Processing Information* and *Staffing Organizational Units*.

Work Context includes 57 descriptors. Examples are *Degree of Automation* and *Exposed to Hazardous Conditions*.

Work Styles include 16 different style descriptors. Examples are *Cooperation* and *Attention to Detail*.

Previous Factor Analyses of Occupations

O*NET is a revision of the Dictionary of Occupational Titles (DOT) by the Department of Labor (DOL) and the Bureau of Labor Statistics. The DOT had grown to over 12,000 entries; the O*NET has a much reduced set (~950, not all titles have complete data sets) with more broadly defined titles. Factor analyses of DOT has been done repeatedly (Hadden, Kravets, and Muntaner 2004). Previous versions of O*NET have had factor analyses completed on some of the domains: version 1.0 (Department of Employment and Economic Development 1999), version 4.0 (Hadden, Kravets, and Muntaner 2004), and version 6.0 (T Smith and Campbell 2006; Crouter et al. 2006).

Factor Analysis Approach for O*NET Domains

This analysis uses version 13.0 of the O*NET, downloaded from the DOL website on February 8, 2008.   The following process description uses the Abilities domain as an example. Each of the other six key domains was handled similarly.

The Abilities domain has 52 variables that measure cognitive abilities, psychomotor abilities, physical abilities, and sensory abilities. In Table 43, the variables are in the lightly shaded rows, with descriptive categories in bold text.

Table 43: Abilities Domain Variables

1.A	Abilities	Enduring attributes of the individual that influence performance
1.A.1	**Cognitive Abilities**	**Abilities that influence the acquisition and application of knowledge in problem solving**
1.A.1.a	**Verbal Abilities**	**Abilities that influence the acquisition and application of verbal information in problem solving**
1.A.1.a.1	Oral Comprehension	The ability to listen to and understand information and ideas presented through spoken words and sentences.
1.A.1.a.2	Written Comprehension	The ability to read and understand information and ideas presented in writing.
1.A.1.a.3	Oral Expression	The ability to communicate information and ideas in speaking so others will understand.

1.A	Abilities	Enduring attributes of the individual that influence performance
1.A.1.a.4	Written Expression	The ability to communicate information and ideas in writing so others will understand.
1.A.1.b	**Idea Generation and Reasoning Abilities**	**Abilities that influence the application and manipulation of information in problem solving**
1.A.1.b.1	Fluency of Ideas	The ability to come up with a number of ideas about a topic (the number of ideas is important, not their quality, correctness, or creativity).
1.A.1.b.2	Originality	The ability to come up with unusual or clever ideas about a given topic or situation, or to develop creative ways to solve a problem.
1.A.1.b.3	Problem Sensitivity	The ability to tell when something is wrong or is likely to go wrong. It does not involve solving the problem, only recognizing there is a problem.
1.A.1.b.4	Deductive Reasoning	The ability to apply general rules to specific problems to produce answers that make sense.
1.A.1.b.5	Inductive Reasoning	The ability to combine pieces of information to form general rules or conclusions (includes finding a relationship among seemingly unrelated events).
1.A.1.b.6	Information Ordering	The ability to arrange things or actions in a certain order or pattern according to a specific rule or set of rules (e.g., patterns of numbers, letters, words, pictures, mathematical operations).
1.A.1.b.7	Category Flexibility	The ability to generate or use different sets of rules for combining or grouping things in different ways.
1.A.1.c	**Quantitative Abilities**	**Abilities that influence the solution of problems involving mathematical relationships**
1.A.1.c.1	Mathematical Reasoning	The ability to choose the right mathematical methods or formulas to solve a problem.
1.A.1.c.2	Number Facility	The ability to add, subtract, multiply, or divide quickly and correctly.
1.A.1.d	**Memory**	**Abilities related to the recall of available information**
1.A.1.d.1	Memorization	The ability to remember information such as words, numbers, pictures, and procedures.
1.A.1.e	**Perceptual Abilities**	**Abilities related to the acquisition and organization of visual information**
1.A.1.e.1	Speed of Closure	The ability to quickly make sense of, combine, and organize information into meaningful patterns.
1.A.1.e.2	Flexibility of Closure	The ability to identify or detect a known pattern (a figure, object, word, or sound) that is hidden in other distracting material.
1.A.1.e.3	Perceptual Speed	The ability to quickly and accurately compare similarities and differences among sets of letters, numbers, objects, pictures, or patterns. The things to be compared may be presented at the same time or one after the other. This ability also includes comparing a presented object with a remembered object.
1.A.1.f	**Spatial Abilities**	**Abilities related to the manipulation and organization of spatial information**
1.A.1.f.1	Spatial Orientation	The ability to know your location in relation to the environment or to know where other objects are in relation to you.

1.A	Abilities	Enduring attributes of the individual that influence performance
1.A.1.f.2	Visualization	The ability to imagine how something will look after it is moved around or when its parts are moved or rearranged.
1.A.1.g	**Attentiveness**	**Abilities related to application of attention**
1.A.1.g.1	Selective Attention	The ability to concentrate on a task over a period of time without being distracted.
1.A.1.g.2	Time Sharing	The ability to shift back and forth between two or more activities or sources of information (such as speech, sounds, touch, or other sources).
1.A.2	**Psychomotor Abilities**	**Abilities that influence the capacity to manipulate and control objects**
1.A.2.a	**Fine Manipulative Abilities**	**Abilities related to the manipulation of objects**
1.A.2.a.1	Arm-Hand Steadiness	The ability to keep your hand and arm steady while moving your arm or while holding your arm and hand in one position.
1.A.2.a.2	Manual Dexterity	The ability to quickly move your hand, your hand together with your arm, or your two hands to grasp, manipulate, or assemble objects.
1.A.2.a.3	Finger Dexterity	The ability to make precisely coordinated movements of the fingers of one or both hands to grasp, manipulate, or assemble very small objects.
1.A.2.b	**Control Movement Abilities**	**Abilities related to the control and manipulation of objects in time and space**
1.A.2.b.1	Control Precision	The ability to quickly and repeatedly adjust the controls of a machine or a vehicle to exact positions.
1.A.2.b.2	Multi-limb Coordination	The ability to coordinate two or more limbs (for example, two arms, two legs, or one leg and one arm) while sitting, standing, or lying down. It does not involve performing the activities while the whole body is in motion.
1.A.2.b.3	Response Orientation	The ability to choose quickly between two or more movements in response to two or more different signals (lights, sounds, pictures). It includes the speed with which the correct response is started with the hand, foot, or other body part.
1.A.2.b.4	Rate Control	The ability to time your movements or the movement of a piece of equipment in anticipation of changes in the speed and/or direction of a moving object or scene.
1.A.2.c	**Reaction Time and Speed Abilities**	**Abilities related to speed of manipulation of objects**
1.A.2.c.1	Reaction Time	The ability to quickly respond (with the hand, finger, or foot) to a signal (sound, light, picture) when it appears.
1.A.2.c.2	Wrist-Finger Speed	The ability to make fast, simple, repeated movements of the fingers, hands, and wrists.
1.A.2.c.3	Speed of Limb Movement	The ability to quickly move the arms and legs.
1.A.3	**Physical Abilities**	**Abilities that influence strength, endurance, flexibility, balance and coordination**
1.A.3.a	**Physical Strength Abilities**	**Abilities related to the capacity to exert force**

1.A	Abilities	Enduring attributes of the individual that influence performance
1.A.3.a.1	Static Strength	The ability to exert maximum muscle force to lift, push, pull, or carry objects.
1.A.3.a.2	Explosive Strength	The ability to use short bursts of muscle force to propel oneself (as in jumping or sprinting), or to throw an object.
1.A.3.a.3	Dynamic Strength	The ability to exert muscle force repeatedly or continuously over time. This involves muscular endurance and resistance to muscle fatigue.
1.A.3.a.4	Trunk Strength	The ability to use your abdominal and lower back muscles to support part of the body repeatedly or continuously over time without 'giving out' or fatiguing.
1.A.3.b	**Endurance**	**The ability to exert oneself physically over long periods without getting out of breath**
1.A.3.b.1	Stamina	The ability to exert yourself physically over long periods of time without getting winded or out of breath.
1.A.3.c	**Flexibility, Balance, and Coordination**	**Abilities related to the control of gross body movements**
1.A.3.c.1	Extent Flexibility	The ability to bend, stretch, twist, or reach with your body, arms, and/or legs.
1.A.3.c.2	Dynamic Flexibility	The ability to quickly and repeatedly bend, stretch, twist, or reach out with your body, arms, and/or legs.
1.A.3.c.3	Gross Body Coordination	The ability to coordinate the movement of your arms, legs, and torso together when the whole body is in motion.
1.A.3.c.4	Gross Body Equilibrium	The ability to keep or regain your body balance or stay upright when in an unstable position.
1.A.4	**Sensory Abilities**	**Abilities that influence visual, auditory and speech perception**
1.A.4.a	**Visual Abilities**	**Abilities related to visual sensory input**
1.A.4.a.1	Near Vision	The ability to see details at close range (within a few feet of the observer).
1.A.4.a.2	Far Vision	The ability to see details at a distance.
1.A.4.a.3	Visual Color Discrimination	The ability to match or detect differences between colors, including shades of color and brightness.
1.A.4.a.4	Night Vision	The ability to see under low light conditions.
1.A.4.a.5	Peripheral Vision	The ability to see objects or movement of objects to one's side when the eyes are looking ahead.
1.A.4.a.6	Depth Perception	The ability to judge which of several objects is closer or farther away from you, or to judge the distance between you and an object.
1.A.4.a.7	Glare Sensitivity	The ability to see objects in the presence of glare or bright lighting.
1.A.4.b	**Auditory and Speech Abilities**	**Abilities related to auditory and oral input**
1.A.4.b.1	Hearing Sensitivity	The ability to detect or tell the differences between sounds that vary in pitch and loudness.
1.A.4.b.2	Auditory Attention	The ability to focus on a single source of sound in the presence of other distracting sounds.
1.A.4.b.3	Sound Localization	The ability to tell the direction from which a sound originated.
1.A.4.b.4	Speech Recognition	The ability to identify and understand the speech of another person.

1.A	Abilities	Enduring attributes of the individual that influence performance
1.A.4.b.5	Speech Clarity	The ability to speak clearly so others can understand you.

Previous factor analyses of the Abilities domain identified from 4 to 12 factors. Starting with an initial principal components analysis, Stata produces the following result (showing only the results for the first 15 components):

```
. pca armhandsteadiness- writtenexpression

Principal components/correlation                   Number of obs    =       807
                                                   Number of comp.  =        52
                                                   Trace            =        52
    Rotation: (unrotated = principal)              Rho              =    1.0000

    --------------------------------------------------------------------------
       Component |   Eigenvalue   Difference         Proportion   Cumulative
    -------------+------------------------------------------------------------
           Comp1 |      22.9339      11.8377             0.4410       0.4410
           Comp2 |      11.0962      7.95877             0.2134       0.6544
           Comp3 |      3.13745      1.11405             0.0603       0.7148
           Comp4 |       2.0234      .786501             0.0389       0.7537
           Comp5 |       1.2369     .0381955             0.0238       0.7775
           Comp6 |       1.1987        .1873             0.0231       0.8005
           Comp7 |       1.0114      .193861             0.0195       0.8200
           Comp8 |      .817544     .0683986             0.0157       0.8357
           Comp9 |      .749145     .0465881             0.0144       0.8501
          Comp10 |      .702557      .104085             0.0135       0.8636
          Comp11 |      .598472     .0931546             0.0115       0.8751
          Comp12 |      .505318     .0417174             0.0097       0.8848
          Comp13 |        .4636    .00915711             0.0089       0.8937
          Comp14 |      .454443     .0751301             0.0087       0.9025
          Comp15 |      .379313     .0299852             0.0073       0.9098
```

Following a standard criterion, I retain only those potential factors with an eigenvalue greater than 1. The eigenvalues represent the percent of total variance captured in the potential factor. The results suggest up to 7 factors which represent a total of 82% of the total variance; 4 factors represent about 75% of the total variance. The actual decision of how many factors to extract is somewhat arbitrary. The Scree plot in Figure 12 shows another view of the same information contained in the code results above – a line is drawn at an eigenvalue of 1 to show the recommended cutoff point.

