THE EFFECT OF EDUCATION ON MORTALITY: EVIDENCE FROM DELAWARE

By

Anli Gu

A thesis submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Master of Science in Agricultural Economics

Winter 2007

UMI Number: 1440586

UMI Microform 1440586

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

THE EFFECT OF EDUCATION ON MORTALITY: EVIDENCE FROM DELAWARE

By

Anli Gu

Approved: ____
Thomas W. Ilvento, Ph.D.
Professor in charge of thesis on behalf of the Advisory Committee

Approved: ____
Thomas W. Ilvento, Ph.D.
Chair of the Department of Food and Resource Economics

Approved: ____
Robin W. Morgan, Ph.D.
Dean of the College of Agriculture and Natural Resources

Approved: ____
Daniel Rich, Ph.D.
Provost

ACKNOLEDGEMENTS

First, I would like to acknowledge Dr. Ilvento, the department chair, and also my advisor, for providing me funding from September 2004 till May 2006, for letting me choose my thesis topic, for giving me constructive advice and suggestions on my thesis, and for reading and revising my thesis.

I would also like to acknowledge the other two committee members: Dr. Bernard and Dr. Awokuse, for providing me insightful and valuable advice on this work. I'd like to thank Dr. Pesek since Dr. Pesek has been an important influence in this thesis work. Thank for all of the committee members for having me in this process of completing a thesis work and I enjoy this experience.

There are other people that have supported me during the last two years: the FREC graduate students, the OR and the STAT graduate students, my friends, my grandmother and my family; thank you for supporting me and walking me through this process.

Lastly, the very important thing that has made this work possible is the understanding and eternal support from my husband, Zhaolin, and this thesis is also devoted to him.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The object of this research is to examine the effect of education on mortality using age-race-sex-education specific death rates. From 1989, educational attainment has been added to standard death certificate. Combining death counts from death certificates and population estimates derived from the Census Public Use Microdata Sample (PUMS), I computed death rates by age group, by race, by sex, and by educational attainment. Two time periods are examined: the 1990 and the 2000.

By fitting Poisson distributed death rates into a Log-Rate model, I tested the main effect and the mortality differentials. Empirical analysis confirms that mortality risk is higher for less educated population; empirical analysis also shows that men face higher mortality risk than women and blacks face higher mortality risk than whites. There exists race difference between blacks and whites on the impact of education in reducing morality risk. This study confirms the importance of education and may contribute to similar mortality studies using Poisson regression models.

Chapter 1

1. INTRODUCTION

1.1 Reasons for Study

"Perhaps no component of social welfare is as highly valued as life itself" (Preston, 1977, p. 1). Decreasing the mortality rate is a universal goal for human society, and governments are expected to spend all that they can to improve the health and longevity of its citizens (Kitigawa, 1977). In Delaware, each year, about one fourth of the annual budget is spent on health and social services (Financial Overview Fiscal year 2002 and various years). From 1989 to 2002, the life expectancy at birth for Delawareans has increased from 74.3 years to 77.3 years and the crude death rate has been relatively stable.

The study of mortality is important because mortality research in recent decades has raised important policy questions for public health, such as restricting cigarette advertising, regulating smoking in public places, obesity and diet, and restricting alcohol sales. Research has also shown a reverse relation between education and mortality (Kitagawa and Hauser, 1973; Christenson and Johnson, 1995; Mark, 1999; Molla, Madans and Wagener, 2004; Muney, 2005). Since expenditures for education accounted for about 40% of Delaware's state annual budget (Financial Overview Fiscal Year 2002) and much work has been done to show the value of education for employment and lifetime earning, it is also reasonable to determine how education might influence life chances and mortality.

Perhaps more importantly, we can now more easily study the relationship between mortality and education. Since 1989, education attainment has become a mandatory item on the standard death certificate in Delaware and other states. This change enables researchers to investigate the effect of education on mortality on a group level by attribute (age, gender, and race) using data from death certificates.

1.2 Background Information

It is important to have a clear understanding of mortality and how mortality is measured. The followings are the accepted demographic definitions of mortality. Death is defined as the permanent disappearance of any evidence of life at any time after a live birth. The Crude Death Rate is expressed as the total deaths per 100,000 populations (Delaware Vital Statistics Annual Report 2002). The crude death rate is misleading if one wants to make comparisons among different populations where the age-race-sex distributions of the populations are not similar. For example, if area A has a higher percentage of elderly population than other areas, then comparing the crude death rate of area A with that of other areas will probably lead to incorrect conclusions. To this point, specific mortality rates are thought to be a better measure of mortality for comparative purposes. For example, an age-specific mortality rate is the number of deaths for a specific age group per 100,000 populations in the same age group. It is possible to generate multiple categories of specific rates, such as an age-gender-race specific rate.

In the United States, state laws require that demographic information on death certificate be recorded by funeral directors, who are asked to report the characteristics of the decedent on the basis of information from the next of kin. Educational attainment has been added to the standard death certificate since 1989, and is recorded on the death certificate as total years of schooling completed by the decedent.

Educational attainment refers to the highest level of education completed in terms of the highest degree or the highest level of schooling completed. The Census has traditionally recorded and tabulated data on education attainment for the population 25 years and over. Twenty-five and older is used as the threshold because usually one has finished his highest level of education after 25 years old, and therefore is relatively fixed at that point.

In Delaware, the rate of unknown education status listed on death certificate was relatively higher (at approximately 9.39%) in the first year this regulation was launched (1989) as funeral directors began to implement the new regulation. However, it dropped to 4.85% by 1990, and has been gradually declining to around 2% since then (see Table 1.1 and Figure 1.1). In national mortality statistics, information on educational attainment is missing for 4.4%. One explanation for missing information on education is that funeral directors have difficulty in eliciting or obtaining accurate information on decedents' educational attainment (Hahn, Wetterhall, Gay, Harsharger, Bernett, Parish and Orend, 2002). By law, funeral directors are responsible for completing and filling the death certificates. The funeral directors obtain the decedents' personal information form the best source available, usually the next of kin. Some researchers suggest that there may exist over-reporting of decedents' educational attainment. (Christenson and Johnson, 1995). However, researchers lack of a direct means to test this kind of misreporting.

Before I present a literature review regarding adult mortality, it is essential to understand the official definition of death rates, the meaning of educational attainment in Census, and how education is added into death certificate, since these issues will affect how death rates are computed, how education would be categorized, and what are the possible data limitation.

1.3 Past Trends in Delaware

From 1989 to 2002, the crude death rate in Delaware did not fluctuate much; however, in 8 out of these 14 years, Delaware's crude death rate was higher than the nation's level. The average crude death rates of the nation and Delaware in this period were 8.674 and 8.694, respectively. The maximum of the crude death rate of nation and Delaware were 8.812 and 8.939, respectively. From these figures, we can see that Delaware has a slightly higher crude death rate than the national level (see Table 1.2 and Figure 1.2).

Over the period from 1980 to 2000, the age structure changed in Delaware. The population of people aged 64 years and over in Delaware has been increasing steadily over this period. The population of people aged 64 years and over accounted for 10% of the total population of Delaware in 1980 and for 13% in 2000, and the respective figures for the nation level are: 11.3% in 1980 and 12.4% in 2000. From these figures we can see that Delaware has been aging during the last two decades and Delaware has been aging faster than the national level. This aging trend of Delaware has greatly contributed to an increase in the elderly female population. Among the elderly population, the population of females significantly exceeds that of males. For instance, in 2000, there were 8700 males and 12693 females for those 75 to 79; there were 4784 males and 8362 females for those 80 to 84; and there were 3013 males and 7385 females for those 85 and over (Delaware Vital Statistics Annual Report, various years.). Much of the increase in the elderly in Delaware occurred in

Sussex County, which had 18.5 percent of its population 65 years and older (1990 Census of Population Social and Economic Characteristics Delaware.).

During the same period, Delawareans have been significantly improving their education levels. The percentage of people who completed only elementary school in total population aged 25 years and over dropped from 7.24 percent in 1990 to 5% in 2000, while the percentage of people who completed 4 year college in total population increased from 19.4 percent in 1990 to 24.4 percent in 2000.

Figure 1.3 and figure 1.4 show a rough pattern of the mortality differences on age, gender, race (simplified into two categories, which are white and black), and education for the year 2000. In general, death rates: increase by age; are invariably higher for black than for white at each age group. For the group of black people aged 35 to 44, death rates are highest for black male and black female who have less than 9 years' education (see figure 1.4). For the group of black people aged 65 to 74, death rates are highest for black male and black females who have 12 years' education, and lowest for black male and black female who have more than 16 years' education (see figure 1.4). For the group of white people aged 35 to 44, death rates are highest for white male who have less than 9 years' education and white male who have 12 years' education, and lowest for white male and white female who have more than 16 years' education (see figure 1.3). For the group of white people aged 65 to 74, death rates are highest for white male who have 12 years' education and white male who have less than 9 years' education, and lowest for white male and white female who have more than 16 years' education (see figure 1.3).

1.4 Purpose of Study

Since mortality studies have important applications in policy making and affect people's daily life and education expenditure is the biggest part of the state's annual budget, I am interested in looking at the linkage between education and mortality, specifically, the effect of education on mortality. The purpose of this thesis is to empirically examine the effect of education on adult mortality and mortality differentials in Delaware from 1989 to 2002. I use grouped data to compute age-race-sex-education specific mortality rates, and then employ a log-rate model to investigate the effect of education. More specifically, I will look at the differences between the model of 1990 death rates and 2000 death rates. Included in the analysis is an examination of the differences of education's effect on mortality between males and females, between white and black, and between younger people and older people.

1.5 Organization of Chapters

This study is organized as follows: in the first chapter, background information regarding the study of mortality and purpose of this study are introduced. The second chapter will review previous mortality studies from the socioeconomic mortality risk factors perspective and a methodological perspective. In the third chapter, I will explain the source of data, the process of computing age-race-sex-education mortality rates, the data issues, the log-rate model, and other related methods used in this study. In the fourth chapter, I will present and interpret the results of the empirical analysis of data for 1990 and 2000. In the last chapter, I will summarize the process and finding of this study and discuss possible further research steps.

Chapter 2

2. LITERATURE REVIEW

This chapter gives a summary of previous mortality studies on education and other socioeconomic status (SES) variables, including theory and methods of analysis. The first section will focus on the effect of education on mortality. The second section will give a brief summary on findings from mortality studies on other SES variables and will introduce the differentials on the effect of education on mortality, for example, the race differential on the effect of education on mortality, the gender differential, and the age differential. The third section will present literature regarding the implication of education on economy and society. The fourth section will review the statistical methods used in previous mortality studies and will focus on the Log-rate model. The last section will summarize the literature review.

2.1 Mortality Studies on the Impact of Education

The major socioeconomic status (SES) indicators used in previous mortality studies are education, income level, occupation, and marital status. There is a debate over the use of education or income as the best SES indicator (Hummer, Rogers, Eberstein, 1998). Christenson and Johnson (1995) argue that education is the optimal measure because educational status is usually fixed early in life, is applicable to all people, is essential in determining employment and income level, and is influential to one's health behavior (Christenson and Johnson, 1995). Considerable previous studies covering different countries and different time periods suggest mortality varies inversely with the level of education. The following section will review some classic mortality studies on the effect of education on mortality.

In a seminal work, Kitigawa and Hauser (1973) used 1960 Matched Records data at the individual level to estimate cause-specific mortality by education, family income, occupation, marital status, and nativity. Most importantly, they focused their analysis on socioeconomic factors such as income and education. For instance, they found that education and income exhibited strong inverse relationships with mortality. Using mostly tables and graphs, they demonstrated that the reverse association between education and mortality is greater among the younger population than among the older population, and greater among women than among men.

Christenson and Johnson (1995) used Michigan's 1989-1991 death certificate data combined with 1990 Census data to compute age-race-sex-education specific death rates and then examined the effect of educational attainment on adult mortality. Their data were group level data which influenced their choice of statistical analysis. Their major findings included: after controlling for age, race, and sex, the mortality rate for those with only a primary education is 15% higher than that of those with a secondary education; the mortality rate of those with a secondary education is 46% higher than that of those with a higher education level. These findings suggest that promoting higher education will lower mortality rates more than reducing the high school dropout rate.

In addition, Christenson and Johnson investigated the age differential, race differential, and sex differential on the effect of education on mortality rates. They found that the negative effect of education on mortality rate diminishes with age: the negative effect of education on mortality is stronger at younger age than at older age. They also found the effect of education on mortality differed between whites and blacks, however this mortality difference only occurs at the lower level of education. In addition, they found moving from primary to secondary education will reduce the mortality rate by 10% more for women than for men and that moving from secondary to postsecondary education will reduce mortality rate by 15% more for men than for women. Therefore, the effect of education on mortality rate is slightly greater for men than for women.

Molla, Madons, and Wagener (2004) confirmed the strong and reverse relation between education and mortality using data of age-sex-education specific death rates from 46 US states in 1998. Although this study does not employ a multivariate statistical model, the mortality rate patterns show a strong association between education and mortality. They found that mortality rates are lowest for those states with the highest education level, and with few exceptions, highest for those with the lowest education levels.

Muney's (2005) study of the effect of education on mortality established a causal relationship between education and mortality. Individual level NHEFS data (referring to the NHANES I Epidemiologic Followup Study; the NHANES refers to the National Health and Nutrition Examination Survey conducted by Centers for Disease Control and Prevention) and aggregated Census data and aggregated NHEFS data were used in this study. The variables used in this study included the following variables: education, year dummy variable and sex; and the individual's state-of-birth characteristics: percentage of urban population, percentage of foreign population, percentage of black population, percentage of population employed in manufacturing, annual manufacturing wage, value of farm per acre, per capita number of doctors, per capita expenditure of education, and number of school buildings per sqr. mile. Muney investigated the effect of compulsory schooling laws and showed that there is a large causal effect of education on mortality: an additional year of education lowers the probability of one dying in the next 10 years by approximately 3.6% and one more year of compulsory schooling decreases mortality after age 35 by about 3%. This effect of education on mortality is stronger and larger than what was found in previous studies.

2.2 Mortality Studies on Other Socioeconomic Factors

Kitigawa and Hauser (1973) used multiple SES indicators and cause-of-death categories to provide the foundation of mortality studies and has dominated the America literature of mortality study ever since. In their study, they used the 1960 Matched Records study to conduct their research. Since the data were individual level data, they were able to look into many SES variables, such as education, income, and occupation. Except for education, the other most commonly used SES variables in mortality studies are: family income, occupation~~s~~, accumulated income (wealth), regions, religions, and marital status.

Income: Some researchers regard income the optimal SES indicator in mortality study because income affects one's accessibility to medical care, nutrition, and quality of life (Hummers; Rogers; Eberstein, 1998). Deaton and Paxson (1999) used both individual data and a panel of aggregated birth-cohorts to conclude that people whose family income was less than 50,000 in 1980 could expect to live about 25% less than people whose family income was greater than 50,000. Rogot, Sorlie, Johnson, and Schmitt (1992) showed that People aged 25 whose family income was $5,000 or less could expect to live 10 years less than those whose family income was more than $50,000.

