THE USE OF ELECTRONIC MEDICAL RECORDS BASED ON A PHYSICIAN DIAGNOSIS OF ASTHMA FOR COUNTY WIDE ASTHMA SURVEILLANCE

by

Jo Ann Marie Glad

BSN, La Roche College, 1992

MPH, Graduate School of Public Health, University of Pittsburgh, 1996

Submitted to the Graduate Faculty of the

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Public Health

University of Pittsburgh

2008

UNIVERSITY OF PITTSBURGH

Graduate School of Public Health

This dissertation was presented

by

Jo Ann Marie Glad

It was defended on

December 9, 2008

and approved by

Dissertation Director: Evelyn Talbott, DrPH, MPH
Professor, Department of Epidemiology
Graduate School of Public Health, University of Pittsburgh

Joseph Schwerha, MD, MPH
Professor, Department of Biostatistics
Graduate School of Public Health, University of Pittsburgh

Joseph Costantino, DrPH
Professor, Department of Environmental Occupational Health
Graduate School of Public Health, University of Pittsburgh

Gerald Barron, MPH
Associate Professor, Department of Health Policy and Management
Graduate School of Public Health, University of Pittsburgh

Luann Brink, PhD
Senior Research Specialist, Department of Epidemiology
Graduate School of Public Health, University of Pittsburgh

Melissa Saul
Adjunct Assistant Professor, Department of Health Information Management,
School of Rehabilitation Sciences, University of Pittsburgh

Copyright © by Jo Ann Marie Glad, BSN, MPH

2008

Evelyn Talbott, DrPH

THE USE OF ELECTRONIC MEDICAL RECORDS BASED ON A PHYSICIAN DIAGNOSIS OF ASTHMA FOR COUNTY WIDE ASTHMA SURVEILLANCE

Jo Ann Marie Glad, DrPH

University of Pittsburgh, 2008

Allegheny County (AC) has limited information on asthma morbidity. In order to improve upon the sensitivity of asthma, a cross sectional study from January 1, 2002 through December 31, 2005 was conducted to determine whether the data received for emergency room visits from a large regional medical center might be a good predictor for quantifying asthma cases for surveillance. An electronic medical record (EMR) abstract using the Council for State and Territorial Epidemiology (CSTE) Asthma Surveillance case definition of an ICD 9 coded physician diagnosis for primary and secondary asthma (n= 18,284), and primary asthma (n = 5,100) were used to define asthma. The analysis used data from a subset of six hospitals from a large regional medical center covering approximately 60% of adult ED visits in AC that use electronic data for reporting. A secondary analysis of the physician diagnosed primary asthma cases (n= 180) was applied against the CSTE Clinical and Laboratory case definition. Statistical software was used to validate these data abstracted from the EMR. Once these data were validated for accuracy, a fourth dataset of any primary asthma emergency room visits (n= 10,183) were used to test the relationship between asthma morbidity and exposure to ozone.

Recent studies have linked asthma hospitalizations in several cities to ozone action days. However, data on the effects of ozone as they relate to asthma emergency room (ER) visits have not been well studied. Electronic medical records from the six hospitals representing the large metropolitan medical center in Allegheny County, PA were obtained on individuals with asthma

based on the ICD-9 discharge diagnosis of (493.0-493.9) for the respective time. Data on ozone, $PM_{2.5}$, and temperature were obtained for same period. A case crossover methodology using conditional logistic regression as the statistical estimator was conducted to assess the relationship between levels of ozone and $PM_{2.5}$ and increases in asthma ER visits. A time stratified sampling strategy was employed assuming a 3:1 case-control ratio.

A total of 6,979 individuals were included in the study, with a mean age of 39.25 ±21.0. The mean ozone exposure for this period was 40.6 ppb (range: 0-126). The effect estimates for year-round data was greatest for a 2-day lag adjusted for temperature (OR= 1.02 (95% CI= 1.01-1.04) ($p<.05$). For each 10-ppb increase in 24-hour maximum ozone, a 2% increase was noted in asthma ER visits. These results reflect the public health significance of ozone on asthma morbidity and indicate a vital source of information that can be used for environmental public health tracking.

TABLE OF CONTENTS

PREFACE xiv

1.0 CHAPTER 1- INTRODUCTION AND SIGNIFICANCE 1

1.1 STATEMENT OF RESEARCH PROBLEM AND SPECIFIC AIMS 3

1.2 DEFINITION AND CLINICAL FEATURES 5

1.2.1 Historical aspect 6

1.2.2 Diagnosing and managing asthma 6

2.0 CHAPTER 2- EPIDEMIOLOGY 9

2.1 ASTHMA PREVALENCE AND BURDEN OF DISEASE 16

2.1.1 Asthma prevalence 16

2.1.2 Burden of disease 21

2.1.3 Economic burden 26

2.2 RISK FACTORS 29

2.3 RELEVANT RESEARCH ON AIR POLLUTION 33

2.4 CASE CROSSOVER DESIGN 35

2.5 SUMMARY OF LITERATURE REVIEW 38

3.0 CHAPTER 3- CRITICAL REVIEW OF LITERATURE 48

3.1 SURVEILLANCE 48

3.1.1 Electronic medical records 50

3.1.2 Classification of clinical free text data 57

3.1.3 Value of the ICD 9 Code 62

3.1.4 Diagnostic accuracy 65

3.1.5 Healthcare utilization 69

3.2 PRELIMINARY ANALYSIS 72

3.3 RESEARCH METHODS AND DESIGN 76

3.3.1 Sample size and power estimate 97

3.4 STATISTICAL ANALYSIS AND RESULTS 99

3.4.1 Populations descriptive analysis 99

3.4.1.1 Population A 99

3.4.1.2 Population B 111

3.4.1.3 Population C 120

3.4.2 Comparison ICD-9 Coded chief complaint: Admitting vs. Hospital discharge diagnosis for asthma 126

3.5 VALIDATION STUDY RESULTS 128

3.5.1 Electronic Medical Record: Composition and Completeness 128

3.5.1.1 Electronic Medical Record: Descriptive analysis 130

3.6 ASTHMA SELECTED INDICATOR RESULTS 134

3.6.1 Manual Review 134

3.6.2 Manual vs. Automation Review 137

3.7 CLASSIFICATION OF ASTHMA POPULATION: CSTE CLINICAL AND LABORATORY CASE CLASSIFICATION DEFINITION RESULTS 147

3.7.1 Confirmed Classification 147

3.7.2 Probable Classification 150

3.7.3 Possible Classification 151

3.7.4 Not Classified 151

3.8 VALIDATION STUDY DISCUSSION 159

4.0 CHAPTER 4- CASE CROSSOVER STUDY 162

4.1 POPULATION D 162

4.2 DISCUSSION 173

APPENDIX A: IRB 178

APPENDIX B: PERL PROGRAM 180

APPENDIX C: AIR POLLUTION DATA: ALL MONTHS 184

APPENDIX D: AIR POLLUTION DATA: SUMMER ALL MONTHS 199

APPENDIX E: AIR POLLUTION DATA: ALLEGHENY COUNTY HEALTH DEPARTMENT, 2008 206

APPENDIX F: COMPARISON OF RANDOM SAMPLE AND EMR TO PRIMARY ASTHMA DATASET BY DEMOGRAPHICS 208

BIBLIOGRAPHY 210

LIST OF TABLES

Table 1: Economic cost of asthma, United States 2007 27

Table 2: Point estimates for lag days associated with respiratory and asthma morbidity 47

Table 3: Diagnostic value of symptoms in the presence and absence of a physician diagnosis of asthma: Survey and chief complaint data 55

Table 4: ICD 9 Coding Process: Pathway for misclassification 64

Table 5: Council and State Territorial Epidemiologist: Asthma surveillance and case definition67