Kaiser-Meyer-Olkin (KMO) statistics provide a measure of sample adequacy. KMO statistics provide values between 0 and 1 with small values meaning that overall the variables have too little in common to warrant a factor analysis; values close to 1 indicate that the factor analysis should yield distinct and reliable factors. The KMO statistics for

Abilities are shown in Table 44, with the overall KMO value of 0.961 – considered an excellent value for continuing a factor analysis.

The next analysis uses the integrated principal factor method to select the appropriate number of factors (the Stata code and results are below). The Abilities analysis was performed by identifying the specific number of potential factors from 3 to 12 – starting with a number just under the minimum literature value and extending to the maximum literature value. Then each result was rotated using oblique promax rotation. The loadings on the factors were limited to values > 0.4. Sample results for a four factor analysis are shown starting on page 200.

```
. screeplot, yline(0) ci
```

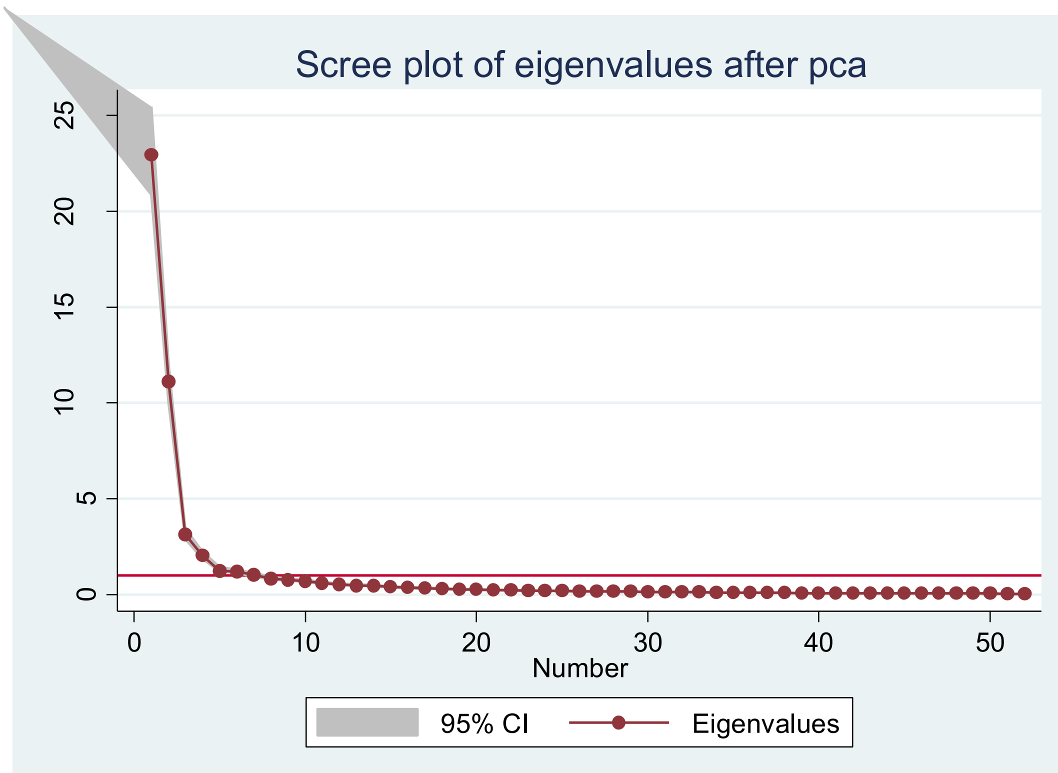

Figure 12: Scree Plot of Abilities Domain after PCA

Table 44: KMO Statistics for Abilities Domain

Variable	KMO	Variable	KMO
Arm-Hand Steadiness	0.9566	Number Facility	0.8751
Auditory Attention	0.9604	Oral Comprehension	0.9738

Variable	KMO	Variable	KMO
Category Flexibility	0.9806	Oral Expression	0.9769
Control Precision	0.967	Originality	0.9377
Deductive Reasoning	0.9723	Perceptual Speed	0.9532
Depth Perception	0.9694	Peripheral Vision	0.9618
Dynamic Flexibility	0.9159	Problem Sensitivity	0.9574
Dynamic Strength	0.9722	Rate Control	0.9645
Explosive Strength	0.8668	Reaction Time	0.9695
Extent Flexibility	0.9834	Response Orientation	0.9605
Far Vision	0.9484	Selective Attention	0.9422
Finger Dexterity	0.9594	Sound Localization	0.9653
Flexibility of Closure	0.9332	Spatial Orientation	0.9632
Fluency of Ideas	0.9358	Speech Clarity	0.9633
Glare Sensitivity	0.9817	Speech Recognition	0.9613
Gross Body Coordination	0.9604	Speed of Closure	0.9503
Gross Body Equilibrium	0.9755	Speed of Limb Movement	0.984
Hearing Sensitivity	0.9409	Stamina	0.962
Inductive Reasoning	0.9468	Static Strength	0.9737
Information Ordering	0.9709	Time Sharing	0.9435
Manual Dexterity	0.9554	Trunk Strength	0.9831
Mathematical Reasoning	0.9122	Visual Color Discrimination	0.9629
Memorization	0.9516	Visualization	0.9354
Multilimb Coordination	0.9768	Wrist-Finger Speed	0.9619
Near Vision	0.9571	Written Comprehension	0.9724
Night Vision	0.9599	Written Expression	0.9796

```
. factor  armhandsteadiness-writtenexpression, ipf factor (4)

Factor analysis/correlation                        Number of obs    =      807
    Method: iterated principal factors             Retained factors =        4
    Rotation: (unrotated)                          Number of params =      186

    --------------------------------------------------------------------------
         Factor  |   Eigenvalue   Difference        Proportion   Cumulative
    -------------+------------------------------------------------------------
        Factor1  |     21.26571     11.49504            0.5980       0.5980
        Factor2  |      9.77068      6.96713            0.2748       0.8728
        Factor3  |      2.80355      1.08216            0.0788       0.9516
        Factor4  |      1.72139      0.89913            0.0484       1.0000
        Factor5  |      0.82226      0.04684            0.0231       1.0231
        Factor6  |      0.77543      0.15985            0.0218       1.0449
```

Factor7	0.61558	0.23596	0.0173	1.0622
Factor8	0.37962	0.05412	0.0107	1.0729
Factor9	0.32550	0.08955	0.0092	1.0821
Factor10	0.23595	0.06327	0.0066	1.0887

LR test: independent vs. saturated: chi2(1128)= 5.9e+04 Prob>chi2 = 0.0000

Factor loadings (pattern matrix) and unique variances

Variable	Factor1	Factor2	Factor3	Factor4	Uniqueness
armhandste~s	0.7629	0.1909	-0.3919	0.2189	0.1801
auditoryat~n	0.5996	0.4580	-0.0658	-0.1199	0.4120
categoryfl~y	-0.5536	0.6367	-0.1069	0.1771	0.2453
controlpre~n	0.8035	0.2969	-0.3036	0.0022	0.1741
deductiver~g	-0.6986	0.6100	0.0849	0.1144	0.1196
depthperce~n	0.7311	0.4698	-0.1148	-0.0728	0.2263
dynamicfle~y	0.4338	-0.0765	0.2168	0.3219	0.6554
dynamicstr~h	0.8766	0.1094	0.1846	0.3086	0.0903
explosives~h	0.3978	0.1165	0.3647	0.2028	0.6540
extentflex~y	0.8789	0.0352	0.0845	0.3025	0.1277
farvision	0.2112	0.6768	0.0666	-0.1490	0.4708
fingerdext~y	0.5502	0.3654	-0.5790	0.0862	0.2210
flexibilit~e	-0.0145	0.8061	-0.1618	0.0013	0.3237
fluencyofi~s	-0.6046	0.6098	0.0629	0.1018	0.2482
glaresensi~y	0.7385	0.3350	0.2854	-0.2617	0.1925
grossbodyc~n	0.8524	0.0764	0.2800	0.3461	0.0694
grossbodye~m	0.8005	0.1771	0.2723	0.2377	0.1971
hearingsen~y	0.5401	0.5235	-0.1939	-0.1484	0.3746
inductiver~g	-0.6323	0.6268	0.0863	0.1732	0.1700
informatio~g	-0.4725	0.6772	-0.0716	0.1644	0.2859
manualdext~y	0.8250	0.1393	-0.3230	0.1867	0.1609
mathematic~g	-0.5379	0.5458	-0.1511	-0.0597	0.3864
memorization	-0.4386	0.5634	0.1941	0.0209	0.4521
multilimbc~n	0.8973	0.2121	-0.0412	0.1282	0.1318
nearvision	-0.5337	0.4172	-0.2499	0.0641	0.4745
nightvision	0.6923	0.3791	0.3370	-0.3533	0.1385
numberfaci~y	-0.3931	0.5330	-0.1848	-0.1149	0.5141
oralcompre~n	-0.7459	0.4971	0.1739	0.1516	0.1434
oralexpres~n	-0.7825	0.4542	0.2314	0.1151	0.1146
originality	-0.5894	0.5759	0.0607	0.1386	0.2980
perceptual~d	0.3074	0.7041	-0.2654	-0.0980	0.3297
peripheral~n	0.7344	0.3432	0.3792	-0.3203	0.0965
problemsen~y	-0.4557	0.6672	0.1187	0.1438	0.3124
ratecontrol	0.8516	0.2568	-0.0717	-0.1792	0.1717
reactiontime	0.8599	0.2886	-0.0117	-0.0989	0.1673
responseor~n	0.8392	0.3366	0.0575	-0.1699	0.1502
selectivea~n	0.1580	0.6587	0.0007	-0.0493	0.5387
soundlocal~n	0.7056	0.3738	0.2785	-0.3644	0.1520
spatialori~n	0.6930	0.3909	0.3420	-0.2542	0.1854
speechclar~y	-0.6899	0.2975	0.3706	0.0408	0.2965
Variable	Factor1	Factor2	Factor3	Factor4	Uniqueness
speechreco~n	-0.6452	0.2774	0.2278	-0.0164	0.4546
speedofclo~e	-0.2211	0.7852	0.0335	-0.0408	0.3318
speedoflim~t	0.8720	0.1577	0.2450	0.0742	0.1493
stamina	0.8658	0.0515	0.2606	0.3437	0.0617

```
    staticstre~h |   0.9036    0.1025    0.1392    0.2624 |      0.0847
     timesharing |   0.1016    0.6170    0.2281   -0.1071 |      0.5455
    trunkstren~h |   0.8182    0.0029    0.1641    0.3301 |      0.1947
    visualcolo~n |   0.4486    0.5796   -0.3910    0.0310 |      0.3090
    visualizat~n |   0.2960    0.6498   -0.3129    0.0998 |      0.3822
    wristfinge~d |   0.6764    0.1318   -0.3473   -0.0061 |      0.4044
    writtencom~n |  -0.7940    0.4941    0.0552    0.0843 |      0.1153
    writtenexp~n |  -0.8073    0.4468    0.1428    0.0558 |      0.1252
    ---------------------------------------------------------------------
. rotate, promax horst blanks(.4)

Factor analysis/correlation                        Number of obs    =      807
    Method: iterated principal factors             Retained factors =        4
    Rotation: oblique promax (Kaiser on)           Number of params =      186

    --------------------------------------------------------------------------
         Factor  |     Variance   Proportion    Rotated factors are correlated
    -------------+------------------------------------------------------------
        Factor1  |     16.44109       0.4623
        Factor2  |     15.71143       0.4418
        Factor3  |     13.62438       0.3831
        Factor4  |     12.35810       0.3475
    --------------------------------------------------------------------------
    LR test: independent vs. saturated: chi2(1128)= 5.9e+04 Prob>chi2 = 0.0000

Rotated factor loadings (pattern matrix) and unique variances

    ---------------------------------------------------------------------
        Variable |  Factor1   Factor2   Factor3   Factor4 |   Uniqueness
    -------------+----------------------------------------+--------------
    armhandste~s |                       0.7917           |      0.1783
    auditoryat~n |                       0.4779           |      0.4256
    categoryfl~y |             0.8567                     |      0.2446
    controlpre~n |                       0.7304           |      0.1797
    deductiver~g |             0.9253                     |      0.1107
    depthperce~n |                       0.5733           |      0.2340
    dynamicfle~y |   0.6765                               |      0.6588
    dynamicstr~h |   0.8048                               |      0.0891
    explosives~h |   0.6335                               |      0.6544
    extentflex~y |   0.7180                               |      0.1285
    fingerdext~y |                       0.9897           |      0.2175
    flexibilit~e |             0.6577    0.5353           |      0.3175
    fluencyofi~s |             0.8640                     |      0.2551
    glaresensi~y |                                 0.7793 |      0.1791
    grossbodyc~n |   0.9066                               |      0.0687
    grossbodye~m |   0.7515                               |      0.1951
    hearingsen~y |                       0.6231           |      0.3828
    inductiver~g |             0.9495                     |      0.1604
    informatio~g |             0.8653                     |      0.2784
    manualdext~y |                       0.7103           |      0.1611
    mathematic~g |             0.6155                     |      0.3815
    memorization |             0.7488                     |      0.4658
    ---------------------------------------------------------------------
        Variable |  Factor1   Factor2   Factor3   Factor4 |   Uniqueness
    -------------+----------------------------------------+--------------
    multilimbc~n |   0.4446              0.4542           |      0.1338
      nearvision |             0.5396                     |      0.4740
     nightvision |                                 0.9157 |      0.1164
    numberfaci~y |             0.4979                     |      0.5147
```