Occupation: Investigations on occupational levels measured by social class and job hazards showed that there is a negative relation between occupational levels and mortality rates. Specific groups, such as the unemployed, service workers and homemakers demonstrated substantially higher mortality risk in comparison to workers classified as professional. For instance, Tuckman, Youngman and Kreizman (1965) used data obtained from a national report of mortality by occupation and cause of death among men 24-60 years of age in United States in 1950 to investigate the relation between occupational types and mortality rates. In their study, data is presented as a Standardized Mortality Ratio for 72 causes of death and 26 occupation groups. They concluded that the mortality rate increased as socioeconomic status decreased from professional workers to unskilled workers. Moore and Hayward

(1990) investigated in the mortality risk for different occupation types. Their data is a cohort of men aged 55 years old and older in United States for the period of 1966 to 1983. Their study indicated that men whose longest occupation ranked low in substantive complexity tended to experience higher mortality.

Marital Status: Hu and Goldman (1990) have shown that married people are healthier and hence they have a lower mortality. Data used in Hu and Goldman's study were obtained from published Vital Statistics and Census for particular countries as well as from the *United Nation Demographic Yearbook*. Death rates are calculated and broken down by countries, by selected years (treated as a continuous variable), by age group (categorized into five groups: 20 to 24, 25 to 34, 35 to 44, 45 to 54, and 55 to 64.), and by marital status (categorized into four groups: married, divorced, widowed, and single). Hu and Goldman stated that men benefited from marriage more than women although, the reason for the marital status differential on mortality is relatively less clear.

Sex differential: One of the most striking trends in mortality has been the continuous divergence of male and female mortality rates. In the United States, for instance, the life expectancy difference between males and females increased from 3.5 years in 1930 to 5.4 years in 2000 for all races (National Vital Statistics Report, 2006). It seems that men suffer a higher mortality than women, especially among age group over 65 years old (Kitigawa and Hauser 1973).

Case and Paxson (2005) showed that men with specific health conditions are more likely to be hospitalized and to die than women with the same conditions.

The data for this study were drawn from the National Health Interview Survey (NHIS) from 1986 to 2001, and from the associated NHIS Multiple Cause of Death Public Use Data File, which contains information on the deaths of individuals who were surveyed between 1986 and 1994. The NHIS survey collects information on self-rated health, chronic health conditions, the use of health care services, and socio demographic characteristics. The final sample consists of 237,140 men and women aged 45-84 whose vital status is known. The determinants of sex difference on mortality were also interesting to the researchers. There is persuasive evidence to show that biological differences favor lower female mortality. Some studies shed some light on the environmental factors which are related to the divergence of male mortality and female mortality. Preston (1970) concluded that smoking is the most promising explanation of the excessive older male mortality: 50.7 percent of the males that smoked heavily would expect to die as a result of cigarette smoking; while only 12.5 percent of the females with same amount daily cigarette consumption would expect to die of this habit.

Sex differential on the effect of education on mortality: Kitigawa and Hauser (1977) found that the effect of education on mortality is larger among the younger population than among the older population, and is larger among women than among men. Christenson and Johnson (1995) studied the education inequality in adult mortality and they argued that women benefit less than men from improved educational status, for subjects holding the same education level. This finding contradicted the previous result.

Race Differential: In general, the non-white population has higher mortality than the white population. Hayward and Heron (1999) examined the racial differences on life expectancy and health condition. They estimated the mortality rates for each racial group by sex and by five-year age groups, following the approach used by National Center for Health Statistics to calculate national life tables. In order to derive the mortality rates, they collected population estimates and number of deaths separately. The population estimates were from PUMS (Public Use Microdata Sample) and the number of deaths was from Mortality Detail File (US Department of Health and Service, 1993). The mortality counts were obtained from death certificates. In the death rate analysis, death rates are broken down by age groups (categorized into 18 groups, 5 years an interval starting from age zero), by gender, and by race (categorized into five major racial groups). In the life expectancy analysis, active life expectancy is also broken by the same classifications. They found that Asian Americans live both longer and healthier than other races; Native Americans live moderately longer but tend to have poorer health; Blacks have the shortest life expectancy and have a high level of chronic health impairment.

Some studies look at the determinants of racial differential on mortality. Rogers (1992) looked into racial differences and argued that if blacks improve their socioeconomic status such as increasing income, increasing their propensity to marry, maintaining marriages, reducing family size, then their life expectancies should increase and the racial difference in mortality should decrease. Rogers used two data sets, one is the 1986 National Health Interview Survey (NHIS) and the other is the 1986 NMFS, and he linked the two data sets to get the sample consisting of individuals who were surveyed in the1986 NHIS and known to have died in the 1986

NMFS. NMFS is the National Mortality Followback Survey that contains 18733 decedents aged 25 years and over. Information of NMFS were obtained from death certificates and data on socio, demographic, and economic characteristics from informants' responses (often surviving spouses). NHIS is the National Health Interview Survey conducted almost annually. The NHIS data included 65,052 individuals, 37,917 of them were aged 25 years and over. Rogers concatenated the two data sets into one data set of individuals in 1986 who either survived or died sometime during the year. So, the NMFS was scaled down to represent the number of death expected in NHIS. Results from logistic regression analysis concluded that mortality is significantly higher for blacks than for whites, for older than for younger adults, and for males than for females; when controlling for age, sex, marital status, family size and income, the racial difference in mortality diminishes.

Bond and Stephanie (2003) concluded that lower asset holding among blacks affects their survival prospects because wealth ultimately buffers against the risk of premature mortality and reduces the race differential on mortality. They used data drawn from the 1992 wave of the Health and Retirement Study (HRS), linked the data to deaths from the National Death Index (NDI) for years 1992–1995 and the HRS Tracker Death File Version 2.0 for the years 1996–1998. The deaths reported in the HRS Tracker File for 1996–1998 were not linked to the NDI but were identified through spousal and other 672 Social Science Quarterly family member reports. The final sample size for their study was 8,633 records.

Potter (1991) claimed that the residential differential of the black population has a strong effect on the total life expectancy differential as it acts through the racial homicide differential. The mortality data used in Potter's study were taken from National Center of Health Statistics mortality detail tapes for 1979 to 1980. Deaths of males aged 0 to 85 plus (in five-year age group) were aggregated by race (white and black) for 16 underlying causes of death; population counts were from 1980 Census and were used as denominators for calculating death rates in life tables. The dependent variable in Potter's analysis is the total white-black life expectancy difference. Potter selected five independent variables to explain the white-black life expectancy variation: the difference of white-black in the proportion of persons living 75 percent below the poverty line; the difference of the race-specific proportion of the male labor force that are unemployed; the probability of black-black residential contact (considered an important index for racial isolation); a dummy variable for residence in the south and a measure of racial difference in age structure (calculated by subtracting the proportion of black males older than 64 from the corresponding proportion for white males).

Race differential on the effect of education on mortality: Christenson and Johnson (1995) found that the mortality rate is nearly 27.5 percent higher for blacks than for whites when moving from primary to secondary education and the mortality rate is about 3.1 percent higher for blacks than for whites when moving from secondary education to postsecondary education. This finding implies that the race differential depends upon the level of education and that the effect of education on mortality diminishes with the level of education.

2.3 The Economic Implication of Education

In order to understand how education affects mortality, it is helpful to look into education from the economic point of view. Researchers believe education can be seen as capital one owns and human capital can increase productivity, in addition, researchers find that education can reduce crimes and save social cost. These economic implications of education have been investigated from different research perspectives. This section lists some of these findings from the human capital perspective, from the productivity perspective, and from the social welfare perspective.

The term “human capital” was first discussed by Pigou (1928) then was associated with Nobel Prize winner Gary Becker. In Gary Becker’s book entitled Human Capital, published in 1964, human capital is similar to "physical means of production", for example, factories and machines: one can invest in human capital (via education, training, medical treatment) and one's income depends partly on the rate of return on the human capital one owns. Thus, human capital is a stock of assets one owns, which allows one to receive a flow of income, which is like interest earned. Becker (1962) presented the following simple equation which relates the costs, return and rate of return:

$$k = \bar{r} * C \tag{2.1}$$

In which k is the total return on the whole investment, $\bar{r}$ is the average rate of return and C is the total foregone earnings.

Becker (1992) stated that schooling raises earnings and productivity mainly by providing knowledge, skills and a way of analyzing problems. An alternative view states that degrees and education convey information about the underlying abilities, persistence, and other valuable traits of people.

From the productivity perspective, Turcotte and Rennison (2004) provided cross-sectional evidence for Canada that university education and computer use are associated with higher productivity. Of particular interest, they found that computer use can augment the qualifications of lower-skilled workers and make firms equally well-off in terms of productivity gain. They used the 1999 Canada Workplace and Employee Survey (WES) data to: first, examine how the use of technology is related to the level of productivity in Canadian firms, controlling for a number of firm- and worker-specific characteristics; second, investigate whether the productivity benefits are indeed greater when technology use is combined with investments in human capital such as education and training; third, examine to what extent the gains in productivity associated with ICT (Information and Communication Technology) and human capital are reflected in better wages for workers. Variables included: technology use (presented as the share of workers using computers); education (presented as the percentage of workers with college degree); training (presented as on-job training, in-class training, and computer training); and some firm characteristics. Empirical analysis shows that the link between education and the level of productivity is robust, with a 10 percentage point increase in the share of workers with a university degree generating a productivity return of about 2 per cent, both with and without the control variables. They found that wage return to workers differs from the productivity return to firms, however, they used WALD test on equality of estimated coefficients to show that the two coefficients are statistically equal, suggesting that higher wage return is associated with higher productivity.

From the social welfare perspective, Lochner and Moretti (2001) found that a one percent increase in high school graduation rates would save approximately $1.4 billion in costs associated with incarceration costs, or about $2,100 for each male high school graduate.

Macallair (1998) argued that it is more expensive to imprison an inmate than fund a college student since prisons and universities have the same target population and the budget on prisons and budget on universities is a tradeoff. “Prisons and universities generally occupy the portion of a state's budget that is neither mandated by federal requirements nor driven by population - like Medicare or K-12 education. Because they dominate a state's discretionary funds, prisons and universities must ‘fight it out’ for the non-mandated portion of the state's budget”; Daniel Macallair presented some interesting statistics: “According to the California Department of Corrections, it currently costs approximately $22,000 to imprison one inmate for a year. With an annual average cost of $4,022 in tuition fees, approximately 5 students could attend the University of California for the cost of housing one inmate”; “A defendant sentenced to life under "Three-Strikes" will cost a minimum of $467,500. This translates into approximately 116 students who could have attended a University of California campus.”

Usher (1997) stated that education conveys a civic externality and when schooling is incorporated into an “anarchy” model where people can choose to be farmers or bandits, schooling inculcates distaste for a life of crime. The assumption of this model is that there exists a society with essentially identical people and free mobility of labor between two occupations: the honorable and productive farmers and the dishonorable and unproductive bandits. In a Cost-Utility equilibrium, when education is incorporated into the model, the "propensity to banditry" will decrease the amount of grain (products from the farmers but stolen by bandits) that the bandits consume, which leads to the conclusion that: schooling inculcates distaste for a life of crime.

Although there is a rich body of empirical research describing the influences of biological, demographic, socioeconomic factors on mortality, there is yet no theoretical model synthesizing the vast amount of literature. Mortality studies are strictly limited by data available; therefore in the next sections I will present some typical mortality studies and highlight each study's data type and data source. The methods are associated with certain type of data, for example, Logistic regression models use individual level data, Cox Hazard Proportional model uses individual level data, Survival analysis uses both individual and grouped data, and Log-rate model uses grouped data.

In order to fit data into Logistic model and Cox Hazard model, researchers need individual level data. Individual level data is typically obtained from national level health surveys and their correspondent follow-up surveys so that the living and the dead are separated and the probability of dying can be estimated. For example, the 1993 National Mortality Follow-back Survey Provisional Data identified the dead from the living. The following information of interviewee are surveyed: age, sex, race, education, occupation, veteran status, marital status, alcohol or drug treat, cancer cause, lived alone, times of doctor visits, psychiatrist visit, pay to Medicare, Medicare covered, problems in getting treatment, getting help at home, paying bills and paying transportation, disease history (high blood pressure, heart attack, stroke, chest pain, Alzheimer, diabetes, cancer, asthma, etc), specific treatments, DUI history, drinking habit, history of specific drug used, life style (going out-door, diet, etc),and income (decedent's income, family income, retirement, interest, etc). Note that in the National Mortality Follow-back survey, for the sake of confidentiality, the interviewee's state of residence is collapsed into region level.

There is not a model synthesizing all the determinants of mortality; mortality, in its nature, is mysterious, unpredictable and unforeseeable. However, researchers from biological and demographical aspects are trying to control death and model the most risky factors in determining death. If there were no data limitation, I would like to examine the effect of three sectors of variables: disease history factors, socioeconomic factors and demographic factors. Specifically, I will look into these variables: age, race, sex, education, longest occupation, income, family wealth, marital status, veteran status, specific cause of death, disease history, family disease history, health habit, life style, social uncertainty and accessibility to Medicare.

In previous sections, I presented some literature that covered some of the above variables, for example, the effect of income on mortality, the effect of occupation type on mortality risk, the mortality differences of marital status, the effect of education on mortality, the racial difference and gender difference on mortality. Literature of cause-of-death studies and disease studies are an important aspect in mortality studies however they are not in my particular interest.

2.4 Methods and Data in Mortality Studies

Mortality is studied from several different perspectives: from the cause of death perspective; from the SES perspectives; from the mortality differentials perspectives, such as age differential, sex differential and race differential. Different perspectives require researches to employ different methods, generally because the characteristics and type of the data vary.

Descriptive Approach Using Matched Records: The milestone to the study of mortality of Americans was Kitagawa and Hauser's 1973 book, in which multiple socioeconomic indicators and broad cause-of-death categories were used. Data were individual level data taken from the 1960 Matched Records Study which matched the decedents from death certificates with individuals surveyed in the 1960 Census. Tables and graphs based on aggregated data were used to illustrate patterns of mortality differentials. Specifically, Kitagawa and Hauser calculated age-sex-race-education specific death rates; showed age differentials, race differentials, sex differentials and education differentials on mortality in forms of mortality ratios; and then used tables and charts to display these differentials. The advantage of this approach was that it was explicit and fundamental. Descriptive studies still dominant approaches in studying mortality and mortality differentials.

