INFANT MORTALITY BY MONTH OF BIRTH: AN ANALYSIS OF CONTEMPORARY COHORTS

A Dissertation

by

RACHEL TRAUT CORTES

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2010

Major Subject: Sociology

INFANT MORTALITY BY MONTH OF BIRTH: AN ANALYSIS OF CONTEMPORARY COHORTS

A Dissertation

by

RACHEL TRAUT CORTES

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Dudley L. Poston, Jr.
Committee Members,	Mark Fossett
	Rogelio Saenz
	Douglas Wunneburger
Head of Department,	Mark Fossett

May 2010

Major Subject: Sociology

ABSTRACT

Infant Mortality by Month of Birth: An Analysis of Contemporary Cohorts.

(May 2010)

Rachel Traut Cortes, B.S., The University of Texas at San Antonio; M.S., Texas A&M University

Chair of Advisory Committee: Dr. Dudley L. Poston, Jr.

There is a well-established connection between adult mortality and the conditions an individual is exposed to while *in utero*. There is a wealth of research that connects conditions such as asthma and allergies, mortality due to heart disease and diagnoses of schizophrenia to conditions during an individual's early life and even their time *in utero.* The aim of this dissertation is to see if this same connection can be made to infant mortality, and further will there be any connection in contemporary cohorts? I use the Linked Birth/Infant Death dataset available from the Centers for Disease Control (CDC) for the years 2000 to 2004. This dissertation specifically uses the dependent variable "cause specific infant death" with various measures of the time the infant was born or was *in utero*. I undertake three multinomial logistic regression models with the dependent variable "cause specific infant death." I then proceed to a multilevel multinomial logistic regression model using state-level climate measures at the second

level. I conclude with the construction of maps displaying the spatial relationship between infant mortality and climate.

The first analysis uses the independent variable of interest “month of birth,” the second analysis uses the independent variable of interest “months of first trimester,” and the last level-one analysis uses the independent variable of interest “months of third trimester.” After running all three models, I determined that the most effective independent variable of interest is “month of birth,” which I use in a multilevel logistic regression model.

The multilevel model uses the month of birth variable at level-one and incorporates state level measures of climate at the second level. I find that the humidity index and the temperature index are negatively associated with the month of birth variable and cause specific infant death variables, meaning that the higher these indices, the more the benefit to an infant’s chances of survival. The wind index is consistently positive, meaning that the interaction of wind with cause specific infant death and month of birth is detrimental to an infant’s survival.

The last methods chapter shows the spatial relationship between infant mortality and climate. In this chapter I find that infant mortality in the United States is concentrated in the Southern U.S., which is also where there is a concentration of high temperature states. The connections between wind and humidity with the infant mortality rate are less consistent.

Dedicated to the memory of my father, John Traut.

ACKNOWLEDGEMENTS

I would first like to thank Dr. Dudley L. Poston, Jr. my committee chair, who is not only my favorite professor, but is one of my favorite people in the world. Not only is he an exceptional mentor, professor and demographer, but he is an exceptional human being. I am continually inspired by him and his love for demography and academics. I hope that in my life I can make him proud and inspire others the way he inspired me over the past five years.

I would also like to thank my committee members Dr. Mark Fossett, Dr. Rogelio Saenz and Dr. Douglas Wunneburger, for their help and input in this dissertation and in my graduate studies. I would also like to thank Christi Ramirez for going above and beyond her duties to ensure everything in the department runs smoothly. Thanks also go to my family and friends, especially my fellow demographers Heather Kincannon, Chris Russell and Bethany DeSalvo who helped me through the writing process.

I would like to thank Dr James Vaupel and all the researchers and staff at the Max Planck Institute for Demographic Research for hosting me as a guest researcher in the spring of 2009. I was able to complete many pages of this dissertation by utilizing the resources at MPI and was given much appreciated input from several researchers in the area of biodemography.

Thanks and gratitude also goes to my friend Dr. Christine Guarneri, without whom I could not have survived the past five years. She is smart, funny, insightful, and

served as a sounding board for many things during my graduate career for all of which I am extremely grateful.

Lastly, I would like to thank my husband, Alberto Cortes, for continually supporting me through my entire college career. He always knows when I need help, when I need to figure things out alone, when I need to laugh and when I need to cry. He has done more for me than words can express.

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

TABLE Page

CHAPTER I

INTRODUCTION

The analysis of the effect of month of birth on an individual's mortality is not a new topic or line of research. In fact, studies linking the month of one's birth to the resulting mortality advantage or disadvantage can be traced back decades. One article in particular that addressed this issue dates back to 1945 and is entitled "Infant Mortality in Relation to Month of Birth" and was published in the *American Journal of Public Health*. Generally, the research reported in this article found that in terms of mortality from all causes

> August babies have the best chance of surviving to their first birthday, while the next in order are those born in September and July. The least favorable months of birth are January, December, and February. Babies born in January suffer a mortality rate in their first year about 15 percent greater than August babies (Eastman 1945:913).

This pattern of better survival for infants born in the summer months is sustained even when the different causes of death are brought into the equation.

This difference between the months that translates into better survival for infants has been attributed to a variety of sources. These include meteorological variables such as rainfall and temperature, maternal nutrition that varies seasonally, and illness that may be caused by exposure to viruses and infections that the mother may contract while pregnant. Whatever the reason for the seasonal variation in infant mortality, the difference between the months has been observed for many years. The purpose of this dissertation is to examine this effect of month of birth on infant mortality using more

---

This dissertation follows the style of *Demography*.

current data from 2000 to 2004 in order to ascertain whether or not the differences in month of birth are still affecting infant's survival as was shown in past research.

In earlier decades, changes in climate may be assumed to have had greater influences on a pregnant mother. Fewer houses had heating or air conditioning, so when temperatures were extreme, the indoor temperatures were more difficult to regulate than is the situation today. Also, although refrigeration has been around for centuries, the home use of refrigerators was new in the early twentieth century. This affects not only the safe storage of perishable foods but also the ability to store foods when some fresh foods might otherwise not be available. When Eastman (1945) shows the differences in infant mortality by month of birth he notes these differences can be seen month by month, following a seasonal variation that could be linked to any of these seasonal differences. The aim of this dissertation is to see if the same differences exist today, despite the technological improvements in recent decades.

It seems evident that advances in medical technology, knowledge of proper heating and refrigeration for food preparation, access to nutritious food regardless of season, and heating and cooling systems in the home or at work should eliminate the temporal risks that pregnant mothers are under. However, there is some evidence in the neurological literature of a link between being diagnosed with schizophrenia and the month of the person's birth. Tramer was the first to show in 1929 that there was an excess of winter to early spring births among schizophrenia patients and a decrease in late spring and summer births. Today there are over 200 articles in the research literature that demonstrate the same finding. This association of winter to early spring births with

schizophrenia coincides with the findings of Eastman showing higher risks of infant mortality in January, February and December.

Current research on schizophrenia and month of birth aims to identify what could be the cause of these increases of schizophrenia among those born in winter to early spring. The most studied variables are temperature and viral causes, but hypothesized links between season and schizophrenia may also revolve around maternal hormones, sperm quality, nutrition, and external toxins (Tochigi et al: 2004). Torrey and colleagues (1997) found in the Northern Hemisphere there was an excess of schizophrenia patients born in the months from January to April, and in the Southern Hemisphere there was an excess of schizophrenia patients born in the months from July to September, showing the complete opposite pattern. Interestingly, they found no pattern in the birth months of schizophrenia patients in equatorial regions. This would seem to indicate that the reason for variation in schizophrenia could be linked to temperature and meteorological variables that vary greatly by the regions discussed. The reason this research on schizophrenia is so important to my dissertation is that it shows that there is something about the month of birth that is still operable. This research on schizophrenia gives merit to the overall objective and approach of my dissertation, namely, examining differences in infant mortality by month of birth with current data.

Among those who believe that the reason for the link between schizophrenia and birthdays in the winter and early spring months is the mother's contraction of infections while pregnant, or by the child in early life, is Dr. Paul Ewald. This is discussed in the book: *The Next Fifty Years: Science in the First Half of the Twenty-First Century*. This

book contains twenty-five essays by leading scientists that discuss where they see science to be headed and what it will discover between the years of 2000 and 2050. The twenty-fifth essay is by Dr. Ewald, a professor at Amherst College and an expert in evolutionary medicine. In this essay he discusses advances in medicine that will result in better indicators of the origins of chronic diseases. Ewald asserts that many chronic diseases that today cause serious health repercussions and sometimes death will be linked to infectious diseases in the next 50 years. As an example Ewald mentions the countless studies that have linked schizophrenia to the presence of infectious diseases of the mother while *in utero*. This means that adults who have been diagnosed with schizophrenia are likely to have mothers who contracted infectious diseases while pregnant (Ewald 2002: 297). This is thought to be due to the increased contraction of colds and influenzas during the winter and early spring months, which are thought to affect the fetus if the mother becomes infected. The reduced likelihood of contracting infectious diseases during the summer months would therefore put those children born in these months at a comparably reduced risk. Although the availability of nutritious food and heating and air conditioning indoors may reduce some of the risks to which pregnant mothers are exposed, the increase in contraction of colds and influenzas can still be prevalent, regardless of decade. If Dr Ewald is correct in his assertion that chronic diseases will be linked to early life infections, we can assume that an individual's early life condition is even more important to their health and survival than is currently thought. The intention of my dissertation is to ascertain whether this effect on health may be seen even in the first year of an individual's life.

As I have already noted, the analysis of variations in mortality by month of birth is not a novel idea. Currently, researchers are looking at the link between early life conditions and later life mortality, for individuals fifty years and older. Years earlier it was evident that there were excess deaths in the winter months and in the extreme heat of summer months. Today, however, "seasonal effects in demographic variables are rarely at the center of attention in population studies" (Rau 2007:2). Some of the current research from the Max Planck Institute for Demographic Research has concentrated on examining this relationship among contemporary cohorts (See Rau 2007; Doblhammer 2004; Doblhammer and Vaupel 2001). The findings of these researchers on the link between childhood and in utero conditions on adult mortality further encourage the study of a similar impact on infant mortality.

My dissertation will consist of nine chapters. Following this introductory chapter, Chapter II will describe the existing literature in the areas of infant mortality, seasonality and the link between season of birth and mortality. Chapter III will provide a discussion of the methods to be used in the analyses by reviewing multinomial logistic regression and multilevel multinomial logistic regression, both of which I will use to examine infant mortality by month of birth. In Chapter IV I will describe the Linked Birth/Infant Death dataset from the Centers for Disease Control (CDC) that will be used in my analyses. Chapter V will describe the results of the multinomial logistic regression using the infant's month of birth as the dependent variable. Chapter VI will repeat the analyses of Chapter V, but will use the infant's month of gestation as the main independent variable of interest. Once it is determined, by the results in Chapters IV and V, which

independent variable measuring the effect of season are most appropriate, the results of the multilevel logistic regression will be presented and analyzed in Chapter VI. Chapter VII will display the results of causes-specific infant mortality by month of birth or gestation using mapping methods. The final chapter will summarize and discuss the findings of the analyses.

This dissertation will contribute to the literature in some important ways. First, through analyzing contemporary data on infant mortality, it will be shown whether or not there is a measurable effect of season on infant mortality. The relationship between month of birth and later life mortality and health is well documented for cohorts aged fifty and older at the time of study; however the same relationship between season and infant mortality has not been demonstrated using current data. This dissertation will look at infant mortality from the years 2000-2004 to see if month of birth is still an important variable impacting an infant's health and survival.

Second, my dissertation will perform analyses using two different main independent variables of interest, namely, month of birth and month of gestation. I hope to be able to ascertain whether the important issue when studying the effects of maternal health on infant survival are the conditions of the month in which the infant was born, or the conditions of the month in which the infant was in utero. These ideas draw from a variety of sources that examine the month that the fetus is most susceptible during gestation. Is the fetus at the highest risk if it is born in the winter? Or is it at the greatest risk if it is utero in the winter? My dissertation will draw on relevant literature to identify the more important variable and include this in the final analyses.

Third, my dissertation will use multilevel models in order to incorporate state-level data in the study of infant mortality by month of birth or month of gestation. Once the most relevant variable measuring the seasonal effect is determined, a multilevel model will be estimated. The strength of multilevel models is that state-level variables can be appropriately incorporated into analyses of individual level data in order to see their effect. At the state level I will be able to incorporate variables on climate, in order to examine hypotheses about month of birth's effect on mortality that are based on temperature variation and the availability of foods that are affected by climate. If these variables have an effect then they will show relationships with infant mortality when included with the individual level variables of interest. Multilevel modeling is the most appropriate way to incorporate state or other non-individual level variables into this analysis.

Fourth, this dissertation will use cause specific infant mortality as the dependent variable in all my analyses. In order to estimate a multinomial logistic regression the dependent variable must be a nominal categorical variable that is not ordered. For the purposes of my dissertation I will group the causes of death as they are listed on the death certificates into several categories and use this as the dependent variable. The advantage of using this type of dependent variable is that I will be able to see not only if there is variation in infant mortality by month of birth, but also if there is any variation between the specific causes of death. This is intended to further elucidate what exactly causes the differences by month of birth, based on the causes of the infant's death.

Infant mortality is an important and multifaceted topic of study in demography. Throughout the world infant mortality is used as a "key social indicator" (Frisbie 2005:251). This means that many times to assess the level of advancement of a population a researcher will look at the infant mortality rate. Countries with high infant mortality rates are thought to be less advanced terms of medical technology and overall levels of modernization and development than populations with low infant mortality rates. With advances in technology and medicine, many infants who would never have a chance at survival in previous decades can now be carried to term and born with minimal, if any, complications. However, despite these advances, the infant mortality rate of the United States still remains high, especially when compared to the rates of other developed nations. Although it is unlikely that determining the most susceptible month of birth will reduce infant mortality, being able to determine the causes of death that are significantly associated with month of birth may shed light on the reasons that infant deaths occur and therefore what measures could be taken to reduce those deaths. By examining in further detail the impact that month of birth has on an infant's chance of survival, I hope to be able to elucidate one aspect of infant mortality in the U.S. and the world.

CHAPTER II

LITERATURE REVIEW

In this chapter I overview several topics related to infant mortality and seasonality. First I will discuss the definition of infant mortality and the variations in this definition throughout the world. Then I will review the literature that examines the impact of infancy on adult mortality. I will also include a discussion of the literature on the general impacts of season on mortality. The chapter will conclude with a discussion of causes of infant death. I will show that these areas of past research support the need for analyses of infant mortality and month of birth such as those I perform later in this dissertation.

Infant Mortality: Definition and Measurement

Before delving into issues of infant mortality by month of birth, I will first define what is meant by infant mortality. It is important to define infant mortality as it is used in the United States because this definition applies to the data used in later analyses of this dissertation. It is also important to discuss the definition of infant mortality because the definitions and adherence to practices of measuring infant deaths and live births do indeed vary by country. Later in this literature review I will touch on the definitions of infant mortality in countries other than the U.S. and the problems these differing definitions imply for cross-national comparisons. I will first focus on the definitions of infant mortality used in the U.S. and the most common ways that infant mortality is measured.

"Infant mortality refers to the death within the first year of life to persons born alive" (Frisbie 2005:255). To further clarify this definition, by live birth is meant "the

complete expulsion or extraction from its mother of a product of conception, irrespective of the duration of pregnancy, which after such separation breathes or shows any other evidence of life"; by death is meant "the permanent disappearance of life any time after live birth has taken place" (Frisbie 2005:255-256). This definition of death "complements that of a live birth" (Siegel and Swanson 2004:372), because it excludes all types of fetal deaths (stillbirths, miscarriages, and abortions) since "fetal death refers to the disappearance of live *prior* to the expulsion or extraction from its mother of a product of conception" (Siegel and Swanson 2004:372).

The World Health Organization (WHO) "recommends that a birth be considered live if the newborn shows any one of the following signs of life: heart beat, breathing, umbilical cord pulsation, or voluntary muscle movement" (Haub 1991:7). In the U.S. the standard definition adheres to the WHOs recommendation and also to the following statement recommended by the American College of Obstetricians and Gynecologists to assist in "determining what should be considered a live birth: 'Heartbeats are to be distinguished from transient cardiac contractions; respirations are to be distinguished from fleeting respiratory efforts or gasps" (Siegel and Swanson 2004:371-372). The purpose of these guidelines is to ensure that the vital registration data derived from birth and death certificates are as reliable and valid as possible and also to ensure that all infant deaths, stillbirths and live births are recorded. The U.S. is known for its strict adherence to these measures.

It is important to note that while this definition of infant mortality and the recommendations of the WHO are widely accepted, not all nations adhere strictly to this

definition of a live birth. Consequently, the reported counts of infant deaths can vary widely between countries based on how that country defines and measures a live birth. Accordingly there may be wide variation in rates derived from these occurrences that can make comparison troublesome. However, for the purposes of the analyses conducted in my dissertation, all the data I use will adhere to these definitions of infant mortality, live birth and infant deaths as mentioned above. The problems with comparability of infant mortality rates between the U.S. and other developed nations arise when countries are compared that do not strictly adhere to these guidelines. These issues will be discussed in a later section of this literature review. Next, I will discuss the different measures of infant mortality.

A common way to express measures of infant mortality is through the use of a rate. Rates that measure infant mortality typically use a combination of "population figures from a census with vital statistics" (Siegel and Swanson, 2004: 10). These "[r]ates or ratios have a vital event as the numerator and a population (count) as the denominator…" (Siegel and Swanson 2004: 10). As such, measures of infant mortality take live births for a given population in a specified time period and are interested in determining of the live births that occurred, how many resulted in deaths before their first year of life? The simplest and most widely used measure is the infant mortality rate. The infant mortality rate (IMR) is calculated as the number of deaths to infants less than one year of age divided by the total live births in a calendar year multiplied by 1,000. In the world, the IMR ranges from the single digits in the developed world to over 150 per 1,000 in the developing nations. Usually, the lower the infant mortality rate, the more

developed the country is considered to be in terms of health care and public health measures.

Although the infant mortality rate is the most widely used, there are several variations of the IMR that distinguish between the different periods of the first year of life. For the infant, the risk of death declines with each second, minute, hour, day, and month that the infants are alive. This means that "deaths are not evenly distributed over the first 12 months" and the risk of death declines the longer the infant lives (Rowland, 2003:199). Because of this phenomenon it is typical for the first year of an infant's life to be separated into different time periods. Each of these time periods have a different rate or ratio that is aimed at measuring the relative occurrence of infant death in this time period. The first distinction is between the neonatal and post-neonatal periods. "Neonatal mortality refers to death in the first 28 days of life" (Rowland 2003:201). It is calculated by dividing the number of deaths in the first 28 days after birth by the total number of live births in a calendar year multiplied by 1,000. "Post-neonatal deaths occur between 28 days and the first birthday" (Rowland 2003:201) and is calculated by dividing the number of deaths between 29 days and one year after birth by the total number of live births in a calendar year multiplied by 1,000. The infant mortality rate can be found by adding the neonatal and post-neonatal mortality rates together, since these two measures encompass deaths between birth and one year of life and they both use the same denominator.

The distinction between the neonatal and post-neonatal periods is important to note because of the rough association of each period with exogenous or endogenous

causes of death. An exogenous cause of death is one due to factors that are outside of the body, such as infections and accidents (Rowland 2003:201). An endogenous cause of death is due to "agents operating within the body, leading to biological defects in the new-born as well as degenerative diseases of later life" (Rowland 2003:201). Although the distinction is not so cut-and-dry, it is usually thought that neonatal mortality is due mostly to endogenous causes, and that post-neonatal mortality is due mostly to exogenous causes. This distinction was thought to be more appropriate in past decades, when most of the causes of post-neonatal mortality could be attributed to exogenous causes. However, Poston and Rogers (1985) have showed that post-neonatal mortality cannot always be used as a proxy for exogenous causes of infant mortality. A further discussion of causes of infant death will be found in a later section of this literature review.

Two other important measures of infant mortality are the perinatal mortality rate and the stillbirth rate. The perinatal mortality rate is calculated as the number of deaths that occur between 28 weeks gestation and 28 days after birth divided by the total number of live births and stillbirths in the year multiplied by 1,000. The stillbirth rate is calculated as the number of deaths from 28 weeks gestation divided by the total number of live births and stillbirths in the year multiplied by 1,000 (Rowland 2003:197). Both of these measures tell us something important about the mortality of infants and stillbirths in a population, although they are not the rates that will be used in this dissertation. Next, I turn to a discussion of the major theoretical foundations of pregnancy outcome research.

Theories of Pregnancy Outcomes

The analyses in this dissertation will involve gathering data on both month of birth and several characteristics of the mother and child in order to determine if a relationship exists between month of birth and infant mortality. These two broad types of independent variables will study the two approaches to pregnancy outcomes that demographers tend to concentrate on: the medical model and the social model (Frisbie 2005:252). The social model stresses "the power of social variables to determine infant survival and the importance of structural change in overcoming disparate outcomes. Medical models stress pathways of frank pathophysiology and their potential interruption through clinical interventions" (McCormack and Wise 1993:555). Until recently, demographic research concentrated mainly on the social aspect of mortality and fertility outcomes. However, current research has expanded to include information from both theoretical perspectives. When both areas are included in analyses, a better idea of the causal mechanisms behind infant mortality can be made. For the purposes of the analyses of this dissertation, both social and biological variables will be included in order to ascertain whether or not month of birth is impacting infant mortality.

A full discussion of the social and biodemographic variables that will be included in the analyses will be found in a later chapter. Next, I will discuss the state of infant mortality in the U.S. and the reason for continuing research in the area of infant mortality.

Infant Mortality in the United States

Infant mortality is a widely studied topic and has been for centuries (Nersesian 1988). In fact, a search of the POPLINE database yielded over 1,000 scholarly journal articles on the subject since 1975, and this may well be an undercount of all studies and articles on the topic of infant mortality. These articles spanned many years and looked at many different populations all with the intent of examining infant mortality in a specific context. In my opinion, infant mortality captures the interest of many researchers because the death of an infant is one of the most tragic events that a family or community can experience. Thus, infant mortality receives a great deal of attention because of the emotional impact of an infant death. It is unlikely that researchers will lose interest in the study of infant deaths. As such, infant mortality will continue to be studied around the world in many different contexts for years to come. This section of the literature review will discuss the importance of studying infant mortality in the United States and also place infant mortality in the U.S. in a worldwide context.

Many public health measures aim to reduce infant mortality worldwide. There have been continuing public health efforts in the U.S. and around the world to reduce infant mortality, although infant mortality will always exist to some extent. One of the lowest infant mortality rates in the world today is in Sweden, a western European country that is well known for its effort to eliminate poverty and provide comprehensive health care to all citizens (Hogue and Hargraves 1993). Sweden's infant mortality rate is 2.75 deaths per 1,000 live births, which, according to the CIA World Factbook (2006), is second only to Singapore whose rate is 2.30 deaths per 1,000 live births. In this same

CIA ranking the U.S. has an infant mortality rate of 6.30, putting the U.S.'s rate at more than twice that of Sweden's rate. It is unlikely that any country will ever report an infant mortality rate of zero or near zero; thus the above measure for Sweden probably represents one of the lowest real-life rates of infant mortality that are likely to exist.

Beyond the emotional impact of an infant death, infant mortality is also widely studied and measured because it is used worldwide as an indicator of social development, health and advancement of a population. Those countries with infant mortality rates that are low are considered to be more advanced and healthier than those countries with high infant mortality rates (Frisbie 2005:251). This is an important reason that infant mortality continues to be studied around the world, and why the rate in the U.S. being higher than expected is considered to be so troublesome. Because of this use of infant mortality rates, national governments have an interest in lowering infant mortality rates and/or maintaining low rates. This usually occurs through public health efforts to ensure mother and child are healthy throughout and after pregnancy.

There is considerable governmental effort committed to the reduction of infant mortality in the United States through health care and public programs. In fact, the initiative set forth by the U.S. Department of Health and Human Services in conjunction with the Centers for Disease Control (CDC), Healthy People 2010 , makes decreasing infant mortality overall in the U.S. one of its major objectives. The program is especially aimed at reducing the disparities in infant mortality between racial and ethnic minorities and the white majority that are prevalent in the United States. By setting such goals regarding infant mortality rates for the overall population and reducing disparities in the

rates among minorities, strides can be taken to improve the health and well-being of the population and serve as an indicator of the betterment of health conditions in the United States.

One of the broad objectives for the year 2010 is to "reduce fetal and infant deaths" with a goal of reducing the fetal and infant deaths during the perinatal period from 7.5 in 1997 to a rate of 4.5 in the year 2010 (U.S. Dept. of Health and Human Services: 16-12). Perinatal deaths are defined as "fetal deaths after 28 weeks gestation and infant deaths within the first 7 days of birth" (U.S. Dept. of Health and Human Services:16-55). Since perinatal deaths only include infant deaths within the first 7 days after birth, the reduction in the rate excludes many of the infant deaths that may occur. In order to incorporate all deaths to infants within the first year of life, the Healthy People 2010 objectives also aim to reduce infant deaths after the perinatal period, specifically all infant deaths, including neonatal deaths and postneonatal deaths. The goal is to reduce "neonatal ([deaths] within the first 28 days of life)" from 4.8 per 1,000 live births in 1998 to 2.9 per 1,000 live births in the year 2010 (U.S. Dept. of Health and Human Services:16-13). The goal is also to reduce "postneonatal ([deaths] between 28 days and 1 year)" from 2.4 per 1,000 in 1998 to 1.2 per 1,000 in the year 2010 (U.S. Dept. of Health and Human Services: 16-13). Lastly, the overall goal is to reduce all infant deaths from 7.2 per 1,000 live births in 1998 to 4.5 per 1,000 live births in 2010 (U.S. Dept. of Health and Human Services:16-13). If these goals were reached, the U.S. would have infant mortality rates that are more in line with rates from other developed nations. Reaching these goals would reflect an improvement in not only the health outcomes for

infants and fetuses, but would also serve as an indicator of overall improvement in health in the United States.

Many of the goals of the Healthy People 2010 program about maternal, infant and child mortality concentrate on preventative measures. Generally, the goals of the program can be attained through comprehensive prenatal care for all pregnant mothers. Prenatal care is "the health care that [a pregnant mother gets] while pregnant" (www.womenshealth.gov/faq/prenatal-care.cfm#a) and includes frequent doctor visits in which the health of the mother and child is evaluated; it also includes educating the mother in different areas that will help ensure that the pregnancy is successful. Prenatal care will not only help prevent infant deaths, but will also reduce maternal deaths and will help guarantee that all infants are born into the healthiest circumstances possible, therefore reducing child mortality as well. Many studies have looked at the impact of prenatal care on the mother and child. Overall findings suggest that there is an increase in positive health outcomes when prenatal care is used (Gortmaker 1979; Gortmaker and Wise 1997; Kiely and Kogan 1994). It is suggested that the most beneficial effects of prenatal care are found when used by "socially disadvantaged women" (Kiely and Kogan 1994: 105).

The goals of Healthy People 2010 target the large differences in infant mortality between minorities and non-minorities because of the significant effect of this differential on the overall infant mortality rate in the United States. The large difference between infant mortality in the United States and in the rest of the world has been touched on in this literature review thus far. The discrepancy between the infant

mortality rates in the U.S. and the rest of the developed world is thought to be due to two main factors: varying definitions of infant mortality among nations, and health disparities between ethnic and racial groups in the U.S. In the next section of this literature review, I will discuss some of the issues involved in comparing infant mortality in the United States with infant mortality rates in the rest of the world. First, I will review the state of infant mortality for racial and ethnic minorities in the United States.

A report from the National Center for Health Statistics (NCHS) from 2008 reviewed the trends in infant mortality for the United States. This report shows that between the years 2000 and 2005 there was no decline in the infant mortality rate. However, preliminary data from the year of 2006 show that there was a decline of about two percent between the years 2005 and 2006. The lack of decline in the infant mortality rate between 2000 and 2005 "represents that first period of sustained lack of decline in the U.S. infant mortality rate since 1950's" (Matthews and MacDorman 2008: 2). In terms of the goals set forth in the Healthy People 2010 the current infant mortality rate in the U.S. is about 50% higher than the goal. The infant mortality rate of 6.71 in 2006 tells us that there were "more than 28,000 deaths to children under 1 year of age each year in the United States" (Matthews and MacDorman 2008:2).

In 2005 the infant mortality rate for non-Hispanic whites was 5.76; the infant mortality rates for non-Hispanic blacks, Puerto Ricans, and American Indian or Alaskan Natives were above the U.S. average IMR of 6.86. Mexicans, Central and South Americans and Cubans all had IMRs that were below the national average and below the

rate for non-Hispanic whites in the year 2005. The highest rate of infant mortality in the United States is among non-Hispanic blacks with an infant mortality rate of 13.63 per 1,000 live births. The lowest rate of infant mortality is among Cubans with a rate of 4.42 per 1,000 live births. In fact this rate has attained the Healthy People 2010 goal of 4.45 per 1,000 live births. The NCHS report acknowledges that some of the discrepancy in infant mortality between racial and ethnic groups is due to differences in risk factors "such as preterm and low birthweight delivery, socioeconomic status, and access to medical care. Even after making these considerations, we find that "many of the racial and ethnic differences in infant mortality remain unexplained" (Matthews and MacDorman 2008:3).

A wealth of literature has looked at these differences in infant mortality between the white majority and the racial and ethnic minorities. There are two important points to consider when looking at this kind of differential infant mortality. The first is the sustained high infant mortality among non-Hispanic blacks in the U.S., and the second is the surprisingly low infant mortality among Hispanics in the U.S. Both topics have been well researched and will only be mentioned briefly here in order to clarify the context of infant mortality in the U.S. between the majority and minority groups.

The situation of infant mortality for non-Hispanic blacks in the U.S. is surely a serious health issue. As stated previously, the IMR for non-Hispanic blacks was 13.63 in 2006, compared to an IMR of 5.76 for non-Hispanic whites in the same year. This means that the IMR for blacks is almost three times as high as that for whites. Researchers have examined this differential at depth in hopes of finding some of the reasons why blacks in

the U.S. lag so far behind white and other minorities. In general these differences are thought to be due to centuries-old consequences of discrimination and racism in the United States that put black Americans at a disadvantage socioeconomically. Overall there is a lack of adequate health insurance and access to health care among blacks that makes obtaining consistent and comprehensive prenatal care near impossible. Compounding the effects of poor health care are the increases in pregnancies among black unmarried and teenage mothers as well as low levels of income and education of black mothers and fathers. Studies have also concentrated on the health effects of mother and child of residential segregation, and have found that being segregated can compound these effects (Polednak 1991; Polednak 1996; LaVeist 1989). For the purposes of this dissertation it is important to note that these differences exist and that the heterogeneity of the U.S. population is responsible for much of the relatively low international ranking of the U.S. among other developed nations (Matteson et al. 1998).

Conversely, the situation of infant mortality for Hispanics in the U.S. is surprisingly favorable. Despite the overall poor socioeconomic standing of Hispanics in the U.S., the infant mortality among many Hispanic nationalities is very low, in fact lower than that of the white majority for some groups in 2006 (Matthews and MacDorman 2008:2). This phenomenon of positive health outcomes for Hispanics despite low levels of education and income is termed the “Hispanic Paradox” and is also an area that it heavily studied. This Hispanic Paradox refers to positive health outcomes for Hispanics throughout the lifespan and includes a lower than expected infant mortality rate. Researchers find that although Hispanics are socioeconomically closer to the black

population in the U.S., they have health outcomes that are closer to those of whites. Possible explanations for this are "cultural practices, family supports, selective migration, diet and genetic heritage" (Markides and Coreil 1986: 253; for more discussion of the Hispanic Paradox, see also Markides and Eschbach 2005). In this dissertation it is important to note these differences in the health of Hispanic groups in the U.S. because of the impact they can have on observed infant mortality. It is also important to note the differences between minority and majority groups in the U.S. because this discrepancy between whites and non-whites is thought to impact the standing of infant mortality of the U.S. in the world. The next section of this literature review will discuss the comparison of infant mortality in the U.S. with other developed nations.

Worldwide Comparisons of Infant Mortality

One of the most important reasons for the continual study of infant mortality is that although the United States is among only a select few nations in the world considered to be developed, there are vast differences between the U.S. and other developed nations with regard to level of infant mortality. Unfortunately, the United States has an infant mortality rate that is closer to some developing nations than to nations of comparable wealth and advancement. The CIA World Factbook (2006) ranks all countries in terms of infant mortality from highest to lowest. According to this source, the United States ranks 181st out of 222 total countries in terms of its infant mortality rate.

Although this puts the U.S. as the country with the 43rd lowest infant mortality, the reason for concern arises from the fact that all countries ranking higher than the U.S

are industrialized nations, putting the U.S.'s infant mortality rate more on par with the developing nations of the world. Since infant mortality is used as an indicator for the health and social condition of a population (Frisbie 2005), this poor ranking of the U.S. seems to indicate that there is something lacking in the U.S. when it comes to the factors that lead to low infant mortality among all other developed nations. As discussed above, part of this discrepancy is due to the high infant mortality among minorities compared to non-minorities. The other portion of this observed difference is the varying definition of infant mortality and live births in the developed world. This section of the literature review will review the problems that arise with the cross-national comparison of infant mortality rates.

Throughout the world the definition of a live birth varies greatly; a less strict definition therefore can deflate the infant mortality rate for a country. If a country is lax in their categorization of a live birth and does not adhere to the WHOs recommended practice, they may categorize an infant death as a still birth. This means that countries that adhere to the WHOs recommended practice would have more infant deaths, and those countries that do not adhere to the WHOs recommended practice would have an increased number of still births and a lower number of infant deaths. Although it cannot be known from looking at vital registration data whether or not a country is following the WHOs recommended practice, some countries "may denote deviations in their publications" (Haub 1991:7), meaning that some countries make note of their divergence from the WHOs recommended practice in their vital registration documentation. Even

so, the vital registration data for a particular nation may or may not make note of these deviations, leaving researchers in the dark about actual practices in many cases.

However, some nations do not deny or conceal the fact that their practices of categorizing live births may stray from the WHOs recommended practice. For example, it is known that France has

> often classified as stillbirths infants that die before their births are registered… [t]he former East Germany required action of the heart *and* lungs, and the Soviet Union excludes infants less than 1,000 grams in weight, 35 centimeters in length, and less than 28 weeks gestation if they die within seven days of birth (Haub 1991:7).

These departures from standard measurements clearly serve to deflate the infant mortality rates for these countries. If we try to compare those countries that follow the recommended practice with those countries that do not, it is clear that will be not be given an equitable comparison. Furthermore, the only way we know of these differences in comparability is if the nation itself makes note of its deviations.

Even with these noted differences in the vital registration systems between the U.S and other developed nations, it can be said that comparing developed nations with one another is preferred to comparing developed nations with developing nations. Developing nations are more likely to lack standard regulations regarding the categorization of infant deaths and live births. In many cases births that occur in developing nations may not occur in hospitals and therefore may not be entered into the vital registration system. Furthermore, those parents who intend to register their births may not do so if the birth is a still born or dies shortly after birth. Another important issue is how likely is the registration of the death of an infant who is born alive and

entered into public record to take place when the infant dies at home and is not taken to a hospital or health center. Hypothetically, the parents of such infants have no reason to register these infant deaths; therefore these infants would still appear to be alive when analyzing the existing vital registration data.

As there is no real way to know if the above scenarios are occurring, the intent of raising these possibilities is to further specify the difficulty in cross-national comparisons of infant mortality. Most comparisons of infant mortality in the U.S. are made with other developed nations because of the well known lack of reliable data from developing nations. Since better and more complete data would most likely only serve to increase the infant mortality rate of most developing nations, the ranking of the U.S. as higher than developing nations of the world is most likely correct. Also, it is usually of interest to researchers why the U.S. ranks below other developed nations, not why the U.S. ranks above developing nations; the issue of poor quality data in developing nations is thus not really a point for discussion here.

Looking again at the CIA ranking of the United States in 2008 as 181st out of 222 countries in terms of having the lowest infant mortality rate, we see that the countries that rank "lower" (that is, they have lower infant mortality rates) are for the most part industrialized nations with reasonably accurate vital registration systems. Less stringent categorizations of infant deaths of other developed nations may contribute to the higher infant mortality rates of the U.S. compared to other industrialized nations. In his article "Infant Mortality: Who's Number One?" Haub (1991) asked whether and how much

these varying definitions of infant mortality were serving to deflate the rates in developed nations other than the U.S.

Haub (1991) first examined this issue of comparability by looking at the number of stillbirths to infant deaths and found that in most cases the "ratio of stillbirths to infant deaths lies between 0.45 and 0.60" (7). He found that Japan, which had the lowest infant mortality for the year 1985, also reported the highest stillbirth to infant death ratio with a value of 0.98. There is no way to know if these stillbirths are actually misclassified infant deaths although Haub noted that culturally there is "a perception that the death of a child or infant would be an undesirable entry in the *Koseki*" (the Japanese vital registration system) (8). However, when contacted, Japanese officials denied that any infant deaths would be intentionally misclassified as stillbirths in the vital registration system.

Since there is no way to ensure that infant deaths are being misclassified Haub developed a new measure to rank countries in terms of their infant mortality. He added stillbirths and infant deaths together, and then divided by all births (live births and stillbirths). This gives the "total infant mortality rate" and now includes the risk of a birth being a stillborn (8), therefore eliminating the problem of still births being misclassified as infant deaths. Performing these calculations using the data from 1989 (the most recent year that was determined to provide complete data for all countries) revealed some changes to the overall ranking of only the infant mortality rate by country. "Finland captures the number one position and Japan drops to third. The United States

[which was previously ranked 22nd], incidentally, picks up a few spots and rises to number 15" (8).

This new calculation shows that one reason why the U.S. has such a high infant mortality is due to differences in classification of infant deaths as stillbirths. The U.S. only advances a few spaces to have the 15th lowest infant mortality/still birth rate in this ranking of data from the late 1980's. This would seem to mean that there are still other reasons why the U.S ranks so poorly when compared to other industrialized nations in terms of infant mortality. Along with the classification of infant deaths, some of the reasons for the consistently poor ranking is thought to the heterogeneity of the U.S. population when compared to the relative homogeneity of other developed countries in Europe and Asia—the countries that consistently hold the top spaces in the ranking of worldwide infant mortality. If we only consider the non-Hispanic white population of the U.S. and make adjustments for the differences in the measurement between populations, the U.S. would be closer to, but likely still behind, many developed nations.

Thus far, this literature review has concentrated on the state of infant mortality in the U.S. and the world. The next section of this dissertation reviews the literature on the impact of conditions an infant or fetus is subjected to and the effects these conditions can have on an individual's health as an adult and ultimately on their longevity. This literature serves as a basis for the later analyses in this dissertation.

Infancy and Longevity

In the history of research between conditions of infancy and *in utero* and later life mortality and morbidity there have been competing hypothesis of the observed

connection. Doblhammer (2004) tells us that from the beginning of the 20th century a major hypothesis was that "infancy and childhood were…the critical phase" to study when trying to determine a person's lifespan (1). In the mid 20th century there was a shift from infant and childhood conditions to conditions experienced in adulthood. For decades research concentrated on adult lifestyle as the main determinant of adult lifespan. However, since the 1970's there has been a turn in the research on lifespan from adulthood, back to infancy and childhood. Studies have shown that conditions such as heart disease and diabetes are linked to the time that the individual is *in utero* (Doblhammer 2004). The main theory in this area is that "a mother's nutrition, and any infectious diseases that she might have had during the pregnancy are responsible for an increased susceptibility of the child to heart disease and diabetes once it reaches adult ages" (Doblhammer 2004:1). The main criticism of this finding is that poor conditions experienced *in-utero* and in childhood are more likely to occur to individuals who will experience deprivation throughout their whole lives. Ben-Shlomo and Smith (1991) point out that "contemporary infant mortality could also be taken as a proxy measure of adult deprivation" (532). This would seem to suggest that the same factors that would lead to deaths due to deprivation in infants would do so for adults in the same state of deprivation. Therefore early life conditions may not be the cause of the disease in later life. Indeed it could well be caused by the conditions throughout the individual's life.

When we think about the conditions in early life that a fetus or infant will experience, it is evident that many of the detrimental effects can be attributed to the experiences and behavior of the parents. Poor pregnancy outcomes occur

disproportionately to mothers who are low income, with low levels of education, who are adolescents, are minorities, live in rural areas or in the inner city, and are unmarried (Kiely and Kogan 1994). The negative effects of all these attributes are unlikely to disappear once the baby is born. This means that babies born to mothers who possess some or all of the characteristics are likely to be disadvantaged throughout their childhood and possibly their lives. So, when studies find that early life conditions have an effect on the longevity of an individual, how can we know that it is not the confounding effects of poor conditions throughout life? Doblhammer (2004) takes a large portion of her monograph "Late Life Legacy of Very Early Life" to examine this question.

Work by Huntington (1938) serves as a starting point for the examination of chronic disease and month of birth. In order to ascertain that infancy is not an indicator of the accumulation of poor living conditions, we must find a variable that can measure early life environment while "controlling" for the effects of poor conditions after gestation and infancy. Huntington proposes that the variable month of birth can provide this very measure. This hypothesis fueled the research that was mentioned earlier in the first chapter of this dissertation, dealing with schizophrenia. I have uncovered over 200 articles that have looked at the association between month of birth and being diagnosed with schizophrenia later in life. In the past twenty years researchers have found connections between month of birth and diseases "such as allergies, insulin dependent diabetes, congenital malformations, Parkinson's and Alzheimer's disease, and breast

cancer" (Doblhammer 2004:19). These studies use month of birth as a gauge of early life conditions. Their findings seem to indicate that the relationships found are legitimate.

Continuing to look at literature from the early 1900's I find that when following the trends of mortality through previous decades one major conclusion is that the health of an individual, when looking at one's entire life span, is highly dependent on their environment up to age fifteen (Kermack et al. 2001). By looking at the death rates in Great Britain and Sweden in the 1900's the above authors find "that improvement in infantile mortality is dependent in large measure on improvement in maternal health" (Kermack et al. 2001:683). The findings of this article set the ground work for subsequent articles that linked later life mortality to early life conditions. And, of particular import to my dissertation, it shows that there is a connection between the health and wellness of the mother and the reduction of infant mortality.

> To further clarify this phenomenon the authors state the following:
> If we remember that before birth and during its first year of life, the child is dependent on its welfare to a very large degree upon the general health and vitality of the mother, then it would be expected that a substantial improvement in the health of the latter would show itself in a reflected improvement in the infantile death-rate. The mothers of 1901 would on the average be born about 1870 or possibly a little later…, the health of females born at that date has so far improved that the death-rate was reduced by about 30 per cent in the case of Scottish and by about 40 per cent in the case of English mothers. It is suggested that this may constitute at least one of the factors conducive to the improvement of infantile mortality in the present century (Kermack et al. 2001:681).

So, by looking at data from Sweden and Great Britain, the authors were able to determine that the improvement in infant mortality coincided with an improvement in death for females of childbearing ages. Since the infant mortality rate was reduced and

the mortality of women was also reduced, the connection that one is related to the other may be appropriate.

One particular subset of the research on early life determinants of mortality in later life looks at the idea of deprivation in childhood and in infancy and the resulting chronic conditions in later life. Several articles have focused on the idea of deprivation in infancy and during gestation and whether a link can be drawn between poor nutrition and health care during this time and chronic diseases in later life. A large share of the literature looks specifically at heart disease in adulthood as a result of deprivation during this important developmental stage (Forsdahl 1977; Williams et al 1979; Barker and Osmond 1986; Osmond and Barker, 2000). Barker (2001) points out that many fetuses may have to "adapt to a limited supply of nutrients, and in doing so they permanently change their physiology and metabolism [and that] these programmed changes may be the origins of a number of diseases in later life…" (2). So the origins of heart disease in an individual may not be their unhealthy behavior in later life such as smoking, weight gain or bad eating habits; rather they may well be due to something about the way the heart is formed, how it regulates blood pressure, and how it handles sugars and fat that is programmed in the womb, and not in adulthood. Barker (2001) tells us that the fetus learns to develop in the womb regardless of the lack of nutrition, although it may do so at the "price of a shortened life" (2).

If these chronic conditions are indeed the result of some conditions that an infant is exposed to *in utero*, then there may well be a connection between the lifespan of an individual and the month of their birth. The influence of month of birth was examined in

terms of its impact on longevity in Doblhammer and Vaupel article “Life Span Depends on Month of Birth” (Doblhammer and Vaupel 2001). In Denmark, the authors found that among those with a life expectation of age fifty and higher, those born in the months of October, November and December lived longer than those born in the months of April, May and June. Interestingly, this pattern was reversed when looking at Australia, which since it is in the Southern Hemisphere, experiences seasons that are opposite those found in the Northern Hemisphere. Their findings supported the hypothesis that “debilitation early in life is the causal mechanism” (2938), and is not a construct of the seasonal distribution of infant deaths or social differences between individuals. They also pointed out that any variation in nutrition that could be the cause of these differences in longevity are not severe malnutrition but only “seasonally inadequate nutrition” which has improved in recent decades and may not be observed in current studies of the same topic (2939).

A continuation of the above work by Doblhammer and Vaupel is Doblhammer’s comprehensive monograph: *The Late Life Legacy of Very Early Life* (Doblhammer 2004); it concentrates on the impact of month of birth on longevity, and mainly examines mortality among individuals of age fifty and older. The author looked at patterns in the United States and Europe and used countries in the southern hemisphere for comparison. In an effort to disprove the hypothesis that month of birth is having an effect on longevity, four alternative hypotheses were proposed as possible explanations for the effect of month of birth on longevity. Although “[w]ith the epidemiological transition from infectious to chronic disease, cohort effects on mortality have largely lost

their importance" (167), in many cases differences were found in mortality based on month of birth in the over-fifty cohorts.

The four hypotheses that Doblhammer set out to examine in order to explain the association of month of birth and longevity, and disprove that month of birth is having a significant effect are the following: 1) the seasonal distribution of deaths, 2) the "Procreational Habits Hypothesis", 3) the 'Deadline Hypothesis", and 4) "Selective Survival or Debilitation". The fifth hypothesis is that "debilitation *in utero* or in the first year of life increases the infant's susceptibility to diseases at adult ages" (Doblhammer 2004:37). If the first four hypotheses are disproven, then the author can attribute the differences in adult longevity to their month of birth. In conclusion we are told that "the causal mechanism of the month of birth pattern in life span is related to debilitating factors that affect either the mother during pregnancy or the infant in the first year of life" (Doblhammer 2004:57).

Using data from a variety of populations, Doblhammer was able to disprove all four of the competing hypotheses. And based on the analyses we are given four important conclusions:

> First, the differences in life span by month of birth are tied to the seasons of the year. This is clearly shown by the reversal of the pattern in the Northern and Southern Hemisphere.
> Second, the pattern also exists in regions close to the equator…This finding rules out the explanation that differences in the length of daylight are the underlying causal mechanism…Third,… [In tropical regions] those who experience a mortality advantage later in life are born at the beginning of the season when temperatures are more moderate. In non-tropical regions…those born in autumn and winter have longer lives.
> [And] fourth, within a country the pattern may vary greatly among different groups [such as racial and ethnic groups] (58).

The findings of the four hypotheses lead to the conclusion that “debilitation early in life is the causal mechanism” (59). This is perhaps the most important finding in the current literature in terms of the hypotheses I test in my dissertation because it establishes that the differences observed in longevity are in fact due to month of birth and not to some confounding variable. As such, if these differences by month of birth are still found today, we are likely to see the differences in the infant mortality of contemporary cohorts.

The conclusions of Doblhammer (2004) indicate that there is a quantifiable impact on longevity of the individual’s month of birth. The final chapter of *The Late Life Legacy of Very Early Life* asks if the impact of month of birth on longevity can still be seen today in younger cohorts. Doblhammer notes that many of the issues affecting maternal health such as poor nutrition during the winter and food-borne illnesses due to poor refrigeration during the summer are not thought to be of major concern to modern pregnant women and or modern parents of infants. Doblhammer shows that these differences can still be seen today. To look into the impact of month of birth for contemporary cohorts, data were used from the Minnesota twin registry. Then the “seasonal pattern in the birth weight of twins born in the 1970’s and 1990’s is compared with the month-of-birth pattern in the mean age at death of decedents aged 50+ who were born in Minnesota” (182). The results of this study showed that a positive correlation exists “between the life span pattern and the birth weight pattern” of the males, but not the females in the twin registry. The author indicates that this may be an

indication that contemporary cohorts may display differences in longevity in terms of their month of birth.

The important work in the areas of infancy and longevity lead to an enhanced understanding in the areas of adult mortality. I hope to show in my dissertation that an investigation of the month of birth phenomenon will yield similar significant results in the area of infant mortality. The next section of this literature review will examine the most common causes of infant death and findings in past literature on the distribution of these causes by month.

Causes of Infant Mortality

An important feature of the analyses in my dissertation is the relationship between certain causes of infant death and season or month of birth. My analyses will examine the relationship between month of birth and infant mortality using causes of infant death that were categorized based on the information provided on the infant's death certificate. The aim of the analyses will be to predict which month of birth is associated with an increased likelihood of a specific cause of infant death. Information on the causes of death for infants is extensive; however nothing is known about the association between being born in a certain month and dying of a certain cause. This is because death is usually not caused by one specific and identifiable cause, but instead a combination of many causes whose origins may or may not be know. As such, associations must be drawn from available literature on causes of infant death and the seasons in which the death occurred. The causes of death used in this dissertation will draw from the literature

on infant deaths. By undertaking analyses using causes of death I hope to draw some conclusions about month of birth and cause of death.

To begin, I will review the leading causes of death for infants in the year 2004, as given by the U.S. Department of Health and Human Services and the Centers for Disease Control (CDC). They list the

> [l]eading causes of infant death …in rank order [as]: Congenital malformations, deformations, and chromosomal abnormalities; Disorders related to short gestation and low birth weight, not elsewhere classified; Sudden infant death syndrome (SIDS); Newborn affected by maternal complications of pregnancy; Accidents (unintentional injuries); Newborn affected by complications of placenta, cord and membranes; Respiratory distress of the newborn; Bacterial sepsis of newborn; Neonatal hemorrhage; and Diseases of the circulatory system (Heron 2007:1).

As we can see from these most frequently reported causes of infant death, the exact origins leading to death may not always be easy to identify. This is why it is important to consider all causes of infant death as potentially being contributed to by the month of birth and the conditions while *in utero* and in the first year of life.

An article by Eastman (1945) provides much of the groundwork for my investigation of cause specific infant mortality. Although Eastman's description of infant deaths by month of birth is from the 1940's and therefore may seem outdated, the aim of this dissertation is to examine the relationships found in this and later time periods with a focus on contemporary infant mortality. Eastman is interested in the infant's cause of death and asks what conditions and diseases could be involved in the relationship between the health of infants and their chances of survival in the first year of life. Although there have been several changes to the classification of diseases in past decades, the causes of death that Eastman chooses to analyze are relatively consistent

with the underlying causes that are given in the linked data set used in this dissertation. He chooses the following broadly classified groups:

> 1. the prenatal causes, which comprise premature births, congenital malformations and congenital debility; 2. the natal and neonatal group, which includes injury at birth and 'other diseases peculiar to early infancy'; 3. the respiratory diseases, chiefly influenza and pneumonia; 4. the gastrointestinal group, which covers diarrhea and enteritis, dysentery and 'other disease of the stomach'; and 5. all other diseases and conditions; among which the more important are whooping cough, syphilis, accidents, and those of ill-defined origin (Eastman 1945:914).

Eastman goes on to provide some seasonal information about how each of the five causes of death is distributed. The first category, prenatal causes, comprises a large portion of the infant deaths in the 1945 sample; "[o]f infant deaths from prenatal causes about 70 percent are chargeable to premature birth, 21 percent to congenital malformations, and 9 percent to congenital debility" (Eastman 1945:915). In these analyses those babies born in winter and spring show the highest death rate from these causes and the author attributes the higher rates to "the harmful effects on both mother and unborn child of the diseases most prevalent in winter" (Eastman 1945:915). He also attributes the increases in mortality due to prenatal causes to the "faulty nutrition in expectant mothers" which has been shown to increase premature births and malformations, which he associates with lack of food in the winter months (Eastman 1945:916). This lack of food in the winter months may account for the rise in prenatal deaths in April and May.

Eastman next describes those causes to the neonatal and natal groups. He points out that the greatest numbers of deaths in these periods are due to injury which would not be thought to be due to season of birth. However, since premature infants are more

likely to die from injuries sustained due to resulting frailty and underdevelopment, the variation in neonatal and natal deaths may be seen by season. Prematurity is most frequent in the winter months, and in the case of the 1945 infants the greatest number of deaths in this category also occurred in the winter months. For the purposes of my dissertation prematurity will be considered as a cause of death, since this category in the dataset is available and detailed. By looking at prematurity as a cause of death, a similar relationship could well be uncovered.

Those deaths caused by respiratory diseases like pneumonia and influenza are associated with cold weather and, as one would expect, infants are less able to recover from these diseases because of their relative frailness. Eastman notes that those babies born in January "suffer excess mortality from these diseases" (Eastman 1945:916). Even these days, mortality due to influenza has a seasonal distribution. According to the CDC, for the 2007-2008 flu season "[a]ctivity increased slowly from mid-December through the end of the year with more rapid increases during January and through the week ending February 16. Flu activity peaked in mid-February and then decreased through the end of the flu season on May 17" (Centers for Disease Control 2008). This distribution of the cases of the flu is similar to the seasonal distribution that Eastman found in infant mortality, although the cases of infant mortality due to respiratory diseases started to peak in October and stayed high through February.

Deaths due to gastrointestinal disease were also common in the early to mid twentieth century. Because of the high number of cases of infant mortality due to gastrointestinal diseases, Eastman includes this cause as one to analyze by month of

birth. This cause of death strikes older infants and is usually more common during the summer months, unlike the characteristics of most of the other common causes of infant death. Gastrointestinal causes of death peak in January and taper off into the summer. Even though the gastrointestinal illnesses are more common in the summer in general, likely due to poor refrigeration in the summer months, the risk of death is more likely for infants who are a few months old in the summer. This is because those infants born in the summer months will likely still be breast-feeding and not eating or drinking those contaminated items that an older infant would.

During the study period of Eastman's article, gastrointestinal illness accounted for a substantial proportion of the infant deaths being studied. There is evidence, however, that these types of illness are not as common currently and as such do not pose as big a threat to infant survival as they did in earlier decades. In a historical investigation of infant mortality in Philadelphia in the 17th to the 20th centuries, Cheney (1984) find that by the early 1920's the pattern of increased mortality due to diarrheal diseases mostly disappeared. However, the Eastman article was published in 1945 and still saw a pattern of gastrointestinal diseases. In these analyses, infant deaths due to gastrointestinal disease will likely be small but will be investigated especially to see if a connection exists for contemporary infants born in the late winter and early spring months.

The last category that Eastman sets forth is the encompassing "all other causes" which includes such a wide variety of causes that it would be more difficult to pinpoint a seasonal distribution for infant deaths. Eastman includes whooping cough,

communicable diseases of childhood, accidents and diseases such as syphilis and tuberculosis. Causes of death such as tuberculosis and syphilis are rare in modern day United States and causes of death due to accidents contradict the theory of disease and nutrition during gestation that the dissertation aims to isolate. Eastman finds that mortality is highest in December, January and February for these causes of death. However, due to the 'catch all' nature of this category, my dissertation will not focus specifically on this category, but will use it as the reference group category in the multinomial logistic regression equations I will estimate. This will be described and discussed in a later chapter of this dissertation.

Although the Eastman article on seasonal distribution of diseases dates back several decades, the causes and distribution of infant deaths will likely be similar in modern data and will be treated as such in this dissertation. Several other sources were found that also separated infant death for analyses (although not by month of birth) and these divisions were similar to Eastman's (see Eberstein et al. 1990). One addition that will be included in the analyses of this dissertation is that of sudden infant death syndrome (SIDS) as a cause of death. Campbell (1994) found that there is an increase in SIDS about two to five days after the temperature goes down in the winter and fall months. This would suggest that there may also be a seasonal distribution to deaths attributed to SIDS (see also Williams et al. 1987; and Murphy and Campbell 1987). The inclusion of SIDS as a cause of death in this dissertation will hopefully shed further light on the association between deaths from SIDS and season.

A more thorough discussion of the causes of death used in the analyses of this dissertation will be found in chapter IV. Next, I turn to a review of the link between season and mortality.

Seasonality and Mortality

The literature that links human exposure to the climatic changes that occur by season is extensive, dates back many years, and concentrates on many different populations. When we think about the causes of death that may be related to seasonal differences throughout the year and in different hemispheres and at different elevations, the link between extreme climatic conditions and increased mortality may be clear. But, beyond any "cut-and-dry" relationship between extreme heat or cold and mortality lies the question of what exactly it is about a particular season that serves to increase mortality in a population. Numerous studies have examined this relationship with the intention of identifying the cause and effect relationship between climate and mortality in order to reduce its impact when possible. This section of my dissertation will examine this literature in order to make connections where possible to the analyses of infant death and season.

The seasonal distribution of deaths is attributed to a variety of causes. The wealth of literature in the area of seasonality and mortality is aimed at determining what causes deaths during certain seasons and what it is about a particular season that increases deaths. This question is not easily answered. Past literature attributes seasonal mortality rates to "a number of physiological parameters, e.g. haemostatic factors, blood pressure, malnutrition, and seasonal variations of the immune system. It is suggested that these

changes, in turn, are a consequence of seasonal fluctuations of temperature, rainfall, or photoperiod. In addition, anthropogenic factors like air pollution also have some seasonal impact" (Lerchl 1998:84). Lerchl (1998) tells us that it should "appear clear why low temperatures are associated with higher mortality rates" (87), because when temperatures are low, it is a biological fact that blood pressure and blood composition change, causing increases in stroke and heart infarcts. Although this is true, it is important to note that this only demonstrates a relationship between winter temperatures and deaths due to stroke and heart infarcts, and that other causes of death may not display such a clear relationship. This is why it is important to consider more than just a few causes of death as being contributed to by the changes in temperature, in order to not overlook the possible effects of heat on other causes.

Several articles that concentrate on seasonal mortality in Europe find that the most excess deaths in the winter occur in the countries with more mild winters than in countries with harsher winters, as measured by very low winter temperatures (McKee 1989; Ballester et al. 2003; Donaldson and Keatinge 1997; Clinch and Healy 2000). This suggests that the ability to survive the winter depends on the individual regulating the indoor temperature of their homes, or of themselves when outside and does not depend on the harsh overall climates in these areas. It makes sense that countries with consistently low average temperatures in the winter would be better equipped to deal with the harshness of the season. This is because the country itself would likely place certain regulations regarding insulation and heating in homes, and the individuals themselves would have learned to cope with low temperatures by wearing layers of

clothes and heating their homes and offices (McKee 1989; Healy 2002). The idea that much of the higher mortality in countries with warmer winters is due to indoor and not outdoor temperatures is supported "by the absence of seasonal variation in Iceland where the average temperature in winter is below freezing point but there is widespread availability of low cost geothermal energy" (McKee 1989:180). This literature points to the investigation of populations based not only on the low average temperatures in the winter, but also on low variability in temperatures throughout the year. Several studies have found support for this reasoning (Ballester et al. 2003).