oralcompre~n		0.8973			0.1367
oralexpres~n		0.8715			0.1110
originality		0.8498			0.3034
perceptual~d			0.7042		0.3302
peripheral~n				0.9024	0.0800
problemsen~y		0.9011			0.3103
responseor~n				0.5312	0.1725
selectivea~n		0.4886			0.5532
soundlocal~n				0.8899	0.1313
spatialori~n				0.8134	0.1721
speechclar~y		0.6988	-0.4964		0.3011
speechreco~n		0.5714			0.4687
speedofclo~e		0.7625			0.3417
speedoflim~t	0.5543				0.1496
stamina	0.8932				0.0611
staticstre~h	0.7277				0.0853
trunkstren~h	0.7963				0.1924
visualcolo~n			0.8421		0.3114
visualizat~n			0.7425		0.3834
wristfinge~d			0.6619		0.4134
writtencom~n		0.8367			0.1067
writtenexp~n		0.8131			0.1174

(blanks represent abs(loading)<.4)

For each of the rotated results, from the analysis with 3 factors to the analysis with 12 factors, the factor loadings were examined for common themes and whether meaningful names could be applied to the potential factors. It was typical, with fewer factors, that there were too many variables assigned to each factor and the meaning of the potential factors was unclear. With too many factors, there were either too few variables assigned to be meaningful or the factors were too similar to be kept. The final choice of factors for each domain was based on the balance between meaningful interpretations of the variable distributions and the complexity of too many factors.

The Scree plot, Figure 13, after the `factor,ipf` command, supports the choice of the four factor analysis as reasonable for the Abilities domain.

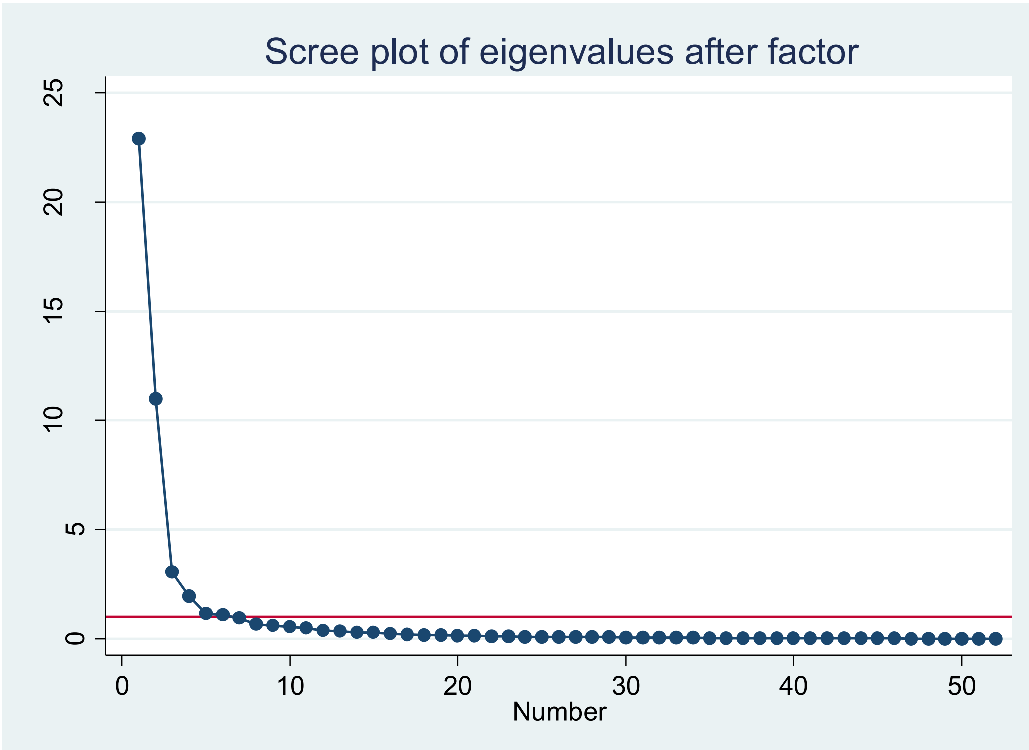

Figure 13: Scree Plot of Abilities Domain after *factor*

During the analyses and rotations, some variables were eliminated - primarily for not having an appreciable factor loading on any of the potential factors, e.g., not having a primary factor loading of 0.4 or above. For Abilities, four variables: Far Vision, Rate Control, Reaction Time, and Time Sharing were eliminated from the final factor analysis.

Naming the factors. Just as the number of final factors chosen is arbitrary, the naming conventions are arbitrary. Names are chosen so as to have a relationship to the primary variables in the factors and to be meaningful to the overall intended use of the factor analysis. For example, the second factor above contains reasoning abilities, problem solving, abilities to express ideas, and creativity related abilities – a clear set of cognitive abilities.

The final rotated Ability factor loadings reordered and sorted by factor loading are in Table 45. The factor labels proposed by (Hadden, Kravets, and Muntaner 2004) suited the extracted factors and were retained.

Table 45: Final Ability Domain Factor Loadings (sorted)

Variable	Sorted factor loadings			
	Cognitive Ability	Fine Motor Abilities	Gross Motor Skills, Strength, and Endurance	Auditory and Visual Processing
Inductive Reasoning	0.9495			
Deductive Reasoning	0.9253			
Problem Sensitivity	0.9011			
Oral Comprehension	0.8973			
Oral Expression	0.8715			
Information Ordering	0.8653			
Fluency of Ideas	0.864			
Category Flexibility	0.8567			
Originality	0.8498			
Written Comprehension	0.8367			
Written Expression	0.8131			
Speed of Closure	0.7625			
Memorization	0.7488			
Speech Clarity	0.6988	-0.4964		
Flexibility of Closure	0.6577	0.5353		
Mathematical Reasoning	0.6155			
Speech Recognition	0.5714			
Near Vision	0.5396			
Number Facility	0.4979			
Selective Attention	0.4886			
Finger Dexterity		0.9897		
Visual Color Discrimination		0.8421		
Arm-Hand Steadiness		0.7917		
Visualization		0.7425		
Control Precision		0.7304		
Manual Dexterity		0.7103		
Perceptual Speed		0.7042		
Wrist-Finger Speed		0.6619		
Hearing Sensitivity		0.6231		
Depth Perception		0.5733		
Auditory Attention		0.4779		
Multilimb Coordination		0.4542	0.4446	
Gross Body Coordination			0.9066	

Variable	Sorted factor loadings: Cognitive Ability	Fine Motor Abilities	Gross Motor Skills, Strength, and Endurance	Auditory and Visual Processing
Stamina			0.8932	
Dynamic Strength			0.8048	
Trunk Strength			0.7963	
Gross Body Equilibrium			0.7515	
Static Strength			0.7277	
Extent Flexibility			0.718	
Dynamic Flexibility			0.6765	
Explosive Strength			0.6335	
Speed of Limb Movement			0.5543	
Night Vision				0.9157
Peripheral Vision				0.9024
Sound Localization				0.8899
Spatial Orientation				0.8134
Glare Sensitivity				0.7793
Response Orientation				0.5312

The complete set of factors from all O*NET domains is shown in Table 46 – a total of 22 factors. The data and results for the factor analyses of the remaining domains are available from the author upon request.

Table 46: Factors from O*NET Domains

#	Domain	Factors
4	Abilities	Gross Motor Skills, Strength, and Endurance
		Cognitive Ability
		Fine Motor Abilities
		Auditory and Visual Processing
2	Education and Experience	Training
		Education and Experience
4	Knowledge	Social Science
		Business
		Bio-Medicine
		Engineering

#	Domain	Factors
3	Skills	Organizational Skills
		Quantitative Skills
		Technical Skills
3	Work Activities	Analyze and Decide
		Interact with Others
		Work with Things
4	Work Context	Physically Challenging
		Office Context
		Socially Challenging
		Repetition
2	Work Styles	Leadership
		Cooperation

Factor Analysis over all O*NET Variables

The second factor analysis included all O*NET variables. The analysis started with a total of 225 O*NET variables and nine demographic variables. During this analysis, another 8 variables were eliminated leaving a total of 226 variables in the final factor analysis and rotation. After exploring a range of factors from three to ten, a four factor solution was chosen that explained 58% of the common variance. The overall KMO statistic is 0.9755. Factor 1 explains 32% of the total variance; Factor 2 explains 15% of the total variance; Factor 3 explains 7% of the total variance; and Factor 4 explains 4% of the total variance.

(Hadden, Kravets, and Muntaner 2004) also presented a four factor solution in an analysis of the complete O*NET variable set, using version 4.0. Their choice of domains was different, and the intervening versions of the O*NET data added a significant number of updated occupations. See Table 47 for description of the factors in this dissertation.

Table 47: Overall O*NET Factor Analysis Results

	Factor	Explanation
1	Reasoning & Complexity	Variables measuring reasoning ability, thinking skills, learning, and information processing work activities
2	Physical Demands	Variables that represent the physical demands of work
3	People versus Things	Variables that represent interpersonal work contexts and activities, and social skills contrasted with work contexts and skills related to electrical or mechanical equipment
4	Attention to Detail	Variables that focus on precision, eye-hand coordination,

	Factor	Explanation
		and attention to detail

Detailed Description of Overall O*NET Variables

Factor 1 – Reasoning and Complexity

Top characteristics:

1	Complex Problem Solving
2	Coordination
3	Developing Objectives and Strategies
4	Active Learning
5	Critical Thinking
6	Scheduling Work and Activities
7	Judgment and Decision Making
8	Monitoring
9	Provide Consultation and Advice to Others
10	Persuasion
11	Speaking
12	Thinking Creatively
13	Time Management
14	Education and Training
15	Coaching and Developing Others

Other key characteristics: Negotiation, Originality, Leadership, Active Listening, Initiative, Analytical Thinking, Innovation, Persistence, Deductive Reasoning, and Inductive Reasoning

The bottom characteristics are physically oriented, or focused on mechanical activities, not cognition or reasoning.