Table 6: Annual hospital admission rates Allegheny County residents 1999-2004 for respiratory disease by age 70

Table 7: Asthma hospital admission rates per 1,000 persons in Allegheny County..... 73

Table 8: Healthcare utilization patterns for individuals with a primary asthma diagnosis 2002-2005..... 75

Table 9: Keywords used for Practical Extraction and Report Language (PERL) matching 88

Table 10: Asthma medications*..... 89

Table 11: UMLS key concept codes for MedLEE..... 94

Table 12: Modified CSTE Clinical and Laboratory case classification criteria 96

Table 13: Primary and secondary asthma diagnosis for all emergency room visits (6 UPMC hospitals) by age, sex, race and disposition 102

Table 14: Insurance type for all primary and secondary asthma emergency room visits (6 UPMC hospitals) by disposition 104

Table 15: Primary and secondary asthma emergency room visits (6 UPMC hospitals) for month and year by disposition 106

Table 16: Primary and secondary asthma emergency room visits for 6 UPMC hospitals by year 107

Table 17: Primary and secondary asthma emergency room visits for the City of Pittsburgh (zip codes) by age, sex and race 2002-2005 (only 6 hospitals of UPMC network)..... 109

Table 18: Primary and secondary asthma emergency room visits for Allegheny County (zip codes) minus the City of Pittsburgh by age, sex and race (2002-2005) (only 6 hospitals in the UPMC network)..... 109

Table 19: First primary asthma emergency room visits (6 UPMC hospitals) by age, sex, and race by disposition 112

Table 20: Insurance type for all first primary asthma emergency visits (6 UPMC hospitals) by disposition 114

Table 21: All first primary asthma emergency room visits (6 UPMC hospitals) by month, year and disposition 116

Table 22: First primary asthma emergency room visit for 6 UPMC hospitals..... 117

Table 23: Random Sample- First Primary asthma emergency room visits (6 UPMC hospitals) by age, sex, and race by disposition..... 121

Table 24: Random Sample- Insurance status for all first primary asthma emergency room visits by disposition 123

Table 25: Random Sample –First Primary asthma emergency room visits by hospital and year 125

Table 26: A comparison of the ICD-9 Coded chief complaint admitting diagnosis to the primary hospital discharge diagnosis 127

Table 27: Electronic Medical Record: First primary asthma ED visit by hospital and disposition 131

Table 28: Electronic Medical Record: First primary asthma ED visit by age, sex and race 133

Table 29: Selected Indicators: Manual Review results..... 136

Table 30: Selected Indicators results: Manual vs. Automation 138

Table 31: MedLEE combined UMLS concept codes: Comparison to manual chart review and PERL..... 141

Table 32: Methodology used to establish degree of concordance 143

Table 33: Degree of Concordance: Manual chart review vs. automation..... 144

Table 34: CSTE Clinical and Laboratory case definition classification results (n=180) 149

Table 35: ICD-9 coded physician diagnosis of asthma: Unable to classify 152

Table 36: CSTE clinical and laboratory classification status by age, sex and race 154

Table 37: Comparison of Selected Indicators to CSTE Clinical and Laboratory Case Classification Definition 156

Table 38: Individuals with any primary asthma emergency room visit (6 UPMC hospitals) by age, sex, and race 163

Table 39: Any primary asthma ED visit by hospital..... 164

Table 40: Distribution of daily concentration of ozone, PM2.5, and temperature for all months (2002-2005)..... 167

Table 41: Table 41: Distribution of daily concentration of ozone, PM2.5, and temperature for summer months (2002-2005)..... 168

Table 42: Correlation coefficients for ozone and PM2.5 and weather variables..... 169

Table 43: Single and multiple lag models for 24-hour maximum ozone, PM2.5 and temperature: All seasons 170

Table 44: Single and multiple lag models for 24-hour maximum ozone, $PM_{2.5}$: Summer months (April-September)..... 171

Table 45: Single and multiple lag models for 24-hour maximum ozone, $PM_{2.5}$ and temperature for adults only: All months 172

LIST OF FIGURES

Figure 1: Asthma Prevalence for the United States, 1980-2004 13

Figure 2: Lifetime Prevalence of Asthma, 2005 17

Figure 3: Current Prevalence by Age, 2005 *(age specific rates)* 18

Figure 4: Current Asthma Prevalence by Sex and Race, 2005 (*sex and race specific rates*) 19

Figure 5: Attack Prevalence, 2005 20

Figure 6: Asthma related emergency room visits, 2004 22

Figure 7: Asthma Hospitalizations, 2004 23

Figure 8: Asthma Physician office visits, 2004 24

Figure 9: Asthma Deaths, 2004 25

Figure 10: Preliminary analysis: Primary asthma diagnosis for emergency room visits (6 hospitals in UPMC network) by disposition 74

Figure 11: Specific Aims 80

Figure 12: Questionnaire 85

Figure 13: PERL script 91

Figure 14: Sample size and power estimate 98

Figure 15: Primary and secondary asthma ED visits (6 UPMC hospitals) by month and year .. 107

Figure 16: Primary and Secondary Asthma Emergency Room Visits in Allegheny County by Zip Code (2002-2005) for 6 UPMC Hospitals 110

Figure 17: Emergency room visits for all first primary asthma by month and disposition 117

Figure 18: Emergency room visits for primary asthma (n=5,100) for 6 UPMC hospitals in Allegheny County 119

Figure 19: Random Sample: First primary asthma visits by month and year..... 124

Figure 20: Any primary asthma ED visit by month and year 165

Figure 21: Any primary asthma visit for 6 UPMC hospitals by zip code in Allegheny County 2002-2005 166

PREFACE

I would like to express my deepest gratitude to Dr. Evelyn Talbott for all of her guidance and support during my journey as a doctoral student. You have been an exceptional advisor and truly inspired me with your passion to the field of epidemiology. Your confidence in me has helped me reach new levels and now provides me the opportunity to work among those, which I believe I will forever owe you. I would like to thank my committee members, Dr. Joseph Costantino, Dr. Joseph Schwerha, Dr. Luann Brink, Melissa Saul and Gerald Barron for serving on my committee and providing much insight into this research.

Last and foremost, this work is dedicated to my parents who have supported me throughout this process. Without your prayers, love and patience I would have never completed this work. Dad and Mom, you are my rock, thank you for holding me up when I was always falling down. Mary Kay, dear sister and Paul, I treasure your counsel and felt encouraged by your support during the most challenging time of my life. Mary Kay, your fight to survive through three episodes of breast cancer taught me to never give up. Just as Dayna told us "strong, strong arm strong" the words of my best friend who lost her life to that dreadful disease. Melanie and Bobbie, your sense of humor in this whole process has kept me going and to both of you I am grateful. Last, in memory of my dear Aunt Rose Marie and Josephine I finally made it. Thank you for cheering me on from above.

1.0 CHAPTER 1- INTRODUCTION AND SIGNIFICANCE

Asthma is a chronic inflammatory disease of the airway that affects over 32 million individuals living in the United States (National Center Health Statistics [NCHS], 2005). Despite the dramatic rise in the prevalence of asthma over the last two decades, limited methodologies to conduct asthma, surveillance continues to challenge the public health professional when studying this disease (Boss, et al., 2001). The inability to identify asthma due to the lack of an operational definition, failure to access data to determine surrogate markers for asthma and the inability to determine asthma triggers for those already ill has hindered the understanding of the disease (Pearce, Douwes, & Beasley, 2000; Peat, Toelle, Marks, & Mellis, 2001; Pekkanen & Pearce, 1999: Pekkanen, Sunyer, Anto & Burney, 2005).

