

ABSTRACT

Title of Document: IDENTIFICATION OF FACTORS THAT RELATE TO GESTATIONAL AGE IN TERM AND PRETERM BABIES USING 2002 NATIONAL BIRTH DATA.

Hoda Tarek Hammad, MPH, 2009

Directed By: Dr. Guangyu Zhang, Department of Epidemiology and Biostatistics

Abstract: Infant mortality and other subsequent handicaps have been found to be correlated with preterm births. The purpose of this study is to investigate which factors relate to gestational age in term and preterm babies using the 2002 Public-Use Natality data file. Using this data, an exploratory data analysis of both the important discrete and continuous variables will be conducted to obtain a general idea of the data set. This will be followed by the use of regression models to determine which explanatory variables best relate to gestational age. The results can be used to establish guidelines for monitoring and treatment plans for expectant mothers who are most susceptible to preterm labor.

IDENTIFICATION OF FACTORS THAT RELATE TO GESTATIONAL AGE IN
TERM AND PRETERM BABIES USING 2002 NATIONAL BIRTH DATA.

By

Hoda Tarek Hammad

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Masters of Public Health
2009

Advisory Committee:
Dr. Guangyu Zhang, Chair
Dr. Mei-Ling Ting Lee
Dr. Jennifer Parker

© Copyright by
Hoda Tarek Hammad
2009

Table of Contents

Table of Contents	ii
List of Tables	iii
Chapter 1: Introduction	1
1.1 Background Information	1
1.2 Importance of Topic	5
1.3 Research Goals	5
Chapter 2: About the 2002 Public-Use Natality File	6
Chapter 3: Methodology	7
3.1 Changes to Dataset	7
3.2 Response Variable	8
3.3 Predictor Variables	8
3.4 Missing Data	8
Chapter 4: Analytic Plan	9
4.1 Analytic Steps	9
4.2 Approach to Model Selection	10
Chapter 5: Preliminary Results of Important Variables	11
5.1 Continuous Explanatory Variables	11
5.2 Categorical Explanatory Variables	12
5.3 Checking Normal Distribution	13
5.4 Bivariate Analysis	13
Chapter 6: Results	15
6.1 Regression Using Stepwise Selection	15
6.2 Interpretation of Results	17
Chapter 7: Discussion	19
7.1 Conclusions	19
7.2 Limitations	19
Appendices	21
Bibliography	41

List of Tables

Table 1: Important Continuous Variables.....	11
Table 2: Important Categorical Variables.....	12
Table 3: Spearman Correlation Coefficients.....	14
Table 4: Parameter Estimates for Regression	16
Table 5: Continuous Variables in 2002 US Birth Data.....	21
Table 6: Categorical Variables in 2002 US Birth Data.....	22
Table 7: Medical Risks Variables in 2002 US Birth Data	26
Table 8: Obstetric Procedures in 2002 US Birth Data.....	29
Table 9: Labor Complications in 2002 US Birth Data	31
Table 10: Newborn Complications in 2002 US birth data.....	34
Table 11: Congenital Anomalies in 2002 US birth data	36
Table 12: Summary Variables	40

Chapter 1: Introduction

1.1 Background Information

Preterm births, classified as those occurring at less than 37 weeks gestation, account for approximately 75% of perinatal deaths (Behrman & Butler, 2007, p. 31). Two-thirds of these deaths occur in preterm infants who were delivered at fewer than 32 gestational weeks (Slattery & Morrison, 2002, p. 1489). The most recent vital statistics report shows that in 2007 preterm births have risen to 12.7% of total births, which is 20% more than the 10.6% reported in 1990 (Hamilton, Martin, & Ventura, 2009, p. 1). That is more than half a million preterm deliveries that places the United States in at least a \$26.2 billion economic burden each year (Behrman & Butler, 2007, p. 31). This is in addition to the emotional impact on the infant's family and the potential complications to neonates.

An average pregnancy lasts about 40 week, and is calculated from the first day of the mother's last normal menstrual cycle (Mattison, Wilson, Coussens, & Gilbert, 2003, p. 14). A preterm birth is when a baby is born prior to the 37th gestation week. They are classified into two distinct categories: a preterm delivery set off for the benefit of the mother and/or fetus by a clinician (indicated preterm birth), or a spontaneous birth which follows unplanned labor or rupture of the membrane (spontaneous preterm birth). About 80% of preterm births are spontaneous preterm labor, while only 20% of them are initiated (Mattison, Wilson, Coussens, & Gilbert, 2003, p. 14). Since there are many distinctive contributing pathways to preterm births, it becomes a challenge to study the mechanisms in order to prevent preterm births

(Mattison, Wilson, Coussens, & Gilbert, 2003, p. 15; Simhan & Caritis, 2007, p. 477).

Preterm birth is a continuing issue both globally and nationally. It is estimated that 9.6% of all global births are preterm, which constitutes about 12.9 million preterm babies around the world (Beck, et al., 2009, pp. 1-2). In reality, this number may be significantly higher for developing countries, but due to not having precise medical records, the actual number is virtually unknown (Beck, et al., 2009, p. 3). In addition, it is possible to assume that most low birthweight babies in poor developing countries may not necessarily be preterm. This may be due to poor nutrition and lack of healthcare, which would cause a term neonate to weigh severely less than expected or what would be considered a healthy weight.

There is a steady increase in the past two decades of preterm births for the United States and other industrialized countries, even though significant improvements in neonatal intensive care has increased survival rates (Mattison, Wilson, Coussens, & Gilbert, 2003, p. 17; Behrman & Butler, 2007, p. 32). This increase is mostly in moderately preterm births, which is between 32 and 36 weeks of gestation (Mattison, Wilson, Coussens, & Gilbert, 2003, p. 15). Within these statistics, there are many health disparities in terms of incidence. Caucasians have a 10.4% incidence rate of preterm births, while African Americans have a higher rate of 17.4%. The factors which explain the variation in incidence rate are mostly unknown (Mattison, Wilson, Coussens, & Gilbert, 2003, pp. 16-17).

On the other hand, there are certain clinical presentations and risk factors that are known to ultimately predispose certain pregnant women to preterm labor. These

may include: multifetal pregnancies, previous preterm delivery, parental socioeconomic status, certain environmental exposures (tobacco and alcohol), medical conditions, biological factors, genetic factors, gene-environmental interactions and more (Simhan & Caritis, 2007, p. 477; Slattery & Morrison, 2002, p. 1489; Behrman & Butler, 2007, p. 37). However, less than half of all preterm births have been shown to be associated with a specific risk factor. This should not deter researchers from the goal and importance of defining risk factors due to the long-term outcomes that preterm delivery poses for the infants (Mattison, Wilson, Coussens, & Gilbert, 2003, p. 20).

There are many long-lasting complications that can arise for a preterm delivery. Some of these problems are diagnosed immediately such as respiratory distress syndrome, brain hemorrhage, jaundice, and infections. Other complications are not detected at birth and may extend over the course of a lifetime such as neurodevelopment disabilities, cerebral palsy, mental retardation, or chronic lung disease. Even among children who do not have any overt neurodevelopment disabilities, they may have subtle problems in behavior and functioning. This can include: poor visual-motor functioning, hyperactivity, poor math skills, and deficient attention skill (Mattison, Wilson, Coussens, & Gilbert, 2003, pp. 18-19). These issues can be assessed through regular visits to the doctor and availability of resources to assist these neonates if needed.