3	Pace Determined by Speed of Equipment
2	Spend Time Using Your Hands to Handle, Control, or Feel Objects, Tools, or Controls
1	Spend Time Making Repetitive Motions

Factor 2 – Physical Demands

Top characteristics:

1	Operating Vehicles, Mechanized Devices, or Equipment
2	Performing General Physical Activities
3	Depth Perception
4	Multi-limb Coordination
5	Gross Body Equilibrium
6	Glare Sensitivity
7	Response Orientation
8	Speed of Limb Movement

9	Very Hot or Cold Temperatures
10	Extremely Bright or Inadequate Lighting
11	Cramped Work Space, Awkward Positions
12	Static Strength
13	Responsible for Others' Health and Safety
14	Exposed to Hazardous Conditions
15	Dynamic Strength

Other key characteristics: Stamina, Handling and Moving Objects, Repairing and Maintaining Mechanical Equipment, Trunk Strength, Peripheral Vision, and Exposed to Contaminants

The bottom characteristics are indoor and structured office type activities, not physical activities.

3	Clerical
2	Indoors, Environmentally Controlled
1	Spend Time Sitting

Factor 3 – People vs. Things

Top characteristics

1	Assisting and Caring for Others
2	Deal With Unpleasant or Angry People
3	Self Control
4	Concern for Others
5	Deal With Physically Aggressive People
6	Exposed to Disease or Infections
7	Social Orientation
8	Contact With Others
9	Stress Tolerance
10	Frequency of Conflict Situations
11	Physical Proximity
12	Medicine and Dentistry
13	Cooperation
14	Deal With External Customers
15	Performing for or Working Directly with the Public

Other key characteristics are: Customer and Personal Service, Dependability, and Resolving Conflicts and Negotiating with Others

The bottom characteristics are oriented around working with equipment, or designing and installing equipment, not dealing with people.

1	Engineering and Technology
2	Troubleshooting
3	Programming
4	Mathematics
5	Quality Control Analysis

6	Equipment Selection
7	Installation
8	Design
9	Operations Analysis
10	Technology Design

Factor 4 – Attention to Detail

The top characteristics:

1	Importance of Being Exact or Accurate
2	Flexibility of Closure
3	Importance of Repeating Same Tasks
4	Near Vision
5	Degree of Automation
6	Attention to Detail
7	Number Facility
8	Perceptual Speed
9	Selective Attention
10	Finger Dexterity
11	Consequence of Error
12	Speed of Closure
13	Information Ordering
14	Documenting/ Recording Information
15	Problem Sensitivity

Occupation List

Based on the Standard Occupation Classification (SOC), from the Bureau of Labor Statistics(Bureau of Labor Statistics 2010a), the O*NET taxonomy includes over 950 occupations. For O*NET version 13, a total of 807 occupations had complete sets of variable data for the domains under consideration. The SOC occupation codes used by O*NET define the detailed, recode, and major occupation groups used in this study, see Table 48 for a brief example.

Table 48: Example of O*NET-SOC Occupation Listing

Major Group	Recode Group	Detailed Group	Occupation Title
11-0000			Management Occupations
	11-1000		Top Executives
		11-1010	Chief Executives
		11-1020	General and Operations Managers
	11-2000		Advertising, Marketing, Promotions, Public Relations, and Sales Managers

Major Group	Recode Group	Detailed Group	Occupation Title
		11-2010	Advertising and Promotions Managers
		11-2021	Marketing Managers
		11-2022	Sales Managers
		11-2030	Public Relations and Fundraising Managers
	11-300		Operations Specialties Managers
		11-3010	Administrative Services Managers
		11-3020	Computer and Information Services Managers
		11-3031	Treasurers and Controllers
		11-3032	Financial Managers, Branch or Department
		11-3040	Human Resources Managers
		11-3041	Compensation and Benefits Managers
		11-3042	Training and Development Managers
		11-3050	Industrial Production Managers
		11-3060	Purchasing Managers
		11-3071	Transportation Managers
		11-3072	Storage and Distribution Managers

To evaluate the relevance of the four overall O*NET factors, Table 49 contrasts the O*NET-SOC occupation titles that have the highest overall factor rating with those occupation titles that have the lowest rating.

Table 49: Occupations Ranking High/Low on O*NET Factors

#	Factor	Highest Ranked Occupations	Lowest Ranked Occupations
1	Reasoning & Complexity	Physicists, medical scientists, physical scientists, physicians and surgeons, engineering teachers, social scientists, administrators	Graders, sorters, crossing guards, sewing machine operators, textile workers, cafeteria attendants, housekeeping workers
2	Physical Demands	Iron and steel workers, derrick operators, commercial divers, fire fighters, mining operators, roofers, drill operators pile-driver operators	Political scientists, sociologists, proofreaders, public relations, benefits managers, law clerks, telemarketers, economists
3	People versus Things	Public health social workers, psychiatrists, nurses, police officers, physician assistants, recreational therapists, flight attendants, emergency medical technicians	Astronomers, software engineers, mathematicians, operations research analysts, electrical engineers, drafters, lathe operators
4	Attention to Detail	Surgeons, air traffic controllers, crime lab officers, clinical laboratory technicians,	Models, tour guides, street vendors, coffee shop attendants,

#	Factor	Highest Ranked Occupations	Lowest Ranked Occupations
		anesthesiologists, oral surgeons, nuclear power reactor operators, mechanical engineers	coatroom attendants, maids, packers, vehicle cleaners

The high and low ranked occupations are similar to those previously reported (Department of Employment and Economic Development 1999; Hadden, Kravets, and Muntaner 2004).

Use of Factors with Occupations to create Regression Variables

The detailed occupation codes in the NLMS database are based on the 1980 and 1990 Census occupation codes. The O*NET detailed, recode, and major group occupation codes were mapped back to the NLMS data to allow the use of the factor analysis results to generate regression terms for occupation characteristics. As cross-walks are not available, I performed the mapping across all datasets in a consistent manner by hand.

Each occupation title in the O*NET-SOC listing is scored in each domain in the O*NET database. For example, in the Abilities domain each occupation title is scored from 0-5 on each of the 52 ability variables (0 means very low capability, 5 means very high capability). The scoring for an occupation title is one of the key updates that occurs from one release of O*NET to the next. Table 50 shows a sample score provided for some of the occupations in Table 48.

Table 50: Example of O*NET Occupation Scoring

Detailed Group	Occupation Title	Arm-Hand Steadiness	Category Flexibility	Control Precision	Deductive Reasoning	Multilimb Coordination	Spatial Orientation
11-1010	Chief Executives	0.000	4.000	1.500	5.000	1.130	0.000
11-1020	General and Operations Managers	2.120	3.500	2.250	4.380	2.000	1.750
11-2010	Advertising and Promotions Managers	1.000	3.750	1.250	4.250	0.750	0.130
11-2021	Marketing Managers	0.000	3.250	0.000	3.620	0.000	0.000
11-2022	Sales Managers	1.380	3.380	2.000	3.880	1.880	0.750
11-2030	Public Relations and	0.000	3.500	0.880	4.000	0.880	0.000

Detailed Group	Occupation Title	Arm-Hand Steadiness	Category Flexibility	Control Precision	Deductive Reasoning	Multilimb Coordination	Spatial Orientation
	Fundraising Managers						

There are several possible ways to combine factor loadings and occupation scoring. For this analysis I have chosen a weighted sum, with weights equal to the estimated factor loadings. So, let's use the following Abilities factor loadings.

	Factor Loadings			
Variable	**Cognitive Ability**	**Fine Motor Abilities**	**Gross Motor Skills, Strength, and Endurance**	**Auditory and Visual Processing**
Arm-Hand Steadiness		0.7917		
Category Flexibility	0.8567			
Control Precision		0.7304		
Deductive Reasoning	0.9253			
Multilimb Coordination		0.4542	0.4446	
Spatial Orientation				0.8134

The regression coefficients for the occupation titles listed above weighted by the factor loadings are calculated as in Table 51.

Table 51: Determination of Factor Coefficients

Detailed Group	Occupation Title	Factor 1: Cognitive Ability	Factor 2: Fine Motor Abilities	Factor 3: Gross Motor Skills, Strength, and Endurance	Factor 4: Auditory and Visual Processing
11-1010	Chief Executives	0.8567*4.00 + 0.9253*5.00 = 8.05	0.7917*0.00 + 0.7304*1.50 + 0.4542*1.13 = 1.61	0.4446*1.13 = 0.50	0.8134*0.00 = 0.00
11-1020	General and Operations Managers	0.8567*3.50 + 0.9253*4.38 = 7.05	0.7917*2.12 + 0.7304*2.25 + 0.4542*2.00 = 4.23	0.4446*2.00 = 0.89	0.8134*1.75 = 1.42
11-2010	Advertising and Promotions	0.8567*3.75 + 0.9253*4.25 = 7.15	0.7917*1.00 + 0.7304*1.25 + 0.4542*0.75 =	0.4446*0.75 = 0.33	0.8134*0.13 = 0.11

Detailed Group	Occupation Title	Factor 1: Cognitive Ability	Factor 2: Fine Motor Abilities	Factor 3: Gross Motor Skills, Strength, and Endurance	Factor 4: Auditory and Visual Processing
	Managers		2.05		
11-2021	Marketing Managers	0.8567*3.25 + 0.9253*3.62 = 6.13	0.7917*0.00 + 0.7304*0.00 + 0.4542*0.00 = 0.00	0.4446*0.00 = 0.00	0.8134*0.00 = 0.00
11-2022	Sales Managers	0.8567*3.38 + 0.9253*3.88 = 6.49	0.7917*1.38 + 0.7304*2.00 + 0.4542*1.88 = 3.41	0.4446*1.88 = 0.84	0.8134*0.75 = 0.61
11-2030	Public Relations and Fundraising Managers	0.8567*3.50 + 0.9253*4.00 = 6.70	0.7917*0.00 + 0.7304*0.88 + 0.4542*0.88 = 1.04	0.4446*0.88 = 0.39	0.8134*0.00 = 0.00

For this small example, Factor 1: Cognitive Ability is the most important factor for these occupations; while Factor 4: Auditory and Visual Processing is the least important factor. There are also clear differences between the occupations with regard to each factor.

For the entire set of occupation titles, the Ability factor regression coefficients are calculated, in the same fashion, using the entire factor loading table (Table 45) and the entire occupation title listing (not shown). Then, using the mappings to the recode, major, and BRG groupings, summarized coefficients are calculated. The calculations are then repeated for the other domain factor analyses and for the overall factor analysis.

As the ranges of the factors are significantly different, I normalized the factors so they fell within the range (0, 1). The adjusted summary statistics for the overall factors are shown below:

Variable	Mean	Std. Dev.	Min	Max
Overall: Reasoning and Complexity	0.464	0.208	0	1
Overall: Physical Demands	0.430	0.248	0	1
Overall: People vs. Things	0.480	0.186	0	1
Overall: Attention to Detail	0.466	0.161	0	1

Finally, individual NLMS Stata databases were created which tie the occupations as reported by the participants and the occupation factors to the appropriate occupation groupings. The result is a set of six Stata data files for NLMS:

Occupation Grouping	NLMS file
Detailed occupations	Occ.dta
Occupation recode male	Recodem.dta
Occupation recode female	Recodef.dta
Major occupations	Majocc.dta
BRG male	Brgm.dta
BRG female	Brgf.dta

Appendix E. Descriptions of Chapter 4 Variables

Table 52 provides definition and source information for the dependent variables used in the NLMS analyses described in chapter 4.

