IMPACT OF AIR POLLUTION

ON STROKE INCIDENCE

IN TEXAS

THESIS

Presented to Graduate Council
of Texas State University – San Marcos
in Partial Fulfillment
of the Requirements

for the degree

Master of SCIENCE

by

Sowmya Anand, B.D.S

San Marcos, Texas
December 2004

DEDICATION

This thesis is dedicated to my Guru, my husband, my brother Sathya, my parents and my in-laws who have supported me all the way since beginning of my studies. My special thanks to my husband Anand. Throughout my thesis-writing period he provided encouragement and great ideas. I would not be able to finish my thesis without him. I am lucky to have such a husband who was a great source of motivation and inspiration. Dedicating this thesis to my husband would be very little for the encouragement he has given me from the beginning of my graduate studies.

ACKNOWLEDGEMENTS

It is difficult to overstate my gratitude to my thesis supervisor, Dr. Ram Shanmugam. With his great effort, he made this thesis simple for me. I am also grateful to my other committee members, Dr. Charles Johnson and Dr. Jean Brender, for sharing their knowledge in the area of epidemiology and biostatistics.

This manuscript was submitted on 12/06/2004.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

ABSTRACT

IMPACT OF AIR POLLUTION ON STROKE INCIDENCE IN TEXAS

by

Sowmya Anand B.D.S

Texas State University-San Marcos

December 2004

SUPERVISING PROFESSOR: RAM SHANMUGAM

Stroke is a cardiovascular disease that affects arteries supplying blood to the brain. A stroke is caused when blood vessels in the brain are blocked by a clot or when the blood vessel bursts. Cardiovascular diseases (heart disease and stroke) have been the leading cause of death in Texas since 1940. About 2 out of every 5 deaths are due to cardiovascular disease in Texas. Risk factors for stroke include high blood pressure, tobacco use, diabetes mellitus, atrial fibrillation, high blood cholesterol, physical inactivity, obesity, excessive alcohol intake and increasing age. The objective of this thesis was to examine the relation between air pollution and stroke death rate in Texas in 1999. Air pollutants considered in this study were carbon monoxide, sulfur dioxide,

nitrogen oxide, particulate matter (PM) 2.5, PM 10, volatile organic compounds and total criteria air pollutant. The stroke death rate was collected for the year 1999. With the stroke death rate as a dependent variable and carbon monoxide, sulfur dioxide, nitrogen oxide, PM 2.5, PM 10, volatile organic compounds pollutant level and total criteria pollutant as independent variables, a linear regression analysis was done. A linear regression analysis with subsets of the data was also performed. In subset analysis, log of the population was added as a covariate. Analyses revealed significant pollutants for stroke mortality. A similar analysis was done using Poisson regression since stroke mortality is a count variable. The population was used as a covariate in Poisson regression. Linear regression analysis with the entire data implied that air pollutant levels were not a good predictor of the stroke death rate. Linear regression analysis was performed with the subset of data in which, counties with stroke deaths less than ten were removed. Log of the population was added as a covariate. The slope associated with carbon monoxide pollutant was positive, which implied that as carbon monoxide pollutant increased by one ton per square mile, stroke death rate would increase by 0.1911. Linear regression analysis was performed with the second subset of data in which counties with stroke deaths less than 15 were removed. The slope associated with the carbon monoxide pollutant was positive, which implied that as carbon monoxide pollutant increased by one ton per square mile, stroke death rate would increase by 0.2434. Subset three was formed with all the counties in which the stroke deaths were equal to or less than 20. No significant variables were identified in subset three. Subset four was formed with all the counties in which the stroke deaths were greater than 20. The subset four model showed that as the volatile organic compound pollutant increased by one unit, the stroke death rate would increase by 1.201 units. Subset five was formed with all the counties in which stroke deaths were equal to or less than 25. Subset six was formed with all the counties in which stroke deaths were greater than 25. The subset five model did not yield any significant variable and the model was insignificant. The model six showed that if carbon monoxide increased by one unit, stroke death rate increased by 0.2269 units. A Poisson regression analysis was performed with full data. Number of stroke deaths was considered as the dependent variable. Carbon monoxide, sulfur dioxide, nitrogen oxide, PM 2.5, PM 10, volatile organic compounds pollutant level and total criteria pollutant were independent variables which were measured in tons per square miles. Log of the population was also considered as an independent variable, which acted as a covariate. The estimated model was the following: y (stroke death) = Exponential (-4.812 + 0.00328 * Volatile organic compounds + 1.75923 * Log of population + 0.00054 * Carbon monoxide). The slope associated with carbon monoxide was positive, which implies that as carbon monoxide increased by one ton per square mile, stroke death increased at the rate of 1.00054. The slope associated with volatile organic compounds was positive, which implies that as volatile organic compounds increased by one ton per square mile, stroke death increased at the rate of 1.00328. The carbon monoxide pollutant level was a significant factor in three linear regression models and the Poisson regression model. Therefore, higher carbon monoxide pollutant level would increase the risk of deaths from stroke. Volatile organic compounds pollution level was a significant factor in the Poisson regression model and subset four. Further studies should be done to understand the association between air pollutants and stroke death.

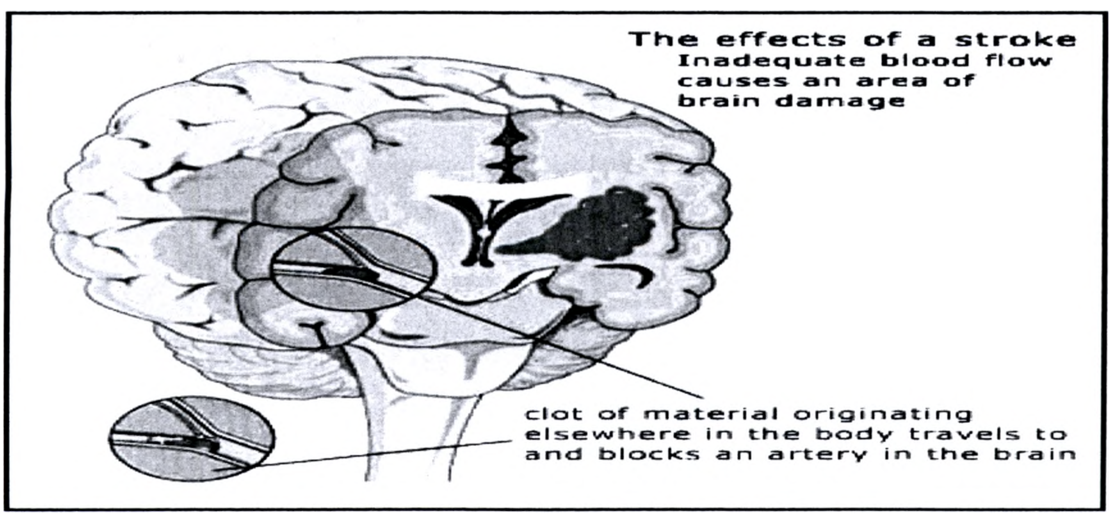
CHAPTER I

INTRODUCTION

Air Pollution and Stroke Incidence

Stroke is a cardiovascular disease that affects arteries supplying blood to the brain. A stroke is caused when blood vessels in the brain are blocked by a clot or when the blood vessel bursts. When blood vessels are blocked, oxygen supply is reduced. The brain cells without oxygen start to die.

FIGURE 1. Causes of a stroke



(Picture reference: http://www.irishhealth.com/content/image/11/Image1.gif)

There are two common types of strokes: ischemic stroke and hemorrhagic stroke. When the cause is due to clotting, it is named as ischemic stroke. This type of stroke accounts 88% of the strokes incidence. When the brain cells in the infarct die, some chemicals are released and those chemicals cause a chain reaction that produces a phenomenon called ischemic cascade. This ischemic cascade affects brain cells and surrounding brain tissues are also deprived of blood supply. If a medical treatment is not done promptly, this brain area called penumbra will die. Some type of intervention treatment should be given within six hours of the stroke because of a rapid pace of ischemic cascade. Once the blood flow is not reestablished or neuro-protective agents are not useful, then the patient will die. When the cause of stroke is ruptured blood vessels, it is called hemorrhagic stroke. If the blood supply is stopped, the affected brain cells die. If any part of the brain dies, the part of the body it controls becomes nonfunctional. The effect of stroke depends on the location and extent of the brain damage that is caused by an obstruction. If a part of the right side of the brain is affected, the left side of the body and the right side of the face become nonfunctional. It could result in any one or more of the following complications depending on the extent of brain damage: paralysis on the left side of the body, problems with vision, quick inquisitive behavioral style, and a loss of memory. If a part of the left side of the brain is affected, it may result in some or all of the following: paralysis on the right side of the body, problems with speech/language, slow or cautious behavioral style and a loss of memory. If a stroke occurs toward the back of the brain, it causes vision problems. These illustrate that different parts of the body become nonfunctional depending on the location of the stroke and extent of the damage.

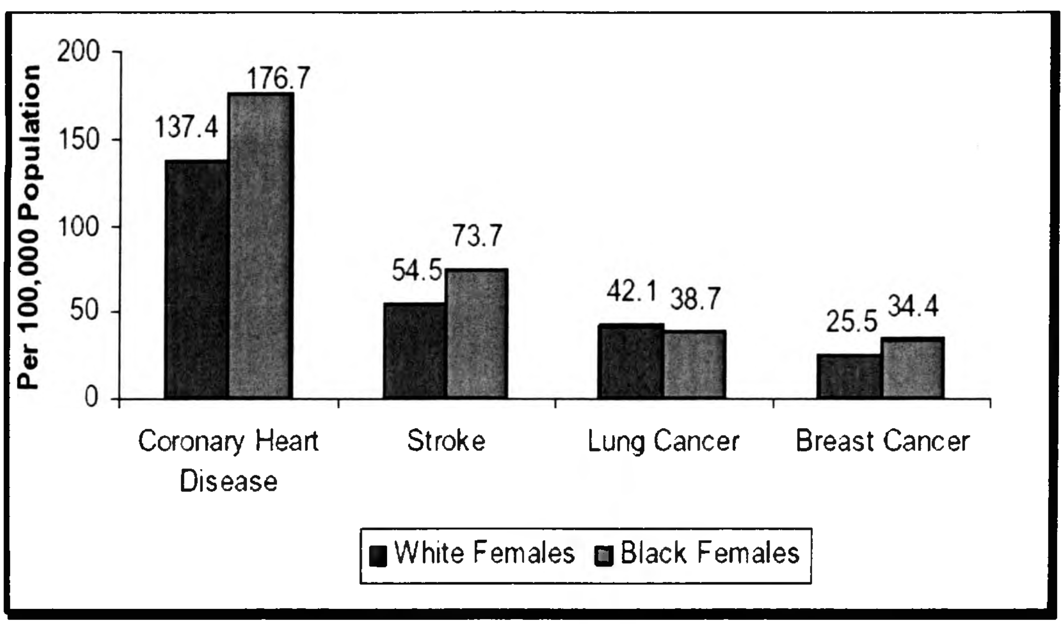
It is helpful to capture some warning signs. Most commonly noted warning signs of a stroke are as follows: sudden numbness of face, arm or leg particularly on one side of the body, sudden confusion, trouble in speaking, seeing, walking, dizziness, loss of body balance and severe headache. The stroke is diagnosed using CT (computerized tomography) scans, ECG (electrocardiogram), chest X-rays and blood tests. In this process, symptoms are identified and used to identify whether a stroke or some type of brain disorder causes it. Consequently, the outcomes also determine the extent of the brain damage.

Risk factors for stroke include high blood pressure, tobacco use, diabetes mellitus, atrial fibrillation, high blood cholesterol, physical inactivity, obesity, excessive alcohol intake and increasing age. Smokers have increased risk of getting heart disease. Being physically active decreases the risk of having stroke. It lowers the total cholesterol and also lowers high blood pressure. Being physically active also increases the good cholesterol (HDL). Thus being physically active decreases the risk of having a stroke. If a body contains too much cholesterol because of improper diet, such excess cholesterol is deposited in arteries, which may cause stroke at a later time. These risks are then avoidable. Unavoidable risks include age, gender and race. Males are more prone to stroke than females. Blacks are more prone to stroke than any other races.

In 2000, about 61% of 167,661 people who died from stroke were women. This accounts for about 1 in every 14 deaths. The death rate in general for stroke was 61 per 100,000 in the population. In 2001, 163,538 persons died from a stroke. Among serious long-term disabilities in United States, stroke is considered to be the leading cause. In 2004, an estimated total treatment costs for stroke is $53.6 billion. This implies that stroke treatment is very costly Each year about 700,000 people experience some form of stroke that includes new and recurrent stroke. During 1991 to 2001, the death rate from stroke had decreased by 3.4 percent, even though the actual number of stroke deaths rose by 7.7 percent. The stroke death rates in 2001 were 56.5 per 100,000 among white males, 85.4 per 100,000 for black males, 54.5 per 100,000 for white females and 73.6 per 100,000 for black females in the population.

The age-adjusted stroke incidence rates for new non-recurrent strokes are about 167 per 100,000 for white males, 138 per 100,000 for white females, 323 per 100,000 for black males and 260 per 100,000 for black females in the population. These statistics imply that the blacks are more affected by stroke than the whites. The blacks in United States have incidence rates twice that of whites. The following graph shows the stroke rates in 2001 for both white and black females.

FIGURE 2. Age-Adjusted death rates for coronary heart disease, stroke, and lung and breast cancer for white and black females, United States 2001

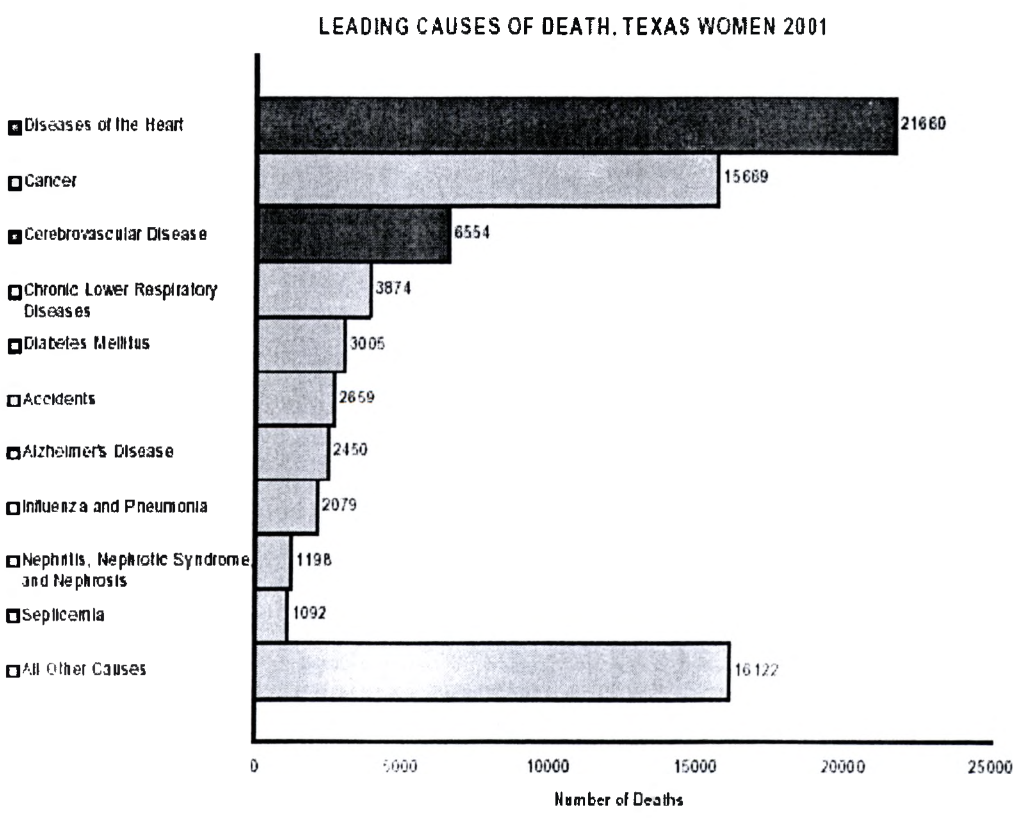


(Picture Reference: http://www.americanheart.org)

This comparison also suggests that the black females are more affected by stroke than the white females. In 1999, around 1.1 million Americans were reported to have functional limitations and reduction in ability to perform activities of daily living as a result of stroke. These statistics also confirm that men are more affected by stroke than women in general.