Logistic regression model using individual level data: There are lots of studies using linear probability models in which mortality is presented as the possibility of dying at a given time. For instance, the probability of a representative individual dying at a certain time is assumed to be determined by his health status and socioeconomic status:

$$P(D) = \beta C_i + \gamma X_i + \varepsilon \qquad (2.2)$$

Where P (D) is the probability of death, C is a sector of health conditions variables and X is a sector of SES variables. For instance, Rogers (1992) employed a polytomous logistic regression model, in which the mortality rate is a function of three sectors of variables: demographic factors, familial factors, socioeconomic factors. The data used in Rogers's analysis were individual level data taken from the 1986 NHIS and the 1986 NMFS. Records of individuals that were interviewed in 1986 NHIS and were known to have died in 1986 are extracted from the two studies. It is a common approach in many mortality studies to extract decedents' information from the National Health Interview Survey (NHIS) and its follow-up survey. For instance, Rogers (1992) used the 1986 NHIS data and its follow-up 1986 NMFS data; Muney (2005) used the National Health and Nutrition Examination Survey and its follow-up survey; the next study used the NHIS as well.

Case and Paxson (2005) used individual level data taken from two data sets: NHIS from 1986 to 2001 and associated NHIS Multiple Cause of Death Public Use Data File. They specified mortality as a function of health conditions (categorized into 18 chronicle diseases) and a set of sociodemographic characteristics such as age, sex and years of schooling.

Contoyannis and Jones (2004) used the multivariate Probit model to investigate the impact of life style on health status. They used individual level data from the Health and Lifestyle Survey (HALS) conducted in United Kingdom over the period from 1984 to 1991. The endogenous behavioral variables employed are diet, smoking, exercise, alcohol, sleep, and weight (for height). The exogenous variables in the model were social class, education, marital status, employment status, ethnicity, types of area, region, physical characteristics, tenure, household characteristics, and parental characteristics. They found that sleeping well, exercising, and not smoking in 1984 had dramatic positive effects on the probability of reporting excellent or good Self-Assessed Health (SAH) in 1991.

Boulier and Paqueo (1988) applied a Logit model in the analysis of the determinants of child mortality in Sri Lanka. The data for the estimates of determinants of child mortality came from birth history of women included in the World Fertility Survey conducted in Sri Lanka. For each birth, they had the information on the date of birth of the child and whether or not the child is alive at the time of the survey. Their Sri Lanka empirical analysis revealed that child survival probability is positively related to the mother's education, lower for births to teenager mothers than mothers aged 20-30, greater for second-order births than for first-order or higher-order births; negligible for rural-urban differences after holding the above conditions constant; and lower for birth of Tamil mothers than non-Tamil mothers.

Cox-hazard proportional model using individual level data: One of the recent trends in examining mortality is to employ the Cox Proportional Hazard Model which is typically used in medical studies. The Cox proportional Hazard Model usually takes the following form:

$$h\{(t),(z_1, z_2, \cdots, z_m)\} = h_0(t) * \exp(b_1 * z_1 + \cdots b_m * z_m) \tag{2.3}$$

Where h(t) denotes the resultant hazard, given the values of the m covariates for the respective case (z_1, z_2, ..., z_m) and the respective survival time (t). The term $h_0(t)$ is called the baseline hazard; it is the hazard for the respective individual when all independent variable values are equal to zero. After taking the logarithm of the both sides of the equation and dividing both sides of the equation by $h_0(t)$, a linear model is created:

$$Log[h\{(t),(z\ldots)\}/h_0(t)] = b_1 * z_1 + \ldots + b_m * z_m \tag{2.4}$$

Using Cox Hazard Model, Moore and Hayward's study (1990) indicated that men whose longest occupation ranked low in substantive complexity tended to experience higher mortality. Their data was a cohort of male individuals aged 55 years old and older in United States for the period of 1966 to 1983. They included these variables: age (as categorical data), year surveyed (as categorical data), income (as categorical data), education (as categorical data), occupation (as categorical data), race (as categorical data), labor force status (as categorical data), and marital status (as categorical data). Data were drawn from National Longitudinal Survey of Mature Males (NLS) and the fourth edition of the *Dictionary of Occupational Titles*. The NLS survey identified death as a separate category, thus allowing researchers to examine the linkage between mortality and characteristics of older men's careers. After recoding and excluding incomplete records, the final data sample consisted of records of 3,092 men who were exactly 55 years old at some time between 1966 and 1976, 576 of whom were known to have subsequently died.

Cohort survival analysis using grouped data: Mark Hill (1999) employed a log probability model to analyze the survival probability of native born America women. The data used in his analysis is pooled sample data for the cohort of women aged 50 to 54 years in 1970 Census and aged 70 to 74 years in 1990 Census. The 1970 data are taken from unweighted 15% long-form sample and the 1990 data are taken from special 1% Unweighted Public Use Sample. One important assumption of this two sample survivorship analysis is that the cohort is closed to all forms of migration during the intercensal period. Specifically in Hill's study, the assumption is that number of women aged 50 to 54 in the 1970 Census represents the total number of cohort at risk of death during the intercensal period, and the number of cohort recorded at the second Census (the 1990 Census) represents the total number surviving the intercensal period. If this assunmption is reasonable, then the probability of survival (P) from first Census to second Census can be calculated by dividing the number of survivors (N_2) at second Census by the number at risk (N_1) at first Census.

$$P = \frac{N_2}{N_1} \tag{2.5}$$

Hill emphasized many times that in his study, the cohort is assumed to be closed to all forms of migration, including pseudo-migration caused by age misreporting and other forms of misclassification, and the only means of leaving the cohort is through "death". Hill then developed a cohort based model of survivorship: let π_i denotes the probability of individual *i* surviving from time t_1 to time $t_{2,}$ then the model is formulated as:

$$Log\pi_i = \alpha + \beta' X_i \tag{2.6}$$

Where X_i is a vector of covariates. The explanatory variables used in Hill's model are race, number of child ever born and education. Hill found that when race and children ever born are controlled, the probability of women surviving the twenty year period from 1970 to 1990 is nearly 47% higher for women with 13 or more years of education than for women with fewer years of education.

Log-rate model using grouped data: This approach uses contingency table data composed of events and exposures with specific characteristics and it assumes death rates follow a Poisson distribution. The key in interpreting the coefficients is we can exponentiate the estimates to express the death rate ratio in a manner similar to exponentiating estimates in Logistic regression to obtain the odds ratio (Stokes, Davis, Koch, 2000). This approach allows the research to estimate the separate or joint effects (interactions) of the characteristics on the specific death rate.

Hu and Goldman (1990) employed a Log-rate model to investigate the mortality differential by marital status for a large number of developed countries. They calculated death rates and broke them down by country, by selected years (treated as continuous variable), by age groups (categorized into five groups: 20 to 24, 25 to 34, 35 to 44, 45 to 54, and 55 to 64.), and by marital status (categorized into four groups: married, divorced, widowed, and single). They assumed that number of death (D) of each age-year-marital category are Poisson random variables with means equal to the product of according subpopulation (N) and an underlying death rate (U). The expected number of deaths then can be expressed as:

$$E(D_{ijk}) = N_{ijk} * U_{ijk} \tag{2.7}$$

Where i, j, k denote age group, year, marital state, respectively. This expression leads to a log model:

$$LogE(D_{ijk}) = \log N_{ijk} + \log U_{ijk} \tag{2.8}$$

Where log U_{ijk} denotes the logarithm of mortality risk.

Then they used four separate linear models for log U_{ijk} (the mortality risk), ranging in complexity of a simple addictive model to one that has several interaction terms.

$$LogE(D_{ijk}) = \log N_{ijk} + \eta + \alpha_i + \beta_j + \gamma_k \quad (2.9)$$

Where α_i denotes the effect of *i* th age group, β_j the effect of *j* th year, γ_k the effect of *k* th marital state.

Four models with different interaction terms are selected:

$$LogU_{ijk} = \eta + \alpha_i + \beta_j + \gamma_k \quad (2.10)$$

The first model is the addictive or the proportional hazard model. This model explores the effect of being in a certain marital state on mortality rate, controlling for age group and year.

$$LogU_{ijk} = \eta + \alpha_i + \beta_j + \gamma_k + (\alpha * \gamma)_{ik} \quad (2.11)$$

The second model considers the interaction of age and marital state.

$$LogU_{ijk} = \eta + \alpha_i + \beta_j + \gamma_k + (\alpha * \gamma)_{ik} + A_k * Y + B_k * Y^2 \quad (2.12)$$

The third model considers the relationship between marital status and year, which is considered as both linear and quadratic terms.

$$LogU_{ijk} = \eta + \alpha_i + \beta_j + \gamma_k + (\alpha * \gamma)_{ik} + A_k * Y + B_k * Y^2 + C_i * Y + D_i * Y^2 + e_k * R \quad (2.13)$$

In the fourth model, R is the logarithm of size of a marital group measured as the percentage of persons of relative age group and gender. e_k denotes the elasticity; that is, e_k measures the percentage change in mortality rate for 1% change in the size of marital group.

Hu and Goldman concluded that the divorced persons have the highest death rates in unmarried groups, especially divorced men and divorced and widowed persons in their twenties and thirties have particularly high risk of dying relative to married persons.

Each method has its uniqueness and its disadvantage, for example, using individual level data from the National Mortality Follow-back survey enables researchers examine many cause-of-death variables and one's economic variables but this survey does not provide one's state of residence information thus the scale of research has to be on region level or nation level.

Although there are many different ways to examine the effect of education on mortality, for example, using logistic regression basing on individual level data, or using the Cox hazard model basing on individual level data, I will choose the Log Rate model basing on grouped data as the methodology in my analysis because this approach fits the data type I have available and it provides a refreshing perspective on the way looking at mortality data. Usually individual level data used in mortality study are assumed to follow a Binomial distribution. However in the Log Rate model approach, mortality is considered a rare event and it follows a Poisson distribution.

2.5 Chapter Summary

The literature review on previous mortality studies provides the basis for identifying critical explanatory variables, on understanding the data structure, on setting up the log-rate model, and on hypothesizing outcomes. According to current available research resources, the approach used in Christenson and Johnson's (1995) article will be the main approach that this study follows. I will assume death rates are rare events and follow a Poisson distribution; I will use age, race, sex, and education, along with several interaction terms, as the explanatory variables; I will employ a Log-rate model to analyze mortality risk for each age-race-sex-education specific group and use goodness-of-fit testing; details about the data structure and the Log-rate model will be provided in the next chapter.

The reasons that I choose Christenson and Johnson's article as the procedure to follow are: firstly, I have similar data as Christenson and Johnson had in their 1995 analysis: the PUMS data which provides population counts for each age-race-sex-education specific group, and the death counts for each corresponding group derived from death certificates. Since death certificates do not record decedents' information like income and wealth, religion and life style, it is unlikely to derive these characteristics from death certificates. Secondly, the Log-rate model using grouped data provides a fresh look at the way studying mortality: this approach enables researchers to study mortality when mortality events are considered to be rare, and the data assumption is different from other mortality studies in which data distribution in my analysis is assumed to be a Poisson distribution instead of a Binomial distribution. Finally, I will go further than Christenson and Johnson's analysis by extending the time period to two time periods and adopt an alternative technique, other than following the same technique used by Christenson and Johnson, as a model selection criterion.

Chapter 3

3. DATA AND METHODOLOGY

This chapter will present the source of data, the process of computing death rates, the data issues, the Log-Rate model used in this study and the methods relating to the modeling. The first section will provide source of data sets used to compute the death rates. The second section will give details of step-by-step process of computing age-race-sex-education specific death rates and procedures of verifying the accuracy of data. The third section will introduce the characteristics of the explanatory variables and the strategy of integrating and collapsing data. The fourth section will focus on deriving the Log-rate model. The last section will summarize the data strategy and model strategy.

3.1 Data source

In order to compute the age-race-sex-education specific death rate, I will need data on number of deaths for each age-race-sex-education group and data on population estimates for each age-race-sex-education group. The equation of death rate is as follows:

$$\text{Death Rate} = \frac{Death_{ijkl}}{Population_{ijkl}} \times 1000 \tag{3.1}$$

Where *i* s for age, j stands for race, k stands for sex and l stands for education.

Source of numerators: A tabulation of deaths in Delaware by age group, sex, race, and education was obtained from the Office of Vital Statistics, Division of Public Heath, of the Delaware Health and Social Services. The data only included the recorded deaths and not the population numbers which are needed to calculate death rates. I divide the total 85972 deaths in Delaware from 1989 to 2002 into 14 tables in which each table represented total number of deaths for each year and each cell represented the number of death for a specific subpopulation (See table 3.1).

Source of denominators: Population estimates for each age-race-sex-education specific category were needed in order to calculate the specific death rates. The population estimates published by the Census Bureau are not specific enough and thus do not satisfy the needs of this analysis. As a result, I needed to compute the estimates for these subpopulations by calculating age, sex, race, and education tables for the 5% Public Use Microdata Sample (PUMS). A similar approach was used by Christenson and Johnson to derive age-race-sex-education specific population estimates.

In the 5% PUMS, a random sample of households are made available to researchers that have individual and household level data stripped of geographic identifiers for the sake of confidentiality. The data file household units and person units are associated with a matched serial number so that the data can be used for a variety of research needs. I derived the percentages of each age-race-sex-education specific group from the person data set using SAS programming (See table 3.2). To check the accuracy of our population estimates, I compared the PUMS results for various groupings with the tabulated Census results. The percentages from my frequency analysis are consistent with those from Census publications.

This approach yielded the percentages of each age-race-sex-education specific group. The next thing to do was to get the population estimates for Black Female, White Females, Black males and White Males. These figures can be obtained from existing Census publications. After that, I simply multiplied each percentage by the according population estimate to get the age-race-sex-education specific population estimates (See table 3.3).

Verifying the accuracy of my population estimates: A simple way to verify the accuracy of these population estimates is to compare our percentages of each race-sex-education subpopulation (See table 3.4) with the percentages computed by Census (See table 3.5). As stated before, the 5% PUMS overweighed women and blacks, therefore a slightly higher percentage of black women in total population is reasonable. By comparing table 3.4 and table 3.5, I find that the percentages derived from 2000 5% PUMS are virtually consistent with those in Census 2000 publications. This suggests that this approach is appropriate and the population estimates are accurate.

Comparing the 1990's and 2000's PUMS provides information on the changes of Delaware's population structure during the ten years (See table 3.6). The percentages of black and white, men and women from these 5% PUMSs are reasonable and do not fluctuate during the ten years.

3.2 The computation of death rates

The numerator of the age-race-sex-education specific death rate (number of death for a subpopulation) were taken from Office of Vital Statistics, Division of Public Heath while the denominator (population estimates for the according subpopulation) was generated from the PUMS data.

When calculating the numerator and denominator, I followed Christenson and Johnson's strategy which is to calculate the specific death rates. A three year average was used for $Death_{ijkl}$ while a midpoint value was used for $Population_{ijkl}$. Christenson and Johnson justified the latter approach by arguing "the 1990 Census was conducted roughly at the midpoint of the three-year period, the denominators are estimated from the 5% Public Use Microdata Sample (PUMS) of the 1990 Census for the State of Michigan." (Christenson and Johnson, 1995).