Regular visits to the pediatrician can help evaluate growth and development in a preterm newborn baby to determine normality in their growth rate. This is especially important for these preterm babies since their organs are not fully

developed. Numerous studies have used preterm babies' head circumference as a measure and general reference of normal growth and development. A study conducted at the Rainbow Babies and Children's Hospital showed that subnormal head circumference in very low birthweight children (mostly due to premature births) had many negative consequences when the child goes to school (Peterson, Taylor, Minich, Klein, & Hack, 2006). Measuring weight, length, and head circumference has been regarded as an important diagnostic tool for discovering different conditions in infants, such as detecting intracranial expansive conditions (Zahl & Wester, 2008). Specifically, monitoring brain growth after birth in order to predict the need for developmental support later on in an infant's life is very important.

There are a few interventions available that may reduce the incidence of spontaneous preterm births. The most general one is providing prenatal care so that it is assured that the mother and fetus are healthy and are taking nutritional supplements, etc. There are also more targeted interventions such as drug, alcohol, and tobacco cessation programs, bed rest, risk screening, hydration, iron supplementation, etc. These interventions are specifically provided to combat the known risk factors that are associated with preterm births. Additional research needs to be conducted to better understand the underlying processes and to provide better interventions for those who need it (Mattison, Wilson, Coussens, & Gilbert, 2003, p. 22).

Overall, these statistics show that preterm birth is a serious concern in the United States. Several organizations, researchers, and a variety of federal agencies have taken steps to address preterm birth and to raise awareness of this problem in

order to reduce the incidence rate of premature births (Behrman & Butler, 2007, p. 34). A major foundation, known as the March of Dimes, has made significant efforts to combat this concern by launching its “Prematurity Campaign”. This campaign focuses on funding research, providing affected families support, educating women and their providers on ways to reduce the risk for preterm delivery, and more. More campaigns such as this need to be emphasized, especially in communities with lower socioeconomic statuses, since they may not be receiving proper prenatal care.

1.2 Importance of Topic

Due to the high incidence rate and consequences of preterm births, the importance of preventing it cannot be stressed enough. There should be attempts made in prioritizing research and in informing the public about the problem of preterm birth. Therefore, it is important to correctly determine the important factors that affect preterm delivery in order to establish guidelines for monitoring and treatment plans for expectant mothers who are most susceptible to preterm labor. Instead of looking at a dichotomous analysis of the response variable, preterm births vs. normal births, as most other studies do, this study will look at weeks of gestational age.

1.3 Research Goals

This project will look to identify which factors that are included in the birth certificates have an important effect on estimating gestational age. There will be special focus on certain variables that are known to be associated with preterm births, such as race, mother’s age, tobacco use, and alcohol use, etc.

Chapter 2: About the 2002 Public-Use Natality File

The National Center for Health Statistics (NCHS) continuously collects and makes public information from birth certificates for all births occurring in the United States. These certificates include a wide assortment of demographic and health characteristics that are registered in the health departments of all the states and territories, and provided through the Vital Statistics Cooperative Program by NCHS (Ventura, 2004). This information, as well as preliminary and final data reports since the 1968 data year, are published annually by NCHS and are available to the public (Ventura, 2004; Martin, et al., 2009).

The 2002 public-use natality data consists of 4,027,376 live births occurring within the United States to residents and non-residents (National Center for Health Statistics, 2002). There are 213 total variables, including the recoded ones. They are broken up into 10 different categories:

1. General Items (year of birth, record type, resident status)
2. Occurrence (state, county, population size of city, region, etc.)
3. Residence (state, county, population size of city, region, etc.)
4. Prenatal Care (number of times, adequacy, month started)
5. Child (sex, birthweight, gestation, month and day of birth, etc.)
6. Mother (age, race, education, marital status, place of birth, etc.)
7. Pregnancy History (birth order, terminations, etc.)
8. Father (age, race, etc.)
9. Other Items (attendant at birth, place of delivery, etc.)

10. Medical and Health Data (method of delivery, risk factors, obstetric procedures, labor complications, abnormal conditions, and congenital anomalies of the newborn.)

Although there are more up-to-date natality datasets available for use on the NCHS website, the 2002 dataset is used due to the implementation of the 2003 revision of the United States Standard Certificate of Live Birth for the later data years (Ventura, 2004). The changes to the birth certificate have been implemented at different times by each State, so there are variables that differ across different states.

Chapter 3: Methodology

SAS software Version 9.1 will be used for all the statistical analyses.

3.1 Changes to Dataset

In order to avoid dealing with the intricacies of misspecified gestational ages, the observations whose observed gestation was different from their clinical estimate by more than 2 weeks were deleted from the dataset. Specifically based on the 2002 dataset, a report from the Centers for Disease Control and Prevention (CDC) showed that preterm births were more likely to have disagreement than term births between the estimates based on the last menstrual period and the clinical approximation (Qin, Hsia, & Berg, 2008). Due to this, California was excluded because information on the clinical estimate is not collected on their birth certificate.

Based on these alterations, the final dataset used in the analysis had a total of 3,029,853 observations. This is about 75% of the complete dataset.

3.2 Response Variable

The response variable is the observed gestational age, which is coded as DGESTAT in the dataset. Observations that are less than 37 weeks are considered preterm births and observations between 37 and 42 are considered term births.

3.3 Predictor Variables

The predictor variables include all the non-recoded variables from all 10 categories mentioned above. Categorical variables that had more than 10 categories, and were not recoded, were excluded from the analysis. This means that there are 163 unique variables in the dataset.

3.4 Missing Data

There following variables had more than 10% missing data points, and will be deleted from the final dataset. They are classified as “Unknown or Not Stated” in the dataset:

1. FMAPS: Five Minute Apgar Score (10.71%)
2. DMAGERPT: Reported age of Mother (91.11%)
3. FRACE: Father’s race (14.08%)
4. ORFATH: Hispanic Origin of Father (14.10%)
5. DFAGE: Age of Father (13.13%)
6. DFAGERPT: Reported age of Father (92.24%)

Chapter 4: Analytic Plan

4.1 Analytic Steps

The following analyses are conducted to determine the important predictors:

1. Descriptive Data Analysis
 - a. Descriptive statistics of both the discrete and continuous variables.
 - b. PROC FREQ/PROC MEANS for all variables to determine outliers, and check for missingness.
 - c. PROC UNIVARIATE for continuous variables to determine distribution.
2. Correlation Analysis
 - a. Bivariate Analysis using PROC CORR.
 - b. Determine crude association using Chi-Square for categorical data, and t-test for continuous data.
3. Model Selection
 - a. Check assumptions and fit a regression model (REG) using stepwise model selection.
 - b. For the categorical variables included, fit a generalized linear model (GENMOD) using the CLASS statement to determine variable significance.

4.2 Approach to Model Selection

The response variable (gestational age) in this dataset is ordinal with more than two levels and possesses intrinsic ordering. Therefore, the response variable used in this analysis is the log of gestational age (LDGESTAT). The parameter estimates will be assessed using the stepwise model selection in the regression procedure (REG). Since there are several categorical variables in the dataset, further analysis will be done using the CLASS statement in the generalized linear models procedure (GENMOD) for each categorical variable. Based on this procedure, the significant categorical variables will be included into a regression model with all the continuous variables.

Several of the categories were summed into one category and added into the model for analysis. These categories include: the total number of medical risk factors, the total number of obstetric procedures, the total number of the complications of labor and/or delivery, the total number of abnormal conditions of the newborn, and the total number of congenital anomalies. The stepwise model selection in the regression procedure, as well as the generalized linear model procedure will be used to calculate these parameter estimates.