In the 2000 report, "*Attack Asthma: Why America Needs a Public Health Defense System to Battle Environmental Threats"* published by the Pew Environmental Health Commission, the need to develop surveillance capacity to detect major public health threats including asthma was highlighted. The ability to detect asthma cases first begins with a proper understanding of what constitutes a diagnosis of asthma. The use of the survey questionnaires to ascertain information about asthma has been widely reported in the literature (Pekkanen & Pearce, 1999). Although the use of a symptom questionnaire provides valuable information on asthma, limitations include recall and selection bias, which reduces the researcher's full understanding of the burden of illness from this disease (Peat et al., 2001; Pekkanen & Pearce, 1999). Since asthma consists of

a group of symptoms, careful consideration to the type of measurement used to capture information on these symptoms needs to be given (Marks, 2005). Using a definition, which can be agreed upon by the physician removes the potential for misclassification of the diagnosis and provides the opportunity to have a consistent approach to validate asthma cases for observation (Peat et al., 2001; Pekkanen & Pearce, 1999).

Describing the trends and burden of disease in individuals diagnosed with asthma is an essential component to public health surveillance. Data collected from national surveys on asthma prevalence and the burden of disease is useful for descriptive epidemiology, but does not provide data on health care utilization, which includes emergency room visits, hospitalizations, doctor office visits and prescription drug usage (Lanphear & Gergen, 2003). These data provide a more comprehensive picture of asthma and thus, serve as a valuable resource for conducting asthma surveillance. Moreover, data collected in the emergency room can be used to evaluate short-term effects caused by environmental factors such as in the case of air pollution exposure and asthma exacerbation.

A surveillance system using information technology to make these data readily available in an automated format, either during the delivery of medical care or with health plans linked to large medical practices that serve a well defined population, has been recommended by the U.S. Department of Health and Human Services [USDHHS] for conducting this type of surveillance (Baxter et al 2000 study cited in Lazarus, Kleinman, Dashevsky, Demaria & Platt, 2001, p. 2).

Thus, the electronic medical record, the automated repository for these data has the capacity to provide immediate access to emergency room visits data for monitoring asthma outcomes and conducting surveillance (Lazarus, et al., 2001; Lombardo et al., 2003; Sanders, Gregg & Arnosky, 2007; Vollmer, et al., 2004; Donahue, et al., 1997; Lanphear & Gergen, 2003;

Panackal et al., 2002; Hung, Posey, Freedman & Thorton, 1998. However, the scientific methods used to extract the data from the electronic medical record, decode it in a manner, which can be meaningfully interpreted with a high degree of statistical validation, must first be tested and the data interpreted for accuracy of diagnosis before these methods can be accepted as valid tools for conducting epidemiological research.

1.1 STATEMENT OF RESEARCH PROBLEM AND SPECIFIC AIMS

There is great need for identifying a consistent approach for conducting asthma surveillance (Peat et al., 2001). The capacity to detect a case of asthma could be improved by using a standard case definition with definable criteria limits for carrying out asthma surveillance (Council of State and Territorial Epidemiologists Case Definition for Asthma [CSTE], 1998). Since no one gold standard is available to diagnosis asthma, collecting data on multiple symptoms along with a physician diagnosis of asthma can be tested to determine which symptoms may be better at predicting an asthma diagnosis (Lanphear & Gergen, 2003; Sistek et al., 2001; Sistek et al., 2006; Sanders et al., 2007; Yu, Wong & Li, 2004). Data abstracted from the EMR and used for scientific conclusion must first be validated beyond the gold standard medical chart review. The use of a natural language processing system, a tool for transforming data from text to code is one methodology that can be used for data interpretation (Manning & Schuetze, 1999; Krauthammer & Hripcsak, 2001; Friedman, Shagina, Lussier & Hripcsak, 2004; Hripcsak, Austin, Alderspcsak, & Friedman, 2002; Chuang, Friedman, & Hripcsak, 2002; Melton & Hripscak, 2005; Chen, Hripcsak, Xu, Markatou, & Friedman, 2008). Moreover, data abstracted from the EMR can be used to test hypotheses about environmental factors, which may

precipitate an asthma event (Barnett, et al., 2005; Burnette, et al., 2001; Lin, Burnett, Villeneuve, Krewski, 2002; Villeneuve, Chen, Rowe & Coates, 2007; Luginaah, Fung, Gorey, Webster, & Wills, 2005; Pope, 1989; Wansoo & Schneider, 2005; Paulu & Smith, 2008). The goal of this research is to evaluate different methodologies for asthma surveillance and demonstrate the sensitivity of their results. These specific aims are:

1. To describe asthma in a population of Allegheny County residents. Characteristics include age, sex, race, hospital, patient type, insurance type, zip code, chief complaint, and admitting and discharge diagnosis.

2. To describe whether the use of electronic medical records are complete enough to use for asthma surveillance. The following questions are posed.

   A. Can the use of automated software extract key discriminating clinical characteristics from the EMR to validate a diagnosis of asthma?

   B. What is the degree of reliability between the different methodologies, manual vs. automation?

   C. Can the use of the Council of State and Territorial Epidemiologists clinical and laboratory case classification definition be used against the robustness of an ICD-9 coded physician diagnosis of asthma?

3. To determine the association between short-term changes in air pollution concentrations and the acute effects of asthma within a specified population using a case crossover study.

1.2 DEFINITION AND CLINICAL FEATURES

Asthma is a chronic disease of the airway that is characterized by a narrowing of the airway passages. The inflammation noted in asthma is produced through a series of "immune-mediated" events causing airway reduction (Merck Medicus Modules: Asthma-Diagnosis, 2001). The clinical symptoms manifested by the asthmatic include wheeze, cough, shortness of breath, and chest tightness. These symptoms occur from the bronchoconstriction created by the inflamed airway in response to variety of asthma triggers. These triggers provoke acute episodes of asthma symptoms which have been reported to be worse in the evening hours and morning hours or when exacerbated by such stimuli as cold air, allergens, air pollution, household dust mites, viruses and tobacco smoke (Merck Medicus Modules: Asthma-Definition, 2001). The symptoms of asthma can be treated with medications or relieved spontaneously on their own

Asthma is not one specific disease rather a spectrum of clinical symptoms, which appear through a number of unconventional pathways. "There are no clinical criteria that are both necessary and sufficient for the diagnosis of asthma" (Marks, 2005, p. 3). Particular to asthma is the inconsistencies associated with this disorder. Asthma can appear as an acute episodic disorder, which can last, from minutes to hours. "Although, asthma may "disappear" in 30-50% of children, it does tend to reappear in adulthood; and even among those who do not have clinical symptoms, lung function may remain altered" (Martin et al. 1982 study cited in Braman, 2006, p. 7). Moreover, the hallmark symptoms of asthma, which include wheezing, shortness of breath and cough, can suggest a diagnosis other than asthma (Marks, 2005). Thus, the considerable amount of overlap in the clinical symptoms of asthma and other respiratory diseases, plus the variability surrounding the manifestation of these symptoms, produces a challenge for not only diagnosing the disease but also controlling the clinical course.

1.2.1 Historical aspect

Asthma is not a new disease. Accounts dating as far back as 2 century AD where noted by Aretaues Cappacdocian, "if from running, gymnastic exercises, or any other work, the breathing becomes difficult, it is called *Asthma*" (D. Skonner, personal communication, October 27, 2007). Almost twenty centuries later, physicians are still focusing on the same clinical context to diagnosis asthma through symptoms of breathlessness induced through exercise, nighttime stress occurrence, and occupation (O'Donnell & Frew, 2002). Different terminology has been used to differentiate the different forms of asthma. These include extrinsic asthma, which is allergen related and intrinsic, a non-allergen type. These two forms are now referred to as atopic or non-atopic asthma (Kelley, Mannino, Homa, Savage-Brown, & Holguin, 2005). Atopy is the genetic marker for asthma, which has been reported in almost 66%-75% of all asthma cases (McFadden 1992 study as cited in May, 1996). Despite knowing the genetic marker to asthma, environmental factors are necessary to the acquisition of the disease.