Studies have also found that the elderly are the most susceptible to these atypical changes in the weather and to seasonal variations in general (Thompson et al. 2003; Keatinge 1986). In a historical study of London it was found that in the winter months when food was less available and therefore more expensive, there was a significant increase in mortality (Galloway 1985). Not surprisingly, in this historical study the author found that "the most striking result is the tremendous and lasting impact of cold winters on the number of deaths in the older age groups" (Galloway 1985: 496). This study also found that decreases in winter temperature also increased the frequency of deaths attributed to 'infancy'; however, we must keep in mind here that infancy as a cause of death in this historical study may refer to children older than one year of age. These and similar results indicate that infants, children and the elderly have been the groups most susceptible for centuries to climate changes (see also Guy and Cantab 1843; Hare et al. 1981).

What seems to be unclear in the study of cold weather and mortality is the exact cause (or causes) of death that contribute to the increase of death during these periods. In most cases only a small fraction of deaths are due to hypothermia—a condition in which the individual dies from a lowering of their body temperature. In a study of England and Wales, Keatinge (1986) found that hypothermia caused only about 300 of the approximately 40,000 excess winter deaths that occurred and were reported in vital statistics data over the study period (732). He found that the majority of the excess winter deaths (to adults) were attributed to "coronary and cerebral thrombosis and respiratory disease" (732). There is also an extensive literature concentrating on the connection between cold weather mortality and cardiovascular diseases (Mercer 2003; Donaldson and Keatinge 1997). Studies considering these causes of death have mostly concentrated on adult mortality because this is the population that is most at risk of dying from cardiovascular causes. However, infant deaths can also be attributed to heart and respiratory causes. I will include in my analyses of infant mortality deaths attributed to heart and respiratory causes.

Some studies have found that "respiratory deaths resulting from influenza epidemics in winter may account for as much as two-thirds of excess winter deaths" (Mercer 2003:9). Studies have concentrated on the increases in deaths from influenza and pneumonia during the winter months because of the peak that is observed in these deaths during the winter months. When looking at the world population it has been found that almost seventy percent of the seasonal variation in mortality is due to influenza and therefore may be used as a surrogate for seasonal mortality (Reichert et al.

2001, Reichert et al 2004). Even in the U.S. and other developed countries the influenza virus has been shown to contribute substantially to the "excess mortality" (the excess number of winter deaths as compared to the number of summer deaths) in the winter months (Reichert et al. 2001). This means that policy directed at influenza vaccines could have huge effects on decreasing the effect of season on mortality.

Increased mortality from extreme heat is also well researched but usually does not lead to a clear cause and effect relationship between heat and death. It has been found that in the summer warm, humid and calm (in terms of wind) conditions are related to the highest mortality (Kalkstein and Davis 1989); and much like the research in the area of cold weather and mortality, the effect of these conditions is strongest in places where hot weather is uncommon. Kalkstein and Davis (1989) call this the "acclimatization effect" (50), meaning that individuals cope with extreme temperatures better when the extreme temperatures are not rare. For example, people are used to the heat in the South and the cold in the North, but when the reverse conditions occur, the highest levels of mortality are seen. Heat related deaths are especially evident during heat waves when temperatures reach highs that are either unprecedented or have not been experienced in an area for several years (Semenza et al. 1996). During heat waves individuals who die from heat related causes usually die from heat stroke, cardiovascular conditions or a combination of these causes with some underlying cause (Semenza et al. 1996). Kalkstein and Davis (1989) found, however, that there is an overall increase in deaths of all causes during extreme summer heat. "Thus it appears that mortality from a wide variety of causes increases during extreme summer weather, and the notion of few

specific weather-related causes in summer appears to be specious" (Kalkstein and Davis 1989:52). This fact about mortality at all ages gives merit to including all causes of infant death that are available in the data.

Conclusion

It seems evident from the literature on infant mortality, seasonality of mortality and on the link between season and adult longevity, that analyses such as those to be performed in this dissertation have merit and will contribute to the literature in all three areas. The extensive literature reported and reviewed on these topics has served as a background for my dissertation research. While performing this literature review, I found that studies looking specifically at the topic of infant mortality and season date back many decades, with few if any contemporary counterparts. Therefore my research will serve as a reexamination of the phenomenon and its relevance and applicability in more modern times. The analyses of my dissertation will also contribute to current literature in two important ways: first, cause of death will be examined in a multinomial dependent variable context; and second, two different types of independent variable will be used to measure season, one measuring month of birth and the other measuring month of gestation. The reason for including these two separate variables is to determine which, if any, has the more important effect on infant mortality. My hope is that the findings of my dissertation will elaborate on past findings as well as contribute to the literature with the inclusion of these innovative measures.

The importance of infant mortality as an indicator of social development guarantees that studies on this subject will continue to be prevalent for the foreseeable

future. The past literature indicates that there are connections between season and mortality, based on numerous studies that concentrate on a variety of geographies, time periods and causes of death. It is also shown that the conditions in early life contribute to the longevity among the population aged 50 and older. The aim of this dissertation is to ascertain whether the past findings in the areas of season, mortality and longevity will translate to significant findings in a study of contemporary cohorts of infants in the United States.

This current chapter is a review of literature in several areas relevant to the research question of this dissertation. Before turning to the analyses intended to examine this research question I will discuss the data and methods that will be used in the analyses. The Linked Birth Infant Death data set from the years 2000-2004 will be discussed as well as the methods of multinomial logistic regression, multilevel multinomial logistic regression, and Geographic Information Systems (GIS) mapping software.

CHAPTER III

DATA AND METHODS

In this chapter, I first describe the Linked Birth/Infant Death records for the years 2000 to 2004; they are the main data I will be analyzing in this dissertation. I address the processes of collecting and linking the data as well as the weighting procedures. A brief discussion of the history of vital registration data will also be included. In addition, I will present the operationalization of my dependent variable, the two independent variables of interest—the month of birth measure and the month of gestation measure, and the several independent variables to be used as controls. I will also discuss the reasons for including the variables to be used in my analyses as well as some potential issues with measurement that may arise in their use. Lastly, I will discuss the methods to be used in this dissertation, namely multinomial logistic regression, multilevel logistic regression, and mapping procedures used in ArcGIS (Geographic Information Systems) that I will use in a later chapter to examine infant mortality geographically in the United States.

Data

The data used in this dissertation are the Linked Birth/Infant Death records for the years 2000 to 2004. These are compiled and made available for public use by the National Center for Health Statistics (NCHS), a division of the Centers for Disease Control (CDC). The data are available from 1983 to 1991 and from 1995 to 2004. (Data from 1992 to 1994 are not available from the NCHS because at the time of this dissertation they had not been linked). The available data date back 25 years and include nineteen individual years. Documentation from the NCHS for the year of 2004 states that 98.9

percent of all infant death certificates in the United States are successfully linked to their corresponding death certificate (see also Prager et al. 1987). In order to overcome any biases that may be found from the fact that one-hundred percent of all records are not able to be linked, a weighting structure is implemented. Usually with vital registration data there is no need for weighting since the data are collected from the whole population, therefore providing a full count of the specified event. However, the nature of the Linked Birth/Infant Death dataset are slightly different and by weighting the data we may obtain a full count of all deaths to infants for in a given year linked to their corresponding death certificates. This weighting process will be discussed in a later section of this chapter. I now turn to a brief discussion of the vital registration data in the United States.

Vital Registration Systems

The analysis of vital registration data can be traced back to John Graunt and his influential book *Natural and Political Observations Mentioned in a Following Index, and Made upon the Bills of Mortality*, first published in London, England in1662. In his book Graunt analyzed information on births and deaths in London during the peak of the plague. In addition to several important demographic facts, Graunt showed how important the collection of data on human populations is to society (Hetzel 1997:45). Graunt may have shown these important facts in the mid seventeenth century, but it took many years to perfect the collection of vital registration data throughout the world. In some countries there is even today a lack of a vital registration system. And even in

other countries where vital registration systems do exist, they may not be comprehensive and therefore lack important components.

In the United States vital registration systems have undergone several stages of improvement, and will likely undergo several more in the coming century. The important issue to note here is that vital registration data have a long history of use by researchers. With improvements in collection and comparability it is likely that vital registration systems will continue to be an important source of data for years to come.

In the United States each of the fifty states, two cities (Washington D.C. and New York City) and five territories (Puerto Rico, the Virgin Islands, Guam, American Samoa, and the Commonwealth of the Northern Mariana Islands) is responsible for collecting vital registration data on deaths, births, marriages, divorces, and fetal deaths that occur within their geography as well as providing individuals with copies of certificates for these events. The NCHS is then responsible for the collection of this information from the states and compiling it into public use datasets for each vital event. The purpose of compiling the state data is not primarily to provide the public with datasets, but to calculate and produce fertility rates and mortality rates for the United States.

The NCHS produces the natality and mortality files annually from birth and death certificates provided to the NCHS from the Vital Statistics Cooperative Program (VSCP). The data are uniformly coded, pass quality control standards, and are edited and reviewed before the U.S. official birth and death statistics are produced. States routinely link birth and infant death certificates for their own use. If the infant dies in a state that is

not their state of birth, the certificates are given to the appropriate state so that a consistent and accurate count can be made. These infant death and birth files take advantage of the matching birth certificate numbers for infant deaths that occur within NCHS jurisdiction. They in turn "[use] this information to extract final, edited mortality and natality data from the NCHS natality and mortality statistical files. Individual birth and death records [are] selected from their respective files and linked into a single statistical record, thereby establishing a national linked record file" (Arias and Smith 2003:2-3).

The Linked Birth/Infant Death files take the compilation methods that the NCHS performs one step further by not only providing statistics on the live births and infant deaths in a year for the entire U.S. and territories, but by also linking those deaths occurring within the first year of life to the corresponding birth certificate. This linkage enables researchers to use information about the infant, the parents and/or geographic information that may only appear on the birth certificate with both the death and birth information. For example, information about the mother's education, race and Hispanic origin does not appear on the death certificate of an infant but does appear on the birth certificate. By linking these two files the researcher is able to look at these maternal characteristics with the information regarding the death of the infant.

Beginning in 1995 the NCHS produces two types of Linked Birth/Infant Death datasets; a cohort file and a period file. Before 1995 the files were produced only in birth cohort format. The period linked file for 2004 contains information on all the deaths that occurred in 2004, whether or not the birth occurred in 2003 or 2004. The birth cohort

file, however, contains all deaths to infants who were born in a specific year, regardless of the year of birth.

The birth cohort file has some advantages for researchers because it allows the investigation of infant mortality for a particular birth cohort. However, the birth cohort data take longer to compile since they cannot be completed until the close of the following data year in order to include all deaths to that specific cohort (Arias and Smith 2003:1). In this dissertation, the period linked files are used since the issue of birth cohort is not of special interest to the research questions I will investigate.

The period data sets contain three files for each year; a numerator file, a denominator file, and an unlinked file. The numerator file is the file discussed above and it "contains all the U.S. infant deaths which occurred in the [particular]… data year linked to their corresponding birth certificates, whether the birth occurred in [that year or the year before]…" (Arias and Smith 2003:1). The denominator file is the NCHS natality file for a specific year, containing all births from that year. This file is included so that in conjunction with the numerator file, infant mortality rates can be computed. The last file is the unlinked file; this file "contains information from the death certificate for all U.S. infant death records which could not be linked to their corresponding birth certificates" (Arias and Smith2003:1).

The approximately one percent of all infant death records that could not be linked to their corresponding birth certificates are contained in the unlinked file. Some differences do exist because of lack of data from the birth certificate, but the layout of the unlinked file mirrors the numerator files of the linked data. If a rate is to be

computed, data from this unlinked file should be added to data in the numerator file and then divided by data in the denominator file. These three files are available for the U.S., Puerto Rico, the Virgin Islands and Guam. In this dissertation it is not necessary to use the denominator or unlinked files since I am not interested in deriving infant mortality rates from these data. I include the brief discussion of these additional files to help explain the use of the Linked Birth/Infant Death files beyond my own dissertation. More information on these files and their use can be found on the NCHS website: www.cdc.gov/nchs.

Weight Procedures

Starting in 1995, several changes were made to the linked file. The most important is the advent of the period file. Another important change was the addition of the weights in order to compensate for the one to three percent of records that are unable to be linked each year. This weight only applies to the linked numerator file for the United States (not the territories) and is able to correct in part for biases in percent linked by major characteristics. The formula for computing the weight is:

$$\frac{\text{Number of Linked Infant Deaths + Number of Unlinked Infant Deaths}}{\text{Number of Linked Infant Deaths}}$$

"A separate weight is computed for each state of residence of birth and each age at death category (less than one day, 1-27 days, 28 days to 1 year). Thus, all weights are 1.0 for states which link all of their infant deaths" (Arias and Smith 2003).

The NCHS provides a table displaying the percent of linked variables for each state; this is important to note because some states are unable to link the majority of their infant death files. For example Texas is able to link about 97 percent of cases; Louisiana

linked the least amount of cases in 2004, 95.6 percent. For this reason the weights are added to the file to correct for the biases that could arise because of the variation in the percent of linked files between states. This is particularly relevant if the data are being used to calculate rates, but is still an important detail to note when using the data for any purpose.

Also important to note is that "a slightly higher percentage of postneonatal than neonatal deaths were linked" (Arias and Smith 2003:5). Although the weighting can correct biases due to the inability to link all files, there is no way to correct all data limitations. Accordingly, the NCHS reminds us that "variations in the percent of records linked should be taken into consideration when comparing infant mortality rates by detailed characteristics" (Arias and Smith 2003:5).

Other Data Considerations

Another important consideration is imputation. Beginning in 1995 is the imputation of the birthweight of the infant in cases where this information is missing. This is done to "reduce the potential bias in the computation of birthweight-specific infant mortality rates. Basically, if birthweight is not stated and the period of gestation is known, birthweight is assigned the value from the previous record with the same period of gestation, race, sex, and plurality" (Arias and Smith 2003). By adding this imputation technique the percent of cases without the information on birthweight drops from 3.15 percent to 1.19 percent in the numerator file, and from 0.10 percent to 0.04 percent in the denominator file. It is also important to note that in the later years of this data set, this reduction in the percent of files that are missing information on birthweight is reduced

even further. For example, “in the 2003 data file, the addition of this imputation has reduced the percent of not-stated responses for birthweight from 3.85 to 0.61 in the numerator file, and from 0.09 to 0.02 in the denominator file…” (Arias and Smith 2003:1-2). The documentation for the 1995 period file tells us that this reduces, but does not eliminate “the potential for underestimation when computing birthweight-specific infant mortality rates” (Arias and Smith 2003). The imputed values are flagged and therefore can be identified by the users as imputed birthweight values when using the data. However, in this dissertation I make no distinction between those values which were imputed and those values which were not. This is because in my analyses birthweight is used as a control variable and is not paid any specific attention. It is important to note these additions to the data, however, since they represent improvements to the files that should be kept in mind.

Operationalization of Variables

The Linked Birth/Infant Death dataset encompass a wealth of information about the infant or parents. Some variables will provide important information for the analyses of this dissertation, while other variables will not be pertinent. As such it is necessary to determine which variables will be used, and of those variables, if any recoding or restructuring is necessary in order to use them in the analyses. In this section of chapter III I will discuss the operationalization of the dependent and independent variables.

Dependent Variable

The dependent variable for this dissertation, which I call ‘cause of infant death’, is constructed by recoding the information on the death certificate about the infant’s cause

of death; the result is a nominal variable with eleven categories that comprise the possible causes of infant death as specified on the death certificate.

The classification of death found on death certificates is based on the International Classification of Diseases, 1992 (10^{th}) Revision. This is a classification endorsed by the World Health Organization (WHO) and is used by all WHO member states. The International Classification of Diseases (ICD) is the "international standard diagnostic classification for all general epidemiological, many health management purposes and clinical use" (http://www.who.int/classifications/icd/en/). The ICD is used to monitor the general health of a population as well as the incidence and prevalence of a specific disease in a population. The Linked Birth/Infant Death dataset uses this classification as well as provides a variable that contains a recode of one-hundred and thirty causes of death. By recoding this information on cause of death, I created the variable used in this dissertation as the categorical dependent variable.

When I initially began exploring the idea of infant mortality by month of birth, I came across the 1945 article by Eastman titled "Infant Mortality in Relation to Month of Birth," which I previously discussed in the literature review of this dissertation. In this article Eastman compiled six categories of infant death—(1) prenatal causes, (2) natal and neonatal causes, (3) respiratory diseases, (4) gastrointestinal causes, and (5) all other causes. These causes seem few when considering all the possible causes of death that are available on the Linked Birth/Infant Death dataset. It is likely that when Eastman wrote his article the categorization of causes of death was not as extensive as it is today. Also, Eastman stated us that he used the principal causes of death, which have changed

significantly since 1945. For example Eastman's category "gastrointestinal causes," (encompassing causes such as diarrhea) contains a bulk of the infant deaths between 1935 and 1937. In the U.S. for the years 2000-2004 in the Linked Birth/Infant Death dataset there are very few deaths that would fall into this category. This may be because of different classifications of death, such as deaths that today may be due to diarrhea are attributed to dehydration, or it may be attributed to an actual decrease in deaths due to these causes in modern times. Either way the categorization of disease using current data is justifiably different from any found in studies from early last century. Because of this, the categorization of causes of infant death that I use is different from those used in the Eastman article.

While I could not use the categories from the Eastman article to examine causes of infant death, I could also not leave the categories the same as those found in the dataset, which was comprised of 130 recoded causes of death that were then categorized into twenty broad categories of causes. Although the multinomial logistic regression will support a model with up to 50 categories in the dependent variable, the interpretation of variables with this many categories would be extremely difficult (Long and Freese 2001). Accordingly, I looked to past literature to find a categorization method of causes of infant death. I was interested in finding a categorization that would be comprised of fewer than twenty categories while maintaining the substantive differences between the causes of death. Meaning that, I did not want to group together any causes that were unrelated in fear that it would result in the wrong conclusions being drawn on the basis of the results.

My investigation and review of the literature led me to use a categorization scheme close to that found in a 1990 article by Dollfus and colleagues titled "Infant Mortality: A Practical Approach to the Analysis of the Leading Causes of Death and Risk Factors." In this article the authors used the following categories of infant death: prematurity and related conditions—extreme immaturity and other preterm infants, intracranial hemorrhage and respiratory problems; congenital anomaly –nervous system, cardiac and circulatory system, pulmonary system, diaphragm, renal, liver and gallbladder, Patau's syndrome, Edward's syndrome, multiple anomaly; Sudden Infant Death Syndrome (SIDS); obstetric conditions—incompetent cervix, premature rupture of the membranes, multiple pregnancy, placental abnormalities; birth asphyxia—unspecified fetal distress, severe and unspecified birth asphyxia; perinatal infections—streptococcal meningitis, maternal infections, chorioamnionitis, other perinatal infections; other infections—infectious and parasitic diseases, meningitis, respiratory infections; and external causes/injuries—motor vehicle accidents, accidental falls, drowning, accidents due to fire/flames, accidental obstruction/suffocation, homicide, neglect, injuries if accidental or purposeful and unspecified accidents (183).

As I discuss below my classification includes eleven categories that are similar, but differ slightly from the above categories. They differ from Dollfus and colleagues classification because I did not drop any of the deaths that did not fit into the above classification; instead I created three extra categories of deaths to incorporate all those deaths occurring to infants between the years 2000 and 2004. Much of the other literature that categorized causes of infant death used fewer than eight categories, so for

this reason I chose to follow the scheme of Dollfus and colleagues. I also chose this classification because it was written in 1990, after the 1989 revision of the ICD coding and therefore lent itself well to replication with the 2000 to 2004 data. Although they included fewer categories, most of the previous literature included classifications that were somewhat similar to the Dollfus and colleagues' categories. (For more information on other classifications of cause of infant death please see: Eberstein, Nam and Hummer 1990; Eberstein and Parker 1984; Buehler et al. 1987; and Ahmed et al 1991).

Another important reason that my analyses use the eleven category 'cause of infant death' variable is due to the nature of multinomial logistic regression models. This model will be discussed in depth in a coming section of this chapter; however in order to explain the rationale for choosing the eleven category cause of death variable I must briefly mention the use of the dependent variable and the method of analysis. As I already mentioned, Long and Freese (2001) tell us that in STATA we may use up to fifty categories in the dependent variable. So my dependent variable is not problematic as far as the STATA algorithms are concerned. The only requirement is that the dependent variable be categorical, nominal and unordered—meaning that the categories do not need to be in any particular order or consist of any specific number of cases. One category of the dependent variable should be 'set to zero' so as to take on the role of the baseline category, or the reference category, to which all other categories of the dependent variable will be compared. "If the base outcome is not specified, the most frequent outcome in the estimation sample is chosen as the base" (Long and Freese

2001:229). In order to ensure that the category of comparison makes the most sense in my analyses, I will set the base category in the estimated models.

In this dissertation the category "external causes of mortality" will be used as the baseline category. This category includes causes of death such as accidents, falls, drowning, and suffocation. This category is used as the baseline to which the other causes of death are compared because it would seem that causes of death due to such circumstances should not be due to any differences in the month in which an infant is born, but purely on random events such as accidents, or in some cases on purposeful events such as neglect or homicide. If this cause of death that is to be used as the baseline did not exclude all other causes of death, making any comparisons between it and the other categories could be problematic. This is why the dependent variable needs to have as many categories as it does and why it is necessary to substantively distinguish between the baseline and the other causes. For the years 2000 to 2004 there were 5,319 infant deaths attributed to external causes in the U.S.

The next cause of death category is prematurity and related conditions. As discussed above, this category is similar to one found in the categorization by Dollfus and colleagues (1990). The causes of death included in this category are hemorrhagic conditions, disorders related to length of gestation and fetal malnutrition—slow fetal growth and fetal malnutrition, disorders related to short gestation and low birthweight not elsewhere classified, extremely low birthweight or extreme immaturity, other low birthweight or preterm; other respiratory conditions originating in the perinatal period—congenital pneumonia, neonatal aspiration syndromes, interstitial emphysema and

related conditions, pulmonary hemorrhage originating in the perinatal period, chronic respiratory disease originating in the perinatal period, atelectasis, and all other respiratory conditions originating in the perinatal period; hemorrhagic and hematological disorders of the newborn—neonatal hemorrhage, hemorrhagic disease of the newborn, hemolytic disease of the newborn due to immunization and other perinatal jaundice, hematological disorders, syndrome of infant of a diabetic mother and neonatal diabetes mellitus, necrotizing enterocolitis of the newborn, hydrops fetalis not due to hemolytic disease, and other perinatal conditions. Once created this category encompasses 26,020 of the infant deaths in the years 2000 to 2004 and is the cause of death category with the most infant deaths.

The third category is congenital anomalies and it also similar to that of Dollfus and colleagues (1990). The causes of death included in this category are congenital malformations, deformations and chromosomal abnormalities—anencephaly and similar malformations, congenital hydrocephalus, spina bifida, other congenital malformations of the nervous system, congenital malformations of the heart, other congenital malformations of the circulatory system, congenital malformations of the respiratory system, congenital malformations of the digestive system, congenital malformations of the genitourinary system, congenital malformations and deformations of the musculoskeletal system, limbs, and integument, Down's syndrome, Edward's syndrome, Patau's syndrome, other congenital malformations and deformations, and other congenital abnormalities not elsewhere classified. This category includes 18,712 of the

infant deaths in the U.S. for the years 2000 to 2004 and is the second largest category of infant deaths to be analyzed in this dissertation.

The fourth category of infant deaths is SIDS and other causes. Sudden Infant Death Syndrome (SIDS) is defined by the CDC as "the sudden death of an infant less than one year of age that cannot be explained after a thorough investigation is conducted, including complete autopsy, examination of death scene, and a review of the clinical history" (Matthews and MacDorman 2008). After a thorough examination as described above has occurred, many times a cause of death other than SIDS can be identified. Those deaths that are attributed to some cause other than SIDS are sometimes referred to as sudden unexpected infant death syndrome or (SUIDS) and are usually causes such as accidental suffocation, hyper/hypothermia, or metabolic conditions. These causes would not fall into the SIDS category; instead they would have been reclassified as into one of the above causes that most likely fall into the external cause category. Deaths classified as "SIDS and cause unknown are examples of those [deaths] that remain unexplained by SUID" (Matthews and MacDorman 2008). The CDC tells us that SIDS is the leading cause of death among infants in the U.S. and the third leading cause overall of infant mortality in the United States (Matthews and MacDorman 2008). In this dataset the SIDS category includes: symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified—SIDS, other symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified, all other diseases (residual). This category of causes of death encompasses 12,801 cases from 2000 to 2004, making it the third largest category of infant deaths in this dataset.

The fifth category is comprised of pregnancy complications. This group of causes includes certain conditions originating in the perinatal period—newborn affected by maternal factors and by complications of pregnancy, labor and delivery, newborn affected by maternal hypertensive disorders, newborn affected by other maternal conditions that may be unrelated to present pregnancy, newborn affected by maternal complications of pregnancy, newborn affected by incompetent cervix, newborn affected by premature rupture of membranes, newborn affected by multiple pregnancy, newborn affected by other maternal complications of pregnancy, newborn affected by complications of placenta, cord and membranes, newborn affected by complications involving placenta, newborn affected by complications involving cord, newborn affected by chorioamnionitis, newborn affected by other an unspecified abnormalities of membranes, newborn affected by other complications of labor and delivery, and newborn affected by noxious influences transmitted via placenta or breast milk. This category includes 8,771 of the infant deaths in the United States in the 2000 to 2004 period.

The sixth category of causes of death is birth asphyxia and birth trauma. This category includes birth trauma, intrauterine hypoxia and birth asphyxia, and respiratory distress of the newborn. This category includes 5,228 of the infant deaths that occurred in the United States between 2000 and 2004.

The seventh category is perinatal infections and it includes infections specific to the perinatal period—bacterial sepsis of the newborn, omphalitis of newborn with or without mild hemorrhage, and all other infections specific to the perinatal period. This

category includes 3,217 of the infant deaths in the United States in the years 2000 to 2004.

Following the categories of Dollfus and colleagues (1990) the eighth category is comprised of “other” infections, to be distinguished from the sixth category of perinatal infections. This category includes certain infections and parasitic diseases—certain intestinal infectious diseases, diarrhea and gastroenteritis of infectious origin, tuberculosis, tetanus, diphtheria, whooping cough, meningococcal infection, septicemia, congenital syphilis, gonococcal infection; viral diseases—acute poliomyelitis, varicella (chicken pox), measles, human immunodeficiency virus (HIV) disease, mumps, other and unspecified viral diseases, candidiasis, malaria, pneumocystosis, and all other specified infectious and parasitic diseases; meningitis; diseases of the respiratory system—acute upper respiratory infections, influenza and pneumonia, acute bronchitis and acute bronchiolitis, bronchitis, chronic and unspecified, asthma, pneumonitis due to solids and liquids, and other and unspecified disease of the respiratory system. This category includes 4,587 infant deaths for the years of 2000 to 2004 in the United States.

The ninth category is endocrine, metabolic and digestive conditions. It is comprised of endocrine, nutritional and metabolic diseases—short stature not elsewhere classified, nutritional deficiencies, cystic fibrosis, volume depletion, disorders of fluid, electrolyte and acid-base balance, and all other endocrine, nutritional and metabolic diseases; diseases of the digestive system—gastritis, duodenitis, and non-infective enteritis and colitis, hernia of abdominal cavity and intestinal obstruction without hernia, and all other and unspecified diseases of the digestive system; disease of the

genitourinary system—renal failure and all other disorders of the kidney, and other and unspecified diseases of the genitourinary system. This category includes 3,397 of the infant deaths in the U.S. for the years 2000 to 2004.

The tenth category of causes of death is neoplasms and blood conditions and it includes neoplasms—malignant neoplasms, Hodgkin's disease and non-Hodgkin's lymphomas, leukemia, other and unspecified malignant neoplasms, in situ neoplasms, benign neoplasms and neoplasms of uncertain or unknown behavior; diseases of the blood and blood forming organs involving the immune system—anemia, and certain disorders involving the immune system. This category includes 612 of the infant deaths in the U.S. for the years of 2000 to 2004.

The last category is diseases of the circulatory and nervous systems and includes diseases of the nervous system—infantile spinal muscular atrophy type I (Werdnig-Hoffman), infantile cerebral palsy, anoxic brain damage, not elsewhere classified, other diseases of the nervous system, and disease of the ear and mastoid process; diseases of the circulatory system—pulmonary heart disease and diseases of pulmonary circulation, pericarditis, endocarditis, and myocarditis, cardiomyopathy, cardiac arrest, cerebrovascular diseases, and all other diseases of the circulatory system. This category includes 3,359 infant deaths for the years of 2000 to 2004.

Independent Variables

The main independent variables in this dissertation are those measuring month of birth and month of gestation. The hypotheses that I intend to test concern whether or not an association between month of birth and infant mortality will be found in contemporary

data. The second goal of this research it to determine whether it is the month that an infant is born that will be of particular interest or the month that an infant is *in utero* is more important. Once the more appropriate independent variable is determined, I will proceed with the multilevel logistic regression that will be discussed later in this chapter. First I turn to a discussion of the rationale and operationalization involved in constructing these two independent variables.

The month of birth variable is relatively straightforward and required minimal manipulation from the layout that can be found in the Linked Birth/Infant Death dataset. In this dataset there already exists a variable measuring the infant's month of birth. In order to include the measure of month of birth I created 12 dummy variables—where a value of one indicates that the infant was born in a particular month and a value of zero indicates that the infant was not born in that particular month. I named these variables born_Jan, born_Feb…born_Dec. When running the multinomial logistic regression I left out one of the variables measuring month of birth in order to compare all other months to that month and because if I included all the months I would have perfect multicollinearity and the model could not be estimated. In chapter IV I will discuss the rationale behind the selection of the most appropriate month of birth variable to be used as the month of comparison.

The variable measuring month of gestation was slightly more involved. When deciding how to create this variable I came across substantive issues that led me to incorporate two measures intended to evaluate month of gestation. Originally the concept of month of gestation became of interest because the hypothesis of this

dissertation are interested in determining whether or not there are significant differences in infant mortality by month of birth. But when thinking about what exactly this means there is the question of whether the month of birth is the important measure, or is it the month of gestation? The literature that finds variation in adult longevity concentrates on the measure of month of birth. For example, as discussed earlier in the literature review, in Denmark individuals born in October, November and December were shown to have lived longer than those born in April, May and June. The reverse pattern was found when the authors examined data from Australia (Doblhammer and Vaupel 2001). Because the hypothesis that these observed differences in life expectancy are due to the seasonal influences that the mother experienced while the individual was *in utero*, it may be more important to know in which months they were *in utero.* Assuming that these individuals were born at term, we can determine which months they were *in utero* if they were born in October, November or December. While it is possible that even in older cohorts we may not be able to assume that all individuals were born at term, the chances for survival for an infant born prematurely were even lower in earlier decades because of lack of sufficient medical care. For the purposes of this dissertation I will assume that when discussing the month of conception for those individuals in older cohorts, we may assume that the majority were born at term.

The literature on embryology discusses the months that a fetus will experience different stages of growth and development. The fetus develops all body parts during the first trimester of the pregnancy (Moore 1974; Rana 1998). This means that during the second and third trimesters the fetus is primarily growing and gaining weight, hopefully

until the end of the pregnancy when it will be able to survive outside the uterus, fully grown and developed. Accordingly, we have to think about which period is the most important to the fetus's development, and is that period also the most important in terms of the survival? If we consider the first trimester the most important then we would be interested in the time that the fetus is developing its organs, body parts, and other tissues, compared to the third trimester when the bulk of weight gain and growth is done. An important issue to consider is the fact that preterm infants that are typically born at very low birthweight have a higher infant mortality than those infants born to term or at 'normal' birthweight. Generally no infant is viable if it is born before 27 weeks of gestation. So the longer the fetus is *in utero* the better its chances of survival; however, if there is a so-called 'dangerous' month for fetuses, maybe being born before that month is beneficial to the chances of the infant's survival.

Obviously, if all infants in the dataset were born at 9 months of gestation, the month of birth and month of gestation in a specified trimester would be simple to calculate. For example, if an infant was born in December then we would know that it was in the first trimester in April, May and June, and in their third trimester in October, November and December. If this were the case for all of the infants there would be no need to construct a separate variable measuring month of gestation because it would be easy to determine which months the infant was *in utero* and it would be the same for all infants, that is, nine months from their date of birth for the first trimester and the last three months of gestation for the third trimester. Creating a separate variable would be

unnecessary because it would be easy to interpret any results by simply subtracting to find the month of gestation of interest.

However, in this dataset there is the important issue to consider, namely that the majority of infants were not in utero for thirty-six (or more) weeks of gestation and are therefore considered preterm. For the years 2000 to 2004, of those infants who died, only 36,570 infants were born at 36 weeks or older; thus almost fifty-five thousand infants were born preterm. This is not unusual since all the cases in this dataset are infant deaths, and it has already been pointed out that infants that are born preterm are at a much higher risk of dying than those that are not. What this means for the creation of the month of gestation variable is that if I want to consider the third trimester the most important period of gestation in terms of an infant's chance of survival after birth, I would have to exclude the approximately 55,000 infants which were born preterm and therefore never made it to their third trimester. With those infants that were born after 28 weeks of gestation I could create 3 variables that specify the months of the first, second and third months of their third trimester. For those infants that did not make it past their first or second trimester a missing value would be given for the relevant dummy variables.

If variation by month of birth is shown to exist in the data, and over half of the infants were born before they reached their third trimester, then maybe it could be said that the differences were due to the month in which the infant was in the first or second trimester instead of the third. About 20,000 infants were born in their second trimester. No infants were born in their first trimester in this dataset; this is probably because any

infants born before the middle of their second trimester are unlikely to be born alive and would be classified as stillbirths or miscarriages. The shortest gestation in this dataset is birth at seventeen weeks gestation. In order to include these and all infant deaths in the dataset I will also create a variable that measures the month in which the infant was in the first, second and third months of their first trimester. I follow this strategy because the literature also suggests that this trimester is very important to the development of the infant (Moore 1974; Rana 1998). The inclusion of this variable will ensure that the first trimester will be examined as the vital period of development for an infant's survival for all infants in the dataset.

Independent Control Variables

The next variables to be described are the control variables. Once I have estimated the models with the two independent variables measuring month of birth and month of gestation I will enter the control variables into the models. If a significant relationship is found, for example, between respiratory causes of death and infants born in (or *in utero* in) January, then I will enter in the control variables into the model. If including the control variables in the model makes the relationship between respiratory causes and being born (or *in utero*) in January no longer significant, then I will have evidence suggesting that month of birth (or gestation) is not having a significant effect on cause specific infant mortality. Instead the important relationship would lie with the control variables and the causes of infant death, and not with the independent variables measuring month of birth or month of gestation. The control variables may be divided into two groups: characteristics of the infant and characteristics of the mother. I will

discuss the operationalization of each of these variables below. For all of the variables used in the analyses of this dissertation, missing values were dealt with before the construction of the variables. When a variable contained missing values such as 99 or 999, those values were assigned a value of '.' so that those cases would not be included in the construction of the variable and therefore inflate the mean. So although the entire dataset from 2000 to 2004 contains 137,951 cases, the exclusion of missing cases will reduce this number to 92,021 cases that will be included in the analyses.

Infant Characteristics

The first group of control variables that I will discuss measures characteristics of the infant. The first variable in this group is the infant's sex. This variable is a dummy variable where a value of one indicates that the infant is a male and a value of 0 indicates that the infant is a female. The second variable is birthweight and it measures, in grams, the birthweight of the infant. It will be considered as a continuous variable in the models. The values of infant birthweight in the dataset range from a low of 227 grams to a high of 8100 grams. In terms of pounds, 227 grams is equivalent to 0.5 lbs, and 8100 grams is equal to 17 lbs and 13 ounces. The mean of the variable birthweight is 1665.65 grams, or about 3.67 lbs.

The last control variable pertaining to infant characteristics is plurality. This variable will be a series of 3 dummy variables, each measuring whether the infant was a single birth, was one of a set of twins, or was one of a set of triplets or more. These three variables will have values of one when the infant is in that category of plurality and a value of zero if not in that category of plurality.

Maternal Characteristics

The next group of control variables pertains to characteristics of the mother. The first maternal variable measures the age of the mother. This variable will be included as a series of four dummy variables of mother's age as follows: fifteen to nineteen, twenty to twenty-nine, thirty to thirty-nine, and forty and higher. The variables will be coded one if the mother is in that age category and zero if the mother is not in that age category. The majority of mothers in this dataset are in the twenty to twenty-nine years age group.

The next variable measuring maternal characteristics is the number of prenatal doctor visits that the mother attended. I also created a series of three dummy variables for this variable. Some researchers claim that measuring prenatal care only by the number of visits is an insufficient measure of prenatal care usage. This owes to the fact that it is also important to consider when the mother began prenatal care as well as the type of care she received (Wise 1994; Fiscella 1995). However, I hold that in the case of this dissertation using the three dummy variables to measure the number of prenatal care will be appropriate. It is generally recommended that a mother have about ten to fifteen visits to a doctor while pregnant (Kogan et al 1998). In order for this to occur the mother would have to start prenatal care early and also continue to visit her doctor throughout the pregnancy. Mothers with at risk pregnancies would be likely to attend more prenatal visits than generally recommended. By creating three dummy variables—one where a value of one indicates that the mother had less than ten visits, one where a value of one indicates that the mother had ten to twenty visits, and a last variable where a value of one indicates that the mother had more than twenty visits—I will be able to accurately

measure the mother's prenatal care usage, while taking these issues into consideration. Also important to consider is that if the infant was preterm the mother may not have attended the recommended ten to fifteen visits before giving birth. Thus, mothers who adequately attended prenatal doctor visits up until the infant's birth may be categorized into the 'less than ten visits' category. However, since the analyses will also be controlling for preterm births, I do not think that this will pose any major issues in the models.

The next maternal characteristic that I will include is mother's race and ethnicity. This will also be measured with a series of dummy variables—white (non-Hispanic), black (non-Hispanic), Hispanic, Asian, and other. A value of one on these dummy variables will indicate that the mother is of that race or ethnicity and a value of zero will indicate that she is not. These variables will be derived from variables in the dataset that measure both race and ethnicity, therefore ensuring that there is no overlapping between the categories.

The next maternal characteristic is mother's education. I will recode this variable into a series of four dummy variables measuring her highest level of completed education as follows: less than high school, where a value of one will indicate that her highest level of education is less than high school; high school education, where a value of one will indicate that her highest level of education is high school; some college, where a value of one will indicate that her highest level of education is one or more years of college; and college or higher, where a value of one indicates that her highest level of education is four years or more of college.

The fifth maternal characteristic is mother's marital status. This variable will be a dummy variable where a value of one indicates that the mother was married at the time of birth, and a value of zero indicates that she was not married at the time of the birth.

The sixth and seventh maternal characteristics are tobacco use and alcohol use during the pregnancy. Both variables will be dummy variables where a value of one will indicate that the mother used alcohol or tobacco while she was pregnant and a value of zero that she did not use tobacco or alcohol while pregnant.

The last maternal characteristic is the amount of weight the mother gained during the pregnancy. This variable will be treated as a continuous variable measuring the weight gain of the mother in pounds during the pregnancy. This variable ranges from zero to 98 pounds.

These control variables will be entered into the model with the independent variables measuring month of birth and month of gestation. I will first ensure that no collinearity exists between the variables; if there is a problematic amount of collinearity, then I will break the control variables into smaller groups to then be entered into the models and only with variables not heavily collinear with the others. By using these control variables I am hoping to find that any relationship uncovered between month of birth and month of gestation will be due to the variation between those variables and the infant's survival and not due to any other factors of the parent or infant, i.e., those entered as controls. In general, by examining these regression results it will be evident whether or not month of birth or month of gestation maintains any effect on infant mortality (as was found in earlier decades), and if this effect varies by the characteristics

of the infant and the parents. Hopefully, the results of this dissertation will permit important conclusions about differences in mortality for contemporary infants. I now turn to a discussion of the methods that will be used to analyze causes specific infant mortality and month of birth and month of gestation.

Methods

This next section of chapter III will be devoted to discussing the methods of analysis that will be used in this dissertation. First I will discuss multinomial logistic regression which will be used to evaluate the degree of association between cause specific infant mortality and the independent variables measuring month of infant birth and month of gestation, as discussed in the first section of this chapter. Then I will turn to a discussion of multilevel logistic regression that will also be used in the analyses. Finally I will discuss the mapping methods using ArcGIS that I will use in a final analysis of the geographic components to infant mortality in the United States. Chapter III will conclude with a brief summary of the data and methods to be used in subsequent chapters.

Multinomial Logistic Regression

In order to examine the phenomenon of infant mortality by month of birth (or gestation) I will employ multinomial logistic regression. The use of hazard or survival analysis would seem to be a logical choice for the analysis of infant survival based on month of birth and is frequently used in infant mortality studies (see Trussell and Hammerslough 1983; Agha 2000). However, the Linked Birth/Infant Death data to be used in this dissertation prohibit the use of hazard or survival analyses. This is because the datasets are created by linking infant death certificates to birth certificates and thus information is

only available on those infants that have died in a specified year. The use of hazard or survival analysis necessitates the use of data in which some individuals have experienced the event of interest (in this case an infant death) and some individuals have not experienced the event of interest, thereby enabling the researcher to examine the risk of death, since only some of the individuals in the dataset would be dead.

The Linked Birth/Infant Death dataset has many advantages that have been discussed previously in chapter III. The ability to use information about the infant, mother and father from the death certificate enables me to determine whether there exists a relationship between month of birth and infant mortality; and if a relationship exists, will it be sustained after entering various control variables? In order to incorporate the variables that are only available on the birth certificate and study the phenomenon of infant deaths, the analyses of this dissertation will use a multinomial logistic regression model, in fact two such models, one in which month of birth is used as the independent variable of interest, and one in which month of gestation is the independent variable of interest. After determining which independent variable has the most impact on infant mortality, a multilevel logistic regression will be performed which will help determine whether the state-level characteristics, i.e., the aggregate contexts, are important to the study of infant mortality. Multilevel logistic regression methods will be discussed in the next section of chapter III. First I turn to a discussion of the multinomial logistic regression.

The multinomial logistic regression model is a generalization of the binomial model where the "number of response categories exceeds two" (Agresti 1996:205). Agresti tells us to

> suppose *Y* is a nominal variable with *J* categories…Let $\{\pi_1, \ldots, \pi_j\}$ denote the response probabilities, satisfying $\Sigma_j \pi_j = 1$. When one takes *n* independent observations based on these probabilities, the probability distribution for the number of outcomes that occur of each of the *J* types is the multinomial. It specifies the probability for each possible way of allocating the *n* observations to the *J* categories (Agresti, 1996:205).

Multinomial logistic regression models (also called polytomous logit model) enable the researcher to consider simultaneously the effects of the X variables on each of the outcome categories relative to the baseline category; the model reports the log odds of the response in one category instead of another.

Formally the model may be written as:

$$\ln \Omega_{m|b}\,(\mathrm{x}) = \ln \frac{\Pr(y = m \mid x)}{\Pr(y = b \mid x)} = x\beta_{m|b} \qquad \text{for } m = 1 \text{ to } J$$

where *b* is the base category, which is also referred to as the comparison group. As $\ln \Omega_{m|b}\,(\mathrm{x}) = \ln 1 = 0$, it must hold that $\beta_{b|b} = 0$.

That is, the log odds of an outcome compared with itself are always 0, and thus the effects of any independent variables must also be 0. These *J* equations can be solved to compute the predicted probabilities:

$$\Pr(y = m \mid x) = \frac{\exp(x\beta_{m|b})}{\sum_{j=1}^{J} \exp(x\beta_{j|b})}$$

(Long and Freese, 2001:227-228).

The multinomial logistic regressions to be estimated in this dissertation use the baseline category of "external causes," in order to compare all the other categories of infant death to those causes classified as external.

> When the last category (*J*) is the baseline the *baseline category logits* are:
> $\log\left(\frac{\pi j}{\pi J}\right)$, $j = 1,\ldots J\text{-}1.$
> Given that the response falls in category *j* or category *J*, this is the log odds that the response is *j*. For $J = 3$, for instance, the logit model uses log (π_1/π_3) and the log (π_2/π_3) (Agresti, 1996:206).

So, as previously mentioned, each of the categories of the outcome is compared to the baseline category. STATA will provide a logit for each of the categories compared to the baseline category.

With my eleven category dependent variable, STATA will provide a coefficient for each of the ten categories compared to the baseline category that I select—external causes of death. Multinomial logistic regression models are estimated using a maximum likelihood model.

> Maximum likelihood (ML) estimates are the values of the parameters that have the greatest likelihood of generating the observed sample of data if the assumptions of the model are true….the likelihood function tells us how likely it is that we would have observed the data that we did observe if these data were the true population parameters (Long and Freese, 2001: 76).

The coefficients are estimated in relation to the baseline category. For my eleven category dependent variable, and thirty-two independent variables that will be included in the full model, STATA will estimate three-hundred and twenty coefficients (or K(J-1); where K= the number of independent variables and J= the number of categories of the dependent variable). This may seem like a large number of coefficients, but since only the variable measuring month of birth and month of gestation are of interest, interpreting all of the coefficients for the variables that are considered controls will not be necessary. The importance of the control variables is only to determine whether a

relationship between month of birth or month of gestation retains its significance with their inclusion. Also, because of multicollinearity it is unlikely that a full model would be able to include all thirty-one variables in one model. This will reduce the overall number of coefficients estimated in the models. Each coefficient will represent the log odds of the categories or metric of the independent variables of being in the dependent category of interest, as compared to being in the baseline category of external causes.

Beyond the difficulty that a large number of coefficients that will be estimated in my full model can pose for interpretation, difficulty can also arise from the interpretation of coefficients in terms of log odds. Hamilton (1992) tells us that log odds or logit interpretation is easy to state, but is not so easily understood. For example if the independent variable measuring whether or not the infant was born in January had a coefficient of 1.00 for those infants that died of respiratory causes, we could say that infants born in January have a log odds of dying from respiratory causes versus external causes that are 1.00 times higher as compared to being born in any other month. It is difficult to imagine what it means to have a 1.00 higher log odds; is this difference large or small? Because of this complexity it is best to interpret the coefficients in terms of odds ratios by exponentiating the log odds. In multinomial logistic regression these exponentiated values are called relative risk ratios or rrr's. For a dummy variable the rrr is the odds of being in the dependent variable category of interest and not being in the base category, for the category of the independent dummy variable with a value of one versus the category with a value of zero. For ease of interpretation I will obtain the rrr's for the multinomial logistic regression models that I will estimate.

Long and Freese (2001) tell us that we should be interested in looking at different combinations of baseline to other category comparisons, beyond the category that we define as the baseline. By using the 'listcoef' command, we can display all combinations of outcome categories without re-estimating the multinomial logit model in STATA. The 'listcoef' command will be used in my analysis for this and other purposes in order to determine that my choosing the baseline category 'external causes' is indeed the best option.

Predicted Probabilities

Because it can sometimes be overwhelming to interpret the many coefficients that are yielded in the multinomial logit model, a preferred way to look at the results of a multinomial logit model is by predicting the probabilities of being in the outcomes of the dependent variable. In this dissertation I can calculate the predicted probabilities of an infant being in each of the eleven total categories of the dependent variable based on the thirty-one total independent variables. After performing the multinomial logit model, STATA will produce the predicted probabilities for each of the categories of the dependent variable. This can be helpful in interpreting the results of the multinomial logit and will also enable me to construct graphs displaying the relationship of being in one of the categories of causes of death based on the independent variables of the model.

Tests for Multicollinearity

An important test to perform in any analysis is testing for multicollinearity between the independent variables used in a model. When variables are collinear with one another it often means that some of the variation in one of the independent variables is explained

by one or more of the other independent variables. While we do want the variation in the dependent variable to be explained by the independent variables in the model, when multicollinearity exists between the independent variables of a model it will frequently result in some inaccuracy in the parameter estimates. “For instance…strong correlations among predictors [may make]…it seem that no one variable is important when all the others are in the model. A variable may seem to have little effect simply because it ‘overlaps’ considerably with other predictors in the model” (Agresti 1996:126). We may find that the overall model has a low P value which would indicate that at least one the independent variables is significantly different from zero; however the individual ‘z’ tests may not show significant results. This is a classic indication of multicollinearity; however in some cases it may not be so obvious and multicollinearity may go unnoticed. This is why it is important to evaluate multicollinearity for all models regardless of whether there may seem to be any in the model at first examination.

In this dissertation I must be aware of the potential between some of the variables used to study the characteristics of the mother and also the variables that measure infant characteristics such as birthweight and gestational age, since these types of variables may be closely related and, if so, perhaps should not be entered into the same model. Entering collinear independent variables into the same model would be redundant and could cause problems with the predictors. If multicollinearity is found between any of the independent variables I will overcome it by estimating several models with several combinations of independent variables that are not collinear with one another.

So how is multicollinearity assessed? After performing a regression in STATA we can enter the command “VIF” which stands for variance inflation factor and STATA gives us a VIF value and a 1/VIF value (or the tolerance) for each of the independent variables. If the tolerance value is less than 0.35 it indicates that multicollinearity may be a problem in estimating the model. Although obtaining the VIF values for the independent variables is the best way to assess multicollinearity, it can also be partly assessed by looking at the zero-ordered correlations among each pair of my independent variables. I will consider all these options to assess collinearity which will ensure that there are no problems with estimating my model.

Multilevel Logistic Regression

After estimating the multinomial logistic regressions, I will have determined the independent variable that is the best indicator of the infant’s experience—month of birth or month of gestation. After these independent variables have been identified I will then perform the multilevel analysis with the same dependent variable—cause of infant death. Multilevel analysis is relatively new to the social sciences, meaning that unlike other methods of analysis it has gained its popularity only in the past few decades. I will now discuss multilevel models and the level-2 (aggregate) variables that I will introduce into my models to further examine infant mortality by month of birth.

Multilevel models have a longer history with researchers interested in education. These researchers found that some micro-analysis did not fully capture the phenomenon that individuals were nested within classes and classes are nested within schools and schools were nested in school districts. “Previously these problems had been approached

by either aggregating individual-level variables to the group level or disaggregating group-level variables to the individual level" (de Leeuw and Meijers 2008:1). Both of these methods are inappropriate as they can severely bias results, often by committing the ecological fallacy. The ecological fallacy tells us that there is a difference between ecological correlations and individual correlations and attributing characteristics to individuals based on data from a level above the individual level is a mistake, oftentimes indicating an incorrect or reversed relationship (Freeman 1999). Multilevel analysis allows the researcher to treat the variables that are collected at a particular level at that level without aggregating or disaggregating, thereby avoiding the ecological fallacy or inferring information based on biased data due to level of collection.

Due to issues with running the multilevel models that I will discuss in detail in a later chapter, I was unable to perform a multilevel multinomial logistic regression in chapter VI. Instead I ran a series of multilevel logistic regression models. The main difference that arises from switching from a multinomial logistic regression to a logistic regression is that instead of running only one model, I ran a series of models that separately incorporated all the causes of infant death shown to be significantly associated with any of the month of birth independent variables in chapter IV.

The multilevel analyses of chapter VI are termed "hierarchical generalized linear model" (HGLM) (Raudenbush and Bryk 2002:291) since the analysis will concentrate on dichotomous outcomes and non-normally distributed errors. HLGM is the appropriate method to use when the dependent variable is dichotomous. This is because dichotomous dependent variables can only take on value of zero or one and are therefore non-linear,

non-normally distributed, without homogenous variance (Raudenbush and Bryk 2002:292). Using HGLM I will be able to assess the influence of the infant's month of birth or month of gestation, as well as the contextual characteristics of the infant's state of birth on the infant's likelihood of dying of a specified cause. This will allow me to examine the effects of the level-one and level-two variables on the odds of an infant dying of a specified cause, in addition to the effects of the level-two variables on the level-one slopes.

The HGLM "uses a binomial sampling model and a logit link" (Raudenbush and Bryk 2002:294). The multilevel analyses of this dissertation will use a sampling model known as the Bernoulli distribution. It is also important to note the iterative process that the HGLM uses, which is distinct from the maximum likelihood estimation process in the HLM (hierarchical linear model). When using the HGLM model the user provides the maximum and minimum number of micro and macro iterations (Bryk and Raudenbush 1992). Because of this iterative process, HGLMs may take longer to converge than HLM equations. As I will discuss in a later chapter, this involved iterative process is the main reason that I could not use a multilevel multinomial logistic regression model in chapter VI.

Bryk and Raudenbush (1992) set forth the following equation as the level-one sampling model:

$$Y_{ij} \mid \varphi_{ij} \sim B\,(m_{ij},\ \varphi_{ij})$$

> where Y_{ij} is the number of "successes" in m_{ij} trials and let φ_{ij} be the probability of successes on each trial... Y_{ij} has a binomial distribution with m_{ij}

trials and the probability of success per trial as φ_{ij}" (294). "According to the binomial distribution, the expected value and variance of Y_{ij} are then:

$$E(Y_{ij} | \varphi_{ij}) = m_{ij} \quad \varphi_{ij}, \text{Var}(Y_{ij} | \varphi_{ij}) = m_{ij} \quad \varphi_{ij} (1 - \varphi_{ij})$$

When m_{ij} is a binary variable taking on a value of either zero or unity. This is a special case known as the Bernoulli distribution (295).

The most common link function for binomial sampling models at level-one is the logit link (295): $n_{ij} = \log\left(\frac{\varphi_{ij}}{1-\varphi_{ij}}\right)$ "…where n_{ij} is the log of the odds of success….[n]ote that while φ_{ij} is constrained to be in the interval (0, 1), n_{ij} can take on any real value" (295). The level-one structural model is shown as:

$n_{ij} = \beta_{oj} + \beta_{1j} X_{1ij} + \beta_{2j} X_{2ij} + ... + \beta_{pj} X_{pij}$ and the level-two structural model is shown as: $\beta_{qj} = \gamma_{q0} + \sum_{s=1}^{S_q} \gamma_{qs} W_{sj} + u_{qj}$ (Raudenbush and Bryk 2002).

In the multilevel logistic regression that I will be estimating in this dissertation, I will be incorporating several level-2 variables intended to measure the state-level influences on infant mortality. The hypothesis that links month of birth to infant mortality is based on the differences in nutrition that a mother may experience throughout the year, as well as differences in climate which may increase the chances that a mother catches communicable illness or that make nutritious food less available. The level two variables I will include will gauge state-specific measures of climate.

I include these state-level (level-two) variables to enable me to examine whether, when including all the level-1 measures discussed up until this point, any of the differences in infant mortality can also be explained by the state-level differences in climate throughout the year. I may also be able to determine whether the information at

the state level enhances or has no effect on the relationships that I observed in the level-one models. For example, if there is a relationship between infant mortality and a particular month of birth, I may hypothesize that the measure of the climate of a state may reduce or increase the effect (the slope) of month of birth on cause specific infant mortality.

Spatial Methods

In the next section of this chapter, I will review the methods I will use to graphically display infant mortality in the United States for the year 2004. I have selected some variables I feel will be important to the infant mortality by month of birth discussion in the United States. I will discuss these variables, how they were created, and the data that generated them. I will also discuss the methods I will use to create the maps using the program ArcGIS. This section will end with a conclusion of the data and methods section of this dissertation.

Mapping Data

In order to examine this relationship of infant mortality by month and see if a spatial relationship exists, data from several sources will be used. First, data from the Centers for Disease Control (CDC) will be used to determine the infant mortality rate by state of residence. On the CDC website there is an application that allows tables to be downloaded that includes various statistics by state, or other specified geographies. For my purposes the total number of infant deaths, the total number of births and the resulting death rates were downloaded by state. The table also includes the ‘statefips’

code which is a unique identifier for each state and will be used when other data to be included in the resulting map are merged.

The state data will be downloaded from the ESRI website. ESRI—Environmental Systems Resource Institute, is a software development and services company that provides Geographic Information Systems (GIS) software and geodatabase management applications. A geodatabase is a database that is designed to store, query and manipulate geographic data. In the ESRI geodatabase there are several sources of information. One type of information available is shape-files of the United States. These shape-files give the map the shape of the United States without any other information. In GIS, maps are created as a compilation of layers. When all these layers are brought together, they create a map that displays the information that the user specifies (Ormsby et al 2001). In the maps to be used in this dissertation, the layers of information will come from these shape-files of the U.S., from the information on infant mortality from the CDC, and by the state-level measures that I will incorporate. This will allow the infant death rates to be displayed by state once the CDC data and the ESRI data are joined together.

Another possibility would be to include information about the climate of each state. These additional data may help me to examine the seasonal patterns of change that exist to varying degrees in different parts of the United States. Possibilities for this type of data would be average rainfall, average summer or winter temperatures, or other state-specific information on climate. Such data may enhance the findings that infant death rate by state would provide.

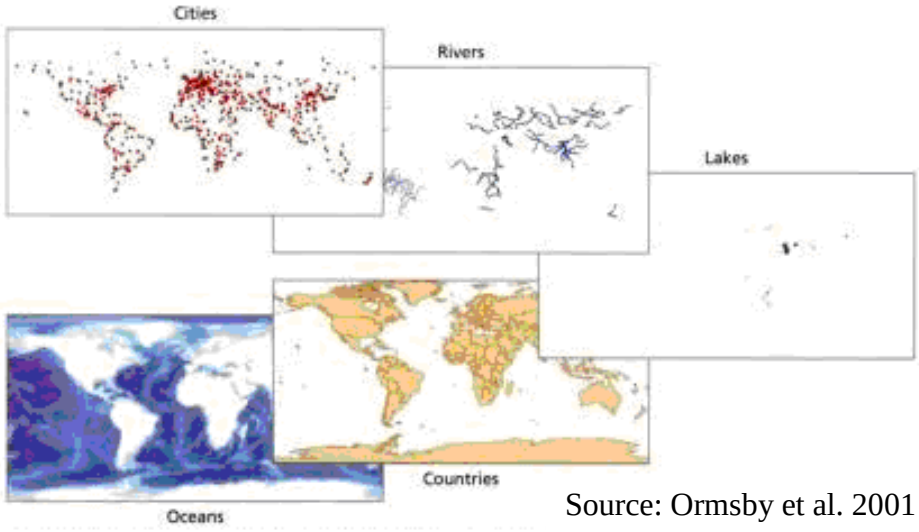
By using these data from a variety of sources, infant death rates by state will be able to be shown on one map. This will allow the observation of any patterns, if patterns do exists, to be examined by state of residence. If a pattern is found, it may give merit to previous research on infant death and month of birth that found cause in the temperature and nutrition (of the mother and infant) to vary geographically.

Mapping Methods

"A GIS is defined as a system for capturing, storing, checking, integrating, manipulating, analyzing, and displaying data that are spatially referenced to the earth" (Bateman et al. 2002:219). In ArcGIS the user can perform all of these actions without having to move data back and forth between programs. All the functions needed in this dissertation to examine infant mortality can be performed and displayed by ArcGIS software.

Basically, a GIS map is a series of layers constructed by entering the data of interest. Figure 3.1 below shows an example of layers that could be put into an ArcGIS map.