Chapter 5: Preliminary Results of Important Variables

Complete tables of all explanatory variables (continuous and categorical) can be found in the Appendix.

5.1 Continuous Explanatory Variables

The mean birth weight is 3303 grams with a standard deviation of 602. The smallest birth weight is 227 grams, and the largest is 8100 grams (Table 1). The mean of the mother's age is 27 years with a standard deviation of 6. The youngest mother's age is 10 years old and the oldest is 54 years old. The mean of the father's age is 31 with a standard deviation of 7. The youngest age is 10 years and the oldest is 90 years old. The total number of prenatal visits has a mean of 11 and a standard deviation of 4. The least amount of visits is zero visits and the most is 49 visits. Lastly, the month for which prenatal care began has a mean of 2 months and standard deviation of 1. The earliest month which prenatal care began is at 0 months and the latest is at 9 months.

Table 1: Important Continuous Variables

<i>Var Name</i>	<i>Var Definition</i>	<i>Mean</i>	<i>STD</i>	<i>Min</i>	<i>Max</i>	<i>N</i>	<i>NMISS</i>	<i>% NMISS</i>
DBIRWT	Birth Weight - Detail in Grams	3303.55	601.69	227	8100	3,027,812	2,041	0.07%
DMAGE	Age of Mother	27.39	6.13	10	54	3,029,853	0	0.00%
DFAGE	Age of Father	30.5	6.79	10	90	30.5	408,282	15.57%

<i>Var Name</i>	<i>Var Definition</i>	<i>Mean</i>	<i>STD</i>	<i>Min</i>	<i>Max</i>	<i>N</i>	<i>NMISS</i>	<i>% NMISS</i>
NPREVIS	Total Number of Prenatal Visits	11.47	3.94	0	49	2,946,486	83,367	2.83%
MONPRE	Detail Month of Pregnancy Prenatal Care Began	2.42	1.42	0	9	2,961,665	68,188	2.30%

5.2 Categorical Explanatory Variables

There are 1,551,155 male infants and 1,478,698 female infants born in 2002 (Table 2). There are 2,403,265 white mothers, 463,429 black mothers, and 163,159 other races. For the father's race, there are 1,689,068 white fathers, 328,363 black fathers, and 576,323 other races. Approximately 14% of the father's data on race is missing.

Table 2: Important Categorical Variables

<i>Var Name</i>	<i>Var Definition</i>	<i>Categories</i>	<i>Category Definitions</i>	<i>Frequency</i>	<i>NMISS</i>	<i>% NMISS</i>
CSEX	Sex	1	Male	1,551,155	-	-
		2	Female	1,478,698		
MRACE	Race of Mother	1	White	2,403,265	-	-
		2	Black	463,429		
		3	Others	163,159		
FRACE	Race of Father	1	White	1,689,068	436,099	14.39%
		2	Black	328,363		
		3	Others	576,323		

5.3 Checking Normal Distribution

To check the distribution of the continuous variables and if they are normally distributed, the UNIVARIATE procedure was used. The Kolmogorov-Smirnov was used to test for normality. If the p-value is less than 0.05, then the data might not be normally distributed. All the continuous variables in this dataset had a Kolmogorov-Smirnov p-value of less than 0.05, which means that the data may not be normally distributed.

5.4 Bivariate Analysis

To check the correlation between the log of gestational age and all other continuous variables, the CORR procedure was used. Specifically, the Spearman correlation was used in this case, since the continuous variables were found to not be normal. The null hypothesis is that there is no linear relationship between the log of gestational age and the continuous variable. If the p-value is less than 0.05, then the null hypothesis is rejected, which means that there is a linear relationship between the log of gestational age and the continuous variables. The correlation between the log of gestational age and all the continuous variables are all significant, which means that we reject the null hypothesis that there is a linear relationship between the log of gestational age and the continuous variables (Table 3).

Table 3: Spearman Correlation Coefficients

<i>Variable</i>	<i>Variable Definition</i>	<i>Spearman Correlation Coefficients</i>	<i>Prob > r under H0: Rho=0</i>
NPREVIS	Total number of prenatal visits	0.10437	<0.0001
CSEX	Sex	0.02699	<0.0001
DBIRWT	Birth weight- detail in grams	0.42606	<0.0001
DMAR	Marital status	0.00881	<0.0001
LMPDAY	Day last normal menses began	-0.02983	<0.0001
DMAGE	Mother's age	-0.05735	<0.0001
NLBNL	Number of live births, now living	-0.08752	<0.0001
NLBND	Number of live births, now dead	-0.03505	<0.0001
NOTERM	Number of other terminations	-0.04743	<0.0001
DTOTORD	Detail total birth order	-0.09533	<0.0001
DLIVORD	Detail live birth order	-0.09049	<0.0001
METHOD	Method of Delivery	0.08808	<0.0001
MEDRISKSUM	Total number of medical risks	-0.13015	<0.0001
OBSTETRCSUM	Total number of newborn complications	0.03491	<0.0001
LABORSUM	Total number of the complications of labor and/or delivery	-0.01576	<0.0001
NEWBORNSUM	Total number of abnormal conditions of the newborn	-0.10817	<0.0001
CONGENITSUM	Total number of congenital anomalies	-0.03201	<0.0001

There are some variables in the dataset which are hypothesized to not be independent of each other. These include the region of occurrence (REGNOCC) versus the region of residence (REGNRES), the division of occurrence (DIVSTOCC) versus the division of residence (DIVRES), and the population size of county occurrence (CNTOCPOP) versus the population size of county residence (CNTRSPOP). To test if these variables are indeed independent of each other, a Chi-

Square statistic will be calculated using the FREQ procedure. The null hypothesis is that these region of occurrence and region of residence variables are independent of each other. In this case, the p-value for the Chi-Square statistic is <0.0001 , which means that it rejects the null hypothesis. Therefore, it can be concluded that region of occurrence and region of residence are not independent of each other. The p-value for the Chi-Square statistics is also <0.0001 for the division of occurrence and division of residence, which means that these two variables are also not independent of each other. Lastly, the p-value for the Chi-square statistics is also <0.0001 for the population size of county occurrence and population size of county residence, which means they are also not independent of each other. Since these variables are not independent of each other, only the residence variables will be used in the analysis.