Table 52: Chapter 4 Explanatory Variables

Variable	Description	Source	Mean	Std. Dev	Min	Max
Occupation/Industry Category Variables						
BRGm	British Registry Groups-male	constructed	2.76	1.20	1	4
BRGf	British Registry Groups-female	constructed	2.34	1.27	1	4
MajorOcc	Major Occupation Code	NLMS	7.48	5.13	1	18
OccupationRecodeMale	Recoded Occupations - Male	constructed	47.3	26.0	1	88
OccupationRecodeFemale	Recoded Occupations - Female	constructed	32.6	16.8	1	59
Occupation	Occupation Code - 1990	NLMS	404.6	249.5	4	905
MajOcc_f1	Factor 1 - Gross Motor Skills	constructed	9.69	5.43	3.4	20.39
MajOcc_f2	Factor 2 - Cognitive Ability	constructed	45.83	6.04	35.92	55.58
MajOcc_f3	Factor 3 - Fine Motor Abilities	constructed	17.12	5.08	11.88	26.27
MajOcc_f4	Factor 4 - Aud. & Vis. Processing	constructed	3.49	2.99	1	11.43
MajOcc_f5	Factor 5 - Training	constructed	6.31	0.92	4.08	8.27
MajOcc_f6	Factor 6 - Education and Exper.	constructed	4.98	1.40	3.57	7.78
MajOcc_f7	Factor 7 - Social Science	constructed	8.58	2.79	5.59	14.94
MajOcc_f8	Factor 8 - Business	constructed	16.89	3.92	11.38	24.37
MajOcc_f9	Factor 9 - Bio-Medicine	constructed	5.50	1.78	3.24	9.64
MajOcc_f10	Factor 10 - Engineering	constructed	11.81	4.15	5.55	19.84
MajOcc_f11	Factor 11 - Organizational Skills	constructed	58.50	8.01	45.51	73.76
MajOcc_f12	Factor 12 - Quantitative Skills	constructed	5.44	1.21	3.13	7.75
MajOcc_f13	Factor 13 - Technical Skills	constructed	19.30	5.85	13.26	32.36
MajOcc_f14	Factor 14 - Analyze and Decide	constructed	41.08	6.72	29.19	53.07
MajOcc_f15	Factor 15 - Interact with Others	constructed	30.08	5.72	21.9	41.45

Variable	Description	Source	Mean	Std. Dev	Min	Max
MajOcc_f16	Factor 16 - Work with Things	constructed	15.99	5.92	9.9	29.18
MajOcc_f17	Factor 17 - Physically Challenging	constructed	31.31	9.31	22.64	55.47
MajOcc_f18	Factor 18 - Office Context	constructed	10.66	4.98	1.9	17.97
MajOcc_f19	Factor 19 - Socially Challenging	constructed	22.33	1.34	19.61	26.86
MajOcc_f20	Factor 20 - Repetition	constructed	10.07	0.79	7.76	11.23
MajOcc_f21	Factor 21 - Leadership	constructed	22.13	1.32	20.56	24.52
MajOcc_f22	Factor 22 - Cooperation	constructed	22.88	1.16	20.74	24.3
MajOcc_f23	Factor 23 - Prestige Ranking	BLS	41.50	10.31	29.2	63.8
MajOcc_f24	Factor 24 - Duncan SE Index	NORC	47.24	15.72	26.3	77.9
MajOcc_f25	Factor 25 - Reasoning & Complexity	constructed	211.30	33.14	169.6	270.9
MajOcc_f26	Factor 26 - People vs. Things	constructed	98.31	34.82	62.04	178.8
MajOcc_f27	Factor 27 - Physical Demands	constructed	32.88	6.91	20.96	43.28
MajOcc_f28	Factor 28 - Attention to Detail	constructed	26.90	2.22	22.38	31.06
Industry	Industry Code	NLMS	543.10	287.24	10	991
MajorInd	Major Industry Code	NLMS	8.70	4.22	1	17

Appendix F. Factor Analysis of State-Level Characteristics

The initial list of 56 State cultural rankings and 25 demographic variables are in Table 53.

Table 53: Initial List of State Characteristics and Demographic Variables

State Ranking		State Demographics
IQ Rank	Gasoline Usage per capita	Latitude
Smartest	UFO Sightings	Longitude
Obesity Rate	Starbucks per capita	Urban percentage
Exercise Rate	Wal-Mart stores per capita	Census Region
Church Attendance	Pollution levels	Census Division
Importance of Religion in Daily Life	Cancer deaths per capita	Population Density
Percentage Going Hungry	Coronary heart disease per capita	Square Miles
Freedom Index	Cardiovascular deaths per capita	Unemployment rate
Tax Burden	Percentage of children under 18 in poverty	Poverty Percentage
Moocher Index	Fruit portions eaten per day	Income per capita
Coincident Index	Outcome Disparity within state	Female percentage
Pro-Business Index	Percentage reporting Poor Health	White percentage
Gini Index	Infectious disease rate	Black percentage
Farming as a percentage of State GDP	Percentage with No Health Insurance	Percentage 0-17 years
Farming Productivity	Unnecessary hospital visits per capita	Percentage 18-24 years
Happiness Index	Primary Care Physicians per capita	Percentage 25-44 years
Well-Being Index	Public Health $ per capita	Percentage 45-65 years
Generosity Index	Mortality rate	Percentage 65+ years
Manufacturing Employment	Autism per capita	High School Graduation rate
Manufacturing Output as a percent of State GDP	Teen Birth rate	College Graduation rate
Teacher Pay Levels	White Mortality rate	Alcohol Use per capita
Education $ Spent per Pupil	Black Mortality rate	Smoking Rate per capita
Percentage 9th Graders Graduating High School	Occupational Death rate	Births per capita
Womens' Status ranking	Years of Potential Life Lost (YPLL)	Men Registered to Vote
Crime Rate - overall	Healthiest	Women Registered to Vote
Violent Crime Rate	Binge Drinking rate	
Speeding – traffic deaths due to speeding	Smoking percentage	
Traffic Deaths - overall	Under-employed percentage	

In the initial state factor analysis, the evaluation of the complete set of characteristics resulted in a set of six (6) factors. Images for each of the factors are shown in Figure 14. In each image, the States with the highest ranking are in dark blue:
and the States with the lowest rank are in pale green:

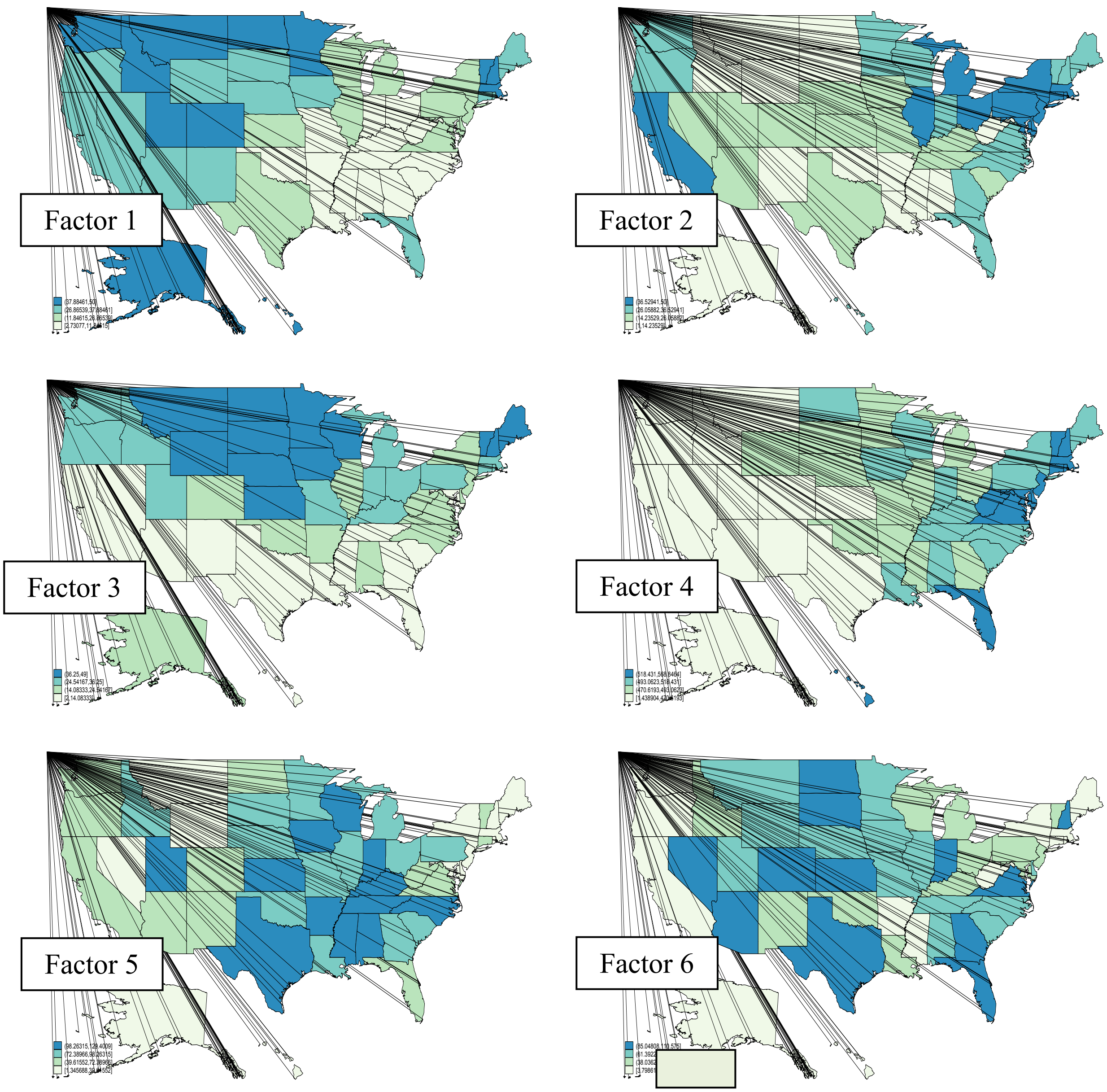

Figure 14: Initial State Factor Images

In this analysis, the six factors account for 72% of the total variance; factor 1 accounts for 27%, factor 2 accounts for 15%, factor 3 accounts for 14%, factor 4 accounts for 6%, factor 5 accounts for 5%, and factor 6 accounts for 4%.

The top characteristics of each factor are listed in Table 54 along with other representative characteristics, and the name identifying the factor (names are generously based on Weiss' lifestyle descriptions (Weiss 2000)).

In this set of state cultural rankings, the most prominent characteristics of factor 1, the factor accounting for the largest percentage of variance, are all related to health. In light of the topics in this dissertation, this is an interesting result. That is, the rankings that most distinguish one state from another are those related to the health of the population within the states.

Table 54: State Level Factor Analysis

Factor	Top Characteristics	Factor Name
1	1. Low Cancer Deaths	Big Sky Country
	2. Low Cardiovascular Deaths	
	3. Low Smoking Rates	
	4. Low Levels of Unnecessary Medical Care	
	5. Low Obesity Rates	
	Other: high well-being index, high exercise rates, healthiest, low mortality rates for blacks and whites	
2	1. Low occupational death rates	Upward Bounders
	2. High in women's rights	
	3. High in primary care physicians per capita	
	4. High in amount of fruit eaten per capita	
	5. Low in percentage on poverty	
	Other: low in teen births, high on $ spent on K-12 education, high $ for teacher salaries, smartest	
3	1. Low rates of infections (HIV, STD)	Big Fish, Small Pond
	2. High in IQ	
	3. Low overall crime rates	
	4. High in graduates	
	5. Low in those having no health insurance	
	Other: Low in violent crime, healthiest, low in percentage urban	
4	1. Low in unemployment rates	Young Influentials
	2. Higher in percentage younger (<25 years) persons	
	3. Low in hunger rates	
5	1. Higher in percentage 45-64 years old persons	Heartlanders
	2. High in church attendance	
	3. High in importance of religion in daily lives	
	4. Higher in alcohol per capita usage	

Factor	Top Characteristics	Factor Name
	5. High in farming output	
6	1. Low in the Moocher index	American Dreams
	2. Highly pro-business	
	3. Low in overall tax rates	
	4. High in the Freedom index	
	5. Low in underemployment rates	

These health-related cultural rankings, however, are also likely to be endogenous with the dependent variables used in chapters 2, 3, and 4. A second factor analysis was performed after removing the cultural rankings most related to mortality outcomes. Table 55 shows the 36 cultural rankings and the 19 demographic variables used in this analysis.