The pathophysiological features of asthma produce a wide range of symptoms that eventually lead to structural changes or airway remodeling (Marks, 2005; Strek, 2006). The many different expressions from the unique clinical phenotypes exhibited by individuals with asthma have given way to describing asthma as a "heterogeneous disorder" (Kelly et al., 2005, p. 726).

1.2.2 Diagnosing and managing asthma

Over the past decade, researchers began to create new guidelines that focused on the treatment and management of asthma. The National Asthma Education and Prevention Program

(NAEPP) Expert Panel 2 Report, "Guidelines for the Diagnosis and Management of Asthma" through the National Heart, Lung, and Blood Institute (NHLBI), National Institute of Health (NIH) provide a structured approach for managing and treating asthma (NHLBI, 1997). The guidelines are prearranged into four areas: (a) measures of assessment and monitoring, (b) influencing factors, (c) medication, and (d) patient education. A diagnosis of asthma is made through an assessment of the patient's symptoms and results of their pulmonary function tests (PFTs). The test includes the use of spirometry to assess the exchange of airflow going in and out of the lungs. Measurements are used to assess normal breathing, forced inhalation and exhalation after a deep breath. This test remains the best test to help the practitioner determine the degree of airway obstruction (Merck Medicus Modules: Asthma-Definition, 2001; Spirometry and Flow Measurements-Standards & Guidelines, 1998). A methacholine challenge is another test, which can be done if there is not sufficient evidence from the patient's symptom history and spirometry results to confirm a diagnosis of asthma. Methacholine is administered through inhalation therapy to determine the degree of airway reactiveness. The results of the lung function measurements taken before and after the challenge is then compared (American Association Respiratory Care Clinical Practice Guidelines, 2001). Depending on the level of impaired lung function and signs and symptoms of the asthmatic determines the severity classification assigned to each individual. This classification of severity includes intermittent, mild, moderate and severe. The treatment goal is to control the asthma by minimizing the symptoms through medication therapy. Individuals with a mild form of asthma may only require a rescue drugs for relief of symptoms while individuals with a more severe form of asthma may use daily medications to control their disease. The more severe asthmatics represent approximately "5% of individuals who are resistant to therapy or difficult to control despite

taking the maximally recommended doses of inhaled medications-in particular, inhaled corticosteroids" (Strek, 2006, p. 116). Despite the severity of classification, the asthma treatment and management guidelines are designed to be an adjunct tool to aid the clinician in proper treatment and management of this disease.

2.0 CHAPTER 2- EPIDEMIOLOGY

Asthma is a public health problem reported worldwide with increasing prevalence noted for both adults and children (Beasley, 2002). Worldwide estimates for asthma have been reported at approximately 300 million people (Braman, 2006). The rise in prevalence appears to be associated with urbanization and westernization. National reports from the United States reveal one of the highest prevalence rates in the world with approximately 22.2 million Americans (7.7%) reporting a case of asthma (NCHS, 2005). Among those individuals diagnosed with asthma, an estimated 4.2% of the population (12.2 million) had at least one attack in the past year and 11.2% of persons (32.6 million) have even been diagnosed with asthma during their lifetime. Children under 18 years of age make up over one third of the asthma cases; 8.9% of children (6.5 million) have asthma compared to 7.2% (15.7 million) of adults.

The economic impact associated with asthma has been estimated at over 19 billion dollars (Weiss, Gergen, & Hodgson, 1992; [NHLBI, unpublished data] cited in the American Lung Association Report, 2007). Asthma affects individuals of all age groups, with the highest prevalence increase reported in children and adolescents (Beasley, 2002). Asthma is more common in school age children than preschool children or adults (NCHS, 2005). Asthma is the leading cause of school absenteeism from a chronic disease, which accounts for 14.7 million lost school days each year and limits some form of activity for school children age 5-17 (NCHS, 2002). Although, the burden of disease is high in children under the age of five symptoms are

frequently overlooked by parents or not recognized until a severe attack occurs (U.S. News & World Report: Asthma & Allergy, 2007).

The diagnosis although usually made in childhood, can occur in adulthood. Asthma in adulthood can be intermittent, persist from childhood, or present as a new diagnosis (Marks, 2005, USDHHS, 2000). A new case of adulthood asthma is to a great extent related to atopy or allergen sensitivity (Merck Medicus Modules: Asthma-Epidemiology, 2001). In working adults, studies have reported occupational factors (Kuschner & Stark, 2003) such as industry type and amount of work place exposure as the cause of 10-20% of the newly diagnosed adulthood asthma cases. The National Institute for Occupational Safety and Health (NIOSH) purport occupational asthma to be the most common occupational lung disease (NIOSH, 2004).

The presentation of asthma prevalence differs significantly by gender and age. Females overall report higher prevalence rates than males. "The male-to-female ratio of asthmatics is 3:2 among children ages 6-11 and increases to an 8:5 ratio among those ages 12-17" (Merck Modules: Asthma-Epidemiology, p.1). The converse is true in adults with females having the highest rates, which continue through adulthood. This gender reversal is not fully understood, however, hormonal differences have been shown to play a role (Strek, 2006).

Asthma is considered to be a disparate disease. A higher disease burden reported has been reported in certain racial and ethnic groups such as African Americans and Hispanics (MMWR, 1998; MMWR, 2002; MMWR, 2007; NCHS, 2005). African Americans are more likely to have asthma than their white counterparts. This finding has been consistent in all asthma categories which includes, lifetime, current and attack prevalence (NCHS, 2005). Ethnic subgroups, which include Hispanics, report lower prevalence rates than non-Hispanic Whites and non-Hispanic Blacks. Puerto Ricans have the highest asthma prevalence rate in the Hispanic

subgroups except for the non-Hispanic Blacks. Certain demographic groups are also at a disadvantage for poorer asthma outcomes. For example, a disproportionate number of Blacks are more likely to die of asthma than Whites (3.1 vs. 1.7 per 10,000) population, with Black females having the highest mortality rates (3.2 vs. 2.2 per 10,000) population, respectively. Asthma related deaths increases with age. A total of 186 children 0-17 years of age died from asthma in 2004 which represents (0.3 per 10,000) population. Individuals 18 years of age and over were more likely to die of asthma (2.5 per 10,000) population (National Vital Statistics System, 1980-2004; MMWR, 2007). Disparities were also reported for hospital admissions with rates being higher in Blacks vs. Whites (33.3 vs. 10.0 per 10,000) population, females vs. males (19.0 vs. 14.5 per 10,000 population) and in the northeast region of the country (23.7 per 10,000) population vs. (13.8 per 10,000) population in the West (MMWR, 2007).

The association between the increase in asthma prevalence and low socioeconomic has been reported for individuals living below the poverty level (MMWR, 2007). Asthma prevalence rates are higher for those individuals who live below the poverty level and live in the inner cities (Wissow, Gittelsohn, Szklo, Starfield & Mussman, 1998; USDHHS, 2000). The National Cooperative Inner-City Asthma project found that poor children living in the inner cities have difficulties gaining access to asthma care, and medications (Kattan 1997 study cited in USDHHS, 2000). Inner city children are exposed to a high level of household allergens (Crain et al., 2002). These allergens produce sensitivities, which precipitate asthma attacks (Rosenstreich, Eggleston, Kattan & Baker, 1997). Lower socioeconomic status places individuals at risk for (a) more frequent emergency room visits (b) higher number of asthma hospitalizations, and (c) non compliance with follow up care (Camargo, Ramachandran, Ryskina, Lewis, & Legorreta, 2007).