3.3 Variables

The independent variables are the age, sex, race, and educational attainment categories. The Census Bureau considers the majority of people to have completed their education by age 25 and uses this threshold as a reasonable level to calculate educational attainment. Because of this, the age threshold for this analysis of deaths is set at 25 years old as well. Race is broken down by white and black since other races account for a very small percentage (about 3%) of the total population.

To reduce the bias caused by over-reporting the completion of high school or college, I collapsed the education categories into three broader groups: primary education (less than 9 years); secondary education (9 to 12 years including having a diploma); and post-secondary education (13 years and over). The same strategy of categorizing educational attainment is adopted by Christenson and Johnson. They argued that, “To minimize the biases introduced by this type of educational misclassification, I re-categorized the educational attainment into three broader groups: a primary education or less, a secondary education (9 to 12 years), and a postsecondary education (13 or more years of school).” (Christenson and Johnson, 1995).

Interaction terms will be included and tested in the analysis in Chapter 4. Interaction terms reflect the differential effects of factors on mortality rate. For example, the interaction of race and education reflects the notion that the effect of education might be different for whites than for blacks. The statistical test for this term in the model provides a way to determine if this differential exists in the data. Interaction terms help answer questions such as, “Do whites benefit more in terms of reducing mortality rates than blacks?” and “Do women benefit more in terms of decreasing mortality rates than men do if both improve education by same level?” Since interaction terms will be determined by a specific model fit criterion introduced in Chapter 4, at this point, neither a positive nor negative effect of these interaction terms on death rates are assumed.

In the PUMS and death certificates, age is recorded as of the date the person is surveyed and education is recorded as years of schooling completed (ranging from 1 year to 16 years and plus). In my analysis, age is categorized into six age groups and education is categorized into five categories. Other minor modifications include excluding records with missing information on age, sex, race and education. In summary, age is treated as a continuous variable, sex as categorical variable, race as categorical variable, and education as categorical variable. In terms of education, I re-categorized education into three broader groups to avoid the biases introduced by over-reporting of completion of high school or college. Interaction terms will be determined by a model fit criterion and the best combination will be selected based on this criterion.

3.4 Log-rate model

Christenson and Johnson (1995) assume that death rates are discrete data which are Poisson distribution. For Poisson distribution, the mean of expected number of death E(D) is equal to the subpopulation (N) times the underlying death rate (U).

$$E(D) = N * U \tag{3.2}$$

After taken logs on both sides of equation (3.2), the Log-Rate model takes the form of the following:

$$Log[E(D)] = Log(N) + Log(U) \tag{3.3}$$

$$\text{Or,}\quad Log[E(D)] - Log(N) = Log(U) \tag{3.4}$$

Where E(D) is number of death of each age-sex-race-education group; N is the population under exposure (the population of the according subpopulation); U is death rate.

$$Log[E(D)] - Log[N] = Log[\frac{E(D)}{N}] \quad = Log(Death\ Rate) \tag{3.5}$$

Since I compute number of death in the past years, E(D) is also the actual number of deaths (*D*).

I then use a separate linear model to express Log(*U*):

$$Log(U) = \alpha_0 + \beta * X_m \tag{3.6}$$

In equation (3.6), Log (U) can be determined by the subgroup's age, sex, race, education and interaction terms. From the death rates pattern, I can see that the effect of age on mortality could be quadratic, so I will consider introducing square term of age into the model.

Substituting equation (3.6) into equation (3.4), I obtain the equation (3.7):

$$Log[E(D)] - Log(N) = \alpha_0 + \beta * X_m \tag{3.7}$$

Moving *Log(N)* to the right hand side of the equation, I obtain the equation (3.8), the formal equation of Log-rate model:

$$Log[E(D)] = Log(N) + \alpha_0 + \beta * X_m \tag{3.8}$$

Christensen and Johnson (1995) used four main factors along with interaction terms in the Log-Rate model to measure mortality risk. The interaction terms include AGE*EDUCATION, SEX*EDUCATION and RACE*EDUCATION. The interaction term AGE*EDUCATION refers to the differential of age on the effect of education on mortality. In other words, does the effect of education on mortality decrease when people are getting older? The interaction term SEX*EDUCATION answers the question whether education has the same effect on mortality for males and females. The interaction term RACE*EDUCATION refers to the differential of the effect of education on mortality between whites and blacks.

Based on the literature review, I will employ eight variables into Log-rate model: age, square term of age, sex, race, education, age*education, race*education and sex*education. I hypothesized that:

H_0: $\beta_1, \beta_2 \ldots\ldots \beta_8 = 0$.

H_1: $\beta_1, \beta_2 \ldots\ldots \beta_8 \neq 0$.

Specifically, I hypothesized that:

β_1: age is positively related to death rates. Subjects in older age categories face a higher death rate.

β_2: The relationship between age and death rate seems non-linear: those in higher age categories face an increasingly higher death rate.

β_3: Males have a higher death rate than females. The sign of risk factor SEX is positive (Female=0; Male=1).

β_4: Blacks have a higher death rate than whites. The sign of risk factor RACE is positive (Whites=0; Blacks=1).

β_5: Education level is negatively related to death rate. The higher education one received, the less mortality risk he faces.

β_6: Interaction of age and education is positive. Although higher education cushions mortality risk, age is still the dominant factor in deciding mortality: especially at older age, the effect of education diminishes.

β_7: Interaction of race and education is positive. I expect that when both blacks and whites improve education, blacks' mortality risk is higher than whites'.

β_8: Interaction of sex and education is positive. I expect that when both men and women improve education, men's mortality risk is higher than women's.

In section 2.4, I have presented Hu and Goldman (1990)'s article using Log-rate model. In section 3.4, I will present Christenson and Johnson (1995)'s article using the same method. They determined interaction terms in the model selection process using BIC statistics and they let the main effect variables interact with education. Since there are debates over whether or not there exists a sex differential and race differential on mortality, interaction terms will be determined by model fit criterion in the next Chapter.

Missing values: Missing data is a common problem in data analysis. To deal with missing data, very often researchers replace missing data with certain values. To replace values with the missing values, the first task is to explore the distribution of the missing values, and learn about the nature of “missingness" in the data. Simple schemes include assigning a fixed value such as the variable mean or median, selecting an existing value at random, or averaging neighboring values (Swayne et al. 2005). In my analysis, I encountered a missing value: the 1990’s average number of death for black female aged 25 to 34 who had finished only primary education is zero. Since I have only one missing value in my analysis and the reason for missing value is simply that the occurrence is zero for this event at this category, I can employ the simplest approach, which is to assign a small fixed value onto the missing count. Because the logarithm of zero is undefined in calculating death rates, assigning a small positive value, like 0.5, on each death count is a reasonable approach (Allison, 2001).

Offset variable: In Poisson regression the number of “exposure” (in my case, the exposure is the population for each subgroup) is usually called an “offset” variable (wikipedia, 2006), where the exposure variable enters on the right-hand side of equation (3.8), but with a parameter estimate constrained to 1. In my analysis, it is the number of deaths, denoted as the expected deaths or E(D), rather than the population N, that is Poisson distribution. As a result, the logarithm of the denominator for each grouping denoted as Population (N) is used as the offset variable with a coefficient equal to 1 for each observation.

Interpretation: The coefficients from the result of empirical analysis represent the logarithm of relative mortality risk, so the actual relative mortality risk is the exponential of the coefficients, e^{β}. The meaning of e^{β} is similar to the odds ratio in logistic regression models, for instance, assuming the reference level for variable RACE is white, then if e^{β} is equal to 1.25, it means that black's mortality risk is 25 percent higher than white's; likewise, assuming the reference level for variable EDUCATION is primary education, then if e^{β} is equal to 0.85, it means that the morality risk is 15 percent less for people with secondary or post secondary education than for people with only primary education.

Overdispersion: Poisson regression has the advantage of being precisely tailored to the discrete, often highly-skewed distribution of the dependent variable. On the other hand, Poisson regression has the disadvantage of being susceptible to problems of overdispersion that do not affect ordinary regression. A characteristic of the Poisson distribution is that its mean is equal to its variance. In certain circumstances, it will be found that the observed variance is greater than the mean; this is known as overdispersion and it indicates that the model is not appropriate. Overdispersion is usually caused by two problems: an incorrectly specified model, which means more interaction terms or nonlinearities are needed; or lack of independence of observations, which means there is heterogeneity that operates at the level of groups rather than individuals (Allison, 2001). The term "Heterogeneous" means that something (an object or system) consists of a diverse range of different items. Therefore, overdispersion happens in death rates data with Poisson distribution partly because death appears to have larger diversity at group level than at individual level; this makes sense because death pattern does not differ much within a group but tends to differ greatly among groups. One important property of Poisson distribution is that the variance is equal to the mean; however, the empirical variance is, in fact, usually much higher than the theoretical variance (The theoretical variance is assumed to be equal to the mean). If the estimate of dispersion, as measured by the deviance or Pearson's chi-square divided by the degrees of freedom, is not near 1, then the data may be overdispersed. One approach to adjust overdispersion is to adjust chi-square and test statistics, leaving the coefficient estimates unchanged.

Model fit criterion: Christenson and Johnson (1995) used the BIC statistic (BIC stands for Bayesian Information Criterion) as a model fit criterion.

$$BIC = L^2 - (df)\log(N) \tag{3.9}$$

Where L^2 is the likelihood-ratio chi-square test statistics, *df* denotes the degrees of freedom, and N is the total sample size. The value of BIC can be either positive or negative. In deciding among several models, the preferred model is the one with the lowest BIC value. The BIC provides a consistent model-selection procedure in that it chooses the correct model with a high probability (Christenson and Johnson, 1995). The BIC statistics is based on the maximum likelihood estimates of the model parameters. In maximum likelihood, the idea is to estimate parameters so that, under the model, the probability of the observed data would be as large as possible. The likelihood is this maximum probability, and will always be between 0 and 1.

However, the function obtained by dividing a log-likelihood function for the binomial or Poisson distribution by a dispersion parameter is not a legitimate log-likelihood function. It is an example of a quasi-likelihood function (SAS Online Documentation Version 8, 1999). Most of the asymptotic theory for log likelihoods also applies to quasi-likelihoods, which justifies computing standard errors and likelihood ratio statistics using quasi-likelihoods instead of proper log likelihoods. Since the BIC statistic is based on log-likelihood function and since the function of Poisson distributed data after adjusting overdispersion is not a log-likelihood function, BIC is not an appropriate model fit criterion in my analysis.

In addition, Lindsey (1997) argued that BIC statistic tends to choose the simplest model with least parameters as the “best” model, hence using BIC criterion will neglect the interaction terms and will not allow me to examine the differentials of education on mortality. According to Lindsey’s theory, using BIC statistics will tend to choose the simplest model and exclude the interaction terms. This contradicts with Christenson and Johnson’s study: they incorporated all the two-way interaction terms in the final model. Because of the above two reasons, I did not follow Christenson and Johnson’s BIC strategy.

Log rate model is a family member of generalized linear model family. Two statistics that are helpful in assessing the goodness of fit of a given generalized linear model are the scaled deviance and Pearson's chi-square statistic. The major problem in Poisson distribution data is overdispersion. Evidence of underdispersion or overdispersion indicates inadequate fit of the Poisson model. Overdispersion can be adjusted by dividing scaled deviance by degrees of freedom; therefore this criterion is a very important measure for goodness-of-fit for Poisson model thus it can be used as model fit criterion in my analysis. To sum up, the deviance or Pearson's chi-square divided by its degrees of freedom is not only used to indicate overdispersion or underdispersion, but also to indicate other problems such as an incorrectly specified model or outliers in the data, therefore Scaled Deviance over degrees of freedom

(SDD) will be used as model fit criterion instead of BIC or likelihood ratio statistics. This approach allows the researcher to compare multiple models of main effects and interaction terms to find a best fit. It also provides a way to gauge whether a variable (in most cases a higher order interaction term) really contributes to the model or not.

The GENMOD procedure in SAS/STAT package has provided some criteria for assessing goodness-of-fit for Poisson regression: the Deviance, the Scaled Deviance, the Pearson Chi-square, the Scaled Pearson Chi-square, and the Log Likelihood. Another statistical knowledge site, the UCLA Technology Service online tutorial also provides an example of using scaled deviance divided by degree of freedom as the goodness-of-fit for Poisson distributed data. Since scaled deviance divided by degree of freedom (SDD) is a recognized measure of goodness-of-fit for Poisson regression and it also reveals information of overdispersion problem, I choose SDD as the model fit criterion.

3.5 Chapter summary

The explained variable, death rates, are assumed to be Poisson distribution and the explanatory variables are assumed to affect mortality risk. Data are carefully cleaned and grouped to compute accurate death rates for two time periods: 1990 and 2000.

Decisions on data include: Data for calculating death rates are the 5% Public Use Microdata Sample (PUMS) and the death counts from death certificates of Delaware from 1989 to 2002; In calculating the death rates, the denominators (the population for each subgroup) is derived from PUMS and the numerators (the number of death) are derived from death certificates; Two time periods will be examined: the 1990 death rates and the 2000 death rates; The death rates are the three year average death rate; explanatory variables are: age, categorized into six 10-year interval age groups; race and sex are treated as categorical variable; and education is treated as categorical variable. Interaction terms will be determined by model fit criterion in chapter 4.

A Log-Rate model, based on Poisson distribution, is used to analyze the data. The approach to deal with the overdispersion problem is to divide the scaled deviance with degrees of freedom. For simplicity, missing values will be replaced by a fixed small value (for instance, 0.5) to avoid the appearance of the logarithm of zero. The criterion for model building is the scaled deviance divided by degrees of freedom as the model fit criterion.

Chapter 4

4. EMPIRICAL ANALYSIS

This chapter presents the analytical results from empirical analysis using Log-rate model. The first section describes the model selection procedure and presents the final model. The subsequent two sections interpret the results from 1990 and 2000's analysis. The last section gives a descriptive summary of the similarities and dissimilarities between 1990 and 2000's results.

4.1 Model Fit

This analysis used the Scaled Deviance divided by Degree of freedom (SDD) as the model selection criterion. The closer this value is to one, the better fit. Since the scaled deviance divided by degree of freedom is between 0 to 1, the actual differences can be very small among these values. For example, in one contrast of models there is little difference between a model with a SDD value of 0.9948 and a SDD value of 0.9922 (table 4.1). When there are several models that had very close SDD values, I chose the most parsimonious one. Results from the empirical analysis showed that although one model had the absolute highest SDD value, few of the parameter estimates were significant and thus it was not chosen as the final model.

To select the best fit model, I tested several one-way models in which age is treated differently. Age in this analysis can be represented by 5 dummy variables or as a continuous variable where the midpoints of the age categories is coded as the values. The later approach allows for the use of a quadratic term for age, which was used in Christenson and Johnson's analysis.