Chapter 6: Results

6.1 Regression Using Stepwise Selection

The following categorical variables were significant when analyzed using generalized linear models using the CLASS statement:

- RESTATUS (Resident Status)
- REGNRES (Region of Residence)
- CNTRSPOP (Population of County of Residence)
- CITRSPOP (Population of City of Residency)
- DPLURAL (Plurality)
- MRACE (Mother's Race)
- DMEDUC (Mother's Education)
- ADEQUACY (Adequacy of Prenatal Care)

The final regression model using stepwise model selection is the following:

Table 4: Parameter Estimates for Regression

<i>Parameter</i>	<i>Definition</i>	<i>Parameter Estimates</i>	<i>p-value</i>
Intercept		3.43446	<.0001
RESTATUS2	Resident Status	-0.00324	<.0001
RESTATUS3	Resident Status	-0.00605	<.0001
REGNRES1	Region of Residency	0.00527	<.0001
REGNRES2	Region of Residency	0.00127	<.0001
REGNRES4	Region of Residency	0.01511	<.0001
MRACE2	Mother's Race	0.00103	<.0001
MRACE3	Mother's Race	0.00651	<.0001
DMEDUC2	Mother's Education	-0.00089410	0.0114
DMEDUC3	Mother's Education	-0.00467	<.0001
DMEDUC4	Mother's Education	-0.00371	<.0001
ADEQUACY2	Adequacy Of Care Recode (Kessner Index)	0.00633	<.0001
ADEQUACY3	Adequacy Of Care Recode (Kessner Index)	0.01431	<.0001
CNTRSPOP1	Population Size of County of Resid	-0.00368	<.0001
CNTRSPOP2	Population Size of County of Resid	-0.00474	<.0001
CNTRSPOP3	Population Size of County of Resid	-0.00547	0.0181
CNTRSPOP9	Population Size of County of Resid	-0.00516	<.0001
CITRSPOP1	Population Size of City of Residence	-0.00356	<.0001
CITRSPOP2	Population Size of City of Residence	-0.00294	<.0001
CITRSPOP3	Population Size of City of Residence	0.00050223	<.0001
CITRSPOP9	Population Size of City of Residence	-0.00094286	<.0001
METRORES	Population Size of City of Residence	-0.00032743	0.0128
DPLURAL2	Plurality	-0.01303	<.0001
DPLURAL3	Plurality	-0.04799	<.0001
NPREVIS	Total Number of Prenatal Visits	0.00179	<.0001
CSEX	Sex of Child	0.00977	<.0001
DBIRWT	Birth Weight - Detail in Grams	0.00006322	<.0001
DMAR	Marital Status	0.00341	<.0001
LMPDAY	Day Last Normal Menses Began	-0.00020428	<.0001
DMAGE	Mother's Age	-0.00041502	<.0001
NLBND	Number of Live Births, Now Dead	-0.00470	<.0001
NOTERM	Number of Other Terminations	0.00046088	
DTOTORD	Detail Total Birth Order	-0.00220	<.0001
METHOD	Method of Delivery	0.00669	<.0001

<i>Parameter</i>	<i>Definition</i>	<i>Parameter Estimates</i>	<i>p-value</i>
MEDRISKSUM	Total number of medical risks	-0.00648	<.0001
NEWBORNSUM	Total number of newborn complications	0.00137	<0.001
LABORSUM	Total number of complications of labor and/or delivery	0.00122	<0.001
NEWBORNSUM	Total number of abnormal conditions of the newborn	-0.02051	<0.0001
CONGENITSUM	Total number of congenital anomalies	-0.00327	<.0001

6.2 Interpretation of Results

Location

From this analysis, it can be seen that the region of residency has an effect on gestation. This means that whether one lives in the Northeast, Midwest, South, or West may determine at what gestation a baby is born. In this case, residing in Midwest (0.00127) and West (0.01511) has a positive effect on the log of gestation, in comparison to residing in the Northeast region, keeping all other variables constant. This may be due to differences in socioeconomic statuses across the States in each particular region. Also, the status of residency at the time of birth, which is whether the neonate is born in their resident state, interstate, intrastate, or overseas, was found to be significant. Residing interstate (-0.00324) or intrastate (-0.00605) also has a negative effect on the log of gestation, in comparison to the neonate being born in their resident state, keeping all other variables equal.

Prenatal Care

The month when pregnancy prenatal care began was not deemed to be significant in the regression model. When looking at other prenatal care variable, adequacy of prenatal care (0.00633 and 0.01431) and total number of prenatal visits (0.00179), they both have a positive effect on the dependent variable.

Mother's Characteristics

A one unit increase in mother's age decreases the log of gestational age by 0.04%, keeping all other variables constant. Also, in comparison to a mother having five years or more college education, having four years of college or less decreases the log of gestational age by 0.00371. Looking at a mother's marital status, in contrast to being a single mother, a mother being married increases the log of gestational age by 0.00341.

Child's Characteristics

Carrying multiple babies at one time decreases the log of gestational age in comparison to carrying only one neonate. With carrying twins, it decreases the log of gestational age by 0.01303, keeping all other variables constant. With carrying triplets or more, it decreases the log of gestational age by 0.04799, keeping all other variables constant. This may explain why multifetal births are usually born prematurely. Also, in comparison to having a male child, having a female child increases gestational age by 0.00977, keeping all other variables constant. Lastly, for every unit increase in the child's birthweight, the log of gestational age changes by 0.006322%, keeping all other variables constant.

Medical and Health Data Summary Variables

The regression model selection showed that most of the medical and health risks show a decrease of the log of gestational age. For example, for every one unit increase in the total number of medical risks, there is a 0.648% decrease in the log of gestational age, keeping all other variables constant. Also, for every one unit increase in the number of abnormal conditions of the newborn, there is a 2.05% decrease in the log of gestational age, keeping all other variables constant. Lastly, for every one unit increase in the total number of congenital anomalies, there is a 0.327% decrease in the log of gestational age, keeping all other variables constant.

Chapter 7: Discussion

7.1 Conclusions

The regression procedure method of model selection calculated 38 total important variables in predicting the log of gestational age. They include the place of residency, the adequacy of prenatal care and the number of prenatal care visits, mother's age, race, marital status, and level of education, plurality, as well as an array of medical and health risks that negatively affect gestational age.

7.2 Limitations

There are some limitations to the study that may have affected the results. Restricting the data to births where the observed gestational age is equal to the clinical estimate of gestation to avoid misspecified observations may have introduced bias into the analysis. If those neonates who were excluded from the dataset due to

not getting proper prenatal care, and thus their observed gestational age is more than 2 weeks from their clinical estimate, then important information may be missing from this subset of the population. Also, the fact that California neonates were not included in this analysis due to clinical estimates not being recorded on their birth certificates may have severely altered the results of this study. Important factors such as distribution of race, marital status, level of education, prenatal care, etc. might differ between these groups of infants. Thus, we may be missing critical overall relationships in an effort to avoid misspecified observations. Therefore, further analysis needs to be conducted in order to determine if there is a difference present.

Appendices

Table 5: Continuous Variables in 2002 US Birth Data

<i>Var Name</i>	<i>Var Definition</i>	<i>Mean</i>	<i>STD</i>	<i>Min</i>	<i>Max</i>	<i>N</i>	<i>NMISS</i>	<i>% NMISS</i>
MONPRE	Detail Month of Pregnancy Prenatal Care Began	2.42	1.42	0	9	2,961,665	68,188	2.30%
NPREVIS	Total Number of Prenatal Visits	11.47	3.94	0	49	2,946,486	83,367	2.83%
DBIRWT	Birth Weight - Detail in Grams	3303.55	601.69	227	8100	3,027,812	2,041	0.07%
FMAPS	Five Minute Apgar Score	8.91	0.73	0	10	2,666,252	363,601	13.64%
DMAGE	Age of Mother	27.39	6.13	10	54	3,029,853	-	0.00%
NLBNL	Number of Live Births, Now Living	1.03	1.2	0	30	3,023,021	6,832	0.23%
NLBND	Number of Live Births, Now Dead	0.02	0.17	0	12	3,021,218	8,635	0.29%
NOTERM	Number of Other Terminations	0.39	0.82	0	29	3,020,088	9,765	0.32%
DTOTORD	Detail Total Birth Order	2.43	1.57	1	31	3,018,235	11,618	0.38%
DLIVORD	Detail Live Birth Order	2.05	1.23	1	22	3,020,802	9,051	0.30%
DFAGE	Age of Father	30.5	6.79	10	90	2,621,571	408,282	15.57%
DFAGERPT	Reported Age of Father					2,796,485	2,796,485	100.00%
CIGAR	Average Number of Cigarettes Per Day	0.99	3.69	0	98	2,939,858	89,995	3.06%
DRINK	Average Number of Drinks Per Week	0.03	0.53	0	98	2,968,774	61,079	2.06%
WTGAIN	Weight Gain	30.85	13.65	0	98	2,813,720	216,133	7.68%