The stroke incidences in Texas are more revealing. Cardiovascular diseases (heart disease and stroke) have been the leading cause of death in Texas since 1940. About 2 out of every 5 deaths are due to cardiovascular disease in Texas.

FIGURE 3. Leading causes of death, Texas women 2001



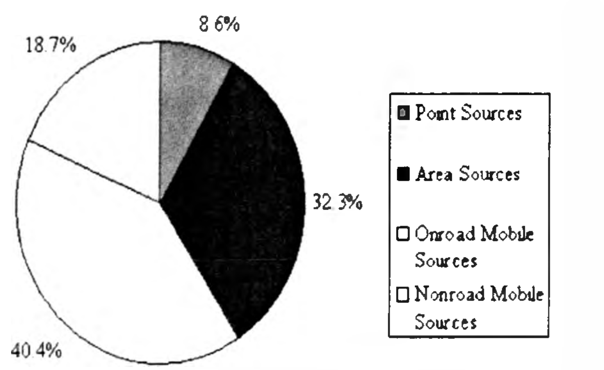
(Picture reference: http://www.tdh.state.tx.us/wellness)

The above graph shows that cardiovascular diseases are the leading cause of death among Texas women during the year 2001.

Some type of prevention and treatment for stroke could reduce the severity of the death rate due to stroke. For example the Stroke Association (in UK) estimates that about 40% of strokes would be prevented by: treatment for high blood pressure (hypertension), smoking cessation, loss of excess weight, regular physical exercise, reducing alcohol intake to within recommended limits, good diet with more fruits and vegetables and eating less salt and fatty foods. Once a person is affected by stroke, there is no known treatment to reverse it. Rehabilitation treatment is the main treatment for stroke, which focuses only on improving physical activities.

This thesis examines the relevance of air pollutants to the incidence of stroke. Some important air pollutants are explained below.

FIGURE 4. Criteria air pollutants sources in United States

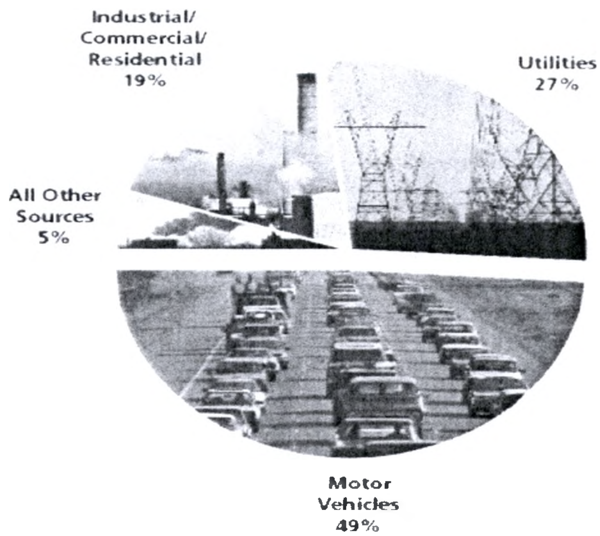


(Picture reference: http://www.pca.state.mn.us/air/criteria-emissiongraphs.html)

The Environmental Protection Agency (EPA) uses six "criteria pollutants" as indicators of air quality, and has provided tolerance limits for each of them. Figure 4 shows that main source of criteria pollutants were the mobile sources. A maximum concentration level above the tolerance limit might affect adversely on the human health may occur. These limits are called as National Ambient Air Quality Standards (NAAQS). These criteria pollutants are ozone, carbon monoxide, nitrogen dioxide, sulfur dioxide, lead and particulate matter. The EPA has declared that many people (in tens of thousands) of people die every year because of inhaling tiny particles (particulate matter) in the air.

The particulate matter includes dust, dirt, soot, smoke and liquid droplets that are emitted into the air by natural or synthetic sources. Particles formed by the condensation or the transformation of emitted gases such as sulfur dioxide and volatile organic compounds (VOC) can also constitute particulate matter. Carbon monoxide is a colorless, odorless gas that is formed when carbon in fuel is not burnt completely. It is a component of motor vehicle exhaust that contributes about 56 percent of all carbon monoxide emissions. Other non-road engines and vehicles (such as construction equipment and boats) could also contribute to air pollution. Nitrogen oxide is a pollutant generally referred to a group of highly reactive gases, all of which contain nitrogen and oxygen but in varying amounts. The sources of nitrogen oxides are from motor vehicles, electric utilities, and other industrial, commercial, and residential sources that burn fuels.

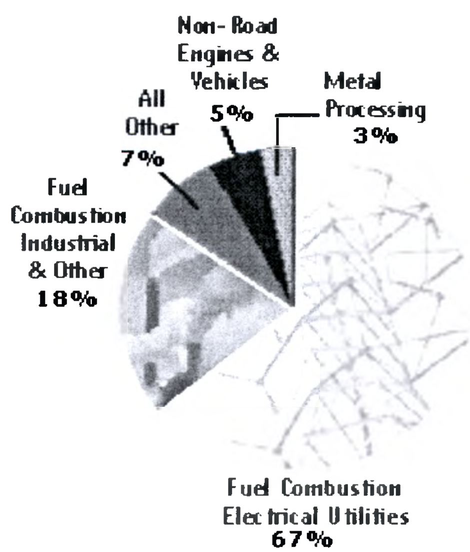
FIGURE 5. Carbon monoxide source, United States, 1999



(Picture reference: http://www.epa.gov/air/urbanair/nox/what.html)

Sulfur dioxide is a pollutant, which belongs to the family of sulfur oxide gases. About 65% of the sulfur dioxide that is released to the air (equivalently about 13 million tons per year) is from utility industries, those in particular that burns coal.

FIGURE 6. Sulfur dioxide source, United States, 1999



(Picture reference: http://www.epa.gov/air/urbanair/nox/what.html)

The volatile organic compounds (VOCs) include a large group of chemicals containing carbon and hydrogen atoms that can react quickly to form other chemicals in atmosphere. The major source of volatile organic compounds in atmosphere is from vehicle exhaust emissions. For each of the above mentioned air pollutants, the EPA has records for two types of air pollution trends: First is an actual pollutant concentration in the ambient air at selected monitoring sites throughout the country. Secondly it is total tons of pollutants released into the air each year. The data on total tons of pollutants in the air each year are used for this thesis.

I now discuss some previous studies that show the relevance of disease mortality to the air pollution. Pope et al (2002) found that fine particulate and sulfur oxide related pollution were associated with cardiopulmonary mortality. However this study did not focus particularly on stroke. It showed a significant association between mortality due to cardiopulmonary disease and air pollution. They also showed an association between mortality due to all causes and fine particulate air pollution. Samet et al (2000) noted a strong association with PM 10 with death from all causes and due to cardiovascular and respiratory disease. In this research, the authors did a time series analysis. Data were collected from 20 U.S cities with a large population. They used time series analysis for these data. They also used Bayesian methods to analyze the data. They found consistent evidence that particulate matter is indeed associated with deaths from all causes and death due to cardiovascular disease. However, this study was not based on deaths due to stroke. The authors considered all cardiovascular deaths in their study.

Hong et al (2002) developed a model that was controlled for time trends and meteorological influence like temperature, humidity and atmospheric pressures. They showed that effects of air pollutants on ischemic stroke mortality were statistically significant. There were increased relative risks of 1.03 (95% CI, 1.00 to 1.06) ischemic stroke mortality for each inter-quartile range increase in total suspended particulates on the same day. There were increased relative risks of 1.04 (95%CI, 1.01 to 1.08) for ischemic stroke mortality for each inter-quartile range increase in sulfur dioxide concentrations on the same day. There was an increased relative risk of 1.04 (95% CI, 1.01 to 1.07) for nitrogen dioxide with a one-day lag. There was an increased relative risk of 1.06 (95% CI, 1.02 to 1.09) for carbon monoxide with a one-day lag. There was an increased relative risk of 1.06 (95% CI, 1.02 to 1.10) for ozone for each inter-quartile range increase with a three-day lag. They also showed that air pollutants were significantly associated with ischemic stroke mortality. This research investigated only stroke with time series analysis. This study established significant association between acute stroke mortality and air pollution.

Kan et al (2003) also used a time series approach to see an association between air pollution and daily stroke mortality in Shanghai. The relative risk was significant for PM10 and NO2 but not for SO2. The relative risk for PM 10 was 1.008 (95% confidence interval 1.000-1.016). The relative risk for SO2 was 1.017 (95% confidence interval .998-1.036). The relative risk for NO2 was 1.029 (95% confidence interval 1.001 -1.057). This study also confirmed a significant association of PM 10 and NO2 with stroke mortality.

Maheswaran and Elliott (2003) associated stroke with air pollution; road traffic was identified as a major source of outdoor air pollution. They concluded that stroke mortality was 7% (95% confidence interval [CI], 4% to 9%) higher in men living within

200 m of a main road compared with men living more than 1000 m away. They also showed an increase in risk for women of 4% (95% CI, 2% to 6%) and the risk for both men and women of 5% (95% CI, 4% to 7%). This also suggests that air pollution mainly caused by traffic is associated with stroke mortality. In this study, they included spatial analysis. Sunyer et al (2003) noted a short-term effect of sulfur dioxide on cardiovascular diseases. They noticed a significant association, showed an increase of 0.7% of hospital admissions with 0.1 to 1.3 per 10-microgram/meter-cube increase of So2 among subjects younger than 65 years for Ischemic heart disease. They did not see such patterns for stroke admission.

Objective

The objective of this thesis is to examine the relation between air pollution and stroke death rate in Texas in 1999. Air pollutants considered in this study are carbon monoxide, sulfur dioxide, nitrogen oxide, PM 2.5, PM 10, volatile organic compounds and total criteria air pollutant. The stroke death rate is collected for the year 1999. The data collected from WebPages is for all the counties in Texas.

CHAPTER II

METHODS

This thesis used the existing data provided from the Texas Department of State Health Services website and also from the Texas Environmental Profile websites. Stroke mortality for all Texas counties was collected for the year 1999. The variables collected from Texas environmental profile websites include carbon monoxide, sulfur dioxide, nitrogen oxide, PM 2.5, PM 10, volatile organic compounds and total criteria pollutant level measured in tons per year for the year 1999. In my data analysis, the stroke death rate was the dependent variable. Carbon monoxide, sulfur dioxide, nitrogen oxide, PM 2.5, PM 10 volatile organic compounds and total criteria pollutant level per square miles were the independent variables. In data analysis, all pollutant levels were examined to assess if they significantly affect stroke incidence.

The Data Collection

The air quality data was collected from the website http://www.texasep.org/. The total amount of criteria air pollutant, carbon monoxide, sulfur dioxide, nitrogen oxide, PM 2.5, PM 10 and volatile organic compounds pollutants amount were collected for each county in this manner. Air pollutant in tons per square miles was calculated using area of each county. I collected the number of stroke deaths in each Texas County from website of Department of State Health Services for the year 1999. Stroke death rate was calculated for 1999 using population data and number of stroke deaths.

Data Analysis

Using Arc view 8.0, several maps were created to assess any unusual stroke death levels, carbon monoxide, sulfur dioxide, nitrogen oxide, PM 2.5, PM 10 and volatile organic compounds pollutant level at various Texas counties.

Descriptive statistics were calculated using SPSS 11.0. Linear regression analysis was done using NCSS-PASS. With stroke death rate as a dependent variable and carbon monoxide, sulfur dioxide, nitrogen oxide, PM 2.5, PM 10, volatile organic compounds pollutant level and total criteria pollutant as independent variables, a linear regression analysis was done. Analyses revealed significant pollutants for stroke mortality. A similar analysis was done using Poisson regression. Population was used as covariate in Poisson regression. Poisson regression analysis was done using software NCSS-PASS. Different types of regression were done to confirm significant pollutant factors that affected stroke mortality in Texas. A comparison of Poisson regression model and linear regression model was performed and the significant air pollutant that influences stroke occurrence was identified.

CHAPTER III

RESULTS

All 254 Texas counties were used in this study. There were no missing information and all the Texas counties were included in the analysis.

Descriptive Statistics

TABLE 1. Descriptive statistics of stroke death rate and air pollutant in Texas, 1999

Variable	**Mean**	**Standard Deviation**	**Minimum**	**Maximum**
Stroke death rate	76.6653	49.23820	0	340.43
Number of Stroke death	40.9252	122.71902	0	1378
PM 2.5 (tons/square miles)	2.5552	3.20538	.06	21.95
PM 10 (tons/square miles)	12.6692	13.80072	.26	88.35
Carbon Monoxide (tons/square miles)	25.9578	62.49987	.17	582.92
Sulfur Dioxide (tons/square miles)	5.3241	23.05572	0	311.30
Volatile organic compounds (tons/square miles)	6.6543	19.29887	.07	250.47
Nitrogen Oxide (tons/square miles)	10.6834	29.27348	.06	305.13
Criteria Air pollutant (tons/square miles)	62 0617	116.85998	.91	904.68

Table one shows mean, standard deviation, minimum and maximum amount of stroke death rate, number of stroke deaths, carbon monoxide, sulfur dioxide, nitrogen oxide, PM 2.5, PM 10, volatile organic compounds pollutant level and total criteria pollutant in tons per square miles for the year 1999.

Maps

FIGURE 7. Stroke death rate (per 100,000 population) in Texas, 1999

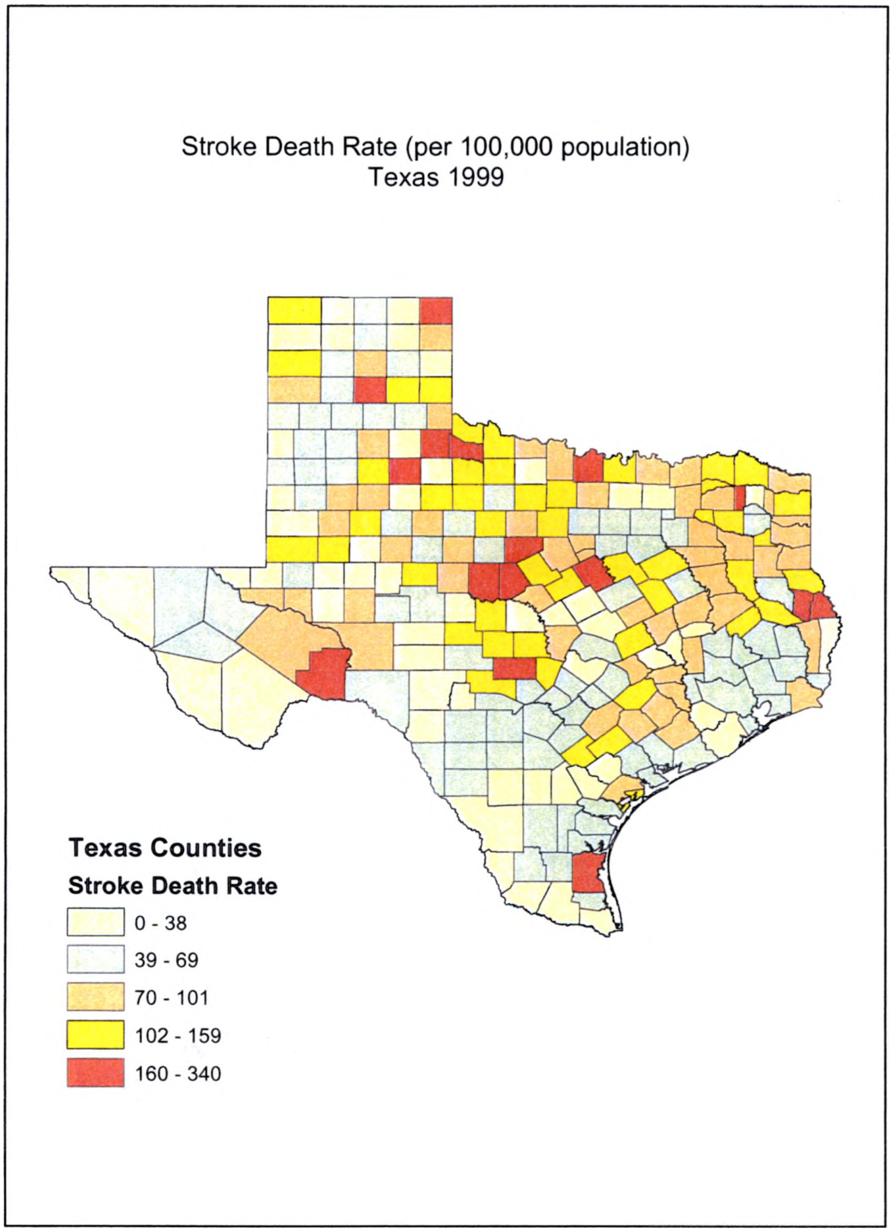


Figure 7 shows the stroke death rate in Texas for the year 1999. Following counties had the highest stroke rate (stroke rate more than 160 per 100,000) in 1999 in Texas: Franklin, Foard, Gillespie, Kenedy, Lipscomb, Montague, Sabine, San August, Brown, Eastland, Terrell, Armstrong, Bosque, Coleman, Cottle and Dickens. Highest stroke death rate was found in Dickens, Texas for the year 1999.