When age is categorized into 5 dummy variables and education is collapsed into 2 dummy variables, the model gives the best model fit value among all of the one-way models, the SDD is equal to 0.9948 (table 4.1). Then I tested a set of two-way models with different combinations of interaction terms based on the selected one-way model. The final selected model has a scaled deviance/degree of freedom of 0.9996, being the best model fit value among the all. The final selected explanatory variables are: age (as categorical); sex (as categorical); race (as categorical); education (as categorical); the interaction between race and education and the interaction between sex and education.

In Christenson and Johnson's (1995) paper, age is treated as categorical in the first order models and then treated as continuous in the second order models (as interaction terms with education). My "best" model from the model selection is consistent with their classification of explanatory variables: age is treated as categorical in main effect and is treated as continuous when it interacts with education; education is collapsed into 3 major categories------primary education (less than 9 years' schooling), secondary education (9 to 12 years' schooling) and post-secondary education (13 years' schooling and over).

4.2 Empirical Analysis of Death Rates of 1990

The coefficients for age (as a categorical variable) show people's mortality risk increased by age, especially at the older age categories. Except for variable AGE1 (refers to age group 25 to 34 years old), all age dummy variables are significant at 1% level, indicating they are significantly different from the reference group (people aged 25 to 34 years old). Being in age group of 55 to 64 years will increase mortality rate by 6.51 times than being in age group of 25 to 34 years, holding other characteristics constant (see table 4.3).

Recall the interpretation of the coefficients from Log-Rate model: The coefficients from the result of empirical analysis represent the logarithm of relative mortality risk, so the actual relative mortality risk is the exponential of the coefficients, e^{β}. The meaning of e^{β} is similar to the odds ratio in logistic regression models, for instance, assuming the reference level for variable RACE is white, then if e^{β} is equal to 1.25, it means that black's mortality risk is (1.25-1)=25 percent higher than white's. (See Chapter 3, section 3.4 Log-rate model).

In table 4.3, the coefficients for education show that people's mortality risk decreases by receiving more years of education. SECONDARY and POSTSECONDARY are both significant at the 1% level. Mortality risk for people with secondary education is 29% (1 – 0.79, e^{β}=0.79) lower than for people with only primary education; likewise, mortality risk for people with post-secondary education is 43% (1 – 0.57, e^{β}=0.57) lower than for people with only primary education (see table 4.3).

Black people usually suffer a higher mortality rate than white people, however, in my analysis, RACE is insignificant at all significance levels, and the value of (e^{β} -1) is roughly equal to 1. However, RACE is significant at second order when it interacts with secondary education. The significance of RACE at second order indirectly shows there is still a racial difference at lower order. The variable SEX is significant at the 5% level which suggests that men's mortality risk is 34% higher than women's risk, holding other factors in the model constant.

The significance of an interaction between RACE and SECONDARY indicates that there is a racial difference between whites and blacks when both improve their education level from primary education to secondary education: black people's mortality risk is 53% higher than white people's. In another words, it suggests that white people benefit more from finishing high school than do black people. The same is not true for education beyond a high school diploma - the interaction of RACE and POSTSECONDARY is insignificant. This suggests that when both blacks and whites improve their education from secondary level to postsecondary level, the effect of education is the same for whites and blacks.

The interaction of SEX and SECONDARY, and SEX and POSTSECONDARY are both insignificant (table 4.3). This suggests that gender does not influence the effect of education on mortality.

4.3 Empirical Analysis of Death Rates of 2000

In the models for 2000, all the age dummy variables are significant: AGE2 is significant at the 5% level while AGE3, AGE4, AGE5, and AGE6 are significant at the 1% level. (See table 4.4) The value of e^{β} for AGE2 suggests that being in the age group of 35 to 44 years increases mortality risk by 1.1 times more than the reference age level (referring to AGE1, 25 to 34 years); being at age group of 45 to 54 years increases mortality risk by 2.8 times more than reference age level; being at age group of 55 to 64 years increases mortality risk by 7 times more than the reference age level; being at age group of 65 to 74 years increases mortality risk by 16.8 times more than the reference age level; and being at age group of 75 years and over increases mortality risk by 54.7 times more than the reference age level (table 4.4). There is a clear mortality risk pattern from these values: as expected, mortality risk increases by age and it increases more dramatically at older ages.

Both SECONDARY and POSTSECONDARY education dummies are significant at the 1% probability level. The value of e^{β} for these variables suggest that finishing secondary education will decrease mortality risk by 37% (37% = 1 - 0.63) compared to having completed only primary education; finishing postsecondary education will decrease mortality risk by 59% than having completed only primary education (table 4.4).

Neither RACE nor SEX were significant in the 2000's death rate analysis (table 4.4). However the signs, while insignificant, indicated that blacks had a higher mortality risk than whites, and men have a higher mortality risk than women.

Combining with the significance of the interaction term of RACE and secondary education, I can conclude that there is an overall racial difference on mortality at lower education level.

Similar to the 1990 analysis, there is a significant interaction between race and education in the 2000 analysis. The interaction of RACE and SECONDARY is significant at the 5% level (see table 4.4). This shows that there exists a racial difference on the effect of completing high school. The effect of completing a high school degree has less of an effect of decreasing mortality for Blacks when compared to Whites. Even though completing a high school degree decreases the mortality rate overall, the model shows that for this level of education, Blacks a have a 75% higher mortality than whites. The interaction of RACE and POSTSECONDARY is insignificant, as was the case in 1990.

As was the case in 1990, the interaction terms of SEX*SECONDARY and SEX*POSTSECONDARY were insignificant (table 4.4). This implies that there are no gender differences on the effect of education on mortality in 2000.

4.4 Comparison of the 1990 Death Rates with the 2000 Death Rates

From comparing the coefficients and the values of e^{β} 's, the dominant effect of age in determining mortality risk is greater for 2000 than for year 1990. The differences between age coefficients between the two time periods (1990 and 2000) are all positive. The signs of differences between education variables from 1990 and 2000 are both negative (see table 4.5), which indicates that education's effect on mortality is greater in 2000 than in 1990. RACE is insignificant in both years. SEX is a significant factor in 1990's analysis, but is insignificant in 2000's analysis. In addition, the e^{β} value of SEX decreases from 1.34 in 1990 to 1.08 in 2000 which also indicates that the effect of gender in determining mortality risk has diminished during the period, once we control for all other variables in the model.

The interaction of RACE*SECONDARY increased by .23 from 1990 to 2000 which indicates the racial gap of the effect of education on mortality became larger during the 10 year period. The interaction of RACE*POSTSECONDARY also increased by 6% from 1990 to 2000, however, since the coefficients are insignificant, it is insufficient to conclude that there exists a racial differential of the effect of education on mortality at a higher education level. There does not seem to exist a gender differential in education over time because the interactions of SEX*SECONDARY and SEX*POSTSECONDARY for both years were insignificant.

In general, the coefficients from the two time periods are basically consistent, but there are some important changes. Some factors show a stronger positive effect on mortality during the 10 years, such as the age variables; some factors show stronger negative effect on mortality during the 10 years, such as education variables; and the racial differential has broadened during this period.

4.5 Chapter Summary

Firstly, the model fit criterion “scaled deviance divided by degrees of freedom” is introduced. Basing on this model fit criterion, the “best” model is selected. Then the empirical results from analysis of 1990 are presented. These results indicate that people with secondary education can expect to decrease mortality risk by 29 percent comparing to people with only primary education; people with postsecondary education can expect to decrease mortality risk by 43 percent comparing to people with only primary education; a male’s mortality risk is 34 percent higher than female’s; and when both improve education from primary to secondary, the mortality risk for blacks is still 53 percent higher than that for whites. The empirical results from analysis of 2000 are virtually consistent with those from analysis of 1990. However, the differences between the results of the two periods indicate that SEX is not a significant term in the 2000 analysis; the dominant effect of age on mortality is stronger in 2000; and the effect of education is weaker in 2000 analysis. Overall, the results from empirical analysis of the two periods confirm the hypothesis that mortality risk decreases when education is improved.

Chapter 5

5. CONCLUSION AND DISCUSSION

5.1 Conclusion

The purpose of this study is to empirically examine the effect on education on mortality and the mortality differentials in Delaware from 1989 to 2002. To do that, I used grouped data to compute the age-race-sex-education specific death rates for 1990 and 2000 in Delaware. I then used this death rates data in a Log-rate model using equation (3.1).

The numerators in the formula are the counts of deaths for each age-race-sex-education group; and the denominators in the formula are the population for each age-race-sex-education group. Death counts for each subgroup are obtained from death certificates from the Office of Vital Statistics, Division of Public Heath, of the Delaware Health and Social Services. Population estimates for each subgroup were derived from PUMS data.

The data strategy and model strategy in my analysis mainly followed the work of Christenson and Johnson's 1995 paper; specifically, the data (death rates) follows a Poisson distribution such that the mortality risk for each subgroup is expressed as a relative risk. For example, assuming the reference level for variable “RACE” is set as blacks, to interpret the mortality risk for whites the parameter estimate for variable “RACE” is exponentiated first, then the exponentiated value is compared with 1 -- the difference is the relative risk.

The model fit criterion used in my analysis is the scaled deviance divided by degrees of freedom. The model fit criterion is different from the BIC statistics used in Christenson and Johnson’s paper because BIC is based on a maximum likelihood assumption. However the Poisson distributed data, after adjusting overdispersion, is a quasi-likelihood function so the BIC is not a suitable criterion in this case.

The results from empirical analysis support the alternative hypothesis that education has an inverse effect on mortality rates. For both periods, moving from primary to secondary education can reduce one’s mortality risk by 30 percent; moving from primary to post-secondary education can decrease one’s mortality risk by at least 40 percent (table 4.4), holding other factors constant. In fact, this inverse effect is even stronger than have been found in Christenson and Johnson’s 1995 article. Moreover this inverse effect of education on mortality tends to increase from 1990 to 2000.

The null hypothesis of RACE is tested and cannot be rejected because for both periods RACE as a main effect is not significant at any confidence level; however, as a matter of fact, blacks do suffer higher death rates relative to whites at each age level. Furthermore, when race interacts with education, particularly, the significance of RACE*SECONDARY in both time periods, suggests that there exists a racial differential on the effect of mortality. This finding implies that policies regarding improving people's education to achieving a high school degree tend to have greater impact on white people than on black people. Although the null hypothesis of RACE cannot be rejected, the significance of risk factor RACE at second order indicates the overall impact of RACE, even though RACE is not significant at first order. Unlike interaction term RACE*SECONDARY, interaction of RACE*POSTSECONDARY is not significant. The changing pattern of the significance of RACE*SECONDARY to RACE*POSTSECONDARY indicates that the racial difference diminishes at higher education level. This finding in turn implies the effect of education on decreasing racial difference on mortality.

The null hypothesis of SEX is tested and is rejected for 1990 but cannot be rejected for 2000. SEX is a significant effect in 1990 so I conclude that mortality risk is higher for men than for women in 1990; SEX is not significant in 2000 but the sign of SEX is positive and the value of (e^{β} -1) is positive, which implies that mortality risk is higher for men than for women in 2000. On the Delaware Vital Statistics Annual Report, men's longevity seems to be a couple of years shorter than women's, in reality men might die of risks but women might live with risks. Again, it might be men's socioeconomic status instead of being men genetically, that cause men live shorter lives than women.

Aside from the above major results, other results drawn from empirical analysis include: For both periods, interaction of RACE with secondary education (RACE*SECONDARY) is significant; interactions of RACE with post-secondary education (RACE*POSTSECONDARY) is not significant; interaction between SEX and education (SEX*SECONDARY, SEX*POSTSECONDARY) is not significant. But overall, empirical results not only pinpoint the effect of education has in reducing one’s mortality risk but also imply the effects of socioeconomic attributes that might cause the race differential and gender differential.

These results should help the policy makers with state budget allocation on education expenditure. Since almost 40 percent of the annual budget is spent on education and the budget of higher education is not a mandatory portion in state budget (unlike the budget on Medicare and K-12), the higher education expenditure must compete with other expenditures, such as the cost associated with incarceration. The result from empirical analysis shows that there is a racial differential of the effect of education on mortality and this difference diminishes at higher education level, which implies that improving college education, rather than improving primary and secondary education, would help reduce the racial difference on death rates. This work may be persuasive in advocating black parents to put more inputs in education in order to reduce mortality risk.

This research may also contribute to similar mortality study using Poisson regression model. When death is considered rare event, as the case of low child mortality in many industrialized countries and areas, using Poisson regression model can be an appropriate approach for this kind of study. The discussion of using SDD (Scaled Deviance divided by Degree of Freedom) over BIC statistics as model fit criterion is a major difference between Christenson and Johnson’s analysis and my analysis. The advantage of SDD is that it adjusts the most common problem in Poisson regression, overdispersion, and it is also a measure of how fit data are for a Poisson model. So far I have not found literature discussing and comparing the usage of SDD over BIC in Poisson model; however this study may shed some light on formalizing the model fit criterion for Poisson model.

5.2 Discussion

Limitations for this study are mainly due to data availability. Although researchers are interested in incorporating more information into mortality study (such as information on the decedents' education, income and wealth, occupation, and marital status), the standard death certificates do not record income and wealth information, thus it is unlikely to compute the age-race-sex-education-income-marital-occupation specific death rates. Current analysis is based on age-race-sex-education specific death rates and data are in form of a contingency table which consists of 120 cells. If, for example, 3 income categories and 5 occupation types are incorporated into the current model, then I will have a contingency table consists of 120*3*5=1800 cells in which each cell representing a age-race-sex-education-income-occupation specific death rate. In that way, I will encounter many blank cells. Since Delaware has a relatively smaller population base, the more specific the subgroup is, the more blank death counts will appear in the calculation of death rate. For example, there is only one blank count in my current analysis: the 1990's average number of death for black female aged 25 to 34 who had finished only primary education is zero; I will encounter many more zeros if I further try to divide the above subgroup into more specific "low income", "middle level income", and "high income" group.

Further research work on adult mortality using Log-Rate model with Poisson distributed data may consider changes or improvements such as re-categorizing data. For example, to divide age into 5 year age groups instead of 10 year age group; to remove age group of 85 years old from the data and over (because some believe that at this age group, death is no longer a rare event, thus Poisson distribution assumption will not hold.). Researchers may also consider approaches like expending data sample by incorporating mortality data from other states with larger population base; or separating the analysis into two analyses: an analysis for the males and the other analysis for the females.

Table 1.1 Numbers and Percentages of Missing Education Information on Death Certificates, Delaware, 1989 to 2002.

	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
Number	523	266	267	224	237	174	159	174	122	150	142	144	163	141
Total deaths	5570	5482	5669	5718	5921	6125	6051	6302	6281	6335	6459	6600	6839	6590
Percentages	9.39%	4.85%	4.71%	3.92%	4.00%	2.84%	2.63%	2.76%	1.94%	2.37%	2.20%	2.18%	2.38%	2.14%

Source: Delaware Vital Statistics Center

Table 1.2 Crude Death Rate of Nationwide and Delaware, 1989 to 2002.