Table 6: Categorical Variables in 2002 US Birth Data

<i>Var Name</i>	<i>Var Definition</i>	<i>Categories</i>	<i>Category Definitions</i>	<i>Frequency</i>	<i>NMISS</i>	<i>% NMISS</i>		
RESTATUS	Resident Status	1	Resident	2,192,970				
		2	Intrastate Resident	756,921				
		3	Interstate Resident	75,416				
		4	Foreign Resident	4,546				
PLDEL	Place or Facility of Birth	1	Hospital	2,998,608				
		2	Freestanding Birthing Center	10,545				
		3	Clinic or Doctor's Office	351				
		4	A Residence	18,636				
		5	Other	1,609				
		9	MISSING	104	104	0.00%		
		BIRATTND	Attendant at Birth	1	Doctor of Medicine (M.D.)	2,625,812		
				2	Doctor of Osteopathy (D.O.)	140,978		
				3	Certified Nurse Midwife (C.N.M.)	230,774		
4	Other Midwife			17,339				
5	Other			14,272				
9	MISSING			678	678	0.02%		
REGNOCC	Region of Occurrence			1	Northeast	605,297		
				2	Midwest	750,823		
				3	South	1,251,633		
		4	West	422,100				
DIVSTOCC	Division of Occurrence	9 Categories						
STSUBOCC	State Subcode of Occurrence	9 Categories						
STATENAT	State of Occurrence	51 Categories						
CNTYNAT	County of Occurrence	247 Categories						
STOCCFIP	State of Occurrence (FIPS)	56 Categories						
CNTOCFIP	County of Occurrence (FIPS)	120 Categories						

Table 6: Categorical Variables Continued

<i>Var Name</i>	<i>Var Definition</i>	<i>Categories</i>	<i>Category Definitions</i>	<i>Frequency</i>	<i>NMISS</i>	<i>% NMISS</i>
CNTOCPOP	Population Size of County of Occurrence	0	County of 1,000,000 or more	631,823		
		1	County of 500,000 to 1,000,000	706,957		
		2	County of 250,000 to 500,000	517,211		
		3	County of 100,000 to 250,000	535,698		
		9	County of Less Than 100,000	638,164		
ADEQUACY	Adequacy Of Care Recode (Kessner Index)	1	Adequate	2,208,189		
		2	Intermediate	533,050		
		3	Inadequate	151,726		
		Missing	Unknown	136,888	136,888	4.52%
CSEX	Sex	1	Male	1,551,155		
		2	Female	1,478,698		
DPLURAL	Plurality	1	Single	2,929,382		
		2	Twin	94,875		
		3	Triplet	5,205		
		4	Quadruplet	334		
		5	Quintuplet or higher	57		
BIRMON	Month of Birth	1	January	246,082		
		2	February	228,079		
		3	March	248,763		
		4	April	243,223		
		5	May	255,277		
		6	June	246,304		
		7	July	270,225		
		8	August	272,126		
		9	September	265,437		
		10	October	262,359		
		11	November	240,838		
		12	December	251,140		
WEEKDAY	Day of Week Child Born	1	Monday	289,931		
		2	Tuesday	451,276		
		3	Wednesday	504,751		
		4	Thursday	484,408		
		5	Friday	486,089		
		6	Saturday	482,579		
		7	Sunday	330,819		

Table 6: Categorical Variables Continued

<i>Var Name</i>	<i>Var Definition</i>	<i>Categories</i>	<i>Category Definitions</i>	<i>Frequency</i>	<i>NMISS</i>	<i>% NMISS</i>
MRACE	Race of Mother	14 Categories				
DMEDUC	Education of Mother	17 Categories				
		Missing	MISSING	38,892	38,892	1.28%
DMAR	Marital Status of Mother	1	2037568			
		2	992285			
MPLBIR	Place of Birth of Mother	59 Categories				
		Missing	MISSING	7,183	7,183	0.24%
MPLBIRR	Place of Birth of Mother Recode	1	Native Born	2,413,021		
		2	Foreign Born	609,649		
		Missing	MISSING	7,183	7,183	0.24%
ORMOTH	Hispanic Origin of Mother	6 Categories				
		Missing	MISSING	23,973	23,973	0.79%
FRACE	Race of Father	14 Categories				
		Missing	MISSING	436,099	436,099	14.39%
ORFATH	Hispanic Origin of Father	6 Categories				
		Missing	MISSING	436,429	436,429	14.40%
VAGINAL	Vaginal	1	The method was used	2,227,370		
		2	The method was not used	781,319		
		8	Method not on certificate	4,546		
		Missing	MISSING	16,618	16,618	0.55%
VBAC	Vaginal birth after previous C-section	1	The method was used	47,271		
		2	The method was not used	2,961,418		
		8	Method not on certificate	4,546		
		Missing	MISSING	16,618	16,618	0.55%
PRIMAC	Primary C-section	1	The method was used	478,475		
		2	The method was not used	2,530,214		
		8	Method not on certificate	4,546		
		Missing	MISSING	16,618	16,618	0.55%

Table 6: Categorical Variables Continued

<i>Var Name</i>	<i>Var Definition</i>	<i>Categories</i>	<i>Category Definitions</i>	<i>Frequency</i>	<i>NMISS</i>	<i>% NMISS</i>
REPEAC	Repeat C-section	1	The method was used	302,844		
		2	The method was not used	2,705,845		
		8	Method not on certificate	4,546		
		Missing	MISSING	16,618	16,618	0.55%
FORCEP	Forceps	1	The method was used	49,340		
		2	The method was not used	2,959,349		
		8	Method not on certificate	4,546		
		Missing	MISSING	16,618	16,618	0.55%
VACUUM	Vacuum	1	The method was used	129,541		
		2	The method was not used	2,879,148		
		8	Method not on certificate	4,546		
		Missing	MISSING	16,618	16,618	0.55%
TOBACCO	Tobacco Use During Pregnancy	1	Yes	324,128		
		2	No	2,651,005		
		9	MISSING	54,720	54,720	1.81%
ALCOHOL	Alcohol Use During Pregnancy	1	Yes	23,907		
		2	No	2,945,834		
		9	MISSING	60,112	60,112	1.98%