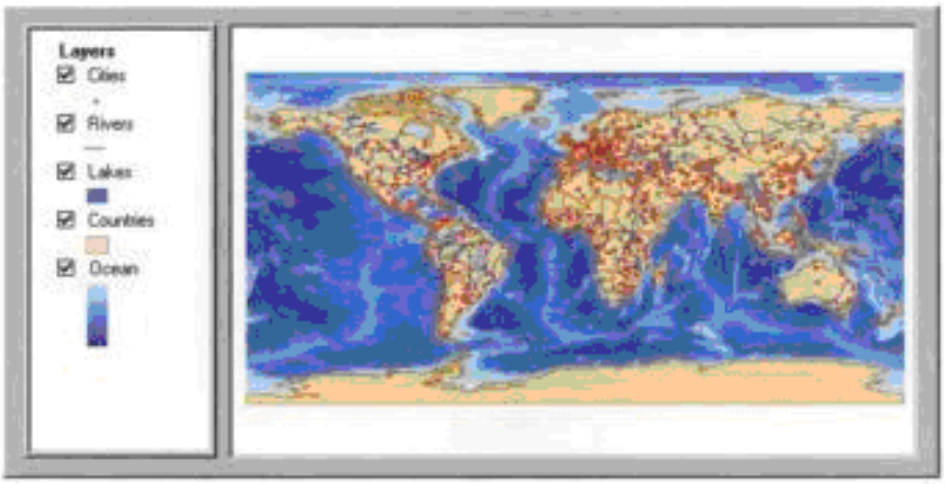
Figure 3.1 An Example of Layers in ArcGIS



Source: Ormsby et al. 2001

The layers are then ‘layered’ on top of one another to create a map, such as that seen below in Figure 3.2.

Figure 3.2 An Example Map



Source: Ormsby et al. 2001

The layers in Figure 3.1 are combined to form the map in Figure 3.2. All maps that are created in ArcGIS are created as a series of layers. In the maps that I will construct the layers will be different, but the process will be the same. I will create layers that together will display infant mortality by month of birth or gestation while incorporating the information about rainfall and average temperatures.

As one might imagine, there are an infinite number of shapes that may be represented on any given map. But in ArcGIS all the features that may be displayed on a map are represented either by a polygon, a line or a point.

> Polygons represent things that are large enough to have boundaries, such as countries, lakes, and tracts of land. Lines represent things too narrow to be polygons, such as rivers, roads and pipelines. Points are used for things too small to be polygons, such as cities, schools, and fire hydrants….Polygons, lines, and points collectively are called vector data (Ormsby et al 2001:3).

When something to be displayed on the map is not able to be displayed in terms of points, lines or polygon then it is called a feature. Features are those items that do not have a distinct shape such as

> elevation, slope, rainfall, and wind speed. What they have instead are measurable values for any particular location on the earth's surface…The most common type of surface is a raster, a matrix of identically sized square cells. Each cell represents a unit of surface area…and contains a measured or estimated value for that location (Ormsby et al. 2001:3).

In ArcGIS, features are linked to information that is contained in tables called attribute tables. This means that not only do features have shapes and locations in ArcGIS, they also contain information. By clicking on a certain place on the map one may bring up the information in the attribute table about that location. The information contained in the map and the information that is contained in the attribute table can help answer questions that may be pertinent to analyses.

For example, I would be able to answer, in which state is infant mortality the highest? And in that area of high mortality, is there also high average temperature? The capabilities of ArcGIS go beyond simple map construction and allow the researcher to investigate important questions about the data contained in that map.

Conclusion

The purpose of chapter III is to introduce the data and methods that I will be using in the analyses of this dissertation. The nature of the question I am investigating and the data that I am using necessitate the use of a multinomial logistic regression model for the level-one analyses, as the dependent variable—cause of infant death—is an unordered, nominal categorical dependent variable. The strengths of the Linked Birth/Infant Death

dataset were discussed as well as the methods that will be used in subsequent chapters of this dissertation. Chapters IV and V will provide the results of the two multinomial logistic regression models that I will perform. Then in chapter VI I will provide the results of the multilevel logistic regression model as well as discuss in further detail the reasons for switching to a logistic regression in the multilevel analyses. The results chapters will conclude with chapter VII, which will provide the geographic investigation of infant mortality using mapping methods in ArcGIS. This dissertation will conclude with overall conclusions of all findings of the three methods chapters, as well as a discussion of the implications of these findings. I will also discuss possible areas of future research in the field of biodemography. Next I turn to a discussion of the first model of this dissertation, using month of birth as the independent variable of interest.

CHAPTER IV

LEVEL-ONE ANALYSES AND RESULTS:

MONTH OF BIRTH AS THE PRINCIPAL INDEPENDENT VARIABLE

In this chapter I discuss the results of the multinomial logistic regressions using the variable 'month of birth' as the principal independent variable of interest. I will estimate three models, as already mentioned in chapter III—one with only the month of birth variables, one with both the month of birth variables and the infant characteristics variables as controls, and a third with the month of birth variables, the infant characteristic variables and the maternal characteristics variables as controls. I hypothesize that after controlling for infant and maternal characteristics that the causes of deaths will have a positive association with the months of the late spring and early summer and a negative association with the months of the late fall and early winter.

First I present and discuss several figures and tables that will provide some perspective and overall legitimacy for the analyses of this dissertation. The main purpose of these figures and tables is to show that the assumptions of the models are founded on the actual distribution of the births and deaths of infants in the United States. I hope that these figures and tables clarify the data and the way that I intend to use these data to study cause specific mortality by month of birth in the United States.

As I discuss later in this chapter, issues of multicollinearity prohibit the inclusion of the variable measuring gestational age and the variable measuring birthweight in the same model. Accordingly, I will estimate two series of models—one with the gestational age variable included as a control, and one with the birthweight variable included as a

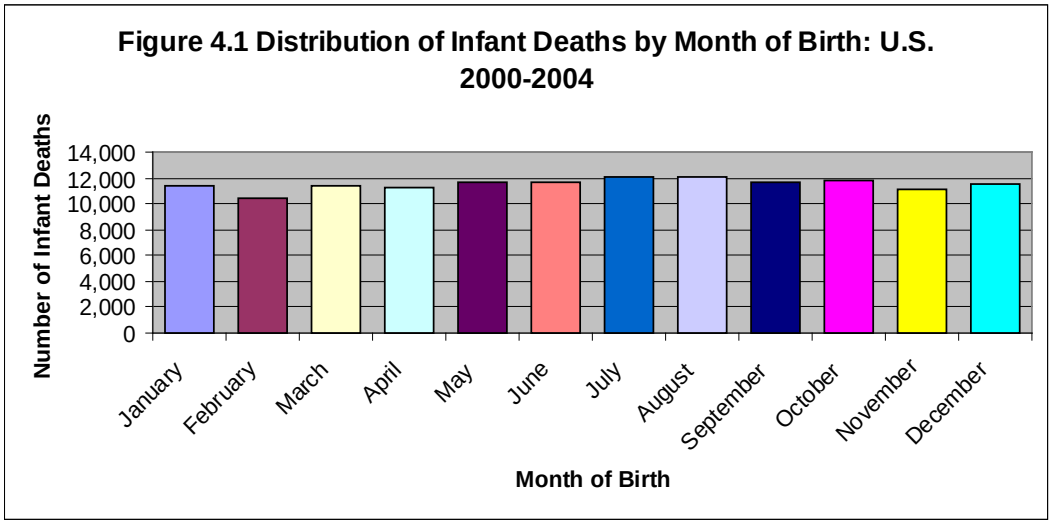
control. This will result in a total of six models, all intended to measure cause specific infant mortality by month of birth. I will then discuss the findings of the likelihood ratio test which tests the hypotheses of the models. I then include a discussion of the results in terms of predicted probabilities. The chapter will be concluded with a brief overview of the findings of the models.

The aim of my dissertation is not to develop a comprehensive explanatory model of infant mortality. Instead my aim is to examine the relationship between an infant's month of birth, or month of gestation, and their odds of dying from a specified cause of death. Then, by incorporating the control variables measuring various infant and maternal characteristics, I hope to see if the observed relationship between cause specific infant mortality and month of birth or month of gestation is maintained. This chapter will discuss the findings of the level-one models.

Data Considerations

As discussed in chapter III, the dependent variable in this dissertation is measured in two ways—month of birth and month of gestation. I will estimate the same models with each of the dependent variables in order to determine if one is a better measure of the relationship that I aim to investigate, namely the link between month of birth or month of gestation and infant mortality. First I will show the distribution of the dependent variable and the independent variables month of birth and the infant and maternal control variables.

Figure 4.1 shows the distribution of the variable measuring the month of birth for all infants who died in the U.S. during the years 2000 to 2004, of all causes.



There were a total of 137,951 infant deaths in the U.S. for the years 2000 to 2004. The month with the highest number of infant deaths is July with 12,132 deaths followed closely by 12,014 deaths in August. Table 4.1 below shows the actual number of infant deaths in each month for the years 2000 to 2004.

Table 4.1 Number of Infant Deaths by Month: U.S. 2000-2004

January	11,403
February	10,365
March	11,443
April	11,271
May	11,661
June	11,598
July	12,133
August	12,014
September	11,618
October	11,767
November	11,099
December	11,578
Total Deaths	137,951

The number of infant deaths by month in the U.S. for the years 2000 to 2004 seems to show only slight differences by month of the year. According to the literature on seasonality of births there has generally been found a peak in the autumn months,

which would correspond with a rise in conceptions in the months of November and December (James, 1990). Although there appears to be variation in human fertility for almost all societies that have been investigated, there is no true consensus on why this pattern exists. (For more information on seasonality of births see Lam and Miron 1994 and Udry and Morris 1967) If there is a certain month or season with a significantly larger number of births, then it may be expected, other things equal, that there would also be an increase in infant deaths in a certain month, associated with the greater number of infant deaths during that, or even an adjacent month. It would be hard to determine from the data used in this dissertation whether or not any of the deaths in the U.S. from 2000 to 2004 are due to an increase in the number of births. More births would make infant deaths more likely, since more infants being born would increase the number of possible infant deaths. However, there seems to be no true and consistent peak in infant deaths for the years of data I am using in my dissertation.

In order to ascertain whether or not there is a true peak in the number of infant deaths owing to a peak in the number of births for those same months, I also downloaded and examined data from the natality file for the years 2000 to 2004. As was the case with the Linked Birth/Infant Death dataset, these files are also available from the Centers for Disease Control (CDC) for individual years and encompass all births that occurred in the United States for a given year. (For more information on the U.S. Natality files please see http://www.cdc.gov/nchs/births.htm.) I compiled the natality files for the years 2000 to 2004 for the U.S. so that the resulting file would contain all births for the U.S. for same years as the Linked Birth/Infant Death file that I am using in the dissertation. My

objective is to describe the seasonal variation of the births in the U.S. for the years 2000 to 2004 in order to see whether or not the peaks and troughs in the number of infant deaths observed for a certain month might be due to overall increases and decreases in the number of births for those same months. Figure 4.2 below shows the distribution of births for the years 2000 to 2004 in the U.S. Again in these data we see that the highest number of births occurred in August, which was the month with the second highest number of infant deaths as shown in Figure 4.1 and in Table 4.1. This gives merit to the idea that some of the reason for increases to the number of infant deaths is due to the overall increase in births for the same period. Table 4.2 shows the overall number of infant deaths by month for the years 2000 to 2004. The data in Table 4.2 indicate that the highest number of births in the U.S. for the years 2000 to 2004 was in the months of July and August.

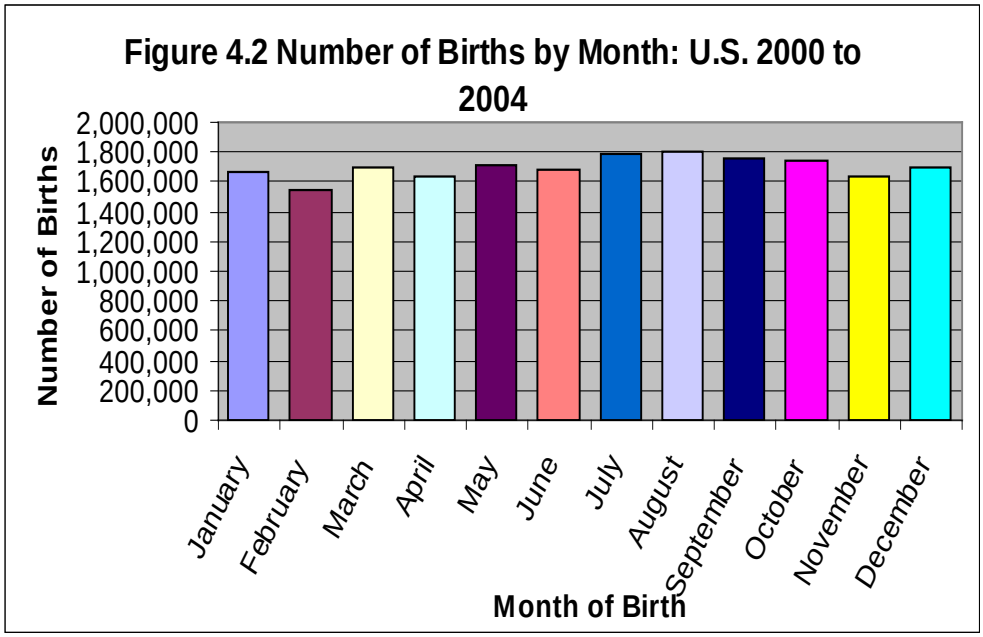
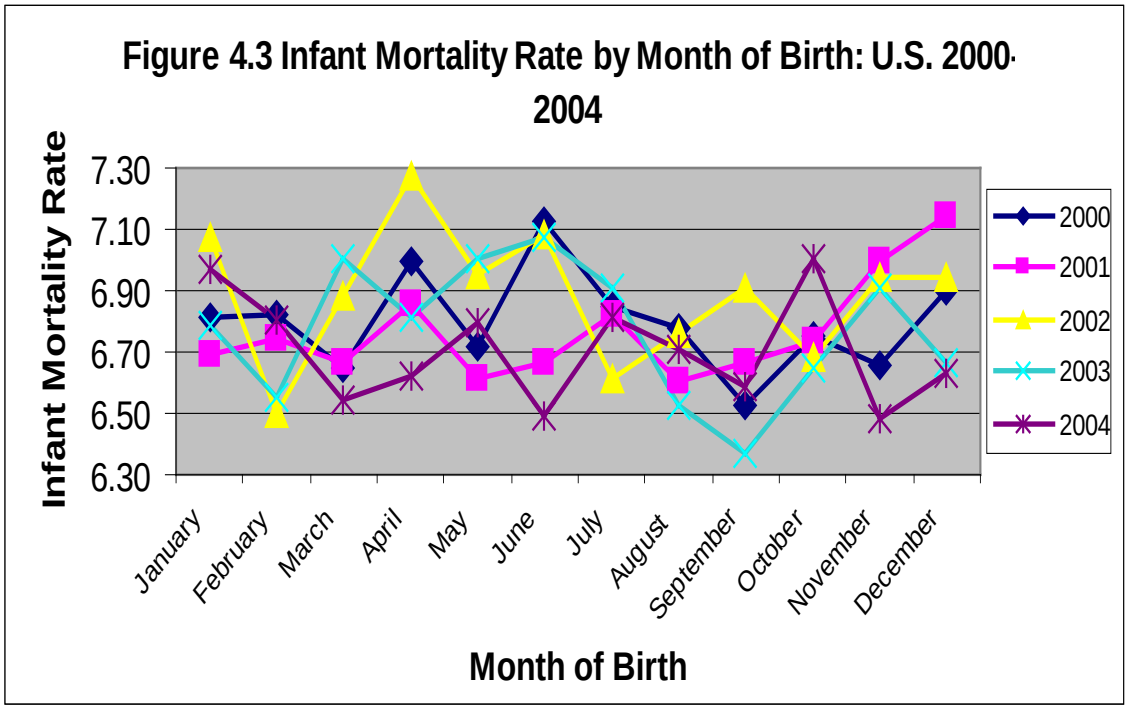


Table 4.2 Number of Births by Month: U.S. 2000 to 2004

January	1,660,926
February	1,549,934
March	1,696,186
April	1,630,868
May	1,710,945
June	1,684,537
July	1,784,085
August	1,799,623
September	1,757,487
October	1,739,005
November	1,634,481
December	1,689,652
Total Deaths	20,337,729

Might the seasonal variation in the number of births be the main reason for the increase in infant deaths? If any variation that can be seen in the number of infant deaths is indeed due to an overall increase in the number of births, then the idea that a certain season or month is the reason for the increase in deaths may only be a part of the seasonal increases. However, in this analysis I believe that the seasonal variation in births has an unimportant impact on the cause specific analysis of infant deaths.

In order to further illustrate the relationship between infant deaths and births by month I calculated the monthly infant mortality rate by month for single years from 2000 to 2004. The resulting graph is shown in Figure 4.3.



The intent of Figure 4.3 is to display the month-specific infant mortality rate for the years for which I will be analyzing infant deaths. As we saw before, the largest number of infant births and infant deaths are in the months of July and August. However, we see in Figure 4.3 that the months of July and August do not have the highest infant mortality rates for the five years shown in the figure. For all five years the infant mortality rate seems to be highest in the late spring and early summer months of April, May and June as well as in the winter months of November, December and January. So although the overall number of infant deaths is high in the late summer, when looking at the infant mortality rates for those months we see several different peaks. In most cases the peaks between the years seem to be within a month or so of each other—the peaks for 2000, 2001 and 2002 take place in April and May and the

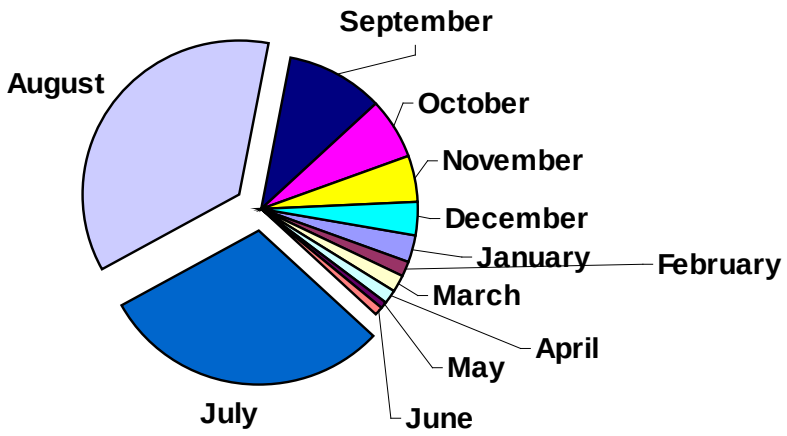
peak for 2003 appears in June. This shows a similar pattern although it does not show conclusively a consistent monthly pattern of infant mortality exists.

The difficult issue to assess here is whether or not the monthly distribution of deaths is truly due to the increases in births or is due to other factors. Even if we do not observe a peak in the infant mortality rate that coincides with the peak in the number of infant births and infant deaths, there still may be a connection. For example, although infant deaths are most likely to occur in the first seconds, minutes, hours, days, and weeks after birth, many deaths still may occur several months after birth. This means that there may not be any real way to tell from looking at the numbers of births and deaths if the increases in infant births are affecting the number of infant deaths since we cannot be sure that an infant death that is occurring months after its birth is somehow connected to its month of birth.

In order to determine if there is any connection between the increase in infant births in the summer months and the increase in infant deaths, I looked at the months of death only for those infants born in the months of July and August, when the overall number of births is the highest. Figure 4.4 below shows the monthly distribution of deaths to infants born in July and August for the years 2000 to 2004. Overall, the majority of infants born in July or August also died in July or August. There were 24,150 infant deaths for the years 2000 to 2004 to infants who were born in the months of July and August. Of this total number of deaths, 7,210 of the infants died in July, and 8,715 of the infants died in August. I believe that this provides some evidence that the majority of the infants died within a month of their birth, especially those infants born in

the months of July and August, when there was a peak in births and deaths. As we can see, the largest portion of the pie chart is August, followed by July. Then the pie slices decrease in order, starting with September and ending with the smallest slice in June—the furthest month from July that a death may occur and still be classified as an infant death (since there is not distinction here regarding year).

Figure 4.4 Month of Death to Infants Born in July or August: U.S. 2000-2004



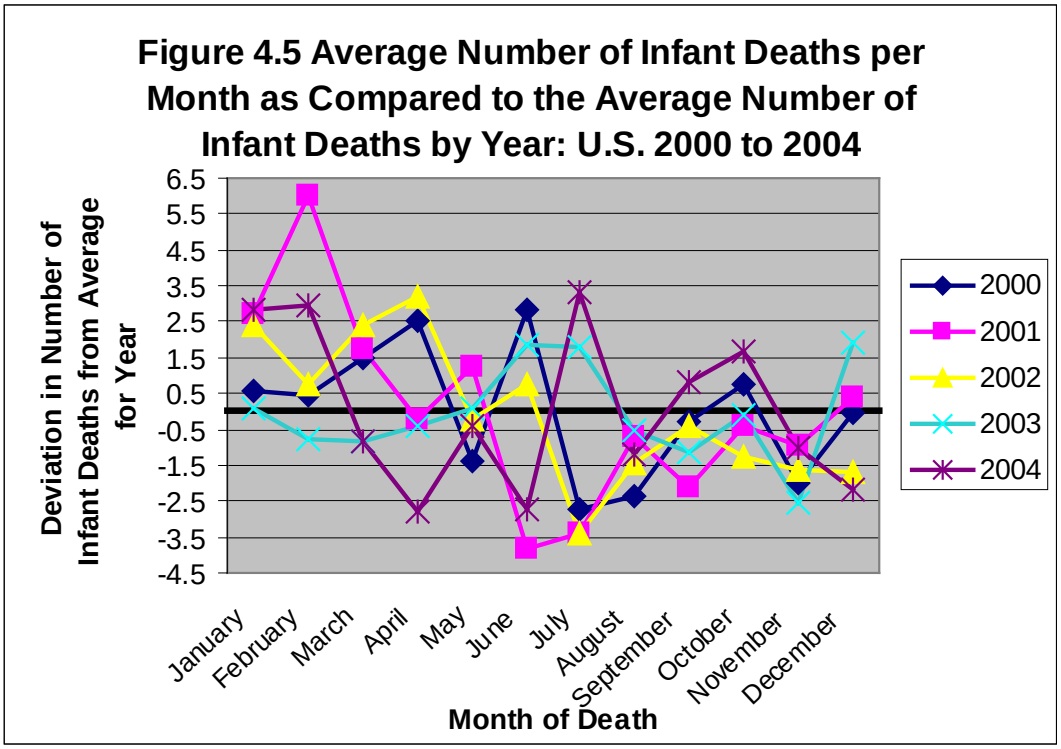
I also tabulated the number of infants who died in the same month in which they were born. These results are shown in Table 4.3. The purpose of this table is to show that in the majority of cases, many infants tend to die within their same month of birth. This further supports the findings of comparing the results found in Figures 4.3 with those in Figure 4.4.

Table 4.3 Frequency and Percent of Infant Deaths within Month of Birth: U.S. 2000-2004			
Month	**Number Infants Dying within Month of Birth**	**Total Number of Infant Deaths in Month**	**Percent of Deaths Occurring in Birth Month**
January	6762	11972	56.5
February	6177	10991	56.2
March	6995	11828	59.1
April	6916	11397	60.7
May	7096	11685	60.7
June	6937	11294	61.4
July	7108	11570	61.4
August	6964	11513	60.5
September	6665	11235	59.3
October	6917	11727	59.0
November	6274	11083	56.6
December	6662	11656	57.2
Total Deaths	**81473**	**137951**	**59.1**

In every month, more than fifty percent of all deaths to infants occur in their same month of birth. This means that the majority of all infants die within their month of birth even without taking into consideration how late or early they were born in that month. So, when comparing the peak months of infant deaths and births in July and August with the finding that the highest infant mortality rates for the years 2000 to 2004 occur in the months of early summer or winter, we can say that it is not necessarily due to those infants being born during the peak and then dying in those months. Instead, in the bulk of cases, the infants in these data are dying within their same month of birth and not in the winter or almost a year after birth in the early summer when we see the peaks in the infant mortality rate by month. In fact, when considering only those infants who did not die of external causes of death, an even higher percentage of infant deaths occur within their month of birth or the month after. For example, for those infants born in November or December, slightly more than seventy-three percent (73.4%) died either in November or December, only considering those infants who did not die from an external cause of death. Even when considering all causes of death, this percentage drops to only

71 percent of all infants born in November dying in either November or December. I believe this provides further evidence to the fact that if the seasonal or monthly increases in births were responsible for the increases to the number of infant deaths, then we would see a sharp increase to the infant mortality rate in the months of July, August and September—the months (or adjacent months) that saw the largest number of births in the U.S. for the years 2000 to 2004.

Another strategy I employed to further justify the analyses of this dissertation was by finding the average number of deaths to infants in the U.S. for the years 2000 to 2004 in order to see if the number of infant deaths for each month was higher or lower than that average. This will show, without the effect of the number of days in the month, whether or not any particular month has more deaths than the daily average for each of the four years. First I took the total number of infant deaths in the U.S. for each of the single years of 2000 to 2004, and divided it by the number of days in that year. For example, in 2000 there were 27,622 infant deaths and there were 366 days in the year. This gave an average number of infant deaths per day of 75.5. Then for each month, I took the number of infant deaths in that month and divided by the number of days in that month. For example, in January of 2000 there were 2,250 infant deaths; dividing this by 31 (the number of days in January) gives an average number of infant deaths per day in January 2000 of 76.0. This is slightly higher than the daily average for the year of 75.5. This removes the influence that the number of days in the month has on the number of deaths in that month because it is divided by the number of days in that month. Figure 4.5 below shows the results of these calculations for each year.



Although it is difficult in Figure 4.5 to see the actual values of each of the months, the purpose of displaying all the years to together is to show that there is no consistent month in the four years that always displays values that are higher or lower than the average number of deaths per day for that year. If there was a constant month that showed a similar difference from the yearly difference for each of the years, then we might need to admit that perhaps this is due to increases in the overall number of births, and not to actual differences in the number of deaths to infants. From Figure 4.5 we can see that the months of the year that have the largest deviations from the average are very different from year to year. For example, in 2001 there is a high value of 6.0 in the month of February. This means that in February in 2001 there were six more infant deaths per day on average than the average number of infant deaths for all of 2001. In

2001 the average number of deaths per day was 74.7 deaths, and in February there were 2,260 infant deaths, or 80.7 deaths per day on average in the month of February for the year 2001. This is not only the highest average for any one month for the year 2001, but is the highest for all five years of data. Interestingly, 2001 is the only year in which such a large value is found for February; in 2004 there were 2.9 more infant deaths in February than the yearly daily average, but this value is still much lower than that found in 2001. Although there seems to be some consistency in the peaks and troughs in the months, overall the results shown in Figure 4.5 indicate that there is no one month where the number of infant deaths is always higher than the average number of infant deaths per day for that year.

I believe that the descriptive results described in this section of my dissertation provide a rather systematic examination of some of the issues that may arise with using the Linked Birth/Infant Death data in my analyses. Although there are still some assumptions that cannot be examined that could well confound my results, by performing the above descriptive investigations I have addressed some of the possible causes for concern in my analyses. In the rest of this chapter and in subsequent chapters, I will address similar issues regarding my data analyses. In the next section of this chapter I discuss the hypotheses and operationalization of the variables that I will use in my models.

Hypotheses and Operationalization

As already discussed in chapter III, the main variables of interest in this dissertation pertain to month of birth and month of gestation. The analyses in this chapter use month

of birth as the principal independent variable of interest in examining cause specific infant death. The next chapter of this dissertation uses the alternative measure, month of gestation, to determine which of these two key independent variables has the greatest impact on cause specific infant mortality. The most appropriate variable will then be used in the multilevel analyses undertaken in chapter VI.

The operationalization of the month of birth variable as well as the dependent variable of cause specific infant mortality and the control variables have already been described in chapter III. One consideration that must be made is that although there is a standard form of the U.S. live birth certificate, not all states adhere completely to the standard form. This means that some states may exclude some of the questions on their birth certificate, hence not producing data for some of the characteristics of the mother and child. For these data and for the variables that I include in the analyses, only California does not ask questions on their certificate for all the variables I decided to use as control variables. Specifically the California birth certificate does not include either the questions about tobacco nor alcohol use during pregnancy nor the question for the mother about how much weight was gained during the pregnancy. Thus, when I include these variables in the analyses, I end up excluding all cases from the state of California.

As previously stated the main objective of the analyses is to determine whether or not the control variables can account for any part of the observed differences between month of birth and cause specific infant death; therefore these control variables are very important. It is also important for me to undertake an analysis using cases from all the states of the United States; especially since the multilevel analyses to be performed in

chapter VI will use the state as the second level of analysis. To determine the best possible solution I have estimated all the models of this chapter and of chapter V two times—once using all the control variables and therefore excluding cases from the state of California, and a second in which I do not use the control variables of alcohol use, tobacco use and mother's weight gain, therefore including cases from the state of California. As we will see in the results of this chapter and later in chapter V there were no drastic differences in the results of the two sets of models. This will be discussed at the end of this chapter in terms of the decision regarding which type of model to be used in the multilevel analyses.

The independent variable measuring month of birth is actually a series of dummy variables for each of the months of the year. In order to include these variables in the models I must exclude one of the months of birth in order to create a reference variable, and in so doing to avoid perfect multicollinearity in the resulting model. In order to exclude the most appropriate variable I took the average number of births for each of the twelve months for all four years. I did this by dividing the number of births in each month by the number of days in that month. Since the number of days varies by year for February, I divided the total number of births for February by 28.4 because both 2000 and 2004 were leap years [(28*3) + (29*2) = 142/5=28.4). I show these averages in Table 4.4 below.

Table 4.4 Average Number of Births by Month of Birth, U.S. 2000 to 2004	
Month of Birth	Average Number of Births
January	367.8
February	365.0
March	369.1
April	375.7
May	376.2
June	386.6
July	391.4
August	387.6
September	387.3
October	379.6
November	370.0
December	373.5
Average	377.5

Adding all the monthly averages and dividing by 12 calculated the average number of births per month for all months, giving an average number of births of 377.5. The month that is closest to this average number of births is May, which has 376.2 births on average for all four years. I therefore use May as the reference month, and exclude its dummy from the models. It has the number of births per month that is closest to the average number of births per month for all years. I now discuss the hypotheses that will be tested using the multinomial logistic regressions that I will estimate in this chapter that focus on the month of birth variable.

In the literature the connection between month of birth and adult mortality points to births occurring in the late spring to the early summer months of April, May and June as being more detrimental to the individual's later life survival than the fall and early winter months of October, November and December. The literature on month of birth

and diagnosis with schizophrenia as an adult finds that there is an excess of individuals born in the months of November and December. As discussed previously in chapter II, the hypothesis in the literature on schizophrenia points to “infectious agents” which produce infections in the central nervous system of the fetus and are thought to be the source of the neurological disorder (Torrey et al 1993). Torrey and colleagues (1993) also noted numerous studies linking maternal contraction of the influenza virus during the fifth or sixth month of pregnancy and later life diagnosis with schizophrenia for those individuals. However, these findings on influenza’s connection to schizophrenia are controversial (see Crow and Done 1992; Brown et al. 2004; Mednick et al. 1994).

In the United States, the peak of the influenza virus is found in the month of February for the years 2000 to 2004 (http://www.cdc.gov/flu/weekly/fluactivity.htm). If a mother is infected with the flu, then we may also be able to suggest that the early spring is detrimental to an infant’s survival.

Based on this previous literature, I hypothesize that infants born in the fall and winter months will have a lower risk of cause specific mortality than those born in the spring and summer months. Namely, the months of October, November, December, January and February should be associated with decreased odds of dying of a specified caused. In contrast, the months of March, April, June, July, August and September are expected to be associated with increased odds of dying of a specified cause. These expectations are based on the fact that infants born in the spring and summer months with be *in utero* during the fall and winter months that are often associated with poorer nutrition, more severe weather and increases to the chances of maternal contraction of

infectious diseases, a conclusion of past studies about the connection between month of birth and mortality.

Before testing these hypotheses, I turn first to a description of the data that will be included in the models.

Descriptive Results

In this section of the chapter I provide descriptive results of the independent and dependent variables. The first point to be mentioned is that the number of infant deaths to be analyzed is 92,021 instead of 137,951 as was shown in the tables at the beginning of this chapter. That occurs because almost 46 thousand infant deaths have missing values on one of more of the infant or mother characteristics. In order to maintain a consistent number of cases, I have restricted my analyses in all three models to those deaths with no missing values. Table 4.5 (below) shows summary information on the dependent variable “cause of infant death” and the independent dummy variables of month of birth.

Table 4.5 Descriptive Statistics of Cause of Infant Death and Month of Birth Variables (N=92,021)

Dependent Variable	Minimum	Maximum	Mean	Standard Deviation
Cause of Infant Death				
11 categories (0-10)	0	10	3.05	2.49
Month of Birth 12 dummy variables 0=not born in month 1=born in month				
January	0	1	0.083	0.275
February	0	1	0.075	0.263
March	0	1	0.082	0.274
April	0	1	0.081	0.274
May	0	1	0.084	0.277
June	0	1	0.082	0.275
July	0	1	0.087	0.281
August	0	1	0.087	0.282
September	0	1	0.083	0.276
October	0	1	0.085	0.279
November	0	1	0.081	0.274
December	0	1	0.085	0.280

Since each of the variables measuring the infant's month of birth is a dummy variable, their minimum values are zero and maximum values are one. The month of birth with the lowest mean value is February with a value of 0.075; this is likely due to the fact that February has the fewest number of days of any month. The dependent cause of death variable is an eleven category nominal variable. Its maximum value is ten and its minimum value is zero. The mean value is 3.05, which is the category "perinatal infections" although this is not the most frequently reported cause of infant death in the dataset. I next present in Table 4.6 similar descriptive statistics for the independent variables to be used as controls.

Table 4.6 Descriptive Statistics of Independent Control Variables

Independent Variables	Minimum	Maximum	Mean	Standard Deviation	N
Infant Characteristics					
Male (1=male, 0=female)	0	1	0.514	0.499	92021
Birthweight (in grams)	227	6521	1750.6	1246.114	92021
Gestational Age (in weeks)	17	47	30.863	7.577	92021
Plurality					
Single Birth (1=yes, 0=no)	0	1	0.852	0.354	92021
Twins (1=yes, 0=no)	0	1	0.129	0.335	92021
Triplets plus (1=yes, 0=no)	0	1	0.018	0.134	92021
Maternal Characteristics					
Age of Mother					
Teen Mom (1=yes, 0=no)	0	1	0.161	0.368	92021
Twenties (1=yes, 0=no)	0	1	0.526	0.499	92021
Thirties (1=yes, 0=no)	0	1	0.283	0.450	92021
Forties plus (1=yes, 0=no)	0	1	0.028	0.165	92021
Number of Prenatal Care Visits					
Nine or less (1=yes, 0=no)	0	1	0.560	0.496	92021
Ten to Twenty (1=yes, 0=no)	0	1	0.424	0.494	92021
Twenty-One plus (1=yes, 0=no)	0	1	0.034	0.182	92021
Mother's Race/Ethnicity					
White (1=yes, 0=no)	0	1	0.522	0.499	92021
Black (1=yes, 0=no)	0	1	0.297	0.457	92021
Hispanic (1=yes, 0=no)	0	1	0.139	0.346	92021
Other (1=yes, 0=no)	0	1	0.039	0.195	92021
Mother's Education					
Less than HS (1=yes, 0=no)	0	1	0.286	0.443	92021
High School (1=yes, 0=no)	0	1	0.358	0.479	92021
Some College (1=yes, 0=no)	0	1	0.202	0.402	92021
Four plus years of College (1=yes, 0=no)	0	1	0.169	0.375	92021
Mother's Marital Status					
Married (1=married, 0=non-married)	0	1	0.516	0.499	92021
Mother's Tobacco Use					
Tobacco (1=used tobacco while pregnant, 0=did not)	0	1	0.183	0.387	92021
Mother's Alcohol Use					
Alcohol (1=used alcohol while pregnant, 0=did not)	0	1	0.012	0.113	92021
Mother's Weight Gain					
Weight Gain (In pounds)	0	98	22.427	14.838	92021

Table 4.6 shows the descriptive results for the independent variables that are included as controls in the analyses. Many of the variables—single birth, twins, triplets plus, male, teen mom, twenties, thirties, forty plus, nine or less prenatal visits, ten to

twenty prenatal visits, twenty-one and more prenatal visits, white, black, Hispanic, other, less than high school, high school, some college, college and higher, married, alcohol, and tobacco—are included as dummy variables. As such, they have minimum values of zero and maximum values of one. When we compare the mean values for each of the sets of dummy variables, we find that the most common characteristics in these data are: male, single birth, twenty year old mother, nine or less prenatal visits, white, high school education, married, non smokers, non drinkers.

The variables birthweight, gestational age and weight gain are included in the analyses as continuous variables. Birthweight ranges from a low of 227 grams to a high of 6,521 grams, with a mean value of 1,750.6 grams. Gestational age ranges from a low of 17 weeks to a high of 47 weeks with a mean value of 30.863 weeks. The mother's weight gain ranges from a low of zero pounds to a high of 98 pounds, with a mean value of 22.427 pounds. All missing values have been dropped from the dataset used for Tables 4.5 and 4.6. Below in Table 4.7 is the distribution of the causes of infant deaths for the nominal dependent variable. As we can see the most common cause of infant death is the category of prematurity and related conditions, followed by congenital anomalies.

Table 4.7 Causes of Infant Death		
External Causes	5319	5.78
Prematurity and Related Conditions	26020	28.28
Congenital Anomaly	18712	20.33
SIDS, Other Unexplained	12801	13.91
Pregnancy Complications	8771	9.53
Birth Asphyxia and Trauma	5228	5.68
Perinatal Infections	3217	3.50
Other Infections	4587	4.98
Endocrine, Metabolic and Digestive System Conditions	3397	3.69
Neoplasms and Blood Conditions	612	0.66
Respiratory, Circulatory and Nervous System Conditions	3359	3.65
Total	92,021	100.00

Multicollinearity Diagnostic

I performed a series of diagnostics for each of the individual-level models that I will be estimating with the independent variable measuring month of birth. In the paragraphs that follow, I discuss the diagnostics, whether there is cause for concern, and the ways that any issues will be dealt with in my models.

First I examined the zero-ordered correlations for all my independent variables. These correlations showed that there may be issues with multicollinearity between the variable measuring birthweight and the variable measuring gestational age. This is not unexpected and as such I will include these variables in separate models. The other variables that seem like they may pose issues for estimation are those measuring age, prenatal care, education, plurality and race/ethnicity. When included in the models I will have to exclude at least one of the dummy variables from each of these categories of infant and maternal characteristics. I will therefore exclude the following variables: “pre10_20,” where a value of one indicates that the woman had between ten and twenty prenatal visits; “white,” where a value of one indicates that the mother is non-Hispanic

white; “HS,” where a value of one indicates that the mother had a high school education; “twenties,” where a value of one indicates that the mother is twenty to twenty-nine years old; “singlebirth,” where a value of one indicates that the child was a single birth. I chose these categories to use as reference categories because they were the most frequently reported of each of the categories measuring that specific infant or maternal characteristic, except for the variable measuring the number of prenatal visits. The most frequently reported category of the prenatal measures is less than 10 visits. However, since ten to twenty visits is the recommended number of visits, I decided to use it as the reference because I believe that the ‘less than ten visits’ and the ‘more than twenty visits’ variables would both be associated with negative outcomes for the infant. This expectation owes to the fact that less than ten visits is less than the recommended number of prenatal visits, and more than twenty visits may be more common for mothers who are having at-risk pregnancies. By leaving out the dummy variable reflecting ten to twenty visits, I will be including in the regressions both of the dummy variables that I expect to have a negative association with infant deaths.

I also calculated the tolerance values for each of the independent variables in order to further assess multicollinearity in my models. As suspected from the zero-ordered correlations, including both gestational age and birthweight in the same model poses problems with collinearity. After including only one of these variables at a time with all other independent variables, the tolerances were all above 0.53. Thus, estimating a model with all variables, and either birthweight or gestational age, should not create major problems of multicollinearity. I now turn to a discussion of the results of my

multinomial logistic regression models. I will display the regression coefficients as odds ratios.

Multinomial Logistic Regression Results (Using Birthweight but Not Gestational Age)

In this section I report the results of the multinomial logistic regression using the independent variable of month of birth. The results are shown in four separate tables; all the results are from just three models. As discussed above, the variables birthweight and gestational age were found to be too collinear to be included in the same analyses; therefore the tables below only include the results using the birthweight variable as a control and omit the gestational age variable. A second series of models have been estimated with the identical variables but switching out the variable birthweight for the variables gestational age. These results are shown in a later section.

The results from the three models shown in Table 4.8 (in four parts) include relative risk ratios. As discussed in chapter III, relative risk ratio (rrr's) are the exponentiated values of the multinomial logistic regression coefficients; these enable the multiplicative interpretation of the odds of being in a cause of death category as opposed to the reference category. Although most months of birth were not associated with increasing the odds of dying of a specific cause of death, some months were shown to be associated with an increased or decreased risk of dying from a specific cause. I now discuss the results.

Table 4.8 Multinomial Logistic Regression Results (Odds Ratios) Using Birthweight: Cause Specific Infant Mortality by Month of Birth, Infant and Maternal Characteristics, U.S. 2000-2004

Month of Birth	Prematurity and Related Conditions			Congenital Anomalies			SIDS		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
January	1.065	1.141	1.195*	1.065	1.081	1.138	1.139	1.138	1.142
February	0.947	1.001	1.031	0.967	0.983	1.016	1.034	1.034	1.036
March	0.994	1.035	1.052	0.962	0.977	0.997	0.937	0.938	0.940
April	1.143	1.132	1.169	1.060	1.065	1.098	1.008	1.010	1.013
May	ref	ref	ref	ref	ref	ref	ref	ref	ref
June	1.037	1.036	1.044	1.034	1.016	1.021	1.126	1.121	1.123
July	0.980	0.994	1.015	1.000	0.996	1.026	1.035	1.032	1.037
August	0.914	0.978	0.996	0.953	0.968	1.001	1.097	1.097	1.097
September	0.865*	0.974	0.992	0.921	0.962	0.996	1.054	1.060	1.060
October	1.045	1.050	1.084	1.004	1.003	1.047	1.122	1.120	1.122
November	0.889	0.968	1.000	0.961	0.977	1.016	1.044	1.046	1.050
December	0.880	0.952	0.982	0.985	1.002	1.037	1.079	1.078	1.079
Infant Characteristics									
Male		1.016	1.021		0.989	1.002		1.020	1.020
Birthweight		0.997***	0.997***		0.999***	0.998***		0.999***	0.999***
Single Birth		ref	ref		ref	ref		ref	ref
Twins		0.977	0.841*		0.684***	0.460***		1.222**	1.205*
Triplets Plus		2.322*	1.034		0.930	0.287***		1.556	1.485
Maternal Characteristics									
Teens			0.814***			0.735***			0.936
Twenties			ref			ref			ref
Thirties			1.568***			1.761***			1.007
Forty and over			1.617***			3.343***			0.837
Prenatal 9 visits			1.351***			0.898**			1.005
Prenatal 10-20 visits			ref			ref			ref
Prenatal 20+ visits			1.213			1.849***			0.968
White			ref			ref			ref
Black			0.830***			0.541***			1.066
Hispanic			1.277			1.343***			0.934
Other			0.941			0.824*			0.908
Less than HS			0.789***			0.829***			0.966
High School			ref			ref			ref
Some College			1.171**			1.121*			0.997
4 plus yrs College			1.821***			1.724***			1.132
Married			1.436***			1.723***			1.108**
Alcohol			0.894			0.844			0.960
Tobacco			0.439***			0.367***			1.188***
Weight Gain			0.991***			1.004***			0.999
N	92021	92021	92021						
LRχ2	246.24	54564.24	63170.51						
df	110	150	300						
prob>χ2	0.0000	0.0000	0.0000						
Pseudo R2	0.007	0.1446	0.1674						

* p<0.05, **p<0.01, ***p<0.001 (Base Outcome=External Causes)

Table 4.8 (continued) Multinomial Logistic Regression Results (Odds Ratios) Using Birthweight: Cause Specific Infant Mortality by Month of Birth, Infant and Maternal Characteristics, U.S. 2000-2004

Month of Birth	Pregnancy Complications			Birth Asphyxia and Birth Trauma			Perinatal Infections		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
January	0.970	1.040	1.085	1.020	1.061	1.106	1.009	1.057	1.108
February	0.844*	0.891	0.915	0.883	0.916	0.941	0.838	0.873	0.900
March	0.913	0.949	0.956	0.877	0.906	0.918	0.982	1.016	1.038
April	1.105	1.090	1.125	1.059	1.052	1.083	0.969	0.963	0.994
May	ref	ref	ref	ref	ref	ref	ref	ref	ref
June	0.861	0.862	0.866	1.008	0.995	1.000	1.152	1.137	1.149
July	0.858	0.871	0.890	0.931	0.932	0.954	1.047	1.052	1.078
August	0.838*	0.898	0.911	0.841	0.878	0.895	0.926	0.971	0.992
September	0.828*	0.935	0.953	0.880	0.963	0.985	0.976	1.074	1.096
October	0.892	0.895	0.922	0.903	0.902	0.933	1.018	1.019	1.056
November	0.861	0.939	0.968	0.836	0.882	0.909	0.912	0.969	0.999
December	0.803**	0.870	0.890	0.896	0.944	0.971	0.894	0.946	0.977
Infant Characteristics									
Male		1.048	1.051		1.103*	1.110		1.011	1.020
Birthweight		0.997***	0.997***		0.998**	0.998***		0.998***	0.998***
Single Birth		ref	ref		ref	ref		ref	ref
Twins		1.171*	1.014		1.227*	1.004		0.972	0.796*
Triplets Plus		2.826**	1.567		2.361*	1.302		2.302*	1.165
Maternal Characteristics									
Teens			0.658***			0.827***			0.924
Twenties			ref			ref			ref
Thirties			1.612***			1.617***			1.535***
Forty and over			1.667***			1.842***			1.802***
Prenatal 9 visits			1.499***			1.264***			1.174**
Prenatal 10-20 visits			ref			ref			ref
Prenatal 20+ visits			1.202			1.042			1.184
White			ref			ref			ref
Black			0.679***			0.708***			0.880*
Hispanic			1.178*			1.189*			1.471***
Other			0.793*			0.763*			0.807
Less than HS			0.723***			0.788***			0.734***
High School			ref			ref			ref
Some College			1.162**			1.187**			1.263***
4 plus yrs College			1.895***			1.842***			1.830***
Married			1.541***			1.508***			1.452***
Alcohol			0.967			0.877			0.798
Tobacco			0.487***			0.418***			0.434***
Weight Gain			0.983***			0.994***			0.998
N	92021	92021	92021						
LRχ2	246.24	54564.24	63170.51						
df	110	150	300						
prob>χ2	0.0000	0.0000	0.0000						
Pseudo R2	0.007	0.1446	0.1674						

* p<0.05, **p<0.01, ***p<0.001 (Base Outcome=External Causes)

Table 4.8 (continued) Multinomial Logistic Regression Results (Odds Ratios) Using Birthweight: Cause Specific Infant Mortality by Month of Birth, Infant and Maternal Characteristics, U.S. 2000-2004

Month of Birth	Other Infections			Endocrine, Metabolic and Digestive System Disorders			Neoplasms and Blood Conditions		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
January	1.224*	1.24*	1.284*	1.055	1.082	1.131	1.191	1.193	1.217
February	1.059	1.076	1.102	0.953	0.978	1.007	1.153	1.155	1.202
March	0.966	0.979	0.991	0.954	0.977	0.996	1.361	1.364	1.404
April	0.954	0.958	0.975	1.217	1.127	1.250*	1.073	1.076	1.117
May	ref	ref	ref	ref	ref	ref	ref	ref	ref
June	1.120	1.102	1.109	0.935	0.919	0.929	1.228	1.220	1.234
July	1.132	1.126	1.138	0.973	0.970	0.991	0.948	0.947	0.985
August	1.114	1.131	1.145	0.931	0.957	0.977	1.152	1.153	1.185
September	1.145	0.197	1.210	0.957	1.023	1.046	0.931	0.938	0.969
October	1.257*	1.254*	1.280*	1.014	1.012	1.049	1.279	1.278	1.348
November	1.236*	1.259*	1.284*	0.889	0.920	0.948	1.160	1.158	1.221
December	1.051	1.068	1.091	0.997	1.030	1.061	1.035	1.036	1.086
Infant Characteristics									
Male		0.973	0.980		1.055	1.066		1.012	1.024
Birthweight		0.999***	0.999***		0.998***	0.998***		0.999***	0.999***
Single Birth		ref	ref		ref	ref		ref	ref
Twins		1.057	0.915		0.983	0.771**		0.596*	0.414***
Triplets Plus		1.894	1.195		2.261*	1.070		1.632	0.543
Maternal Characteristics									
Teens			0.791***			0.837**			0.811
Twenties			ref			ref			ref
Thirties			1.329***			1.548***			1.594***
Forty and over			1.611**			1.727**			1.151
Prenatal 9 visits			1.050			1.075			1.032
Prenatal 10-20 visits			ref			ref			ref
Prenatal 20+ visits			1.370**			1.135			1.129
White			ref			ref			ref
Black			1.036			0.927			0.540***
Hispanic			1.392***			1.471***			1.221
Other			1.229*			1.019			1.300
Less than HS			0.902*			0.892*			0.791
High School			ref			ref			ref
Some College			1.074			1.194**			1.165
4 plus yrs College			1.319***			1.760***			2.198***
Married			1.248***			1.493***			1.929***
Alcohol			0.986			0.652			1.009
Tobacco			0.624***			0.451***			0.461***
Weight Gain			1.000			1.004**			1.002
N	92021	92021	92021						
LRχ2	246.24	54564.24	63170.51						
df	110	150	300						
prob>χ2	0.0000	0.0000	0.0000						
Pseudo R2	0.007	0.1446	0.1674						

* p<0.05, **p<0.01, ***p<0.001 (Base Outcome=External Causes)

Table 4.8 (continued) Multinomial Logistic Regression Results (Odds Ratios) Using Birthweight: Cause Specific Infant Mortality by Month of Birth, Infant and Maternal Characteristics, U.S. 2000-2004

Month of Birth	Respiratory, Circulatory and Nervous System Disorders		
	Model 1	Model 2	Model 3
January	1.047	1.052	1.105
February	0.769*	0.774*	0.797*
March	0.726**	0.733*	0.746*
April	0.903	0.907	0.932
May	ref	ref	ref
June	0.961	0.948	0.958
July	0.956	0.951	0.975
August	0.852	0.858	0.878
September	0.827	0.850	0.872
October	0.883	0.881	0.915
November	0.875	0.881	0.916
December	0.766*	0.771	0.796
Infant Characteristics			
Male		1.020	1.030
Birthweight		0.999***	0.999***
Single Birth		ref	ref
Twins		0.907	0.673***
Triplets Plus		2.189	0.905
Maternal Characteristics			
Teens			0.742***
Twenties			ref
Thirties			1.647***
Forty and over			1.988***
Prenatal 9 visits			0.976
Prenatal 10-20 visits			ref
Prenatal 20+ visits			1.183
White			ref
Black			0.816***
Hispanic			1.155**
Other			1.122
Less than HS			0.834**
High School			ref
Some College			1.147*
4 plus yrs College			1.842***
Married			1.496***
Alcohol			0.860
Tobacco			0.464***
Weight Gain			1.002
N	92212	92212	92212
LRχ2	245.61	54456.27	63092.63
df	110	150	300
prob>χ2	0.0000	0.0000	0.0000
Pseudo R2	0.006	0.1441	0.1669

* p<0.05, **p<0.01, ***p<0.001 (Base Outcome=External)

First, we see that the causes of death SIDS, congenital anomalies, birth asphyxia and birth trauma, neoplasms and blood conditions and perinatal infections were not associated in any of the three models with an increased risk of dying from one of the specified causes during any of the months, compared to the base outcome of external causes of death. This was contrary to my hypotheses that stated that all causes of death would be associated with an increased risk of dying during the months of the late spring and early summer.

In the first part of Table 4.8 we see the rrr's for the cause of death "prematurity and related causes"; in model one the data indicate that there is an increased risk of dying of that cause for those born in the month of September. However, the September month of birth loses its significance once the models incorporate the characteristics of the infant in model two and of both the infant and the mother in model three. This indicates that the negative association in model one with the month of September and the risk of dying of prematurity and related conditions is accounted for when the variables that measure characteristics of the infant and the infant and the mother are included.

Of particular interest with respect to the causes of death due to prematurity and related causes is the third model that includes all of the control variables; here the month of January gains significance at the $p<0.05$ level. The odds ratio of 1.195 means that infants born in January, compared to those born May (the reference month), have a risk of dying of prematurity and related causes that is multiplied by 1.195, compared to dying of external causes (the base outcome). One may also interpret this coefficient in terms of the percent change in the odds ratio, which is calculated as the value of the odds ratio,

minus one, multiplied by one-hundred, or (1.195 – 1.00)*100= 19.5%. So the odds of infants dying in January are 19.5 percent higher compared to dying in May of prematurity or related causes, compared to external causes. The most interesting point about this finding is that in models one and two the month of January was not associated with a significant increased or decreased risk of dying; it only gained significance once the infant and maternal characteristic control variables were included in the model.

The cause of death of "pregnancy complications" produced a similar result as the "prematurity" cause of death. In the first model which included only the month of birth variables, the months of February, August, and December are found to be negatively and significantly associated with an increased risk of dying from "pregnancy complications" compared to external causes of death. Again, however, these months of birth lost their significance in models one and two when the infant and maternal characteristics were added into the models. This suggests that any negative relationship that an infant born in the months of February, August or December may have of dying from pregnancy complications as compared to external causes of death may be explained by various characteristics of the infant and/or of the mother.

Again, similar to the findings in the category of prematurity and related causes, endocrine, metabolic and digestive system disorders had one significant risk ratio in the third model. The month of April is found to be positively and significantly associated with an increased risk of death due to this cause as compared to external causes of death. This means that infants born in April versus the reference month of May are 25 percent more likely to die of endocrine, metabolic and digestive system disorders than from

external causes of death. This month of birth is significantly associated with an increased risk of dying after controlling for variables tapping the characteristics of the infant and mother.

Both the causes of death of “other infections” and “respiratory, circulatory and nervous system disorders” had a few months of birth that yielded significant associations with some of the month of birth variables in all three models. Specifically, the other infections cause of death shows a positive and significant relationship with the birth months of January, October and November in all three models. This means that the increased risk of dying of other infections in January, October or November is not explained by any of the infant or maternal characteristics. Thus being born in the months of January, October or November puts an infant at an increased risk of dying of other infections as compared to their risk of dying from external causes of death. Specifically, the positive association in the month of January indicates that infants born in the month of January, as compared to those infants born in May, are at a 28.4 percent greater risk of dying of other infections as compared to all external causes of death. Almost identical increased risks were found for the months of birth of October and November.

Similarly, the association of the months of birth and the cause of death respiratory causes yielded significant findings. For the months of birth of February and December these negative and significant findings were in the hypothesized direction. However, the negative and significant association with the month of birth of March was opposite that the expected association.

These significant findings associated with the months of birth of October, November and January with the other infections cause of death are in the opposite direction of the hypotheses, which expected negative associations with these months reflecting lower risks of dying of the specified causes. Also, the negative effect of the month of birth March with the cause of death respiratory causes indicates better survival for those infants who are born in March as compared to May of dying from respiratory causes. This may indicate that the hypotheses that are used in this dissertation that were derived from the literature on month of birth and adult longevity are not fully applicable to the study of infants. It could also be possible that the monthly differences in the cause of infant death categories are dictating the direction of the association that may be found. This would mean that although for most causes of death those infants born in March should expected increased risks of dying, the small number of infant deaths due to respiratory causes in March may well be overshadowing this relationship. In chapter V I will examine whether or not these findings that were opposite of those hypothesized are due to the fact that the time of interest is the infant's fetal development and not their month of birth.

The next cause of death, "respiratory, circulatory and nervous system disorders," produced similar findings as the "other infections" category in all three models. The months of February, March and December were all found to be significantly associated with dying of respiratory, circulatory and nervous system disorders as compared to external causes of death. However, in this case the association was indeed negative, not positive, as was the case with the "other infections" category. For example, being born

in, the month of birth of February, compared to May, is associated with a 20.3 percent decreased risk of dying of respiratory, circulatory and nervous system disorders compared to external causes of death. The months of birth supported my hypothesis because they were found to be negatively associated with the 'other infections' cause of death.

Results of Model Including Cases from California (with Birthweight)

I then predicted the full model including all infant and mother variables as controls except the variables alcohol, tobacco and weight gain. As discussed in an earlier section of this chapter the exclusion of these variables means that cases from the state of California would now be included in these models. Including these variables in the previous models excludes cases from California, as in Table 4.8.

In the full model that includes the variable birthweight (and not gestational age) there were only slight differences in the associations of the month of birth variables. First the causes of death congenital anomalies, pregnancy complications, birth asphyxia and birth trauma, perinatal infections, endocrine causes and neoplasms and blood conditions are not significantly associated with any of the month of birth variables in the model that included cases from the state of California. These causes of death were also not significantly associated with any of the month of birth variables in the models that excluded cases from California as seen in Table 4.8. The significant association of the month of birth January and the cause of death prematurity was observed in both the model that excluded California and the model that included California.

However, there were some changes to the significant findings that are shown in Table 4.8. First, the cause of death SIDS is now significantly associated with the month of birth January when the alcohol, tobacco and weight gain variables are excluded. Similarly, the cause of death of other infections is now significantly associated with the month of birth September in the model that includes cases from California. Lastly, the cause of death respiratory causes is significantly associated with the months of birth of March and December. The month of birth March was significant in the model that excluded California, but the month of December is only significant in the models which include California.

Although there are some changes in the models that include cases from California, overall the significant associations were not drastically different from those models which excluded cases from California. Therefore I believe it is not detrimental to my models to use the control variables alcohol and tobacco use and weight gain in the level one models of chapters IV and V and exclude those controls to include California in the multilevel models of chapter VI. Nest, in order to further examine the relationship between month of birth and cause specific infant mortality, I will next turn to a discussion of the predicted probabilities for the full model as shown in Table 4.8.

Predicted Probabilities

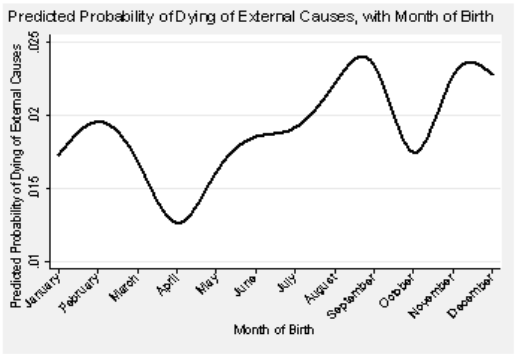
As stated in chapter III, an additional way to interpret the results of a multinomial logistic regression model is via predicted probabilities. Each value shows the predicted probability of being in a specific category of the dependent variable according to the full model with all independent variables. I obtained the predicted probabilities based on

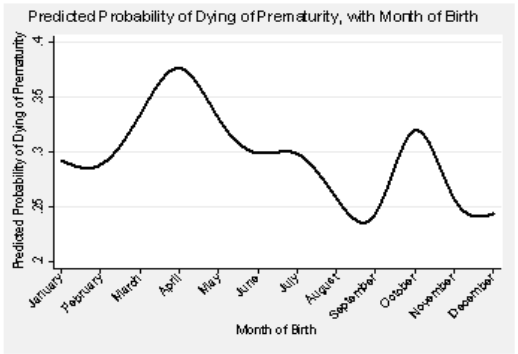
model three (with the birthweight variable) and then graphed these values by the infant's month of birth. Figure 4.6 presents a series of predicted probability values for each of the eleven causes of death by the infant's month of birth. A similar series of graphs is shown later for the final model using the gestational age variable instead of the birthweight variable.