Table 55: Final List of State Characteristics and Demographic Variables

State Ranking		State Demographics
IQ Rank	Crime Rate - overall	Latitude
Smartest	Violent Crime Rate	Longitude
Obesity Rate	Traffic Deaths - overall	Urban percentage
Exercise Rate	Gasoline Usage per capita	Population Density
Church Attendance	Starbucks per capita	Square Miles
Importance of Religion in Daily Life	Wal-Mart stores per capita	Poverty Percentage
Freedom Index	Pollution levels	GSP per capita
Coincident Index	Percentage of children under 18 in poverty	Income per capita
Gini Index	Fruit portions eaten per day	Female percentage
Farming as a percentage of State GDP	Percentage reporting Poor Health	White percentage
Farming Productivity	Infectious disease rate	Black percentage
Happiness Index	Percentage with No Health Insurance	Percentage 0-17 years
Generosity Index	Unnecessary hospital visits per capita	Percentage 18-24 years
Manufacturing Employment	Primary Care Physicians per capita	Percentage 25-44 years
Manufacturing Output as a percent of State GDP	Autism per capita	Percentage 45-65 years
Teacher Pay Levels	Teen Birth rate	Percentage 65+ years

Percentage 9th Graders Graduating High School	Healthiest	High School Graduation rate
Womens' Status ranking	Binge Drinking rate	College Graduation rate

In this factor analysis, the evaluation of the 55 characteristics results in a set of four (4) factors. Images for the factors are shown in Figure 15. In each image, the States with the highest rankings are in dark blue: and the states with the lowest rank are in pale green:

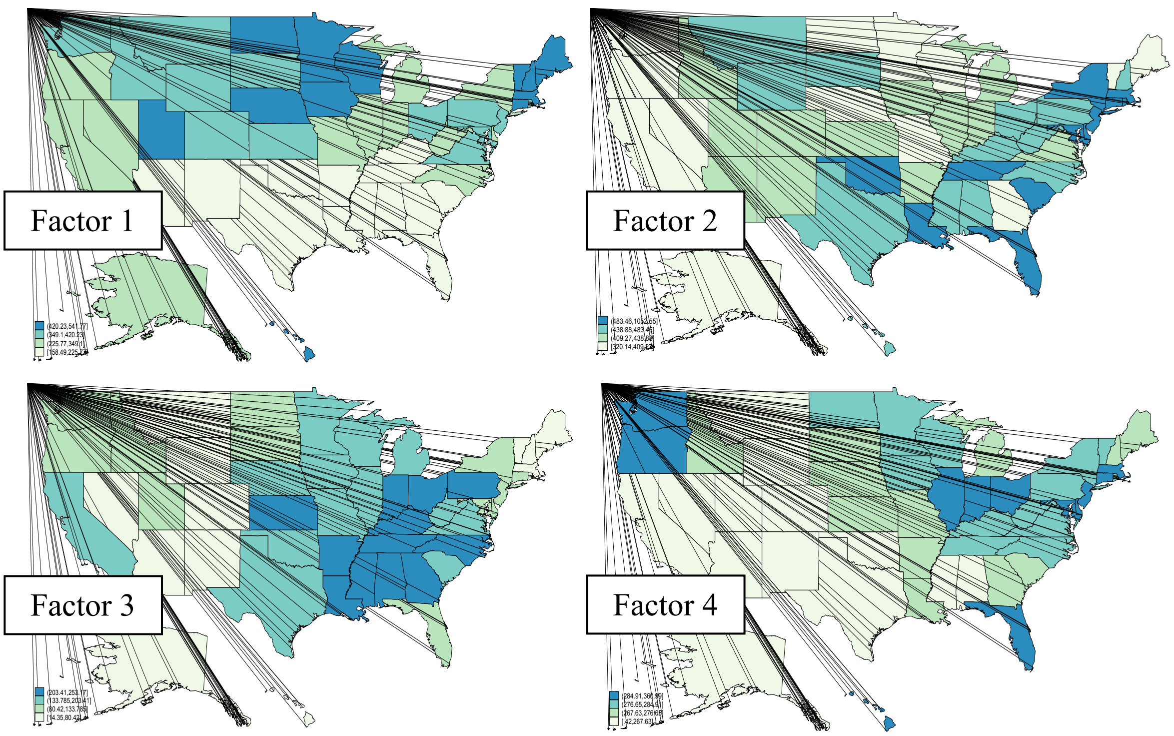

Figure 15: Final State Factor Images

The four factors account for 70% of the total variance; factor 1 accounts for 32%, factor 2 accounts for 17%, factor 3 accounts for 13%, and factor 4 accounts for 8%.
The top characteristics of each factor are listed in Table 56 along with other representative characteristics and the name identifying the factor (again, names are based on Weiss' lifestyle descriptions (Weiss 2000)).

Table 56: State Level Factor Analysis

Factor	Top Characteristics	Factor Name
1	1. High IQ Rank	Big Fish, Small Pond
	2. Healthiest	
	3. High in Percentage of Graduates	
	4. Low Crime Rate	
	5. Low Gini Index	
	6. Low in Teen Birth rate	
	7. High in Average Exercise per capita	
	8. Smartest	
	Other: low in those with no insurance, low child poverty rate, high in binge drinking, more northern latitudes preferred	
2	1. High Income per capita	Up-and-Comers
	2. High in Urban Percentage	
	3. High in Percentage in the 25-44 age group	
	4. High Generosity	
	5. Low in Autism per capita	
	6. High in population density	
	7. High in Happiness Index	
	Other: high in college graduate percentage, high in Freedom index, low in fruit portions eaten per day, low in women's rights	
3	1. High in Manufacturing Employment	Heartlanders
	2. High in Manufacturing Output as a Percentage of GSP	
	3. High in Church Attendance	
	4. High in Importance of Religion in Daily Lives	
	5. High in Farming Percentage	
	6. High in the number of Wal-Mart Stores per capita	
	Other: High in level of unnecessary medical care, high in obesity, high in pollution	
4	1. High in Starbucks per capita	Empty Nesters
	2. High in Percentage of 45-64 age group	
	3. High in Percentage of 65-plus age group	
	4. High in Percentage of Females	
	Other: more western longitudes preferred, low in percentage of 18-24 age group, low in percentage of 0-17 age group, low in births per capita	

In the second analysis, with the mortality related variables excluded, the primary characteristics in Factor 1 are related to education, crime rates, and the remaining (non-mortality) health rankings. These categories of cultural characteristics appear to be those variables that best define differences in state groupings.

As the ranges of the variables are different, I normalized the variables so they fell within the range (0, 1). The adjusted summary statistics are shown below:

Variable	Mean	Std. Dev.	Min	Max
Factor 1: Big Fish, Small Pond	0.432	0.263	0	1
Factor 2: Up-and-Comers	0.216	0.191	0	1
Factor 3: Heartlanders	0.740	0.117	0	1
Factor 4: Empty Nesters	0.561	0.255	0	1

References

References

Adams, Peter, Michael Hurd, Daniel McFadden, Angela Merrill, and Tiago Ribeiro. 2003. "Healthy, Wealthy, and Wise? Tests for Direct Causal Paths between Health and Socioeconomic Status." *Journal of Econometrics* 112: 3-56.

Adler, Nancy, and Katherine Newman. 2002. "Socioeconomic Disparities in Health: Pathways and Policies." *Health Affairs* 21 (2): 60-76.

Anderson, G, Jeremy Hurst, Peter Hussey, and Melissa Jee-Hughes. 2000. "Health Spending and Outcomes: Trends in OECD Countries, 1960-1998." *Health Affairs* 19 (3): 150.

Anderson, G, Uwe Reinhardt, Peter Hussey, and Varduhi Petrosyan. 2003. "It's the Prices, Stupid: Why the United States is so Different from Other Countries." *Health Affairs* 22 (3): 89.

Anderson, Richard, Paul Sorlie, Eric Backlund, Norman Johnson, and George Kaplan. 1997. "Mortality Effects of Community Socioeconomic Status." *Epidemiology* 8 (1): 42-47.

Arah, Onyebuchi, Gert Westert, Diana Delnoij, and Niek Klazinga. 2005. "Health System Outcomes and Determinants Amenable to Publi Health in Industialized Countries: A Pooled Cross-Sectional Time Series Analysis." *BMC Public Health* 5.

Ariste, Ruolz, and Jeff Carr. 2003. "New Considerations on the Empirical Analysis of Health Expenditures in Canada: 1966-1998." *Health Canada Working Paper* 02-06.

Auster, Richard, Irving Leveson, and Deborah Sarachek. 1969. "The Production of Health, An Exploratory Study." *Journal of Human Resources* 4 (4): 411-436.

Bac, Catherine, and Yannick Le Pen. 2002. An International Comparison of Health Care Expenditure Determinants. In *10th International Conference on Panel Data*. Berlin.

Backlund, Eric, Paul Sorlie, and Norman Johnson. 1996. "The Shape of the Relationship between Income and Mortality in the United States: Evidence from the NLMS." *AEP* 6

(1): 12-20.

———. 1999. “A Comparison of the Relationships of Education and Income with Mortality: the National Longitudinal Mortality Study.” *Social Science and Medicine* 49: 1373-1384.

Barros, Pedro. 1998. “The Black Box of Health Care Expenditure Growth Determinants.” *Health Economics* 7: 533.

Beck, Nathaniel. 2001. “Time-Series–Cross-Section Data: What Have We Learned in the Past Few Years?” *Annual Review of Political Science* 4: 271-293.

Beck, Nathaniel, and Jonathan Katz. 1995. “What To Do (and Not To Do) with Time-Series Cross-Section Data.” *American Political Science Review* 89 (3): 634-647.

Bednarek, Heather, Rowena Pecchenino, and Sally Stearns. 2003. “Fat and Happy: Health Choices and Health Outcomes.” *mimeo, Michigan State University*.

Berger, Mark, and J. Paul Leigh. 1989. “Schooling, Self-Selection, and Health.” *Journal of Human Resources* 24 (3): 433-455.

Berger, Mark, and Jodi Messer. 2002. “Public Financing of Health Expenditures, Insurance, and Health Outcomes.” *Applied Economics* 34: 2105-2113.

Berk, Marc, and Alan Monheit. 2001. “The Concentration of Health Care Expenditures, Revisited.” *Health Affairs* 20 (2): 9.

Biggs, Brian, L King, S Basu, and David Stuckler. 2010. “Is Wealthier always Healthier? The Impact of National Income Level, Inequality, and Poverty on Public Health in Latin America.” *Social Science & Medicine* 71: 266-273.

Bodenheimer, Thomas. 2005a. “High and Rising Health Care Costs. Part 1: Seeking an Explanation.” *Annals of Internal Medicine* 142 (10): 847.

———. 2005b. “High and Rising Health Care Costs. Part 2: Technological Innovation.” *Annals of Internal Medicine* 142 (11): 932.

———. 2005c. “High and Rising Health Care Costs. Part 3: The Role of Health Care Providers.” *Annals of Internal Medicine* 142 (12): 996.

Bodenheimer, Thomas, and Alicia Fernandez. 2005. “High and Rising Health Care Costs. Part 4: Can Costs be Controlled While Preserving Quality.” *Annals of Internal Medicine* 143 (1): 26.

Bosma, Hans, Stephen Stansfield, and Michael Marmot. 1998. “Job Control, Personal

Characteristics, and Heart Disease." *Journal of Occupational Health Psychology* 3 (4): 402-409.

Braveman, Paula, Catherine Cubbin, Susan Egerter, Sekai Chideya, Kristen Marchi, Marilyn Metzler, and Samuel Posner. 2005. "Socioeconomic Status in Health Research: One Size Does Not Fit All." *JAMA* 294 (22): 2879-2888.