	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
Nationwide	8.716	8.618	8.610	8.543	8.812	8.765	8.810	8.739	8.655	8.660	8.782	8.747	8.476	8.501
Delaware	8.874	8.583	8.684	8.555	8.709	8.832	8.607	8.780	8.663	8.616	8.605	8.770	8.939	8.497

Notes:

Crude death rate is expressed as total deaths per 100,000 populations.

Source: http://www.census.gov/popest/archives/1990s/nat-total.txt;
Delaware Vital Statistics Annual Report 2002;
http://www.census.gov/popest/states/tables/NST-EST2004-08.pdf;

Table 3.1 Three-year-average Number of Death by Education Attainment, by Age, by Race, and by Gender for 1990 (Three Year average from 1989 to 1991).

		White		Black	
Education Attainment	Age Group	Male	Female	Male	Female
<9 years	25-34	3.7	1.3	1.7	0.0
	35-44	6.7	2.7	1.0	1.0
	45-54	12.7	6.3	6.3	3.0
	55-64	57.0	25.0	25.0	10.7
	65-74	121.7	84.0	45.7	35.7
	75+	294.7	411.0	71.3	87.0
	Unknown	…	…	…	…
9-11 years	25-34	10.7	4.3	6.0	3.0
	35-44	16.3	3.3	6.7	5.0
	45-54	26.3	13.7	16.0	10.0
	55-64	60.0	43.0	29.0	16.7
	65-74	119.3	87.3	27.0	25.0
	75+	113.0	170.3	15.7	22.3
	Unknown	…	…	…	…
12 years (including diploma)	25-34	45.7	14.0	18.3	4.7
	35-44	38.0	20.3	18.3	9.7
	45-54	67.7	39.0	22.0	18.3
	55-64	145.3	113.7	26.3	25.7
	65-74	239.0	217.7	30.7	30.7
	75+	243.7	451.3	21.0	33.7
	Unknown	…	…	…	…
13-15 years	25-34	9.0	7.0	4.0	2.0
	35-44	15.7	6.0	4.0	3.0
	45-54	15.3	16.7	2.7	4.0
	55-64	33.7	31.0	5.3	4.0
	65-74	77.0	74.3	3.7	4.0
	75+	94.3	172.3	2.0	4.7
	Unknown	…	…	…	…
16 years and over	25-34	11.3	4.0	2.7	1.0
	35-44	19.0	4.0	2.0	2.0
	45-54	20.0	2.7	7.3	2.7
	55-64	50.3	5.3	2.3	3.3
	65-74	98.7	3.7	4.3	3.3
	75+	136.0	2.0	5.0	7.7
	Unknown	…	…	…	…
Total		2318.7	2059.0	469.0	422.7

Source: Death certificates from the Office of Vital Statistics, Division of Public Heath, of the Delaware Health and Social Services.

Table 3.2 Percentages of Each age-race-sex-education Specific Population in Total Race-Sex population for 1990, Derived from 1990 5% PUMS.

Black Female			Education Attainment		
Age Group	<9	9 to 11	12	13 to 15	16 and over
25-34	0.54%	5.48%	12.78%	8.23%	3.34%
35-44	1.27%	5.48%	8.23%	7.36%	3.01%
45-54	0.87%	3.81%	5.69%	4.15%	2.34%
55-64	2.61%	3.88%	3.14%	0.94%	0.94%
65-74	2.88%	4.01%	1.61%	0.47%	0.33%
75+	3.41%	1.40%	1.14%	0.33%	0.33%
White Female			Education Attainment		
Age Group	<9	9 to 11	12	13 to 15	16 and over
25-34	0.45%	1.99%	7.78%	7.36%	5.95%
35-44	0.46%	1.50%	7.90%	6.24%	5.61%
45-54	0.46%	2.26%	6.18%	3.97%	3.07%
55-64	0.81%	3.24%	6.64%	2.86%	1.99%
65-74	1.57%	3.02%	5.34%	2.12%	1.48%
75+	2.55%	2.41%	2.44%	1.48%	0.86%
Black Male			Education Attainment		
Age Group	<9	9 to 11	12	13 to 15	16 and over
25-34	0.82%	5.48%	12.85%	6.46%	3.76%
35-44	0.65%	6.38%	7.86%	9.00%	3.60%
45-54	1.15%	5.07%	5.56%	2.62%	2.37%
55-64	4.91%	3.85%	2.86%	0.98%	1.06%
65-74	4.01%	2.45%	1.88%	0.33%	0.33%
75+	2.29%	0.82%	0.49%	0.08%	0.00%
White Male			Education Attainment		
Age Group	<9	9 to 11	12	13 to 15	16 and over
25-34	0.59%	2.99%	9.44%	7.13%	5.98%
35-44	0.55%	1.96%	6.86%	6.52%	7.40%
45-54	0.79%	2.17%	5.32%	3.70%	4.66%
55-64	1.58%	2.89%	4.74%	2.82%	3.40%
65-74	1.90%	2.70%	3.31%	2.17%	2.51%
75+	1.60%	1.33%	1.15%	0.74%	1.12%

Table 3.3 Population Estimates of Each age-race-sex-education Specific Population for 1990.

Black Female, total = 34460	Education Attainment				
Age Group	<9	9 to 11	12	13 to 15	16 and over
25-34	186	1888	4404	2836	1151
35-44	438	1888	2836	2536	1037
45-54	300	1313	1961	1430	806
55-64	899	1337	1082	324	324
65-74	992	1382	555	162	114
75+	1175	482	393	114	114
White Female, total = 185974	Education Attainment				
Age Group	<9	9 to 11	12	13 to 15	16 and over
25-34	837	3701	14469	13688	11065
35-44	855	2790	14692	11605	10433
45-54	855	4203	11493	7383	5709
55-64	1506	6026	12349	5319	3701
65-74	2920	5616	9931	3943	2752
75+	4742	4482	4538	2752	1599
Black Male, total = 28260	Education Attainment				
Age Group	<9	9 to 11	12	13 to 15	16 and over
25-34	232	1549	3631	1826	1063
35-44	184	1803	2221	2543	1017
45-54	325	1433	1571	740	670
55-64	1388	1088	808	277	300
65-74	1133	692	531	93	93
75+	647	232	138	23	0
White Male, total = 169211	Education Attainment				
Age Group	<9	9 to 11	12	13 to 15	16 and over
25-34	998	5059	15974	12065	10119
35-44	931	3317	11608	11033	12522
45-54	1337	3672	9002	6261	7885
55-64	2674	4890	8021	4772	5753
65-74	3215	4569	5601	3672	4247
75+	2707	2251	1946	1252	1895

Source: Total population for Black Female, White Female, Black Male and White Male are taken from 1990 Census publication.

Table 3.4 Population Estimates by Education Attainment, Age, Race, and Sex, derived from 1990 5% PUMS.

Black Female						
Education Attainment						
	<9	9 to 11	12	13 to 15	16 +	Total
Population	3990	8291	11231	7402	3546	34460
Percentage	0.95%	1.98%	2.69%	1.77%	0.85%	8.25%
White Female						
Education Attainment						
	<9	9 to 11	12	13 to 15	16 +	Total
Population	11716	26817	67471	44690	35261	185955
Percentage	2.80%	6.42%	16.15%	10.69%	8.44%	44.50%
Black Male						
Education Attainment						
	<9	9 to 11	12	13 to 15	16 +	Total
Population	3908	6797	8902	5502	3143	28252
Percentage	0.94%	1.63%	2.13%	1.32%	0.75%	6.76%
White Male						
Education Attainment						
	<9	9 to 11	12	13 to 15	16 +	Total
Population	11862	23757	52151	39054	42421	169245
Percentage	2.84%	5.68%	12.48%	9.35%	10.15%	40.50%

Table 3.5 Population Estimates by Education Attainment, Age, Race, and Sex from 1990 Census Publication.

			Black Female			
			Education Attainment			
	<9	9 to 11	12	13 to 15	16 +	Total
Population	3441	8904	11026	7510	3579	34460
Percentage	0.82%	2.13%	2.64%	1.80%	0.86%	8.25%
			White Female			
			Education Attainment			
	<9	9 to 11	12	13 to 15	16 +	Total
Population	11328	25682	66345	46207	36412	185974
Percentage	2.71%	6.15%	15.88%	11.06%	8.71%	44.50%
			Black Male			
			Education Attainment			
	<9	9 to 11	12	13 to 15	16 +	Total
Population	3511	7220	9195	5283	3051	28260
Percentage	0.84%	1.73%	2.20%	1.26%	0.73%	6.76%
			White Male			
			Education Attainment			
	<9	9 to 11	12	13 to 15	16 +	Total
Population	11045	22087	51306	39567	45202	169207
Percentage	2.64%	5.29%	12.28%	9.47%	10.82%	40.49%

Table 3.6 Comparing the 1990 5% PUMS with 2000 5% PUMS.

	1990 PUMS		2000 PUMS	
	sample size	percentage	sample size	percentage
Total records	48153		57292	
housing records	16239		18386	
persons records	31914		38906	
total white males	12683	39.74%	14361	36.91%
total white females	13576	42.54%	15434	39.67%
total black males	2243	7.03%	3164	8.13%
total black females	2541	7.96%	3593	9.24%
total other races	871	2.73%	2354	6.05%
total persons aged 25 years and over (only white and black)	20332		24527	
males	9556	47.00%	11478	46.80%
females	10776	53.00%	13049	53.20%
total persons aged 25 years and over (only white and black)	20332		24527	
white	17615	86.64%	20606	84.01%
black	2717	13.36%	3921	15.99%
total persons aged 25 years and over (only white and black)	20332		24527	
white males	8334	40.99%	9749	39.75%
white females	9281	45.65%	10857	44.27%
black males	1222	6.01%	1729	7.05%
black females	1495	7.35%	2192	8.94%

Table 3.7 Explanatory Variables.

Variables	Category	Expected effect on Death Rates
Age	Six categories: 25 to 34; 35 to 44; 45 to 54; 55 to 64; 65 to 74; 75 and over	Positive
Sex	Two categories: 1=Male; 0=Female	Positive
Race	Tow categories: 1=Black; 0=White	Positive
Education	Five categories: less than 9 years; 9 to 11 years; 12 years;13 to 15 years; 16 years and over	Negative
Interaction Terms	Description	Expected effect on Death Rates
Age*Education	Age will be treated as continuous in interaction term.	Positive
Sex*Education	Sex has two categories.	Positive
Race*Education	Race has two categories.	Positive

Source: Data are taken from 1990 and 2000 PUMS and Office of Vital Statistics, Division of Public Heath, of the Delaware Health and Social Services.

Table 4.1 Model Fit Selection Using Death Rate Data of 1990.

No.	Explanatory Variables One-Way	DF	Scaled Deviance	Scaled Deviance/DF
1	age educ race and sex	115	105.917	0.9210
2	age educ[2] race and sex	114	105.381	0.9244
3	age educ[4] race and sex	112	98.898	0.8830
4	age[5] educ race and sex	111	109.175	0.9836
5	age[5] educ[2] race and sex	110	109.427	0.9948
6	age[5] educ[4] race and sex	108	101.887	0.9434
7	age age^2 educ race and sex	114	111.884	0.9814
8	age age^2 educ[2] race and sex	113	112.117	0.9922
9	age age^2 educ[4] race and sex	111	104.799	0.9441

Notes:
1. The criterion used in model selection is scaled deviance/DF; the closer it is to 1, the better fit.
2. One way model #5's scaled Deviance/DF is closest to 1 thus it is selected as the main effect model.
3. In the second column, the "Explanatory Variables", "age" refers to continuous age using midpoint coding; "age[5]" refers to 5 categorical age dummy variables; "educ" refers to continuous years of schooling; "educ[2]" refers to 2 categorical education dummy variables; "age^2" refers to the interaction of continuous age with itself.

Table 4.2 Model Fit Selection Using Death Rate Data of 1990, with Two-Way Interaction Terms Based on One-Way model # 5.

No.	Explanatory Variables Two-Way Interaction Terms Based on Model #5	DF	Scaled Deviance	Scaled Deviance/DF
10	age age^2 race and sex interact with educ[2]	102	100.406	0.9844
11	age race and sex interact with educ[2]	104	102.842	0.9889
12	age^2 race and sex interact with educ[2]	104	102.539	0.9860
13	age age^2 and race interact with educ[2]	104	103.022	0.9906
14	age age^2 and sex interact with educ[2]	104	101.277	0.9738
15	race and sex interact with educ[2]	106	105.954	0.9996
16	age and age^2 interact with educ[2]	106	103.844	0.9797
17	age and race interact with educ[2]	106	105.526	0.9955
18	age and sex interact with educ[2]	106	103.642	0.9778
19	age^2 and race interact with educ[2]	106	105.232	0.9928
20	age^2 and sex interact with educ[2]	106	103.394	0.9754

Notes:
1. Two way model #15's scaled deviance/DF is closest to 1 thus it is selected as the final model for analysis.
2. Other combinations of interaction terms do not give higher scaled deviance/DF thus they are not listed here.
3. In the second column, the “Explanatory Variables”, “age” refers to continuous age using midpoint coding; “age[5]” refers to 5 categorical age dummy variables; “educ” refers to continuous years of schooling; “educ[2]” refers to 2 categorical education dummy variables; “age^2” refers to the interaction of continuous age with itself.
4. The final selected explanatory variables are: age[5], educ[2], race, sex, race*educ[2], and sex*educ[2], in which age is treated as categorical and education is collapsed into dummy variables.

Table 4.3 Empirical Analysis Using Death Rates of 1990.

Variables	Parameter Estimates (β)	e^{β}	Pr>ChiSq
Main effect			
Intercept	-6.3457***		<.0001
AGE1 (age between 25 and 34 is the reference level for age.)			
AGE2 (35 to 44)	0.334	1.40	0.136
AGE3 (45 to 54)	1.1422***	3.13	<.0001
AGE4 (55 to 64)	2.0165***	7.51	<.0001
AGE5 (65 to 74)	2.8072***	16.56	<.0001
AGE6 (75 +)	3.9084***	49.82	<.0001
Education (Primary education is reference)			
secondary	-0.3387***	0.71	0.0016
postsecondary	-0.5585***	0.57	<.0001
Race (White is reference)	-0.0007	1.00	0.9962
Sex (Female is reference)	0.2942**	1.34	0.012
Interaction Terms			
Race*Education			
Race*Secondary	0.4253**	1.53	0.0169
Race*postsecondary	0.3925	1.48	0.1384
Sex*Education			
Sex*Secondary	0.2038	1.23	0.1512
Sex*Postsecondary	-0.0573	0.94	0.7362

Notes:
Observation=120. Scaled Deviance/DF=0.9996.
*** ---indicates variable is significant at 1% level;
**--- indicates variable is significant at 5% level;
*---indicates variables is significant at 10% level.
e^{β} is the exponential of parameter estimate (β).

Table 4.4 Empirical Analysis Using Death Rates of 2000.