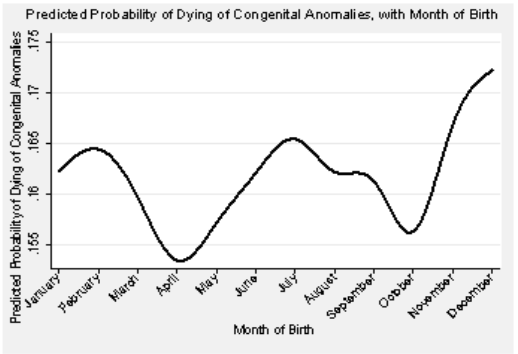
If we looked at each of the values for the predicted probabilities for a specific case, we would see that the values add to one for each infant. This is a way to think about what these predicted probabilities are demonstrating—the prediction based on the independent variables in the model for an infant being in each of the eleven categories of the dependent variable cause of death. From Figure 4.6 we can see that the scale of the predicted probabilities changes for each of the eleven graphs. This shows that the overall chances of dying from some of the causes are low for all months of birth. For example, the predicted probability of dying of external causes only ranges from 0.01 to 0.025. The predicted probability that an infant in this dataset will die of external causes of death is very low, especially when we compare this value to the probability of dying from prematurity, which ranges from 0.2 to 0.4.

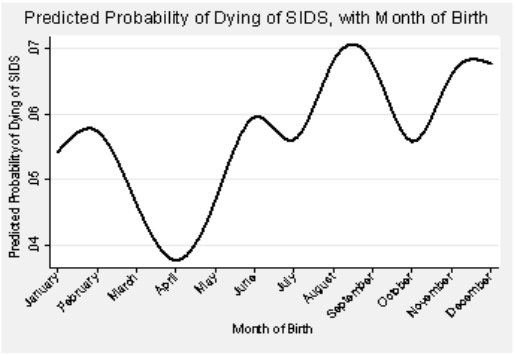
We can see that there is much variation in the predicted probabilities of each of the months of birth in the graphs of Figure 4.6. For example if we look at the month of January we see that the probability of dying of external causes is only about 0.017 but the probability of dying of prematurity is almost 0.3. Looking only at the probability of dying of prematurity, we see that the value is a low of 0.24 in the months of August to September and a high of about 0.29 in January to February.

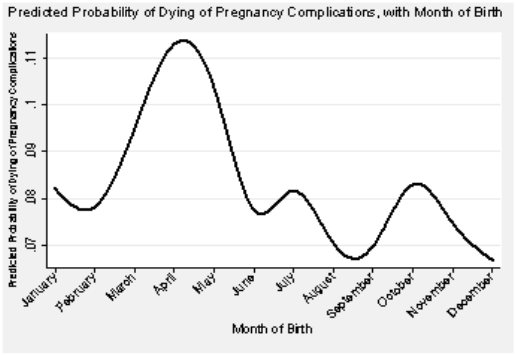
Figure 4.6 Predicted Probabilities: Cause of Infant Death by Month of Birth (Birthweight), U.S. 2000-2004











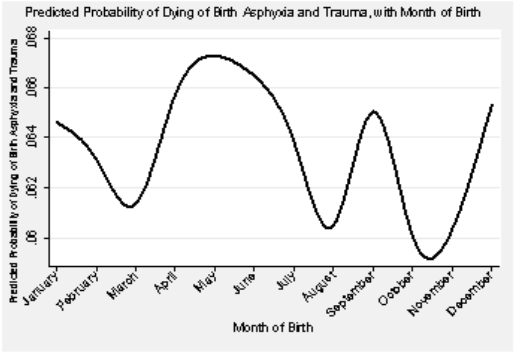
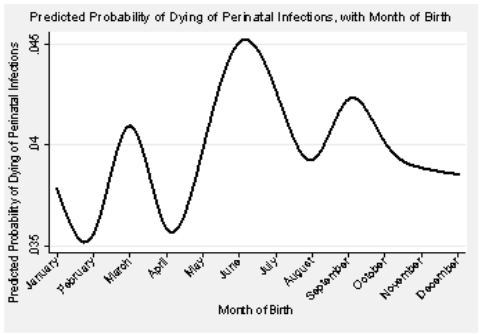
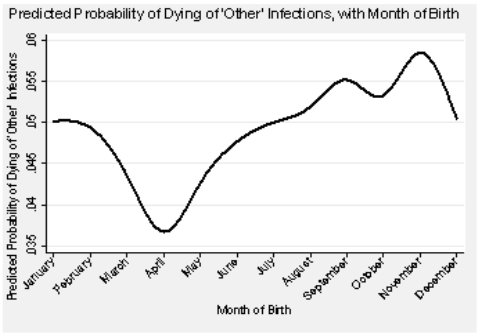
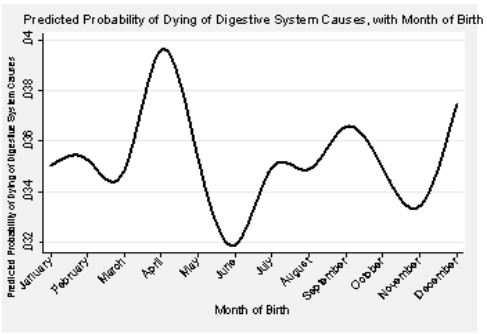
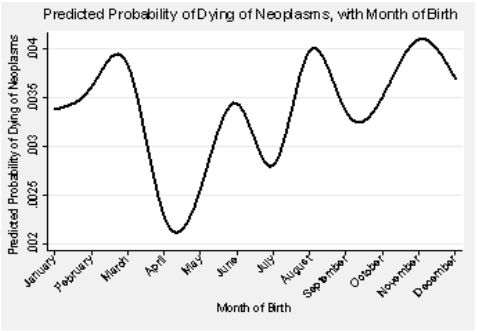


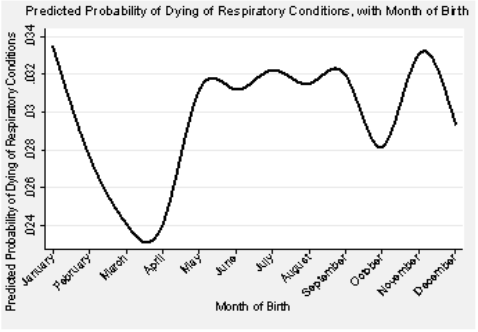
Figure 4.6 (Continued) Predicted Probabilities: Cause of Infant Death by Month of Birth (Birthweight), U.S. 2000-2004











Multinomial Logistic Regression Results (Using Gestational Age but Not Birthweight)

As previously discussed, the variables measuring the infant's birthweight and its gestational age could not be estimated in the same model because of multicollinearity. Accordingly, in the previous section I estimated the first three models with the variable measuring birthweight, the month of birth variables and all independent control variables except gestational age. The results were shown above in Table 4.8. I next estimated the same models, but omitted the birthweight variable and included the gestational age variable. The results of these models are shown below in Table 4.9. As we can see from Table 4.9 few of the cause specific infant death variables were significantly related to any of the month of birth variables. The results shown in Table 4.9 are also similar to the results found in Table 4.8, the identical model using birthweight. As we see below, few of the variables are significantly related to the cause specific infant death dependent variable.

.

Table 4.9 Multinomial Logistic Regression Results (Odds Ratios) Using Gestational Age: Cause Specific Infant Mortality by Month of Birth, Infant and Maternal Characteristics, U.S. 2000-2004

Month of Birth	Prematurity and Related Conditions			Congenital Anomalies			SIDS		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
January	1.065	1.176	1.223*	1.065	1.085	1.137	1.139	1.140	1.144
February	0.947	0.999	1.024	0.967	0.977	1.006	1.034	1.034	1.035
March	0.994	1.047	1.061	0.962	0.979	0.993	0.937	0.939	0.942
April	1.143	1.157	1.190*	1.060	1.070	1.104	1.008	1.010	1.014
May	ref	ref	ref	ref	ref	ref	ref	ref	ref
June	1.037	1.056	1.061	1.034	1.031	1.035	1.126	1.125	1.126
July	0.980	1.007	1.026	1.000	1.001	1.028	1.035	1.034	1.039
August	0.914	0.993	1.003	0.953	0.969	0.990	1.097	1.098	1.098
September	0.865*	0.967	0.976	0.921	0.949	0.967	1.054	1.059	1.062
October	1.045	1.053	1.078	1.004	0.999	1.032	1.122	1.120	1.122
November	0.889	0.994	1.024	0.961	0.975	1.012	1.044	1.047	1.051
December	0.880	0.956	0.978	0.985	1.002	1.028	1.079	1.080	1.081
Infant Characteristics									
Male		0.970	0.977		0.959	0.968		1.014	1.015
Gestational Age		0.673***	0.681***		0.867***	0.864***		0.973***	0.975***
Single Birth		ref	ref		ref	ref		ref	ref
Twins		1.516**	1.413***		1.015	0.837*		1.276**	1.276**
Triplets Plus		3.736***	2.373*		1.536	0.669		1.652	1.629
Maternal Characteristics									
Teens			0.840***			0.784***			0.943
Twenties			ref			ref			ref
Thirties			1.489***			1.729***			0.998
Forty and over			2.020***			3.867***			0.845
Prenatal 9 visits			0.808			0.476***			1.060
Prenatal 10-20 visits			ref			ref			ref
Prenatal 20+ visits			0.728*			0.480***			1.060
White			ref			ref			ref
Black			0.904*			0.616***			1.080
Hispanic			1.318***			1.372***			0.939
Other			0.955			0.811**			0.905
Less than HS			0.810***			0.822***			0.969
High School			ref			ref			ref
Some College			1.109			1.103*			0.992
4 plus yrs College			1.702***			1.685***			1.124
Married			1.380***			1.648***			1.101*
Alcohol			1.003			0.874			0.965
Tobacco			0.521***			0.419***			1.213***
Weight Gain			0.989***			0.995***			0.999
N	92021	92021	92021						
LRχ2	246.24	59256.14	66321.36						
df	110	150	300						
prob>χ2	0.0000	0.0000	0.0000						
Pseudo R2	0.0007	0.1571	0.1758						

* p<0.05, **p<0.01, ***p<0.001 (Base Outcome=External Causes)

.

Table 4.9 (continued) Multinomial Logistic Regression Results (Odds Ratios) Using Gestational Age: Cause Specific Infant Mortality by Month of Birth, Infant and Maternal Characteristics, U.S. 2000-2004

Month of Birth	Pregnancy Complications			Birth Asphyxia and Birth Trauma			Perinatal Infections		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
January	0.970	1.076	1.112	1.020	1.089	1.128	1.009	1.081	1.126
February	0.844*	0.888	0.907	0.883	0.919	0.939	0.838	0.873	0.896
March	0.913	0.958	0.962	0.877	0.919	0.928	0.982	1.029	1.048
April	1.105	1.116	1.146	1.059	1.068	1.098	0.969	0.979	1.009
May	ref	ref	ref	ref	ref	ref	ref	ref	ref
June	0.861	0.878	0.878	1.008	1.016	1.019	1.152	1.162	1.170
July	0.858	0.883	0.902	0.931	0.945	0.965	1.047	1.066	1.088
August	0.838*	0.913	0.920	0.841	0.892	0.901	0.926	0.983	0.994
September	0.828*	0.928	0.940	0.880	0.959	0.971	0.976	1.065	1.076
October	0.892	0.900	0.918	0.903	0.901	0.924	1.018	1.017	1.045
November	0.861	0.974	0.997	0.836	0.896	0.921	0.912	0.981	1.008
December	0.803**	0.876	0.888	0.896	0.951	0.969	0.894	0.949	0.972
Infant Characteristics									
Male		0.999	1.006		1.059	1.067		0.970	0.978
Gestational Age		0.646***	0.655***		0.745***	0.751***		0.743***	0.750***
Single Birth		ref	ref		ref	ref		ref	ref
Twins		1.814***	1.655***		1.775***	1.600***		1.466***	1.353***
Triplets Plus		4.515***	2.669		3.908***	2.281*		3.615***	2.281*
Maternal Characteristics									
Teens			0.677***			0.854*			0.962
Twenties			ref			ref			ref
Thirties			1.514***			1.549***			1.479***
Forty and over			2.066			2.265***			2.238***
Prenatal 9 visits			0.865			1.139			1.030
Prenatal 10-20 visits			ref			ref			ref
Prenatal 20+ visits			0.757			1.041			0.960
White			ref			ref			ref
Black			0.723***			0.774*			0.984
Hispanic			1.211**			1.235**			1.528***
Other			0.806*			0.779*			0.819
Less than HS			0.749***			0.810***			0.747***
High School			ref			ref			ref
Some College			1.099			1.129*			1.207**
4 plus yrs College			1.756***			1.739***			1.737***
Married			1.487***			1.444***			1.387***
Alcohol			1.101			0.960			0.871
Tobacco			0.579***			0.494***			0.512
Weight Gain			0.985***			0.992***			0.993***
N	92021	92021	92021						
LRχ2	246.24	59256.14	66321.36						
df	110	150	300						
prob>χ2	0.0000	0.0000	0.0000						
Pseudo R2	0.0007	0.1571	0.1758						

* p<0.05, **p<0.01, ***p<0.001 (Base Outcome=External Causes)

Table 4.9 (continued) Multinomial Logistic Regression Results (Odds Ratios) Using Gestational Age: Cause Specific Infant Mortality by Month of Birth, Infant and Maternal Characteristics, U.S. 2000-2004

Month of Birth	Other Infections			Endocrine, Metabolic and Digestive System Disorders			Neoplasms and Blood Conditions		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
January	1.224*	1.256*	1.294*	1.055	1.100	1.144	1.191	1.204	1.277
February	1.059	1.076	1.098	0.953	0.978	1.002	1.153	1.160	1.198
March	0.966	0.987	0.996	0.954	0.987	1.003	1.361	1.376	1.412
April	0.954	0.964	0.983	1.217	1.229	1.264*	1.073	1.082	1.124
May	ref	ref	ref	ref	ref	ref	ref	ref	ref
June	1.120	1.119	1.122	0.935	0.936	0.944	1.228	1.223	1.233
July	1.132	1.136	1.148	0.973	0.981	1.001	0.948	0.948	0.989
August	1.114	1.141	1.146	0.931	0.966	0.978	1.152	1.163	1.195
September	1.145	1.193	1.198	0.957	1.017	1.029	0.931	0.946	0.977
October	1.257*	1.250*	1.268*	1.014	1.008	1.035	1.279	1.273	1.338
November	1.236*	1.266*	1.287*	0.889	0.927	0.951	1.160	1.164	1.226
December	1.051	1.076	1.090	0.997	1.035	1.057	1.035	1.046	1.088
Infant Characteristics									
Male		0.944	0.950		1.017	1.026		1.002	1.011
Gestational Age		0.841***	0.844***		0.798***	0.799***		0.904***	0.891***
Single Birth		ref	ref		ref	ref		ref	ref
Twins		1.362***	1.308**		1.388***	1.237*		0.520**	0.404***
Triplets Plus		2.517*	1.930		3.328**	1.961		1.278	0.504
Maternal Characteristics									
Teens			0.821***			0.876*			0.817
Twenties			ref			ref			ref
Thirties			1.285***			1.495*			1.563**
Forty and over			1.822***			2.064***			1.215
Prenatal 9 visits			0.783			0.781			0.780
Prenatal 10-20 visits			ref			ref			ref
Prenatal 20+ visits			0.794			0.760			0.871
White			ref			ref			ref
Black			1.116*			1.028			0.539***
Hispanic			1.428***			1.525***			1.237
Other			1.224*			1.028			1.275
Less than HS			0.917			0.907			0.810
High School			ref			ref			ref
Some College			1.043			1.150*			1.137
4 plus yrs College			1.273**			1.685***			2.219***
Married			1.202***			1.428***			1.887***
Alcohol			1.038			0.698			1.061
Tobacco			0.706***			0.525***			0.487***
Weight Gain			0.996*			0.999			1.003
N	92021	92021	92021						
LRχ2	246.24	59256.14	66321.36						
df	110	150	300						
prob>χ2	0.0000	0.0000	0.0000						
Pseudo R2	0.0007	0.1571	0.1758						

* p<0.05, **p<0.01, ***p<0.001 (Base Outcome=External Causes)

Table 4.9 (continued) Multinomial Logistic Regression Results (Odds Ratios) Using Gestational Age: Cause Specific Infant Mortality by Month of Birth, Infant and Maternal Characteristics, U.S. 2000-2004

Month of Birth	Respiratory, Circulatory and Nervous System Disorders		
	Model 1	Model 2	Model 3
January	1.047	1.062	1.11
February	0.769*	0.774*	0.793*
March	0.726**	0.737**	0.748**
April	0.903	0.911	0.938
May	ref	ref	ref
June	0.961	0.958	0.966
July	0.956	0.957	0.982
August	0.852	0.863	0.877
September	0.827	0.848	0.864
October	0.883	0.878	0.906
November	0.875	0.885	0.918
December	0.766*	0.776*	0.795*
Infant Characteristics			
Male		0.997	1.006
Gestational Age		0.881***	0.878***
Single Birth		ref	ref
Twins		1.075	0.91
Triplets Plus		2.659*	1.385
Maternal Characteristics			
Teens			0.769***
Twenties			ref
Thirties			1.602***
Forty and over			2.171***
Prenatal 9 visits			0.978
Prenatal 10-20 visits			ref
Prenatal 20+ visits			1.038
White			ref
Black			0.867*
Hispanic			1.176
Other			1.109
Less than HS			0.845**
High School			ref
Some College			1.121
4 plus yrs College			1.793***
Married			1.449***
Alcohol			0.902
Tobacco			0.513***
Weight Gain			0.999
N	92021	92021	92021
LRχ2	246.24	59256.14	66321.36
df	110	150	300
prob>χ2	0.0000	0.0000	0.0000
Pseudo R2	0.0007	0.1571	0.1758

* p<0.05, **p<0.01, ***p<0.001 (Base Outcome=External)

The results in Table 4.9 are similar to those in Table 4.8. The causes of death of birth asphyxia and birth trauma, perinatal infections, congenital anomalies, SIDS, and neoplasms and blood conditions showed no significant associations with any of the month of birth variables. Also similar to the results of Table 4.8, those in Table 4.9 show that the months January and April are significantly associated with an increased risk of death due to prematurity and related causes as compared to external causes of death. As in the first models, the month of January only showed a significant association in the third model, after all control variables were entered. This indicates that infants born in January, compared to those born in May, have a 22.3 percent higher risk of dying of prematurity and related causes than of dying of external causes of death. The similar finding regarding born in the month of April is new to these models, although like January, it loses its significance in models two and three.

In Table 4.9 we can also see that, similar to the results found in Table 4.8, the significant and negative association of the months of February, August, September and November with the cause of death "pregnancy complications" lost their significance in models two and three when the control variables measuring characteristics of the mother and infant were entered into the model.

Another similarity with the earlier analyses is that the month of April gains significance in the third model with the causes of death due to endocrine, metabolic and digestive system conditions. This can be interpreted as: For those infants born in April compared to those born in May, there is a 26.4 percent greater risk of dying of endocrine, metabolic, digestive system disorders as compared to dying from external

.

causes. This significant and positive association was also found in the models estimated earlier with the birthweight variable.

Also similar to the earlier models using the birthweight variable are the positive and significant associations in Table 4.9 for the months of January, October and November for the "other infections" causes of death, and the negative and significant associations for the months of February, March, and December for the respiratory, circulatory, and nervous system disorders causes of death. Irrespective of whether the model included birthweight or gestational age, the significant findings with these causes of death were maintained in all three models; neither birthweight, gestational age, nor the other control variables included in the models are able to "explain away" the associations of these months of birth with these causes of death. However, only the significant and positive association with the month of January supported my hypothesis.

Lastly, the cause of death of "respiratory, circulatory and nervous system disorder" was again found to have a negative and significant association with the months of birth of February, March and December. But only the months February and December were hypothesized to have a negative association with any of these causes of infant death, so once again the month of March's negative association was in the opposite direction of the hypotheses.

Results of Models Including Cases from California (with Gestational Age)

As in the models that include birthweight, I next estimated the full model that includes the variable gestational age and excludes the variables alcohol, tobacco and weight gain, therefore including cases from the state of California. For the full model the cause of

.

death prematurity is significant with the month of birth January in both the previous model that excluded California and in these analyses that includes California. However, the month of birth April was significantly associated with prematurity in the models that excluded California but loses significance in the models that include California. The cause of death SIDS is now significantly associated with the month of birth January in the models which include California but not in the previous models which excluded California. The cause of death of other infections is significantly associated with January, September, October, and November. In the models that excluded California the month of birth September was not significantly associated with the cause of death other infections. For the cause of death of endocrine conditions the month of birth April is significantly associated in the model which includes California but not in the model which excludes California. Lastly, the cause of death respiratory conditions is significantly associated in both the model that included California and in the model which excluded California with the month of birth March and December. However in the model that includes California, the month of birth February is no longer significant as it was in the models that excluded California.

Again, the causes of death of congenital anomalies, pregnancy complications, birth asphyxia and birth trauma, perinatal infections and neoplasms were not significantly associated with any of the months of birth in the models which included California and the models that excluded California. As in the model that included the variable birthweight, there are some changes in the significant findings when California cases are included in the model are found, but they are not drastic.

.

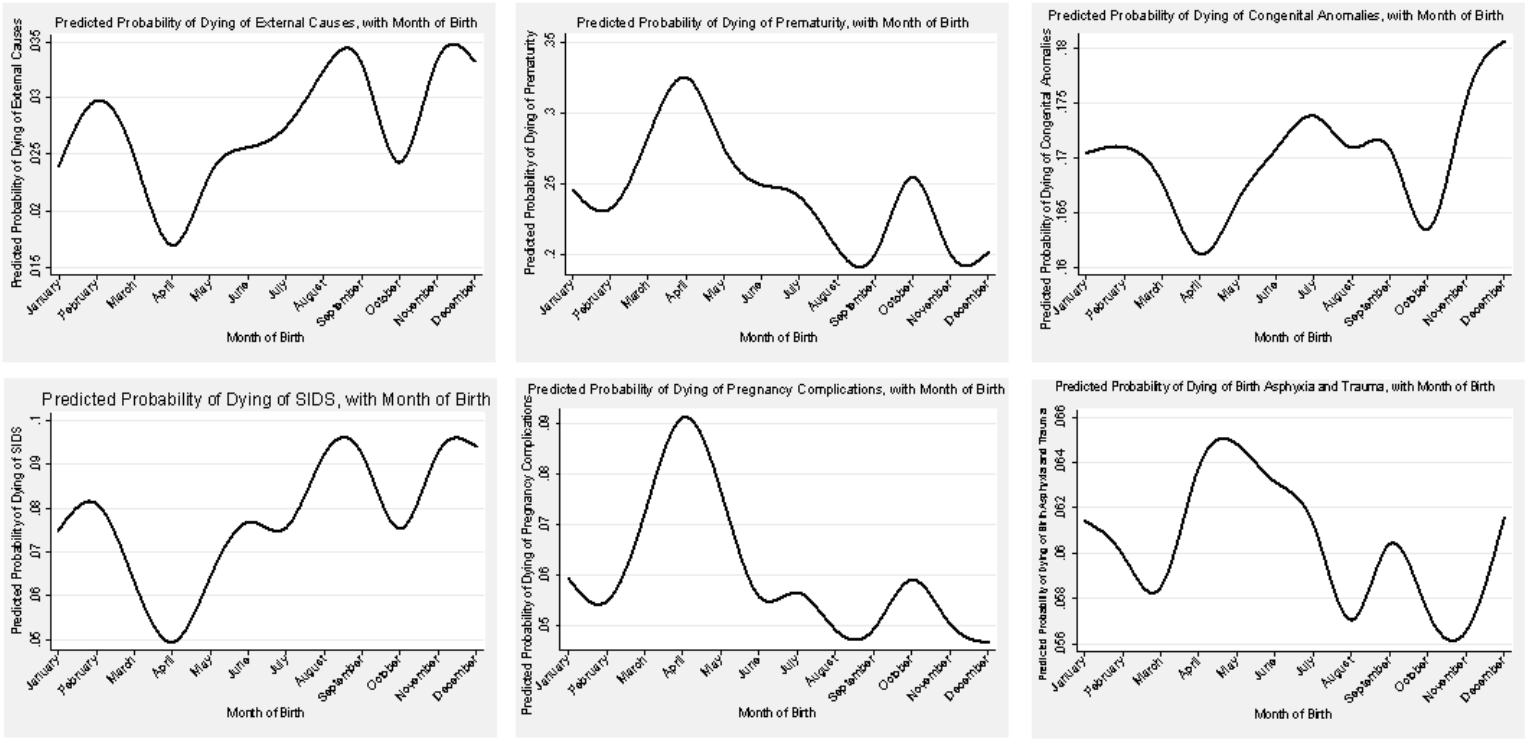
Predicted Probabilities

Below in Figure 4.6 are a series of eleven graphs that display the predicted probabilities based on the multinomial logit model including all control variables and the gestational age variable. The eleven graphs show the predicted probabilities for an infant to be in each category by their month of birth. The predicted probabilities for each of the causes of death are based on the full multinomial logit model according to the infant's month of death. As discussed in chapter III, Long and Freese (2001) tell us that using the predicted probabilities is a simple way to the display the results of a multinomial logit model, since with several categories of the dependent variable and several independent variables, the number of possible comparisons can be numerous.

Again, we can see from the graphs of the predicted probabilities that there is much variation in the probabilities of each of the months of birth in the graphs of Figure 4.7, just as in Figure 4.6. The month of January again has a high value of about 0.24 for the cause of death prematurity and a low value of 0.0046 for dying of neoplasms and blood conditions. If we look at the predicted probability of dying of prematurity, we see that the value is a low of 0.19 in the months of August to September and a high of about 0.32 in March to April. The months that see highs and lows in this model that included the gestational age variable as opposed to the birthweight variable have changed. This means that when using the gestational age variable and not the birthweight variable to derive the predicted probabilities, the months of birth are now associated with different probabilities of dying of the specified causes

.

Figure 4.7 Predicted Probabilities: Cause of Infant Death by Month of Birth (Gestational Age), U.S. 2000-2004



Conclusions

In chapter IV I have displayed and discussed the results of the first series of models examining cause specific infant mortality by month of birth. The analyses were split into two separate series of models, one that included the measure of the infant's birthweight, and one that included the measure of the infant's gestational age. The chapter also includes a discussion of the multicollinearity diagnostics. I also include a series of graphs that show my results in terms of predicted probabilities.

Since I hypothesized that all causes of death would be associated significantly with the month of birth variables, my hypotheses were not confirmed in the case of the causes of death of birth trauma and asphyxia, perinatal infections, congenital anomalies, SIDS, and neoplasms and blood conditions; none of these causes of death were significantly associated with any of the month of birth variables in any of the models. My hypotheses were also not confirmed when significant associations with the cause of death "pregnancy complications" lost significance in models two and three when the control variables were introduced to the models. A similar case was found with the month of September (in both series of models) for the cause of death "prematurity" which was only found to be significant in the first model.

An interesting finding is with the cause of death "prematurity" where the months of January (in both the birthweight and the gestational age models) and April (the gestational age model only) showed a significant association in model three when all control variables were introduced, but not in models one and two. This finding was also

.

seen in the cause of death category “endocrine, metabolic and digestive system disorders” (in the models with gestational age) for the month of April.

Only two causes of death were significantly associated in all three models with the month of birth variables, namely, “other infections” and “respiratory, circulatory, and nervous system conditions.” For the cause of death “other infections” the months January, October and November were found to be significantly and positively associated with this cause of death (in both the models that included the gestational age variable and the models that included the birthweight variable). This means that infants born in the months of January, October, and November, have an increased risk of dying from other infections as compared to external causes of death. Even though the significant relationship was maintained in all three models for these months, only the month of January performed in the hypothesized direction.

As for the cause of death category “respiratory, circulatory, and nervous system conditions” the months of February, March and December were found to be negatively and significantly associated with this cause in all three models. Again, however, only the months of February and December were in the direction hypothesized.

In both models that were estimated in this chapter—one including birthweight and one including gestational age—I have also estimated the full model without the variables alcohol, tobacco and weight gain. As discussed previously this is done because including these variables in the models excludes the cases from the state of California where these questions are not asked on the birth certificate. In the majority of cases the models with and without California had similar findings of association between month

.

of birth and cause specific infant mortality. Since the models were so similar when the control variables alcohol, tobacco and weight gain were and were not included, I believe that in the multilevel analyses in chapter VI, the exclusion of these variables in order to use all states as the second level of analysis will be reasonable and not effected by the exclusion of these variables in order to include the state of California.

In the next chapter I estimate similar models but I use the main independent variable measuring month of gestation rather than month of birth. I will then compare the overall results in chapter V with those reported in chapter IV and, hope to be able to determine if month of birth or month of gestation is the more accurate measure of the timing of an infant's risk of death. The more appropriate variable will then be used in the multilevel analyses to be reported in chapter VI. These multilevel analyses will take the results of chapters IV and V further by including state level measures of climate, specifically temperature, humidity and wind. Next, I turn to the discussion of the level-one analyses using months of third trimester.

.

CHAPTER V

LEVEL-ONE ANALYSES AND RESULTS: USING MONTHS OF FIRST TRIMESTER AND MONTHS OF THIRD TRIMESTER AS THE PRINCIPAL INDEPENDENT VARIABLES

Chapter V of this dissertation is similar to chapter IV. But instead of using "month of birth" as one of the principal independent variables in my multinomial logistic regression models I use the two principal independent variables that reflect the time of the year the infant was *in utero*, namely, the first and the third trimesters of their fetal development. My reason for distinguishing between these two types of independent variables is that I wish to ascertain whether the meaningful measurement of the association between seasonality and infant mortality is the infant's month of birth or its month of gestation.

In the literature previously discussed in chapter II I found that when studying the month of birth and adult longevity there is a consistent pattern of better survival for those individuals born in the months of October, November and December, and a consistent pattern of worse survival for those born in April, May and June. The hypotheses tested in chapter IV thus predicted that negative associations would be found with the months of October, November and December and the causes of infant death, and that positive associations would be found for the months of April, May and June.

However, the issue to consider is what exactly is it about the month of birth that could have a positive or negative impact on the infant's risk of death? If we think it is the actual month in which the infant was born that has the influence on mortality, then

.

the analyses of chapter IV are sufficient. However, if we believe there is something about the time of the year the infant is in the mother's womb, then we are interested in measuring gestation.

In the literature on fetal development there is consensus that the first trimester of the fetus's development is when the majority of the organs, tissues and body parts are formed. Then, in the third trimester, the necessary growth to the fetus, its body and organs take place. In this chapter I thus use two measures—months of first trimester and months of third trimester; I hope to ascertain whether the month in which an infant is born or these periods of development is more important to its risk of death from specified causes. In this chapter I discuss the construction and use of these independent variables of interest, and then present the analyses using both measures. The chapter concludes with my decision about which of the three measures—month of birth, months of first trimester, or months of third trimester—will be used in the multilevel analyses presented in chapter VI.

Operationalization

In the analyses undertaken in chapter IV, the variable month of birth was the major independent variable of interest. In the present chapter the independent variables of interest reflect not the month in which the infant was born, but instead the time the fetus was *in utero*. This will be accomplished by using a series of variables set to measure the months of the year in which the fetus was in his or her first trimester, and his or her third trimester. First I will discuss the variable that measures the months that the fetus was in utero for its first trimester.

.

The first step to creating this variable was to create a measure of the month of conception. This variable was constructed from the Linked Birth/Infant Death dataset using the variables measuring the infant's gestational age and their month of birth. For each infant I do not know the exact day of birth, only the month of birth or the day of the week of the birth (i.e. Monday, Tuesday, Wednesday, and so forth) so it was not possible to find the exact month of conception from the given information. Instead I considered each infant to be born in the middle of the reported month of birth and then counted backwards based on the value on the "gestational age in weeks" variable to find the estimated month of conception. Although counting all the infants as being born in the middle of the month has the potential to make mistakes on the month of conception variable that I constructed, without exact information about the date of birth and gestational age in days, this measure is the best that could be created.

This month of conception variable was used to create a measure of months of the year that the fetus was in the first trimester. As discussed previously, the literature on embryology considers the first trimester to be the most important to the development of tissues, organs and body parts of the fetus. If we consider the first trimester the most important to the chances of the fetus's survival, then this would be the time period that would be of greatest interest in these analyses. All fetuses in these data were born after their first trimester, so any measure that is meant to identify this time period would necessarily include all infants in the dataset.

As we know, by definition the first trimester of gestation is comprised of the first three months after conception that the fetus is *in utero*. In order to measure this time

.

period I created 12 dummy variables to measure the 12 possible three-month periods that a fetus could be in the first trimester—January/February/March, February/March/April, March/April/May, May/June/July, and so on. These categories are mutually exclusive since it is only possible for the first trimester to fall into one of these twelve three-month periods. Then using the month of conception variable that I discussed above, I grouped the fetuses into one of the twelve dummy variables of the months of the first trimester. For example, if a fetus was conceived in January, then it was in its first trimester in the months of January, February and March; therefore for the variable Jan/Feb/Mar those conceived in January would be given a value of one and for all other dummy variables, a value of zero.

The operationalization discussed in chapter IV regarding the dependent variable "cause of infant death" and the independent control variables are used again in the analyses of this chapter. To avoid repetition, I do not discuss the creation of these variables here. It is also important to note that as in chapter IV the models performed in this chapter will have to be repeated excluding the control variables alcohol use, tobacco use and weight gain during pregnancy to include the state of California in the analyses. As in chapter for the exclusion of these control variables had little effect on the significance of the month of first or third trimester variables with cause specific infant death Now I will discuss the hypotheses of the first analyses of this chapter, which use months of first trimester as the independent variables of interest.

.

Hypotheses

In chapter IV the hypotheses regarding the association of cause specific infant mortality and month of birth were based on findings of past literature on adult longevity; the hypotheses expected that being born in the months of October, November, December, January and February would be found to be negatively associated with cause specific infant death, and that being born in the months of March, April, May, June, July, August and September would be positively associated with cause specific infant mortality.

Since I am changing the measure from month of birth to a measure of when the infant was *in utero,* the months with which I expect to find an association with cause specific infant death will also change. The past literature found that those individuals born in the months of October, November and December were at an advantage in terms of their longevity and those born in April, May and June were at a disadvantage. If we assume that these individuals were born at term we would translate the advantage of the months of October, November and December into an advantage in their first trimester in the months of February/March/April, March/April/May, and April/May/June respectively. Conversely if we also assume that those individuals born in the disadvantaged months of April, May and June were born to term, then they would be in their first trimester in the months of August/September/October, September/October/November and October/November/December respectively. As such I hypothesize that the months of the first trimester of February/March/April, March/April/May, and April/May/June will be associated negatively with cause specific infant mortality, and that the months of August/September/October,

.

September/October/November and October/November/December will be positively associated with cause specific infant mortality.

Descriptive Results

In this section of chapter V I further discuss the variables that will be used in the analyses of this chapter. I provide descriptive results of the variable measuring months of the first trimester, the independent variable of major interest in the first series of analyses of this chapter. The descriptions of the independent variables that will be used as controls in these analyses are the same as those used in chapter IV. As such, please refer to chapter IV for a complete description of these variables.

Table 5.1 Descriptive Statistics of Months of First Trimester Variable, U.S. 2000-2004

Months of First Trimester 12 dummy variables 0=first trimester in those months 1= first trimester not in those months	Minimum	Maximum	Mean	Standard Deviation	N
Jan/Feb/Mar	0	1	0.084	0.278	92021
Feb/Mar/Apr	0	1	0.085	0.280	92021
Mar/Apr/May	0	1	0.084	0.278	92021
Apr/May/June	0	1	0.084	0.277	92021
May/June/July	0	1	0.083	0.275	92021
June/July/Aug	0	1	0.080	0.271	92021
July/Aug/Sept	0	1	0.079	0.270	92021
Aug/Sept/Oct	0	1	0.077	0.267	92021
Sept/Oct/Nov	0	1	0.079	0.270	92021
Oct/Nov/Dec	0	1	0.084	0.277	92021
Nov/Dec/Jan	0	1	0.088	0.283	92021
Dec/Jan/Feb	0	1	0.088	0.283	92021

In Table 5.1 we see that all variables are dummy variables with maximum values of one and minimum values of zero. The highest mean of these variables is in the first trimester of November/December/January and December/January/February with values

.

of 0.088 for both. The lowest mean is for the months of first trimester of August/September/October with a value of 0.077.

Next I will discuss the diagnostics for collinearity in my models with month of first trimester as the independent variable of interest. This will be followed by the results and a discussion of the six models analyzed in this section.

Multicollinearity Diagnostics

Although the independent control variables that will be used in the analyses of this chapter are the same as those in chapter for, the use of the variable measuring months first trimester instead of month of birth changes the models from those in chapter IV. Although the independent control variables remain the same, the use of the new independent variable of interest—months of third trimester—mean that I will have to again conduct the multicollinearity diagnostics for the models of this chapter.

First I examined the zero-order correlations for the independent variables of the models. These correlations show that there may be issues, as in chapter IV, with the estimation of the model with both the birthweight and the gestational age variables together. Again, as was the case in chapter IV, there also seems to be issues with estimating the models that contain all the dummy measures of the months of first trimester, race/ethnicity, plurality, education, and age. To ascertain the degree to which the variables are collinear I calculated tolerance values.

As was suspected from the results of the zero-order correlations, the birthweight and gestational age variables showed a high degree of collinearity with tolerance values of 0.18. These tolerance values suggest there would likely be a problem using both

.

variables in my models. As in chapter IV I will thus estimate two sets of models, one with the birthweight variable, and one with the gestational age variable. Also like the models estimated in chapter IV , I will leave the following variables out of the models, in order to minimize collinearity—ten to twenty prenatal visits (pre10_20), high school education (HS), whether the birth is a single birth (single birth), mothers aged twenty to twenty-nine (twenties), and white mothers (white). For a discussion of the rationale for excluding the above variables from the analyses, please see chapter IV.

I also need to determine which of the dummy variables measuring the infant's months of their first trimester to use as the reference dummy variable. In chapter IV I decided to exclude the month of birth "May" because it had the number of births in the dataset that was closest to the average number of monthly births for all four years. In this section of chapter V I chose to exclude the dummy variable "Dec/Jan/Feb" because it was the month for which I expect to see a negative association with the cause of infant death dependent variable—that is, those fetuses who experienced their third trimester in the months of December, January and February would have a decreased likelihood of experiencing a specified cause of death as compared to external causes of death. In both series of models shown below (one using gestational age, and one using birthweight) I will exclude the dummy variable Dec/Jan/Feb as the first trimester of reference.

.

Next, I turn to a discussion of the results from the multinomial logistic regression models using the variables measuring month of gestation.

Multinomial Logistic Regression Results (Using Birthweight but Not Gestational Age)

Below in Table 5.2 are the results of the first three models including the birthweight variable and the months of the first trimester as the principal independent variable of interest. Due to the issues with multicollinearity discussed previous, these results will be repeated later in this chapter with the variable gestational age and not birthweight. As with the previous models, the results when including birthweight are similar to the results for the models including gestational age. This is important to note because it means that in further analyses, using only one of these control variables will not mean that findings are derived from models that lack one or more important factor to the cause specific death to infants in the United States.

.

Table 5.2 Multinomial Logistic Regression Results (Odds Ratios) Using Birthweight: Cause Specific Infant Mortality by Months of First Trimester, Infant and Maternal Characteristics, U.S. 2000-2004

Months of First Trimester	Prematurity and Related Conditions			Congenital Anomalies			SIDS		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Jan/Feb/Mar	1.070	1.038	1.051	1.029	1.020	1.037	1.056	1.056	1.052
Feb/Mar/Apr	1.158*	1.090	1.100	1.130	1.112	1.131	1.024	1.021	1.021
Mar/Apr/May	0.995	1.012	1.030	1.044	1.045	1.069	1.072	1.073	1.072
Apr/May/June	1.006	0.972	0.991	1.006	0.993	1.019	0.989	0.987	0,984
May/June/July	1.220**	1.157	1.173	1.102	1.083	1.100	1.086	1.080	1.081
June/July/Aug	1.077	1.065	1.064	1.091	1.087	1.081	0.992	0.992	0.992
July/Aug/Sept	1.078	1.072	1.071	1.096	1.097	1.085	0.928	0.929	0.930
Aug/Sept/Oct	1.170*	1.184*	1.188*	1.160	1.180*	1.167	1.021	1.025	1.026
Sept/Oct/Nov	1.058	1.042	1.031	1.108	1.097	1.070	0.979	0.977	0.974
Oct/Nov/Dec	1.000	1.007	1.026	0.984	0.983	1.008	0.992	0.991	0.992
Nov/Dec/Jan	1.083	1.085	1.072	1.056	1.063	1.042	1.005	1.005	1.003
Dec/Jan/Feb	ref	ref	ref	ref	ref	ref	ref	ref	ref
Infant Characteristics									
Male		1.017	1.022		0.989	1.003		1.021	1.021
Birthweight		0.997***	0.997***		0.999***	0.998***		0.999***	0.999***
Single Birth		ref	ref		ref	ref		ref	ref
Twins		0.976	0.840*		0.684***	0.460***		1.221*	1.205*
Triplets Plus		2.318*	1.306		0.926	0.287**		1.559	1.486
Maternal Characteristics									
Teens			0.814***			0.735***			0.936
Twenties			ref			ref			ref
Thirties			1.568***			1.761***			1.007
Forty and over			1.612***			3.339***			0.832
Prenatal 9 visits			1.351***			0.898**			1.005
Prenatal 10-20 visits			ref			ref			ref
Prenatal 20+ visits			1.210			1.846***			0.966
White			ref			ref			ref
Black			0.829***			0.51***			1.065
Hispanic			1.275***			1.342***			0.933
Other			0.940			0.824*			0.909
Less than HS			0.790***			0.829***			0.966
High School			ref			ref			ref
Some College			1.170**			1.121*			0.997
4 plus yrs College			1.818***			1.723***			1.132
Married			1.436***			1.723***			1.108**
Alcohol			0.896			0.845			0.960
Tobacco			0.439***			0.367***			1.188***
Weight Gain			0.991***			1.004***			0.999
N	92021	92021	92021						
LRχ2	174.32	54555.26	63148.51						
df	110	150	300						
prob>χ2	0.0001	0.0000	0.0000						
Pseudo R2	0.0005	0.1446	0.1674						

* p<0.05, **p<0.01, ***p<0.001 (Base Outcome=External Causes)

Table 5.2 (Continued) Multinomial Logistic Regression Results (Odds Ratios) Using Birthweight: Cause Specific Infant Mortality by Months of First Trimester, Infant and Maternal Characteristics, U.S. 2000-2004

Season of First Trimester	Pregnancy Complications			Birth Asphyxia and Birth Trauma			Perinatal Infections		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Jan/Feb/Mar	1.067	1.033	1.041	1.121	1.098	1.114	1.078	1.053	1.072
Feb/Mar/Apr	1.052	0.983	0.992	1.166	1.113	1.125	1.092	1.043	1.056
Mar/Apr/May	1.052	1.069	1.091	1.152	1.165	1.190	1.110	1.123	1.143
Apr/May/June	1.013	0.975	0.994	1.131	1.099	1.222	1.028	1.000	1.022
May/June/July	1.182	1.114	1.124	1.186	1.132	1.149	1.104	1.056	1.074
June/July/Aug	1.131	1.115	1.112	1.136	1.124	1.121	1.115	1.105	1.103
July/Aug/Sept	1.056	1.046	1.042	1.147	1.139	1.131	1.048	1.042	1.043
Aug/Sept/Oct	1.186*	1.195	1.187	1.200	1.217	1.212	1.007	1.023	1.028
Sept/Oct/Nov	1.087	1.065	1.049	1.179	1.162	1.144	0.938	0.926	0.916
Oct/Nov/Dec	0.981	0.987	1.002	1.018	1.021	1.042	0.994	0.997	1.020
Nov/Dec/Jan	1.172	1.168	1.150	1.26*	1.104*	1.245*	1.130	1.135	1.121
Dec/Jan/Feb	ref	ref	ref	ref	ref	ref	ref	ref	ref
Characteristics									
Male		1.047	1.051		1.104*	1.112*		1.013	1.020
Birthweight		0.997***	0.997***		0.998***	0.998***		0.998***	0.998***
Single Birth		ref	ref		ref	ref		ref	ref
Twins		1.171*	1.015		1.225*	1.005		0.972	0.796*
Triplets Plus		2.821**	1.569		2.621*	1.297		2.283*	1.162
Maternal Characteristics									
Teens			0.657***			0.828**			0.924
Twenties			ref			ref			ref
Thirties			1.611***			1.618***			1.534***
Forty and over			1.670***			1.841***			1.798***
Prenatal 9 visits			1.501***			1.264***			1.175**
Prenatal 10-20 visits			ref			ref			ref
Prenatal 20+ visits			1.201			1.043			1.186
White			ref			ref			ref
Black			0.678***			0.707***			0.878*
Hispanic			1.177**			1.189*			1.471***
Other			0.791*			0.763*			0.808
Less than HS			0.724***			0.788***			0.734***
High School			ref			ref			ref
Some College			1.163**			1.188**			1.262***
4 plus yrs College			1.891***			1.840***			1.828***
Married			1.541***			1.507***			1.450***
Alcohol			0.966			0.877			0.799
Tobacco			0.488***			0.418***			0.434***
Weight Gain			0.982***			0.994***			0.998
N	92021	92021	92021						
LRχ2	174.32	54555.26	63148.51						
df	110	150	300						
prob>χ2	0.0001	0.0000	0.0000						
Pseudo R2	0.0005	0.1446	0.1674						

* p<0.05, **p<0.01, ***p<0.001 (Base Outcome=External Causes)

.

Table 5.2 (Continued) Multinomial Logistic Regression Results (Odds Ratios) Using Birthweight: Cause Specific Infant Mortality by Months of First Trimester, Infant and Maternal Characteristics, U.S. 2000-2004

Season of First Trimester	Other Infections			Endocrine, Metabolic and Digestive System Disorders			Neoplasms and Blood Conditions		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Jan/Feb/Mar	0.984	0.976	0.990	0.961	0,947	0.965	1.181	1.179	1.208
Feb/Mar/Apr	1.198	1.175	1.186	1.094	1.061	1.075	1.234	1.230	1.264
Mar/Apr/May	1.090	1.092	1.105	1.036	1.042	1.063	1.216	1.212	1.259
Apr/May/June	1.095	1.080	1.100	0.963	0.943	0.962	1.311	1.305	1.335
May/June/July	1.041	1.016	1.033	1.148	1.109	1.133	1.074	1.072	1.100
June/July/Aug	0.883	0.878	0.880	0.908	0.901	0.899	1.275	1.273	1.278
July/Aug/Sept	0.856	0.854	0.861	1.172	1.168	1.166	1.426	1.426	1.425
Aug/Sept/Oct	0.900	0.914	0.924	1.077	1.095	1.099	1.044	1.048	1.046
Sept/Oct/Nov	0.907	0.897	0.889	0.950	0.938	0.926	1.211	1.206	1.175
Oct/Nov/Dec	0.912	0.911	0.924	0.981	0.981	1.002	1.167	1.165	1.210
Nov/Dec/Jan	0.977	0.980	0,977	0.986	0.991	0.979	1.363	1.366	1.344
Dec/Jan/Feb	ref	ref	ref	ref	ref	ref	ref	ref	ref
Infant Characteristics									
Male		0.974	0.981		1.056	1.066		1.012	1.024
Birthweight		0.999***	0.999***		0.999***	0.998***		0.999***	0.999***
Single Birth		ref	ref		ref	ref		ref	ref
Twins		1.057	0.915		0.982	0.772**		0.596*	0.415***
Triplets Plus		1.899	1.196		2.260*	1.077		1.643	0.548
Maternal Characteristics									
Teens			0.792***			0.837**			0.811
Twenties			ref			ref			ref
Thirties			1.328***			1.549***			1.592***
Forty and over			1.607**			1.725**			1.149
Prenatal 9 visits			1.05			1.076			1.030
Prenatal 10-20 visits			ref			ref			ref
Prenatal 20+ visits			1.370**			1.316			1.135
White			ref			ref			ref
Black			1.034			0.926			0.540***
Hispanic			1.392***			1.468***			1.219
Other			1.228*			1.019			1.304
Less than HS			0.901*			0.892			0.793
High School			ref			ref			ref
Some College			1.074			1.195**			0.167
4 plus yrs College			1.320***			1.757***			2.199***
Married			1.247***			1.493***			1.928***
Alcohol			0.988			0.653			1.006
Tobacco			0.623***			0.452***			0.461***
Weight Gain			1.000			1.004			1.001
N	92021	92021	92021						
LRχ2	174.32	54555.26	63148.51						
df	110	150	300						
prob>χ2	0.0001	0.0000	0.0000						
Pseudo R2	0.0005	0.1446	0.1674						

* p<0.05, **p<0.01, ***p<0.001 (Base Outcome=External Causes)

.

Table 5.2 (Continued) Multinomial Logistic Regression Results (Odds Ratios) Using Birthweight: Cause Specific Infant Mortality by Months of First Trimester, Infant and Maternal Characteristics U.S. 2000-2004

Season of First Trimester	Respiratory, Circulatory and Nervous System Disorders		
	Model 1	Model 2	Model 3
Jan/Feb/Mar	1.000	0.995	1.012
Feb/Mar/Apr	1.155	1.141	1.158
Mar/Apr/May	0.941	0.943	0.961
Apr/May/June	0.947	0.938	0.958
May/June/July	1.039	1.023	1.043
June/July/Aug	0.893	0.890	0.884
July/Aug/Sept	1.007	1.006	1.001
Aug/Sept/Oct	1.006	1.019	1.014
Sept/Oct/Nov	1.023	1.014	0.995
Oct/Nov/Dec	1.003	1.001	1.033
Nov/Dec/Jan	1.005	1.008	0.991
Dec/Jan/Feb	ref	ref	ref
Infant Characteristics			
Male		1.022	1.033
Birthweight		0.999***	0.999***
Single Birth		ref	ref
Twins		0.905	0.673***
Triplets Plus		2.177	0.906
Maternal Characteristics			
Teens			0.743***
Twenties			ref
Thirties			1.646***
Forty and over			1.986***
Prenatal 9 visits			0.976
Prenatal 10-20 visits			ref
Prenatal 20+ visits			1.183
White			ref
Black			0.813***
Hispanic			1.153*
Other			1.119
Less than HS			0.833**
High School			ref
Some College			1.148*
4 plus yrs College			1.839***
Married			1.494***
Alcohol			0.862
Tobacco			0.463***
Weight Gain			1.002
N	92021	92021	92021
LRχ2	174.32	54555.26	63148.51
df	110	150	300
prob>χ2	0.0001	0.0000	0.0000
Pseudo R2	0.0005	0.1446	0.1674

* p<0.05, **p<0.01, ***p<0.001 (Base= External Causes)

As in the analyses of chapter IV, several of the causes of infant death were not shown to have any significant association between the months of first trimester variables and the causes of infant death. The causes of death of perinatal infections, respiratory, circulatory and nervous system causes, SIDS, other infections, endocrine, metabolic and digestive system causes and neoplasms and blood conditions had no significant associations with the months of first trimester variables.

For the causes of death in the category “prematurity and related causes” the months of the first trimester August/September/October were significantly associated with an increased risk of death as compared to external causes of death. This significance was maintained in all three models. For model three the odds ratio is 1.188, which may be interpreted as follows: for those fetuses whose first trimester was in the months of August, September and October, compared to those whose first trimester was in December, January and February, have a risk of dying of prematurity and related causes that is 18.8 percent higher than for dying of external causes of death. For the other months of the first trimester, there were significant associations in the first model that did not include the control variables. However for the months of February, March and April this positive and significant finding was only found in model one. For the months of the first trimester May, June and July the odds ratio in model one was 1.220 and was significant at the 0.01 level. However, for models two and three these months of the first trimester maintained the positive association but lost significance, with a p value of 0.079 in model two and a p value 0.059 in model three.

.

The cause of death of “congenital anomalies” shows in model one the months of the first trimester August, September and October to be significantly associated with congenital anomalies only in model two. In model one the odds ratio had a z value of 1.94 and a p>|z| of 0.052 and in model three the odds ratio will be considered significant with a p>|z| value of 0.054.

The causes of death ‘pregnancy complications’ also had a case of a loss of significance in models two and three, after showing a significant association in model one. In model one the odds ratio is 1.186, which may be interpreted as indicating that for those infants whose first trimester was in the months of August, September and October as compared to those whose first trimester fell in the months of December, January and February, had a 18.6 percent increased risk of dying of pregnancy complications as compared to external causes of death, and this was significant at the p>0.05 level. In models one and two the significant association is lost (although barely so); indeed both odds ratios are significant at the p>0.1 level with p>|z| values of 0.06 in model two and 0.075 in model three.

The causes of ‘birth asphyxia and birth trauma’ were found to be positively and significantly associated with the months of the first trimester of November, December and January in all three models. In model three the odds ratio is 1.245 meaning that those infants whose first trimester fell into the months November, December and January, compared to the months of December, January and February, had a risk of dying of birth asphyxia or birth trauma that was 24.5 percent higher than dying of external causes. Also interesting in the case of birth asphyxia and birth trauma is that in

.

all three models the months of August, September and October, although not found to be significant at the $p>0.05$ level, were approaching significance with p values of 0.059, 0.054 and 0.062 respectively. Indeed this first trimester may be said to be associated with an increased risk of death from birth asphyxia and birth trauma as compared to external causes at the $p>0.1$ level.

It seems that based on the models with the months of first trimester variable, as compared to the models with the month of birth variable (as presented and discussed in the previous chapter), there is no clear finding as to which principal independent variable is the more appropriate measure.

Results of the Model Including Cases from California (with Birthweight)

As discussed earlier, these models with duplicated excluding the control variables alcohol, tobacco and weight gain so that the state of California would be included in the analyses. Similar to chapter IV, there were some differences between the models that excluded California and the models that excluded California; however the differences were not drastic. For the full model that includes the variables birthweight but not gestational age only the cause of death prematurity and related causes showed any significant relationships between the cause specific infant death variable and the month of first trimester variables. The months of first trimester May/June/July were found to be significantly associated with causes of death due to prematurity. In the original analyses of chapter V this significant relationship in model three was not observed, although it was observed in the first model which included only the months of first trimester variables. The r-squared value in the models which included California was 0.1614

.

which is only slightly lower than the r-squared observed in the models which excluded California which was 0.1674. This very small decrease to the r-squared value shows that the inclusion of the variables alcohol, tobacco and weight gain does not give the model any substantial predictive power.

Predicted Probabilities

I now display the results of the models estimated with the season of first trimester variable using the predicted probabilities of an infant being in one of the categories of cause specific infant death. Below in Figure 5.1 are the graphical representations of the predicted probabilities of the cause specific infant mortality variable and the twelve monthly measures of when the infant was in its first trimester. These predicted probabilities are based on the full model which included all control variables and the birthweight variable. A similar set of graphs will be shown for the full model that included the gestational age variable.

As we can see from the figures below, just as was shown in the models in chapter IV, the predicted probabilities change from month to month for each of the causes of infant death. Also similar to the models shown in chapter IV, the highest probabilities are found for all months for the cause of death of “prematurity” and the lowest probabilities are found for the cause of death of “neoplasms and blood conditions.” These graphs do not seem to show a pattern of overall lower probability of dying of the specified cause for those infants whose first trimester was in the months of October to February and higher probabilities for those infants whose first trimester was in the months of March to September. Instead the figures show for the most part that for each

.

cause of death the probabilities rise and fall by months of first trimester. For example, the cause of death of prematurity shows a moderately high rate for January/February/March and February/March/April, and then there is a dip in the probability for the months March/April/May, which again sharply rises in the months May/June/July.

The “other infections” cause of death is an interesting case. Its graph seems to show that there is an overall high probability of dying for the months from January to June, and then a lowering for the summer months, and then another rise in the probabilities for the fall and winter months. It will be interesting to see if these same patterns are seen with the model which includes gestational age, and later in this chapter for the models that use months of third trimester. Next, I turn to the multinomial logistic regression results for the model using gestational age.

.