FIGURE 8. Carbon monoxide pollutant (tons per square miles) in Texas, 1999

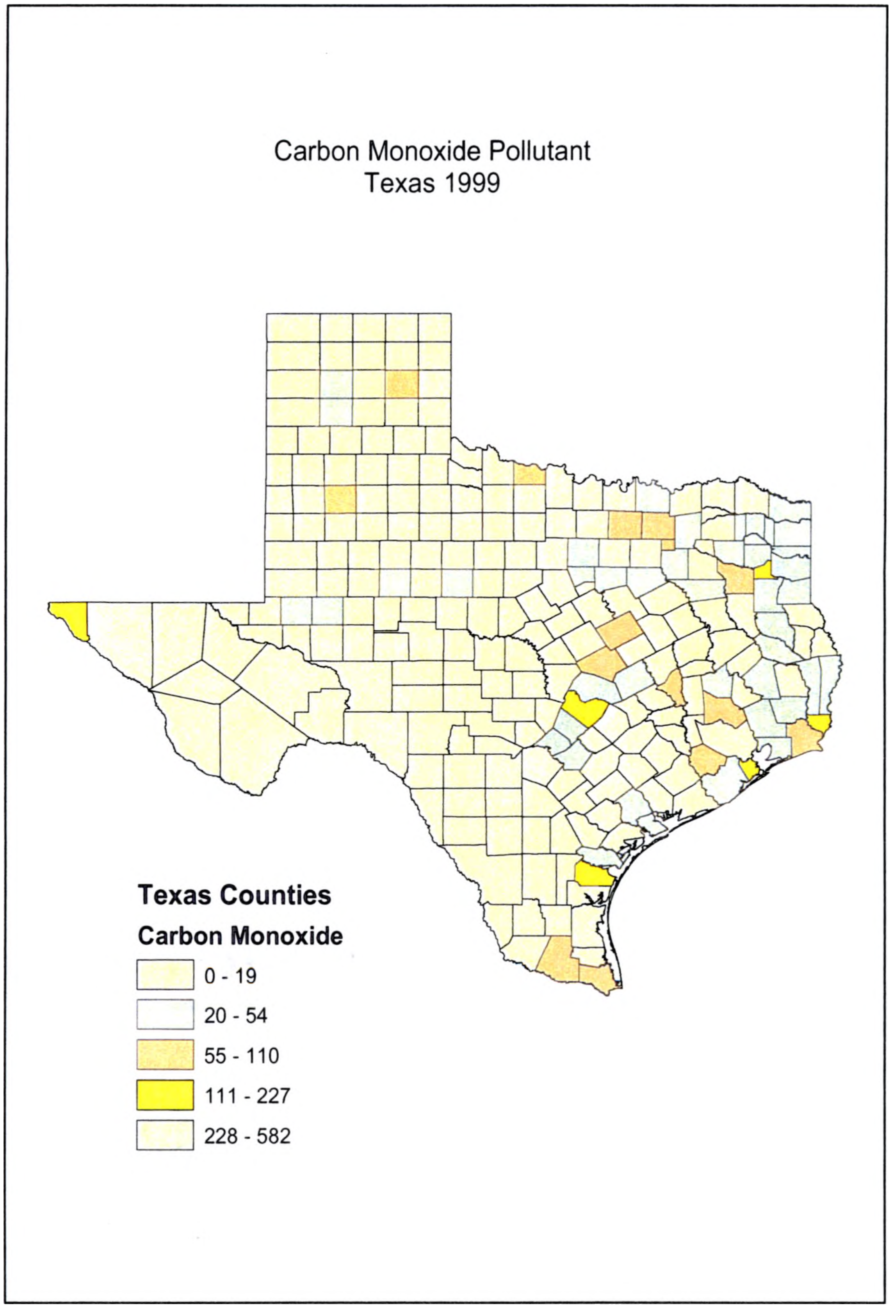


Figure 8 shows the carbon monoxide pollutant (measured in tons/square miles) in Texas for the year 1999. Four counties with the high carbon monoxide pollutant were Bexar, Dallas, Harris and Tarrant. Dallas had the highest carbon monoxide pollutant amount in Texas for the year 1999.

FIGURE 9. Nitrogen dioxide pollutant (tons per square miles) in Texas, 1999

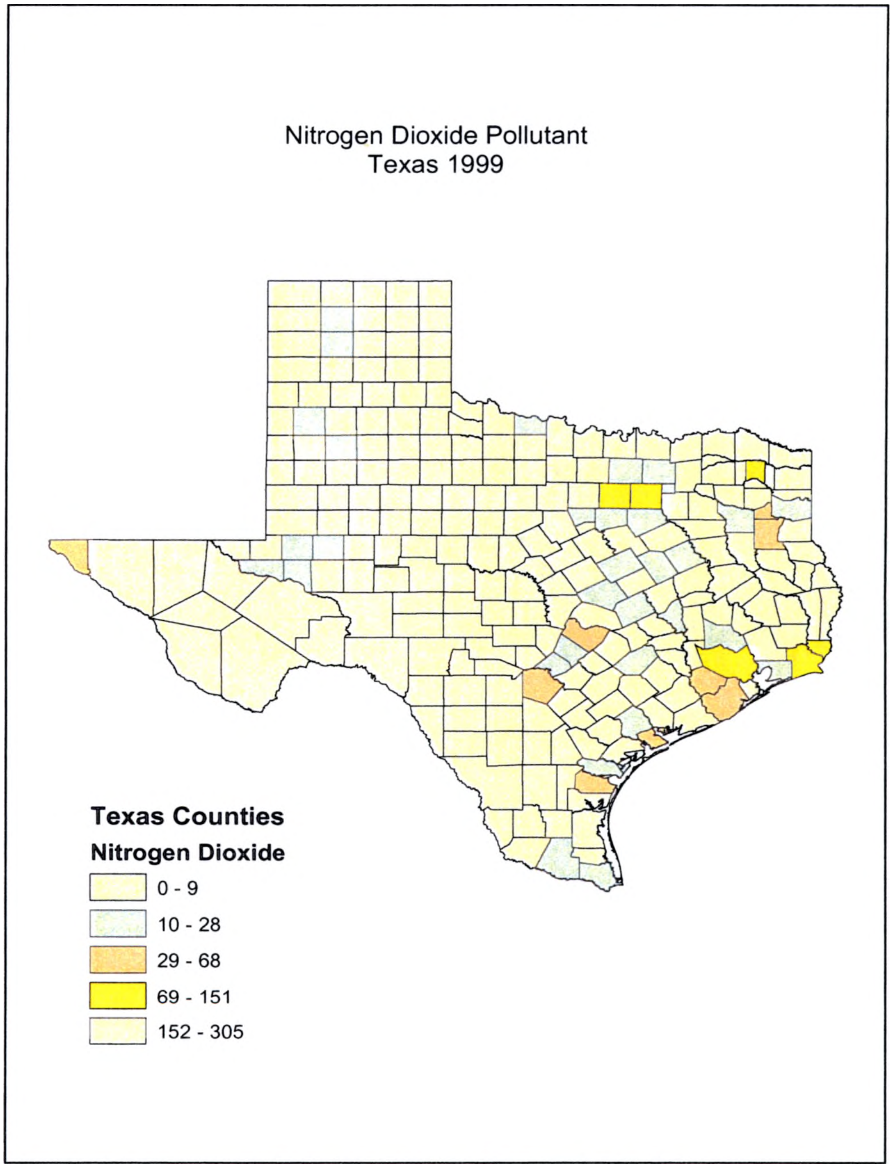


Figure 9 shows that the highest nitrogen dioxide pollutant was found in Galveston in Texas for the year 1999.

FIGURE 10. PM 2.5 pollutant (tons per square miles) in Texas, 1999

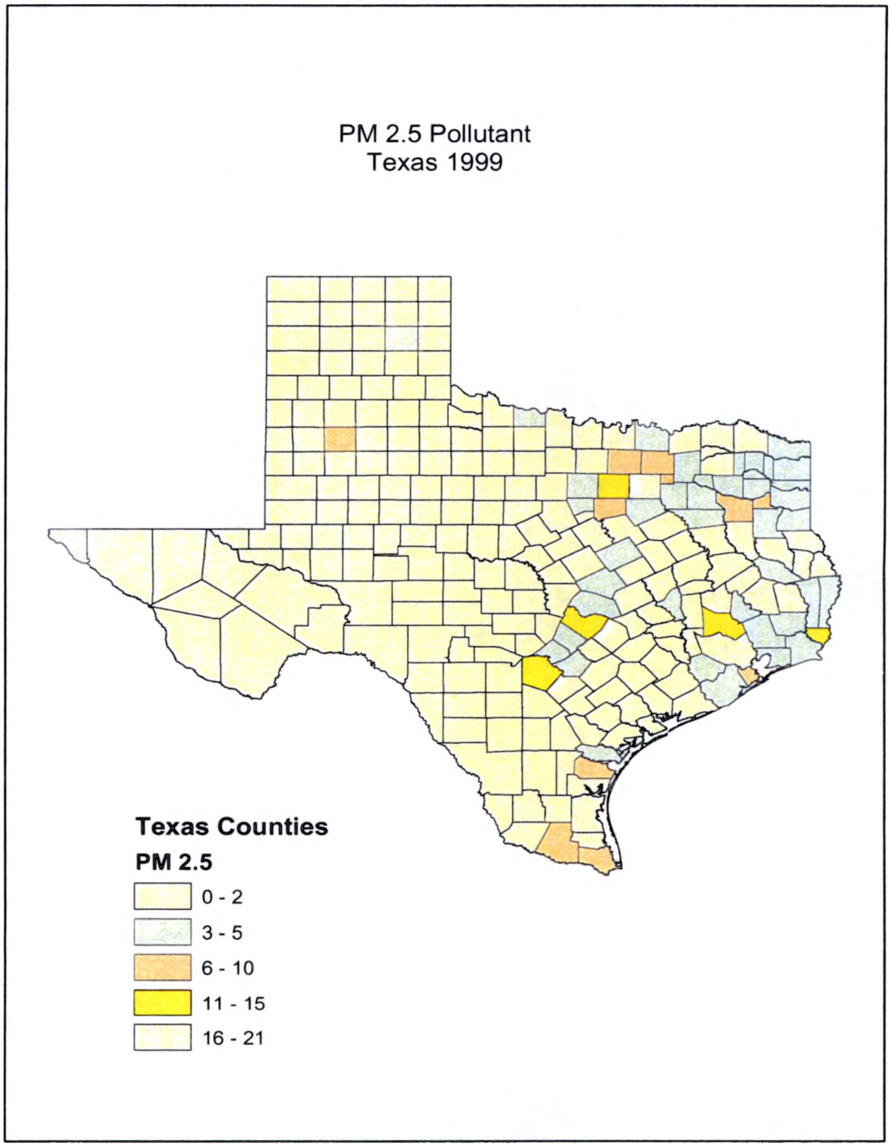


Figure 10 shows that Dallas, Harris, Mitchell were the counties with highest PM 2.5 pollution in Texas for the year 1999.

FIGURE 11. PM 10 pollutant (tons per square miles) in Texas, 1999

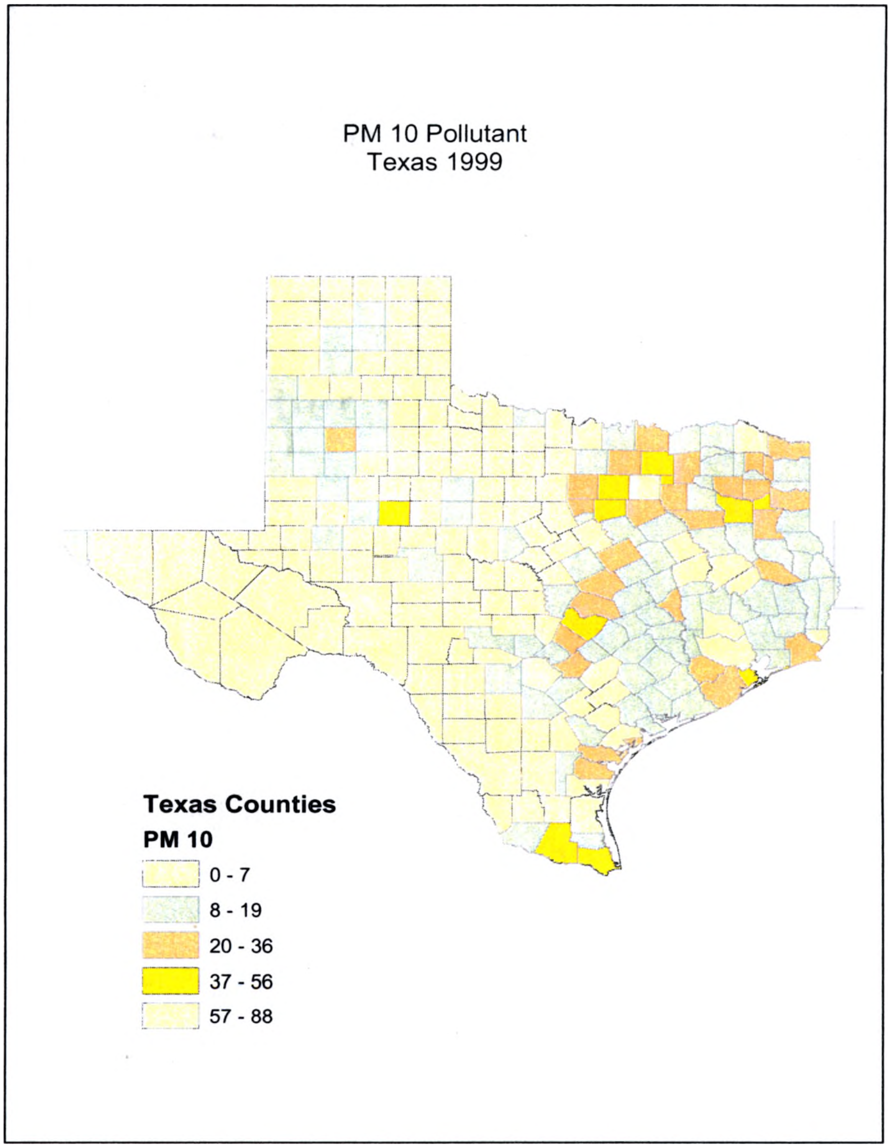


Figure 11 shows that Montgomery, Orange, Harris, Bexar, Dallas were the counties with highest PM 10 pollution in Texas for the year 1999. Harris county had the highest PM 10 level for the year 1999.