Variables	Parameter Estimates(β)	e^{β}	Pr>ChiSq
Main effect			
Intercept	-6.2846***		<.0001
AGE1 (age between 25 and 34 is the reference level for age)			
AGE2 (35 to 44)	0.7292**	2.07	0.0472
AGE3 (45 to 54)	1.3368***	3.81	0.0001
AGE4 (55 to 64)	2.0853***	8.05	<.0001
AGE5 (65 to 74)	2.8816***	17.84	<.0001
AGE6 (75 +)	4.0197***	55.68	<.0001
Education (Primary education is reference)			
secondary	-0.4576***	0.63	0.0068
Postsecondary	-0.8921***	0.41	<.0001
Race (White is reference)	-0.2953	0.74	0.2357
Sex (Female is reference)	0.0731	1.08	0.7211
Interaction Terms			
Race*Education			
Race*Secondary	0.5679**	1.76	0.0496
Race*postsecondary	0.4305	1.54	0.2394
Sex*Education			
Sex*Secondary	0.3337	1.40	0.1489
Sex*Postsecondary	0.1482	1.16	0.5678

Notes:
Observation=120. Scaled Deviance/DF=1.047.
*** ---indicates variable is significant at 1% level;
**--- indicates variable is significant at 5% level;
*---indicates variables is significant at 10% level.
e^{β} is the exponential of parameter estimate (β).

Table 4.5 Comparison of Two Period's Analysis.

	1990			2000			Differences between 1990 and 2000	
Variables	Parameter Estimates	e^{β}	Pr>ChiSq	Parameter Estimates	e^{β}	Pr>ChiSq	ΔParameter Estimates	Δe^{β}
Main effect								
Intercept	-6.3457***		<.0001	-6.2846***		<.0001	0.0611	
AGE1 (age between 25 and 34 is the reference level for age)								
AGE2 (35 to 44)	0.334	1.40	0.136	0.7292**	2.07	0.0472	0.3952	0.68
AGE3 (45 to 54)	1.1422***	3.13	<.0001	1.3368***	3.81	0.0001	0.1946	0.67
AGE4 (55 to 64)	2.0165***	7.51	<.0001	2.0853***	8.05	<.0001	0.0688	0.54
AGE5 (65 to 74)	2.8072***	16.56	<.0001	2.8816***	17.84	<.0001	0.0744	1.28
AGE6 (75 +)	3.9084***	49.82	<.0001	4.0197***	55.68	<.0001	0.1113	5.87
Education (Primary education is reference)								
secondary	-0.3387***	0.71	0.0016	-0.4576***	0.63	0.0068	-0.1189	-0.08
postsecondary	-0.5585***	0.57	<.0001	-0.8921***	0.41	<.0001	-0.3336	-0.16
Race (White is reference)	-0.0007	1.00	0.9962	-0.2953	0.74	0.2357	-0.2946	-0.25
Sex (Female is reference)	0.2942**	1.34	0.012	0.0731	1.08	0.7211	-0.2211	-0.27
Interaction Terms								
Race*Education								
Race*Secondary	0.4253**	1.53	0.0169	0.5679**	1.76	0.0496	0.1426	0.23
Race*postsecondary	0.3925	1.48	0.1384	0.4305	1.54	0.2394	0.038	0.06
Sex*Education								
Sex*Secondary	0.2038	1.23	0.1512	0.3337	1.40	0.1489	0.1299	0.17
Sex*Postsecondary	-0.0573	0.94	0.7362	0.1482	1.16	0.5678	0.2055	0.22

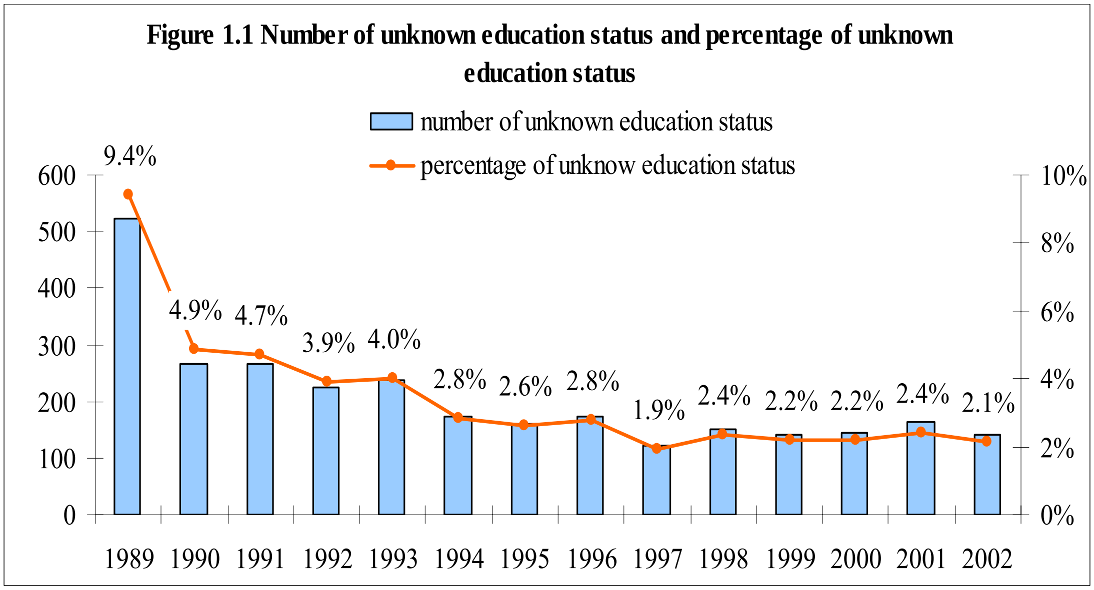

Figure 1.1 Numbers and Percentages of Missing Education Information on Death Certificates, Delaware, 1989 to 2002.

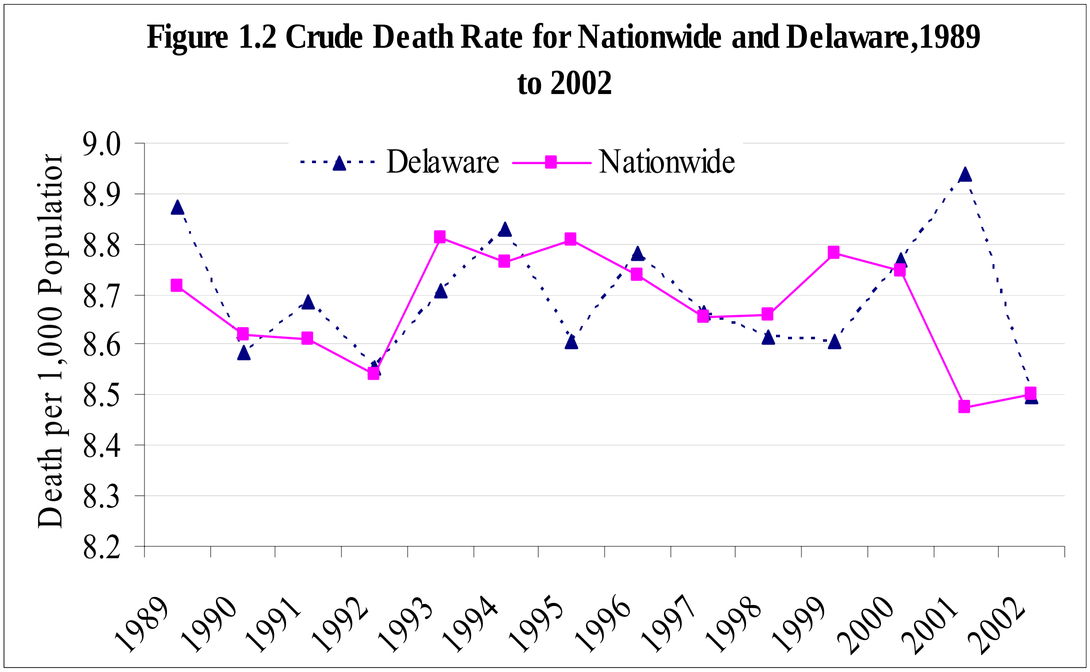

Figure 1.2 Crude Death Rate of Nationwide and Delaware, 1989 to 2002.

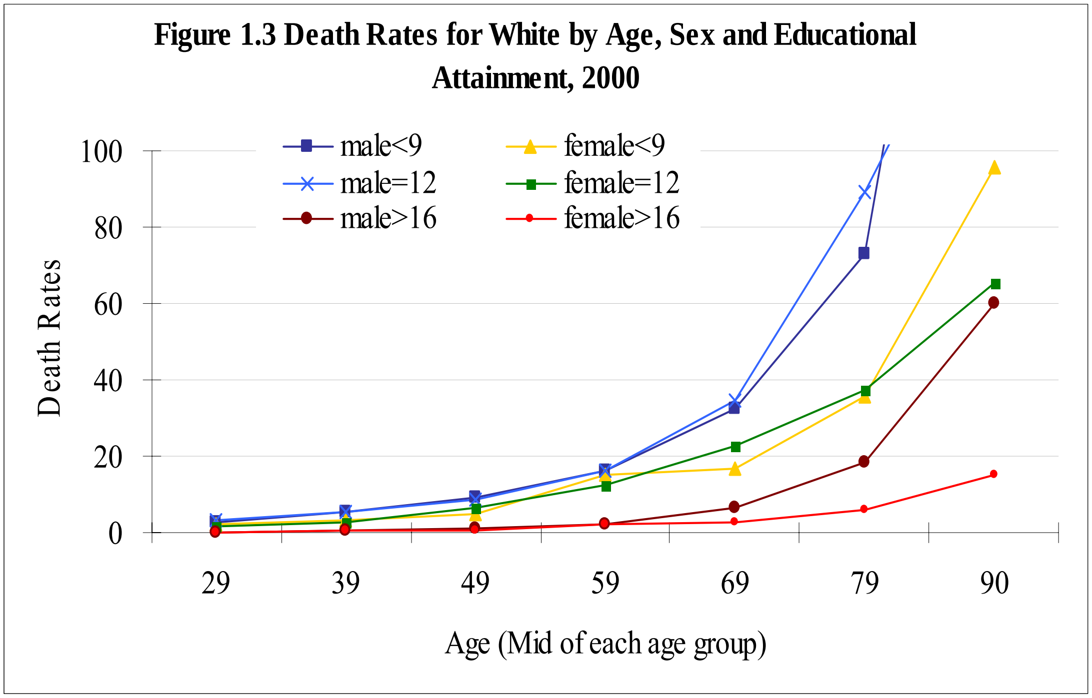

Figure 1.3 Death Rates for White by Age, Sex and Educational Attainment, 2000.

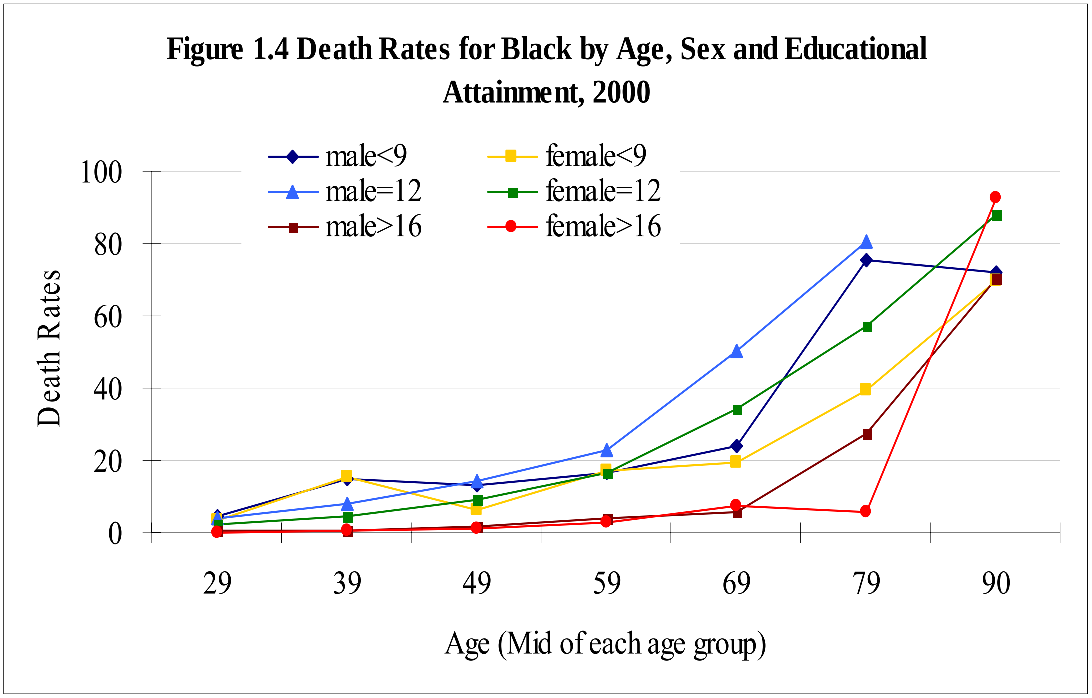

Figure 1.4 Death Rates for Black by Age, Sex and Educational Attainment, 2000.

DOC. NO. 35-05-20-98/07/03

OFFICE OF VITAL STATISTICS

LOCAL REG NO.

CERTIFICATE OF DEATH

State of Delaware (107)

DEPARTMENT OF HEALTH AND SOCIAL SERVICES

STATE FILE NUMBER

DECEDENT

1. DECEDENT'S NAME (FIRST, MIDDLE, LAST) | 2. SEX | 3. DATE OF DEATH (MO., DAY, YR)

4. SOCIAL SECURITY NO. | 5A. AGE (YRS) | 5B. UNDER 1 YEAR MONTHS DAYS | 5C. UNDER 1 DAY HOURS MINUTES | 6. DATE OF BIRTH (MO., DAY, YR.) | 7. BIRTHPLACE (CITY AND STATE OR FOREIGN COUNTRY)

8. WAS DECEDENT EVER IN U.S. ARMED FORCES? ☐ YES ☐ NO | 9. ANATOMICAL GIFT ☐ CONSENT GRANTED ☐ NOT GRANTED | 10A. PLACE OF DEATH (CHECK ONLY ONE, SEE INSTRUCTIONS ON OTHER SIDE) HOSPITAL ☐ INPATIENT ☐ ER/OUTPATIENT ☐ DOA OTHER ☐ NURSING HOME ☐ RESIDENCE ☐ OTHER (SPECIFY)

10B. FACILITY NAME (IF NOT INSTITUTION GIVE STREET AND NUMBER) | 10C. CITY, TOWN, OR LOCATION OF DEATH | 10D. COUNTY OF DEATH

11. MARITAL STATUS — MARRIED, NEVER MARRIED, WIDOWED, DIVORCED (SPEC.) | 12. MOST RECENT SPOUSE ☐ LIVING ☐ DECEASED (MAIDEN NAME IF WIFE) | 13A. DECEDENT'S USUAL OCCUPATION (KIND OF WORK DURING MOST OF WORKING LIFE. DO NOT USE RETIRED) | 13B. KIND OF BUSINESS/INDUSTRY

14A. RESIDENCE — STATE | 14B. COUNTY | 14C. CITY, TOWN, OR LOCATION | 14D. STREET AND NUMBER

14E. INSIDE CITY LIMITS? (YES OR NO) | 14F. ZIP CODE | 15. WAS DECEDENT OF HISPANIC ORIGIN? (SPECIFY NO OR YES, SPECIFY CUBAN, MEXICAN, PUERTO RICAN, ETC.) ☐ NO ☐ YES (Specify) | 16. RACE — AMERICAN INDIAN, BLACK, WHITE, ETC. (SPECIFY) | 17. DECEDENT'S EDUCATION (SPECIFY ONLY HIGHEST GRADE COMPLETED) ELEMENTARY/SECONDARY (0-12) COLLEGE (1-4 OR 5+)

TO FUNERAL DIRECTOR: After certificate has been signed by attending physician and completely filled in by funeral director remove carbons, file parts 1 and 2 with Registrar within 72 hrs. after death and then use Burial-Transit Permit for disposition of body.