Figure 5.1 Predicted Probabilities: Cause of Infant Death and Months of First Trimester (Birthweight), U.S. 2000-2004

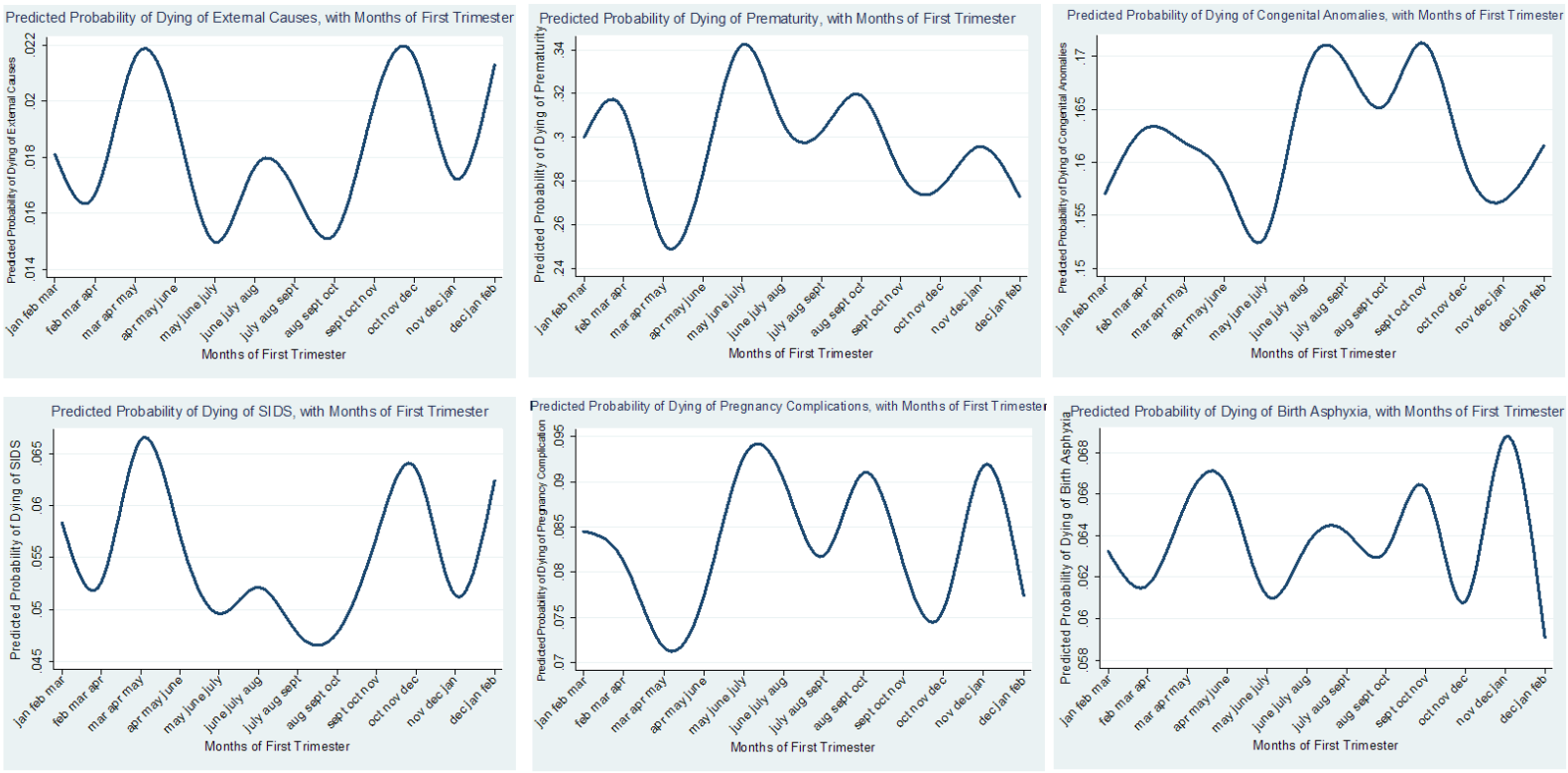
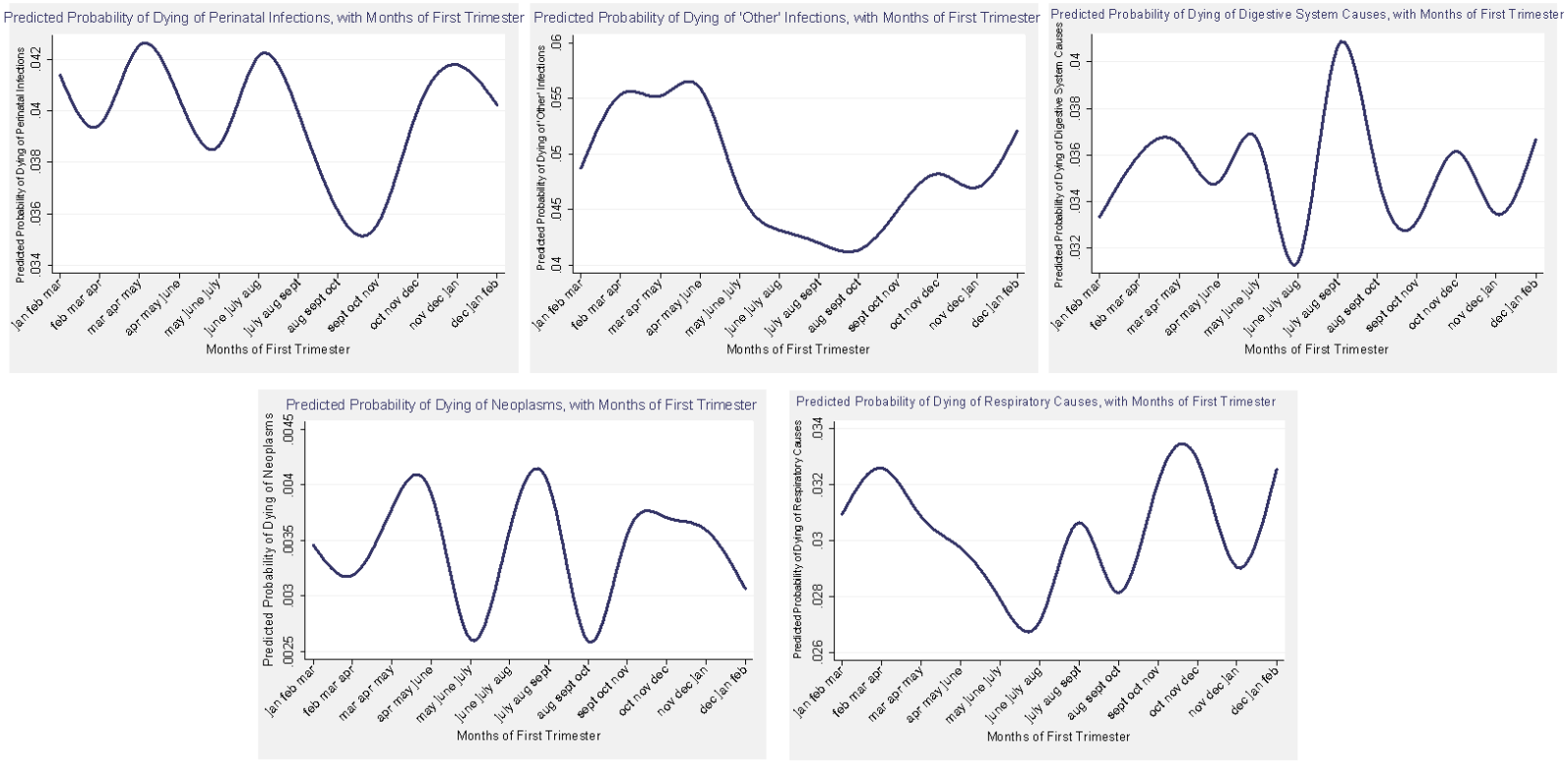


Figure 5.1 (Continued) Predicted Probabilities: Cause of Infant Death and Months of First Trimester (Birthweight), U.S. 2000-2004



Multinomial Logistic Regression Results (Using Gestational Age but Not Birthweight)

The next series of models to be presented uses the gestational age variable, instead of the birthweight variable, and examines once again cause specific infant mortality by the months of the infant's first trimester. These models contain the identical variables as those described in the previous section except for the exclusion of the birthweight variable and the inclusion of the gestational age variable.

Below in Table 5.3 are the results of the multinomial logistic regression using the main independent variable, months of first trimester, and the independent control variables including gestational age but excluding birthweight. As we can see, as in the other models, the results are similar to previous models in terms of the significant associations and in terms of the direction of those significant findings.

.

Table 5.3 Multinomial Logistic Regression Results (Odds Ratios) Using Gestational Age: Cause Specific Infant Mortality by Months of First Trimester, Infant and Maternal Characteristics, U.S. 2000-2004

Season of First Trimester	Prematurity and Related Conditions			Congenital Anomalies			SIDS		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Jan/Feb/Mar	1.070	1.063	1.075	1.029	1.029	1.043	1.056	1.056	1.052
Feb/Mar/Apr	1.158*	1.075	1.088	1.130	1.117	1.137	1.024	1.023	1.022
Mar/Apr/May	0.995	1.051	1.073	1.044	1.061	1.091	1.072	1.076	1.075
Apr/May/June	1.006	1.006	1.025	1.006	1.004	1.029	0.989	0.988	0.984
May/June/July	1.220**	1.144	1.159	1.102	1.076	1.091	1.086	1.079	1.078
June/July/Aug	1.077	1.097	1.104	1.091	1.096	1.100	0.992	0.993	0.994
July/Aug/Sept	1.078	1.121	1.127	1.096	1.108	1.107	0.928	0.929	0.930
Aug/Sept/Oct	1.170*	1.189*	1.194*	1.160	1.178*	1.172*	1.021	1.024	1.023
Sept/Oct/Dec	1.058	1.082	1.079	1.108	1.114	1.101	0.979	0.980	0.977
Oct/Nov/Dec	1.000	1.043	1.063	0.984	1.001	1.025	0.992	0.995	0.996
Nov/Dec/Jan	1.083	1.090	1.084	1.056	1.066	1.051	1.005	1.005	1.005
Dec/Jan/Feb	ref	ref	ref	ref	ref	ref	ref	ref	ref
Infant Characteristics									
Male		0.971	0.975		0.960	0.970		1.015	1.016
Gestational Age		0.673**	0.681***		0.867***	0.865***		0.973***	0.975***
Single Birth		ref	ref		ref	ref		ref	ref
Twins		1.517***	1.409***		1.015	0.828*		1.275**	1.275**
Triplets Plus		3.737***	2.383*		1.535	0.662		1.662	1.631
Maternal Characteristics									
Teens			0.840***			0.784***			0.943
Twenties			ref			ref			ref
Thirties			1.487***			1.723***			0.997
Forty and over			2.011***			3.854***			0.843
Prenatal 9 visits			1.110**			1.015			0.999
Prenatal 10-20 visits			ref			ref			ref
Prenatal 20+ visits			1.250*			1.898***			0.965
White			ref			ref			ref
Black			0.903*			0.616***			1.079
Hispanic			1.317***			1.375***			0.938
Other			0.956			0.816*			0.905
Less than HS			0.811***			0.822***			0.968
High School			ref			ref			ref
Some College			1.108*			1.100*			0.992
4 plus yrs College			1.700***			1.683***			1.124
Married			1.381***			1.649***			1.101*
Alcohol			1.008			0.878			0.966
Tobacco			0.521***			0.414***			1.213***
Weight Gain			0.989***			0.995***			0.999
N	92021	92021	92021						
LRχ2	174.32	59243.46	66353.89						
df	110	150	300						
prob>χ2	0.0001	0.0000	0.0000						
Pseudo R2	0.0005	0.1570	0.1759						

* p<0.05, **p<0.01, ***p<0.001 (Base Outcome=External Causes)

.

Table 5.3 (Continued) Multinomial Logistic Regression Results (Odds Ratios) Using Gestational Age: Cause Specific Infant Mortality by Season of First Trimester, Infant and Maternal Characteristics, U.S. 2000-2004

Season of First Trimester	Pregnancy Complications			Birth Asphyxia and Birth Trauma			Perinatal Infections		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Jan/Feb/Mar	1.067	1.061	1.071	1.121	1.115	1.127	1.078	1.071	1.086
Feb/Mar/Apr	1.052	0.968	0.980	1.166	1.102	1.117	1.092	1.034	1.049
Mar/Apr/May	1.052	1.113	1.140	1.152	1.204	1.231*	1.110	1.159	1.182
Apr/May/June	1.013	1.012	1.033	1.131	1.127	1.149	1.028	1.026	1.046
May/June/July	1.182	1.105	1.116	1.186	1.115	1.130	1.104	1.040	1.057
June/July/Aug	1.131	1.155	1.158	1.136	1.144	1.148	1.115	1.123	1.311
July/Aug/Sept	1.056	1.103	1.102	1.147	1.172	1.174	1.048	1.073	1.082
Aug/Sept/Oct	1.186*	1.210	1.195	1.200	1.220*	1.217	1.007	1.025	1.033
Sept/Oct/Dec	1.087	1.109	1.102	1.179	1.196	1.189	0.938	0.953	0.953
Oct/Nov/Dec	0.981	1.026	1.043	1.018	1.051	1.072	0.994	1.025	1.050
Nov/Dec/Jan	1.172	1.172	1.161	1.26*	1.274*	1.262*	1.130	1.143	1.136
Dec/Jan/Feb	ref	ref	ref	ref	ref	ref	ref	ref	ref
Infant Characteristics									
Male		0.999	1.007		1.060	1.069		0.971	0.979
Gestational Age		0.646***	0.655**		0.745***	0.751***		0.743***	0.750***
Single Birth		ref	ref		ref	ref		ref	ref
Twins		1.816***	1.651**		1.775***	1.596***		1.466***	1.347***
Triplets Plus		4.507***	2.673*		3.893***	2.276*		3.590***	2.249*
Maternal Characteristics									
Teens			0.677***			0.855*			0.962
Twenties			ref			ref			ref
Thirties			1.512***			1.549***			1.477***
Forty and over			2.058***			2.264***			2.234***
Prenatal 9 visits			1.146**			1.094			1.083
Prenatal 10-20 visits			ref			ref			ref
Prenatal 20+ visits			1.246			1.051			1.210
White			ref			ref			ref
Black			0.723***			0.773***			0.983
Hispanic			1.211**			1.236**			1.529***
Other			0.806*			0.780**			0.821
Less than HS			0.749***			0.811***			0.747***
High School			ref			ref			ref
Some College			1.098			1.128*			1.204**
4 plus yrs College			1.752**			1.738***			1.735***
Married			1.488**			1.444***			1.386***
Alcohol			1.101			0.960			0.871
Tobacco			0.580**			0.494***			0.512***
Weight Gain			0.985**			0.992***			0.993***
N	92021	92021	92021						
LRχ2	174.32	59243.46	66353.89						
df	110	150	300						
prob>χ2	0.0001	0.0000	0.0000						
Pseudo R2	0.0005	0.1570	0.1759						

* p<0.05, **p<0.01, ***p<0.001 (Base Outcome=External Causes)

.

Table 5.3 (Continued) Multinomial Logistic Regression Results (Odds Ratios) Using Gestational Age: Cause Specific Infant Mortality by Season of First Trimester, Infant and Maternal Characteristics, U.S. 2000-2004

Season of First Trimester	Other Infections			Endocrine, Metabolic and Digestive System Disorders			Neoplasms and Blood Conditions		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Jan/Feb/Mar	1.067	0.982	0,994	1.121	0,957	0.972	1.078	1.181	1.215
Feb/Mar/Apr	1.052	1.175	1.186	1.166	1.058	1.073	1.092	1.228	1.260
Mar/Apr/May	1.052	1.115	1.130	1.152	1.069	1.092	1.110	1.227	1.274
Apr/May/June	1.013	1.092	1.110	1.131	0.959	0.976	1.028	1.310	1.347
May/June/July	1.182	1.005	1.020	1.186	1.095	1.115	1.104	1.065	1.084
June/July/Aug	1.131	0.885	0.893	1.136	0.911	0.916	1.115	1.281	1.290
July/Aug/Sept	1.056	0.866	0.876	1.147	1.191	1.198	1.048	1.439	1.435
Aug/Sept/Oct	1.186*	0.914	0.924	1.200	1.096	1.103	1.007	1.056	1.047
Sept/Oct/Dec	1.087	0.913	0.912	1.179	0.960	0.957	0.938	1.215	1.189
Oct/Nov/Dec	0.981	0.930	0.943	1.018	1.005	1.027	0.994	1.182	1.221
Nov/Dec/Jan	1.172	0,986	0.987	1.26*	0.998	0.992	1.130	1.376	1.036
Dec/Jan/Feb	ref	ref	ref	ref	ref	ref	ref	ref	ref
Infant Characteristics									
Male		0.945	0.951		1.017	1.027		1.002	1.011
Gestational Age		0.841***	0.845***		0.798***	0.799***		0.903***	0.890***
Single Birth		ref	ref		ref	ref		ref	ref
Twins		1.361***	1.298**		1.390***	1.238		0.520**	0.404***
Triplets Plus		2.522*	1.904		3.333**	1.986		1.281	0.510
Maternal Characteristics									
Teens			0.823***			0.876*			0.816
Twenties			ref			ref			ref
Thirties			1.283***			1.494***			1.560***
Forty and over			1.816***			2.060***			1.213
Prenatal 9 visits			0.999			1.026			0.895
Prenatal 10-20 visits			ref			ref			ref
Prenatal 20+ visits			1.364**			1.145			1.097
White			ref			ref			ref
Black			1.115*			1.028			0.539***
Hispanic			1.430***			1.522			1.235
Other			1.227*			1.029			1.279
Less than HS			0.917			0.908			0.813
High School			ref			ref			ref
Some College			1.04			1.149*			1.138
4 plus yrs College			1.273**			1.684***			2.131***
Married			1.210***			1.429***			1.888***
Alcohol			1.040			0.699			1.060
Tobacco			0.705***			0.526***			0.487***
Weight Gain			0.996*			0.999			1.003
N	92021	92021	92021						
LRχ2	174.32	59243.46	66353.89						
df	110	150	300						
prob>χ2	0.0001	0.0000	0.0000						
Pseudo R2	0.0005	0.1570	0.1759						

* p<0.05, **p<0.01, ***p<0.001 (Base Outcome=External Causes)

.

Table 5.3 (Continued) Multinomial Logistic Regression Results (Odds Ratios) Using Gestational Age: Cause Specific Infant Mortality by Months of First Trimester, Infant and Maternal Characteristics, U.S. 2000-2004

Season of First Trimester	Respiratory, Circulatory and Nervous System Disorders		
	Model 1	Model 2	Model 3
Jan/Feb/Mar	1.000	1.000	1.017
Feb/Mar/Apr	1.155	1.143	1.167
Mar/Apr/May	0.941	0.955	0.979
Apr/May/June	0.947	0.945	0.962
May/June/July	1.039	1.017	1.033
June/July/Aug	0.893	0.896	0.897
July/Aug/Sept	1.007	1.016	1.016
Aug/Sept/Oct	1.006	1.020	1.017
Sept/Oct/Dec	1.023	1.027	1.042
Oct/Nov/Dec	1.003	1.019	1.002
Nov/Dec/Jan	1.005	1.013	1.008
Dec/Jan/Feb	ref	ref	ref
Infant Characteristics			
Male		0.999	0.984
Gestational Age		0.881***	0.878***
Single Birth		ref	ref
Twins		1.075	0.904
Triplets Plus		2.657*	1.361
Maternal Characteristics			
Teens			0.770***
Twenties			ref
Thirties			1.601***
Forty and over			2.170***
Prenatal 9 visits			0.95
Prenatal 10-20 visits			ref
Prenatal 20+ visits			1.172
White			ref
Black			0.865*
Hispanic			1.176*
Other			1.108
Less than HS			0.844**
High School			ref
Some College			1.121
4 plus yrs College			1.789***
Married			1.447***
Alcohol			0.902
Tobacco			0.512***
Weight Gain			0.999
N	92021	92021	92021
LRχ2	174.32	59243.46	66353.89
df	110	150	300
prob>χ2	0.0001	0.0000	0.0000
Pseudo R2	0.0005	0.1570	0.1759

* p<0.05, **p<0.01, ***p<0.001 (Base= External Causes)

The results shown in Table 5.3 are similar to those in Table 5.2. Again, the causes of death of SIDS, perinatal infections, neoplasms and blood conditions, and respiratory, circulatory and nervous system causes were all found to have no significant association with any of the month of first trimester variables in any of the three models.

For the prematurity and related causes, however, it is shown again that the months of the first trimester of August, September and October are significantly associated with an increased risk of death from these causes. This significance was maintained in all three models. For model three, the odds ratio is 1.194, indicating that infants whose first trimester was in the months of August, September and October, compared to those whose first trimester was in December, January and February, have a risk of dying of prematurity and related causes that it 19.4 percent higher than for dying of external causes of death.

Also for prematurity and related causes there were significant associations with the months of February, March and April and for May, June and July, but this was only so in model one. When the control variables were introduced in models two and three these months of the first trimester were no longer significant at the $p<0.05$ level. For the months of the first trimester May, June and July the odds ratio in model one was 1.220 and was significant at the 0.01 level. But in models two and three, these months of the first trimester maintained their positive association, but the statistical significance values rose to 0.107 in model two and to 0.08 in model three.

The months of the first trimester August, September and October were significantly associated with congenital anomalies only in models one, two and three. In

.

model one the odds ratio had a z value of 1.94 and a p>|z| of 0.052 and in models two and three these months of the first trimester are significant at the p>0.05 level.

The cause of death of “pregnancy complications” was significantly impacted by the month variable in model one, but lost the significant impact in models two and three. Just as in the models with the birthweight variable, in model one the odds ratio was 1.186, which may be interpreted as indicating that those infants whose first trimester was in the months of August, September and October, compared to those whose first trimester fell onto the months of December, January and February, had a 18.6 percent increased risk of dying of pregnancy complications as compared to external causes of death, and this was significant at the p>0.05 level. But when the control variables were introduced in models two and three, the significant association is lost albeit only barely; indeed both odds ratios are significant at the p>0.1 level with p>|z| values of 0.057 in model two and 0.067 in model three.

The cause of death “birth asphyxia and birth trauma” were positively and significantly associated with the months of the first trimester November, December and January in all three models. In model three the odds ratio is 1.262 meaning that infants whose first trimester fell into the months November, December and January, compared to the months of December, January and February, had a risk of dying of birth asphyxia and birth trauma that was 26.2 percent higher than dying of external causes. Also interesting in the case of birth asphyxia and birth trauma is that in all three models the months of August, September and October although not found to be significant at the p>0.05 level, were approaching significance with p values of 0.059, 0.054 and 0.062

.

respectively. This first trimester may be said to be associated with an increased risk of death from birth asphyxia and birth trauma as compared to external causes at the $p>0.1$ level. The months of the first trimester March/April/May showed a positive and significant association with the causes of death of birth asphyxia and birth trauma only in model three. For models one and two, the odds ratios were not significant at the $p>0.05$ level as they were in model three, although they approached statistical significance.

One last difference between the series of models that included birthweight, as in the previous section, and the models here that included gestational age pertains to the positive and significant association with the months of the first trimester of August/September/October with the cause of death of "other infections," and with the months of the first trimester November/December/January with the cause of death of "endocrine, metabolic and digestive system causes." In both cases the significant association was lost in models two and three when the control variables were introduced.

Results of the Model Including Cases from California (with Gestational Age)

The results of the full model that included CA and the variable gestational age also showed little change from the analyses which excluded CA. Again the r-squared value only drops from 0.1759 to 0.1714, showing only a small drop in the predictive power of the models that exclude the control variables alcohol, tobacco and weight gain. Also like the model which included the birthweight measure only one cause of death—birth asphyxia—was found to be significantly associated with an increased risk of dying for a

.

group of months. This groups of months—March/April/May—was also significant in the original models which excluded CA.

Predicted Probabilities

In Figure 5.2 I now show graphs reflecting the predicted probabilities of cause specific infant mortality based on the predictive model with the prime measure of when the infant was in its first trimester and the other control variables. Again, these graphs are similar to those presented in a previous section of this chapter with the difference that the full model that was used to calculate the predicted probabilities used the variable gestational age instead of the variable birthweight. As we can see the different causes of infant death show different probabilities in the graphs below. This is expected, however, based on the findings in the models show previously. The predicted probability for each cause of death is shown in the two parts of Figure 5.2.

.

Figure 5.2 Predicted Probabilities: Cause of Infant Death and Months of First Trimester (Gestational Age), U.S. 2000-2004

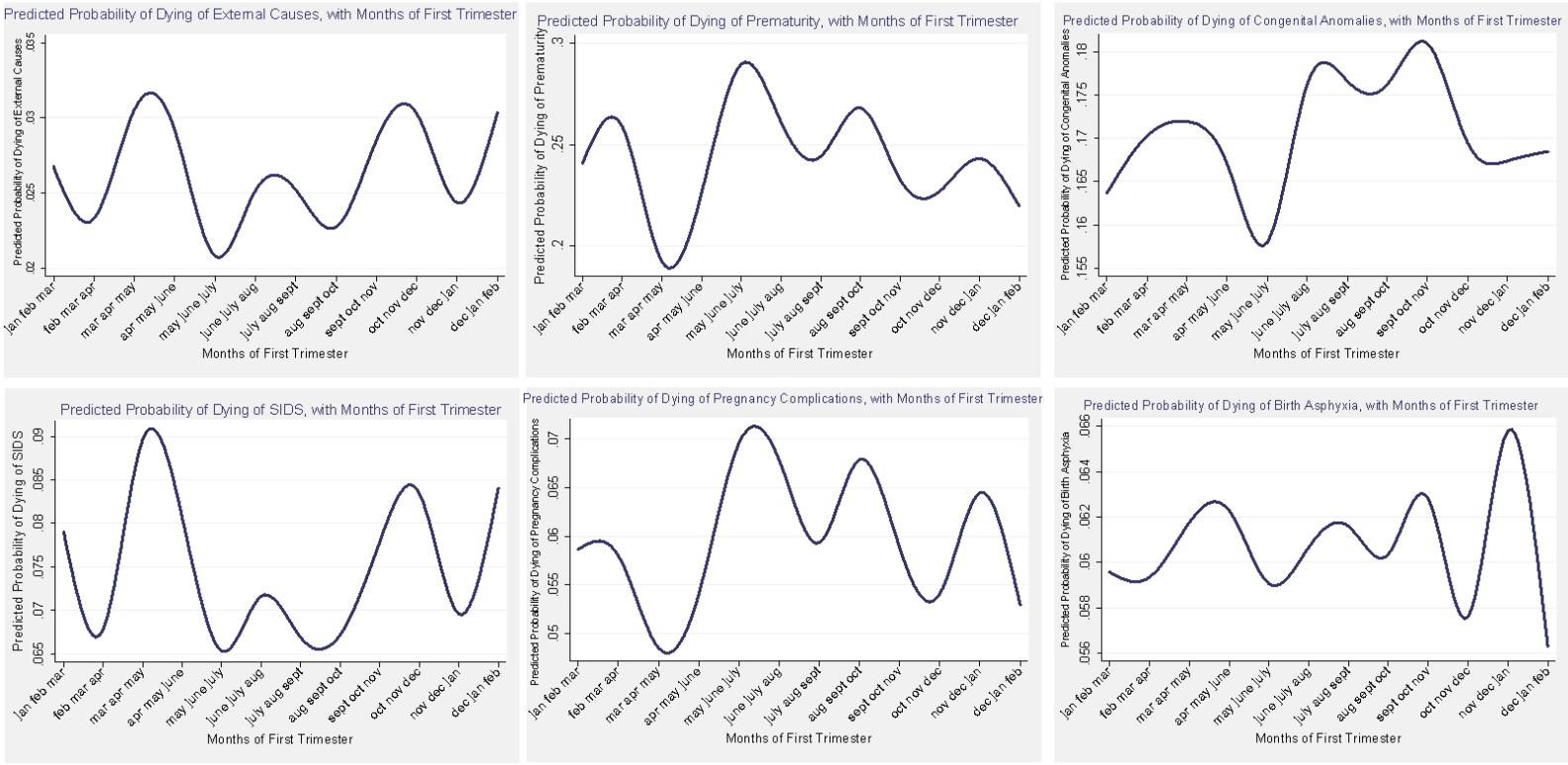
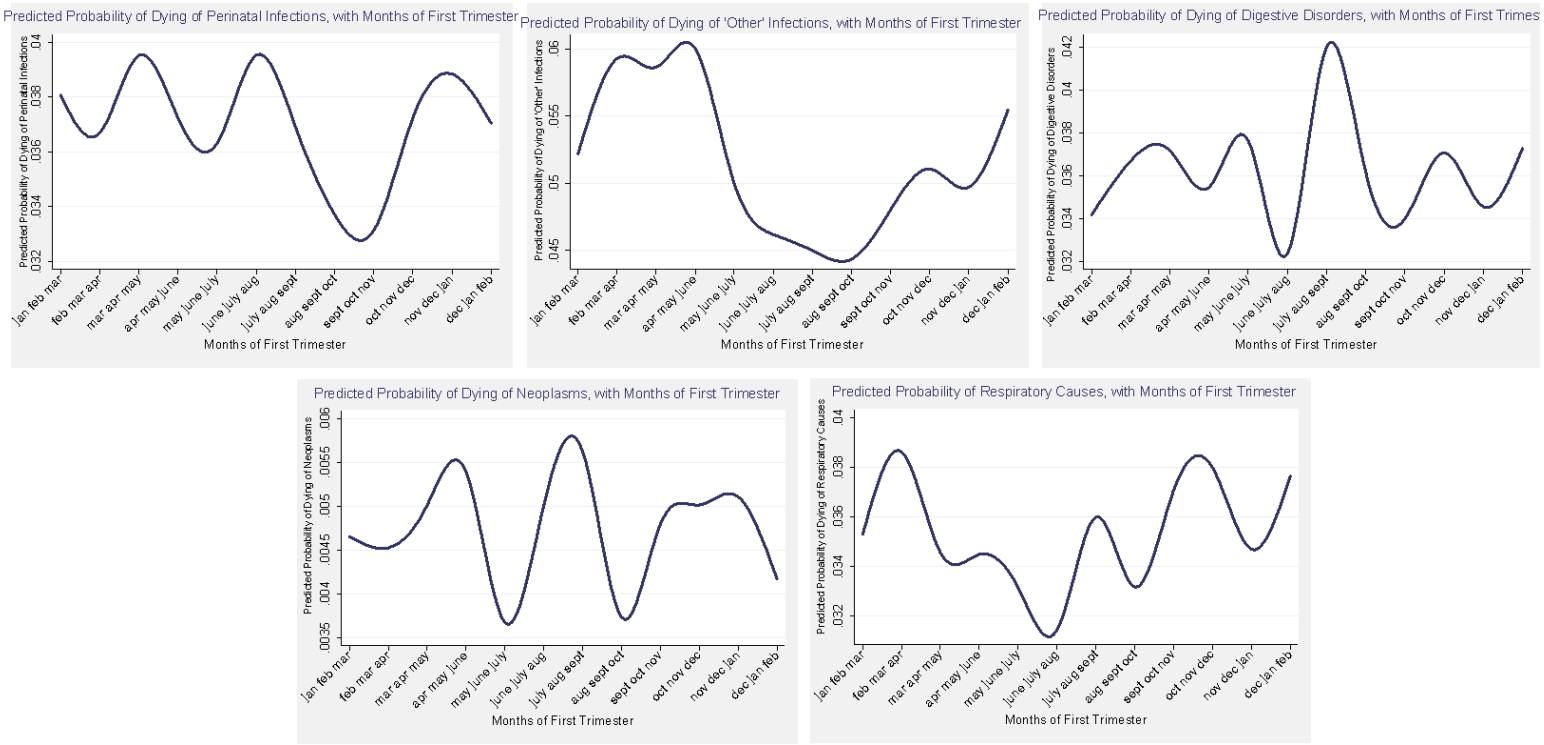


Figure 5.2 (Continued) Predicted Probabilities: Cause of Infant Death and Months of First Trimester (Gestational Age), U.S. 2000-2004



Once again, the predicted probabilities vary for each of the eleven causes of death by the months of the infant's first trimester. And as might be expected these graphs of predicted probabilities look similar to those in Figure 5.2 which used the model with birthweight. And again, the "other infections" cause of death shows the most consistent pattern of higher probabilities in the spring and early summer months, and lowest for the summer and early fall. Also as in Figure 5.2, the highest probabilities are found for the cause of death of prematurity, and the lowest probabilities are found in the cause of death of neoplasms and blood conditions.

Conclusions for 'Months of First Trimester' Models

The models shown and discussed above appear to indicate that the seasonal measure of when the infant was in its first trimester is not the optimum measure for identifying seasonal variation in cause specific infant mortality. This could mean that either the variable that I created did not adequately measure this time of fetal development, or it could mean that the time that the infant was in its first trimester is not the theoretically best time period of interest in terms of the analysis of cause specific infant mortality.

Only the cause of death of "other infections" showed any significant relationship with the season of first trimester variable. Although the findings were maintained after the inclusion of the control variables, the lack of an association in any of the models with any other seasonal measure was not encouraging.

In the next chapter I will undertake the multilevel analyses with the most appropriate independent variable of interest measuring month. In order to be sure that

.

month of birth is the best measure I will first estimate the multinomial logistic regression models with one more series of independent variable—months of third trimester.

Analyses Using Months of Third Trimester as the Principal Independent Variable

The next section of this chapter will show the results and discuss the findings of these models with the "months of third trimester" included as the monthly measure of interest. I will calculate the same models with the same variables used as controls in order to determine the most appropriate measure of the seasonal variation in cause specific infant mortality.

Operationalization

In the literature, the third trimester is considered to be the period that the fetus grows and completes the development that was started in the previous two trimesters. If we consider this time period to be the most important, then, obviously, we would be interested in when the fetus was in the last three months of gestation. In the dataset I use in this dissertation, many of the infants were born before their third trimester. In a restricted data, therefore, only 53,394 of the 92,021 total infants were born at or after twenty-eight weeks gestation. This means that 38,627 of the infants were born in their second trimester (none were born in their first trimester in these data). This made the construction of the variable to measure month of third trimester a little more difficult than the measure of the first trimester. Because so many of the infants were born preterm, it was not as easy as counting backwards to obtain their month of gestation.

Many infants in the data set made it all the way to 37 weeks of gestation or higher which is considered being born at term. Of those that made it to their 28th week of

.

gestation (the first week of the third trimester) many still did not make it to the 37th week mark. Below in Table 5.4 are data showing the numbers of infants who made it to each of the weeks of gestation in the dataset.

Table 5.4 Number of Infants Born at Each Week of Gestation, U.S. 2000-2004

Weeks of Gestation	Number of Infants	Weeks of Gestation	Number of Infants
17	579	**33**	1,919
18	1,097	**34**	2,472
19	1,826	**35**	2,960
20	3,201	**36**	3,676
21	4,292	**37**	5,118
22	5,397	**38**	7,247
23	6,071	**39**	8,121
24	5,703	**40**	6,256
25	4,347	**41**	3,209
26	3,440	**42**	1,408
27	2,674	**43**	733
28	2,212	**44**	432
29	1,919	**45**	171
31	1,732	**47**	71
32	1,692	**Total**	92,021

From Table 5.4 we can see that 7,781 infants made it to the first month of their third trimester (weeks 28 to 31) before birth, 9,043 infants made it to the second month of their third trimester (32 weeks to 35 weeks) before birth, and 32,894 infants were born at 37 weeks and older, therefore making it to at least some of their third trimester and are the only infants in the dataset that would not be considered preterm.

For those infants that were born in the first month of their third trimester creating a variable that measured the month that they experienced their third trimester would essentially be the month of birth variable. Including this measure would only differ slightly from the analyses of chapter IV because it would use the same variables, but

.

would exclude all infants not born in between 28 and 31 weeks of gestation—7,781 infants. [1]

The measure that I decided to create for the third trimester consists of twelve dummy variables that include the possible combinations for the last three months of gestation. This measure will therefore only apply to the 32,894 infants who were born at term and therefore experienced some or all of the last three months of their gestation. These dummy variables will measure the three months that the infant was in its third trimester, such as January/February/March, February/March/April, March/April/May, and so on for all possible three month combinations. The analyses that will be conducted with these "months of third trimester" variables will only include those infants born at or after 37 weeks of gestation.

Again, the operationalization that was discussed in chapter IV regarding the dependent variable of cause of infant death and the independent control variables still applied for these analyses. Thus there is no need for me to discuss again the creation of these variables; such has already been discussed in chapter V. Also as previously seen in this chapter and the analyses of chapter IV, I will discuss but not show the results of the models estimated excluding the variables alcohol, tobacco and weight gain and including California. I will now discuss the hypotheses of the analyses of this section, which use months of the third trimester as the independent variables of interest.

[1] I ran the full model with all controls with the month of birth measure only for those infants born at 28 to 31 weeks gestation. The results showed positive and significant associations of the month of birth January with the neoplasms causes of death and with the month of birth September with the perinatal infections causes of death. The LR chi^2 value was 1528.05 with 300 degrees of freedom and the $p>chi^2$ of 0.0000.

.

Hypotheses

The hypotheses of the models to be estimated with the month of third trimester variable are variations of the hypotheses of the previous models of this chapter and chapter IV. As discussed in the literature, the months of October, November and December were found to positively impact an individual's adult longevity for those 50 and over. If we assume that these individuals were born at term than that would mean that the months of their third trimester would be August, September and October, September, October and November, and October, November and December, respectively. Also the literature on adult longevity consistently showed that being born in the months of April, May and June negatively impacted an individual's longevity. Again, assuming that these individuals were born at term, they would have been in their third trimester in the months of February, March and April, March, April and May and April, May and June respectively. As such I would expect that infants whose third trimester fell into the months between August and December would be at a lower risk of non-external causes of mortality than infants whose third trimester fell into the months of February to June. I expect to find this pattern of positive associations with the months of February to June and negative associations with the months of August to December for all causes of death (compared to the base outcome of external causes). As in the analyses with the months of the first trimester variable reported earlier in chapter V I treated the winter months of December, January and February as the reference months of the third trimester. I chose these months because I expect that other groups of months, compared to these months, will have an increased risk of dying of non-external causes.

.

Descriptive Results

Since the dataset is now restricted to the 32,894 infants who were born after 37 weeks gestation, the descriptive statistics of the dependent and independent variables need not necessarily be the same as the descriptive statistics for infants in the full dataset. Therefore, I show below in Table 5.5 the distribution of the cause of infant death variable for the 32,894 infants born in their third trimester.

Table 5.5 Distribution of Causes of Infant Death for Those Infant Born at 37 Weeks Gestation and Higher, U.S. 2000-2004		
External Causes	4,259	12.95
Prematurity and Related Conditions	1,636	4.97
Congenital Anomalies	9,535	28.99
SIDS, Other Unexplained	9,680	29.43
Pregnancy Complications	542	1.65
Birth Asphyxia and Trauma	1,014	3.08
Perinatal Infections	485	1.47
Other Infections	2,196	6.68
Endocrine, Metabolic and Digestive System Conditions	1,140	3.47
Neoplasms and Blood Conditions	404	1.23
Respiratory, Circulatory and Nervous System Conditions	2,003	6.09
Total	**32,894**	**100.00**

As we can see, when using the restricted data, there is a relative redistribution in the number of cases in each cause of infant death. It makes sense that prematurity and related causes will no longer be the most frequently reported cause of infant death since I have restricted the cases to only those born to term. Now, congenital anomalies are the most frequent cause of infant death, followed closely by SIDS. However, as was the case with the non-restricted sample, neoplasms and blood conditions are the least frequently reported causes of infant death.

.

Next, in Table 5.6 descriptive statistic are presented for the independent control variables measuring infant and maternal characteristics. Again these values have changed somewhat compared to the descriptive statistics based on the full sample. As in previous models the variables male, single birth, twins, triplets plus, teen mom, twenties, thirties, forties and higher, nine or less prenatal visits, ten to twenty prenatal visits, twenty-one and more prenatal visits, white, black, Hispanic, other, married, alcohol and tobacco are all dummy variables with minimum values of zero and maximum values of one. The variables gestational age, birthweight and weight gain are continuous variables. Gestational age ranges from 36 weeks to forty-seven weeks with a mean value of 39.2. Birthweight ranges from 230 grams to 6,521 grams with a mean value of 3070.5 grams. The weight gain variable ranges from zero to 98 pounds and the mean value is 29.7. It is important to note these changes to the independent control variables since restricting the sample, since it shows that restrictions to the deaths may substantively change the infants included in the analyses.

.

Table 5.6 Descriptive Statistics of Independent Control Variables of Chapter V					
Independent Variables	**Minimum**	**Maximum**	**Mean**	**Standard Deviation**	**N**
Infant Characteristics					
Male (1=male, 0=female)	0	1	0.514	0.499	32894
Birthweight (in grams)	230	6521	3070.509	645.537	32894
Gestational Age (in weeks)	36	47	39.212	1.721	32894
Plurality					
Single Birth (1=yes, 0=no)	0	1	0.972	0.162	32894
Twins (1=yes, 0=no)	0	1	0.026	0.160	32894
Triplets plus (1=yes, 0=no)	0	1	0.000	0.025	32894
Maternal Characteristics					
Age of Mother					
Teen Mom (1=yes, 0=no)	0	1	0.172	0.377	32894
Twenties (1=yes, 0=no)	0	1	0.561	0.496	32894
Thirties (1=yes, 0=no)	0	1	0.24	0.427	32894
Forties plus (1=yes, 0=no)	0	1	0.025	0.158	32894
Number of Prenatal Care Visits					
Nine or less (1=yes, 0=no)	0	1	0.295	0.456	32894
Ten to Twenty (1=yes, 0=no)	0	1	0.686	0.463	32894
Twenty-One plus (1=yes, 0=no)	0	1	0.042	0.200	32894
Mother's Race/Ethnicity					
White (1=yes, 0=no)	0	1	0.582	0.493	32894
Black (1=yes, 0=no)	0	1	0.229	0.420	32894
Hispanic (1=yes, 0=no)	0	1	0.144	0.351	32894
Other (1=yes, 0=no)	0	1	0.044	0.205	32894
Mother's Education					
Less than HS (1=yes, 0=no)	0	1	0.309	0.462	32894
High School (1=yes, 0=no)	0	1	0.361	0.480	32894
Some College (1=yes, 0=no)	0	1	0.182	0.385	32894
Four plus years of College (1=yes, 0=no)	0	1	0.146	0.353	32894
Mother's Marital Status					
Married (1=married, 0=non-married)	0	1	0.514	0.499	32894
Mother's Tobacco Use					
Tobacco (1=used tobacco while pregnant, 0=did not)	0	1	0.230	0.420	32894
Mother's Alcohol Use					
Alcohol (1=used alcohol while pregnant, 0=did not)	0	1	0.012	0.112	32894
Mother's Weight Gain					
Weight Gain (In pounds)	0	98	29.699	14.662	32894

Below in Table 5.7 I are the descriptive results for the cause of infant death dependent variable and the variable "months of third trimester" to be used in these analyses. The operationalization for the dependent variable—cause of infant death, as well as for the independent control variables measuring infant and maternal characteristics are the same as found in chapter IV . Again as in the previous models, the cause of infant death variable ranges from zero to ten with a mean value of 3.284. But since this is a nominal variable with eleven unordered categories, this mean value does not have any statistically meaningful interpretation. The months of third trimester variables are all dummy variables which range from zero to one.

Table 5.7 Descriptive Statistics of Cause of Infant Death and Month of Birth Variables for Infants Born at 37 Weeks of Gestation and Higher, U.S. 2000-2004

Dependent Variable	Minimum	Maximum	Mean	Standard Deviation	N
Cause of Infant Death					
11 categories (0-10)	0	10	3.284	2.684	32894
Months of Third Trimester					
12 dummy variables					
0=third trimester in those months					
1= third trimester not in those months					
Jan/Feb/Mar	0	1	0.080	0.271	32894
Feb/Mar/Apr	0	1	0.075	0.263	32894
Mar/Apr/May	0	1	0.081	0.273	32894
Apr/May/June	0	1	0.081	0.273	32894
May/June/July	0	1	0.085	0.280	32894
June/July/Aug	0	1	0.090	0.287	32894
July/Aug/Sept	0	1	0.089	0.284	32894
Aug/Sept/Oct	0	1	0.082	0.275	32894
Sept/Oct/Nov	0	1	0.085	0.278	32894
Oct/Nov/Dec	0	1	0.089	0.285	32894
Nov/Dec/Jan	0	1	0.083	0.277	32894
Dec/Jan/Feb	0	1	0.074	0.263	32894

.

Multicollinearity Diagnostics

I next calculated a zero-order correlation matrix of the independent variables in the model, in order to find out which, if any, of the variables may pose problems due to collinearity. I suspected from my previous findings regarding multicollinearity that were presented earlier in this chapter and in chapter IV that the variable birthweight and gestational age might well pose problems if included in the same analyses. My results here revealed, just as suspected, the variable birthweight and gestational age were highly correlated with one another, suggesting that I would need to estimate two series of models with the months of third trimester variable in order to overcome the collinearity problem. Also as in earlier investigations of collinearity, the series of dummy variables measuring plurality, age, prenatal visits, race/ethnicity and education all show high correlations when all measures are included. Again, I will leave out the variables twenties, ten to twenty prenatal visits, white, and high school education when estimating my models.

I also calculated tolerance values for all the independent variables. The resulting tolerance values also give support to my exclusion of one variable in each of the dummy series measuring infant or maternal characteristics as well as undertaking two analyses—one with birthweight, and one with gestational age. All the other independent variables had tolerance values of 0.50 or above, suggesting little danger including them together in the model.

.

Multinomial Logistic Regression Results (Using Birthweight but Not Gestational Age)

Below in Table 5.8 are the results of the first series of models using the months of third trimester variable with the infant variable of birthweight and the other controls. An identical model will be estimated in a later section of this chapter with the gestational age variable.

When estimating the model with the infant characteristics variable of "triplets plus," where a value of one indicates that the infant was one of a set of triplet infants or higher plurality, the odds ratios for the variable were very large, and in some other cases were omitted from the results. This is because of those infants who gestated for 37 weeks or more, only 21 were triplets or higher. Thus, some of the causes of death do not have values in Table 5.8 for the 'triplets and higher' variable since of those infants born at 37 weeks or higher, there were no infants of triplets or higher plurality that died from that specified cause. I could have deleted the triplet infants from the analysis, but opted not to do so; there are so few cases that keeping or deleting them will not seriously impact the results. Moreover, I was not comfortable with a decision to delete so-called "troublesome" cases.

.

Table 5.8 Multinomial Logistic Regression Results (Odds Ratios) Using Birthweight: Cause Specific Infant Mortality by Months of Third Trimester, Infant and Maternal Characteristics, U.S. 2000-2004

Months of Third Trimester	Prematurity and Related Conditions			Congenital Anomalies			SIDS		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Jan/Feb/Mar	1.167	1.175	1.148	1.052	1.042	1.005	0.921	0.922	0.923
Feb/Mar/Apr	1.136	1.153	1.158	1.184	1.200	1.208	1.007	1.010	1.013
Mar/Apr/May	0.953	0.956	0.931	1.076	1.079	1.038	0.955	0.955	0.955
Apr/May/June	1.104	1.106	1.072	1.275*	1.259*	1.212	1.149	1.146	1.143
May/June/July	0.996	1.004	0.994	1.088	1.082	1.078	1.002	1.002	1.004
June/July/Aug	1.121	1.128	1.116	1.126	1.140	1.128	1.127	1.127	1.128
July/Aug/Sept	1.034	1.060	1.050	1.075	1.099	1.094	1.001	1.004	1.005
Aug/Sept/Oct	0.996	1.005	1.005	1.050	1.056	1.058	1.070	1.070	1.070
Sept/Oct/Nov	0.905	0.923	0.931	1.087	1.088	1.094	1.028	1.028	0.129
Oct/Nov/Dec	0.807	0.795	0.792	1.093	1.081	1.073	1.072	1.070	1.068
Nov/Dec/Jan	1.321*	1.322*	1.353*	1.170	1.150	1.179	1.087	1.083	1.081
Dec/Jan/Feb	ref	ref	ref	ref	ref	ref	ref	ref	ref
Infant Characteristics									
Male		1.022	1.039		1.002	1.021		1.035	1.036
Birthweight		0.999***	0.999**		0.998***	0.998*		0.999***	0.999***
Single Birth		ref	ref		ref	ref		ref	ref
Twins		2.246***			0.532***	0.356*		1.003	0.990
Triplets Plus		2.71e+10***			9.55e+08***	6.04e+08***		1.64e+09***	2.20e+09***
Maternal Characteristics									
Teens			0.818			0.704***			0.916
Twenties			ref			ref			ref
Thirties			1.651**			1.745***			0.968
Forty and over			1.643*			2.991***			0.760
Prenatal 9 visits			0.965			0.859***			1.015
Prenatal 10-20 visits			ref			ref			ref
Prenatal 20+ visits			1.392*			1.945***			0.983
White			ref			ref			ref
Black			0.865			0.698***			0.999
Hispanic			1.271**			1.316***			0.892
Other			0.920			0.807*			0.863
Less than HS			0.821*			0.895*			0.989
High School			ref			ref			ref
Some College			1.088			1.151*			1.046
4 plus yrs College			1.732***			1.739***			1.185*
Married			1.462***			1.729***			1.060
Alcohol			1.690*			0.742			0.918
Tobacco			0.390***			0.316**			1.156***
Weight Gain			1.000			1.004***			1.000
N	32894	32894	32894						
LRχ2	153.84	2713.37	7682.77						
df	110	150	300						
prob>χ2	0.0037	0.0000	0.0000						
Pseudo R2	0.0012	0.0218	0.0617						

* p<0.05, **p<0.01, ***p<0.001 (Base Outcome=External Causes)

.

Table 5.8 (Continued) Multinomial Logistic Regression Results (Odds Ratios) Using Birthweight: Cause Specific Infant Mortality by Months of Third Trimester, Infant and Maternal Characteristics, U.S. 2000-2004

Season of First Trimester	Pregnancy Complications			Birth Asphyxia and Birth Trauma			Perinatal Infections		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Jan/Feb/Mar	0.714	0.720	0.704	1.069	1.075	1.048	1.498	1.496	1.470
Feb/Mar/Apr	1.187	1.206	1.220	1.137	1.139	1.144	1.578	1.589	1.569
Mar/Apr/May	1.260	1.269	1.230	1.016	1.018	0.979	1.193	1.193	1.154
Apr/May/June	0.886	0.889	0.865	1.303	1.307	1.288	1.721	1.716*	1.651*
May/June/July	0.915	0.923	0.928	1.060	1.065	1.065	1.344	1.348	1.324
June/July/Aug	0.879	0.886	0.872	0.981	0.984	0.976	1.445	1.447	1.430
July/Aug/Sept	0.669	0.684	0.680	1.040	1.039	1.038	1.113	1.129	1.103
Aug/Sept/Oct	0.970	0.974	0.978	0.946	0.946	0.952	1.135	1.134	1.314
Sept/Oct/Nov	0.999	1.021	1.035	0.958	0.967	0.986	0.820	0.824	0.816
Oct/Nov/Dec	0.814	0.813	0.808	1.175	1.180	1.168	1.254	1.239	1.231
Nov/Dec/Jan	0.860	0.860	0.883	1.226	1.231	1.267	1.630*	1.617*	1.643*
Dec/Jan/Feb	ref	ref	ref	ref	ref	ref	ref	ref	ref
Characteristics									
Male		1.081	1.099		0.985	1.005		1.034	1.043
Birthweight		0.999***	0.999***		1.000**	0.999		0.999***	0.999***
Single Birth		ref	ref		ref	ref		ref	ref
Twins		2.340***	1.744*		2.166***	1.477		1.377	1.151
Triplets Plus		1.12e+10***	--		--	--		--	--
Maternal Characteristics									
Teens			0.619**			0.778*			1.285
Twenties			ref			ref			ref
Thirties			1.573***			1.929***			1.132
Forty and over			1.78			2.957***			1.670
Prenatal 9 visits			0.982			0.889			1.090
Prenatal 10-20 visits			ref			ref			ref
Prenatal 20+ visits			0.839			0.862			0.985
White			ref			ref			ref
Black			0.714*			0.739**			0.777
Hispanic			1.25			1.033			1.495***
Other			0.676			0.676**			0.814
Less than HS			0.814			0.824			0.589***
High School			ref			ref			ref
Some College			1.168			1.229*			0.996
4 plus yrs College			2.052***			1.879***			1.420*
Married			1.490***			1.524***			1.486***
Alcohol			0.568			0.316*			1.168
Tobacco			0.492***			0.456***			0.544***
Weight Gain			0.997			1.002			1.002
N	32894	32894	32894						
LRχ2	153.84	2713.37	7682.77						
df	110	150	300						
prob>χ2	0.0037	0.0000	0.0000						
Pseudo R2	0.0012	0.0218	0.0617						

* p<0.05, **p<0.01, ***p<0.001 (Base Outcome=External Causes)

Table 5.8 (Continued) Multinomial Logistic Regression Results (Odds Ratios) Using Birthweight: Cause Specific Infant Mortality by Months of Third Trimester, Infant and Maternal Characteristics, U.S. 2000-2004

Season of First Trimester	Other Infections			Endocrine, Metabolic and Digestive System Disorders			Neoplasms and Blood Conditions		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Jan/Feb/Mar	0.972	0.973	0.960	1.090	1.052	1.025	1.107	1.109	1.095
Feb/Mar/Apr	1.011	1.018	1.016	1.395*	1.376	1.377	0.960	0.952	0.958
Mar/Apr/May	0.939	0.941	0.926	0.913	0.894	0.865	0.794	0.788	0.759
Apr/May/June	1.372*	1.369*	1.338*	1.090	1.151	1.120	1.129	1.129	1.090
May/June/July	1.217	1.219	1.206	0.984	1.000	0.996	0.858	0.857	0.858
June/July/Aug	1.304*	1.308*	1.294	1.033	1.051	1.039	1.236	1.230	1.211
July/Aug/Sept	1.303*	1.314	1.295*	1.045	1.084	1.075	0.878	0.871	0.856
Aug/Sept/Oct	1.435**	1.438**	1.423**	1.062	1.038	1.038	1.091	1.087	1.088
Sept/Oct/Nov	1.486**	1.488**	1.475**	1.045	1.062	1.063	1.180	1.177	1.182
Oct/Nov/Dec	1.288	1.281	1.268	1.118	1.084	1.072	0.949	0.949	0.949
Nov/Dec/Jan	1.459**	1.455**	1.467**	1.165	1.174	1.195	1.002	1.001	1.028
Dec/Jan/Feb	ref	ref	ref	ref	ref	ref	ref	ref	ref
Infant Characteristics									
Male		0.969	0.977		1.052	1.069		1.099	1.122
Birthweight		0.999***	0.999***		0.999***	0.999***		1.000*	0.999
Single Birth		ref	ref		ref	ref		ref	ref
Twins		1.051	0.932		0.671	0.493**		0.725	0.513
Triplets Plus		3.16e+09***	3.69e+09***		3.97e+09***	3.49e+09***		--	--
Maternal Characteristics									
Teens			0.836*			0.772**			0.937
Twenties			ref			ref			ref
Thirties			1.14			1.490***			1.507***
Forty and over			1.249			2.047**			0.640
Prenatal 9 visits			0.959			0.887			0.825
Prenatal 10-20 visits			ref			ref			ref
Prenatal 20+ visits			1.401*			1.104			1.037
White			ref			ref			ref
Black			0.892			0.686***			0.569***
Hispanic			1.426***			1.218			1.413*
Other			1.274*			1.017			0.964
Less than HS			0.931			1.069			0.852
High School			ref			ref			ref
Some College			1.035			1.286**			1.181
4 plus yrs College			1.215			1.695***			2.458***
Married			1.158*			1.423***			1.848**
Alcohol			1.045			0.421*			0.419
Tobacco			0.718***			0.438***			0.550***
Weight Gain			0.998			1.003			1.000
N	32894	32894	32894						
LRχ2	153.84	2713.37	7682.77						
df	110	150	300						
prob>χ2	0.0037	0.0000	0.0000						
Pseudo R2	0.0012	0.0218	0.0617						

* p<0.05, **p<0.01, ***p<0.001 (Base Outcome=External Causes)

.

Table 5.8 (Continued) Multinomial Logistic Regression Results (Odds Ratios) Using Birthweight: Cause Specific Infant Mortality by Months of Third Trimester, Infant and Maternal Characteristics U.S. 2000-2204

Months of First Trimester	Respiratory, Circulatory and Nervous System Disorders		
	Model 1	Model 2	Model 3
Jan/Feb/Mar	0.945	0.946	0.926
Feb/Mar/Apr	1.187	1.192	1.195
Mar/Apr/May	1.246	1.246	1.207
Apr/May/June	1.344*	1.338*	1.301
May/June/July	1.164	1.163	1.156
June/July/Aug	1.094	1.095	1.085
July/Aug/Sept	1.064	1.069	1.059
Aug/Sept/Oct	1.133	1.133	1.135
Sept/Oct/Dec	1.227	1.222	1.232
Oct/Nov/Dec	1.019	1.016	1.009
Nov/Dec/Jan	1.405*	1.399*	1.432**
Dec/Jan/Feb	ref	ref	ref
Infant Characteristics			
Male		1.076	1.096
Birthweight		0.999***	0.999***
Single Birth		ref	ref
Twins		0.696	0.495***
Triplets Plus		3.46e+09***	2.98e+09***
Maternal Characteristics			
Teens			0.701***
Twenties			ref
Thirties			1.665***
Forty and over			1.643*
Prenatal 9 visits			0.889
Prenatal 10-20 visits			ref
Prenatal 20+ visits			1.081
White			ref
Black			0.827**
Hispanic			1.107
Other			1.102
Less than HS			0.861*
High School			ref
Some College			1.091
4 plus yrs College			1.670***
Married			1.475***
Alcohol			1.024
Tobacco			0.429***
Weight Gain			1.002
N	32894	32894	32894
LRχ2	153.84	2713.37	7682.77
df	110	150	300
prob>χ2	0.0037	0.0000	0.0000
Pseudo R2	0.0012	0.0218	0.0617

* p<0.05, **p<0.01, ***p<0.001 (Base= External Causes)

The results of the models shown in Table 5.8 show that the causes of infant death of prematurity and related causes, SIDS, pregnancy complications, birth asphyxia and birth trauma and neoplasms and blood conditions had no significant associations with the month of third trimester variables. The causes of death of congenital anomalies and perinatal infections show a significant and positive association with the months of the third trimester April/May/June. However, this significant association is not sustained in model three when all the infant and maternal characteristics are entered as controls.

The causes of death of endocrine, metabolic and digestive system disorders show a positive and significant association with the months of the third trimester February/March/April. In model three the odds ratio is 1.410, which means that those infants whose third trimester falls in the months February/March/April, compared to the months of December/January/February, have a risk of dying of endocrine, metabolic and digestive system causes that is 41.0 percent higher than that of dying of external causes of death.

The cause of death of respiratory, circulatory and nervous system disorders also shows a significant and positive association with one group of months, namely, November/December/January. Its odds ratio means that infants that experience their third trimester in the months of November/December/January, compared to the third trimester in the months of December/January/February, have a risk of dying of respiratory causes that is 36.4 percent higher than dying of external causes of death.

The cause of death of “other” infections shows a positive and significant association with the months of the third trimester July/August/September,

.

August/September/October, September/October/November, October/November/December and November/December/January. In all three models these months of the third trimester maintained significant and positive relationships with the respiratory cause of death. These results suggest that when the third trimester occurs in the months of the late summer, fall and early winter, there is a higher risk of dying of other infections, compared to external causes for infants that are born at thirty-seven or more weeks.

Results of the Model Including Cases from California (with Birthweight)

Next, I re-estimated the models excluding the variables alcohol, tobacco and weight gain so that the state of California would be included. For the model which included the variable birthweight, the causes of death prematurity, SIDS, pregnancy complications and neoplasms were not significantly associated with any of the months of third trimester variables. It is also important to note that the r-squared value was very low in both types of models—only 0.0617 in the models that include weight gain, alcohol and tobacco variables and dropping to 0.0467 in the models that exclude these variables and therefore include the state of California.

The cause of death congenital anomalies shows the most change in this model which includes California when compared to the model which excludes California. Three months of third trimester variables—Nov/Dec/Jan, Feb/Mar/Apr and Apr/May/June were all shown to be significantly associated with the risk of dying of congenital anomalies in the model which includes California but were not found to be significant in the models that excluded California A in the original chapter V analyses.

.

The cause of death birth asphyxia was found to be significantly associated with the months Apr/May/June in the model with California but was not found to be significant in the model that excluded California. The cause of death perinatal infections were found to be significantly associated with the months of third trimester Nov/Dec/Jan and April/May/June in both the model that included California and the model that excluded California. The cause of death other infections was found to be significantly associated with several of the months of third trimester variables—Nov/Dec/Jan, April/May/June, June/July/Aug, July/Aug/Sept, Aug/Sept/Oct, September/Oct/Nov and October/November/December. The months of Oct/Nov/Dec were not found to be significant only in the model which included California as was the months of June/July/August. However the months of June/July/Aug were significantly associated in the model that excluded California in models one and two. Lastly, the cause of death endocrine and digestive disorder was found to be significantly associated with the months of Feb/Mar/Apr in the models that included California but not in the model that excluded California. Lastly, the cause of death respiratory conditions was found to be significantly associated with the months of April/May/June only in the models that included California, however in the models that excluded California these months of birth were significantly associated with respiratory conditions in models one and two.

.

Predicted Probabilities

Again, as in pervious chapters, I show the predicted probabilities for the model using months of third trimester. Looking at these graphs we can see that there are several differences in the models of this chapter when compared to previous chapters. Owing to the fact that the analyses reported here are based on a restricted number of cases, the predicted probabilities seen below in Figure 5.3 for this multinomial model are somewhat different from those shown earlier in this chapter and in chapter IV. We especially see these differences in the prematurity causes of death because we have restricted the observations to infants born to term. The probability of dying of prematurity appears to gradually decline for the later winter, spring and fall months, dipping to its lowest value in the third trimester October, November and December then rising sharply to its highest value in November, December and January.

.