Table 7: Medical Risks Variables in 2002 US Birth Data

<i>Var Name</i>	<i>Var Definition</i>	<i>Categories</i>	<i>Category Definitions</i>	<i>Frequency</i>	<i>NMIS S</i>	<i>% NMISS</i>
MEDRISK	Medical Risk Factors	1	Factor reported			
		2	Factor not reported			
		8	Factor not on certificate			
		9	Factor not classifiable			
ANEMIA	Anemia (Hct.<30/Hgb.<10)	1	Factor reported	83,643		
		2	Factor not reported	2,917,575		
		8	Factor not on certificate	4,546		
		9	Factor not classifiable	24,089	24,089	0.80%
CARDIAC	Cardiac disease	1	Factor reported	17,531		
		2	Factor not reported	2,983,687		
		8	Factor not on certificate	4,546		
		9	Factor not classifiable	24,089	24,089	0.80%
LUNG	Acute or chronic lung disease	1	Factor reported	41,544		
		2	Factor not reported	2,959,674		
		8	Factor not on certificate	4,546		
		9	Factor not classifiable	24,089	24,089	0.80%
DIABETES	Diabetes	1	Factor reported	101,578		
		2	Factor not reported	2,899,640		
		8	Factor not on certificate	4,546		
		9	Factor not classifiable	24,089	24,089	0.80%
HERPES	Genital herpes	1	Factor reported	27,316		
		2	Factor not reported	2,656,322		
		8	Factor not on certificate	321,762		
		9	Factor not classifiable	24,453	24,453	0.81%

Table 7: Medical Risks Continued

<i>Var Name</i>	<i>Var Definition</i>	<i>Categories</i>	<i>Category Definitions</i>	<i>Frequency</i>	<i>NMISS</i>	<i>% NMISS</i>
HYDRA	Hydramnios/Oligohydramnios	1	Factor reported	45,289		
		2	Factor not reported	2,955,929		
		8	Factor not on certificate	4,546		
		9	Factor not classifiable	24,089	24,089	0.80%
HEMO	Hemoglobinopathy	1	Factor reported	2,435		
		2	Factor not reported	2,998,783		
		8	Factor not on certificate	4,546		
		9	Factor not classifiable	24,089	24,089	0.80%
CHYPER	Hypertension, chronic	1	Factor reported	27,042		
		2	Factor not reported	2,974,176		
		8	Factor not on certificate	4,546		
		9	Factor not classifiable	24,089	24,089	0.80%
PHYPER	Hypertension, pregnancy-associated	1	Factor reported	119,364		
		2	Factor not reported	2,881,854		
		8	Factor not on certificate	4,546		
		9	Factor not classifiable	24,089	24,089	0.80%
ECLAMP	Eclampsia	1	Factor reported	10,593		
		2	Factor not reported	2,990,625		
		8	Factor not on certificate	4,546		
		9	Factor not classifiable	24,089	24,089	0.80%
INCERVIX	Incompetent cervix	1	Factor reported	9,535		
		2	Factor not reported	2,991,683		
		8	Factor not on certificate	4,546		
		9	Factor not classifiable	24,089	24,089	0.80%

Table 7: Medical Risks Continued

<i>Var Name</i>	<i>Var Definition</i>	<i>Categories</i>	<i>Category Definitions</i>	<i>Frequency</i>	<i>NMISS</i>	<i>% NMISS</i>
PRE4000	Previous infant 4000+ grams	1	Factor reported	33,743		
		2	Factor not reported	2,967,475		
		8	Factor not on certificate	4,546		
		9	Factor not classifiable	24,089	24,089	0.80%
PRETERM	Previous preterm or small-for-gestational-age infant	1	Factor reported	36,424		
		2	Factor not reported	2,964,794		
		8	Factor not on certificate	4,546		
		9	Factor not classifiable	24,089	24,089	0.80%
RENAL	Renal disease	1	Factor reported	10,207		
		2	Factor not reported	2,991,011		
		8	Factor not on certificate	4,546		
		9	Factor not classifiable	24,089	24,089	0.80%
RH	Rh sensitization	1	Factor reported	21,871		
		2	Factor not reported	2,942,443		
		8	Factor not on certificate	39,430		
		9	Factor not classifiable	26,109	26,109	0.86%
UTERINE	Uterine bleeding	1	Factor reported	16,841		
		2	Factor not reported	2,666,797		
		8	Factor not on certificate	321,762		
		9	Factor not classifiable	24,453	24,453	0.81%
OTHERMR	Other Medical Risk Factors	1	Factor reported	628,229		
		2	Factor not reported	2,372,989		
		8	Factor not on certificate	4,546		
		9	Factor not classifiable	24,089	24,089	0.80%

Table 8: Obstetric Procedures in 2002 US Birth Data

<i>Var Name</i>	<i>Var Definition</i>	<i>Categories</i>	<i>Category Definitions</i>	<i>Frequency</i>	<i>NMISS</i>	<i>% NMISS</i>
OBSTETRC	Obstetric Procedures	1	Procedure reported			
		2	Procedure not reported			
		8	Procedure not on certificate			
		9	Procedure not classifiable			
AMNIO	Amniocentesis	1	Procedure reported	62,754		
		2	Procedure not reported	2,950,037		
		8	Procedure not on certificate	4,546		
		9	Procedure not classifiable	12,516	12,516	0.41%
MONITOR	Electronic fetal monitoring	1	Procedure reported	2,627,753		
		2	Procedure not reported	385,038		
		8	Procedure not on certificate	4,546		
		9	Procedure not classifiable	12,516	12,516	0.41%
INDUCT	Induction of labor	1	Procedure reported	675,855		
		2	Procedure not reported	2,336,936		
		8	Procedure not on certificate	4,546		
		9	Procedure not classifiable	12,516	12,516	0.41%
STIMULA	Stimulation of labor	1	Procedure reported	555,848		
		2	Procedure not reported	2,456,943		
		8	Procedure not on certificate	4,546		
		9	Procedure not classifiable	12,516	12,516	0.41%
TOCOL	Tocolysis	1	Procedure reported	64,917		
		2	Procedure not reported	2,947,874		
		8	Procedure not on certificate	4,546		
		9	Procedure not classifiable	12,516	12,516	0.41%

Table 8: Obstetric Procedures Continued

<i>Var Name</i>	<i>Var Definition</i>	<i>Categories</i>	<i>Category Definitions</i>	<i>Frequency</i>	<i>NMISS</i>	<i>% NMISS</i>
ULTRAS	Ultrasound	1	Procedure reported	2,088,917		
		2	Procedure not reported	923,874		
		8	Procedure not on certificate	4,546		
		9	Procedure not classifiable	12,516	12,516	0.41%
OTHEROB	Other Obstetric Procedures	1	Procedure reported	234,247		
		2	Procedure not reported	2,778,544		
		8	Procedure not on certificate	4,546		
		9	Procedure not classifiable	12,516	12,516	0.41%

Table 9: Labor Complications in 2002 US Birth Data

<i>Var Name</i>	<i>Var Definition</i>	<i>Categories</i>	<i>Category Definitions</i>	<i>Frequency</i>	<i>NMISS</i>	<i>% NMISS</i>
LABOR	Complication of Labor and/or Delivery	1	Procedure reported			
		2	Procedure not reported			
		8	Procedure not on certificate			
		9	Procedure not classifiable			
FEBRILE	Febrile (>100 degrees F. or 38 degrees C.)	1	Procedure reported	47,836		
		2	Procedure not reported	2,961,198		
		8	Procedure not on certificate	4,546		
		9	Procedure not classifiable	16,273	16,273	0.54%
MECONIUM	Meconium, moderate/heavy	1	Procedure reported	156,056		
		2	Procedure not reported	2,852,978		
		8	Procedure not on certificate	4,546		
		9	Procedure not classifiable	16,273	16,273	0.54%
RUPTURE	Premature rupture of membrane (>12 hours)	1	Procedure reported	70,550		
		2	Procedure not reported	2,938,484		
		8	Procedure not on certificate	4,546		
		9	Procedure not classifiable	16,273	16,273	0.54%
ABRUPTIO	Abruptio placenta	1	Procedure reported	16,745		
		2	Procedure not reported	2,992,289		
		8	Procedure not on certificate	4,546		
		9	Procedure not classifiable	16,273	16,273	0.54%
PREPLACE	Placenta previa	1	Procedure reported	10,344		
		2	Procedure not reported	2,998,690		
		8	Procedure not on certificate	4,546		
		9	Procedure not classifiable	16,273	16,273	0.54%