Bureau of Labor Statistics. 2010a. Standard Occupational Classification (SOC) System. http://www.bls.gov/soc/.

———. 2010b. O*NET Resource Center. http://www.onetcenter.org/.

Chandra, Amitabh, Elliott Fisher, and Jonathan Skinner. 2007. "Pitfalls in the Analysis of Regional Variation in Health Care: A Response to Hadley, Berenson, Waidmann, and Zuckerman."

Clark, Damon, and Heather Royer. 2010. "The Effect of Education on Adult Health and Mortality: Evidence from Britain." *NBER Working Paper 16013*.

Connelly, Luke, and Darrel Doessel. 2004. "Medical Expenditures and Health Status in Australia: A Story of Increasing or Decreasing Returns?" *Australian Economic Review* 37 (1): 12-30.

Corman, Hope, Theodore Joyce, and Michael Grossman. 1987. "Birth Outcome Production Function in the United States." *Journal of Human Resources* 22: 339-360.

Cremieux, Pierre-Yves, Denise Jarvinen, Genia Long, and Phil Merrigan. 2005. Pharmaceutical Spending and Health Outcomes. In *International Conference on Pharmaceutical Innovation*. Taiwan.

Cremieux, Pierre-Yves, Marie-Claude Meilleur, Pierre Ouellette, Patrick Petit, Martin Zelder, and Ken Potvin. 2005. "Public and Private Pharmaceutical Spending as Determinants of Health Outcomes in Canada." *Health Economics* 14 (107).

Cremieux, Pierre-Yves, Pierre Ouellette, and Caroline Pilon. 1999. "Health Care Spending as Determinants of Health Outcomes." *Health Economics* 8: 627.

Crouter, Ann, Stephanie Lanza, Amy Pirretti, W. Benjamin Goodman, and Eloise Neebe. 2006. "the O*NET Jobs Classification System: A Primer for Family Researchers." *Family Relations* 55: 461-472.

Crum, Alia, and Ellen Langer. 2007. "Mind-Set Matters. Exercise and the Placebo Effect." *Psychological Science* 18 (2): 165-171.

Dartmouth Team. 2010. Dartmouth Atlas of Health Care. http://www.dartmouthatlas.org.

Davey Smith, George, Carole Hart, David Hole, Pauline MacKinnon, Charles Gillis, Graham Watt, David Blane, and Victor Hawthorne. 1998. "Education and Occupation Social Class: Which is the More Important Indicator of Mortality Risk?" *Journal of Epidemiology and Community Health* 52: 153-160.

Department of Employment and Economic Development. 1999. *Minnesota's Most Marketable Skills*. Minnesota.

Department of Labor. 2009. O*NET Resource Center. http://www.onetcenter.org/.

Diez Roux, Ana, Sharon Merkin, Donna Arnett, Lloyd Chambless, Mark Massing, Javier Nieto, Paul Sorlie, Moyses Szklo, Herman Tyroler, and Robert Watson. 2001. "Neighborhood of Residence and Incidence of Coronary Heart Disease." *NEJM* 345 (2): 99-106.

DiMatteo, L, and R DiMatteo. 1998. "Evidence on the Determinants of Canadian Provincial Government Health Expenditures, 1965-1991." *Journal of Health Economics* 17: 211-228.

Duncan, Greg, Mary Daly, Peggy McDonough, and David Williams. 2002. "Optimal Indicators of Socioeconomic Status for Health Research." *American Journal of Public Health* 92 (7): 1151-1157.

Ellen, Ingrid, Tod Mijanovich, and Keri-Nicole Dillman. 2001. "Neighborhood Effects on Health: Exploring the Links and Assessing the Evidence." *Journal of Urban Affairs* 23 (3-4): 391-408.

Elo, Irma, and Samuel Preston. 1996. "Educational Differentials in Mortality: United States, 1979-1985." *Social Science & Medicine* 42 (1): 47-57.

Ennett, Susan, and Karl Bauman. 1993. "Peer Group Structure and Adolescent Cigarette Smoking: A Social Network Analysis." *Journal of Health and Social Behavior* 34: 226.

Erbsland, Manfred, Walter Ried, and Volker Ulrich. 1995. "Health, Health Care, and the Environment. Econometric Evidence from German Micro Data." *Health Economics* 4: 169-182.

Fillmore, Kay, William C Kerr, Tim Stockwell, Tanya Chikritzhs, and Alan Bostrom. 2006. "Moderate Alcohol Use and Reduced Mortality Risk: Systematic Error in Prospective Studies." *Addiction Research and Theory* 14 (2): 101.

Fisher, Elliott, D Wennberg, Therese Stukel, Daniel Gottlieb, F. L Lucas, and Etoile Pinder. 2003a. "The Implications of Regional Variations in Medicare Spending. Part 1: The Content, Quality, and Accessibility of Care." *Annals of Internal Medicine* 138 (4): 273.

———. 2003b. “The Implications of Regional Variations in Medicare Spending. Part 2: Health Outcomes and Satisfaction with Care.” *Annals of Internal Medicine* 138 (4): 288.

Fletcher, Jason, and Jody Sindelar. 2009. “Estimating Causal Effects of Early Occupational Choice on later Health: Evidence using the PSID.” *NBER Working Paper 15256*.

Fletcher, Jason, Jody Sindelar, and Shintaro Yamaguchi. 2008. “Cumulative Effects of Job Characteristics on Health.” *SSRN Working Paper*.

Freedman, D. 1984. “On Bootstrapping Two-Stage Least Squares Estimates in Stationary Linear Models.” *The Annals of Statistics* 12 (3): 827-842.

Gerdtham, Ulf, and Magnus Johannesson. 1999. “New Estimates of the Demand for Health: Results Based on a Categorical Health Measure and Swedish Micro Data.” *Social Science & Medicine* 49: 1325-1332.

Gerdtham, Ulf, and Christopher Ruhm. 2005. “Deaths Rise in Good Economic Times: Evidence from the OECD.” *NBER Working Paper*.

Gravelle, H.S.E., and M.E. Backhouse. 1987. “International Cross-Section Analysis of the Determination of Mortality.” *Social Science & Medicine* 25 (5): 427-441.

Gregorio, David, Stephen Walsh, and Deborah Paturzo. 1997. “The Effects of Occupation-Based Social Position on Mortality in a Large American Cohort.” *American Journal of Public Health* 87 (9): 1472.

Grossman, Michael. 1972a. “The Demand for Health: A Theoretical and Empirical Investigation.” *NBER Book*.

———. 1972b. “On the Concept of Health Capital and the Demand for Health.” *Journal of Political Economy* 80 (2): 223-255.

———. 1973. “The Correlation between Health and Schooling.” *NBER Working Paper 22*.

———. 1999. “The Human Capital Model of the Demand for Health.” *NBER Working Paper 7078*.

———. 2000. The Human Capital Model. In *Handbook of Health Economics*. Vol. 1. Handbook in Economics 17. The Netherlands: Elsevier Science.

———. 2005. “Education and Non-Market Outcomes.” *NBER Working Paper 11582*.

Gueorguieva, Ralitza, Jody Sindelar, Tracy Falba, Jason Fletcher, Patricia Keenan, Ran

Wu, and William Gallo. 2009. “The Impact of Occupation on Self-Rated Health: Cross-Sectional and Longitudinal Evidence from the Health and Retirement Survey.” *The Journals of Gerontology* 64B (1): 118-124.

Guindon, G. Emmanuel, and Paul Contoyannis. 2008. “A Second Look at Pharmaceutical Spending as Determinants of Health Outcomes in Canada.” *Health Economics* 10.1002/hec.1415 (Online 28 Oct 2008).

Hadden, Wilbur, Nataliya Kravets, and Carles Muntaner. 2004. “Descriptive Dimensions of US Occupations with Data from the O*NET.” *Social Science Research* 33: 64-78.

Hadley, Jack. 1982a. *More Medical Care, Better Health?* Urban Institute.

———. 1988. “Medicare Spending and Mortality Rates of the Elderly.” *Inquiry* 25: 485-493.

Hadley, Jack, Robert Berenson, Timothy Waidmann, and Stephen Zuckerman. 2006b. “Variations in Medical Care Spending per Medicare Beneficiary: The First Stage of an Instrumental Variable Analysis.” *The Urban Institute* (May 2006).

Hadley, Jack, Timothy Waidmann, Stephen Zuckerman, and Robert Berenson. 2011. “Medical Spending and the Health of the Elderly.” *personal communication*.

Hansen, Paul, and A King. 1996. “The Determinants of Health Care Expenditure: A Cointegration Approach.” *Journal of Health Economics* 15: 127.

Haveman, Robert, Barbara Wolfe, Brent Kreider, and Mark Stone. 1994. “Market Work, Wages, and Men’s Health.” *Journal of Health Economics* 13: 163-182.

Hayward, Mark, and Bridget Gorman. 2004. “The Long Arm of Childhood: the Influence of Early-Life Social Conditions on Men’s Mortality.” *Demography* 41 (1): 87.

House, James, James Lepkowski, David Williams, Richard Mero, Paula Lantz, Stephanie Robert, and Jieming Chen. 2000. “Excess Mortality Amoung Urban Residents: How Much, for Whom, and Why?” *American Journal of Public Health* 90 (12): 1898.

Howard, G, Roger Anderson, Gregory Russell, V Howard, and Gregory Burke. 2000. “Race, Socioeconomic Status, and Cause-Specific Mortality.” *Annals of Epidemiology* 10 (4): 214-223.

Hsu, Laura, Jaewoo Chung, and Ellen Langer. 2010. “The Influence of Age-Related Cues on Health and Longevity.” *Perspectives on Psychological Science* 5: 632-648.

Huber, Manfred, and Eva Orosz. 2003. “Health Expenditure Trends in OECD Countries, 1990-2001.” *Health Care Financing Review* 25 (1): 1-22.

Hummer, Robert, Richard Rogers, Charles Nam, and Christopher Ellison. 1999. “Religious Involvement and U.S. Adult Mortality.” *Demography* 36 (2): 273.

Johnson, Norman, Paul Sorlie, and Eric Backlund. 1999. “The Impact of Specific Occupation on Mortality in the U.S. National Longitudinal Mortality Study.” *Demography* 36 (3): 355-367.

Joyce, Theodore. 1994. “Self-Selectcion, Prenatal Care, and Birthweight among Blacks, Whites, and Hispanics in New York City.” *Journal of Human Resources* 29 (3): 762-794.

Kiiskinen, Urpo. 2003. A Health Production Approach to the Economic Analysis of Health Promotion. Univeristy of York.

Kiuila, Olga, and Peter Mieszkowski. 2007. “The Effects of Income, Education, and Age on Health.” *Health Economics* 16 (8): 781-798.

Kposowa, Augustine. 2001. “Unemployment and Suicide: A Cohort Analysis of Social Factors Predicting Suicide in the US National Longitudinal Mortality Study.” *Psychological Medicine* 31: 127-138.

Lantz, Paula, James House, James Lepkowski, David Williams, Richard Mero, and Jieming Chen. 1998. “Socioeconomic Factors, Health Behaviors, and Mortality: Results from a Nationally Representative Prospective Study of US Adults.” *JAMA* 279: 1703-1708.

Lantz, Paula, J Lynch, James House, James Lepkowski, Richard Mero, Mark Musick, and David Williams. 2001. “Socioeconomic Disparities in Health Change in a Longitudinal Study of US Adults: the Role of High-Risk Behaviors.” *Social Science & Medicine* 53: 29-40.

Levy, David, and Sandra Peart. 2008. “Inducing Greater Transparency: Towards the Establishment of Ethical Rules for Econometrics.” *Eastern Economic Journal* 34 (103-114).

Lleras-Muney, Adriana. 2005. “The Relationship Between Education and Adult Mortality in the United States.” *Review of Economic Studies* 72: 189-221.