PARENTS

18. FATHER'S NAME (FIRST, MIDDLE, LAST) | 19. MOTHER'S NAME (FIRST, MIDDLE, MAIDEN SURNAME)

INFORMANT

20A. INFORMANT'S NAME | 20B. MAILING ADDRESS (STREET AND NUMBER OR RURAL ROUTE NUMBER, CITY OR TOWN, STATE, ZIP CODE)

DISPOSITION

21A. METHOD OF DISPOSITION ☐ BURIAL ☐ CREMATION ☐ REMOVAL FROM STATE ☐ DONATION ☐ OTHER (SPECIFY) | 21B. PLACE OF DISPOSITION (NAME OF CEMETERY, CREMATORY, OR OTHER PLACE) | 21C. LOCATION (CITY, TOWN, STATE)

22A. SIGNATURE OF FUNERAL SERVICE LICENSEE ▶ | 22B. LICENSE NUMBER (OF LICENSEE) | 23. NAME AND ADDRESS OF FACILITY

DRAFT

24. REGISTRAR'S SIGNATURE ▶ | 25. DATE FILED (MO., DAY, YR.)

PRONOUNCING OFFICIAL

COMPLETE ITEMS 26 A-C ONLY WHEN CERTIFYING PHYSICIAN IS NOT AVAILABLE AT TIME OF DEATH TO CERTIFY CAUSE OF DEATH | 26A. TO THE BEST OF MY KNOWLEDGE DEATH OCCURRED AT THE TIME, DATE, AND PLACE STATED. SIGNATURE AND TITLE ▶ | 26B. LICENSE NUMBER | 26C. DATE SIGNED (MO., DAY, YR.)

ITEMS 27-29 MUST BE COMPLETED BY PHYSICIAN OR HOSPICE NURSE WHO PRONOUNCES DEATH

27. TIME OF DEATH ☐ AM ☐ PM | 28. DATE PRONOUNCED DEAD (MO., DAY, YR.) | 29. WAS CASE REFERRED TO MEDICAL EXAMINER? (YES OR NO)

SEE DEFINITION ON OTHER SIDE

CERTIFIER

30A. CERTIFIER (CHECK ONLY ONE)

☐ CERTIFYING PHYSICIAN (Physician certifying cause of death when another physician has pronounced death and completed item 26) **To the best of my knowledge, death occurred due to the cause(s) and manner as stated.**

☐ PRONOUNCING AND CERTIFYING PHYSICIAN (Physician both pronouncing death and certifying the cause of death) **To the best of my knowledge, death occurred at the time, date, and place, and due to the cause(s) and manner as stated.**

☐ MEDICAL EXAMINER **On the basis of examination and/or investigation, in my opinion, death occurred at the time, date, and place, and due to the cause(s) and manner as stated.**

30B. SIGNATURE AND TITLE OF CERTIFIER ▶ | 30C. LICENSE NUMBER | 30D. DATE SIGNED (MO., DAY, YR.)

31. NAME AND ADDRESS OF CERTIFIER WHO COMPLETED CAUSE OF DEATH (ITEM 40) (TYPE/PRINT)

32A. WAS AN AUTOPSY PERFORMED? ☐ YES ☐ NO | 33. MANNER OF DEATH ☐ NATURAL ☐ ACCIDENT ☐ SUICIDE ☐ HOMICIDE ☐ PENDING INVESTIGATION ☐ UNDETERMINED | 34. INJURY AT WORK? ☐ YES ☐ NO | 35. DATE OF INJURY (MO., DAY, YR.) | 37. DESCRIBE HOW INJURY OCCURRED

32B. WERE AUTOPSY FINDINGS AVAILABLE PRIOR TO COMPLETION OF CAUSE OF DEATH? ☐ YES ☐ NO | 36. TIME OF INJURY ☐ AM ☐ PM | 38. PLACE OF INJURY (AT HOME, FARM, STREET, FACTORY, OFFICE BUILDING, ETC. (SPECIFY)) | 39. LOCATION (STREET AND NUMBER OR RURAL ROUTE NUMBER, CITY OR TOWN, STATE)

40. PART I DO NOT ENTER THE MODE OF DYING SUCH AS CARDIAC OR RESPIRATORY ARREST, SHOCK, OR HEART FAILURE. LIST ONLY ONE CAUSE PER EACH LINE | APPROXIMATE INTERVAL BETWEEN ONSET AND DEATH

IMMEDIATE CAUSE (FINAL DISEASE, INJURY OR CONDITION THAT IN YOUR OPINION CAUSED THE DEATH) — IMMEDIATE CAUSE (A)

DUE TO (B)

SEQUENTIALLY LIST CONDITIONS, IF ANY, LEADING TO IMMEDIATE CAUSE. ENTER UNDERLYING CAUSE (DISEASE OR INJURY WHICH INITIATED EVENTS RESULTING IN DEATH) LAST — DUE TO (C)

DUE TO (D)

PART II OTHER SIGNIFICANT CONDITIONS— CONTRIBUTING TO CAUSE OF DEATH

TO HOSPITAL OR PHYSICIAN — DELAWARE LAW REQUIRES THAT THE DEATH CERTIFICATE BE EXECUTED WITHIN 72 HOURS AFTER DEATH

REV. 8/98

(1) ORIGINAL COPY—STATE

APPENDIX A STATE OF DELAWARE CERTIFICATE OF DEATH

REFERENCES

1990 Census of Population Social and Economic Characteristics Delaware.

2000 Census of Population Social and Economic Characteristics Delaware.

Allison, P. D. (2001). "Logistic Regression Using the SAS System: Theory and Application." Wiley-SAS. Chapter 9.

Bannister, J., and S. H. Preston (1981). "Mortality in China." Population and Development Review 7, 1, 98-110.

Becker, G. S. (1962). "Investment in Human Capital: A Theoretical Analysis." The Journal of Political Economy 70, 5, Part 2: Investment in Human Beings, 9-49.

Becker, G. S. (1992). "Human Capital and the Economy." Proceedings of the American Philosophical Society 136, 1, 85-92.

Boulier, B.L., and V. B. Paqueo (1988). "On the Theory and Measurement of the Determinants of Mortality." Demography 25, 2, 249-63.

Cantoyannis, P., and A. M. Jones (2004). "Socio-economic Status, Health and Lifestyle." Journal of Health Economics 23, 5, 965-995.

Case, A., and C. Paxson (2005). "Sex Differences in Morbidity and Mortality." Demography 42, 2, 189-214.

Chakraborty, S., and M. Das (2005). "Mortality, Human Capital and Persistent Inequality." Journal of Economic Growth 10, 20, 159-192.

Christenson, B. A., and N. E. Johnson (1995). "Educational Inequality in Adult Mortality: An Assessment with Death Certificate Data from Michigan. " Demography 32, 2, 215-229.

Deaton, A., and C. Paxson (1999). "Mortality, Education, Income, and Inequality among American Cohorts." NBER Working Papers N0 7140.

Delaware Vital Statistics Annual Report. Various years.

Financial Overview Fiscal Year 2002. State of Delaware, Office of the Governor.

Hahn, R. A., S. F. Wetterhall, G. A. Gay, D. S. Harshbarger, C. A. Burnett, R. G. Parrish, and R. J. Orend (2002). "The Recording of Demographic Information on Death Certificate: A National Survey of Funeral Director." Public Health Report 117.

Hayward, M. D., and M. Heron (1999). "Racial Inequality in Active Life among Adult Americans." Demography 36, 1, 77-91.

Hill, M. E. (1999). "Multivariate Survivorship Analysis Using Two Cross-Sectional Samples." Demography 36, 4, 497-503.

Hobbs, F., and N. Stoops (2002). "Demographic Trends in the 20th Century." Census 2000 Special Reports. Available online at: http://www.cjcj.org/pubs/classdis/classdis.html (accessed Sep 2006.)

Hu, Y., and N. Goldman (1990). "Mortality Differentials by Marital Status: An International Comparison." Demography 27, 2, 233-250.

Huie, B., P. M. Kruger, R. G. Rogers, and R. A. Hummer (2003). "Wealth, Race, and Mortality." Social Science Quarterly 84, 3, 667-84.

Hummer, R. A., R. G. Rogers, and I. W. Eberstein (1998). "Sociodemographic Differentials in Adult Mortality: A Review of Analytic Approaches." Population and Development Review 24, 3, 553-578.

Johnson, G. "SAS Software to Fit the Gneralized Model." SAS Institute Inc.

Johnson, N. J., P. D. Sorlie, and E. Backlund (1999). "The Impact of Specific Occupation on Mortality in the U.S. National Longitudinal Mortality Study." Demography 36, 3, 355-367.

Kitagawa, E. M., and P. M. Hauser (1973). Differential Mortality in the United States: A Study in Socioeconomic Epidemiology. Harvard University Press. Cambridge, MA.

Kitagawa, E.M. (1977). "On Mortality." Demography 14, 4, 381-389.

Lindsey, J. K. (1997). "Applying Generalized Linear Models." Springer. Page 209.

Lochner, L., and E. Moretti (2001). "The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-reports." National Bureau of Economic Research Working Paper Series 8605. Available online at http://www.nber.org/papers/w8605.pdf (accessed Nov 2006).

Macallair, D. (1998). “Class Dismissed: Higher Education vs. Corrections During the Wilson Years.” Center on Juvenile and Criminal Justice. Available online at http://www.cjcj.org/pubs/classdis/classdis.html (accessed Sep 2006).

Martikainen, P., T. Martelin, E. Nihtila, K. Majamaa, and S. Koskinen (2005). “Differences in Mortality by Marital Status in Finland from 1976 to 2000: Analyses of Changes in Marital-status Distributions, Socio-demographic and Household Composition, and Cause of Death.” Population Studies 59, 99-115.

McCullagh, P., and J. A. Nelder (1989). “Generalized Linear Models.” First CRC Press.

Molla, M. T., J. H. Madans, and D. K. Wagener (2004). “Differentials in Adult Mortality and Activity Limitation by Years of Education in the United States at the End of the 1990s.” Population and Development Review 30, 4, 625–646.

Moore, D. E., and M. D. Hayward (1990). “Occupational Careers and Mortality of Elderly Men.” Demography 27, 1, 31-53.

Muney, A. L. (2005). “The Relationship between Education and Adult Mortality in the United States.” Review of Economic Studies 72, 189–221.

National Vital Statistics Report 54, 14, 30-33. (2006). Available online at http://www.cdc.gov/nchs/data/nvsr/nvsr54/nvsr54_14.pdf (accessed Nov 2006).

Pigou, A. C. (1928). “A Study in Public Finance.” Macmillan. London.

Potter, L. B. (1991). “Socioeconomic Determinants of White and Black Males' Life Expectancy Differentials, 1980.” Demography 28, 2, 303-21.

Powers, D. A., and Y. Xie (2000). “Stitistical Methods for Categorical Data Analysis.” Academic Press. Chapter 5.

Preston, S. H. (1970). “The Age-Incidence of Death from Smoking.” Journal of American Statistics Association 65, 331, 1125-1130.

Preston, S. H. (1975). “The Changing Relation between Mortality and Level of Economic Development.” Population Studies 29, 2, 231-248.

Preston, S. H. (1977). “Mortality Trends.” Annual Review of Sociology 3, 163-178.

Preston, S. H. (1996). “Population Studies of Mortality.” Population Studies 50, 3, 525-536.

Rogers, R. G. (1992). “Living and Dying in the U.S.A.: Sociodemographic Determinants of Death among Blacks and Whites.” Demography 29, 2, 287-303.

Rogers, R. G., R. A. Hummer, P. M. Krueger, and F. C. Pampel (2005). “Mortality Attributable to Cigarette Smoking in the United States.” Population and Development Review 31, 2, 259-292.

Rogot, E., P. D. Sorlie, N. J. Johnson, C. S. Glover, and D. W. Treasure (1988). “Mortality Study of One Million Persons by Demographic, Social and Economic Factors: 1979-1981 Follow-up. “ First Data Book. NIH Publication No 88-2896.

SAS Online Documentation Version 8 (1999). Available at http://www.tau.ac.il/cc/pages/docs/sas8/ (accessed Mar 2006).

Schultz, T. P. (2003). “Human Capital, Schooling and Health.” Economics and Human Biology 1, 2, 207-221.

Smith, T. E. (1967). “The Control of Mortality.” Annals of the American Academy of Political and Social Science 369, World Population, 16-25.

Smits, J., I. Keij-Deerenberg, and G. Westert (2005). “Effects of Socio-economic Status on Mortality: Separating the Nearby from the Farther Away.” Health Economics 14, 6, 595-608.

Stokes, M. E., C. S. Davis, and G. G. Koch (2000). “Categorical Data Analysis Using the SAS System.” SAS Publishing.

Swayne, D. F., D. Cook, A. Buja, H. Hofmann, and D. T. Lang (2005). “Interactive and Dynamic Graphics for Data Analysis: With Examples Using R and GGobi” Iowa State University. Available online at http://www.public.iastate.edu/~dicook/ggobi-book/chap-miss.pdf (accessed Sep 2006).

Tuckman, J., W. F. Youngman, and G. B. Kreizman (1965). “Occupational Level and Mortality” Social Forces 43, 4, 575-577.

Turcotte, J., and L. W. Rennison (2004). “The Link between Technology Use, Human Capital, Productivity and Wages: Firm-Level Evidence.” International Productivity Monitor 9, 25-36.

_______. (2006). “Annotated Output Poisson Regression.” UCLA Academic Technology Service. Available online at http://www.ats.ucla.edu/stat/sas/output/sas_poisson_output.htm (accessed Oct 2006).

Usher, D. (1997). “Education as a Deterrent of Crime.” The Canadian Journal of Economics 30, 2.

Wikipedia. Available online at http://en.wikipedia.org/wiki/ (accessed Mar 2006).