Table 9. Labor Complications Continued

<i>Var Name</i>	<i>Var Definition</i>	<i>Categories</i>	<i>Category Definitions</i>	<i>Frequency</i>	<i>NMISS</i>	<i>% NMISS</i>
EXCEBLD	Other excessive bleeding	1	Procedure reported	10,344		
		2	Procedure not reported	2,998,690		
		8	Procedure not on certificate	4,546		
		9	Procedure not classifiable	16,273	16,273	0.54%
SEIZURE	Seizures during labor	1	Procedure reported	958		
		2	Procedure not reported	3,008,076		
		8	Procedure not on certificate	4,546		
		9	Procedure not classifiable	16,273	16,273	0.54%
PRECIP	Precipitous labor (<3 hours)	1	Procedure reported	57,987		
		2	Procedure not reported	2,951,047		
		8	Procedure not on certificate	4,546		
		9	Procedure not classifiable	16,273	16,273	0.54%
PROLONG	Prolonged labor (>20 hours)	1	Procedure reported	21,713		
		2	Procedure not reported	2,987,321		
		8	Procedure not on certificate	4,546		
		9	Procedure not classifiable	16,273	16,273	0.54%
DYSFUNC	Dysfunctional labor	1	Procedure reported	91,396		
		2	Procedure not reported	2,917,638		
		8	Procedure not on certificate	4,546		
		9	Procedure not classifiable	16,273	16,273	0.54%
BREECH	Breech/Malpresentation	1	Procedure reported	119,983		
		2	Procedure not reported	2,889,051		
		8	Procedure not on certificate	4,546		
		9	Procedure not classifiable	16,273	16,273	0.54%

Table 9: Labor Complications Continued

<i>Var Name</i>	<i>Var Definition</i>	<i>Categories</i>	<i>Category Definitions</i>	<i>Frequency</i>	<i>NMISS</i>	<i>% NMISS</i>
CEPHALO	Cephalopelvic disproportion	1	Procedure reported	48,449		
		2	Procedure not reported	2,960,585		
		8	Procedure not on certificate	4,546		
		9	Procedure not classifiable	16,273	16,273	0.54%
CORD	Cord prolapse	1	Procedure reported	5,595		
		2	Procedure not reported	3,003,439		
		8	Procedure not on certificate	4,546		
		9	Procedure not classifiable	16,273	16,273	0.54%
ANESTHE	Anesthetic complications	1	Procedure reported	1,970		
		2	Procedure not reported	2,687,535		
		8	Procedure not on certificate	321,762		
		9	Procedure not classifiable	18,586	18,586	0.61%
DISTRESS	Fetal distress	1	Procedure reported	110,604		
		2	Procedure not reported	2,578,901		
		8	Procedure not on certificate	321,762		
		9	Procedure not classifiable	18,586	18,586	0.61%
OTHERLB	Other Complication of Labor and/or Delivery	1	Procedure reported	499,940		
		2	Procedure not reported	2,509,094		
		8	Procedure not on certificate	4,546		
		9	Procedure not classifiable	16,273	16,273	0.54%

Table 10: Newborn Complications in 2002 US birth data

<i>Var Name</i>	<i>Var Definition</i>	<i>Categories</i>	<i>Category Definitions</i>	<i>Frequency</i>	<i>NMISS</i>	<i>% NMISS</i>
NEWBORN	Abnormal Conditions of the Newborn	1	Condition reported			
		2	Condition not reported			
		8	Condition not on certificate			
		9	Condition not classifiable			
NANEMIA	Anemia (Hct.<39/Hg b.<13)	1	Condition reported	3,293		
		2	Condition not reported	2,998,832		
		8	Condition not on certificate	4,546		
		9	Condition not classifiable	23,182	23,182	0.77%
INJURY	Birth injury	1	Condition reported	8,718		
		2	Condition not reported	2,650,340		
		8	Condition not on certificate	344,316		
		9	Condition not classifiable	26,479	16,273	0.54%
ALCOSYN	Fetal alcohol syndrome	1	Condition reported	107		
		2	Condition not reported	2,942,680		
		8	Condition not on certificate	62,979		
		9	Condition not classifiable	24,087	16,273	0.54%
HYALINE	Hyaline membrane disease	1	Condition reported	19,465		
		2	Condition not reported	2,982,660		
		8	Condition not on certificate	4,546		
		9	Condition not classifiable	23,182	23,182	0.77%
MECONSYP	Meconium aspiration syndrome	1	Condition reported	4,332		
		2	Condition not reported	2,997,793		
		8	Condition not on certificate	4,546		
		9	Condition not classifiable	23,182	23,182	0.77%

Table 10: Newborn Complications Continued

<i>Var Name</i>	<i>Var Definition</i>	<i>Categories</i>	<i>Category Definitions</i>	<i>Frequency</i>	<i>NMISS</i>	<i>% NMISS</i>
VENL30	Assisted ventilation, less than 30 minutes	1	Condition reported	68,720		
		2	Condition not reported	2,819,573		
		8	Condition not on certificate	110,155		
		9	Condition not classifiable	31,405	31,405	1.04%
VEN30M	Assisted ventilation, 30 minutes or more	1	Condition reported	29,621		
		2	Condition not reported	2,858,672		
		8	Condition not on certificate	110,155		
		9	Condition not classifiable	31,405	31,405	1.04%
NSEIZ	Seizures	1	Condition reported	1,475		
		2	Condition not reported	3,000,650		
		8	Condition not on certificate	4,546		
		9	Condition not classifiable	23,182	23,182	0.77%
OTHERAB	Other Abnormal Conditions of the Newborn	1	Condition reported	142,408		
		2	Condition not reported	2,859,717		
		8	Condition not on certificate	4,546		
		9	Condition not classifiable	23,182	23,182	0.77%

Table 11: Congenital Anomalies in 2002 US birth data

<i>Var Name</i>	<i>Var Definition</i>	<i>Categories</i>	<i>Category Definitions</i>	<i>Frequency</i>	<i>NMISS</i>	<i>% NMISS</i>
CONGENIT	Congenital Anomalies	1	Condition reported			
		2	Condition not reported			
		8	Condition not on certificate			
		9	Condition not classifiable			
ANEN	Anencephalus	1	Condition reported	310		
		2	Condition not reported	2,977,180		
		8	Condition not on certificate	28,093		
		9	Condition not classifiable	24,270	24,270	0.80%
SPINA	Spina bifida/ Meningocele	1	Condition reported	632		
		2	Condition not reported	2,976,858		
		8	Condition not on certificate	28,093		
		9	Condition not classifiable	24,270	24,270	0.80%
HYDRO	Hydrocephalus	1	Condition reported	729		
		2	Condition not reported	2,976,761		
		8	Condition not on certificate	28,093		
		9	Condition not classifiable	24,270	24,270	0.80%
MICROCE	Microcephalus	1	Condition reported	162		
		2	Condition not reported	2,977,328		
		8	Condition not on certificate	28,093		
		9	Condition not classifiable	24,270	24,270	0.80%
NERVOUS	Other central nervous system anomalies	1	Condition reported	704		
		2	Condition not reported	2,976,786		
		8	Condition not on certificate	28,093		
		9	Condition not classifiable	24,270	24,270	0.80%