FIGURE 12. Sulfur dioxide pollutant (tons per square miles) in Texas, 1999

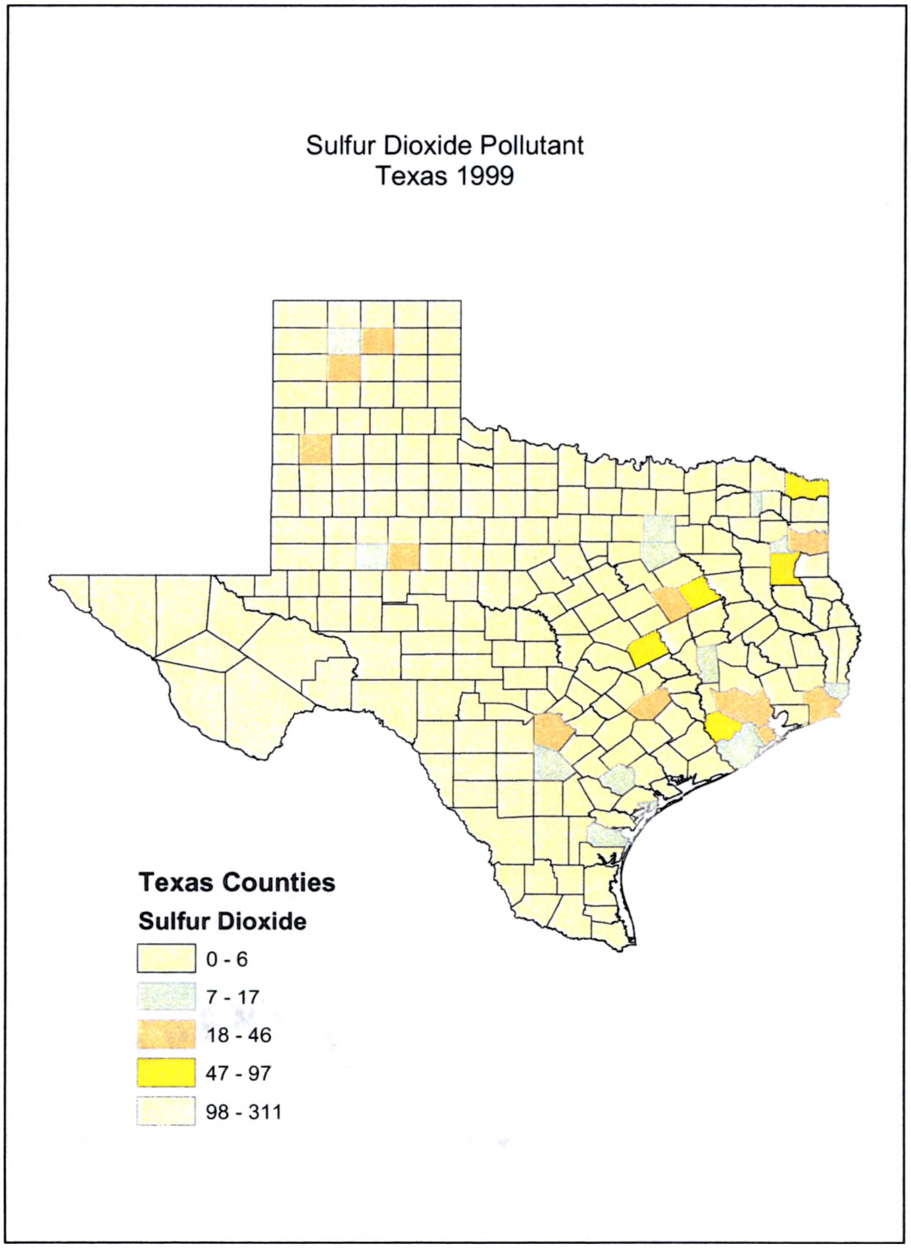


Figure 12 shows that Titus County had the highest Sulfur dioxide pollutant in Texas during 1999.

FIGURE 13. Volatile organic compounds pollutants (tons per square miles) in Texas, 1999

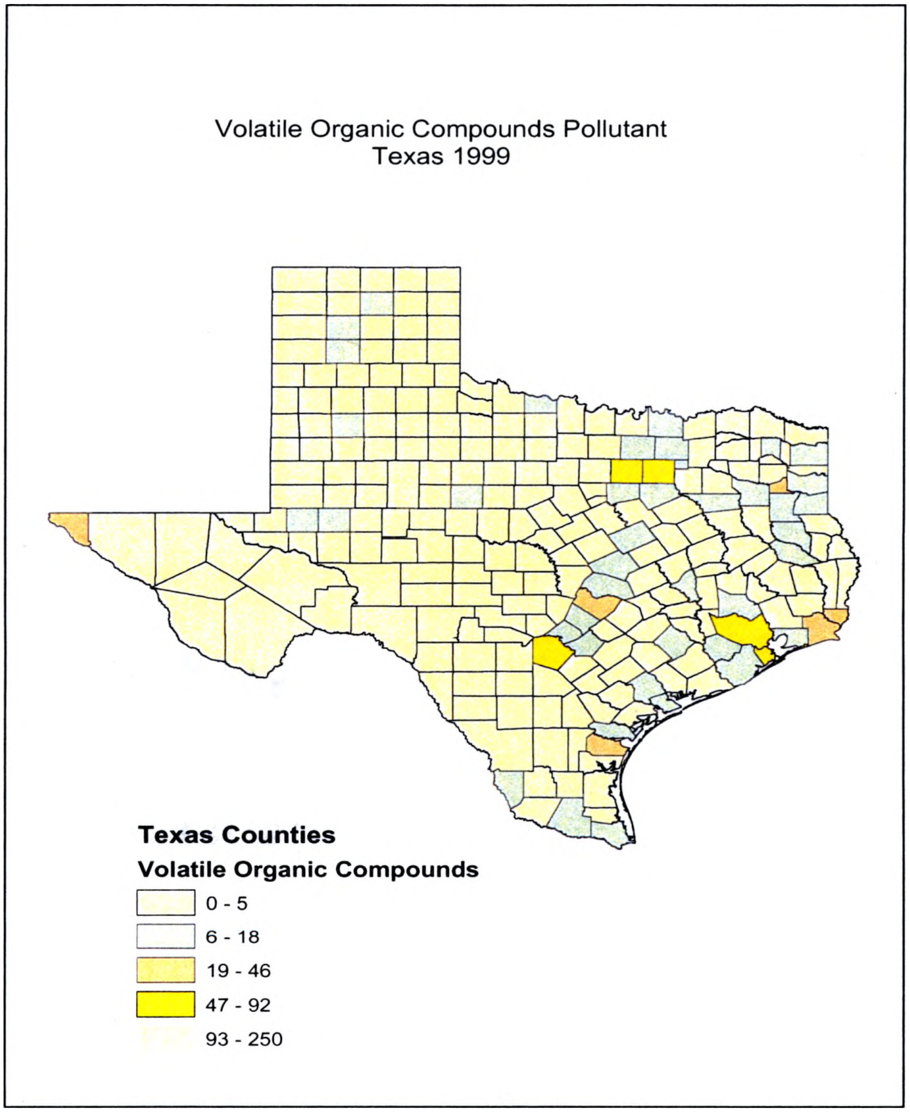


Figure 13 shows that Mitchell County had the highest volatile organic compounds pollutant in Texas for the year 1999.

FIGURE 14. Total criteria air pollutant (tons per square miles) in Texas, 1999

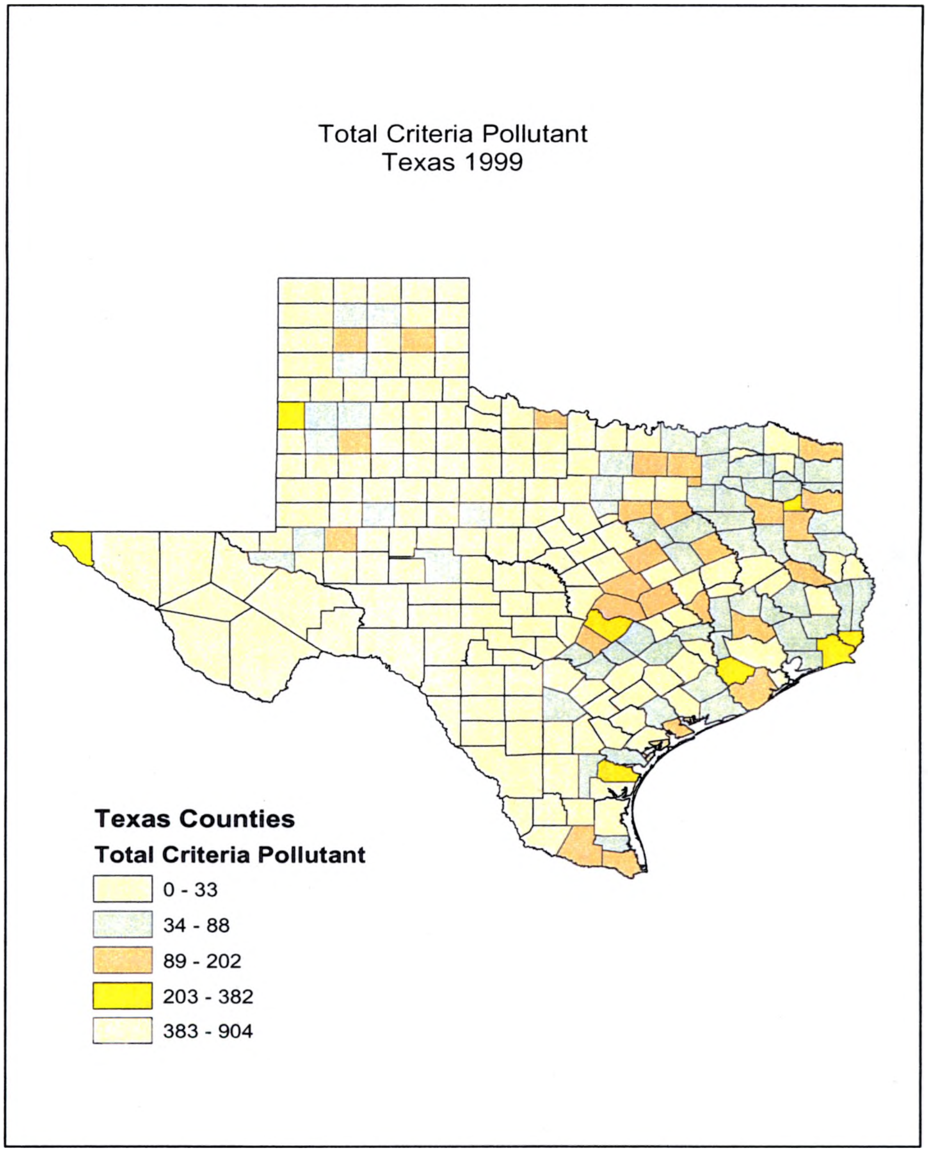


Figure 14 shows that the highest total criteria air pollutants amount was found in Galveston, Harris, Bexar, Tarrant and Dallas in Texas for the year 1999.

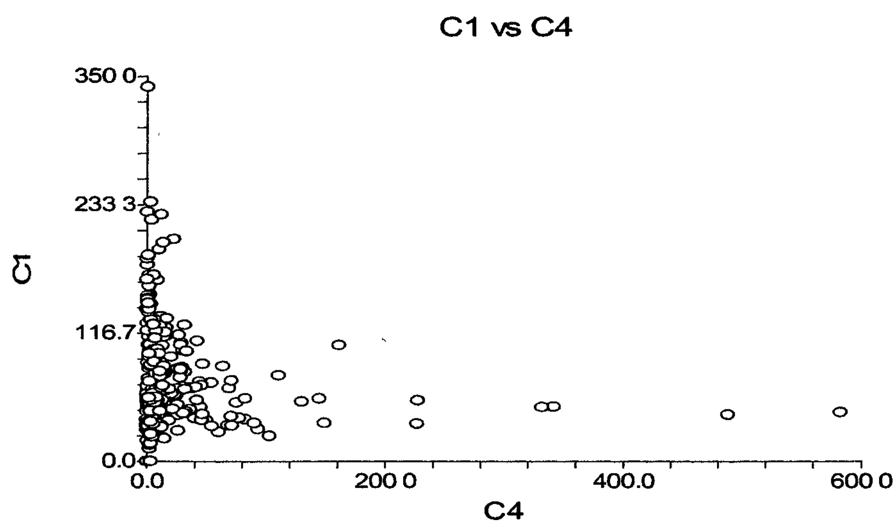
Correlations

1. Correlation between carbon monoxide and stroke death rate.

TABLE 2. Correlation between carbon monoxide and stroke death rate, Texas 1999

	Pearson correlation with Carbon Monoxide	P value
Stoke Death Rate	-0.148	< 0.05

FIGURE 15. Scatter Plot of carbon monoxide and stroke death rate



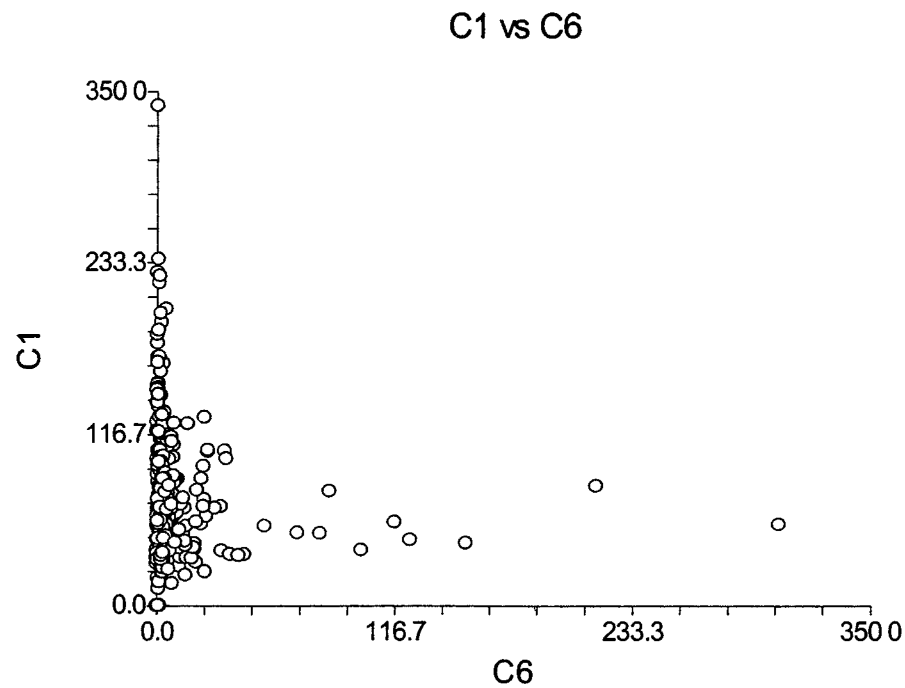
X axis in the above graph has variable C4 which is carbon monoxide and Y axis has C1 which is the stroke death rate. The correlation between carbon monoxide and the stroke death rate was -0.148. The p value was less than 0.05 which implies that there was a significant negative correlation between carbon monoxide and the stroke death rate. The table shows whether the probability levels for variables selected were lesser than 0.05, 0.01 and 0.001. All tables included in this thesis follow the same format. If probability value was greater than .05 it was labeled as a non significant value.

2. Correlation between Nitrogen dioxide and stroke death rate.

TABLE 3. Correlation between nitrogen dioxide and stroke death rate, Texas 1999

	Pearson correlation with nitrogen dioxide	**P value**
Stoke death rate	-0 105	Not Sıgnıficant

FIGURE 16. Scatter plot of nitrogen dioxide and stroke death rate



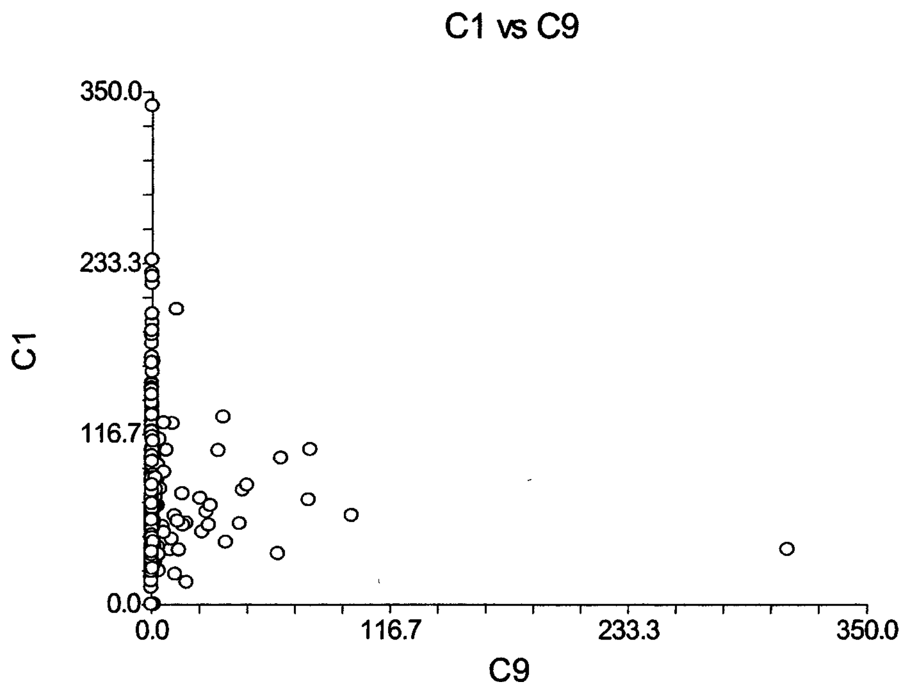
X axis in the above graph has variable C6 which is nitrogen dioxide and Y axis has C1 which is the stroke death rate. The correlation between nitrogen dioxide and the stroke death rate was -0.105. The p value was greater than 0.05 which implies that there was no significant correlation between nitrogen oxide and the stroke death rate.