Figure 5.3 Predicted Probabilities: Cause of Infant Death and Months of Third Trimester (Birthweight), U.S. 2000-2004

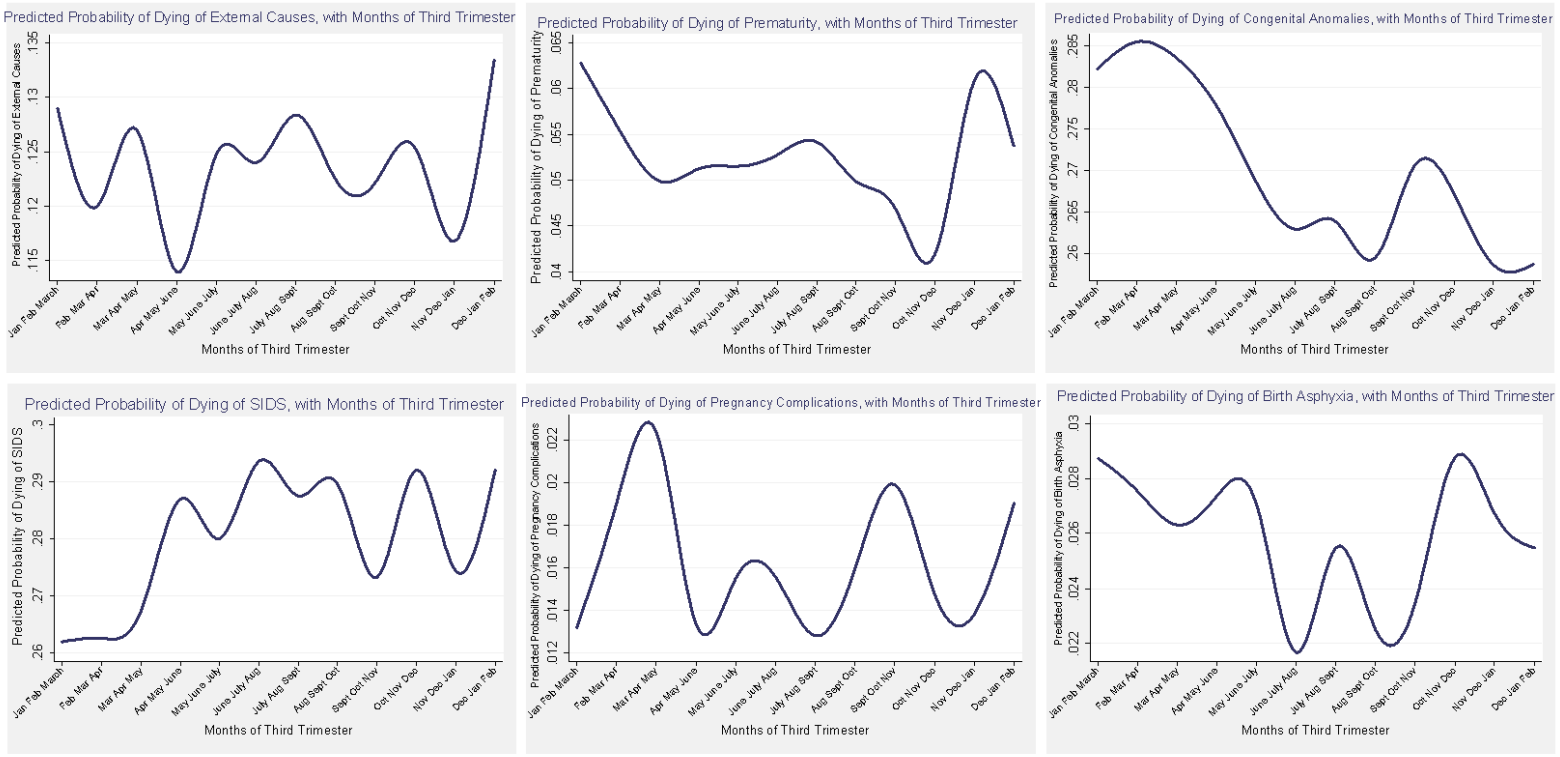
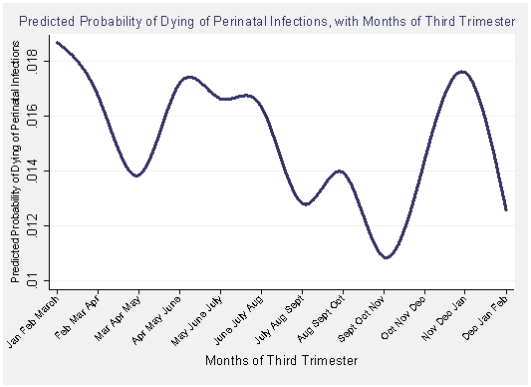
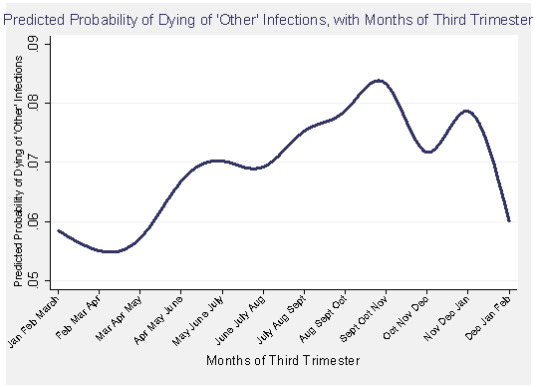
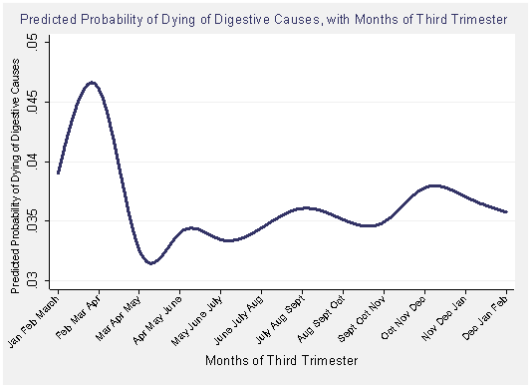
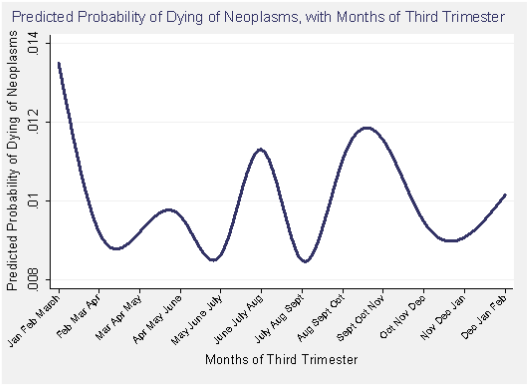


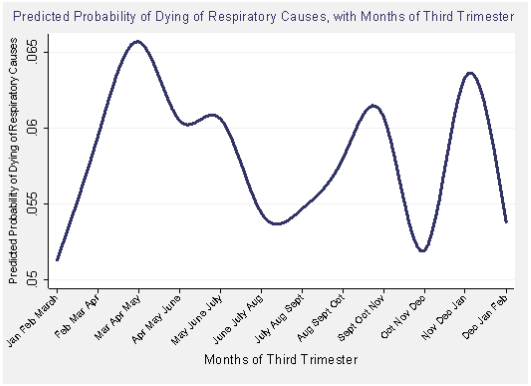
Figure 5.3 (Continued) Predicted Probabilities: Cause of Infant Death and Months of Third Trimester (Birthweight), U.S. 2000-2004











Multinomial Logistic Regression Results (Using Gestational Age but Not Birthweight)

I now present the results of the multinomial logistic regression model which includes the variable “gestational age” instead of birthweight as in the previous model. Again, this is due to multicollinearity between the birthweight and gestational age variables. This necessitates the use of two series of models in order to see the influence of both of these very important independent control variables. The results are shown below in Table 5.9 As with the results using other principal independent variables of interest; the findings shown in Table 5.9 have few significant relationships in the hypothesized direction. However, where significant relationships are observed, the patterns are similar to the findings of the multinomial logistic regression model which included the birthweight variable instead of the gestational age variable as seen below.

.

Table 5.9 Multinomial Logistic Regression Results (Odds Ratios) Using Gestational Age: Cause Specific Infant Mortality by Months of Third Trimester, Infant and Maternal Characteristics, U.S. 2000-2004

Months of Third Trimester	Prematurity and Related Conditions			Congenital Anomalies			SIDS		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Jan/Feb/Mar	1.167	1.171	1.154	1.052	1.058	1.020	0.921	0.912	0.924
Feb/Mar/Apr	1.136	1.122	1.147	1.184	1.181	1.207	1.007	0.990	1.014
Mar/Apr/May	0.953	0.939	0.933	1.076	1.084	1.048	0.955	0.942	0.957
Apr/May/June	1.104	1.103	1.086	1.275*	1.211*	1.239*	1.149	1.138	1.147
May/June/July	0.996	1.012	1.001	1.088	1.079	1.089	1.002	1.016	1.005
June/July/Aug	1.121	1.058	1.107	1.126	1.094	1.111	1.127	1.086	1.127
July/Aug/Sept	1.034	1.057	1.033	1.075	1.067	1.061	1.001	1.030	1.004
Aug/Sept/Oct	0.996	0.999	0.999	1.050	1.077	1.043	1.070	1.075	1.070
Sept/Oct/Nov	0.905	0.931	0.933	1.087	1.091	1.094	1.028	1.025	1.028
Oct/Nov/Dec	0.807	0.823	0.799	1.093	1.094	1.088	1.072	1.063	1.072
Nov/Dec/Jan	1.321*	1.269*	1.361*	1.170	1.119	1.205	1.087	1.073	1.084
Dec/Jan/Feb	ref	ref	ref	ref	ref	ref	ref	ref	ref
Infant Characteristics									
Male		1.007	1.015		0.964	0.980		1.032	1.030
Gestational Age		0.932***	1.002		0.873***	0.916***		0.973***	0.975***
Single Birth		ref	ref		ref	ref		ref	ref
Twins		3.206***	3.042***		1.276*	1.103		1.210	1.099
Triplets Plus		5.02e+10***	5.52e+10***		1.12e+10***	3.88e+09***		2.49e+09***	2.49e+09***
Maternal Characteristics									
Teens			0.857			0.759***			0.926
Twenties			ref			ref			ref
Thirties			1.598***			1.735***			0.961
Forty and over			1.807**			3.702***			0.769
Prenatal 9 visits			1.004			0.935			1.019
Prenatal 10-20 visits			ref			ref			ref
Prenatal 20+ visits			1.405*			1.981***			0.978
White			ref			ref			ref
Black			0.934			0.803***			1.015
Hispanic			1.303			1.388***			0.896
Other			0.927			0.833			0.861
Less than HS			0.818*			0.890*			0.990
High School			ref			ref			ref
Some College			1.066			1.104			1.041
4 plus yrs College			1.697***			1.650***			1.175*
Married			1.425***			1.644***			1.054
Alcohol			1.749*			0.772			0.926
Tobacco			0.433***			0.378***			1.183***
Weight Gain			0.995*			0.995***			0.999
N	32894	32894	32894						
LRχ2	153.84	499.91	5104.01						
df	110	150	300						
prob>χ2	0.0037	0.0000	0.0000						
Pseudo R2	0.0012	0.004	0.041						

* p<0.05, **p<0.01, ***p<0.001 (Base Outcome=External Causes)

Table 5.9 (Continued) Multinomial Logistic Regression Results (Odds Ratios) Using Gestational Age: Cause Specific Infant Mortality by Months of Third Trimester, Infant and Maternal Characteristics, U.S. 2000-2004

Season of First Trimester	Pregnancy Complications			Birth Asphyxia and Birth Trauma			Perinatal Infections		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Jan/Feb/Mar	0.714	0.691	0.707	1.069	1.113	1.058	1.498	1.492	1.480
Feb/Mar/Apr	1.187	1.084	1.209	1.137	1.147	1.144	1.578	1.458	1.576
Mar/Apr/May	1.260	1.197	1.233	1.016	1.037	0.986	1.193	1.115	1.160
Apr/May/June	0.886	0.806	0.876	1.303	1.221	1.298	1.721	1.576*	1.671*
May/June/July	0.915	0.858	0.925	1.060	1.103	1.080	1.344	1.387	1.329
June/July/Aug	0.879	0.886	0.867	0.981	0.915	0.977	1.445	1.406	1.427
July/Aug/Sept	0.669	0.707	0.672	1.040	1.042	1.046	1.113	1.077	1.091
Aug/Sept/Oct	0.970	0.904	0.978	0.946	0.947	0.962	1.135	1.189	1.139
Sept/Oct/Nov	0.999	1.115	1.035	0.958	0.988	0.995	0.820	0.929	0.819
Oct/Nov/Dec	0.814	0.824	0.811	1.175	1.191	1.172	1.254	1.218	1.247
Nov/Dec/Jan	0.860	0.822	0.887	1.226	1.158	1.273	1.630*	1.571*	1.659*
Dec/Jan/Feb	ref	ref	ref	ref	ref	ref	ref	ref	ref
Infant Characteristics									
Male		1.093	1.080		1.013	1.004		1.083	1.019
Gestational Age		0.949*	1.028		1.012	1.059**		0.880***	0.938*
Single Birth		ref	ref		ref	ref		ref	ref
Twins		2.821***	2.763***		2.049***	1.673*		1.747*	1.977**
Triplets Plus		--	--		--	--		1.58e+10***	2.55e+10***
Maternal Characteristics									
Teens			0.640**			0.781*			1.344*
Twenties			ref			ref			ref
Thirties			1.560***			1.943***			1.118
Forty and over			1.912*			3.025***			1.825
Prenatal 9 visits			1.007			0.893			1.127
Prenatal 10-20 visits			ref			ref			ref
Prenatal 20+ visits			0.852			0.874			0.984
White			ref			ref			ref
Black			0.753*			0.738**			0.835
Hispanic			1.271			1.029			1.526**
Other			0.673			0.656*			0.810
Less than HS			0.812			0.820*			0.593***
High School			ref			ref			ref
Some College			1.154			1.231*			0.978
4 plus yrs College			2.029***			1.892***			1.383
Married			1.464***			1.529***			1.444***
Alcohol			0.583			0.324*			1.213
Tobacco			0.531***			0.456***			0.601***
Weight Gain			0.994			1.002			0.998
N	32894	32894	32894						
LRχ2	153.84	499.91	5104.01						
df	110	150	300						
prob>χ2	0.0037	0.0000	0.0000						
Pseudo R2	0.0012	0.004	0.041						

* p<0.05, **p<0.01, ***p<0.001 (Base Outcome=External Causes)

.

Table 5.9 (Continued) Multinomial Logistic Regression Results (Odds Ratios) Using Gestational Age: Cause Specific Infant Mortality by Months of Third Trimester, Infant and Maternal Characteristics, U.S. 2000-2004

Season of First Trimester	Other Infections			Endocrine, Metabolic and Digestive System Disorders			Neoplasms and Blood Conditions		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Jan/Feb/Mar	0.972	0.969	0.961	1.058	1.090	1.036	1.107	1.298	1.098
Feb/Mar/Apr	1.011	0.988	1.018	1.369	1.409*	1.381	0.960	0.974	0.972
Mar/Apr/May	0.939	0.959	0.927	0.897	0.915	0.872	0.794	0.897	0.765
Apr/May/June	1.372*	1.284*	1.346*	1.159	1.089	1.133	1.129	1.065	1.096
May/June/July	1.217	1.233	1.207	1.005	0.983	1.005	0.858	0.894	0.865
June/July/Aug	1.304*	1.245*	1.290	1.048	1.033	1.033	1.236	1.195	1.220
July/Aug/Sept	1.303*	1.308*	1.284	1.073	1.048	1.061	0.878	0.872	0.867
Aug/Sept/Oct	1.435**	1.416**	1.421**	1.038	1.064	1.036	1.091	1.166	1.104
Sept/Oct/Nov	1.486**	1.490***	1.475**	1.067	1.043	1.069	1.180	1.201	1.182
Oct/Nov/Dec	1.288	1.268	1.278	1.095	1.119	1.087	0.949	0.975	0.959
Nov/Dec/Jan	1.459**	1.470**	1.472**	1.184	1.167	1.208	1.002	0.994	1.032
Dec/Jan/Feb	ref	ref	ref	ref	ref	ref	ref	ref	ref
Infant Characteristics									
Male		0.969	0.964		1.033	1.044		1.103	1.121
Gestational Age		0.935***	0.954**		0.920***	0.952*		0.931**	0.946
Single Birth		ref	ref		ref	ref		ref	ref
Twins		1.262	1.213		1.145	0.886		0.628	0.514
Triplets Plus		3.51e+09***	5.18e+09***		--	--		--	--
Maternal Characteristics									
Teens			0.856*			0.809*			0.943
Twenties			ref			ref			ref
Thirties			1.126			1.465***			1.512***
Forty and over			1.29			2.231***			0.655
Prenatal 9 visits			0.973			0.920			0.818
Prenatal 10-20 visits			ref			ref			ref
Prenatal 20+ visits			1.393*			1.003			1.015
White			ref			ref			ref
Black			0.927			0.741***			0.567***
Hispanic			1.440***			1.248*			1.412*
Other			1.269			1.020			0.930
Less than HS			0.934			1.070			0.853
High School			ref			ref			ref
Some College			1.024			1.260*			1.178
4 plus yrs College			1.194			1.650***			2.428***
Married			1.141*			1.384***			1.832***
Alcohol			1.072			0.440			0.430
Tobacco			0.759***			0.448***			0.549***
Weight Gain			0.996			0.999			1.000
N	32894	32894	32894						
LRχ2	153.84	499.91	5104.01						
df	110	150	300						
prob>χ2	0.0037	0.0000	0.0000						
Pseudo R2	0.0012	0.004	0.041						

* p<0.05, **p<0.01, ***p<0.001 (Base Outcome=External Causes)

Table 5.9 (Continued) Multinomial Logistic Regression Results (Odds Ratios) Using Gestational Age: Cause Specific Infant Mortality by Months of Third Trimester, Infant and Maternal Characteristics U.S. 2000-2004

Months of First Trimester	Respiratory, Circulatory and Nervous System Disorders		
	Model 1	Model 2	Model 3
Jan/Feb/Mar	0.945	0.950	0.928
Feb/Mar/Apr	1.187	1.193	1.197
Mar/Apr/May	1.246	1.232	1.209
Apr/May/June	1.344*	1.286*	1.310
May/June/July	1.164	1.173	1.162
June/July/Aug	1.094	1.098	1.081
July/Aug/Sept	1.064	1.064	1.052
Aug/Sept/Oct	1.133	1.155	1.135
Sept/Oct/Dec	1.227	1.209	1.234
Oct/Nov/Dec	1.019	1.014	1.018
Nov/Dec/Jan	1.405*	1.331*	1.440*
Dec/Jan/Feb	ref	ref	ref
Infant Characteristics			
Male		1.060	1.079
Gestational Age		0.936***	0.948***
Single Birth		ref	ref
Twins		0.791	0.692
Triplets Plus		3.95e+09***	4.65e+09***
Maternal Characteristics			
Teens			0.724***
Twenties			ref
Thirties			1.646***
Forty and over			1.728**
Prenatal 9 visits			0.904
Prenatal 10-20 visits			ref
Prenatal 20+ visits			1.068
White			ref
Black			0.865*
Hispanic			1.122
Other			1.095
Less than HS			0.864*
High School			ref
Some College			1.077
4 plus yrs College			1.639***
Married			1.447***
Alcohol			1.059
Tobacco			0.459***
Weight Gain			1.000
N	32894	32894	32894
LRχ2	153.84	499.91	5104.01
df	110	150	300
prob>χ2	0.0037	0.0000	0.0000
Pseudo R2	0.0012	0.004	0.041

* p<0.05, **p<0.01, ***p<0.001 (Base= External Causes)

.

The results in Table 5.9 show that the causes of infant death of SIDS, pregnancy complications, birth asphyxia and birth trauma, and neoplasms and blood conditions are not significantly associated with any of the months of the third trimester variables in any of the models estimated.

As in the model with the birthweight variable (discussed above) the cause of death of “other infections” has several significant and positive associations with the months of third trimester variables. Specifically, the months of April/May/June, August/September/October, September/October/November and November/December/January all have positive and significant associations with the COD of ”other infections” in all three models. We can say that experiencing fetal development in the third trimester in the above months (for those infants born at term) puts the infant at an increased risk of dying of other infections as compared to dying of external causes of death. Also for the “other infections” cause of death the months of June/July/August and July/August/September as the infant’s third trimester showed significant and positive associations in models one and one and two respectively, however once the maternal characteristics were included in the third model the significant association was lost.

The cause of death of “endocrine, metabolic and digestive system disorders” showed a significant association with the months of February/March/April, but this significant association was not seen in models one and three. In model one the odds ratio is associated with a p value of 0.054, and in model three the odds ratio for this cause of death was associated with a p value of 0.052.

.

In the previous model with the birthweight variable the cause of death of "perinatal infections" was positively and significantly associated with the months of third trimester of April/May/June. In this model with gestational age, "perinatal infections" shows a significant and positive association with the months of April/May/June as well as with the months of November/December/January. This significant and positive association was seen in all three models and may be interpreted as meaning that for those infants whose third trimester was in the months of April/May/June or November/December/January, their risk of dying of "perinatal infections" as compared to dying of external causes is higher, is 67.1 percent and 65.9 percent greater, respectively, than for those infants whose third trimester was in the months of December/January/February.

The cause of death "respiratory, circulatory and nervous system disorders" is shown to be significantly associated with the months of third trimester of April/May/June in all models one, two and three. The months of November/December/January in all three models were significantly associated with an increased risk of dying of respiratory causes as compared to dying of external causes.

The causes of death "prematurity and related causes" show a significant and positive association with an increased risk of dying for the months of November/December/January as compared to the months of December/January/February in all three models. The causes of death "congenital anomalies" also shows a significant association for one group of months of third trimester—April/May/June—where we see an increased risk of dying of congenital anomalies as compared to external causes of

.

death as compared to those infants who have their third trimester in the months of December/January/February.

Results of the Model Including Cases from California (with Gestational Age)

The models that included the variable gestational age yielded similar results as those discussed which included the variable birthweight. Again, the r-squared value, although very low in both models, drops even lower from 0.041 in the models that excluded California to 0.0272 in the models that included California. Also like the model using birthweight the causes of death prematurity, SIDS, pregnancy complications, and neoplasms were not significantly associated with any of the months of third trimester variables.

The cause of death congenital anomalies was significantly associated with the months of third trimester Nov/Dec/Jan, Feb/Mar/April and April/May/June. Only the months of third trimester April/May/June were also significant in the models that excluded California. The cause of death birth asphyxia and birth trauma was found to be significantly associated with the months of birth April/May/June in the model with California but not in the model without California. The cause of death of perinatal infections was found to be significantly associated with the months of third trimester Nov/Dec/Jan and April/May/June both in the models that included California and in the models that excluded California. The cause of death respiratory conditions was significantly associated with the month of third trimester April/May/June in the model that excluded California, but not in the model that included California, although these months of third trimester were significant in the models one and two that excluded

.

California. The months of third trimester Nov/Dec/Jan was significant in the models that excluded California but in the model that includes California these months lose significance. Similarly, the cause of death endocrine and digestive disorder is found to be significantly associated with the months of third trimester Feb/Mar/Apr in the model that included California but not in the model the excluded California, although in the model that includes California these months of third trimester are significantly associated with the endocrine causes of death in models one and two.

Lastly the cause of death other infections is again associated with several of the months if third trimester variables—Nov/Dec/Jan, Apr/May/June, June/July/Aug, July/Aug/Sept, Aug/Sept/Oct, Sept/Oct/Nov and Oct/Nov/Dec. Of these months of third trimester variables only June/July/Aug, July/August/Sept and Oct/Nov/Dec are not also significant in the models that excluded California. However, June/July/Aug and July/August/Sept were found to be significantly associated with the other infections cause of death in the models which excluded California for models one and two.

Predicted Probabilities

As in the models that used the birthweight variable, we see in Figure 5.4 that the "prematurity" cause of death is no longer the cause with the highest probability for all the months of the third trimester. For all causes of death there does not seem to be an overall pattern of high probability in the spring and summer months as hypothesized; instead there seems to be increases and decreases for each month of third trimester group for each of the eleven causes.

.

Figure 5.4 Predicted Probabilities: Cause of Infant Death and Months of Third Trimester (Gestational Age), U.S. 2000-2004

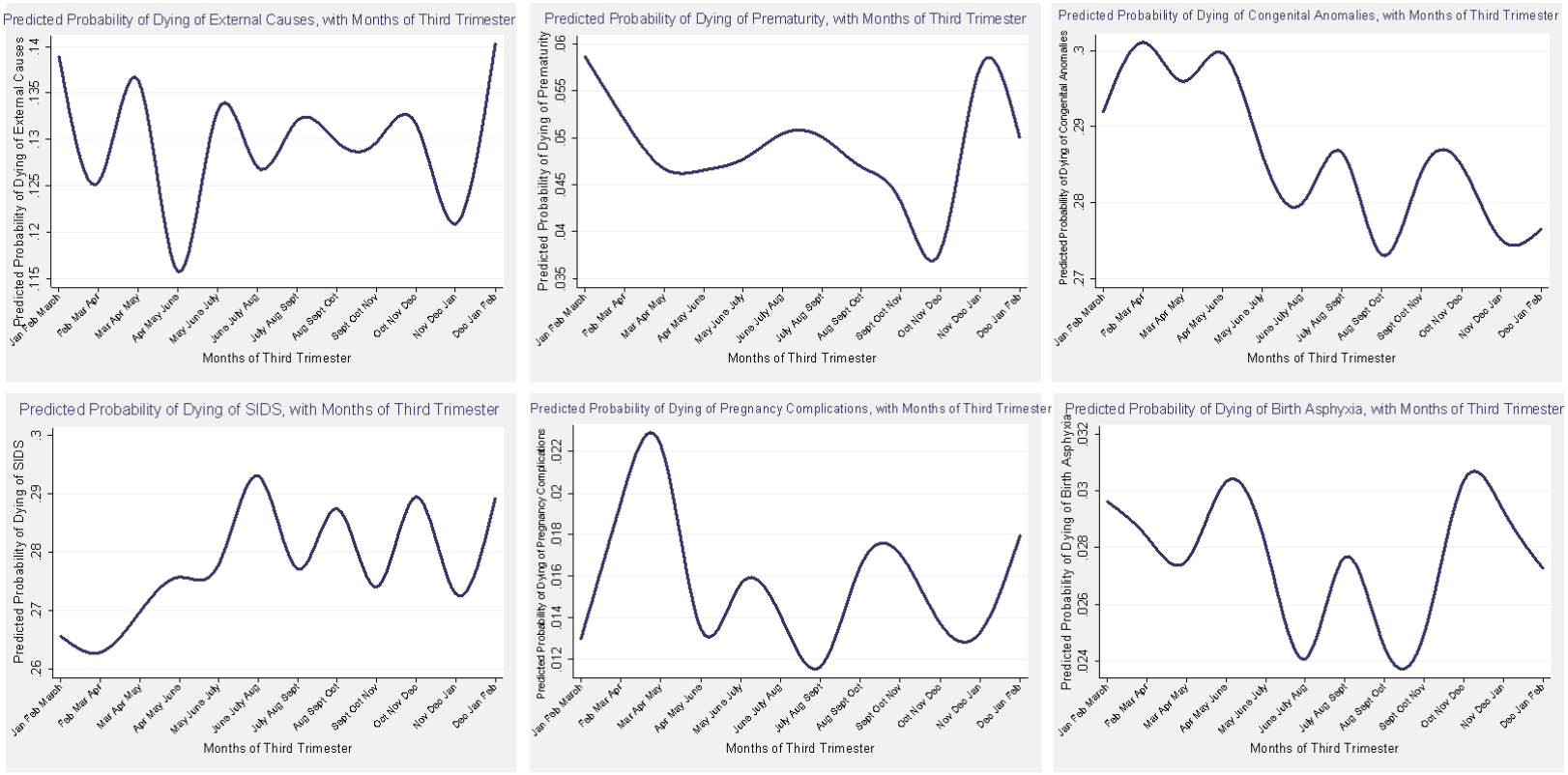
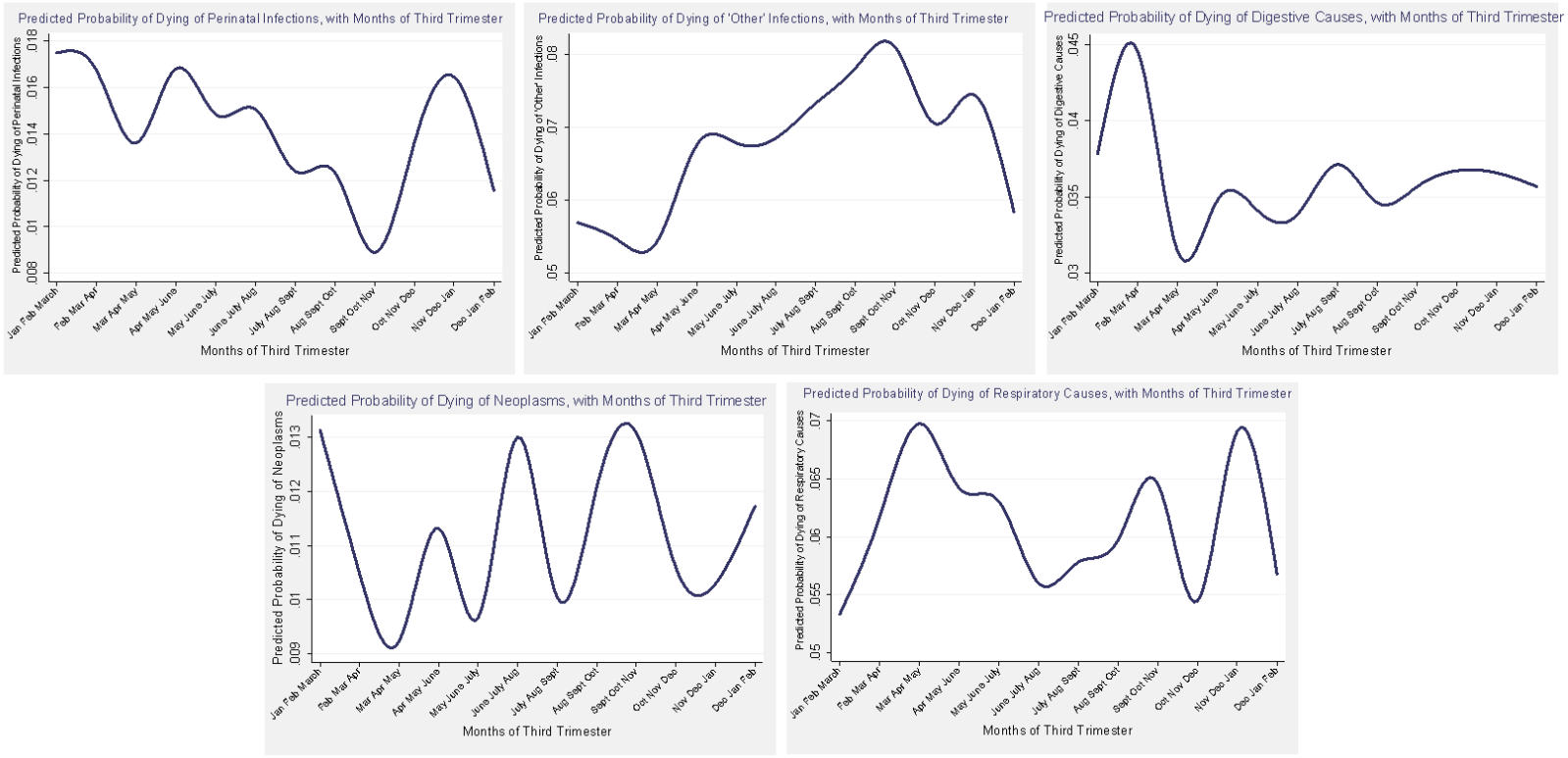


Figure 5.4 (Continued) Predicted Probabilities: Cause of Infant Death and Months of Third Trimester (Gestational Age), U.S. 2000-2004



Conclusions for 'Month of Third Trimester' Models

In the case of many of the causes of death such as SIDS, pregnancy complications, birth trauma and birth asphyxia and neoplasms and blood conditions showed few if any significant associations with the months of first trimester variables. This lack of association was also found in the months of third trimester models and in some cases also in the month of birth models. Also in several cases a month of first trimester variable would lose significance when the control variables were introduced.

When comparing the models that include California with those models that exclude California we see again that there is some change in the significance of the month of first or third trimester and the variable cause specific infant death. There seems to be more changes between the models with and without California when the month of third trimester is the independent variable of interest. The small differences between the two types of models also seem to indicate that excluding the control variables alcohol, tobacco and weight gain in the multilevel models in order to include California will not drastically reduce the predictive power of the models. Therefore regardless of which principal independent variable is used in chapter VI, I can be confident that the exclusion of these variables will not cause any major issues with the estimation of the models.

The findings of chapter IV and the models of this chapter will have to be considered as a whole to determine what the best and most appropriate measure of the seasonal or monthly variation in cause specific infant mortality. Next I will discuss the conclusion of this chapter as a whole and the decision of which variable will be used in the multilevel analyses of chapter VI.

.

Conclusions for Chapter V

Since the decision of which variable to base the multilevel results on cannot be based purely on the number of significant associations found in the respective models (due to the small difference in the number of significant associations between the models) I will have to consider all aspects of the models of chapters IV and V. My conclusion based on the finding of this chapter is to use the variable month of birth in the multilevel analyses. I believe that the reduction of the number of cases in the models using the months of third trimester variable causes many problems with the explanatory power of the models which should be avoided.

In chapter VI I will use the month of birth variables as the independent variables of interest and include these variables in the analyses which will use state level measures of climate as the second level of analyses. From the multilevel model I hope to examine the variation in cause specific infant mortality at both the state level and the individual level. Chapter VII will then use ArcGIS mapping methods to look at the geographic distribution of cause specific infant mortality in the United States.

.

CHAPTER VI

MULTILEVEL ANALYSES AND RESULTS

In this chapter I will discuss the results of the multilevel analyses that were undertaken to estimate the log odds of an infant dying of a specified cause. The advantage of multilevel analyses is that it will allow me to introduce variables into the analyses that measure contextual circumstances to which the infant has been exposed along with the infant's month of birth. In this dissertation the contextual measures are state level variables of climate. By using multilevel analyses the researcher is able to look at the influences of the infant's month of birth on cause specific infant death as well as the effects, if any, of the state level measures of climate. The direct effects of the state level variables on the cause specific infant death dependent variable may be appraised. In addition, the effects of the state level variables on the slopes of the infant variable effects on infant death, i.e., the cross-level interactions, may also be appraised. By estimating multilevel analyses I hope to be able to better understand the possible effects of climate on an infant's log odds of dying of a specified cause.

Four hierarchical generalized linear models (HGLMs) will be estimated to evaluate the log likelihood of an infant dying of a specified cause of death. The models will be estimated for infants in the United States for the years 2000 to 2004. The level-one independent variables will be the month of birth of the infant and the level-two independent variables will be three state level measures of climate. I will next discuss the HGLM models to be estimated in this chapter.

.

Hierarchical Generalized Linear Models

First I will take some time to discuss the background for the models that I will be estimating. This is important because the HGLM model is different from the regression models estimated and discussed in previous chapters of this dissertation. Also, in this chapter I will not continue to use a multinomial logistic regression model; instead I will be estimating a series of multilevel logistic regression models. I will discuss the multilevel logistic regression model and the reasons for using this type of statistical model as opposed to the multinomial logistic regression model used in previous chapters. In chapter III I discussed the HGLM model. Here I will discuss the rationale for changing from the multinomial logistic regression model to the logistic regression model, which uses a dichotomous dependent variable, and specifically why this method is more appropriate for the analyses of this chapter.

As discussed in chapter III a HGLM model is used when we are interested in looking at the impacts on a dependent variable that is nonlinear (dichotomous, count, and so forth) of contextual level variables and individual level variables. In this dissertation the dependent variable “cause of infant death” is an individual level categorical variable representing the infant’s cause of death, as listed on the death certificate. Earlier chapters of this dissertation discussed link between death in adulthood to older cohorts and their month of birth. Many of the hypotheses associated with this research center on the conditions to which the individual is exposed during gestation. More specifically, the literature concentrates on variables of exposure of the mother to harsh conditions of the winter or summer, the contraction of communicable diseases of

.

the mother, and the malnutrition of the mother during pregnancy. The common element in these variables is either the direct or indirect impacts of climate the mother experiences while she is pregnant. When the winter is harsh, food may be in short supply, colds or influenzas may be more readily contracted, and the overall exposure to adverse climate may negatively affect the mother and in turn her fetus.

Based on these hypothesized effects, I decided to incorporate state level measures of climate into my analyses. As discussed in chapter III the most appropriate way to incorporate variables measured at the state level is through a multilevel model. The goal of this chapter is to look at the interaction effects of the climate variables at the state level, with the month of birth variables at the individual level, to see if weather enhances the relationship between month of birth and the selected causes of infant death. It may well be the case that there will be no interaction between the climate variables and the slopes of month of birth on cause specific infant mortality. If this is the case I may well be able to say that it is not the variations in the climate of the state that is having the important effect on infant mortality. However, if state-level climate measures are shown to impact the slopes of month of birth on infant mortality I may say that the connection is indeed due in part to the climate a fetus experiences. However, whether or not a relationship is found, the analyses of this chapter will shed further light on the impact of month of birth on cause specific infant mortality.

In chapters IV and V, multinomial logistic regression models s were used to analyze the cause specific infant death dependent variable and the independent variable of month of birth, along with several control variables measuring maternal and infant

.

characteristics. A multinomial logit model was used because of the nature of the dependent variable cause of infant death, which is an eleven category variable with ten causes of death and a reference category of “external causes of death.” The multinomial logit results were discussed in chapters IV and V in terms of the log odds of an infant dying of the specified cause as compared to the reference category of external causes of infant death. As seen in chapter IV only four of the ten causes of infant death, namely, prematurity, respiratory causes, other infections, and endocrine disorders, were shown to be significantly associated with either an increased or decreased risk for any of the months of birth.

In this multilevel model the intention was to recreate the models of chapter IV, but to also incorporate level-two measures of climate in a multilevel multinomial logistic regression model. However, problems arose when trying to estimate the models. When completely recreating the models of chapter IV in HLM (Hierarchical Linear Models—the name of the software used to estimate the multilevel models, the term HLM will also be used to refer to the model itself as well as the software used in this dissertation) with the state level measures of climate, the large number of level-one and level-two independent variables proved to be too numerous; the models would not converge. Even after five thousand (5,000) iterations, convergence was not obtained for any of the models. Typically when HLM requires a large number of iterations, this indicates that there are one or more problems with the data and/or the model.

Raudenbush and Bryk (2002:436) tell us that there are three distinct issues to consider when running a multilevel model in HLM:

.

> the *model*, which defines the population parameters of substantive interest, the *estimation theory*, which enables us to make statistical inferences about those population parameters based on sample data, and the *computational algorithm*, which implements the estimation theory.

The decision of the model to be used varies based on the variables, data and research question. There are also a few well-known estimation theories that are used in multilevel research, which, when used with large samples, give "convergent results" (436). Once a model and an estimation theory are selected, a computational algorithm can be chosen. "In principal, any of [the] algorithms applied to the same data and the same model should produce identical results. However, the performance of an algorithm, including its rate of convergence and the reliability of that convergence across difficult applications can vary" (437). This convergence process that Raudenbush and Bryk discuss is essentially where the original multilevel model of this dissertation ran into problems. The iterative process, via a maximum likelihood estimation theory, never converged, most likely because of the size and associated difficulty of the model.

I now discuss the steps I followed to address these problems. First, I restricted the data in the hope that a simpler model would converge. It is not hard to understand why the original multilevel model had trouble converging when we consider that the original model (as presented in chapter IV) consisted of an 11 category dependent variable, a primary independent variable (month of birth) with twelve categories and numerous independent control dummy variables measuring maternal and infant characteristics. When the level-two measures of weather were then introduced it meant that for each interaction between the cause of death category and the month of birth there were three interaction effects of the three climate variables with all of the possible

.

combinations of these level-one measures. Thus my first restriction was to estimate a model in which the control variables were excluded, and only the twelve month of birth variables were included. Although this greatly reduced the number of variables in the model, the same problem—thousands of iterations without convergence—was encountered.

The second approach I followed was to only include the four causes of death that showed significant relationships with any of the month of birth variables in the full model, as estimated and discussed in chapter IV. I thus created a new dependent variable with five categories—one for each of the four causes of infant death that showed significant associations with any of the month of birth variables, and a reference category of "external causes." Also, I only included the month of birth variables found in the earlier chapter to be significantly related to any of the causes of infant death in the multilevel model along with the climate measures. Only seven of the twelve months of birth were included as separate dummy variables therefore the months May, June, July, August and September become the reference months since they are all excluded from the analyses. This relatively smaller number of variables did allow the program to fully converge, but again it only finished after thousands of iterations and with only one of the three weather measures included at level-two. Although the model did eventually converge, the large number of iterations indicated that the model was still not completely effective and that further restrictions should be made to simplify the model.

After trying to estimate the multilevel model that compliments the models from chapter IV several different ways, Dr. Poston and I decided that an alternate approach

.

would be to change the way the dependent variable—cause of infant death—was entered into the model. Instead of using a four category dependent variable with the reference category of "external causes" and respiratory causes, endocrine conditions, other infections and prematurity as the only causes of death, we decided to use these four causes of infant death as four separate dichotomous variables, and estimate four separate multilevel logistic regression models; in each model, the specific cause of death would be predicted, as compared to the reference category "external causes." Using these four causes of infant death is an appropriate strategy. I know from the analyses presented in chapter IV that even after including all the control measures, these four causes of infant death maintained significance with one or more of the month of birth variables. To further reduce the models, I only used the level-one independent variables of months of birth with which the causes of infant death categories were significantly related. This reduced the number of months of birth from 12 to 8, still using May as the month of reference, as in chapter IV.

Finally I estimated a series of four multilevel logistic regression models, instead of the single multilevel multinomial logistic regression model and the iterative process in HLM quickly converged and useable results were obtained for each of the four models, even when including all three measures of climate at the state level. In order to further determine whether using the logistic regression was appropriate and did not potentially change the outcome of the models, I estimated a level-one logistic regression model and compared the results to those found in chapter IV. When comparing the effect parameters (b's) in the logistic regression model to the results of chapter IV, there was

.

little to no difference in the strength and magnitude of the parameters. In fact, in almost every case, the exact same months of birth variables were significant as those in the multinomial model, and the value of the effect parameters only differed slightly, if at all. The similarity of the results provided further justification that switching to a logistic regression model in this chapter would not lead to different results.

Logistic regression and multinomial logistic regression models are very similar. In fact, the multinomial logistic regression model is a variation of the logistic regression model. In the logistic regression the dependent variable is dichotomous, coded as zero and one. In the multinomial logistic regression model the dependent variable is nominal, coded 0, 1, 2, 3… and so forth. And in the multinomial model the results are shown as contrasts between the various categories of the dependent variable and the category assigned as the reference, usually the zero value. So, in effect, the multinomial model is indeed a series of logistic regression models that are run simultaneously. By running these models simultaneously a single parameter of model fit is produced, rather than separate values of model fit when the dichotomous logistic regression models are estimated separately. Obtaining a single statistic of model fit is one of the reasons why it is preferable to run one multinomial logistic regression model as compared to running several logistic regression models when the dependent variable is categorical. However, due to the size of the dataset and number of variables that are used in the analyses in this dissertation, running several logistic regression models was necessary to obtain the multilevel model.

.

Although the single model fit statistic is an advantage of the multinomial logistic regression model when analyzing a multinomial dependent variable, in this dissertation obtaining a single statistic indicating model fit is not of much concern. In these analyses I am not interested in the overall fit of the model estimating the log odds of cause specific infant death. Specifically, the goal of these analyses is to assess the effect of month of birth on infant mortality and the effects of climate on that relationship, not to create an overall explanatory model of infant mortality. I would not have been able to introduce the three level-two climate variables if I had used a multinomial logistic regression model in HLM, and including all three measures of climate was important to the multilevel analyses. Based on the research questions of this dissertation and the level-one analyses that were reported in chapters IV and V, the strategy I have chosen is the most appropriate option that was available. Being able to include all three of the climate indices is enough of a benefit to justify my decision to estimate a series of multilevel logistic regression model in this chapter. Next, I turn to a discussion of the hypotheses of this chapter and the operationalization of the variables used to measure cause specific infant mortality, month of birth and climate at the state-level.

Hypotheses and Operationalization

In chapter IV, the independent variable of interest was month of birth, and the analyses used the month of May as the reference month, along with the control variables of the mother and child. From these analyses, when compared to the analyses of chapter V, which used the independent variable of interest of the infant's month of conception, I found that the month of birth is the more appropriate variable to use when assessing the

.

odds of an infant dying of a specified cause. Therefore in this chapter I will use as my main independent variable the month of birth. However, I will only include in the analyses the months of birth that showed significant associations with the categories of the cause of infant death dependent variable as seen in the results of chapter IV. Secondly, I will only include the causes of infant death that showed significant associations in the analyses reported in chapter IV. The end result will be four multilevel logistic regression models, each with one of the four cause specific infant death categories with the reference category being all "external causes" (the same reference category used in the analyses reported in chapters IV and V) as the dichotomous dependent variable.

The four causes of infant death are respiratory causes, endocrine conditions, prematurity and related causes, and other infections. The months of birth that will be included in the multilevel analyses are January, February, March, April, October, November and December. No other months or causes of infant death had any significant associations in the full models presented in the analyses of chapter IV. Including only these seven months of birth causes the reference category to change. In the analyses of chapters IV and V only the month of May was excluded, therefore making it the reference category. Similarly, the reference category of these multilevel analyses will be the five months that are excluded from the multilevel analyses, namely May, June, July, August and September.

Below in Table 6.1 are the frequencies of each of the four causes of death that will be included as separate dichotomous dependent variables in the multilevel analyses.

.

I also include the number of cases of "external causes" because this will be the reference category for each of the logistic regressions. All other cause of infant death categories as seen in chapters IV and V are excluded from these analyses.

Table 6.1 Causes of Infant Death: Multilevel Analyses	
	N
Respiratory, Circulatory and Nervous System Conditions	4,213
Other Infections	5,736
Endocrine, Metabolic and Digestive System Conditions	4,335
Prematurity and Related Conditions	35,208
External Causes	6,299

We can see that the most common cause of death is prematurity and related conditions and the least common of the four causes is respiratory, circulatory and nervous system conditions. These numbers are the same as the number of cases from chapters IV and V; the only difference is that the overall number of cases in each model will be the number of deaths from the specified cause and with the number of deaths from external causes as the reference category.

As discussed earlier in this chapter none of the control variables measuring characteristics of the mother or of the infant will be included in the multilevel analyses. By excluding these variables a simpler model is yielded, with a smaller number of coefficients and interactions than would be obtained had they been included. I will only be using those months of birth and causes of infant death that were shown to be significantly related in the full multinomial logistic regression models of chapter IV. Because I made this restriction I feel confident that I will not make any inferences from the results that are based on not controlling for characteristics of the mother or infant.

.

To incorporate the measures of climate at level-two I will be using the indices created by Poston, Zhang, Gotcher and Zu (2009) in their analysis of state-level migration. In this study the authors use eleven different climate variables for each of the fifty states. “These climate variables are based on population weighted climate data for the cities in each state that serve as the ‘major weather observation stations’ of the National Climatic Data Center” (Poston et al. 2009:745). After compiling the eleven measures of climate, the authors calculated correlations between the eleven variables. They found that of the eleven variables, three of the measures were not correlated with one another. “For instance, the variables tapping temperature are not related much with the three variables measuring humidity…[a]lso the wind measure does not have a correlation with any of the other climate measures above ±0.5. These relationships among the climate variables suggest that there may well be three underlying dimensions of climate captured by the 11 climate variables, namely temperature, humidity and wind” (Poston et al. 2009:747). The authors assess this possible connection by performing a factor analysis. In short, they find that there are ‘three statistically independent sources of climate variability characterizing the 50 states of the U.S. These represent TEMPERATURE, HUMIDITY and WIND” (Poston et al. 2009: 747). The authors then use these three measures of climate among their predictors of migration. In this dissertation I will use these three measures as my state-level variables measuring climate.

.

Descriptive Statistics of Contextual Variables

A brief discussion of the state-level variables measuring climate will now be given. Poston and colleagues remind us that climate is not the same as weather—"climate typically refers to average weather conditions, so it takes into consideration the variability in weather" (2009: 745). The climate variables are only available for the fifty states; therefore Washington, D.C. was excluded from their analysis, and, similarly, will be excluded from the multilevel logistic regression models I estimate in this chapter. Below in Table 6.2 are the descriptive statistics for the three climate factors.

Table 6.2 Descriptive Statistics of Climate Factors

	Min.	Max.	Mean	Std Dev.	N
Temperature	-2.491	2.132	-3.95E-09	1.00	50
Humidity	-2.952	1.371	2.38E-09	1.00	50
Wind	-1.836	5.065	1.21E-09	1.00	50

The minimum value for the temperature dimension is -2.491, which is for the state of Alaska, and the maximum value of 2.132 is for Hawaii. The humidity dimension shows a low of -2.952 for Arizona and a maximum value of 1.371 for New Hampshire. Lastly, the wind factor has a low of -1.836 for Oregon and a high of 5.07 for New Hampshire. Next, I will discuss the results of the multilevel analyses, starting with the one-way ANOVA results.

Multilevel Results

This next section of this chapter will discuss the findings of the multilevel models. First I will discuss the one-way ANOVA models that I performed in order to assess the appropriateness of the model. I will follow this with the findings of the four multilevel logistic regression models.

.

One-Way ANOVA Models

It is appropriate to think about a hierarchical model as a one-way ANOVA (analysis of variance) with random effects. Therefore, before estimating any multilevel model it is useful to run an ANOVA with the dependent variable in order to determine whether or not there is statistically significant variance in the dependent variable at level-two to justify the use of a multilevel model. This model is called a fully unconditional model because no independent variables are included at level-one or at level-two. The results of the ANOVA model will tell us how much variation occurs between-groups and how much occurs within groups by providing the intra-class correlation (Raudenbush and Bryk 2002: 24). However, when using a nonlinear link function, as with the logistic regression model to be used in this chapter, the intra-class correlation is less informative because "the level-1 variance is now heteroscedastic" (Raudenbush and Bryk 2002:298). Therefore it is necessary to calculate the intra-class correlation in terms of a latent variable $Z_{ij} = \eta_{ij} + r_{ij}$ (Raudenbush and Bryk 2002:334). "Under this model, the intra-class correlation can be computed as $\rho = \tau_{00} / (\tau_{00} + \pi^2 / 3)$. This conception of ρ depends on the choice of η_{ij}as the logit link and the assumption that a latent r_{ij} follows a logistic distribution" (Raudenbush and Bryk 2002:334).

Below in Tables 6.3 through 6.6 I show the results of the one-way ANOVA models for the four dependent variables to be used in the multilevel models of this chapter—respiratory causes, prematurity, other infections and endocrine conditions.

.

Table 6.3 One-Way ANOVA Results for Bernoulli Non-Linear Multilevel Model: Respiratory Causes (N=10,512)			
τ_{00}	P value	$(\tau_{00} + \pi^2/3)$	$(\tau_{00}/(\tau_{00} + \pi^2/3))*100$
0.23409	0.000	3.52409	6.65%

In Table 6.3 we see that although the majority of the variance occurs at the individual level, about 6.65 percent of the variance in respiratory related deaths to infants occurs between states.

Table 6.4 One-Way ANOVA Results for Bernoulli Non-Linear Multilevel Model: Endocrine Conditions (N=10,634)			
τ_{00}	P value	$(\tau_{00} + \pi^2/3)$	$(\tau_{00}/(\tau_{00} + \pi^2/3))*100$
0.23417	0.000	3.52417	6.64%

Similar to the results just shown in Table 6.3, the results of the ANOVA model for endocrine causes as in Table 6.4 shows that about 6.64 percent of the variance in infant mortality by endocrine causes occurs between states.

Table 6.5 One-Way ANOVA Results for Bernoulli Non-Linear Multilevel Model: Prematurity and Related Causes (N=41,507)			
τ_{00}	P value	$(\tau_{00} + \pi^2/3)$	$(\tau_{00}/(\tau_{00} + \pi^2/3))*100$
0.12583	0.000	3.41583	3.68%

The results of the ANOVA model for prematurity and related causes found in Table 6.5 show that only 3.68 percent of the variance is found between states.

Table 6.6 One-Way ANOVA Results for Bernoulli Non-Linear Multilevel Model: Other Infections (N=12,035)			
τ_{00}	P value	$(\tau_{00} + \pi^2/3)$	$(\tau_{00}/(\tau_{00} + \pi^2/3))*100$
0.2422	0.000	3.5322	6.85%

.

Lastly, we see that for deaths to infants due to other infections (results shown in Table 6.6), about 6.85 percent of the variance is found between states.

The results of these ANOVA models show that although the majority of the variance in the dependent variables is found at the individual level, statistically significant amounts of variation exist at the state level. Therefore continuing to the multilevel models in this chapter is merited. Next, I will show the results of the hierarchical generalized linear models for the four specified causes of infant death with the state-level measures of climate.

Multilevel Models

In Tables 6.7 through 6.10 I show the results of the multilevel models. My results will only address the interaction effects of the level-two variables on the slopes of the main effects. Below is Table 6.7 which contains the results of the first multilevel model with the dichotomous dependent variable “prematurity and related causes” and the months of birth January, February, March, April, October, November and December at level-one and the level-two measures of climate. As we see from the table, many of the coefficients were not significantly related to the dependent variable of “prematurity and related causes”. Also the months that were excluded from this model—May, June, July, August and September will serve together as the reference period for the month of birth. I have entered the level-one variables uncentered and the level-two variables have been centered around their grand mean. I will next interpret the significant findings of the four models.

.

Table 6.7 Effects (Gamma Coefficients) with Robust Standard Errors, Month of Birth and State-Level Climate Measures on the Likelihood of Dying of Prematurity; United States 2000-2004 (N=41,507)					
Fixed Effects	**Coefficient**	**Std. Error**	**Odds Ratio**	**t**	**p>0**
Intercept γ_{00}	1.480*	0.064	4.369	23.113	0.000
Temperature γ_{01}	0.149	0.100	1.161	1.488	0.143
Humidity γ_{02}	0.147*	0.057	1.158	2.573	0.014
Wind γ_{03}	0.046	0.082	1.047	0.563	0.576
Slope--January Births					
Intercept γ_{10}	0.109*	0.044	1.115	2.463	0.018
Temperature γ_{11}	-0.044	0.034	0.956	-1.300	0.200
Humidity γ_{12}	-0.078*	0.028	0.924	-2.763	0.009
Wind γ_{13}	0.106	0.082	1.112	1.291	0.204
Slope--February Births					
Intercept γ_{20}	0.040	0.447	1.041	0.900	0.373
Temperature γ_{21}	-0.013	0.052	0.986	-0.261	0.795
Humidity γ_{22}	-0.121*	0.039	0.885	-3.108	0.004
Wind γ_{23}	0.074	0.070	1.076	1.053	0.298
Slope--March Births					
Intercept γ_{30}	0.073	0.048	1.076	1.511	0.137
Temperature γ_{31}	0.028	0.049	1.029	0.584	0.562
Humidity γ_{32}	-0.069	0.046	0.933	-1.475	0.147
Wind γ_{33}	0.154*	0.077	1.166	1.979	0.053
Slope--April Births					
Intercept γ_{40}	0.097*	0.035	1.102	2.756	0.009
Temperature γ_{41}	-0.032	0.031	0.967	-1.054	0.298
Humidity γ_{42}	0.041	0.031	1.042	1.301	0.200
Wind γ_{43}	-0.095	0.061	0.908	-1.567	0.124
Slope--October Births					
Intercept γ_{50}	0.022	0.039	1.022	0.566	0.574
Temperature γ_{51}	-0.037	0.052	0.934	-0.714	0.479
Humidity γ_{52}	-0.543	0.045	0.947	-1.185	0.242
Wind γ_{53}	0.012	0.078	1.012	0.158	0.875
Slope--November Births					
Intercept γ_{60}	-0.029	0.029	0.970	-0.990	0.328
Temperature γ_{61}	-0.132*	0.033	0.875	-3.949	0.000
Humidity γ_{62}	-0.097*	0.041	0.906	-2.364	0.022
Wind γ_{63}	0.082	0.063	1.086	1.304	0.199
Slope--December Births					
Intercept γ_{70}	0.033	0.041	0.966	-0.814	0.420
Temperature γ_{71}	-0.128*	0.035	0.878	-3.624	0.001
Humidity γ_{72}	-0.074*	0.032	0.928	-2.269	0.028
Wind γ_{73}	-0.009	0.078	1.009	0.114	0.910

* Values significant at 0.05 or above

The first value of γ_{02} is the main effect of the humidity level-two variable on mean infant mortality from prematurity. For every unit increase on the humidity index the state's average expected odds of an infant dying of prematurity versus external causes would be multiplied by 1.158, all else equal. That is the odds would increase by about 16 percent. This value is significant with a t value of 2.573 and a p-value of 0.014. The values of $\gamma_{01,}$ the main effect of the temperature level-two variable, and $\gamma_{03,}$ the main effect of the wind level-two variable were not significant.

The effects of the level-two variables on each of the slopes of the month of birth variables on the odds of dying of prematurity, i.e., the cross-level interactions, will be interpreted next. These values are of more interest than the direct effects of the level-2 variables because they show the effect of the level-two variables on the slope of month of birth of dying of the specified cause. This is because I am mainly interested in obtaining information on the way that the level-two climate variables interact with the month of birth variables to effect cause specific infant mortality.

The value of γ_{10} is the main effect of being born in January on the probability of dying of prematurity versus dying of external causes. Those infants born in January are 1.115 times more likely to die of prematurity than those born in May through September, all else equal. This value is significant with a t value of 2.463 and p=0.018. The value of $\gamma_{11,}$ the interaction effect of the level-two temperature variable and $\gamma_{13,}$ the level-two wind variable were not significant. The value of $\gamma_{12,}$ is the interaction effect of the humidity level-two variable on the slope of being born in January on the odds of dying of prematurity versus dying of external causes. For every one unit increase on the humidity

.

index, the January-prematurity slope decreases by 0.08, that is it becomes less steep. This value is significant with an associated t value of -2.763, p=0.009.

The next interaction effects are between the level-two variables and the month of birth of February. Again, only the level-two variable of humidity is significant. The value of γ_{22} is the cross level interaction effect between being born in February and the level-two variable humidity is -0.121; this may be interpreted to mean that for every one unit increase on the humidity index, the February-prematurity slope decreases by 0.121 (t= -3.108, p=0.004).

The interaction effect of the variable wind on the slope of being born in March on dying of prematurity as compared to external causes is the only significant level-two variable. For every unit increase on the wind index, the March-prematurity slope increases by 0.15, all else equal (t= 1.979, p=0.053). The interaction effects of the level-two variables temperature and humidity were non-significant.

Next, for the effect of being born in April on the prematurity dependent variable, none of the interaction effects were significant. Only the intercept was significant—this value, $\gamma_{40,}$ may be interpreted as the main effect of being born in April versus being born in May through September, of dying of prematurity versus external causes. For those born in April the odds of dying of prematurity increase by 10.2 percent, all else equal (t=2.756, p=0.009).

None of the interaction effects between being born in October and dying of prematurity and related causes were significant. I will therefore move to the interpretation of the month of birth November. The cross level interaction of temperature

.

with the slope of being born in November on dying of prematurity, $\gamma_{61,}$ is -0.132 and may be interpreted as follows: for every unit increase on the temperature index, the November-prematurity slope is decreased by 0.13, all else equal (t=-3.949, p=0.000). The interaction effect for the humidity variable (γ_{62}) is also significant; it is the cross level interaction of humidity with the slope of born in November on dying of prematurity. For every one unit increase on the humidity index, the November-prematurity slope declines by 0.10, all else equal (t= -2.364, p=0.022). The cross level interaction for the level-two variable wind was non-significant.