Table 11: Congenital Anomalies Continued

<i>Var Name</i>	<i>Var Definition</i>	<i>Categories</i>	<i>Category Definitions</i>	<i>Frequency</i>	<i>NMISS</i>	<i>% NMISS</i>
HEART	Heart malformation	1	Condition reported	4,148		
		2	Condition not reported	2,973,342		
		8	Condition not on certificate	28,093		
		9	Condition not classifiable	24,270	24,270	0.80%
CIRCUL	Other circulatory/respiratory anomalies	1	Condition reported	3,716		
		2	Condition not reported	2,973,774		
		8	Condition not on certificate	28,093		
		9	Condition not classifiable	24,270	24,270	0.80%
RECTAL	Rectal atresia/stenosis	1	Condition reported	266		
		2	Condition not reported	2,977,224		
		8	Condition not on certificate	28,093		
		9	Condition not classifiable	24,270	24,270	0.80%
TRACHEO	Tracheo - esophageal fistula/Esophageal atresia	1	Condition reported	301		
		2	Condition not reported	2,977,189		
		8	Condition not on certificate	28,093		
		9	Condition not classifiable	24,270	24,270	0.80%
OMPHALO	Omphalocele/Gastroschisis	1	Condition reported	948		
		2	Condition not reported	2,976,542		
		8	Condition not on certificate	28,093		
		9	Condition not classifiable	24,270	24,270	0.80%
GASTRO	Other gastrointestinal anomalies	1	Condition reported	1,181		
		2	Condition not reported	2,976,309		
		8	Condition not on certificate	28,093		
		9	Condition not classifiable	24,270	24,270	0.80%

Table 11: Congenital Anomalies Continued

<i>Var Name</i>	<i>Var Definition</i>	<i>Categories</i>	<i>Category Definitions</i>	<i>Frequency</i>	<i>NMISS</i>	<i>% NMISS</i>
GENITAL	Malformed genitalia	1	Condition reported	2,799		
		2	Condition not reported	2,974,691		
		8	Condition not on certificate	28,093		
		9	Condition not classifiable	24,270	24,270	0.80%
RENALAGE	Renal agenesis	1	Condition reported	3,716		
		2	Condition not reported	2,973,774		
		8	Condition not on certificate	28,093		
		9	Condition not classifiable	24,270	24,270	0.80%
UROGEN	Other urogenital anomalies	1	Condition reported	3,385		
		2	Condition not reported	2,974,105		
		8	Condition not on certificate	28,093		
		9	Condition not classifiable	24,270	24,270	0.80%
CLEFTLP	Cleft lip/palate	1	Condition reported	2,451		
		2	Condition not reported	2,975,039		
		8	Condition not on certificate	28,093		
		9	Condition not classifiable	24,270	24,270	0.80%
ADACTYLY	Polydactyly/Syndactyly/Adactyly	1	Condition reported	2,621		
		2	Condition not reported	2,974,869		
		8	Condition not on certificate	28,093		
		9	Condition not classifiable	24,270	24,270	0.80%
CLUBFOOT	Club foot	1	Condition reported	1,949		
		2	Condition not reported	2,975,541		
		8	Condition not on certificate	28,093		
		9	Condition not classifiable	24,270	24,270	0.80%

Table 11: Congenital Anomalies Continued

<i>Var Name</i>	<i>Var Definition</i>	<i>Categories</i>	<i>Category Definitions</i>	<i>Frequency</i>	<i>NMISS</i>	<i>% NMISS</i>
HERNIA	Diaphragmatic hernia	1	Condition reported	386		
		2	Condition not reported	2,977,104		
		8	Condition not on certificate	28,093		
		9	Condition not classifiable	24,270	24,270	0.80%
MUSCULO	Other musculoskeletal/integumental anomalies	1	Condition reported	7,391		
		2	Condition not reported	2,970,099		
		8	Condition not on certificate	28,093		
		9	Condition not classifiable	24,270	24,270	0.80%
DOWNS	Down's syndrome	1	Condition reported	1,489		
		2	Condition not reported	2,976,001		
		8	Condition not on certificate	28,093		
		9	Condition not classifiable	24,270	24,270	0.80%
CHROMO	Other chromosomal anomalies	1	Condition reported	1,038		
		2	Condition not reported	2,976,452		
		8	Condition not on certificate	28,093		
		9	Condition not classifiable	24,270	24,270	0.80%
OTHERCON	Other Congenital Anomalies	1	Condition reported	12,460		
		2	Condition not reported	2,965,041		
		9	Condition not classifiable	52,352	52,352	1.73%

Table 12: Summary Variables

<i>Variable</i>	<i>Variable Definition</i>	<i>N</i>	<i>Mean</i>	<i>Std Dev</i>	<i>Min</i>	<i>Max</i>
MEDRISKSUM	Total number of medical risks	3001218	0.4108948	0.6673149	0	10
OBSTETRCSUM	Total number of newborn complications	3012791	2.0945001	0.9594637	0	7
LABORSUM	Total number of the complications of labor and/or delivery	3009034	0.4252209	0.6728824	0	8
NEWBORNSUM	Total number of abnormal conditions of the newborn	3002125	0.0926474	0.340297	0	6
CONGENITSUM	Total number of congenital anomalies	2977501	0.0177269	0.1669546	0	11

Bibliography

- Behrman, R. E., & Butler, A. S. (2007). *Preterm Birth: Causes, Consequences, and Prevention*. Washington, D.C.: National Academies Press.
- Hamilton, B. E., Martin, J. A., & Ventura, a. S. (2009). *Births: Preliminary Data for 2007*. Hyattsville, MD: National Vital Statistics Reports.
- Martin, J. A., Hamilton, B. E., Sutton, P. D., Ventura, S. J., Menacker, F., Kirmeyer, S., et al. (2009). *Births: Final Data for 2006*. Hyattsville, MD: National Vital Statistics Reports.
- Mattison, D. R., Wilson, S., Coussens, C., & Gilbert, D. (2003). *The Role of Environmental Hazards in Premature Birth*. Washington, D.C.: The National Academics Press.
- National Center for Health Statistics. (2002). *Natality 2002*. Hyattsville, MD: National Center for Health Statistics.
- Peterson, J., Taylor, H. G., Minich, N., Klein, N., & Hack, M. (2006). Subnormal head circumference in very low birth weight children: Neonatal correlates and school-age consequences. *Early Human Development* , 82 (5), 325-334.
- Qin, C., Hsia, J., & Berg, C. J. (2008). Variation between Last-Menstrual-Period and Clinical Estimates of Gestational Age in Vital Records. *American Journal of Epidemiology* , 167 (6), 646–652.
- Simhan, H. N., & Caritis, S. N. (2007). Prevention of Preterm Delivery. *The New England Journal of Medicine* , 357, 477-87.
- Slattery, M. M., & Morrison, J. J. (2002). Preterm delivery. *Lancet* , 360, 1489–97.

Ventura, S. (2004). Reproductive Statistics Branch: Natality Data. *American Sociological Association*. San Francisco.

Zahl, S. M., & Wester, K. (2008). Routine Measurement of Head Circumference as a Tool for Detecting Intracranial Expansion in Infants: What Is the Gain? A Nationwide Survey. *Pediatrics* , 416-420.