3. Correlation between sulfur dioxide and stroke death.

TABLE 4. Correlation between sulfur dioxide and stroke death rate, Texas 1999

	Pearson correlation with sulfur dioxide	P value
Stroke death rate	-0.059	Not Significant

FIGURE 17. Scatter plot of sulfur dioxide and stroke death rate



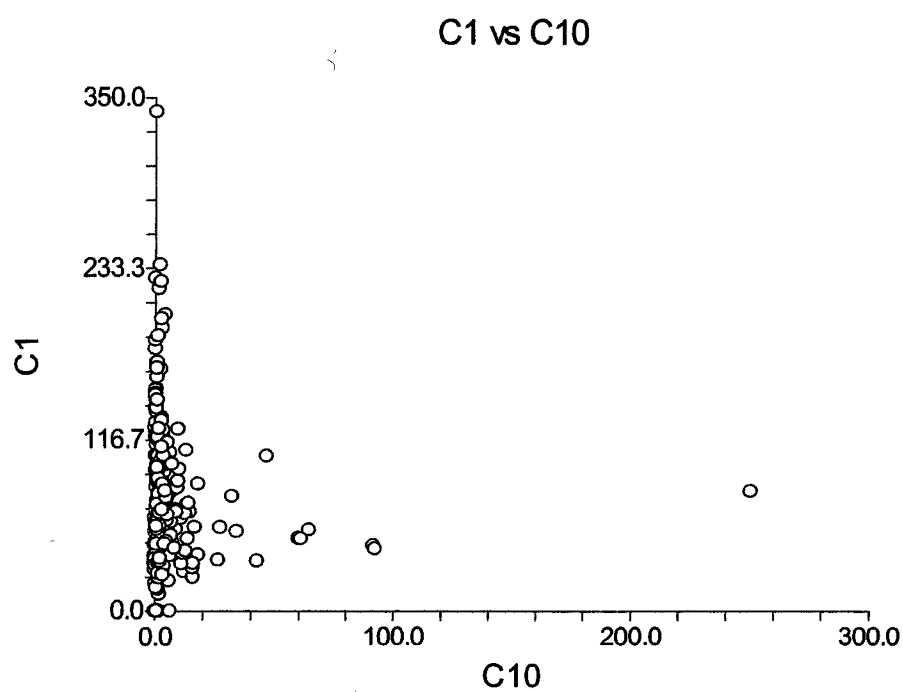
X axis in the above graph has C9 variable which is sulfur dioxide and Y axis has C1 variable which is the stroke death rate. The correlation between sulfur dioxide and the stroke death rate was -0.059. The p value was greater than 0.05 which implies that there was no significant correlation between sulfur dioxide and the stroke death rate.

4. Correlation between volatile organic compounds and stroke death.

TABLE 5. Correlation between volatile organic compounds and stroke death rate, Texas 1999

	Pearson correlation with volatile organic compounds	P value
Stroke death rate	-0.083	Not Significant

FIGURE 18. Scatter plot of volatile organic compounds and stroke death rate



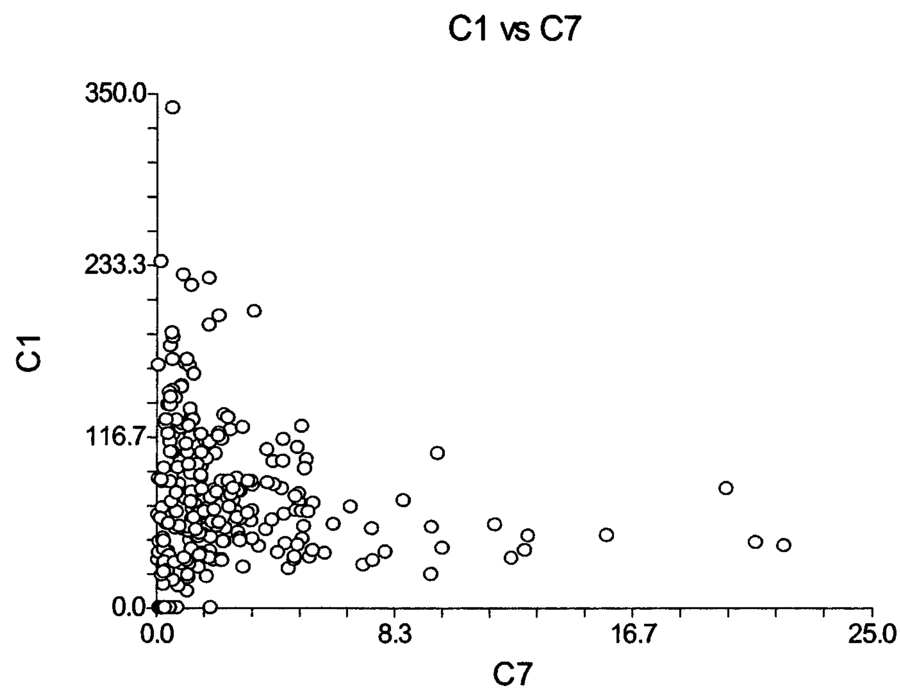
X axis in the above graph has C10 variable which is volatile organic compounds and Y axis has C1 which is the stroke death rate. The correlation between volatile organic compounds and the stroke death rate was -0.083. The p value was greater than 0.05 which implies that there was no significant correlation between the volatile organic compounds and the stroke death rate.

5. Correlation between PM 2.5 and stroke deaths.

TABLE 6. Correlation between PM 2.5 and stroke death rate, Texas 1999

	Pearson correlation with PM 2.5	P value
Stroke Death Rate	-0.142	< 0.05

FIGURE 19. Scatter plot of PM 2.5 and stroke death rate



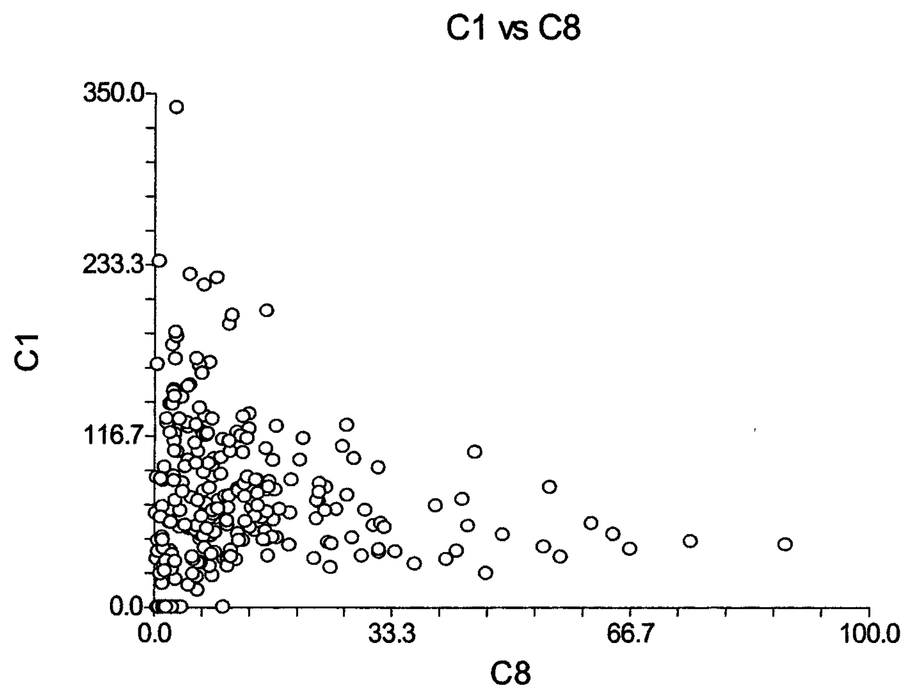
X axis in the above graph has C7 variable which is PM 2.5 and Y axis which has C1 is the stroke death rate. The correlation between PM 2.5 and the stroke death rate was -0.142. The p value was less than 0.05 which implies that there was significant correlation between PM 2.5 and the stroke death rate.

6. Correlation between PM-10 and stroke deaths.

TABLE 7. Correlation between PM 10 and stroke death rate, Texas 1999

	Pearson correlation with PM 10	P value
Stroke Death Rate	-0.150	< 0.05

FIGURE 20. Scatter plot of PM 10 and stroke death rate



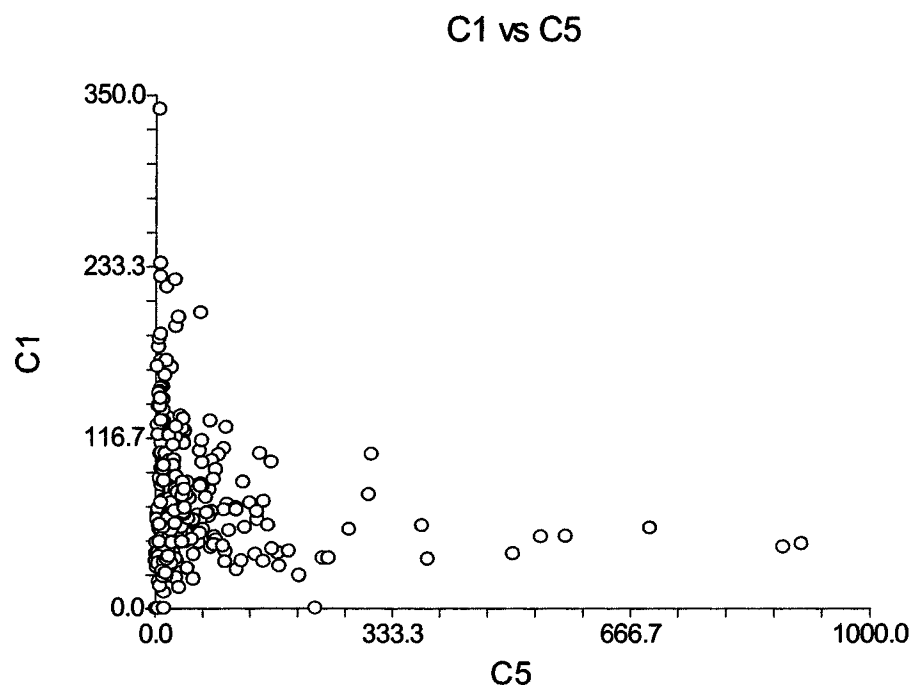
X axis in the above graph has C8 which is PM 10 and Y axis has C1 which is the stroke death rate. The correlation between PM 10 and the stroke death rate was -0.150. The p value was less than 0.05 which implies that there was significant negative correlation between PM 10 and the stroke death rate.

7. Correlation between total criteria air pollutant and stroke deaths.

TABLE 8. Correlation between total criteria air pollutant and stroke death rate, Texas 1999

	Pearson correlation with total criteria air pollutant	P value
Stroke Death Rate	-0.165	< 0.05

FIGURE 21. Scatter plot of total criteria air pollutant and stroke death rate



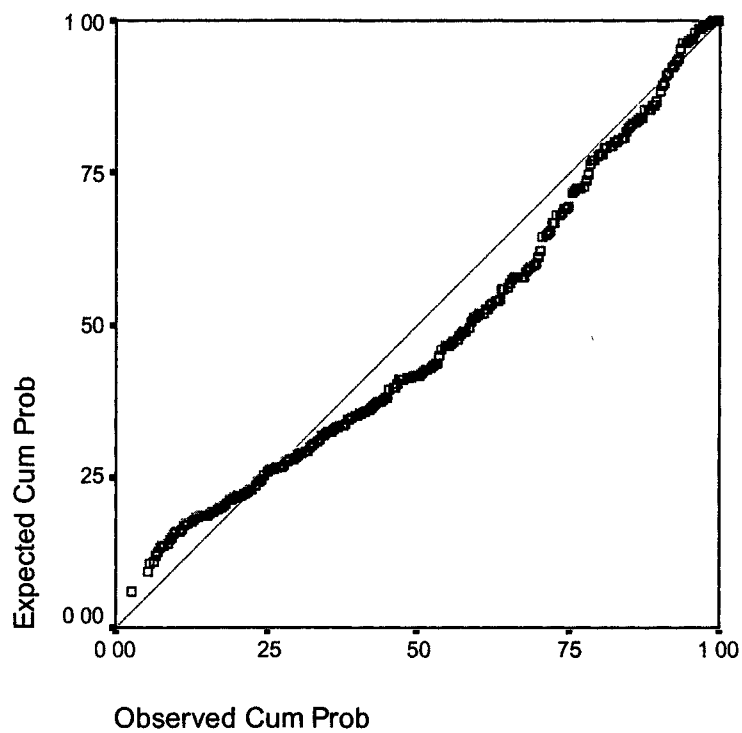
X axis in the above graph has C5 which is criteria air pollutant and Y axis is C1 which is the stroke death rate. The correlation between total criteria air pollutant and the stroke death rate was -0.165. The p value was less than 0.05 which implies that there was a significant negative correlation between total criteria air pollutant and the stroke death rate.

Linear Regression Analysis

Normality assumption of the dependent variable:

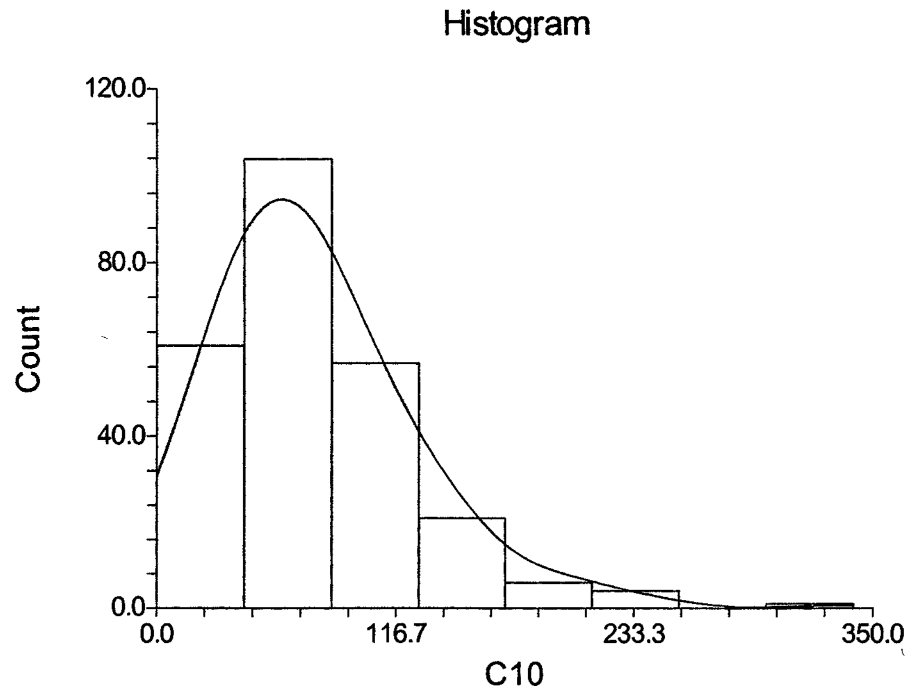
The probability plot was done with SPSS to check the normality assumption of the dependent variable.

FIGURE 22. P-P plot of stroke death rate



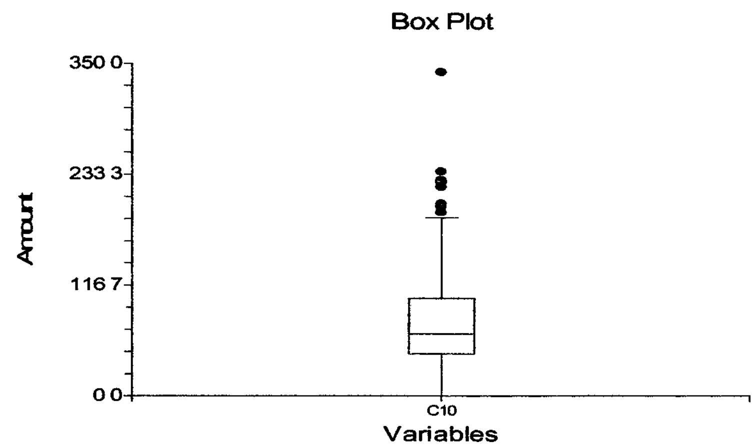
The probability plot above shows that the stroke death rate distribution is slightly skewed. However, it is almost normal in distribution. The linear regression analysis could withstand the slight skewness present with this data. This normality assumption was also verified with the histogram.