A cohort study examining enrollees in a State Medicaid program found recidivism rates to be higher for emergency room visits and hospital admissions in children who failed to take their asthma medications. Controlling asthma through medication compliance was the key to reducing risk (OR =0.55), (p = < .001) (Camargo et al., 2007). The outcome of not controlling asthma results in a more severe form of disease. Volmer (2001) reports that individual of lower socioeconomic status are more likely to have a severe form of asthma and die prematurely from the consequences. The reduction of income further complicates the risk by increasing hospitalization risk by 20% [OR= 1.20 (95% CI 1.02-1.4) for a reduction in income by $10,000 decrements (Eisner, Katz, Yellin, Shiboski & Blanc, 2001).

In order to understand the disparities noted for asthma, better data are needed to describe the differences in the demographic subgroups of age, race and sex. Currently, the best estimate of asthma prevalence comes from survey data. The primary survey most commonly used to report trends and track progress toward meeting national objectives on asthma is the National Health Interview Survey (NHIS). Three estimates are currently used to describe asthma prevalence (a) lifetime asthma, (b) current asthma, and (c) asthma attack. Until 1997, prevalence estimates were based a 12-month self-reporting measure which at the time did not require a physician diagnosis of asthma. Data reported after this time period (1997) reflects the use of two new measures, which are based on a physician diagnosis of asthma; lifetime prevalence or cumulative prevalence, and attack prevalence, which is based on someone reporting they had an asthma attack within the past 12 months. A third measure, current asthma prevalence was added in 2000 to establish whether someone still had asthma, given they reported a previous asthma history (Centers for Disease Control and Prevention, NHIS Asthma Prevalence; MMWR, 2007).

TRENDS

The increase in asthma prevalence over the past 40 years has become a global public health problem. The unanswered questions about what is causing this increase continue to challenge the public health practitioner (Beasley, 2002). Although this increase is not fully understood, multiple factors have been suggested to be responsible for this global issue (Buist, & Vollmer 1990 study cited in Balmes, 1993; Beasley, 2002). A review of the current trends in asthma prevalence for the United States is presented below.

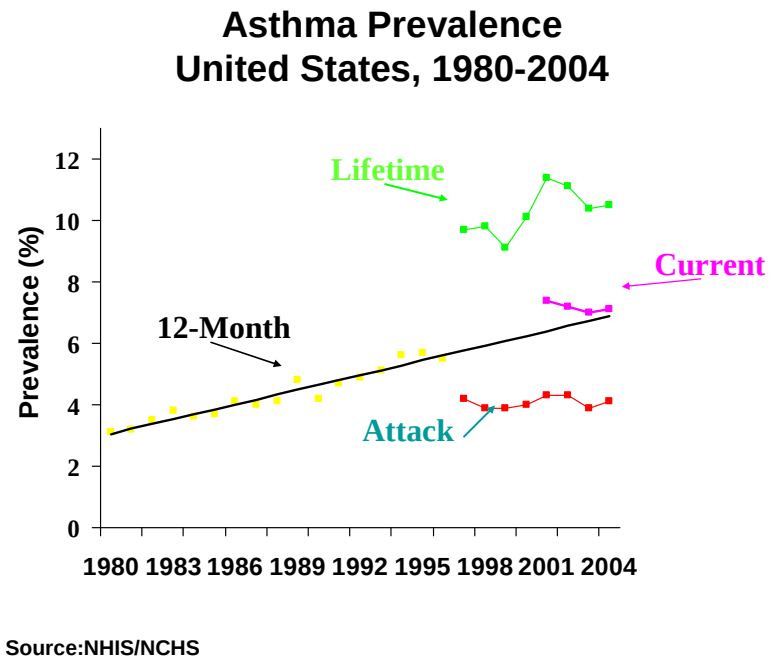


Figure 1: Asthma Prevalence for the United States, 1980-2004

[1] Reproduced from "Asthma Prevalence" by the Centers Disease Control and Prevention: Asthma Data and Surveillance. Source: NHIS, NCHS 1980-2004.

An overall assessment of asthma rates between 1980 and 1996 demonstrates a rise in asthma prevalence (Figure 1). This rise was based on the 12-month self-report asthma prevalence estimate. The lifetime and attack prevalence estimates following this time period fluctuated through 2004 (NCHS, 1980-2004; MMWR, 2007). Attack prevalence in 1997 to 2004 were 4.2%, 4.0%, 3.9%, 4.0%, 4.3%, 4.3%, 3.9% and 4.1%, respectively. Following this time period, attack prevalence has been at 4.2%, 2005-2007 (NHIS: [January-June 2007]), NCHS, 1997-2007). Lifetime asthma prevalence from 1997-2004 has also fluctuated (Figure 1), with estimates at 10.4% in 2000, increasing up to 13% in 2006. Current prevalence estimates in 2001 were 7.6%, decreased to 7.1% in 2003, increased to 7.3 % in 2004, 7.8% in 2005 and 8.0% in 2006.

Available data for US asthma deaths were also increasing from 1980-1995 (MMWR, 2002). This increase was highest for Blacks, females and older adults. There was a gradual decline in US death rates, which was noted after 1996 and continues up to present time (MMWR, 2007). The 1999 ICD-9 coding classification change from ICD-9 to ICD-10 may have likely influenced this trend. However, a "comparability ratio" was used to adjust rates to make data comparable from previous years. Following the adjustment to the classification code, disparities continued to persist with mortality rates reflecting similar patterns in the demographic subgroups.

Average annual numbers of deaths from 2001-2003 with asthma as the underlying cause were (2.1 per 10,000) population. The female death rate was higher (2.3 per 10, 000) than the males (1.8 per 10,000) population. Racial differences were noted with mortality rates being highest for Blacks (3.4 per 10,000) than Whites (1.9 per 10,000), Others (1.3 per 10,000), and

Hispanics (1.5 per 10,000) population. Black females had the highest mortality rate overall (3.5 per 10,000) population ((NCHS, 1980-2004; MMWR, 2007).

An admission to the hospital for asthma serves as a marker to assess trends in asthma morbidity. The asthma admission rates in the United States have increased and, similarly in other westernized countries over the past 40 years (Beasley, 2002). "Between the mid 1960s and the mid 1980s, hospitalization rates for asthma increased by more than 200% in children and 50% in adults" (Evans et al. 1989 study cited in Beasley, 2002, p. 483). Hospital admissions rates in 1980 were (18.5 per 10,000) and (17.9 per 10,000) population in 2004. Rates increased in 2003 to (19.9 per 10,000) population and decreased in 2004 (17.0 per 10,000) population (MMWR, 2007). Asthma hospitalization rates were consistently higher for females from 1985-2004 (18.7- 23.2 per 10,000 population) than males (16.8-17.4 per 10,000) population, respectively. Hospital admission rates have consistently been higher for Blacks (33.3 per 10,000) than Whites (10.0 per 10,000) and Others (19.0 per 10,000) population (US 1980-2004: MMWR, 2007). Contrasting to emergency room visit rates which were (66.2 per 10,000) population in 2000, (59.8 per 10,000) population in 2001, (68 per 10,000) population in 2002, and (63.4 per 10,000) population in 2004. Emergency room visits for asthma are at least 270% times greater than hospital admissions for asthma. This difference demonstrates that emergency room visits represent the largest at risk population versus hospital admissions, which reflect rates of the sickest population who are only at the highest level of risk due to a more severe form of illness.