Lastly, the cross level interaction between temperature with the slope of being born in December on dying of prematurity was significant. This may be interpreted as follows: for every one unit increase on the temperature index, the December-prematurity slope declines by 0.13, all else equal (t=-3.624, p=0.001). The cross level interaction for the level-two variable of humidity (γ_{72})is also significant and may be interpreted as follows: for every one unit increase on the humidity index, the December-prematurity slope drops by 0.07, all else equal (t=-2.269, p=0.028). The cross level interaction for the wind level-two variable was not significant.

Summary of Results of the Prematurity Model

Overall the results of the multilevel model with prematurity and related causes as the dependent variable yielded both significant and non-significant coefficients. The significant effects of the model generally performed as expected. For example, the value of -0.078 for humidity on the slope of January on prematurity means that as humidity increases, the slope of being born in January on prematurity decreases, or the slope

.

becomes less steep. This may be interpreted as meaning that the more humid it is, the less of an effect being born in January has on dying of prematurity and related causes. This may mean that drier months are more detrimental to an infant's survival than months that are more humid. For all the cross-level interactions the variable humidity yielded a negative value. So we may say that an increase in the humidity index tends to decrease the danger that an infant is exposed to as demonstrated by their chances of dying of prematurity.

The results of the cross level interaction effects of the level-two variable wind for the slope of March on prematurity is the only case in which the level-two climate variable is positive. This means that for every one unit increase on the wind index, the March-prematurity slope increases; that is it gets steeper. This may mean that windier months may be more detrimental to the associated risk of infant mortality from prematurity. This effect was also expected, since wind can make temperatures colder, thus providing a rationale for the hypothesized connection between mortality and month of birth.

The findings for the winter months also coincide with the hypothesized connection between month of birth and infant mortality. For the months of November and December the variables temperature and humidity are negatively and significantly associated with the cross-level interaction on the slope of those months of birth and dying of prematurity. This means that for November and December, increases in the temperature and humidity indices tend to make the slope of those months of birth on dying of prematurity less steep. This tells us that these months would be less detrimental

.

to the infant's odds of dying of prematurity if the states in which the infants were born had greater values on the temperature and/or humidity indices.

These findings seem to support the idea that one of the reasons that some months are more detrimental to an infant's survival is the effect of the weather or climate. This is promising since this is the hypothesis from the literature on adult mortality that I intended to build on for this dissertation. Next, I will show and interpret the results of the model which uses respiratory conditions as the dependent variable.

Below in Table 6.8 are the results of the multilevel model with the dichotomous dependent variable "respiratory causes". This model contains the same month of birth level-one variables and level-two climate variables as the model for the cause of death prematurity as shown in Table 6.7. As in the interpretations of Table 6.7 I will only interpret the significant gamma coefficients.

.

Table 6.8 Effects (Gamma Coefficients) with Robust Standard Errors, Month of Birth and State-Level Climate Measures on the Likelihood of Dying of Respiratory Causes; United States 2000-2004 (N=10,512)					
Fixed Effects	**Coefficient**	**Std. Error**	**Odds Ratio**	**t**	**p>0**
Intercept γ_{00}	-0.317*	0.059	0.727	-5.309	0.000
Temperature γ_{01}	0.023	0.060	1.023	0.038	0.705
Humidity γ_{02}	-0.016	0.042	0.983	-0.383	0.703
Wind γ_{03}	-0.078	0.058	1.081	1.355	0.182
Slope--January Births					
Intercept γ_{10}	0.043	0.056	1.044	0.781	0.439
Temperature γ_{11}	0.078	0.053	1.023	1.480	0.146
Humidity γ_{12}	-0.027	0.075	0.983	-0.367	0.715
Wind γ_{13}	-0.040	0.114	1.081	-0.359	0.721
Slope--February Births					
Intercept γ_{20}	-0.205*	0.085	0.814	-2.397	0.021
Temperature γ_{21}	0.126	0.076	1.134	1.651	0.105
Humidity γ_{22}	-0.001	0.088	0.998	-0.015	0.988
Wind γ_{23}	0.043	0.117	1.044	0.037	0.714
Slope--March Births					
Intercept γ_{30}	-0.247*	0.074	0.780	-3.311	0.002
Temperature γ_{31}	0.062	0.055	1.064	1.127	0.266
Humidity γ_{32}	-0.020	0.080	0.979	-0.260	0.796
Wind γ_{33}	-0.099	0.142	0.905	-0.696	0.490
Slope--April Births					
Intercept γ_{40}	-0.058	0.091	0.924	-0.647	0.521
Temperature γ_{41}	0.017	0.082	1.017	0.207	0.831
Humidity γ_{42}	0.177*	0.087	1.194	2.035	0.047
Wind γ_{43}	0.132	0.128	1.141	1.030	0.309
Slope--October Births					
Intercept γ_{50}	-0.052	0.061	0.948	-0.858	0.396
Temperature γ_{51}	-0.015	0.049	0.984	-0.308	0.759
Humidity γ_{52}	-0.008	0.039	0.991	-0.209	0.835
Wind γ_{53}	-0.008	0.099	0.991	-0.089	0.930
Slope--November Births					
Intercept γ_{60}	-0.025	0.057	0.975	-0.438	0.663
Temperature γ_{61}	-0.008	0.050	0.991	-0.165	0.870
Humidity γ_{62}	-0.031	0.064	0.969	-0.485	0.629
Wind γ_{63}	-0.088	0.130	1.092	0.679	0.500
Slope--December Births					
Intercept γ_{70}	-0.123*	0.063	0.883	-1.953	0.057
Temperature γ_{71}	-0.138*	0.060	0.870	-2.277	0.027
Humidity γ_{72}	-0.014	0.053	0.985	-0.275	0.785
Wind γ_{73}	0.010	0.121	1.010	0.087	0.931

* Values significant at 0.05 or above

The results of the model shown in Table 6.8 with the dependent variable respiratory causes have fewer significant results than the prematurity model shown in Table 6.7. The cross level interactions between the months of February and March only yielded significant associations with the intercept, not with the level-two climate variables. The cross level interactions for the months of October and November yielded no significant associations. I now discuss the significant findings of this model.

The value of γ_{20} is the main effect of being born in February, and is interpreted as follows: those born in February are 0.814 times more likely to die of respiratory causes than those not born in February , all else equal; t= -2.397, p=0.021. The value γ_{30} is the value of the main effect of being born in March on the risk of dying of respiratory causes and may be interpreted as those born in March are 0.780 times as likely to die of respiratory causes than those not born in March, all else equal; t = -3.311, p=0.002.

The cross level interaction between humidity with the slope of being born in April on dying of respiratory causes was significant. This may be interpreted as follows: for every unit increase on the humidity index, the April-prematurity slope increases by 0.17, all else equal (t=2.035, p=0.047). This means that for every one unit increase in humidity the slope of being born in April with the risk of dying of respiratory causes becomes steeper. The cross level interaction for the wind and temperature level-two variables were not significant.

Lastly, $\gamma_{71,}$ the cross level interaction between temperature with the slope of being born in December on dying of respiratory causes was significant. This may be interpreted in the following way: for every one unit increase in the temperature index,

.

the December-respiratory causes slope declines by 0.13, all else equal (t=-2.277, p=0.027). The value γ_{70} was also significant and is the value of the main effect of being born in December on the risk of dying of respiratory causes and may be interpreted as those born in December are 0.883 times as likely to die of respiratory causes than those not born in December, that is less likely, all else equal (t = -1.953, p=0.057). The interaction effects for the level-two variables of wind and humidity were not significant.

Summary of Results of the Respiratory Causes Model

Overall the results of the respiratory causes model did not yield as many significant results as did the prematurity model. This may be due to the actual differences in the ways that the level-two climate variables interact with the slopes of month of birth variables on dying of respiratory causes. Specifically this model shows two significant interaction effects—one for the month of April and the level-two variable humidity and one for the month of December and the level-two variable temperature. Both these variables show statistically significant interaction effects in the expected direction. For instance, the effect of humidity on the slope of April on respiratory causes shows that the higher the value on the humidity index, the steeper is the slope of being born in April on dying of respiratory causes. This result is opposite of the results of humidity in the model with prematurity as the dependent variable, but here the month of birth is April instead of January, February, November or December, which is where we saw negative and significant associations with the prematurity model. However, since April is a generally milder month, the fact that humidity is having the opposite effect in this model

.

is understandable. Specifically, when the humidity index is higher, the effect of being born in April lessens the risk of dying of respiratory causes.

The second significant association is found on the month of December's slope, where we see a negative association with the level-two variable temperature. Here the interaction effect is negative, meaning that for a one unit increase in the temperature index the slope of being born in December on the risk of dying of respiratory causes declines. The direction of this association is the same as the association of temperature for the months of birth November and December in the prematurity model.

Next I will discuss the results of the third multilevel model with the dichotomous dependent variable of endocrine conditions as the cause death. Below Table 6.9 shows the results of this model. Again, I will only interpret the significant associations. Similar to the results of the respiratory causes of death, there are few significant cross level interactions in this model. The months of March and October yielded no significant results.

.

Table 6.9 Effects (Gamma Coefficients) with Robust Standard Errors, Month of Birth and State-Level Climate Measures on the Likelihood of Dying of Endocrine Conditions; United States 2000-2004 (N=10,634)					
Fixed Effects	**Coefficient**	**Std. Error**	**Odds Ratio**	**t**	**p>0**
Intercept γ_{00}	-0.451*	0.065	0.063	-6.863	0.000
Temperature γ_{01}	0.109	0.065	1.115	1.672	0.101
Humidity γ_{02}	0.058	0.059	1.060	0.993	0.326
Wind γ_{03}	0.003	0.061	1.003	0.063	0.951
Slope--January Births					
Intercept γ_{10}	0.047	0.080	1.049	0.598	0.552
Temperature γ_{11}	-0.143*	0.061	0.866	-2.353	0.023
Humidity γ_{12}	0.000	0.080	1.000	0.005	0.996
Wind γ_{13}	-0.243	0.138	0.783	-1.760	0.085
Slope--February Births					
Intercept γ_{20}	-0.140*	0.065	0.868	-2.135	0.038
Temperature γ_{21}	0.058	0.074	1.060	0.779	0.440
Humidity γ_{22}	0.115	0.073	1.122	1.566	0.124
Wind γ_{23}	-0.027	0.108	0.973	-0.252	0.802
Slope--March Births					
Intercept γ_{30}	0.054	0.840	1.056	0.648	0.520
Temperature γ_{31}	-0.015	0.085	0.984	-0.181	0.857
Humidity γ_{32}	-0.048	0.078	0.952	-0.619	0.539
Wind γ_{33}	0.226	0.134	1.254	1.687	0.098
Slope--April Births					
Intercept γ_{40}	0.185*	0.654	1.204	2.842	0.007
Temperature γ_{41}	0.009	0.063	1.009	0.143	0.888
Humidity γ_{42}	0.133*	0.067	1.142	1.974	0.054
Wind γ_{43}	0.114	0.104	1.120	1.087	0.283
Slope--October Births					
Intercept γ_{50}	0.018	0.091	1.018	0.204	0.839
Temperature γ_{51}	0.015	0.074	1.015	0.204	0.839
Humidity γ_{52}	0.100	0.075	1.105	1.329	0.190
Wind γ_{53}	0.123	0.122	1.131	1.008	0.319
Slope--November Births					
Intercept γ_{60}	-0.022	0.074	0.978	-0.297	0.768
Temperature γ_{61}	-0.074	0.068	0.927	-1.092	0.281
Humidity γ_{62}	-0.185*	0.086	0.830	-2.143	0.037
Wind γ_{63}	0.018	0.124	1.018	0.147	0.884
Slope--December Births					
Intercept γ_{70}	0.114	0.094	1.121	1.208	0.233
Temperature γ_{71}	-0.120	0.081	0.886	-1.484	0.145
Humidity γ_{72}	-0.022	0.067	0.977	-0.340	0.735
Wind γ_{73}	0.290*	0.115	1.337	2.506	0.016

* Values significant at 0.05 or above

The cross level interaction effect of temperature with the slope of being born in January on dying of endocrine causes, $\gamma_{11,}$ has a value of -0.143. This means that for every unit increase in the temperature index, the January-endocrine slope declines by 0.14, all else equal (t=-2.353, p=0.023). The cross level interactions for the level-two variables of wind and humidity were not significant.

Only the intercept for the month of birth February was significant in this model. The value of γ_{20} is 0.868 and is the main effect of being born in February on the probability of dying of endocrine causes versus dying of external causes. Those infants born in February have a risk of dying of endocrine causes that is multiplied by 0.868, that is their risk decreases by 14 percent compared to those born in all other months, all else equal (t=2.463, p=0.018). The cross level interaction effects for temperature, wind and humidity were not significant.

Similar to the results for February the month of April also yielded statistically significant results only for the intercept, or the main effect of being born in April on the probability of dying of endocrine causes, i.e. $\gamma_{40.}$ This is the main effect and may be interpreted as meaning that those infants born in April are 1.204 times more likely to die of endocrine causes than those infants not born in April, all else equal (t= 2.842, p=0.007). That is their odds increase by about 20.4 percent.

The value of γ_{62} is the cross level interaction involving the humidity level-two variable on the slope of being born in November on dying of endocrine causes. For every one unit increase in the humidity index, the November-endocrine slope is decreased by 0.185, that is the slope becomes less steep, all else equal (t=-2.143,

.

p>=.037). The cross level interactions for the temperature and wind indices were not significant.

Lastly, the value of γ_{73} for the month of December is 0.290. This is the cross level interaction effect of the level-two wind variable on the slope of being born in December on dying of endocrine causes. For every one unit increase in the wind index, the December-endocrine slope increases by 0.290, that is the slope becomes steeper, all else equal (t=2.506, p=0.016). The cross level interactions for temperature and humidity were not significant.

Summary of Results of the Endocrine Causes Model

This model with the dependent variable of endocrine causes yielded only three significant interaction effects. In January, an increase in the temperature index results in a decline in the slope of dying of endocrine causes. This finding coincides with the idea that severe weather in the colder months is, in part, the cause of the increased risk of infant mortality. The interaction effect of the humidity index on the slope of being born in November on the risk of dying of endocrine causes was negative; this means that increases to the humidity index is related negatively to the slope of being born in November and dying of endocrine causes of death. This too seems to coincide with the hypothesis that more humid months in the late fall may be milder and therefore not cause as much of a threat to infant survival.

The last significant interaction effect is for the month of birth of December and the level-two variable wind. This interaction tells us that as the wind index is increased the slope of being born in December and dying of endocrine causes is also increased;

.

meaning windier months are more detrimental to the infant's chances of survival. This is an interesting finding since although these climate factors are independent of one another, we tend to think of wind as being related to cold. In this circumstance, where is it completely independent from temperature, we still see a detrimental effect on the survival of an infant. So even when removed from the lower temperatures we expect to be detrimental, we still see negative effects of wind.

Next I will discuss the last multilevel model with "other infections" as the dichotomous dependent variable. This model yielded more significant results when compared to the models for the dependent variables respiratory causes and endocrine causes. As with the other models, I will only interpret the significant results. Below in Table 6.10 are the results of this fourth model. The month of birth February yielded no significant results.

.

Table 6.10 Effects (Gamma Coefficients) with Robust Standard Errors, Month of Birth and State-Level Climate Measures on the Likelihood of Dying of Other Infections; United States 2000-2004 (N=12,035)

Fixed Effects	Coefficient	Std. Error	Odds Ratio	t	p>0
Intercept γ_{00}	-0.207*	0.061	0.812	-3.378	0.002
Temperature γ_{01}	0.186*	0.051	1.205	3.655	0.001
Humidity γ_{02}	0.018	0.048	1.018	0.375	0.709
Wind γ_{03}	-0.105	0.077	0.899	-1.357	0.181
Slope--January Births					
Intercept γ_{10}	0.207*	0.092	1.230	2.234	0.030
Temperature γ_{11}	-0.096	0.065	0.908	-1.461	0.151
Humidity γ_{12}	-0.088	0.068	0.915	-1.279	0.207
Wind γ_{13}	0.159	0.143	1.173	1.112	0.272
Slope--February Births					
Intercept γ_{20}	-0.042	0.083	0.958	-0.512	0.611
Temperature γ_{21}	0.033	0.072	1.034	0.460	0.647
Humidity γ_{22}	-0.128	0.083	0.879	-1.536	0.131
Wind γ_{23}	0.111	0.107	1.117	1.035	0.306
Slope--March Births					
Intercept γ_{30}	-0.093	0.069	0.910	-1.345	0.185
Temperature γ_{31}	0.014	0.064	1.014	-0.230	0.819
Humidity γ_{32}	-0.132*	0.060	0.875	-2.172	0.035
Wind γ_{33}	0.180	0.126	1.197	1.425	0.161
Slope--April Births					
Intercept γ_{40}	-0.144*	0.065	0.865	-2.193	0.033
Temperature γ_{41}	-0.038	0.071	0.962	-0.525	0.595
Humidity γ_{42}	0.086	0.058	1.090	1.492	0.142
Wind γ_{43}	-0.066	0.133	0.935	-0.501	0.618
Slope--October Births					
Intercept γ_{50}	0.179*	0.076	1.196	2.346	0.023
Temperature γ_{51}	0.001	0.065	1.001	0.027	0.979
Humidity γ_{52}	-0.061	0.063	0.939	-0.980	0.332
Wind γ_{53}	0.182*	0.092	1.199	1.960	0.056
Slope--November Births					
Intercept γ_{60}	0.210*	0.059	1.234	3.561	0.001
Temperature γ_{61}	-0.114	0.068	0.891	-1.679	0.099
Humidity γ_{62}	-0.127*	0.055	0.880	-2.274	0.028
Wind γ_{63}	0.277*	0.096	1.320	2.884	0.006
Slope--December Births					
Intercept γ_{70}	0.066	0.077	1.069	0.858	0.395
Temperature γ_{71}	-0.132*	0.067	0.876	-1.962	0.055
Humidity γ_{72}	-0.012*	0.041	0.882	-3.034	0.004
Wind γ_{73}	0.163	0.119	1.177	1.362	0.180

* Values significant at 0.05 or above

The month of birth January only yielded significant results in terms of the intercept or main effect $\gamma_{10.}$ This value is the main effect of being born in January on the risk of dying of other infections versus dying of external causes. Those infants born in January are 1.230 times more likely to die of other infections as compared to those not born in January, all else equal (t=2.234, p=0.030). The cross level interactions for the month of birth January were not significant.

The month of March yielded a significant association with the level-two variable humidity. Unlike the findings of the endocrine causes the association is negative—γ_{32} is -0.132. This cross level interaction of the humidity level-two variable on the slope of born in March on the odds of dying of other infections versus dying of external causes may be interpreted as follows: for every one unit increase in the humidity index, the March-other infections slope is decreased by 0.132, that is, it becomes less steep (t= -2.172, p=0.035). The cross level interactions between the temperature and wind level-two variables were not significant.

The next gamma coefficient I focus on is γ_{40}; this is the main effect of the infant being born in April on the probability of dying of other infections. Those infants born in April have odds that are lower, i.e., they are multiplied by 0.865, of dying of other infections as compared to those infants born in May-Sept, all else equal (t = -2.193, p=0.033). All three cross level interactions were non-significant.

The results of the month of birth October yielded a significant intercept and cross level interaction with the level-two variable wind. This gamma coefficient, γ_{50}, may be interpreted as the main effect of being born in October on the probability of dying of

.

other infections versus external causes. Those infants born in October are 1.196 times more likely to die of other infections than those not born in October, all else equal (t=2.346, p=0.023). The cross level interaction with the level-two variable wind is also significant for the month of birth October. The value for γ_{53} of 0.182 may be interpreted as follows: for every one unit increase on the wind index the October-other infections slope increases by 0.182, that is it becomes steeper, all else equal (t=1.196, p=0.056). The cross level interactions for humidity and temperature were not significant.

The month of birth November yielded several significant coefficients. First, the main effect of being born in November, $\gamma_{60,}$ is 1.234; this means that for those infants born in November as compared to those not born in November, the odds of dying of other infections increase by 23.4 percent, all else equal (t=3.561, p=0.001). The cross level interactions with the level-two variables of both humidity and wind were also significant. The cross level interaction with the level-two variable humidity, $\gamma_{62,}$ is -0.127, and means that for every one unit increase in the humidity index, the November-other infections slope is decreased by 0.127, that is it becomes less steep, all else equal (t=-2.274, p=0.028). The cross level interaction with the level-two variable wind, $\gamma_{63,}$ is 0.277, indicating that with a one unit increase in the wind index, the November-other infections slope increases by 0.277, i.e., the slope becomes more steep, all else equal (t=2.884, p=0.006). The cross level interaction effect for the level-two variable temperature was not significant.

Lastly, the cross level interaction of the level-two variable temperature on the slope of being born in December on dying of other infections is significant. This

.

indicates that for every one unit increase in the temperature index, the December-other infections slope is decreased by 0.132, that is the slope becomes less steep, all else equal (t=-1.962, p=0.050). The cross level interactions for the variables wind and humidity were not significant.

Summary of Results of the Other Infections Causes Model

The results of the model with the dependent variable other infections performed similarly to the other models in this chapter. In March the humidity variable was associated with a decrease in the slope of being born in March on the other infections dependent variable. This may well indicate that in the milder spring months humidity can decrease the risk of death to the infant. In November the humidity variable was also negative, meaning that also in the fall months increases to humidity will decrease the slope of the effect of being born in November on dying of other infections. Conversely, for the month of November the level-two variable of wind is positive; this means that increases in the wind index results in increases in the slope of the effect of being born in November on dying of other infections.

Surprisingly both the level-two variables of temperature and humidity are negatively associated with the month of December. This means that as the temperature and humidity indices increase, the slope of the effect of being born in December on dying of other infections is decreased, that is the slope becomes less steep. If climate were the causal factor in this case we would expect increases in the temperature index to increase the slope, but for those born in December the slope is decreased.

.

Conclusions for the Analyses with Multilevel Models

In chapter VI of this dissertation I used Hierarchical Generalized Linear Modeling (HGLM) to perform multilevel analyses of the likelihood of an infant dying of a specified cause. My purpose was to include both the infant's month of birth as well as state-level measures of climate to assess the likelihood of dying for infants in the U.S. for the years 2000 to 2004. I was also interested in the effect that the state-level climate measures have on the relationship between month of birth and cause specific infant mortality—the interaction effects.

For the cause of death of prematurity, the humidity index was negatively associated with the months of January, February, and December but positively associated with November. For the cause of death of other infections, the humidity index was negatively associated with the months of March, November and December. For the cause of death of endocrine causes humidity was negatively associated in November, and April was positively associated. Lastly, for the cause of death of respiratory causes, humidity was positively associated with the month of April. Overall these findings do not indicate one concrete interaction effect of humidity on the four causes of infant death used in the multilevel analyses. The negative associations in the cross-level interactions for the majority of the months may indicate that when humidity is high, the chance of an infant dying of the specified cause decrease. This, however, does not tell us why humidity is positively associated with the month of April for the cause of death of respiratory causes, and with the month of November for the cause of death of

.

prematurity, since these months show negative associations for the other two causes of death.

The temperature index performed more closely to my expectations than did the humidity index. In all cases where the temperature index was significantly associated with any of the months of birth in the four multilevel models, the association was negative. This means that for all four causes of death, as the temperature index increases the chances of an infant dying of the specified cause decrease; that is the slope of that association becomes less steep. The months where this association is obtained are January, December and November, all months of the late fall or winter. This negative association is expected since it means that when these months are warmer—indicated by an increase in the temperature index—the slope of the association between infant mortality from a specified cause and that particular month of birth becomes weaker. This finding coincides with the idea that an infant's chance of survival is influenced by the climate that they are exposed to as a fetus or in very early life.

Similar to the humidity index, the wind index did not perform as expected in the four multilevel models. First, the wind index was only significant in three of the models, namely those where the dependent variable was endocrine causes, other infections, and prematurity and related causes. Second, the index was negatively associated with the month of December for endocrine causes, positively associated with November in the other infections model, and negatively associated with March for the prematurity model. These results do not indicate a consistent association between the wind index and the slope of cause of infant death and month of birth.

.

Overall I believe the findings of the impact of the temperature index on the slope of month of birth and cause specific infant mortality give general support to the idea that early life conditions impact an infant's chance of survival. The negative associations show that when temperature is higher, especially in months that are typically colder, the slope of month of birth on infant mortality is decreased. This means that the association is weakened as a result of higher temperatures. This finding may give support to the idea that if the climate that the mother is exposed to during pregnancy and the infant is exposed to in early life is less severe, the chances of survival of that infant are better than if the climate is more severe.

To further explore the relationship of infant mortality and month of birth, chapter VII will present maps of the United States portraying the infant mortality rate of that state as well as the value for the climate index for each state. These maps will provide a visual representation of the associations found in this chapter between climate and the effect of month of birth on cause specific infant mortality.

.

CHAPTER VII

SPATIAL ANALYSES OF INFANT MORTALITY IN THE STATES OF THE UNITED STATES

The previous three chapters of this dissertation used multivariate statistical models to examine the effects of month of birth, and of month of gestation on an infant's chance of dying of a specified cause, as well as the extent to which the physical climate of the state in which the infant was born affects the relationship in that state between infant mortality and month of birth or gestation. In this final analysis chapter of my dissertation, I will examine in another way the relationship between physical climate and infant mortality in a series of maps displaying the association. The purpose of this chapter is to visually display some of the relationships found in the previous analyses and to further clarify and test the hypothesis that an infant's cause of death is in some part determined by the conditions to which it is exposed during its fetal development.

I will be showing seven separate maps. The first three maps will show the three climate indices, state by state, that were discussed in chapter VI. The fourth map will show the infant mortality rate for the years 2000 through 2004 by state. And the next three maps will display two variables—each climate index separately by state along with graduated symbols by state that represent the infant mortality rate of that state. Showing more than one variable on one map can be difficult because the more information one tries to show on one map, the less clear the map becomes. Multivariate maps are possible, but their effectiveness is debatable. Keeping in mind that the purpose of a map is to show information visually, the more information that is included in one map, the

.

more difficult it becomes to read. In this chapter I choose to show the two state-level variables of interest—the infant mortality rate and the three climate indices —first separately on four univariate maps, and secondly together on three bivariate maps. By creating and describing the results in these two types of maps, I hope that the relationship between the infant mortality rate and state-level climate will be elucidated and, hopefully, not obfuscated with the inclusion of two variables on the same map.

In this chapter I first briefly discuss mapping methods used to develop the best possible representation of my data. Next, I will discuss the data used in the maps as well as the descriptive statistics of these data. Then I will turn to a discussion of the hypotheses to be tested with the spatial maps. The fourth section of this chapter will present the seven maps, as discussed above. Lastly I will conclude chapter VII with a discussion of the findings of this chapter, i.e., the specific data displayed in the seven maps I created, and the major findings produced by the maps with respect to the relationship between climate and infant mortality.

Mapping Infant Mortality and Climate

As discussed in chapter III, I use the mapping software ArcGIS to create the maps for this chapter. The ability to incorporate numeric data into my maps allowed me to incorporate the climate indices that I used as the state-level variables in the multilevel statistical analyses presented in chapter VI. I use the same three indices—temperature, wind and humidity—at the state level in the maps to be shown here. Next, I decided to use the infant mortality rate for each of the fifty states in place of the cause specific infant mortality variables that I used throughout this dissertation. The reason for using

.

the infant mortality rate, instead of using any of the cause specific infant mortality rates, owed to the small number of cases when the causes of death were separated by state. Also, using the infant mortality rate allowed me to compare results between states since the measure takes the population of the state into consideration. Although I believe that using a different measure of infant mortality in this chapter is justifiable, caution should be taken when interpreting the resulting maps. Comparing the results of the maps with the results of the analyses of chapters III, IV and V is not entirely appropriate. The previous analyses use cause specific infant mortality instead of infant mortality rates. Also, the analyses of previous chapters used control variables to assure that any relationships found were not instead due to characteristics of the mother or child. The maps of this chapter do not incorporate any control measures, and therefore we cannot be certain that relationships observed are not instead due to unmeasured variables. The maps of this chapter should be used to supplement the findings of the previous chapters and considered apart from the previous findings.

Below in Table 7.1 are the values of the infant mortality rate by state that will be used in the maps to be presented here. These data were produced by the Centers for Disease Control's National Center for Health Statistics and the National Vital Statistics System, which is the same federal unit that compiled and provided the Linked Birth/Infant Death dataset used earlier in this dissertation. I took the infant mortality rate for years 2000 through 2004 and calculated the average rate for all four years; the average annual rate thus coincides with the years of annual data used in the previous chapters.

.

Table 7.1 Infant Mortality Rate by State, 2000-2004

State	Infant Mortality Rate	State	Infant Mortality Rate
Alabama	8.82	Montana	6.42
Alaska	6.36	Nebraska	6.34
Arizona	6.55	Nevada	6.00
Arkansas	8.47	New Hampshire	4.93
California	5.25	New Jersey	5.62
Colorado	6.11	New Mexico	6.11
Connecticut	5.75	New York	6.08
Delaware	8.88	North Carolina	8.35
Florida	7.33	North Dakota	6.48
Georgia	8.65	Ohio	7.74
Hawaii	6.95	Oklahoma	7.95
Idaho	6.14	Oregon	5.59
Illinois	7.53	Pennsylvania	7.40
Indiana	7.78	Rhode Island	6.40
Iowa	5.36	South Carolina	8.98
Kansas	7.04	South Dakota	7.11
Kentucky	6.94	Tennessee	9.05
Louisiana	9.95	Texas	6.37
Maine	5.01	Utah	5.26
Maryland	8.09	Vermont	4.68
Massachusetts	4.80	Virginia	7.48
Michigan	8.09	Washington	5.62
Minnesota	4.85	West Virginia	7.98
Mississippi	10.32	Wisconsin	6.43
Missouri	7.95	Wyoming	6.99

For the years 2000 to 2004 the average annual infant mortality rate ranges from a low of 4.68 deaths to infants of less than 1 year of age per 1,000 births in Vermont, to a low of 10.32 per 1,000 in Mississippi. By looking at these values alone we see that it does not seem that higher infant mortality rates are concentrated in the colder states and lower infant mortality in the warmer states; in fact the opposite is observed in Table 7.1. However, from the previous analyses undertaken in this dissertation we know that there

.

are statistically significant relationships between the month an infant is born and the cause of infant death dependent variable. So, although it would appear that a relationship is unlikely to be observed, I will create maps with the data as described above in hopes that it will elaborate in part some of the relationships observed in previous chapters.

Because the state-level infant mortality rates shown in Table 7.1 were not used in previous chapters, I first estimated three zero-ordered correlations between the infant mortality rate and each of the three climate indices. The purpose of this exercise was to determine whether there is a relationship between the infant mortality rate and the temperature indexes, thus to provide a justification for examining the relationships spatially via maps. Below I show the zero-order correlation coefficients.

Table 7.2. Zero-Order Correlation Coefficients between the Infant Mortality Rate and the Climate Indexes: United States 2000 to 2004

Climate Index	Correlation Coefficient
Temperature	0.468
Wind	-0.234
Humidity	0.344

In Table 7.2 we see that the humidity and temperature indexes are positively correlated with the infant mortality rate and the wind index is negatively associated. The highest correlation is between the temperature index and the infant mortality rate with a correlation coefficient of 0.468. The humidity index shows a correlation coefficient of 0.344 and the wind index has the lowest correlation coefficient with a value of -0.234. These correlations are moderate, but I will still continue on to the mapping of infant mortality with each of the three temperature indexes so to visually display the

.

relationships. Next, I turn to a discussion of the hypotheses and the operationalization of the data used in this chapter.

Hypotheses and Operationalization

The state-level climate variables that will be shown on the maps of this chapter are the same data that were used in chapter VI. They include a value on each of the three climate indices for each of the fifty states (excluding Washington, D.C.). Each state will therefore have a value for each of the indices that will be shown on the maps by a color indicating whether this is a high or low value on the index. The index was categorized into six groups. This makes the map easier to read. Having a different color for each of the values of the index is not possible, since this would mean fifty different colors. Having the colors categorized into six groups makes the colors and values easier to read and allows the data to be shown more effectively. The maps will also contain the infant mortality rate by state. These data will also be broken into categories that will be shown on the maps. The infant mortality rate will be shown in five categories on the univariate map and on the bivariate maps.

The hypothesized connection between exposure to adverse conditions as a fetus and adult mortality is well established, as discussed in earlier chapters of this dissertation. However, the connection between exposure to adverse conditions and infancy as well as between adverse conditions and *in utero* existence are less well accepted and understood, especially with respect to the experiences of contemporary cohorts. I believe that the results of the level one and multilevel analyses reported in earlier chapters of this dissertation provide evidence to support the idea that there is

.

indeed some connection between month of birth and cause specific infant death. If we accept this connection, we may next be interested in understanding better why exactly it exists. The aim of chapter VI was to see how the state level measures of climate effect the relationship between infant mortality and month of birth. In some cases the climate variable made the effect stronger, and in some cases the climate variable made the connection weaker. In this chapter the use of the same state level climate measures will be used to further examine the findings of the multilevel analyses.

Take for instance the finding that in states that are warmer in the typically colder months the effect of month of birth on cause specific infant mortality is weakened. The implication here is that the warmer temperatures (those states with higher scores on the temperature index) are less detrimental to an infant's chances of dying of a specified cause. Extending this finding further, we may expect that those states that are warmer, showing higher values on the temperature index, will have lower levels of infant mortality. This connection between climate and the infant mortality rate was not directly studied in the level one and multilevel analyses I conducted earlier in this dissertation; instead it is an extension of the findings of previous chapters. By mapping these two variables in this chapter, I aim to further clarify the connection between climate and infant mortality.

Specifically, I expect to find that those states with *higher* values on the temperature index to have *lower* levels of infant mortality, and those states with *lower* values on the temperature index to have higher levels of infant mortality. Second, I expect to find that those states with *higher* values on the humidity index to have *lower*

.

levels of infant mortality and those states with *lower* values on the humidity index to have *higher* values of infant mortality. Lastly, I expect to find that those states with *higher* values on the wind index to have *higher* levels of infant mortality and those states that have *lower* values on the wind index to have *lower* values of infant mortality. It is important to keep in mind that the relationships that will be shown on the maps of this chapter will only include those two variables discussed here—the climate indices and the infant mortality rate by state. This means that unlike the analyses of chapter IV, all the relationships observed on the maps do not control for any characteristics of the mother or infant. Nor do they control for characteristics of the state. The inability to include controls is a notable shortcoming of these maps. But, the intent of the maps is to compliment, not replicate exactly, the findings of the previous conducted multivariate statistical analyses where I was able to control for variables that are known to impact infant mortality. Next I will discuss some of the data issues that should be considered when looking at the maps in this chapter. I will then show the seven maps that I created using the above-mentioned data. I will conclude this chapter with a summary of findings and an overall conclusion.

Data Considerations

As stated above the infant mortality data that I use in this chapter differ from the data used in the previous analyses chapters. This means that the analyses of the previous chapters should be thought of as being separate from the results to be reported in the maps. There are at least three ways to overcome this shortcoming. First, I could rerun the analyses of chapters IV through VI with the infant mortality rate in place of the cause

.

specific infant mortality variable. A second option would be to produce the maps of this chapter with the cause specific infant death variables in place of the infant mortality variable. I discuss the third option in the next paragraph. Following one or more of these options would make the analyses of chapters IV through VI more consistent with the maps. Due to time constraints, however, I have not undertaken any of these options, and therefore the maps of this chapter should be considered as separate from the analyses of the previous chapters. If more time were available, it would have been beneficial to this dissertation to include the models presented in the previous chapters to be better able to compare the results. This would give the maps of this chapter more weight in terms of their ability to add to the analyses of the previous chapters.

It is also important to note that the maps of this chapter do not take into account any of the control variables employed in the previous chapters, specifically those pertaining to infant and maternal characteristics. This is the third option; it pertains to a shortcoming of this chapter in terms of the ability of the maps to show the relationship between the infant mortality rate and the climate indices. This shortcoming, however, could be addressed by controlling for these variables before the construction of the maps. A way to overcome this would be to undertake a statistical analysis using the infant mortality rate and the control variables along with the state level measures of climate, and to use these multivariate regression results to calculate a predicted or so-called “hat” measure of the “infant mortality rates by state”; this would then be mapped in place of the infant mortality rate and state-level climate variable. The resulting map would be able to account for the variation in the infant and maternal characteristics as

.

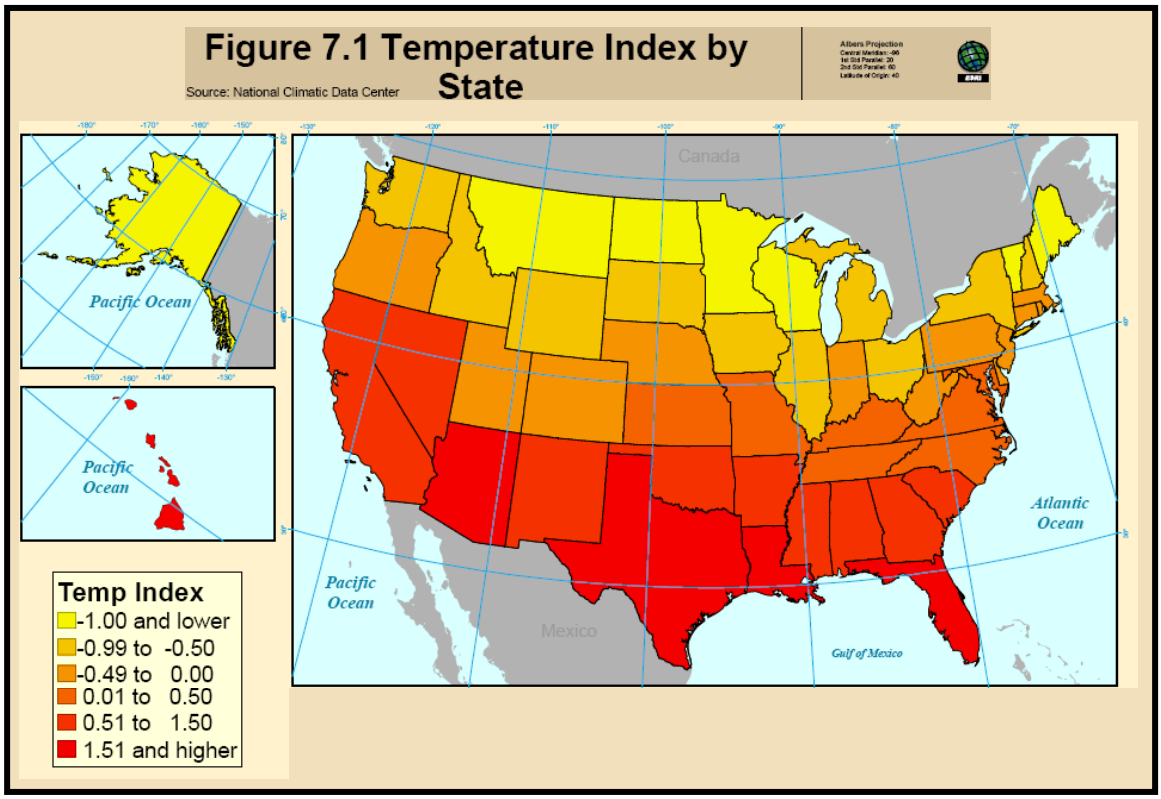
well as the climate measures, since these variables would have been used to calculate the predicted infant mortality rates. A map of this type would add to the analyses of the previous chapters because it would control for the same variables used in chapters IV and V. The map would be a univariate map that takes into account multiple variables. Typically the fewer variables displayed on a map, the simpler it is to read, therefore this potential map would also be improve upon the maps that are shown in this chapter because they would be easier to read. Again, due to time constraints I will be unable to create a map with the predicted probabilities as discussed above. However, this type of map would greatly add to these analyses and will be an important direction in my future research. Next I will show the maps that I did create, using the infant mortality rate and the state level measures of climate.

Maps

The first series of maps I show are the four univariate maps, one for each of the three climate indices for the states of the United States, and a fourth map of the 2000-2004 average annual infant mortality rate for the states. As discussed above, these maps will only show one variable at a time to ensure that the maps are as clear as possible. A later section will then take the first three of the four univariate maps and add to them the infant mortality rate. The first map shown below in Figure 7.1 is the temperature index for the United States.

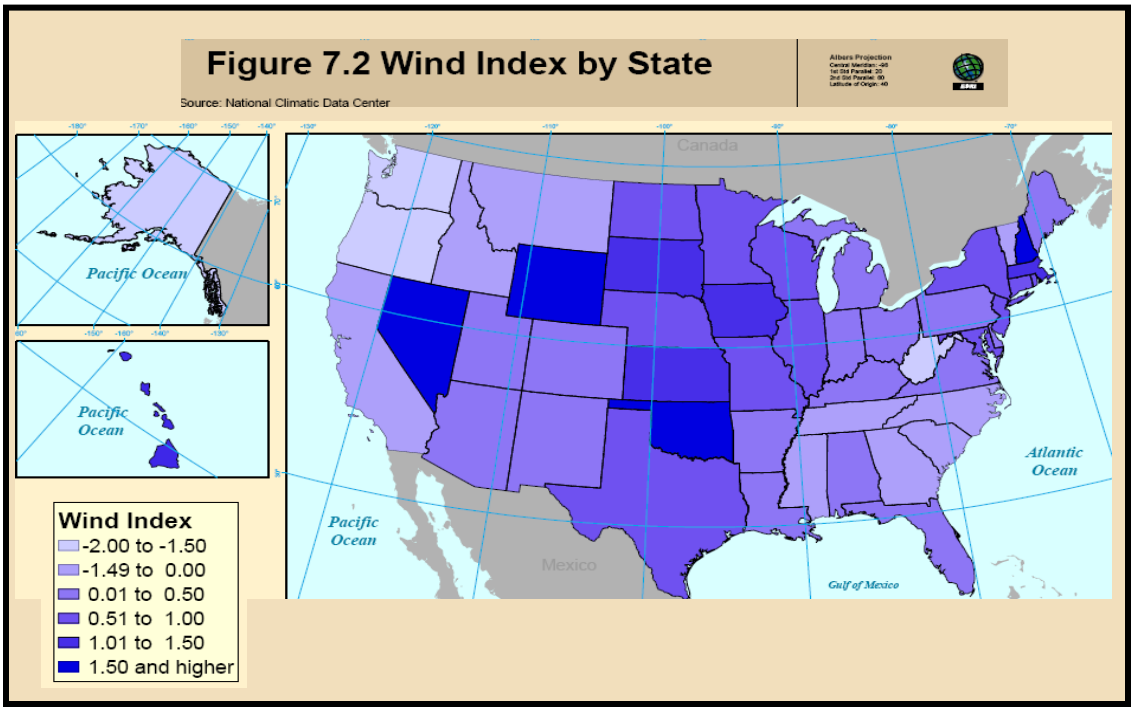
Figure 7.1 shows the temperature index for each of the fifty states of the United States. As would be expected, those states with higher values on the temperature index are concentrated in the southern part of the United States; the state with the lowest value

.



on the temperature index is Alaska, shown in Figure 7.1 as the lightest yellow color.

Next, in Figure 7.2 I show the wind index by state.



.

As we can see in Figure 7.2, unlike the temperature index, there is not as predictable a North-South pattern among the fifty states for the wind index. The highest values are spread throughout the map with no clear concentration. However, it does seem that overall the higher values of the wind index are seen in the states in the middle of the country, with lower values on either coast. Specifically there seem to be lower values in the Pacific Northwest and the coastal Southeast. The last climate index that I show as a univariate map is the humidity index. Figure 7.3 shows the values of the humidity index for the fifty states.

As we can see from Figure 7.3, the higher humidity states seem to be concentrated on the coasts of the United States, with the lower humidity states in the southwestern region of the country. High values on the humidity index are also seen in the Pacific Northwest, Hawaii and Alaska.

The fourth and final univariate map contains the 2000-2004 average annual infant mortality rate by state. The infant mortality rate data were taken from the National Center for Health Statistics (NCHS). The purpose of this fourth map is to show the data on infant mortality without the climate data. Since the infant mortality rate is hypothesized to be related to the climate variables shown in Figures 7.1 to 7.3, it is useful to also show these data alone in Figure 7.4 below. By showing the IMR data separately, we get an idea of the distribution of the rates throughout the United States. This will be especially useful when I next include the climate variables with the infant mortality rate in Figures 7.5 through 7.7. In Figure 7.4 the infant mortality rate is

.

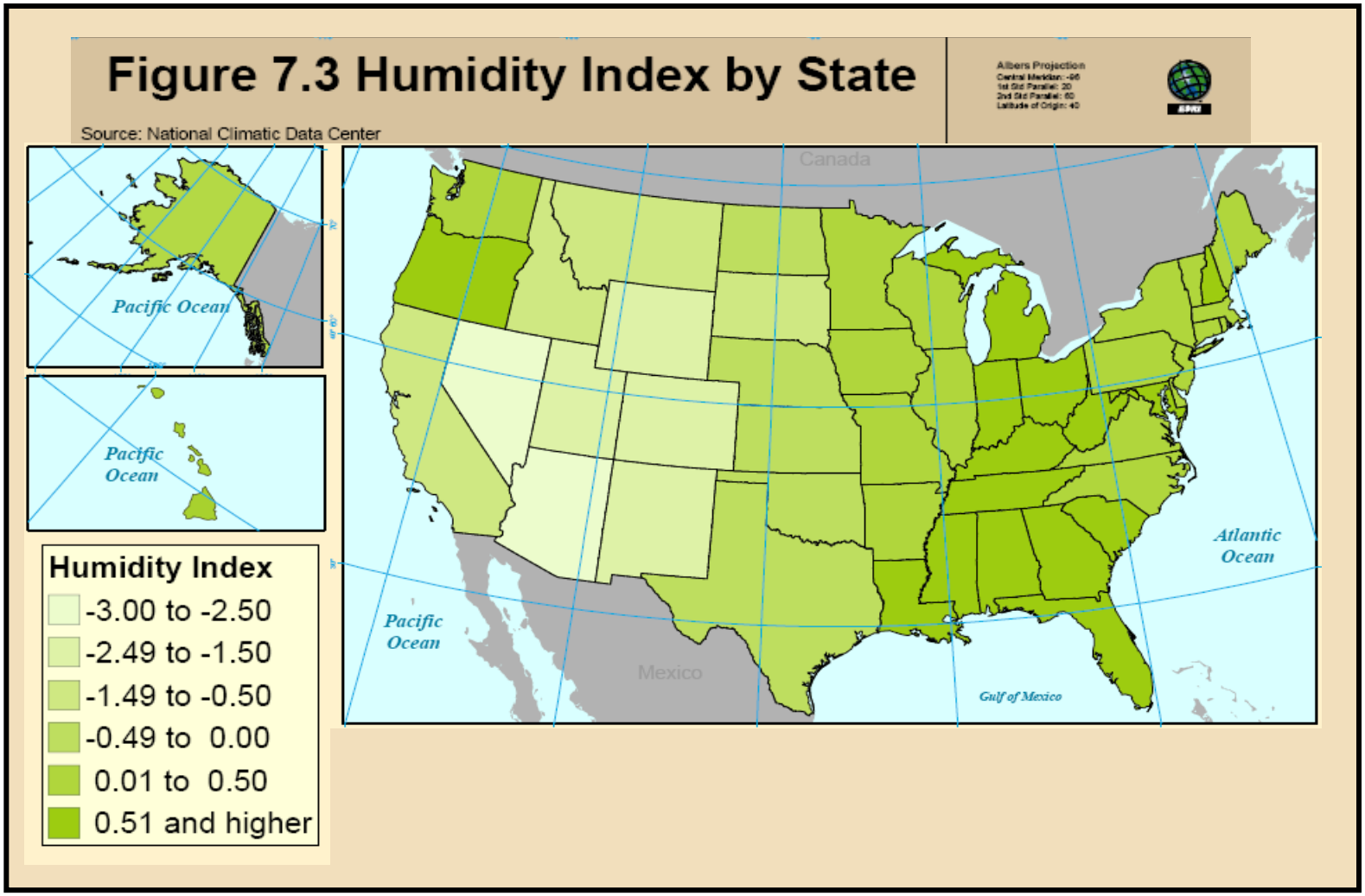


Figure 7.3 Humidity Index by State
Source: National Climatic Data Center
Albers Projection
Central Meridian: -96
1st Std Parallel: 20
2nd Std Parallel: 60
Latitude of Origin: 40
ESRI
Canada
Pacific Ocean
Pacific Ocean
Pacific Ocean
Atlantic Ocean
Mexico
Gulf of Mexico
Humidity Index
-3.00 to -2.50
-2.49 to -1.50
-1.49 to -0.50
-0.49 to 0.00
0.01 to 0.50
0.51 and higher

symbolized by graduated triangles, where each increase in the category of the infant mortality rate is shown with a larger triangle.

The average annual infant mortality rate for the years 2000 to 2004 shown in Figure 7.4 seems to be highest in the southern states, as seen by the larger triangles, especially concentrated in the states of Mississippi, Louisiana and Tennessee. We see smaller triangles concentrated in the western portion of the map, indicating lower levels of infant mortality in these areas.

In the next three figures we will view the exact same data shown in Figures 7.1 to 7.4. The difference will be that each of the climate variables will be shown in a map along with the infant mortality rate. The result will be a demonstration of the spatial relationship between the infant mortality rate and each of the three climate indices used in this dissertation. First, I will take the temperature index and infant mortality and show the resulting map in Figure 7.5. The hypothesized relationship between temperature and infant mortality, as stated in an earlier section of this dissertation, is that those states with *higher* values on the temperature index will have *lower* levels of infant mortality, and those states with *lower* values on the temperature index will have *higher* values of infant mortality. Figure 7.5 below tests this hypothesized spatial relationship.

The results seen in Figure 7.5 show the higher temperature index values are concentrated in the Southern United States. We also see a high concentration of higher infant mortality rates in the Southeastern United States. This relationship is the opposite of the relationship found in the multilevel models of chapter VI, where the higher

.

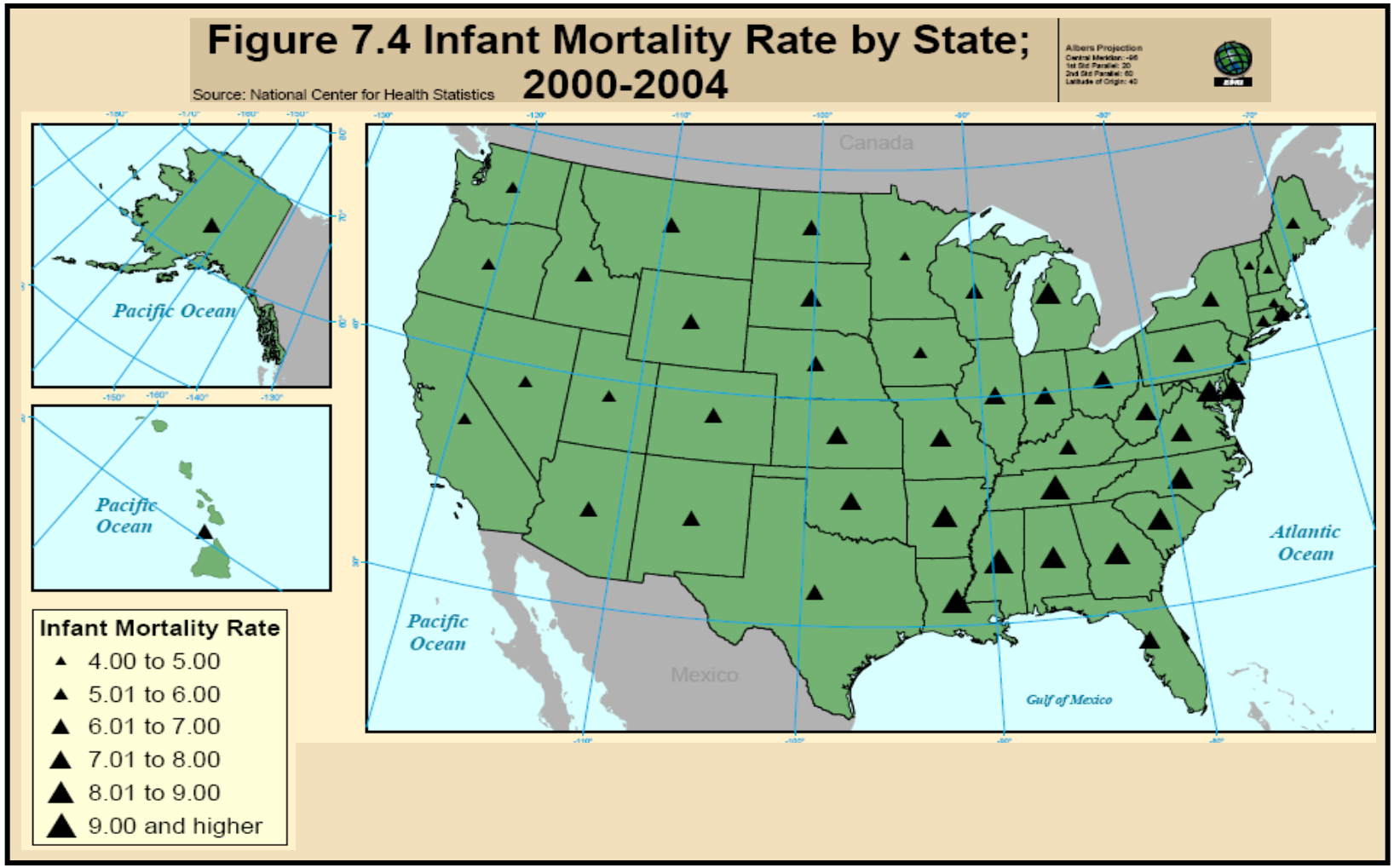


Figure 7.4 Infant Mortality Rate by State; 2000-2004
Source: National Center for Health Statistics
Albers Projection
Central Meridian: -96
1st Std Parallel: 20
2nd Std Parallel: 60
Latitude of Origin: 40
Canada
Pacific Ocean
Pacific Ocean
Pacific Ocean
Atlantic Ocean
Mexico
Gulf of Mexico
Infant Mortality Rate
4.00 to 5.00
5.01 to 6.00
6.01 to 7.00
7.01 to 8.00
8.01 to 9.00
9.00 and higher

temperatures reduced the relationship between cause specific infant mortality and month of birth. This is likely because I was not able to control for the various socioeconomic and sociodemographic variables. Factors other than climate are likely at play but their effect on the infant mortality rate cannot be observed in this map. Also, smaller values for the infant mortality rate in the Southwest coincide with higher values on the temperature index for these states. This pattern is more consistent with the hypothesized connection between the infant mortality rate and temperature that was taken from the previous chapters.

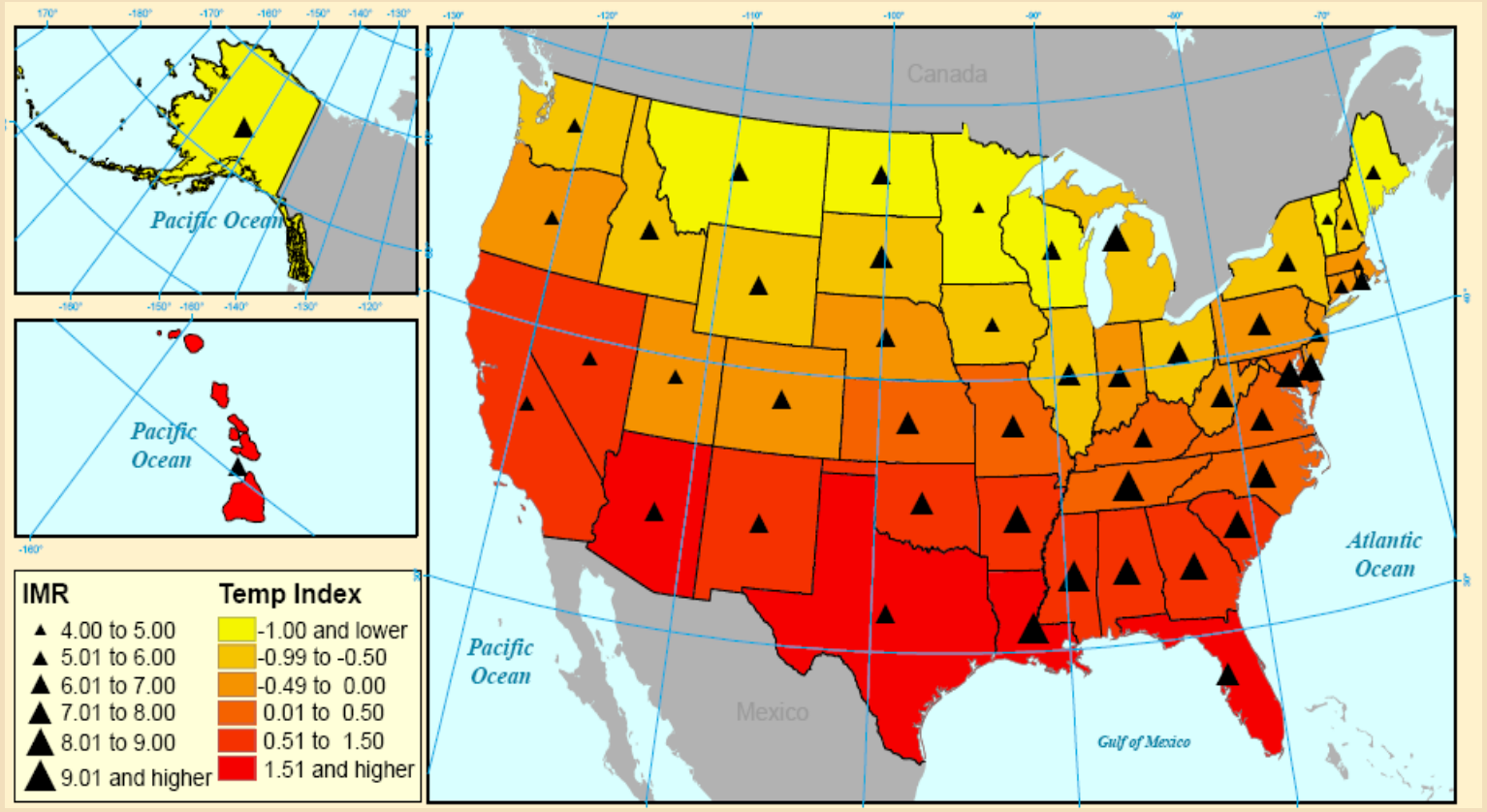
Figure 7.6 shows the wind index and the infant mortality rate by state. Again, we see the higher infant mortality rates concentrated in the Southeastern United States. However, this does not seem to coincide with higher values on the wind index as was hypothesized. The hypothesized connection between higher values on the wind index and higher values of the infant mortality rate seems not to be supported by the distributions of the map in Figure 7.6.

Figure 7.7 below shows the third and final bivariate map of the humidity index and infant mortality rate by state. Higher values on the humidity index seem to be concentrated in the Southeast to Northeast regions of the United States. The southern states are also where we observe the higher values of infant mortality. The hypothesized connection between infant mortality and the humidity index is that higher values of humidity will lead to lower levels of the infant mortality rate. From Figure 7.7 this does not seem to be the case.

.