FIGURE 23. Histogram of stroke death rate for all the counties



The histogram plot above shows that the stroke death rate is almost normal in distribution. A linear regression analysis was done with the stroke death rate as the dependent variable and carbon monoxide, sulfur dioxide, nitrogen oxide, PM 2.5, PM 10, volatile organic compounds pollutant level and total criteria pollutant as the independent variables. The independent variables were measured in tons per square miles.

FIGURE 24. Box plot of stroke death rate for all the counties



C10 in the above graph is the variable stroke death rate. The box plot above shows the range of stroke death rate for all the Texas counties. The range for the stroke death rate with full data set was 340.43.

TABLE 9. Linear regression equation table for entire Texas data

Independent variable	Regression Coefficient	Standard Error	T-Value	P value	Reject null hypothesis at .05 level	Power
Intercept	83.2806	4.5460	18.320	< 0.05	Yes	1.0000
Volatile organic compounds	-0.4872	0.6084	-0.801	Not Significant	No	0.1255
Carbon monoxide	0.3127	0.3317	0.942	Not Significant	No	0.1555
Criteria pollutant	-0.2924	0.2096	-1.395	Not Significant	No	0.2845
Nitrogen Oxide	0.3859	0.3322	1.162	Not Significant	No	0.2119
PM 2.5	2.8630	10.5397	0.272	Not Significant	No	0.0584
PM 10	-0.4607	1.8555	-0.248	Not Significant	No	0.0570
Sulfur dioxide	0.1981	0.2338	0.847	Not Significant	No	0 1348

Table 9 shows that at alpha level 0.05, all the independent variables were not significant as the p value is greater than 0.05. The linear regression analysis implied that air pollutant levels were not good predictors of stroke death rate. The linear regression model was the following: y (stroke death rate) = 83.2806 - 0.4872 * volatile organic compound + 0.3126 * carbon monoxide - 0.2924 * criteria pollutant + 0.3859 * nitrogen oxide + 2.8630 * PM 2.5 -0.4607 * PM 10 + 0.1981 * sulfur dioxide. The volatile organic compounds, criteria pollutant and PM 10 had negative associations (negative beta) with stroke death rate. However, these negative associations were not significant. Carbon monoxide, nitrogen oxide, PM 2.5 and sulfur dioxide had positive associations (positive beta) with stroke death rate. Again, these positive associations were not significant.

TABLE 10. Analysis of variance for linear regression model

Source	DF	R^2	Sum of square	Mean Square	F-Ratio	Probability Level	Power (5%)
Intercept	1		1492917	1492917			
Model	7	0.0350	21463.23	3066.176	1.274	Not Significant	0.5410
Error	246	0.9650	591914.3	2406.156			
Total (Adjusted)	253	1.0000	613377.6	2424.417			

Table 10 shows that the p value is greater than 0.05. This shows that the predicted model was not a significant model. R square is 3.5% which again shows that the air pollutants were not good predictors of stroke death rate in Texas in 1999.

Linear Regression Model with subsets of the data:

Overlapping Subsets:

Overlapping subsets were subset one and subset two. They are explained below.

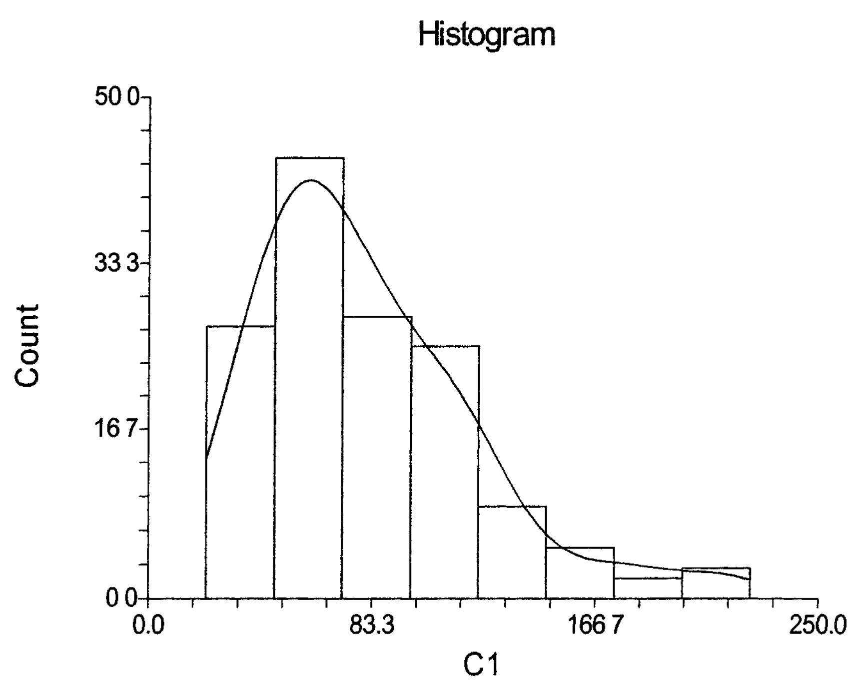
Subset one:

The linear regression model above had a very poor R square and the model was also not significant. In the subset one, the counties which had the stroke death number less than 10 was removed to verify if small stroke deaths were misleading. The stroke death rate was again considered as the dependent variable in this model also. All the pollutants and log of population were the independent variables. The log of population was added as a covariate to verify if population was affecting the air pollution also. A linear regression analysis was performed with hierarchical forward selection. This yielded a better model.

Normality assumptions:

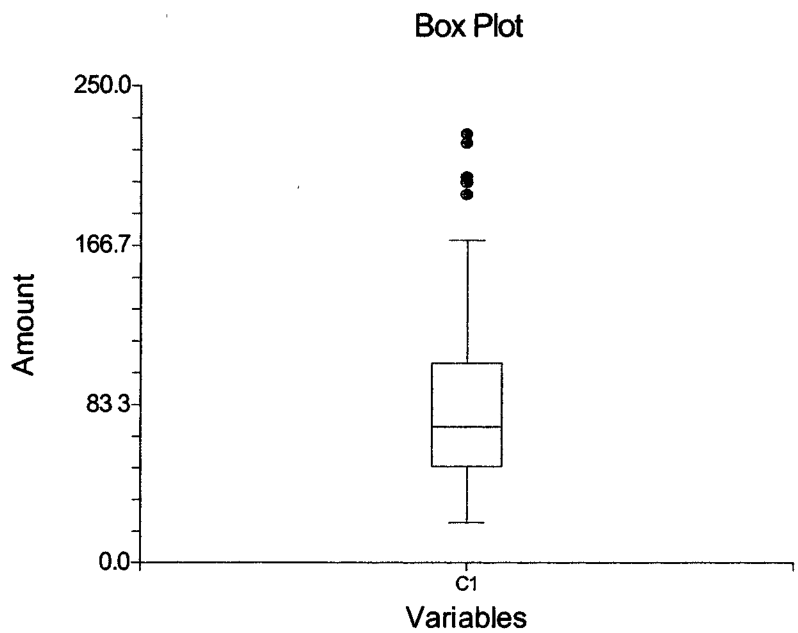
Normality assumption was verified with histogram plot and P-P plot.

FIGURE 25. Histogram of the stroke death rate for subset one



The histogram plot above shows that the stroke death rate was almost normal in distribution.

FIGURE 26. Box plot of the stroke death rate for subset one



The box plot above gives the range of stroke death rate for subset one. The range was 203.18 for the subset one.

TABLE 11. Linear regression equation table for the subset one

Independent Variable	Regression Coefficient b(i)	Standard Error Sb(i)	T-Value to test H0:B(i)=0	Probability Level	Reject H0 at 5%?	Power of Test at 5%
Intercept	401.2816	35.8312	11.1990	< 0.001	Yes	1
Carbon monoxide	0.1911	0.0516	3.7070	< 0.05	Yes	0.9574
Log population	-69.9577	7.9735	-8.7740	< 0.001	Yes	1

The forward selection model yielded carbon monoxide pollutant as a significant variable at p value less than 0.05. The slope associated with the carbon monoxide

The forward selection model yielded carbon monoxide pollutant as a significant variable at p value less than 0.05. The slope associated with the carbon monoxide pollutant was positive which implies that as the carbon monoxide pollutant increased by 1 ton per square mile, stroke death rate would increase by 0.1911. The estimated model was the following: y (stroke death rate) = 401.2816 + 0.1911 * carbon monoxide - 69.9577 * log of population.

TABLE 12. Analysis of variance for the subset one model

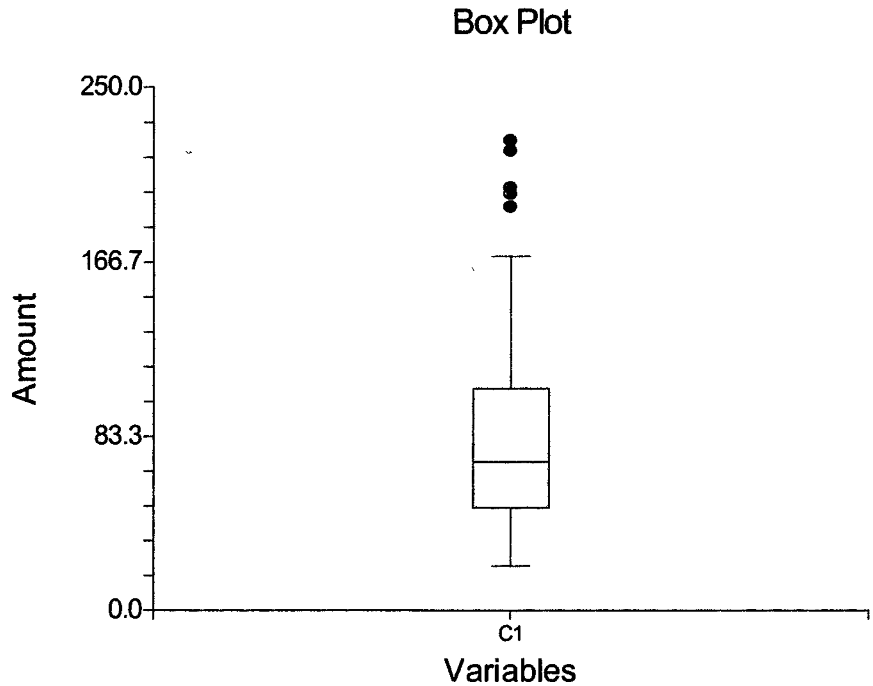
Source	DF	R^2	Sum of Squares	Mean Square	F-Ratio	Probability Level	Power 5%
Intercept	1.0000		950141.7000	950141.7000			
Model	2.0000	0.4115	97924.7400	48962.3700	48.9540	< 0.001	1.0000
Error	140.0000	0.5885	140024.5000	1000.1750			
Total (Adjusted)	142.0000	1.0000	237949.2000	1675.6980			

This shows that this yielded a significant model with p value less than 0.05. The R square was 41% which was highly improved from the previous model which used the entire data.

Subset two:

In the subset two, the counties which had the stroke death number less than 15 was removed to verify if small stroke deaths were misleading. The stroke death rate was the dependent variable. All the pollutants and log of the population were the independent variables. Log of the population was added again as a covariate to verify if population is affecting the air pollution also. A linear regression analysis was performed with hierarchical forward selection. This yielded a better model than the other two models.

FIGURE 27. Box plot of the stroke death rate for subset two



C1 in the above graph is the stroke death rate. The box plot above gives the range of stroke death rate for subset two. Maximum stroke death rate was 224.58 and minimum stroke death rate was 21.4. The range of the stroke death rate was 203.18.

Normality assumption for the stroke death rate:

This was again verified with probability plot and histogram plot. Figure 28 and 29 shows that normality assumption was satisfied even though there was a slight skewness.

FIGURE 28. Probability plot for stroke death rate for subset two

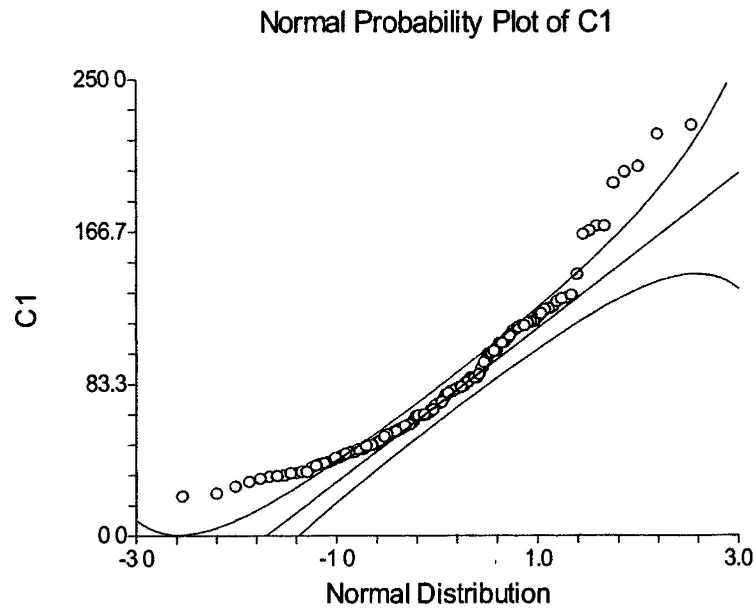
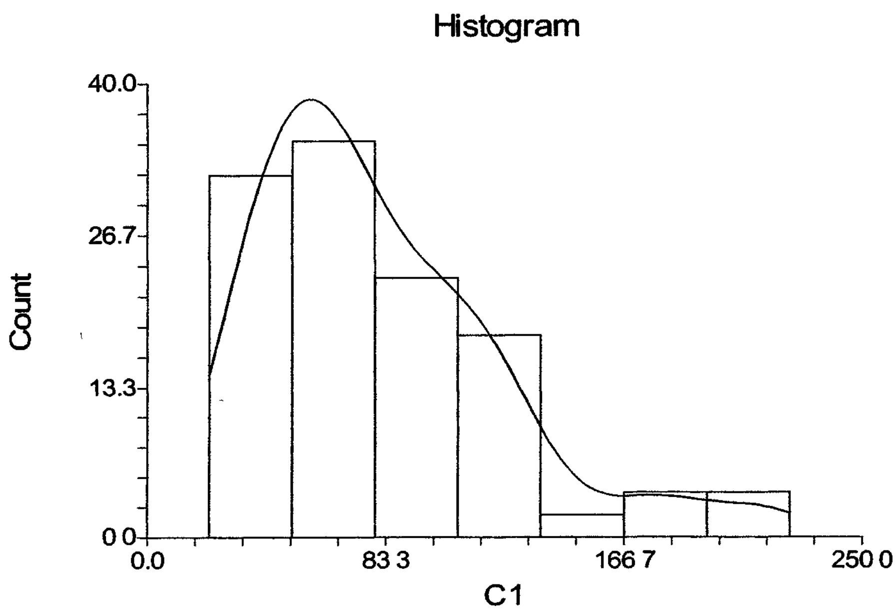


FIGURE 29. Histogram for stroke death rate for subset two



C1 in the above graphs is the stroke death rate. The graphs show a slight skewness in the data for subset two. A linear regression analysis procedure can withstand this type of slight skewness.

C1 in the above graphs is the stroke death rate. The graphs show a slight skewness in the data for subset two. A linear regression analysis procedure can withstand this type of slight skewness.

TABLE 13. Linear Regression Equation Table for the subset two

Independent Variable	Regression Coefficient b(i)	Standard Error Sb(i)	T-Value to test H0:B(i)=0	Probability Level	Reject H0 at 5%?	Power of Test at 5%
Intercept	484.9563	39.1572	12.385	< 0.001	Yes	1
Carbon monoxide	0.2434	0.051	4.772	< 0.001	Yes	0.9972
Log of population	-86.7942	8.5612	-10.138	< 0.001	Yes	1

The forward selection model yielded the carbon monoxide pollutant as a significant variable at p value less than 0.001. The slope associated with the carbon monoxide pollutant was positive which implies that as the carbon monoxide pollutant increased by 1 ton per square mile, stroke death rate would increase by 0.2434. The estimated model was the following: y (stroke death rate) = 484.9563 + 0.2434 * carbon monoxide -86.7942 * log of population.