Lockhart, Charles, Joanne Green, and Jean Giles-Sims. 2010. Do Women Legislators have a Positive Effect on the Supportiveness of States toward Older Citizens? In San Francisco, CA.

MacDonald, Leslie, A Cohen, Sherry Baron, and Cecil Burchfiel. 2009. “Occupation as Socioeconomic Status or Environmental Exposure? A Survey of Practice Among Population-based Cardiovascular Studies in the United States.” *American Journal of Epidemiology* 169: 1411-1421.

Macinko, James, Barbara Starfield, and Leiyu Shi. 2003. "The Contribution of Primary Care Systems to Health Outcomes within OECD Countries, 1970-1998." *Health Services Research* 38 (3): 831-865.

MacLeod, John, George Davey Smith, Chris Metcalfe, and Carole Hart. 2005. "Is Subjective Social Status a more Important Determinant of Health than Objective Social Status? Evidence from a Prospective Observational Study of Scottish Men." *Social Science & Medicine* 61: 1916-1929.

Marmot, M., G Smith, S. Stansfeld, C. Patel, F. North, J. Head, I. White, E. Brunner, and A. Feeney. 1991. "Health Inequalities among British Civil Servants: the Whitehall II Study." *Lancet* 337: 1387.

Marmot, Michael, Carol Ryff, Larry Bumpass, Martin Shipley, and Nadine Marks. 1997. "Social Inequalities in Health: Next Questions and Converging Evidence." *Social Science and Medicine* 44 (6): 901-910.

Martin, Stephen, Nigel Rice, and PC Smith. 2008. "Does Health Care Spending Improve Health Outcomes? Evidence from English Programme Budgeting Data." *Journal of Health Economics* 27: 826-842.

McClellan, Mark, Barbara McNeil, and Joseph Newhouse. 1994. "Does More Intensive Treatment of Acute Myocardial Infarction in the Elderly Reduce Mortality? Analysis Using Instrumental Variables." *JAMA* 272 (11): 859-866.

Miech, Richard, and Robert Hauser. 2001. "Socioeconomic Status and Health at Midlife: A Comparison of Educational Attainment with Occupation-Based Indicators." *AEP* 11 (2): 75-84.

Muntaner, Carles, Paul Sorlie, Patricia O'Campo, Norman Johnson, and Eric Backlund. 2001. "Occupational Heirarchy, Economic Sector, and Mortality from Cardiovascular Disease Among Men and Women: Findings from the National Longitudinal Mortality Study." *Annals of Epidemiology* 11 (3): 194-201.

Murray, Christopher, Sandeep Kulkarni, Catherine Michaud, Niels Tomijima, Maria Bulzacchelli, Terrell Iandiorio, and Majid Ezzati. 2006. "Eight Americas: Investigating Mortality Disparities Across Races, Counties, and Race-Counties in the United States." *Public Library of Science* 3 (9): 1513.

Newhouse, Joseph, and Lindy Friedlander. 1980. "The Relationship between Medical Resources and Measures of Health: Some Additional Evidence." *Journal of Human Resources* 15 (2): 200-218.

Nocera, Sandra, and Peter Zweifel. 1998. "The Demand for Health: An Empirical Test of the Grossman Model using Panel Data." *Developments in Health Economics and Public Policy* 6: 35-49.

Oakes, J. Michael. 2004. "The (Mis)estimation of Neighborhood Effects: Causal Inference for a Practicable Social Epidemiology." *Social Science and Medicine* 58 (10): 1929-1952.

Or, Zeynep. 2000. "Determinants of Health Outcomes in Industrialized Countries: A Pooled, Cross-Country, Time-Series Analysis." *OECD Economic Studies* (30): 50.

Philadelphia Fed. 2010. State Coincident Indexes. http://www.philadelphiafed.org/research-and-data/regional-economy/indexes/coincident/.

Pincus, Theodore, Robert Esther, Darren DeWalt, and Leigh Callahan. 1998. "Social Conditions and Self-Management are More Powerful Determinants of Health than Access to Care." *Annals of Internal Medicine* 129 (5): 406-411.

Pritchett, Lant, and Lawrence Summers. 1996. "Wealthier is Healthier." *Journal of Human Resources* 31 (4): 841-868.

Rajkumar, Andrew, and Vinaya Swaroop. 2008. "Public Spending and Outcomes: Does Governance Matter?" *Journal of Development Economics* 86: 96-111.

Rehavi, M. Marit. 2007. *Sex and Politics: Do Female Legislators Affect State Spending?*

Roberts, Brent, Nathan Kuncel, Rebecca Shiner, Avshalom Caspi, and Lewis Goldberg. 2007. "The Power of Personality. The Comparative Validity of Personality Traits, Socioeconomic Status, and Cognitive Ability for Predicting Important Life Outcomes." *Perspectives on Psychological Science* 2 (4): 313-345.

Rothberg, Michael, J Cohen, Peter Lindenauer, Judith Maselli, and Andy Auerbach. 2010. "Little Evidence of Correlation Between Growth in Health Care Spending and Reduced Mortality." *Health Affairs* 29 (8): 1523-1531.

Sacker, Amanda, David Firth, Ray Fitzpatrick, K Lynch, and Mel Bartley. 2000. "Comparing Health Inequality in Men and Women: Prospective Study of Mortality 1986-96." *British Medical Journal* 320: 1303-1307.

Salas, Christian. 2002. "On the Empirical Association Between Poor Health and Low Socioeconoimic Status at Old Age." *Health Economics* 11: 207-220.

Schaffer, M. E. 2007. *XTIVREG2: Stata module to perform extended IV/2SLS, GMM and AC/HAC, LIML and k-class regression for panel data models*. http://ideas.repec.org/c/boc/bocode/s456501.html.

Sen, Anindya. 2005. "Is Health Care a Luxury? New Evidence from OECD Data." *International Journal of Health Care Finance and Economics* 5: 147-164.

Shang, Baoping, and Dana Goldman. 2007. "Does Age or Life Expectancy Better Predict Health Care Expenditures?" *Health Economics* 17: 487-501.

Shaw, James, William Horrace, and Ronald Vogel. 2005. "The Determinants of Life Expectancy: An Analysis of the OECD Health Data." *Southern Economic Journal* 71 (4): 768-783.

Sickles, Robin, and Abdo Yazbeck. 1998. "On the Dynamics of Demand for Leisure and the Production of Health." *Journal of Business & Economic Statistics* 16 (2): 187-197.

Sindelar, Jody, Jason Fletcher, Tracy Falba, Patricia Keenan, and William Gallo. 2007. "Impact of First Occupation on Health at Older Ages." *NBER Working Paper Series* (Working Paper 13715).

Skinner, Jonathan, Amitabh Chandra, Douglas Staiger, Julie Lee, and Mark McClellan. 2005. "Mortality after Acute Myocardial Infarction in Hospitals that Disproportionately Treat Black Patients." *Circulation* 112: 2634-2641.

Skinner, Jonathan, Douglas Staiger, and Elliott Fisher. 2006. "Is Technological Change in Medicine Worth It? The Case of Acute Myocardial Infarction." *Health Affairs. Suppl Web Exclusive* 7 February: W34-W46.

Skinner, Jonathan, and J Wennberg. 1998. "How Much is Enough? Efficiency and Medicare Spending in the Last Six Months of Life." *NBER Working Paper 6513*.

Smith, J. 1999. "Health Bodies and Thick Wallets: The dual Relation between Health and Economic Status." *Journal of Economic Perspectives* 13 (2): 145-166.

Smith, T, and Cynthia Campbell. 2006. "The Structure of O*NET Occupational Values." *Journal of Career Assessment* 14: 437-448.

Sorlie, Paul, Eric Backlund, and Jacob Keller. 1995. "US Mortality by Economic, Demographic, and Social Characteristics: The National Longitudinal Mortality Study." *American Journal of Public Health* 85 (7): 949.

Sorlie, Paul, Sean Coady, Charles Lin, and Elizabeth Arias. 2004. "Factors Associated with Out-of-Hospital Coronary Heart Disease Death: the National Longitudinal Mortality Study." *Annals of Epidemiology* 14 (7): 447-452.

Sorlie, Paul, Norman Johnson, Eric Backlund, and Douglas Bradham. 1994. "Mortality in the Uninsured Compared with that in Persons with Public and Private Health Insurance." *Archives of Internal Medicine* 154 (14): 2409-2416.

StataVersion 11.1. 2010. StataCorp, College Station, TX. http://www.stata.com/.

Stock, J.H., and M. Yogo. 2005. “Testing for Weak Instruments in Linear IV Regression.” *NBER Technical Working Paper 284*. http://www.nber.org/papers/TO284.

Subramanian, S. V, Ichiro Kawachi, and Bruce Kennedy. 2001. “Does the State You Live In Make a Difference? Multilevel Analsysis of Self-Rated Health in the US.” *Social Science and Medicine* 53: 9-19.

Terza, Joseph, A Basu, and Paul Rathouz. 2008. “Two-Stage Residual Inclusion Estimation: Addressing Endogeneity in Health Econometric Modeling.” *Journal of Health Economics* 27: 531-543.

Thornton, James. 2002. “Estimating a Health Production Function for the US: Some New Evidence.” *Applied Economics* 34: 59-62.

Torgerson, D. J, and Alan Maynard. 1998. “International Comparisons of Health Care Expenditure: A Dismal Science?” *QJM: An International Journal of Medicine* 91: 69-70.

U.S. Census Bureau. 2010. National Longitudinal Mortality Study (NLMS). http://www.census.gov/did/www/nlms/index.html.

Volkers, Anita. 2005. “How Occupationsl Position and Educational Level Modify Each Other’s Relationships with Poor Health Status and Morbidity.” *European Journal of Public Health* 15.

Vork, Andres. 2000. An Empirical Estimation of the Grossman Health Demand Model using Estonian Survey Data. Working Paper, University of Tartu.

Wada, Roy. 2010. OUTREG2: Stata module to arrange regression outputs into an illustrative table. *Statistical Software Components S456416*. http://ideas.repec.org/c/boc/bocode/s456416.html.

Wagstaff, Adam. 1986a. “The Demand for Health: A Simplified Grossman Model.” *Bulletin of Economic Research* 38 (1): 93-95.

———. 1986b. “The Demand for Health. Some New Empirical Evidence.” *J* 5: 195-233.

———. 1993. “The Demand for Health: an Empirical Reformulation of the Grossman Model.” *Health Economics* 2: 189o-198.

Warren, John Robert, and Hsiang-Hui Kuo. 2003. “How to Measure ‘What People do for a Living’ in Research on the Socioeconomic Correlates of Health.” *AEP* 13 (5): 325-334.

Weiss, Michael. 2000. *The Clustered World: How We Live, What We Buy, and What It All Means About Who We Are*. 1st ed. Little, Brown and Company.

Wennberg, J. 2004. "Practice Variation and Health Care Reform: Connecting the Dots." *Health Affairs. Suppl Web Exclusive*: VAR140-4.

Wennberg, J, Elliott Fisher, and Jonathan Skinner. 2002. "Geography and the Debate over Medicare Reform." *Health Affairs*: W96 (Web Exclusive).

Wilper, Andrew, Steffie Woolhandler, Karen Lasser, Danny McCormick, David Bor, and David Himmelstein. 2009. "Health Insurance and Mortality in US Adults." *American Journal of Public Health* 99 (12): 1-7.

Zhou, Xueguang. 2005. "The Institutional Logic of Occupational Prestige Ranking: Reconceptualization and Reanalyses." *American Journal of Sociology* 111: 90-140.

Curriculum Vitae

Kenneth Lee earned a Bachelor of Science degree in Chemistry in 1977 from Carleton College and a Masters in Business Administration from the University of Pittsburgh Executive MBA program in 1995. After working at the Software Engineering Institute at Carnegie Mellon University and Lockheed Martin Mission Systems Division, Ken has worked the last nine years at the MITRE Corporation as a Principal, Information Systems Engineer.