Figure 7.5 Temperature Index and Infant Mortality by State

Source: National Climatic Data Center
National Center for Health Statistics



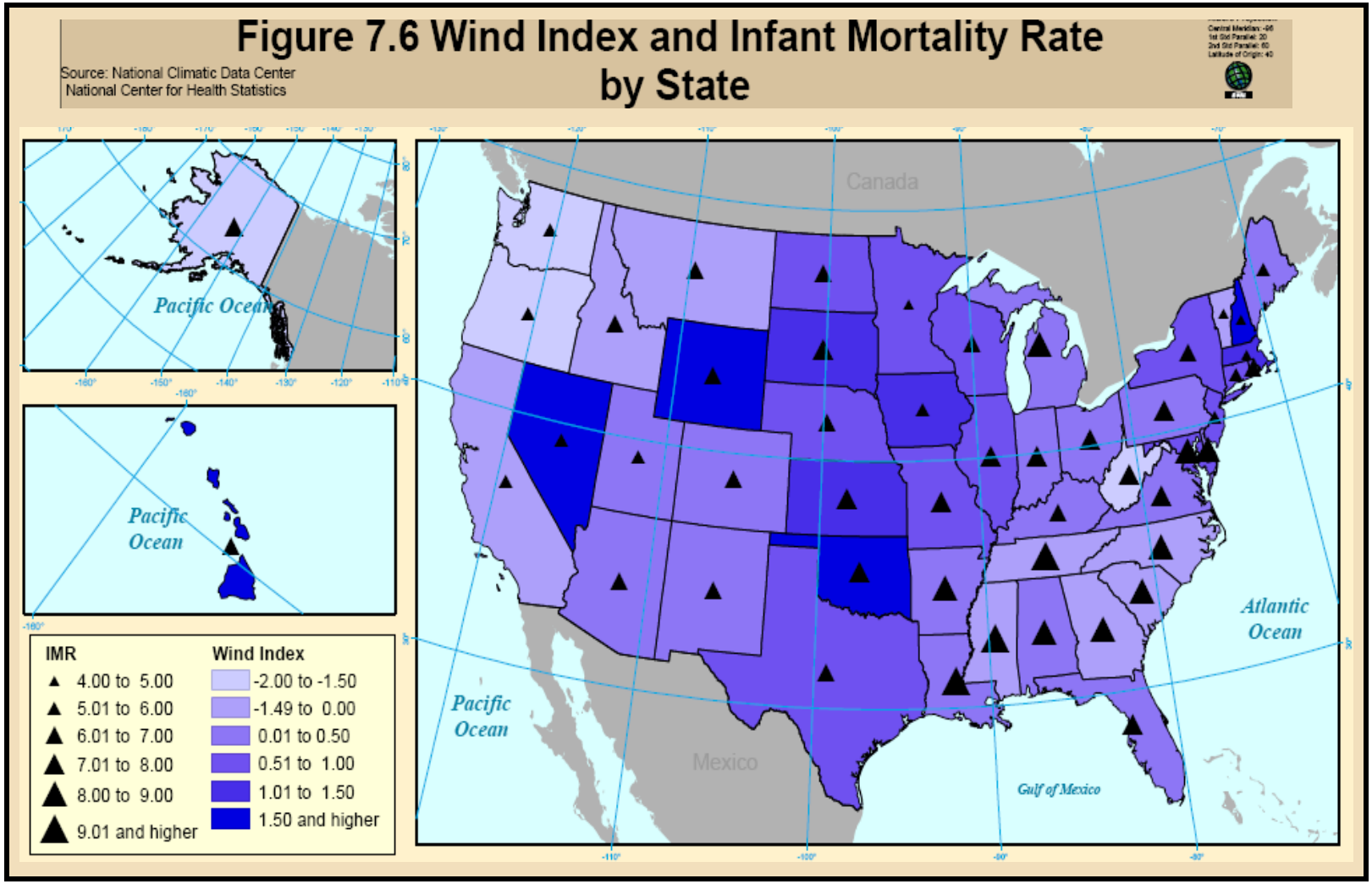


Figure 7.6 Wind Index and Infant Mortality Rate by State
Source: National Climatic Data Center
National Center for Health Statistics
Canada
Pacific Ocean
Pacific Ocean
Pacific Ocean
Atlantic Ocean
Mexico
Gulf of Mexico
IMR
4.00 to 5.00
5.01 to 6.00
6.01 to 7.00
7.01 to 8.00
8.00 to 9.00
9.01 and higher
Wind Index
-2.00 to -1.50
-1.49 to 0.00
0.01 to 0.50
0.51 to 1.00
1.01 to 1.50
1.50 and higher

Conclusions

The purpose of this chapter was to display the spatial relationship between the infant mortality rate and climate among the states of the United States. The intended outcome of the analyses of this chapter was to supplement the multivariate statistical analyses in chapters IV through VI. Although the infant mortality data used in this chapter were not identical to the data used in previous chapters, the infant mortality rate data were used here in conjunction with the climate indices; so we are able to observe the overall or general relationship among the states of the United States that was examined in earlier chapters.

Overall the spatial relationship between the infant mortality rate in the United States and the climate indices is not consistent. In some cases the relationship is shown as expected, with higher rates being observed in the states where the indices are also low. However, in most cases the instances where the relationships seem to perform as hypothesized, it is not supported for all states. For example, although it seems that the higher temperature values are concentrated in the southern United States and the higher infant mortality rates are also found in the southern United States, there are also some high values of the infant mortality rate that are observed in the central United States. So although it may appear that these hypotheses are supported, contradictions to the hypotheses are also observed.

Although my hypotheses were not fully supported with the spatial relationships shown in this chapter, the addition of these maps can tell us something about the connection between infant mortality in the United States.

.

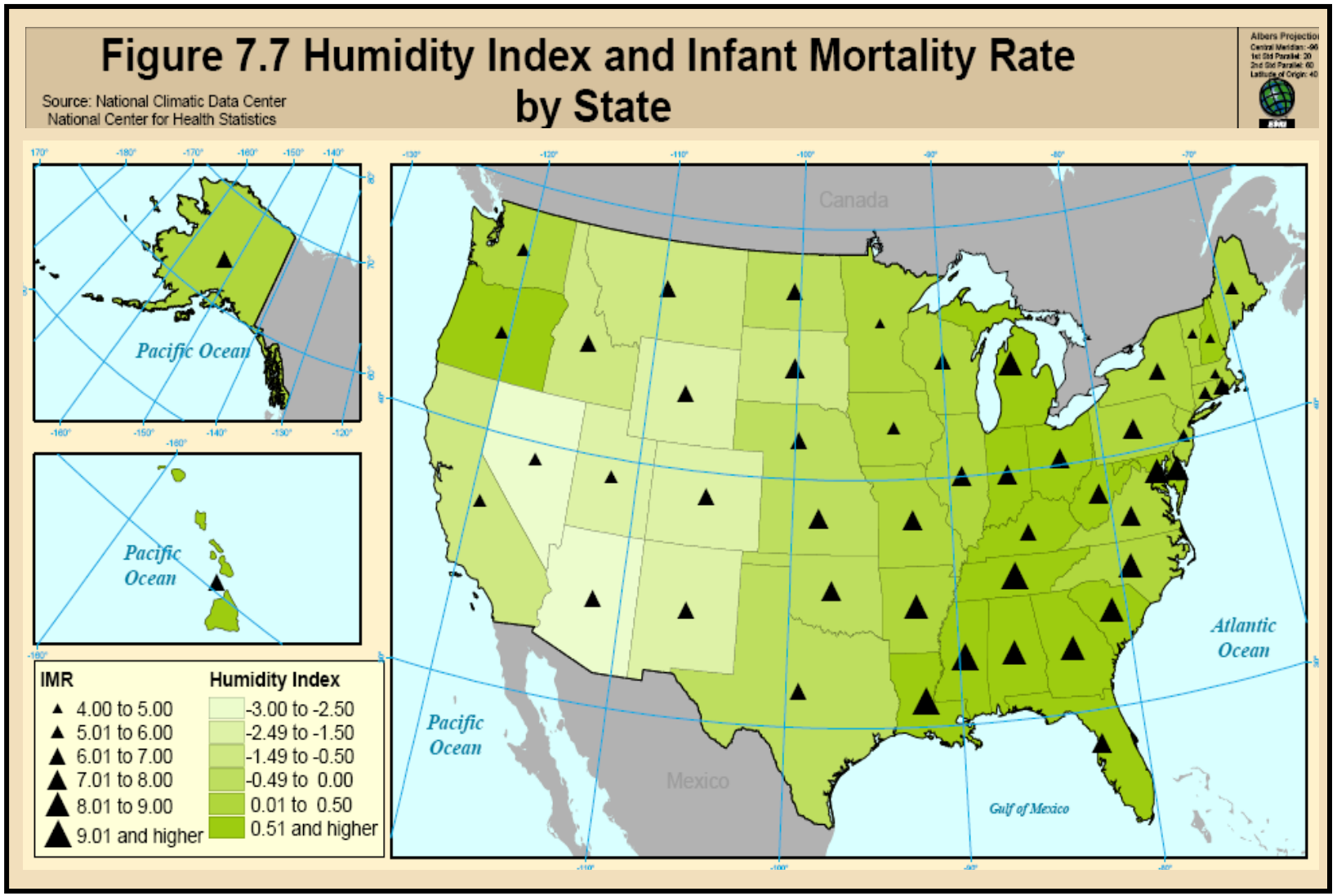


Figure 7.7 Humidity Index and Infant Mortality Rate by State
Source: National Climatic Data Center
National Center for Health Statistics
Albers Projection
Central Meridian: -96
1st Std Parallel: 20
2nd Std Parallel: 60
Latitude of Origin: 40
Pacific Ocean
Pacific Ocean
Canada
Atlantic Ocean
Pacific Ocean
Mexico
Gulf of Mexico
IMR
4.00 to 5.00
5.01 to 6.00
6.01 to 7.00
7.01 to 8.00
8.01 to 9.00
9.01 and higher
Humidity Index
-3.00 to -2.50
-2.49 to -1.50
-1.49 to -0.50
-0.49 to 0.00
0.01 to 0.50
0.51 and higher

First, we know that although the multilevel analyses showed that the state-level climate measures had an impact on the slopes of month of birth and cause specific infant mortality, the same type of connection with the infant mortality rate and the state-level climate measures overall was not observed. The detrimental effects of climate on cause specific infant mortality may not be observed without taking into account the infant's month of birth as was done in previous chapters. Perhaps the climate that an infant or fetus is exposed to only impacts its survival during certain months of birth. This idea coincides with the findings in chapter VI which show the beneficial impact of increases to the temperature and humidity index in months that are typically colder. Without the connection of climate to the infant's month of birth in the maps presented in this chapter, the connections found in chapter VI are not enhanced in the analyses of this chapter.

Second, the maps of this chapter show that there is an observable pattern of infant mortality in the United States. Also we see that the temperature index has a predictable pattern in the United States. The wind and humidity indices also show patterns, although there are exceptions to these patterns. I am referring to those cases where we observe high values of the wind index in Nevada or Maine, as well as high values of the humidity index in Washington and Oregon. Perhaps this less consistent spatial relationship between the wind and humidity indices is a reason that the connection between these indices and the infant mortality rate is not as readily observed in Figures 7.6 and 7.7.

Future research that I wish to undertake in the area of infant mortality and climate will likely address the shortcomings of these maps by incorporating the changes

.

that I addressed in the “data considerations” section of this chapter. Without controlling for any variables, it is hard to find a spatial relationship between infant mortality and climate; however the maps of this chapter do demonstrate that spatial relationships exist for infant mortality and for temperature. While being able to map these variables does have the advantage of displaying this information in a different way, it cannot replace the statistical analyses of the previous chapters.

The real strength of this chapter is the ability to look at state-level variables spatially. I will address the shortcoming of this chapter—namely the inability to control for various geographic and sociodemographic variables in the conclusions of this dissertation. Namely, I will discuss how I would be able to control for various characteristics of the state in similar analyses. Due to time constraints I was unable to undertake these changes in this dissertation, but this is something that I hope to address in future research. Chapter VIII will conclude this dissertation with a summary of findings and a discussion of future research.

.

CHAPTER VIII

CONCLUSIONS AND FUTURE RESEARCH

My dissertation had four main objectives: the first was to examine the relationship between month of birth and infant mortality; the second was to examine the relationship between infant mortality and month of gestation; the third was to employ multilevel modeling to examine the relationship between infant mortality and month of birth with the context, i.e., level-two variables measuring climate; and the last objective was to use maps to visually display the relationship between infant mortality and month of birth while incorporating the contextual climate variables. In this chapter I will discuss the hypotheses of the models tested in this dissertation, summarize my main research and findings, and address the potential for future research in the area of infant mortality and month of birth. I will first discuss some of the principal results from chapter IV, which reported the results of the multinomial logistic regression model using month of birth as the dependent variable.

Level-One Analyses: Month of Birth

Chapter IV used the dependent variable month of birth to examine the relationship between early life and gestation and an infant's chance of survival. The analyses of this chapter tested the hypothesis of adult mortality that finds that those individuals born in the months of October, November and December have lower mortality than those born in the months of April, May and June. This would mean that being born in October, November or December is favorable to an individual's survival as compared to being born in April, May or June. If this same hypothesized association

.

were applied to infant mortality, then we would expect to find negative associations with being born in the months of October, November and December and positive associations with being born in the months April, May and June. This hypothesized connection has found support in the study of adult mortality as well as in the study of schizophrenia. The aim of this chapter and this dissertation as a whole was to ascertain whether this same connection can be applied to cause specific infant mortality in the United States.

I estimated the models of chapter IV in a series of three stages. The distinction between the three stages or models is the way the independent control variables were entered into the model. Following the format of the models of Rogers et al (2000) in their influential book *Living and Dying in the U.S.A.* where the authors studied adult mortality based on different demographic and social variables, I entered the independent variables incrementally. First I entered only the month of birth variables into the model, second I included the month of birth variables and the characteristics of the infant, lastly I included the month of birth variables, the infant characteristics and the maternal characteristics. The rationale behind entering the independent variables in this manner was that first I am able to see the relationship between month of birth and cause specific infant mortality without controls, then when the controls are entered I can observe whether the relationship is sustained after the controls are entered or whether it disappears.

As expected, most of the independent variables measuring infant and maternal characteristics were significantly related to cause specific infant mortality. However, these relationships were not of direct interest in this dissertation and therefore were not

.

interpreted in any detail; their main role in the analyses to include them to determine whether they “explain away” the significant relationships with the month of birth variables alone in model one. The only relationships of interest are those between cause specific infant mortality and month of birth that were maintained when the controls were entered into the models. This is because these significant associations show the true relationship, taking out the effects of the sociodemographic characteristics of the mother or child, between month of birth and cause specific infant mortality that is the focus of this dissertation.

In chapter IV I estimated the series of multinomial logistic regression models twice, once with the independent variable birthweight and next with the independent variable gestational age. There was a good deal of multicollinearity between these two variables, not surprisingly, and therefore they could not be entered into the same model. However, I thought it necessary to estimate two series of models since I believe both gestational age and birthweight were important control variables and excluding either one could possibly lead to a significant relationship between one or more of the month of birth variables and the cause of death that would not be present were these variables included. When estimating both models that were identical otherwise, I found only slight changes between the model with gestational age and the model with birthweight. In terms of significant findings, the month of April was significantly associated with the cause of death of “prematurity” in model three which included gestational age, but not in the model which included birthweight. Otherwise all significant associations were identical in the two models.

.

Overall in chapter IV only four of the ten causes of death showed any significant association between month of birth and infant mortality. These causes were “other infections,” “respiratory causes,” “prematurity and related causes” and “endocrine causes.” Of these four causes of death, the month of January performed as hypothesized (was positively associated) with other infections, the months of February and December performed as hypothesized (were negatively associated) with respiratory causes, April and January performed as hypothesized (were positively associated) with prematurity, and April also performed as hypothesized (was positively associated) with endocrine causes.

Lastly, issues with the non-uniformity of the content of the birth certificates across the states of the U.S. required that the models of chapter IV be rerun. The basic information on the birth certificate is uniform throughout the states; however, certain states choose not to include questions pertaining to the mother or the father. Specifically, in California the birth certificate does not include information on the mother’s alcohol and tobacco use or on her weight gain during pregnancy. I felt that including these variables as controls in the models was important because of the known impacts of alcohol and tobacco use on a fetus. Also, mothers who gain too little or too much weight may be adversely impacting their infant. Since the models I estimated in chapter VI would include state-level measures of climate, I wanted to therefore include all states in the micro-analyses. Therefore the models of this dissertation should include births from California even that that means excluding important control variables. Therefore, in order to assess the impact on the model of excluding the maternal control variables of

.

alcohol and tobacco use and weight gain, I estimated the same models as earlier in chapter IV which excluded California but this time excluded the just mentioned two controls and included California.

There were some differences between the two types of models, although overall the same months of birth were significantly associated in both the models which included California births and the models with excluded California births. Specifically I found that by including California births and therefore excluding the maternal controls, the cause of death of SIDS, which was not significantly associated in any of the models of chapter IV which excluded California, was now significantly associated with the month of birth January. Also the causes of death of other infections and respiratory causes, which did have some significant associations in the previous models, were now significantly associated with September and December respectively. Although these three cases are important to note, I do not believe that it indicates that including California in the multilevel model would be a mistake; therefore I excluded these three control variables in chapter VI.

The analyses undertaken in chapter IV leads to some interesting results. I hypothesized that relationships would be observed between all the months of birth and all the causes of infant mortality in these models. In theory some months should be beneficial to the infant and some months should be detrimental to the infant. However, the majority of the months of birth showed no significant associations in the two versions of model three. I was also surprised to find that more of the months of birth were not significantly related in model one. I expected to find more significant

.

relationships in model one that would ultimately be “explained away” once the infant and maternal characteristics were entered. This turned out not to be the case; in fact the causes of infant death of perinatal infections, birth asphyxia and birth trauma, SIDS, congenital anomalies and neoplasms showed no significant relationships in any of the three models for either series. Only pregnancy complications showed significant associations in model one, which were not found in models two or three. However, finding significant relationships between the four causes of death of other infections, prematurity, endocrine causes and respiratory causes was promising.

In previous literature the associations between month of birth and adult mortality have not been examined among later cohorts. This is primarily because these cohorts do not have completed mortality and therefore cannot be completed, but it is also because the reason for these associations in later cohorts is thought to be not applicable to them. Issues of lack of heating or air conditioning, poor health care for mothers and availability of nutritious foods year round are not seen to be an issue for modern families. The findings of the models I estimated in chapter IV seem to indicate that even though modern families do not have to worry about these basic elements, there is still something about an infant’s month of birth that may be positively or negatively affecting its health. The findings of this chapter also provide a justification for examining cause specific infant mortality and month of birth in the multilevel models of chapter VI, and undertaking and reporting the results of mapping methods of chapter VII. Next, I will discuss the models of chapter V which are a slight variation of the models of chapter IV.

.

The models of chapter V were estimated in an effort to find still additional significant relationships between the month of birth variables and cause specific infant mortality. I suggested that it was possible that any lack of significant relationships could be due to the fact that the underlying mechanism that makes certain months more hazardous to infants is not the months per se of their birth, but instead the months in which they were in utero. I now discuss the main results of those models. The results influenced my decision to ultimately use the month of birth variable instead of the month of gestation variable in the multilevel analyses I undertook in chapter VI.

Level-One Analyses: Months of Gestation

After estimating the two series of three models with the dependent variable of month of birth (as in Chapter IV), I estimated the same series of models with the dependent variable of month of gestation. As just noted, my rationale for changing the dependent variable was to see if the connection between month of birth and infant mortality was an effect of the time the infant was in utero instead of an effect of the month in which the infant was born. In chapter V I undertook the analyses of chapter IV twice, once with the dependent variable months of first trimester, and once with the dependent variable month of third trimester. To operationalize these two variables I took the month of birth variable and the variable measuring the infant's gestational age in weeks and created a variable measuring its month of conception. Because I do not know the infant's exact week or day of birth I considered its day of birth to be in the middle of the month. Although this may lead to potential problems, considering the available information it was the best and really only option. Once this "month of conception"

.

variable was created I could construct the two months of gestation variables to be used in the models of chapter V.

First I created the months of first trimester variable. I consider the first trimester to be important to fetal development because this is the time in which the fetus develops its tissues, organs and body parts. Also the during the first trimester the mother may still not be aware that she is pregnant and therefore may not be making adjustments to her lifestyle that may be beneficial to the infant, such as eating properly, obtaining prenatal care, and avoiding hazardous situations. Accordingly, the first trimester may well be the most important time in a fetus' development, and we may therefore see this in the association between its months of first trimester and cause of death. This variable consists of twelve overlapping measures of the possible combinations of the infant's months of first trimester.

Next, I created the infant's months of third trimester. This variable was also creating using the month of conception variable that I discussed above; however the three month period of interest was the last three months of the fetus' gestation. The most important issue to remember when creating this variable is that not all infants in this dataset were carried to term and therefore never experienced a third trimester. The third trimester is when the fetus does the majority of growth; hence it is not difficult to imagine how important this developmental period should be to a fetus' survival. The months of third trimester variable restricted the number of infants in the data from 92,021 in the months of first trimester models to 32,894 in the months of third trimester

.

models. These 32,894 infants in the months of third trimester models were those who experienced at least one month of their third trimester.

The results of the models of chapter V were not as consistent or as promising as the results of chapter IV. The month of first trimester models only yielded one significant association with the months of birth November, December and January for the cause of death of birth asphyxia and birth trauma that was in the hypothesized direction. All other significant associations were not in the hypothesized direction. The months of third trimester models had many more significant relationships than the months of first trimester models. Again, however, only a couple of the associations were in the hypothesized direction. In the gestational age and birthweight models, perinatal infections were positively associated with the months of April, May and June. Also in the gestational age model, the cause of death of other infections was significantly and positively associated with the months of birth of August, September and October (in the hypothesized direction). Overall, in the birthweight model, the cause of death of other infections yielded the most significant associations of any of the models, in chapter IV or V. However, only one of these relationships, April, May and June, was in the hypothesized direction.

The lack of significant associations in the hypothesized direction found in the months of first and third trimester models led me to use the month of birth as the dependent variable in the models of chapter VI. I was also more confident in the results of chapter IV because there was less variation between the significant findings of the model using birthweight and the model using gestational age. The models of chapter IV

.

were far more consistent in this respect; this indicates a more reliable and robust general model.

The next major step in my dissertation was to estimate the multilevel models using the month of birth as the dependent variable, but this time adding state-level measures of climate at level-two. I will briefly discuss the issues involved in estimating this model and I will also summarize the results and discuss the major findings.

Multilevel Analyses: Month of Birth

The multilevel analyses of this dissertation were undertaken in an effort to elucidate the connection between infant mortality and month of birth. As has already been discussed, the proposed connection between mortality and month of birth lies in the association between the conditions while *in utero* and the long-term detrimental effects on the fetus. By incorporating the level-two (state-level) measures of climate, the analyses of chapter VI provide more of a test of the hypotheses introduced in chapters IV and V.

As I discussed in more detail in chapter VI, I was unable to estimate multilevel multinomial logistic regression with the cause of death data of this dissertation. This occurred because of the size and associated difficulty of the model. As a result I decided to run four logistic regression models, each one with a dichotomous dependent variable for each of the four cause of infant death that were shown to be significantly associated with the month of birth variables in chapter IV. These four logistic regression models included the state level climate variables at level-two. I believe that my decision to change from a multinomial to a logistic regression model is justified and does not pose

.

any severe problems with respect to comparing the results of the analyses of chapters IV and V with the analyses of chapter VI.

Before estimating the multilevel models, I estimated four one-way ANOVA models with the four cause of infant death dependent variables and no independent variables at level-1 and at level-2 so to ascertain whether there was a statistically significant amount of variation in the dependent variables at the state level. Estimating ANOVA models before estimating any multilevel models is allows me to confirm that not all of the variation in the dependent variable occurs at the individual level. If all the variation occurs at the individual level, there is no need to undertake a multilevel model strategy.

The results of the four one-way ANOVA models indicated that although the majority of the variation in cause specific infant mortality occurs at the individual level, enough variation occurs at the state level to justify the use of multilevel methods to examine the impact of climate on infant mortality at the state-level. Specifically, about 6.65 percent of the variation in respiratory related deaths occurs between states; about 6.64 percent of the variation in deaths due to endocrine causes occurs between states; about 3.68 percent of the variation in deaths due to prematurity and related causes is found between states; and about 6.85 percent of the variance in deaths due to “other” infections is found between states.

I will now discuss the results of the four multilevel models reported in chapter VI. Here I will limit my discussion to only those months of birth that were significantly associated in the four models, and I will also only discuss the interaction effects since

.

these are the effects of main interest. Owing to the inherent complexity of multilevel models, i.e., reporting direct effects of the level-1 independent variables, the direct effects of the level-2 independent variables, and the cross-level interactions of the effects of the level-2 independent variables on the slopes of the level-1 variables, I kept to a minimum the total number of level-1 and level-2 independent variables in my models. Thus the models estimated in chapter VI do not include any of the control variables, and only include the month of birth variables of January, February, March, April, October, November and December. The exclusion of the months of May, June, July, August and September means that these five months will serve as the months of comparison for the models. This is a slight variation from the models of chapters IV and V, where only the month of May served as the comparison month. The months of June, July August and September did not show any significant associations in the analyses I undertook in chapter IV, and the month of May was excluded because it was the reference month in the earlier analyses.

I will discuss the findings in terms of the hypotheses of the multilevel models. Since the principal feature of the multilevel models of interest in this dissertation is the interaction effects between the infant's month of birth and the slope of dying of the specified cause, the hypotheses vary from those seen in chapters IV and V. Instead of hypothesizing a direct effect of the month of birth on the infant's chances of dying of a specified cause, in chapter VI I am interested in whether and how the level two variables interact with the month of birth variable to alter the slope of month of birth on the chances of dying. I hypothesized that *increases* in the temperature index would *decrease*

.

the slope of the month of birth on dying of the specified cause for all causes of death. Next, I hypothesized that *increases* in the humidity index would *increase* the risk of dying of the specified cause for all causes of death. Lastly, I hypothesized that *increases* in the wind index would *decrease* the slope of the month of birth on dying of the specified cause for all causes of death.

The first multilevel logistic regression model uses the dichotomous dependent variable of prematurity and related causes as the cause of infant death. In this model the months of January, February, March, November and December showed significant effects. Of these effects the level-two humidity variable did not behave as expected with any of the month of birth variables. Specifically, the interaction effect of the humidity variable was found to negatively impact the slope of the month of birth and prematurity slope. The months of birth of January, February, November and December yielded negative coefficients with the humidity index. This means that each one-unit increase on the humidity index, the month of birth-prematurity slope was decreased, not increased as I hypothesized.

In the prematurity model the month of birth March yielded a significant association with the level-two variable wind. This association was positive, which is the hypothesized direction of association. Specifically this means that for every one-unit increase in the wind index, the slope of being born in March on dying of prematurity is increased, that is the slope becomes steeper. For the months of birth of November and December, the level-two variable of temperature was negatively associated, which was the direction I hypothesized; this means that for each unit increase in the temperature

.

index the slope of being born in November or December is decreased, meaning it becomes less steep.

The second multilevel logistic regression model used the dichotomous dependent variable of respiratory and related causes as the cause of infant death. In this model only the months of April and December yielded significant interaction effects. Of these effects the level-two variables performed as hypothesized. Specifically, the month of birth of April showed a significant interaction effect with the level-two variable humidity. In this case, the humidity index was positively related, which is the hypothesized direction. Also, the month of birth of December showed a significant interaction effect with the level-two variable temperature variable. The direction of the interaction was negative, meaning that increases in the temperature index would decrease the slope of the month of birth on the specified cause of death.

The next multilevel logistic regression model used the dichotomous dependent variable of endocrine conditions as the cause of infant death. In this model the months of January, April, November and December showed significant interaction effects. For the months of birth of January and April, I found the significant interaction effects as hypothesized. Specifically, for the effect on death of the month of birth of January, the temperature index is negatively associated; meaning that for every one-unit increase in the temperature index the slope of January-endocrine causes is decreased. For the month of birth of April slope, the level-two variable humidity is positively related, meaning that with every one-unit increase in the humidity index, the slope of April-endocrine causes is increased. For model three the month of birth of November slope has a significant

.

interaction effect with the level-two variable humidity; however the direction of the association is negative, which is the opposite of the hypothesized direction. Similarly, the level-two variable wind is significantly and positively related with the month of birth of December slope, which is in the opposite of what I hypothesized.

The last multilevel logistic regression model uses the dependent variable of "other infections" as the cause of infant death. In this model the slopes of months of March, October, November and December yielded significant associations with the level-two climate variables. However, only the effect of the variable of temperature and the slope involving the month of birth of December is in the hypothesized direction. This means that for every one-unit increase in the temperature index, the slope of December on the "other infections" cause of death is decreased, that is it becomes less steep. The slopes involving the months of birth of March, October and November showed significant interaction effects with the level-two variables of humidity and/or wind; however all these interactions were in the opposite direction of those hypothesized.

Similar to the results of chapters IV and V, the results of the multilevel analyses are mixed. In some cases the hypotheses were supported; but there were also many instances where the coefficients for the interaction effects were not significant, and there were also instances where the interaction effects were in the opposite direction predicted by the hypotheses. Here I will discuss the findings of the four multilevel models and what I believe these findings mean with regard to the overall hypotheses of this dissertation, namely, that the conditions of early life affect an infant's likelihood of dying from a specified cause.

.

The first model of chapter VI uses the dependent variable of prematurity and related causes. In this model, in those cases where the level-two variable of humidity was significantly associated with the slopes of the month of birth variables, the association was negative, not positive as hypothesized. This is also seen in the "endocrine" and "other infections" models. Only in the "respiratory causes" model is the humidity variable positively associated with a month of birth slope, and in this case it is only positively associated with the month of birth slope involving April. This may indicate that the hypothesized connection between humidity and month of birth may need to be revised. It may be the case that increases in humidity will serve to warm the temperatures of usually cold months such as January, February, November and December. These also happen to be the months whose slopes show negative associations in the models of chapter VI. Only the month of April slope in the models for respiratory causes and endocrine causes shows a positive association, meaning that the slope of these months on the likelihood of dying of the specified cause is increased with increases in the humidity index. The only instance where this hypothesis is not confirmed is with regard to the slope involving the month of March in the "other" infections model, where humidity is negatively associated with the slope.

The level-two variable temperature shows more consistent results. Specifically in all cases where the temperature variable is significantly related to the month of birth slopes in all of the four models, the associations are negative. The associations are also found exclusively in the months of November, December and January in all four models. This is interesting because these are the months in which in the U.S., the temperatures

.

are typically low. In these months, increases in the temperature index tend to decrease the slope of the month of birth on the specified cause of infant death.

The last level-two variable, wind, performs as expected in the three of the four models where significant associations are found. In all cases wind is positively associated with the month of birth slopes with infant death. This means that with increases in the wind index, the slope of the given month of birth and the specified cause of infant death increases, meaning the slope becomes steeper. These findings are also concentrated in the winter and late fall months, when we would expect conditions to be colder and harsher overall. Increases in the wind index for a state may well adversely affect an infant's cause of death because it makes the conditions more severe in the winter months.

Overall I believe the findings of the multilevel analyses of chapter VI show support for the idea that there is a legitimate connection between climate, the month of birth and the infant's ultimate likelihood of dying of a specified cause as compared to external causes of death. In most cases, the level-two climate variables were associated with the level-1 slopes in the hypothesized direction and support the fact that more severe weather conditions—heavier winds, or colder temperatures—apparently tend to have a detrimental effect on an infant's health. These multilevel models support the fact that the microlevel relationships observed in chapters IV and V may well be due in part to such contextual (state-level) characteristics as climate conditions, and not to some unmeasured relationships involving unspecified social variables and the infant's month of birth.

.

It may be the case that the relationship between humidity and the slopes of months of birth and cause specific infant mortality reflects reality, even though the reported relationship is in the opposite of that hypothesized. Instead of concluding that the reason for the lack of significant relationships in the hypothesized direction is because humidity does not affect cause specific infant mortality, I would conclude that there is an important relationship that I need to consider in my rethinking. Specifically, in the colder months, increases to humidity can be beneficial to the infant's cause of death, instead of increases being detrimental to an infant's cause of death in the warmer months. This makes sense when we consider the overall hypotheses of this dissertation, which state that difficulties during the time of gestation in the months of fall and/or winter are detrimental to a fetus, and may well influence its cause of death. If the colder months are those that we consider to be detrimental, then increases to humidity will make these months milder and therefore the influence on the infant may be seen as positive (a negative coefficient which would mean a decreasing slope) and not as negative (a positive coefficient which would mean an increasing slope).

The final substantive chapter of this dissertation used mapping methods to visually display the relationship between the infant mortality rate in the United States and the state-level measures of climate developed and used in the models in chapter VI. I now discuss the main findings of chapter VII and what I believe these maps can tell us about the relationship between infant mortality and climate in the United States.

.

Mapping of Infant Mortality and Climate

The last analysis chapter of this dissertation used mapping methods to spatially examine climate and the infant mortality rate in the United States. This chapter used maps to visually display the relationship between the state level measures of climate used in chapter VI and the infant mortality rate. As I discussed in chapter VII, I chose to use the infant mortality rate as opposed to any of the cause specific infant mortality rates used in chapters IV through VI because of the small number of cases in many of the states for the majority of the causes of death. Although this means that the maps of chapter VII do not exactly mirror the analyses reported in chapter VI, I believe the choice to include these maps with the infant mortality rate is justified, since these maps will serve as another way to show the relationship between infant mortality and climate. It is important, however, to note the differences between the data used in this chapter so that direct comparisons will not be made.

The climate measure temperature, not surprisingly, shows a pattern of higher temperatures in the southern part of the United States with decreasing temperatures into the northern part of the country. The wind and humidity indices are less consistently distributed throughout the country. However, it does seem that, overall, higher values of the wind index are seen in the middle of the states, with lower values on either coast. High values on the humidity index are also seen in the Pacific Northwest, in Hawaii and in Alaska.

The fourth and final univariate map contains the average infant mortality rate by state for the years 2000 to 2004. The infant mortality rate for the years 2000 to 2004

.

seems to be the highest in the southern states, as seen by the larger triangles especially concentrated in the states of Mississippi, Louisiana and Tennessee. We can see the smaller triangles concentrated in the western portion of the map, indicating lower levels of infant mortality in the states of this area.

The real focus of chapter VII is seen in the bivariate maps which displayed the infant mortality rate with each of the three climate indices. The first of these bivariate state maps shows the temperature index with the infant mortality rate. From this map we observe higher temperature index values are concentrated in the Southern United States. We also see a high concentration of higher infant mortality rates in the Southeastern United States. This relationship is the opposite of the relationship reported in the multilevel models of chapter VI, where the higher temperatures reduced the relationship between month of birth and cause specific infant mortality. Of course it must be remembered that the map relationship is reporting the spatial association between temperature and the infant mortality rate, and, therefore, does not include the month of birth data

The second bivariate map displays the wind index and the infant mortality rate. Again, we see the higher infant mortality rates concentrated in the Southeastern United States. However, this does not seem to coincide with higher values on the wind index as was hypothesized. The third and final bivariate map displays the humidity index and infant mortality rate by state. Higher values on the humidity index seem to be concentrated in the Southeast to Northeast of the United States. The hypothesized connection between infant mortality and the humidity index is that higher values of

.

humidity will lead to lower levels of the infant mortality rate does not seem to be supported by the findings of this map.

The purpose chapter VII was to display the spatial relationships between the infant mortality rate and climate in the United States. Overall the spatial relationships between the infant mortality rate in the United States and the climate indices are not fully consistent with the results of my earlier analyses. In some cases the relationship is as expected, but, in most cases the results seemed to be mixed, that is, the relationships hypothesized were not found in all states. While the mapping of these variables does have the advantage of displaying this information in a different way, it cannot fully replace the statistical analyses of the previous chapters. The real strength of the maps of chapter VII is the ability to look at variables that are measured by state, spatially in the United States.

Discussion and Future Research

When thinking about this dissertation now that it is completed, there are several things that I would have done if better data had been available. First, I would have liked to have estimated a hazard model, which would have allowed me to compare infants based on their month of birth. For this type of model however, the data need to contain all infants, some of whom died and some of whom did not die during their first year of life. This type of data would be matched as opposed to linked. The linked file data used in this dissertation do not allow me to estimate hazard models because the data only contain those infants who died, not all infants born in a specific year. So although the Linked Birth/Infant Death dataset has many advantages, one disadvantage is that it does

.

not contain subjects who survived, hence not permitting using its data to estimate hazard models.

Another area in which better data would have enhanced my models would be with regard to the socioeconomic status of the mother and father. In this dissertation I used the mother's education as a proxy for socioeconomic status (SES). However, if it were available, information on income or occupation of the mother and information on the education, occupation and income of the father would add significantly to my operationalization of SES. I would like to see if lower SES of the mother and/or father enhanced the relationship between an infant's month of birth and its ultimate cause of death. I suspect that those parents with lower SES would have steeper slopes of month of birth on cause specific infant mortality than parents with higher SES. More extensive socioeconomic data in this area would have contributed greatly to my analyses in this dissertation.

Also, a better measure of month of gestation could well have dramatically changed the outcome of the models reported in chapter V. Even though the results of chapter V indicate that month of birth is the more important measure of infant's chance of survival, I believe that a finer measure of month of gestation could well have shown that the period of gestation turns out to be the most important for the infant's survival. If the connection between month of birth and infant mortality is indeed due to the conditions *in utero*, then month of gestation should also show this relationship. If exact date of birth were available to construct a more precise measure, it would surely have allowed the estimation of better models in chapter V.

.

I hope that my future research in the area of infant mortality and month of birth will expand and extend the analyses of this dissertation in several ways. First, I would like to make further restrictions based on age of the infant and model these age groups separately. This would entail making restrictions based on age of the infants at death and looking at their cause of death. This may well necessitate combining more years of data in order to obtain a sufficiently large sample. Also, it would be useful to only look at infants who were born to term. This would drastically reduce the number of infant deaths since the majority of the infants were born preterm. However, I believe looking only at those infants who were born at term would allow me to draw better conclusions about the association between month of birth and cause specific infant mortality.

Another area that I would like to explore in the future has to do with extending the models of my dissertation to child mortality. By looking at the birth cohorts of 2000 to 2004 in childhood, I should be able to extend my hypotheses accordingly. Once this cohort has "completed" their childhood years, I would be able to make similar inferences about mortality in childhood. This would be especially interesting with regard to the study of chronic ailments in childhood, such as allergies, asthma, and mental disorders. Although mortality is low throughout childhood, especially from chronic conditions, being able to examine these infants through their childhood would expand research in this area.

Available data from the linked birth infant death dataset would also allow me to look at alternate health outcomes, which will be another of my further research objectives. For example, in this dissertation I used infant's birth weight and gestational

.

age as control variables in the multinomial logistic regression models, the results of which I reported and discussed in chapters IV and V. In future research it may be interesting to use these variables not as controls, but as outcomes in multiple regression models. Information is also available on the APGAR score at the time of the infant's birth. "At birth the APGAR score is used to assess neonatal condition and frequently is correlated with the data obtained by fetal monitoring techniques" (Silverman et al. 1985:332). This score is included on all birth certificates; analysis of APGAR variability would be yet another strategy for extending this dissertation beyond my analyses of infant's cause of death. Research in the area of the effects of month of birth on infant mortality may benefit from the research on alternate health outcomes.

As discussed in chapter VII a shortcoming of the maps used to display the spatial relationship between infant mortality and month of birth is the inability to control for characteristics of the infant and mother as I did in the multinomial logistic regression of chapters IV and V. There is, however, a way to overcome this shortcoming and display maps that do control for the influences that the characteristics of the mother and child have on an infant's chance of survival. It was suggested that in order to control for variable known to effect infant mortality, I could compute the predicted infant mortality rates by state. This could be done by performing the multivariate analysis with the selected control variables, and then from the generated coefficients calculating the predicted infant mortality rate by the state of death. This would also allow me to include the state-specific measures of the climate variable. The resulting predictions would be the state-specific predictions for infant mortality that vary with climate which hold the

.

effects of the control variables constant. This ability to control for sociodemographic variables would be a great strength of taking this type of step in the mapping of infant mortality. In the future, if I use maps to expand my research on infant mortality by month of birth and climate, I will use the method of generating predicted infant mortality rates by state. I believe this type of map would greatly contribute to the analyses of my dissertation, as well as contribute to the literature in the area of infant mortality.

Conclusion

Infant mortality is and will continue to be an important area of research in health and demographic studies. Not only is infant mortality a sensitive and crucially relevant issue, it is also used as an indication of the overall health and wellness of a nation. Causes and ways to reduce infant mortality throughout the world will likely be the emphasis of research for many years. By looking at the association of an infant's cause of death and its month of birth, this dissertation shed some light on the conditions experienced during gestation as well as in very early life on the likelihood that an infant will die of a specific cause. While it is unlikely that knowledge of the association between month of birth and cause specific infant mortality will lead to any great reductions in infant mortality, it may lead to a better understanding of the early life mechanisms that affect an infant's chance of survival.

.

REFERENCES

Agha, S. 2000. "The Determinants of Infant Mortality in Pakistan." *Social Science and Medicine* 51(2):199-208.

Agresti, A. 1996. *Categorical Data Analysis* (2nd edition). New York: Wiley & Sons.

Ahmed, F., O. Shisana and F. Saadatmand. 1991. "Leading Causes of Infant Deaths in the District of Columbia." *Journal of the National Medical Association* 83(8): 665-673.

Arias, E. and B. Smith. 2003 "Deaths: Preliminary Data for 2001." *National Vital Statistics Reports.* 51(5):1-45.

Ballester, F., P. Michelozzi and C. Iniguez. 2003. "Weather Climate and Public Health." *British Medical Journal.* 57:759-760.

Barker, D.J. 2001. "Fetal and Infant Origins of Adult Disease." *Monatsschrift Kinderheilkunde* 149(1):S2-S6.

Barker, D.J. and C. Osmond. 1986. "Infant Mortality, Childhood Nutrition, and Ischemic Heart Disease in England and Wales." *The Lancet* i:1077-1081.

Bateman, I.J., A.P. Jones, A.A. Lovett, I.R. Lake and B.H. Day. 2002. "Applying Geographical Information Systems (GIS) to Environmental and Resource Economics." *Environmental and Resource Economics* 22(1-2):219-269.

Ben-Shlomo, Y. and G.D. Smith. 1991. "Deprivation in Infancy or in Adult Life: Which Is More Important for Mortality Risk?" *The Lancet* 337:530-534.

.

Brown, A.S., M.D. Begg, S. Gravenstein, C.A. Schaefer, R.J. Wyatt, M. Bresnahan, V.P. Babulas and E.S. Susser. 2004. “Serologic Evidence of Prenatal Influenza in the Etiology of Schizophrenia.” *Archives of Geriatric Psychiatry* 61:774-780.

Buehler, J.W., J.C. Kleinman, C.J. Hogue, L.T. Strauss and J.C. Smith. 1987. “Birth Weight-Specific Infant Mortality, United States, 1960 and 1980. “ *Public Health Reports* 102(2):151-161.

Bryk A. and S.W. Raudenbush. 1992. *Hierarchical Linear Models: Applications and Data Analysis Methods: Advanced Qualitative Techniques in the Social Sciences.* Thousand Oaks, CA: Sage.

Campbell, M.J. 1994. “Time Series Regression for Counts: An Investigation into the Relationship between Sudden Infant Death Syndrome and Environmental Temperature.” *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 157(2): 191-208.

Centers for Disease Control. 2008. “2006-2007 U.S. Influenza Season Summary.” Available online at: http://www.cdc.gov/flu/about/qa/season.htm (accessed on 9/1/2008).

Cheney, Rose A. (1984). “Seasonal Aspects of Infant and Childhood Mortality: Philadelphia, 1865-1920.” *Journal of Interdisciplinary History* 14(3):561-585.

Central Intelligence Agency. 2006. *World Factbook*. Available online at: https://www.cia.gov/library/publications/the-world-factbook/

Clinch, J.P. and J.D. Healy. 2000. “Housing Standards and Excess Winter Mortality.” *Journal of Epidemiology and Community Health* 54:719-720.

.

Crow, T.J. and D.J. Done. 1992. “Prenatal Exposure to Influenza Does Not Cause Schizophrenia.” *The British Journal of Psychiatry* 161:390-393.

de Leeuw, J. and E. Meijers 2008. “Introduction to Multilevel Analysis.” Pp. 1-76 in *Handbook of Multilevel Analysis,* edited by J. de Leeuw and E. Meijers. New York, NY: Springer.

Doblhammer, G. 2004. *The Late Life Legacy of Very Early Life.* Berlin, Germany: Springer.

Doblhammer, G. and J. Vaupel. 2001. “Lifespan Depends on Month of Birth.” *Proceedings of the National Academy of Sciences of the United States of America*. 9(5):2934-2939.

Dollfus, C., M. Patetta, E. Siegel and A.W. Cross. 1990. “Infant Mortality: A Practical Approach to the Analysis of the Leading Causes of Death and Risk Factors.” *Pediatrics* 86(2): 176-183.

Donaldson, G.C. and W.R. Keatinge. 1997. “Early Increases in Ischemic Heart Disease Mortality from and Later Changes Associated with Respiratory Mortality after Cold Weather in South East England.” *Journal of Epidemiology and Community Health.* 51:643-648.

Eastman, P.R. 1945. “Infant Mortality in Relation to Month of Birth.” *American Journal of Public Health* 35:913-922.

Eberstein, I.W., C.B. Nam and R.A. Hummer. 1990. “Infant Mortality by Cause of Death: Main and Interaction Effects.” *Demography* 27(3):413-430.

.

Eberstein, I.W. and J.R. Parker. 1984. “Racial Differences in Infant Mortality by Cause of Death: The Impact of Birth Weight and Maternal Age.” *Demography* 21(3):309-321.

Ewald, P. 2002. “Mastering Disease.” Pp. 289-301 in *The Next Fifty Years: Science in the First Half of the Twenty-First Century,* edited by J. Brockman. New York: Vintage Books.

Fiscella, K. 1995. “Does Prenatal Care Improve Birth Outcomes? A Critical Review” *Obstetrics and Gynecology* 85(3):468-479.

Forsdahl, A. 1977.” Are Poor Living Conditions in Childhood and Adolescence an Important Risk Factor for Atherosclerotic Heart Disease?” *British Journal of Preventative Social Medicine* 31:91-95.

Frisbie, W. P. 2005. “Infant Mortality.” Pp. 251-282 in *Handbook of Population,* edited by D.L. Poston and M. Micklin. New York: Kluwer Academic.

Galloway, P.R. 1985. “Annual Variations in Deaths by Age, Deaths by Cause, Prices and Weather in London 1670 to 1830.” *Population Studies* 39(3):487-505.

Gortmaker, SL. 1979. “The Effects of Prenatal Care upon the Health of the Newborn.” *The American Journal of Public Health.* 69(7): 653-660.

Gortmaker, S.L. and P.H. Wise. 1997. “The First Injustice: Socioeconomic Disparities, Health Services Technology, and Infant Mortality. *Annual Review of Sociology* 23:147-170.

.

Guy, W.A. and M.B. Cantab. 1843. “An Attempt to Determine the Influence of the Seasons and Weather on Sickness and Mortality.” *Journal of the Statistical Society of London* 6(2):133-150.

Hamilton, L.C. 1992. *Regression with Graphics: A Second Course in Applied Statistics*. Belmont, CA: Duxbury Press.

Hare, E.H., PA. Moran and A. MacFarlane. 1981. “The Changing Seasonality of Infant Deaths in England and Wales 1912-1978 and Its Relation to Seasonal Temperature.” *Journal of Epidemiology and Community Health* 35:77-82.

Haub, C. 1991. “Infant Mortality: Who’s Number One?” *Population Today* 19(3):6-8.

Healy, J.D. 2002. “Excess Winter Mortality in Europe: A Cross Country Analysis Identifying Key Risk Factors.” *Journal of Epidemiology and Community Health* 57:784-789.

Heron, M. 2007. “Deaths: Leading Causes for 2004.” *National Vital Statistics Report.* 56(5): 1-95.

Hetzel, A.M. 1997. *History and Organization of the Vital Statistics System*. Hyattsville, MD: National Center for Health Statistics Press.

Hogue, C.J. and M.A. Hargraves. 1993. “Class, Race and Infant Mortality in the United States.” *American Journal of Public Health.* 83(1):9-12.

Huntington, E. 1938. *Season of Birth.* Oxford, England: Wiley & Sons.

Kalkstein, L.S. and R.E. Davis. 1989. “Weather and Human Mortality: An Evaluation of Demographic and Interregional Responses in the United States.” *Annals of the Association of American Geographers* 79(1):44-64.

.

Keatinge, W.R. 1986. “Seasonal Morality among Elderly People with Unrestricted Home Heating.” *British Medical Journal* 293:732-733.

Kermack, W.O., A.G. McKendrick and P.L. McKinlay. 2001. “Death-Rates in Great Britain and Sweden: Some General Regularities and Their Significance.” *International Journal of Epidemiology* 30: 678-683.

Kiely, J.L. and M.D. Kogan. 1994. “Prenatal Care.” Pp.105-118 in *From Data to Action: CDC’s Public Health Surveillance for Women, Infants and Children.* edited by L.S. Wilcox and J.S. Marks. Atlanta, GA: U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control (CDC).

Kogan, M.D., J.A. Martin, G.R. Alexander, M. Kotelchuck, S.J. Ventura and F.D. Frigoletto. 1998. “The Changing Pattern of Prenatal Care Utilization in the United States, 1981-1995, Using Different Prenatal Care Indices.” *Journal of the American Medical Association* 279(20):1623-1628.

Lam, D.A. and J.A. Miron. 1994. “Global Patterns of Seasonal Variation in Human Mortality.” *Annals of the New York Academy of Sciences* 709(1):9-28.

LaVeist, T.A. 1989. “Linking Residential Segregation and Infant Mortality in U.S. Cities.” *Sociology and Social Research* 73:90-94.

Lerchl. A. 1998. “Changes in the Seasonality of Mortality in Germany from 1946-1995: The Role of Temperature.” *International Journal of Biometeorology* 42(2):84-88.

Long, J. S. and J. Freese. 2001. *Regression Models for Categorical Dependent Variables Using Stata*. College Station, TX: Stata Press.

.

Markides, K.S. and J. Coreil. 1986. “The Health of Hispanics in the Southwestern United States: An Epidemiological Paradox.” *Public Health Reports* 101(3):253-265.

Markides K.S. and K. Eschbach. 2005. “Aging, Migration and Mortality: Current Status on the Hispanic Paradox.” *The Journals of Gerontology: Series B; Psychological Sciences and Social Sciences* 60(2): 68-75.

Matteson, D.W., J.A. Burr and J.R. Marshall. 1998. “Infant Mortality: A Multilevel Analysis of Individual and Community Risk Factors.” *Social Science and Medicine* 47(11):1841-1854.

Matthews, T.J. and M.F. MacDorman. 2008. “Infant Mortality Statistics from the 2005 Period Linked Birth/Infant Death Data Set.” *National Vital Statistics Reports* 57(2):1-32.

McCormack, M.C. and P.H. Wise. 1993. “Infant Mortality.” *Current Opinion on Pediatrics.* 5(5):552-557.

McKee, C.M. 1989. “Deaths in Winter: Can Britain Learn from Europe?” *European Journal of Epidemiology* 5(2):178-182.

Mednick, S.A., M.O. Huttunen and R.A. Machon. 1994. “Prenatal Influenza and Adult Schizophrenia.” *Schizophrenia Bulletin* 20(2):263-267.

Mercer, J.B. 2003. “Cold—An Underrated Risk Factor for Health.” *Environmental Research* 92(1):8-13.

Moore, Keith L. 1974. *Before We Are Born: Basic Embryology and Birth Defects*. Philadelphia: Saunders Company

.

Murphy, M.F. and M.J. Campbell. 1987 “Sudden Infant Death Syndrome and Environmental Temperature: An Analysis Using Vital Statistics.” *Journal of Epidemiology and Community Health* 41:63-71.

National Center for Health Statistics. 2010. Data available online at: http://www.cdc.gov/nchs/linked.htm#Microdata. Accessed on April 2010.

National Climatic Data Center. 2010. Data available online at: http://www.ncdc.noaa.gov/oa/mpp/freedata.html#FREE. Accessed on April 2010.

Nersesian, W. 1988. “Infant Mortality in Socially Vulnerable Populations.” *Annual Review of Public Health* 9(1): 361-377.

Ormsby, T. E. Napoleon, R. Burke, C. Groessl and L. Feaster. 2001. *Getting to Know ArcGIS Desktop: Basic View of ArcView, ArcEditor and ArcInfo.* Redlands, CA: ESRI Press.

Osmond, Clive and D.J. Barker. 2000. “Fetal, Infant and Childhood Growth Are Predictors of Coronary Heart Disease, Diabetes and Hypertension in Adult Men and Women.” *Environmental Health Perspectives* 108(3):545-553.

Polednak, A.P. 1991. “Black-White Differences in Infant Mortality in 38 Standard Metropolitan Statistical Areas.” *American Journal of Public Health* 81(11):1480-1482.

Polednak, A.P. 1996. “Trends in U.S. Urban Black Infant Mortality, by Degree of Residential Segregation.” *American Journal of Public Health* 86(5): 723-726.

.

Poston, D.L. and R.G. Rogers. 1985. “Toward a Reformation of the Neonatal Mortality Rate.” *Social Biology* 32:1-12.

Poston, D.L., L. Zhang, D.J. Gotcher and Y. Gu. 2009. “The Effect of Climate on Migration: United States, 1995-2000.” *Social Science Research* 38:743-753.

Prager, K. G. A. Flinchum and D.P. Johnson. 1987. “The NCHS Pilot Project to Link Birth and Infant Death Records: Stage 1.” *Public Health Report.* 102(2):216-223.

Rana, M. W. 1998. *Human Embryology Made Easy*. Amsterdam, The Netherlands: Overseas Publishing Association.

Rau, R. 2007. *Seasonality in Human Mortality: A Demographic Approach.* Berlin, Germany: Springer.

Raudenbush, S.W. and A.S. Bryk 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods.* Thousand Oaks, CA: Sage.

Reichert, T.A., A. Sharma and S. Pardo. 2001. “A Global Pattern for Influenza Activity.” *International Congress Series* 1219:87-94.

Reichert, T.A., L. Simmonsen, A. Sharma, S.A. Pardo, D.S. Fedson and M.A. Miller. 2004. “Influenza and the Winter Increase in Mortality in the United States, 1959-1999.” *American Journal of Epidemiology* 160(5):492-502.

Rowland, D. 2003. *Demographic Methods and Concepts.* Oxford, England: Oxford University Press.

Semenza, J.C., C.H. Rubin, K.H. Falter, J.D. Selanikio, D. Flanders, H.L. Howe and J.L. Wilhelm. 1996. “Heat Related Deaths During the July 1995 Heat Wave in Chicago.” *The New England Journal of Medicine* 335:85-90.

.

Siegel, J. and Swanson, D. eds. 2004. *The Methods and Materials of Demography (*2nd edition). San Diego, CA: Elsevier Academic Press.

Silverman, F., J. Suidan, J. Wasserman, C. Antoine and B.K. Young. 1985. "The Apgar Score: Is It Enough?" *Obstetrics and Gynecology* 66(3):331-336.

Thompson, W. W., D.K. Shay, E. Weintraub, L. Brammer, N. Cox, L.J. Anderson and K. Kukuda. 2003. "Mortality Associated with Influenza and Respiratory Syncytial Virus in the United States." *Journal of the American Medical Association* 289:179-186.

Tochigi, M., Y. Okazaki, N. Kato and T. Sasaki. 2004. "What Causes Seasonality of Birth in Schizophrenia?" *Neuroscience Research* 48:1-11.

Torrey, E.F., A.E. Bowler, R. Rawlings and A. Terrazas. 1993. "Seasonality of Schizophrenia and Stillbirths." *Schizophrenia Bulletin* 19(3):557-562.

Torrey, E.F., J. Miller, R. Rawlings, R.H. Yolken. 1997. "Seasonality of Births in Schizophrenia and Bipolar Disorder: A Review of the Literature. *Schizophrenia Research* 28(1):1-38.

Tramer, M. 1929. "Uber die Biologische bedeutung des Geburtsmonates, Imbesodere fur die Psychoseerkrankung Schweiz." *Archives of Neurological Psychiatry* 34:17-24.

Trussell, J. and C. Hammerslough, 1983. "A Hazards-Model Analysis of the Covariates of Infant and Child Mortality." *Demography* 2(1):1-26.

Udry, J.R. and N.M. Morris. 1967. "Seasonality of Coitus and Seasonality of Birth." *Demography* 4(2):673-679.

.

U.S. Department of Health and Human Services. *Healthy People 2010: Understanding and Improving Health*. 2nd ed. Washington, DC: U.S. Government Printing Office, November 2000.

Williams, A.L., E.C. Uren and D.P. Southall. 1987. "Respiratory Viruses and Sudden Infant Death Syndrome." *British Medical Journal* 288:1491-1493.

Williams D.R., S.J. Robert and T.W. Davies. 1979. "Deaths from Ischemic Heart Disease and Infant Mortality in England and Wales." *Journal of Epidemiology and Community Health* 33:199-202.

Wise, P.H. 1994. "What You Measure Is What You Get: Prenatal Care and Women's Health." *American Journal of Public Health* 84(9):1374-1375.

.

VITA

Rachel Traut Cortes
909 New Jersey Ave SE Apt 615 Washington, D.C. 20003
(210)325-1173 racheltcortes@gmail.com

Education

2010 **Ph.D. in Sociology** Dept. of Sociology, Texas A&M Univ., College Station, TX
2007 **M.S. in Sociology** Dept of Sociology, Texas A&M Univ., College Station, TX
2004 **B.S. in Sociology** Dept of Sociology, The Univ. of Texas at San Antonio, TX

Publications

Dudley L. Poston, Jr., Bethany DeSalvo, and Rachel Traut Cortes. 2009. "Neonatal Deaths." In Dennis L. Peck and Clifton D. Bryant (editors), *Encyclopedia of Death and the Human Experience.* Thousand Oaks, CA: Sage Publications, forthcoming.

Dudley L. Poston, Jr. and Rachel Traut Cortes. 2008 "Patterns of Lesbian and Gay Male Cohabiting Partners in the Nonmetropolitan United States in 2000" *Rural Sociology,* forthcoming.

Ginny E. Garcia, Rachel T. Cortes, and Dudley L. Poston, Jr. "Patterns of Sexually Transmitted Infections in China and the United States." Chapter 8 in Joseph Tucker, Dudley L. Poston, Jr., Qiang Ren, Baochang Gu, Xiaoying Zheng, Stephanie Wang, and Chris Russell (editors), *Gender Policy and HIV in China: Catalyzing Policy Chang*e. New York: Springer Publishers, forthcoming.

Rachel T.Cortes and Dudley L. Poston, Jr., 2008. "Immigrants to North America" Pp. 576-580 in William A. Darity (Editor-in-Chief), *International Encyclopedia of the Social Sciences*, Vol. 3, Second Edition. Detroit, MI: Macmillan Reference.

Work Experience

2009 - **Statistician/Demographer at the U.S. Census Bureau;** Suitland, MD
2006-09 **Teaching/Research Assistant to Dudley L. Poston, Jr**.; Texas A&M Univ.
2005 **Research Assistant to Carol Albrecht and Sam Cohn;** Texas A&M Univ.
2004-05 **Research Assistant to Ruben Martinez;** The Univ. of Texas at San Antonio

Additional Experience

Spring 2009 Guest Researcher in the Laboratory for Biodemography at the Max Planck Institute in Rostock, Germany.

.