TABLE 14. Analysis of Variance for the subset two

Source	DF	R^2	Sum of Squares	Mean Square	F-Ratio	Probability Level	Power 5%
Intercept	1		787774.3	787774.3			
Model	2	0.5219	113981.1	56990.57	62.763	< 0.001	1.000
Error	115	0.4781	104423.3	908.0286			
Total (Adjusted)	117	1	218404.4	1866.705			

This shows that the subset two yielded a significant model with p value less than 0.001. The R square was 52% which was highly improved from both subset 1 and overall linear regression model.

Non-Overlapping Subsets:

The first two subsets analyzed were overlapping subsets. To confirm the results derived, non overlapping subsets were analyzed. First non-overlapping subsets were labeled subset three and four.

Subset three analysis results:

The subset three was formed with all the counties in which the stroke deaths were equal to or less than 20. The stroke death rate was the dependent variable. All pollutants level and log of the population were the independent variables. A square root transformation of all the pollutant levels were done and used as independent variables. A linear regression analysis was done with hierarchical forward selection.

Normality assumption for the dependent variable:

Normality assumption was checked with histogram and probability plots.

FIGURE 30. Histogram for subset three

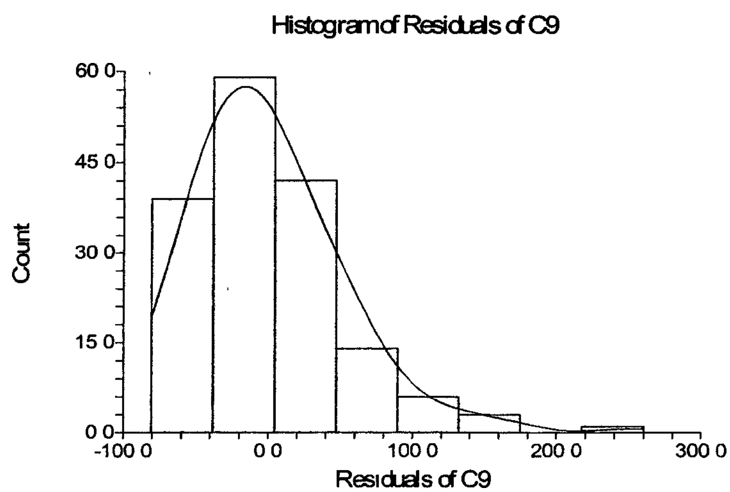
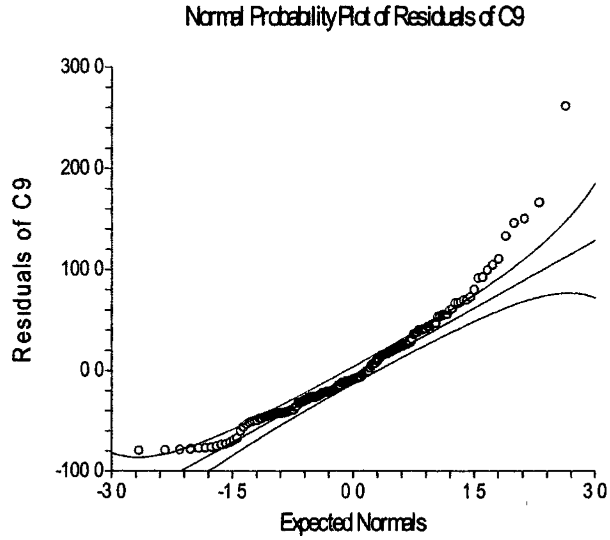


FIGURE 31. Probability plot for subset three



C9 in the above graphs is the stroke death rate. The stroke death rate is almost normal in distribution. The normal probability plot and histogram confirms the normality assumption for the stroke death rate.

TABLE 15. Linear regression equation table for the subset three

Independent Variable	**Regression Coefficient b(i)**	**Standard Error Sb(i)**	**T-Value to test H0:B(i)=0**	**Probability Level**	**Reject H0 at 5%?**	**Power of Test at 5%**
Intercept	77.4383	9.674	8.005	< 0.05	Yes	1
Square root of total criteria pollutant	-5.0273	2.2824	-2.203	< 0.05	Yes	0.5908
Square root of PM 10	8.2435	5.2978	1.556	Not Significant	No	0.3399

The total criteria pollutant was a significant variable at p value less than 0.05. The predicted model was the following: y (stroke death rate) = 77.4383 - 5.0273 * square root of total criteria pollutant + 8.2435 * square root of PM10. This model shows that as the total criteria pollutant increases by one unit, the stroke death rate would decrease by

5.0273 units. PM 10 was not a significant factor. This model shows that as the PM 10 pollutant increases by one unit, the stroke death rate would increase by 8.2435 units.

TABLE 16. Analysis of variance for the subset three

Source	DF	R^2	Sum of Squares	Mean Square	F-Ratio	Probability Level	Power 5%
Intercept	1		933381.4	933381.4			
Model	2	0 0293	13587.64	6793.822	2.427	Not Sıgnıficant	0.4836
Error	161	0.9707	450602.4	2798.773			
Total (Adjusted)	163	1	464190.1	2847.792			

The table above shows that R square of the model is 2.9%. This subset three did not yield a sıgnıficant model.

Subset four analysis results:

Subset four was formed with all counties in which the stroke deaths were greater than 20. The stroke death rate was the dependent variable. All pollutants level and log of the population were independent variables. A linear regression analysis was done with hierarchical forward selection.

Normality assumption for the dependent variable:

FIGURE 32. Histogram for subset four

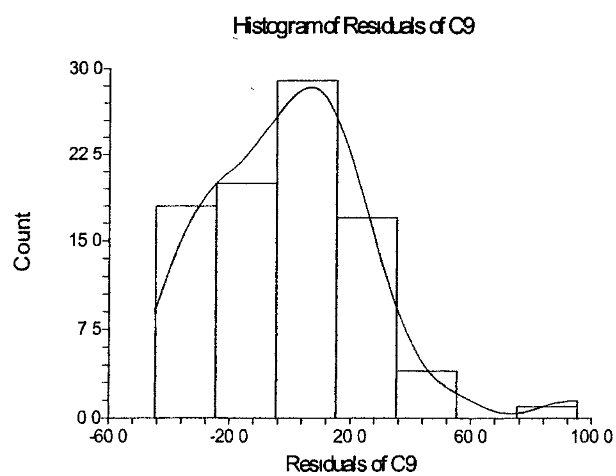
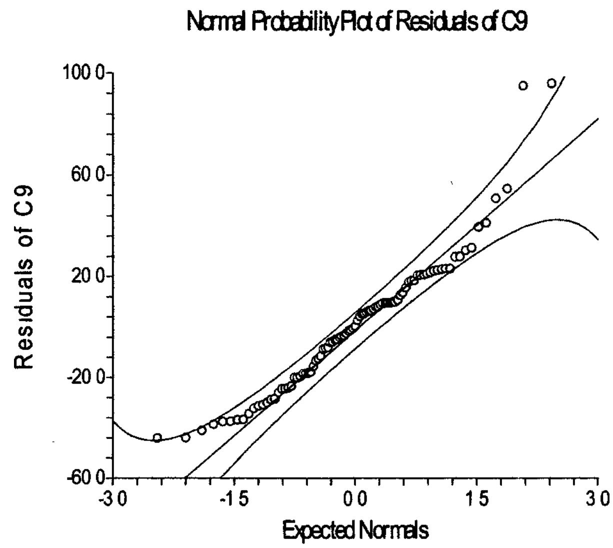
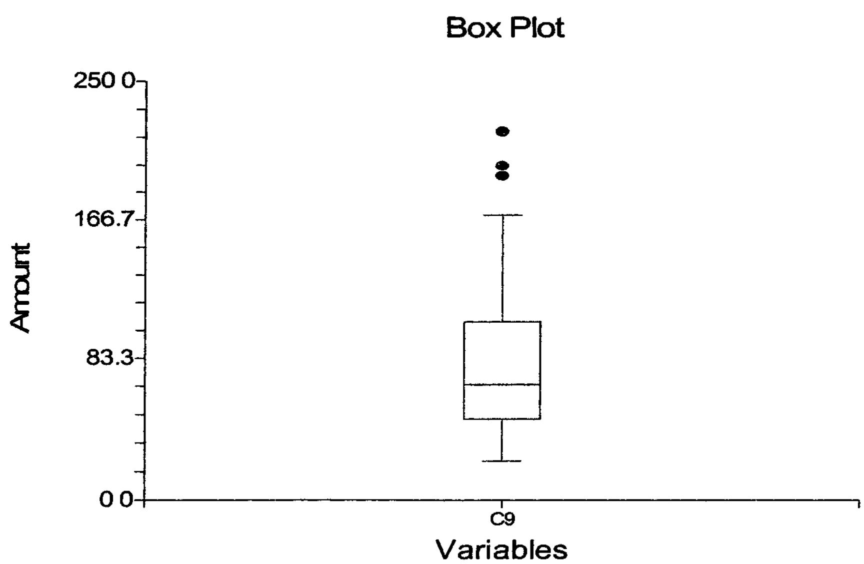


FIGURE 33. Probability plot for subset four



C9 in the above graphs is the stroke death rate. The stroke death rate was almost normal in distribution. The normal probability plot and histogram confirms the normality assumption for the stroke death rate.

FIGURE 34. Box plot for subset four



C9 in the above graph is the stroke death rate. The box plot above gives the range of stroke death rate for subset four.

TABLE 17. Linear regression equation table for the subset four

Independent Variable	Regression Coefficient b(i)	Standard Error Sb(i)	T-Value to test H0:B(i)=0	Probability Level	Reject H0 at 5%?	Power of Test at 5%
Intercept	499.315	42.5565	11.733	< 0.001	Yes	1
Volatile organic compounds	1.201	0.2594	4.63	< 0.001	Yes	0.9956
Log of population	-88.524	9.1325	-9.693	< 0.001	Yes	1

The volatile organic compounds and log of the population were the significant variables. The predicted model was the following: y (stroke death rate) = 499.315 + 1.201 * volatile organic compounds - 88.524 * log of population. This model showed that as the volatile organic compound pollutant increased by one unit, the stroke death rate would increase by 1.201 units.

TABLE 18. Analysis of variance for the subset four

Source	DF	R^2	Sum of Squares	Mean Square	F-Ratio	Probability Level	Power 5%
Intercept	1		560229.4	560229.4			
Model	2	0.5707	84747.04	42373.52	57.831	< 0.001	1
Error	87	0.4293	63746.32	732.7164			
Total (Adjusted)	89	1	148493.4	1668.465			

The R square was 57.07% and the model was significant as the p value was less than 0.001. This again verified that counties with higher stroke deaths models were better than the model where stroke deaths were smaller.

Second non-overlapping subsets:

Second non-overlapping subsets had subset five and subset six. Subset five was formed with all counties in which stroke deaths were equal to or less than 25. Subset six was formed with all counties in which stroke deaths were greater than 25.

Subset five analysis results:

The stroke death rate was the dependent variable. The square root of all the pollutants and log of the population were the independent variables. A linear regression analysis was done with hierarchical forward selection.

Normality Assumption of stroke death rate:

Histogram and P-P plots below show the normality assumption as follows.

FIGURE 35.Histogram of subset five

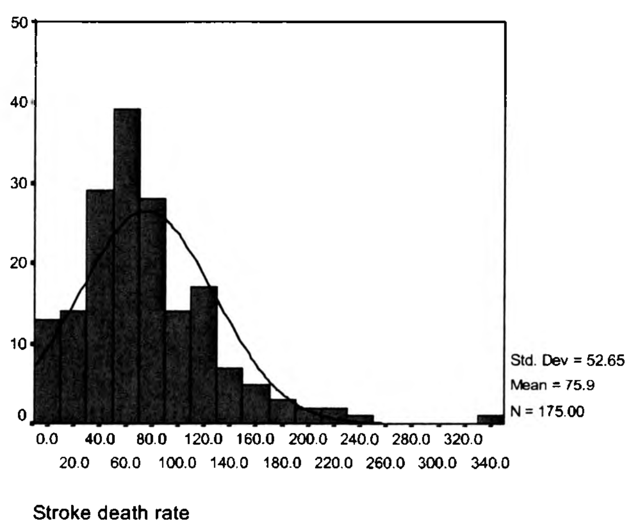
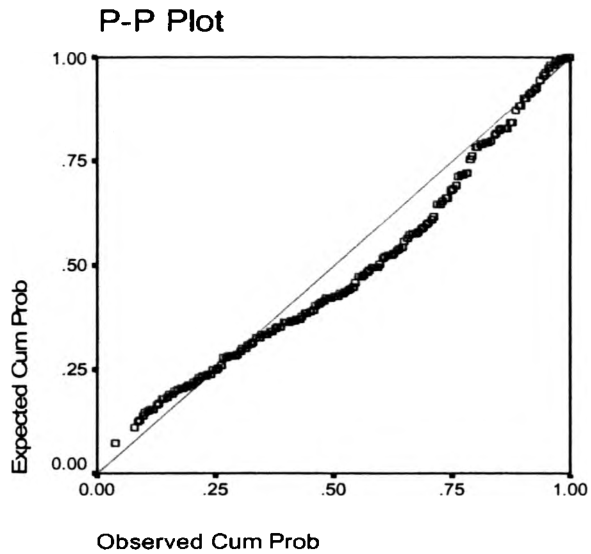
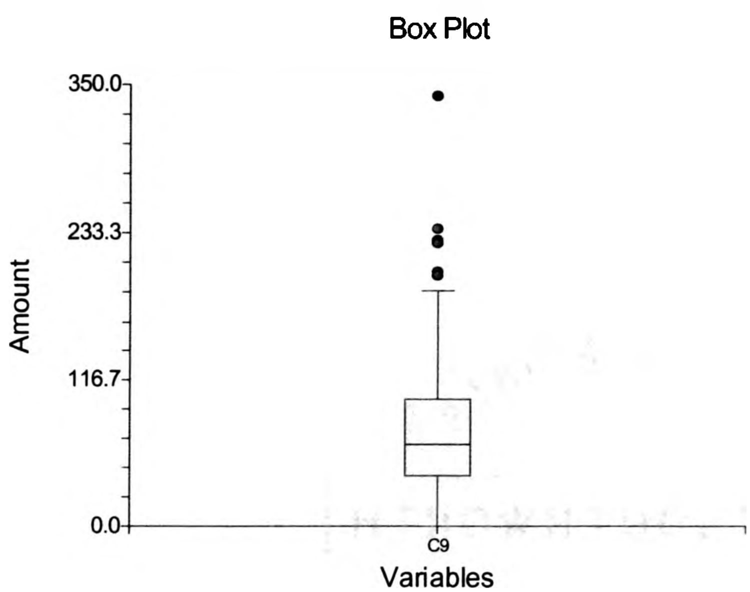


FIGURE 36.Probability plot for subset five



The above normal probability plot and histogram confirms that the normality assumption for the stroke death rate was satisfied.

FIGURE 37. Box plot for subset five



C9 in the above graph is the stroke death rate. The box plot shows the range of stroke death rate. The range for subset five was 340.43.

TABLE 19. Linear regression equation table for the subset five

Independent Variable	Regression Coefficient b(i)	Standard Error Sb(i)	T-Value to test H0:B(i)=0	Probability Level	Reject H0 at 5%?	Power of Test at 5%
Intercept	142.3914	36.5303	3.898	< 0.001	Yes	0.9723
Log of population	-19.9229	10.5789	-1.883	Not Significant	No	0.4653
Square root of PM 10	4.8018	4.3268	1.11	Not Significant	No	0.197

All the independent variables were insignificant in the subset five. The predicted model was y (stroke death rate) = 142.3914 - 19.9229 * log of population + 4.8018 * square root of PM 10. The PM 10 pollutant level had positive association with the stroke death rate which was not significant.