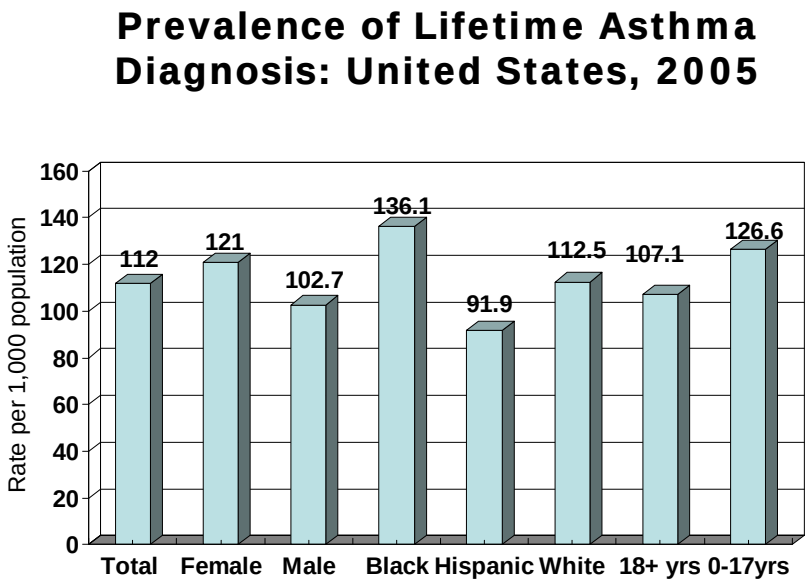
The variations noted for the trends in asthma mortality and morbidity over the last half of the decade have been suggested to result from several factors. These factors include " (a) the 1979 change in the International Classification of Diseases (ICD) coding of asthmatics bronchitis

as asthma rather than bronchitis, (b) a shift in physician diagnosis away from bronchitis to asthma, (c) an improved ability of physicians to diagnose asthma through greater availability and use of pulmonary function tests, (d) increased toxicity due to asthma medications, and (e) true increase in the prevalence and /or severity of asthma" (Buist & Vollmer 1990 study cited in Balmes, 1993, p. 221). Additionally, the change in the 1999 ICD classification code from ICD-9 to ICD-10 and environmental pollutants have likely played a role (MMWR, 2007, Balmes, 1993).

2.1 ASTHMA PREVALENCE AND BURDEN OF DISEASE

2.1.1 Asthma prevalence

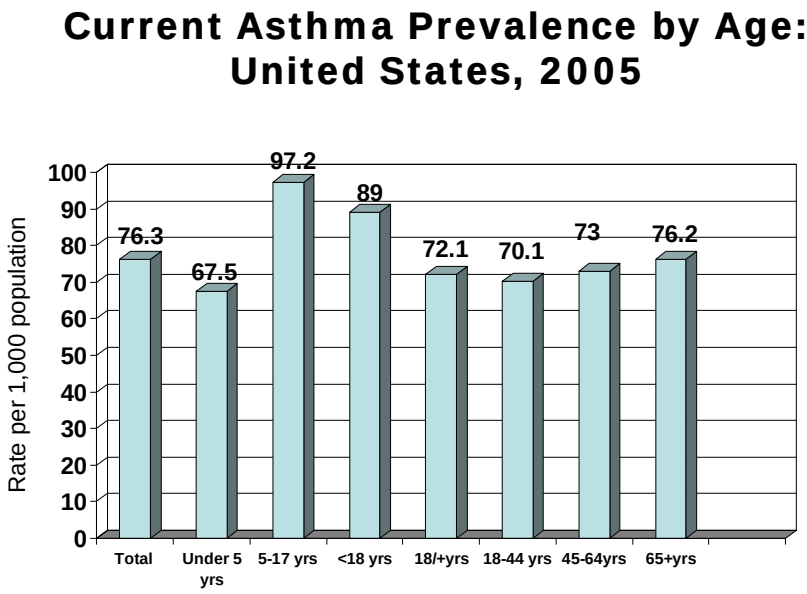
The following information describes asthma prevalence and burden of disease in the United States for 2005. Data from the 2005 NHIS, National Center for Health Statistics were used to create the following graphs and report on these data.



Source: (NCHS, 2005)

Figure 2: Lifetime Prevalence of Asthma, 2005

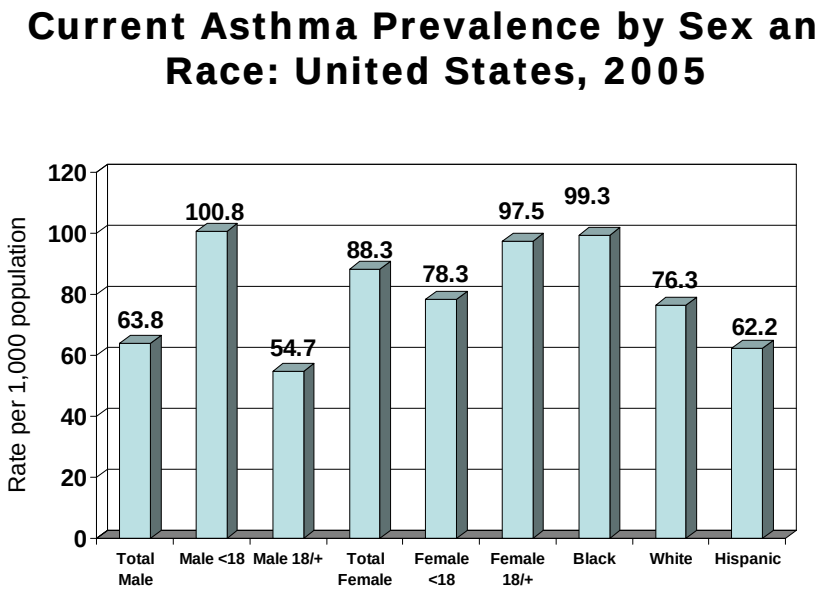
1. An estimated 32.6 million Americans or (112 per 1,000) persons report a lifetime diagnosis of asthma.
2. More females report a lifetime prevalence of asthma than males (121 vs. 102.7 per 1,000) persons, respectively.
3. Children less than 18 years of age had higher prevalence rates (126.6 per 1,000) persons than individuals 18 years of age and older (107.1 per 1,000) persons.
4. Children 5-17 were diagnosed with lifetime prevalence at a rate of (142.2 per 1,000) persons.
5. The rate in Blacks (136.1 per 1,000) persons was higher in both Whites (112.5 per 1,000) and Hispanics (91.9 per 1,000) persons.



Source: (NCHS, 2005)

Figure 3: Current Prevalence by Age, 2005 ***(age specific rates)***

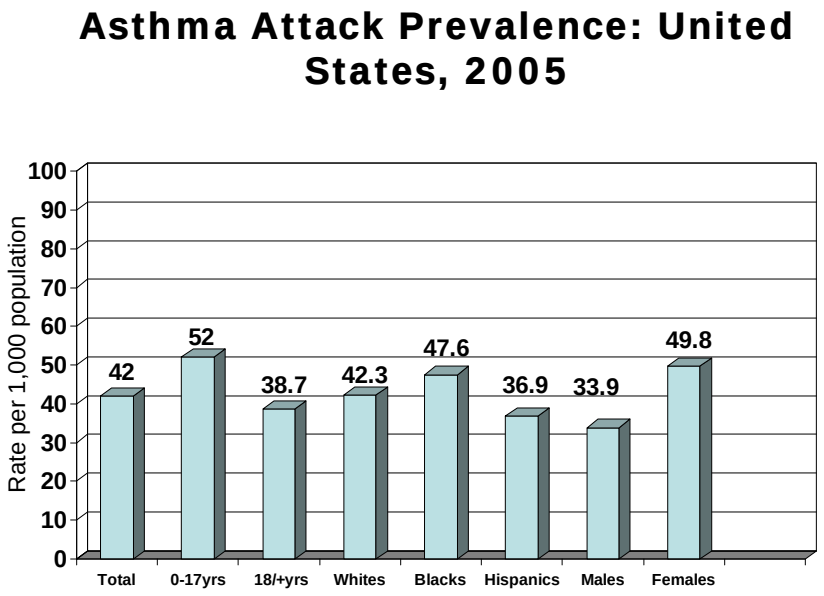
1. A total of 22.2 million Americans report they currently have asthma (76.3 per 1,000) persons.
2. Children under 5 years of age report the lowest current asthma prevalence rate (67.5 per 1,000) persons compared to all other age groups.
3. Children and adolescents under 18 years of age (89.0 per 1,000) persons reported the highest current asthma prevalence than those 18 years of age and older (72.1 per 1,000) persons.
4. Current asthma prevalence rates increased from childhood to adulthood (67.5 per 1,000 persons to 76.2 per 1,000) persons, respectively.