TABLE 20. Analysis of variance for the subset five

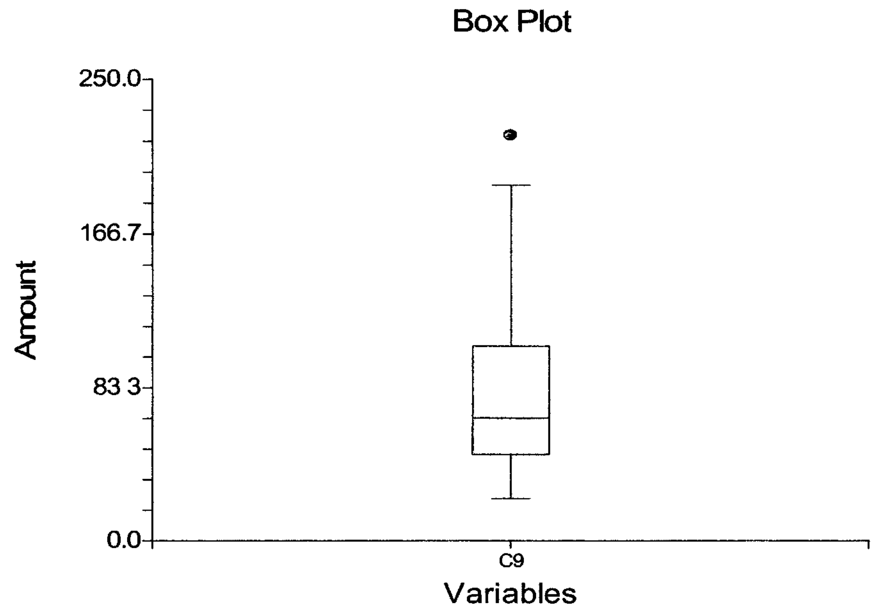
Source	DF	R^2	Sum of Squares	Mean Square	F-Ratio	Probability Level	Power 5%
Intercept	1		1007803	1007803			
Model	2	0.0202	9745.683	4872.841	1.773	Not Significant	0.3676
Error	172	0.9798	472623.7	2747.812			
Total (Adjusted)	174	1	482369.3	2772.238			

The R square was 2% and this model was not significant. This again shows that the subset with lesser stroke deaths did not yield a good model.

Subset six analysis results:

The stroke death rate was the dependent variable. All the pollutants level and log of the population were independent variables. A linear regression analysis was done with hierarchical forward selection.

FIGURE 38. Box plot for subset six



The box plot above gives the range of stroke death rate for subset six. The range was 196.73 for the subset six.

Normality assumption of stroke death rate:

FIGURE 39. Histogram for subset six

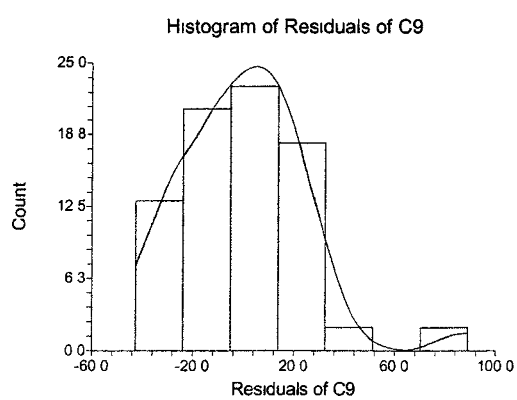
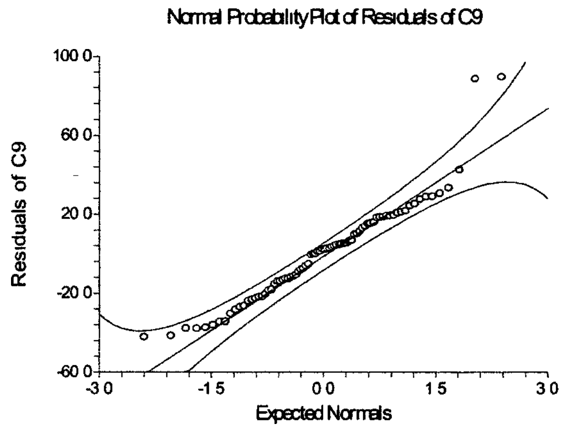


FIGURE 40. Probability plot for subset six



C9 in the above graphs is the stroke death rate. The normal probability plot and histogram above confirmed the normality assumption for the stroke death rate.

TABLE 21. Linear regression equation table for the subset six

Independent Variable	Regression Coefficient b(i)	Standard Error Sb(i)	T-Value to test H0:B(i)=0	Probability Level	Reject H0 at 5%?	Power of Test at 5%
Intercept	531.862	44.1284	12.053	< 0.05	Yes	1
Carbon monoxide	0.2269	0.047	4.826	< 0.05	Yes	0.9975
Log of population	-94.1856	9.3427	-10.081	< 0.05	Yes	1.000

The above table shows that the significant variables were carbon monoxide pollutant and log of the population. The predicted model was the following: y (stroke death rate) = 531.862 + 0.2269 * carbon monoxide - 94.1856 * log of population. This shows that if carbon monoxide level increased by one unit, the stroke death rate would increase by 0.2269 units.

TABLE 22. Analysis of variance for the subset six

Source	DF	R^2	Sum of Squares	Mean Square	F-Ratio	Probability Level	Power 5%
Intercept	1		485454.4	485454.4			
Model	2	0.6286	82135.24	41067.62	64.311	< 0.05	1
Error	76	0.3714	48532.04	638.5795			
Total (Adjusted)	78	1	130667.3	1675.222			

The R square for subset six was 62.86% and it was a significant model as p value was less than .05. This again shows that for the counties with higher stroke deaths than 25, carbon monoxide was a significant pollutant which was also confirmed by using the overlapping model.

Poisson Regression Analysis Results

A Poisson regression analysis was performed with entire data. Number of stroke deaths was considered as the dependent variable. Since Poisson regression needs count variable as the dependent variable, number of stroke deaths was considered as the dependent variable. Carbon monoxide, sulfur dioxide, nitrogen oxide, PM 2.5, PM 10, volatile organic compounds pollutant level and total criteria pollutant were the independent variables which were measured in tons per square miles. Log of the population was also considered as an independent variable which acted as a covariate. This model was analyzed with the entire data. Hierarchical forward selection was used to select the regression variables in the Poisson regression model.

TABLE 23. Poisson regression coefficient

Independent Variable	Regression Coefficient (B)	Standard Error	Wald's Chi^2 (Ho:B=0)	Probability Level	Lower 95.0% Confidence Limit	Upper 95.0% Confidence Limit
Intercept	-4.81204	0.11294	1815.46	< 0.001	-5.03339	-4.59069
Volatile organic compounds	0.00328	0.0008	16.84	< 0.001	0.00171	0.00485
Log population	1.75923	0.0237	5511.78	< 0.001	1.71279	1.80568
Carbon monoxide	0.00054	0.00016	11.18	< 0.001	0.00022	0.00086
Dispersion Phi		4.5706				

Table 23 shows that the significant variables were the carbon monoxide pollutant, log of the population and the volatile organic compounds pollutant.

TABLE 24. Rate ratio

Independent Variable	Regression Coefficient (B)	Rate Ratio [Exp(B)]	Lower 95.0% Confidence Limit	Upper 95.0% Confidence Limit
Intercept	-4.81204	0.00813	0.00652	0.01015
Volatile organic compounds	0.00328	1.00328	1.00171	1.00486
Log of population	1.75923	5.80798	5.5444	6.08408
Carbon monoxide	0.00054	1.00054	1.00022	1.00086

From table 24, we could get the predicted model. The estimated model was y (stroke death) = Exponential (-4.812 + 0.00328 * Volatile organic compounds + 1.75923 * Log of population + .00054 * Carbon monoxide). The slope associated with carbon monoxide was positive which implied that as carbon monoxide increased by one ton per square mile, stroke death would increase at the rate of 1.00054. The slope associated with the volatile organic compounds was positive which implied that as volatile organic compounds increased by one ton per square mile, stroke death would increase at the rate 1.00328.

TABLE 25. Poisson regression model summary

Number of terms	**Log Likelihood**	**R-Squared**
1	-14029.72	0
2	-1187.205	0.9524
3	-1122.533	0.9572
4	-1116.536	0.9576

Table 25 shows that the R square was very high. The R square in the above table is around 95% which implies that it was a good model to predict.

CHAPTER IV

DISCUSSION

Hong et al (2002) showed that there was an increased relative risk of 1.06 (95% CI, 1.02 to 1.09) for carbon monoxide with a one-day lag. This model developed was controlled for time trends and meteorological influence like temperature, humidity and atmospheric pressures. Kan et al (2003) also used a time series approach to see an association between air pollution and daily stroke mortality in Shanghai in which the relative risk was not significant for sulfur dioxide. This study examined the relation between air pollution and stroke death rate in Texas in 1999. Air pollutants considered in this study were carbon monoxide, sulfur dioxide, nitrogen oxide, PM 2.5, PM 10, volatile organic compounds and total criteria air pollutant. Different types of regression analysis were performed for the Texas 1999 data. In model one, linear regression analysis was done with stroke death rate as a dependent variable and carbon monoxide, sulfur dioxide, nitrogen oxide, PM 2.5, PM 10, volatile organic compounds pollutant level and total criteria pollutant as the independent variables in all the Texas counties. In the first subset, the stroke death rate is again the dependent variable. However, the counties which had the stroke death number less than 10 were removed to verify if small stroke deaths were misleading. Alternatively, another model using all the pollutants and log of population as the independent variable was analyzed. A linear regression analysis was performed with hierarchical forward selection. In the second subset, stroke death rate was the dependent variable. The counties which had stroke death number less than 15 were removed to verify if small stroke deaths were misleading. A third model using all the pollutants and log of population as the independent variables was attempted. A linear regression analysis was performed with the hierarchical forward selection. Subset one and subset two were overlapping subsets. To confirm the results, non-overlapping subsets were also analyzed. First non-overlapping subsets were subset three and subset four. Subset three was formed with all counties in which the stroke deaths were equal to or less than 20. Subset four was formed with all counties in which the stroke deaths were greater than 20. Second non-overlapping subsets had subset five and subset six. Subset five was formed with all counties in which stroke deaths were equal to or less than 25. Subset six was formed with all the counties in which stroke deaths were greater than 25. A Poisson regression analysis was performed with entire data. Number of stroke deaths was considered as the dependent variable. Carbon monoxide, sulfur dioxide, nitrogen oxide, PM 2.5, PM 10, volatile organic compounds pollutant level and total criteria pollutant were the independent variables. Log of the population was also considered as an independent variable which acted as a covariate. This model was analyzed with the entire data. Hierarchical forward selection was used to select the regression variables in the Poisson regression model.

Findings and Relation to the Previous Studies

In the linear regression analysis performed in this thesis, carbon monoxide pollutant was a significant variable that affected stroke death rate in subset one, two and six which is consistent with the previous studies. In the subset four, volatile organic compound was significant. In the Poisson regression analysis conducted in this thesis, carbon monoxide and volatile organic compound pollutants were found as significant variables. However, all the other variables were not significant in both linear and Poisson regression analysis.

Study Strengths and Limitation

The results of this study confirm the previous study that carbon monoxide pollutant increases the risk of stroke death rate. Also, this data gives more knowledge on Texas based data. Two models used entire Texas data. Six subsets of full data were also analyzed. Further more, population was added as a covariate in seven models. The variables included in this study such as volatile organic compound pollutant and total criteria pollutant were not considered in most of the previous studies which would be a major strength of this study.

Nevertheless, this study also has some limitations. In linear regression model with the entire data, a significant model was not established. Only the subsets had significant models. Also, Poisson regression model did not have a good fit. More significant models should be created in new studies. Data used were for the year 1999. This should be done with the latest data in future studies. Models were not controlled for time trends and spatial trends which should be considered in further studies. This study did not consider ozone level. This study did not consider other potential confounders such as ethnicity, race and age.

Conclusion

According to the results of this study, carbon monoxide pollutant level was a significant factor in three linear regression models and Poisson regression model.

Therefore, higher carbon monoxide pollutant level increases the risk of stroke death rate. Volatile organic compounds pollution level was a significant factor with Poisson regression model and in one subset model. Further studies should be done to develop models on the association of volatile organic compounds and stroke death.

Health Implications

Four counties with high carbon monoxide pollutant were Bexar, Dallas, Harris and Tarrant. The Dallas had the highest carbon monoxide pollutant amount in Texas for the year 1999. According to the results, higher carbon monoxide pollutant level is related with stroke death rate. A formation of carboxyhemoglobin in the blood due to carbon monoxide inhibits oxygen intake and it could be the reason for its health impact. The carbon monoxide emission should be controlled to reduce the risk of increasing stroke death. First, an emission control should start with the counties like Bexar, Dallas, Harris and Tarrant where the highest carbon monoxide levels were found. Motor vehicle exhaust contributes about 56 percent of all carbon monoxide emissions. These counties should have higher standards of motor vehicle emissions control. Some policies for controlling the carbon monoxide emission are as follows: 1) Implementing destruction programs to vehicles older than 1980, 2) Have strict rules to change the vehicle components that are high emitters of pollutants while testing, 3) Promote car pooling by having many incentives like lower tolls, 4) Gas filters should be used to decrease carbon monoxide levels, 5) Convert the bus system to natural gas or a cleaner fuel, 6) Educate employees to combine trips to avoid unnecessary trips, 7) More emphasis should be given on reporting polluting vehicles, 8) Oxygenated gasoline program should be emphasized for all government vehicles because wintertime carbon monoxide emissions is decreased through the use of cleaner burning gasoline, 9) Emphasize the public about winter season use of engine-block heaters which helps motorists warm their car engine quicker, thus reducing the amount of emissions. Apart from these, there are a few indoor sources for carbon monoxide pollutants including gas space heaters, leaking chimneys and furnaces, gas water heaters, wood stoves, fireplaces, gas stoves, generators, and other gasoline powered equipment. Some policies to control these indoor emissions are educating the public to install and use an exhaust fans vented to outdoors over gas stoves, educating about proper sized wood stoves that are certified to meet EPA emission standards, educating the public on advantages of a vented space heater, and implementing mandatory trained professional inspection for central heating systems (furnaces, flues, and chimneys) every year.

REFERENCES

Hong, Y., Lee, J., Kim, Ho., Kwon, H., Hee Ha, E., Schwartz, J., & Christiani, D. (2002). Air Pollution, a New Risk Factor in Ischemic Stroke Mortality. Environmental Health Prospectives, 110, 2

Kan, H., Jia, J., & Chen, B., (2003). Acute stroke mortality and air pollution: new evidence from Shangai, China. Journal of occupational health, 45:321-323

Maheswaran, R., & Elliott, P. (2003). Stroke Mortality Associated With Living Near Main Roads in England and Wales, A Geographical Study. American Heart Association Journal, 34, 2776

Pope, C., Burnet, R., Thun, M., Calle, E., Krewski, D., Ito, K., & Thurston, G. (2002). Lung Cancer, Cardiopulmonary Mortality, and Long-term Exposure to fine particulate Air Pollution. Journal of American Medical Association, 287, 9

Samet, J., Dominici, F., Curriero, F., Coursac, I., & Zeger, S. (2000). Fine Particulate Air Pollution and Mortality in 20 U.S Cities, 1987 –1994. The New England Journal of Medicine, 343, 24

Sunyer, J., Ballester, F., Tertre, AL., Atkinson, R., Ayres, JG., Forastiere, F., Forsberg, B., Vonk, JM., Bisanti, L., Tenias, JM., Medina, S., Schwartz, J., & Katsouyanni, K. (2003). The association of daily sulfur dioxide air pollution levels with hospital admissions for cardiovascular disease in Europe. European Heart Journal, 24(8): 752-60

http://edugreen.teri.res.in/explore/air/major.htm

http://www.casadata.org/pollutants/vol_compounds.asp

http://www.cdc.gov/cvh/library/fs_stroke.htm http://www.irishhealth.com/?level=4&con=8

http://www.irishhealth.com/content/image/11/Image1.gif

http://www.tdh.state.tx.us/wellness/facts/facts03.pdf

http://www.pca.state.mn.us/air/criteria-emissiongraphs.html

http://www.tnrcc.state.tx.us/exec/sbea/air/initiatives.html

VITA

Sowmya was born in Thanjavur, India, on May 25, 1978, the daughter of Mohandoss and Bhanumathy. After completing her work at Sacred Heart School, Thanjavur, in 1996 she entered Saveetha Dental College in Chennai, India. She received a bachelor of dental surgery degree from M.G.R University in 2001. In 2003 she entered the Graduate College of Texas State University-San Marcos.

Permanent Address: 29/14 Second Main Street

Indian Bank Colony

Chennai-600053

This thesis was typed by Sowmya Anand.