Source: (NCHS,2005)

Figure 4: Current Asthma Prevalence by Sex and Race, 2005 (*sex and race specific rates*)

1. The current asthma prevalence rate in females vs. males was (88.3 per 1,000) to (63.8 per 1,000) persons, respectively.
2. The rate for adult females 18 years of age and older was (97.5 per 1,000) persons and greater than males in this same age category (54.7 per 1,000) persons. Contrasting the younger males under 18 years of age who have higher current asthma prevalence rate (100.8 per 1,000) persons than females (78.3 per 1,000) persons.
3. Black had higher current prevalence rates (99.3 per 1,000) persons compared to Whites (76.3 per 1,000) and Hispanics (62.2 per 1,000) persons.



Source: (NCHS, 2005)

Figure 5: Attack Prevalence, 2005

1. 12.2 million Americans reported having an asthma attack within the past 12 months (42.0 per 1,000) persons.
2. Children 5-17 years of age had the highest attack prevalence (55 per 1,000) persons vs. the lowest rate which was found in the oldest age group (65+) (30.4 per 1,000) persons.
3. Asthma attack prevalence rates are higher in the 0-17 year old age group (52.0 per 1,000) persons than in individuals 18 years of age and over (38.7 per 1,000) persons.
4. More females than males had an asthma attack within the past 12 months (49.8 per 1,000 vs. 33.9 per 1,000) persons.
5. Blacks have higher asthma attack prevalence rates (47.6 per 1,000) persons than Whites (42.3 per 1,000) and Hispanics (36.9 per 1,000) persons.

2.1.2 Burden of disease

The burden of disease and mortality from asthma can be characterized by a review of data on health care use, which includes emergency room visits, asthma hospitalizations, physician office visits, and mortality data. Asthma emergency room visits are far greater in number than asthma hospitalizations. For example, in 2004 the numbers of emergency room visits for asthma were 1.8 million compared to the number of asthma hospitalizations, which were 504,000. This represents (63.4 per 10,000) vs. (17.0 per 10,000) population, respectively or a 270% greater difference (US 1980-2004: MMWR, 2007). A rise in emergency service utilization and hospitalizations are a direct result of poorly controlled asthma (Camargo, 2007; Strek, 2006). Although asthma hospitalization can be used to characterize the burden of disease, they only capture the highest at risk population or those with the most severe form of the disease. Furthermore, individuals with a more severe form of asthma are at risk for premature death (Volmer, 2001).

The National Ambulatory Medical Care Survey, National Hospital Ambulatory Care Survey, National Hospital Discharge Survey, and National Vital Statistic System were all data sources used to report on these findings. The National Surveillance for Asthma: United States 1980-2004 Surveillance Summary Report (MMWR, 2007) was the main document referenced for this section of the review.

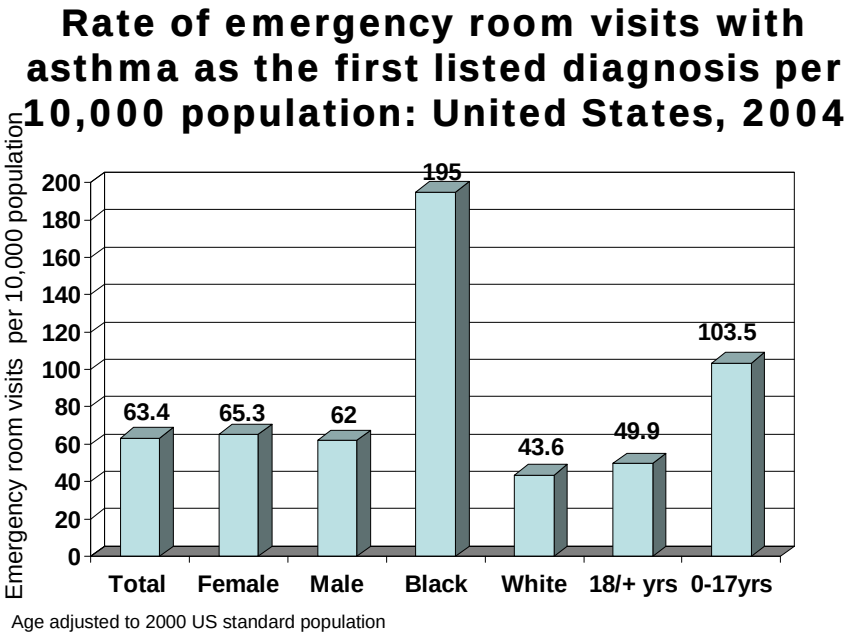


Figure 6: Asthma related emergency room visits, 2004

1. Asthma accounted for 1.8 million emergency room visits in 2004. This represents 63.4 per 10,000 populations.
2. Females have a higher usage of emergency room services than males (65.3 per 10,000 vs. 62 per 10,000) population, respectively.
3. Rates for Blacks (195 per 10,000) are much greater than Whites (43.6 per 10,000) population.
4. Children aged 0-17 were the highest users of emergency room care, a rate of (103.5 per 10,000) population.
5. The highest rate was among children 0-4 years of age at (168.3 per 10,000) population.

6. In 2004, adults 18 years of age and older used the emergency room less frequently (49.9 per 10,000) populations than individuals less than 17 years of age (103.5 per 10,000) population.
7. Use of the emergency room by the individuals 5-14 years of age were (155.1 per 10,000) population and the elderly at (138.4 per 10,000) population.

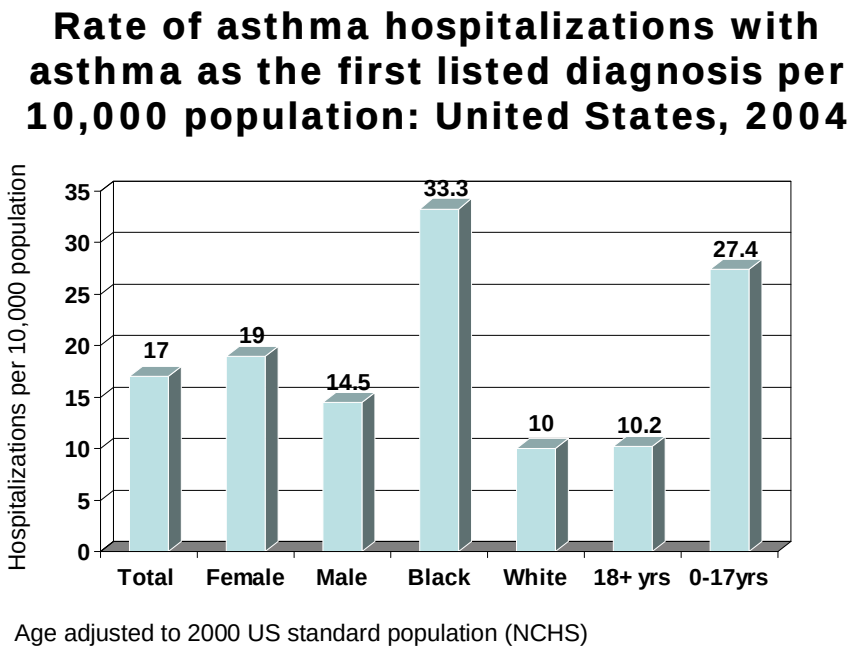


Figure 7: Asthma Hospitalizations, 2004

1. There were 504, 000 asthma hospitalizations in 2004 due to asthma, which represents (17 per 10,000) population.
2. A total of 204,700 hospitalizations were in children under the age of 17 or (27.4 per 10,000) population.
3. The hospitalization rate (59.9 per 10,000) population was highest in children 0-4 years of age followed by individuals 65+ (28.7 per 10,000) population.

4. Blacks were over 3 times more likely to be admitted (33.3 per 10,000) vs. that of their white counterparts (10 per 10,000) population.
5. Females had a higher admission rate than males (19 vs. 14.5 per 10,000) population.

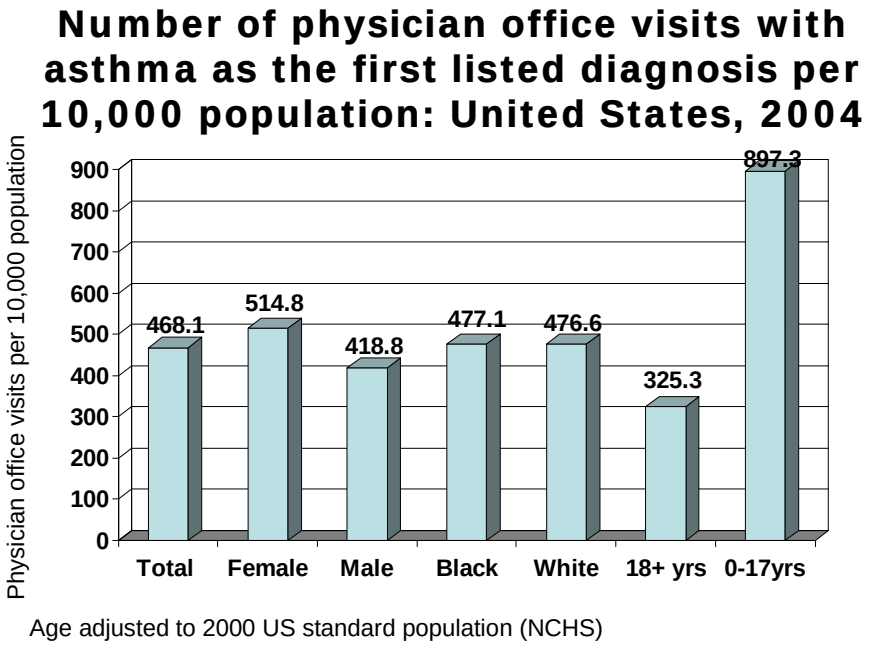
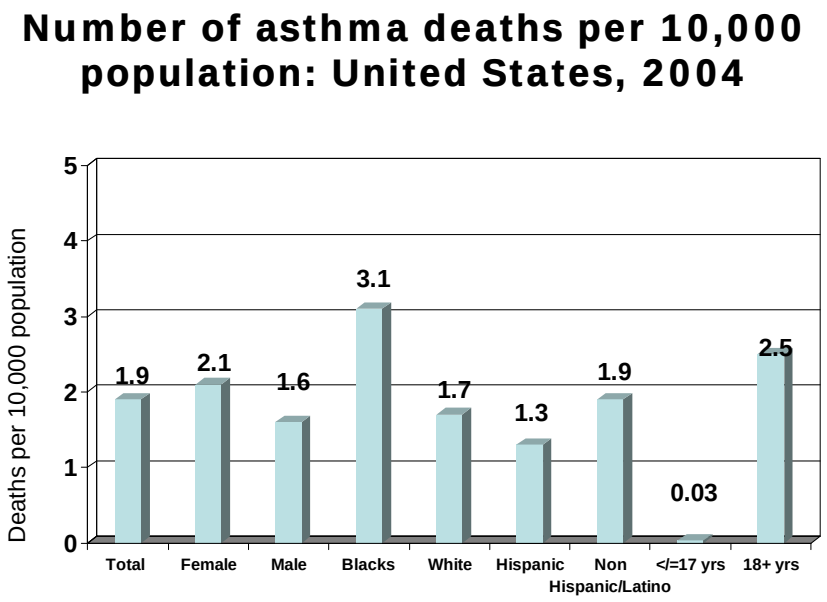


Figure 8: Asthma Physician office visits, 2004

1. There were 13.6 million outpatient visits related to asthma care in 2004. This represents (468.1 per 10,000) population.
2. Children 17 years of age and younger utilized services at a higher rate (897 per 10,000) population than adults 18 years of age and older (325.3 per 10,000) population.
3. Blacks' rates were slightly higher than Whites (477.1 vs. 476.6 per 10,000) population, respectively.
4. Females had a higher rate of utilization of outpatient services than males (514.8 vs. 418.8 per 10,000), respectively.



Age adjusted to 2000 US standard population (NCHS)

Figure 9: Asthma Deaths, 2004

1. A total of 3,816 people died of asthma in 2004. This represents (1.9 per 10,000) population.
2. A total of 186 children 0-17 years of age died from asthma in 2004 which represents (0.3 per 10,000) population. Individuals 18 years of age and over are more likely to die of asthma (2.5 per 10,000) population.
3. Death rates were 31% higher in females than males (2.1 v. 1.6 per 10,000) population, respectively.
4. Blacks were more likely to die of asthma (3.1 per 10,000) than Whites (1.7 per 10,000), Hispanics (1.3 per 10,000) and Non-Hispanics (1.9 per 10,000) population.

2.1.3 Economic burden

Asthma is a major public health problem with significant economic consequences. The estimated cost of one case of asthma results in a price tag of approximately "$300 to $1300 dollars per-person per-year" with the largest burden coming from those individuals whose asthma can not be controlled (Sullivan et al 1996 study cited in Beasley, 2006, p. 487); Barnes, Johnson & Kilm, 1996). A rise in emergency service utilization and asthma hospitalizations has been reported as a direct result of poorly controlled asthma (Camargo, 2007; Strek, 2006). Although, the number of individuals with poorly controlled asthma represents the smallest proportion of asthmatics (10-20%), they are considered to be "high cost-patients," who are responsible for using over 50% of the direct medical care expenditures (Beasley, 2006, p. 488; Sullivan et al 1996 study cited in Beasley, 2006, p. 487; Braman, 2006). The data reported in Table 1 provides an overview by category for asthma cost. Direct expenditures account for 14.7 billion or 77% of the total economic burden for asthma. It appears that if asthma could be controlled through medication management, the financial burden would be drastically reduced.

Table 1: Economic cost of asthma, United States 2007

Category	Costs (Billions)
Direct medical expenditures	
Hospital Care	4.7
Physicians Services	3.8
Prescription Drugs	6.2
All Direct Expenditures	14.7
Indirect Costs	
Morbidity	3.1
Mortality	1.9
All Indirect Costs	5.0
All Costs	19.7

Source: Weiss, et al., New England Journal of Medicine. (March 26, 1992): Vol. 326;13.
NHLBI chart book on cardiovascular, lung and blood disease, 2007 [unpublished data] cited in NCHS, 2007

Reproduced from the 2007 American Lung Association Report, "Trends in Asthma Morbidity and Mortality", p. 32.

Reprinted with permission ©2009 American Lung Association
For more information about the American Lung Association or to support the work it does, call 1-800-LUNG-USA (1-800-586-4872) or log on to www.lungusa.org.

In summary, several factors have been reported that describe the burden of disease for asthma. Although each factor is important for understanding the disease, the emergency room visit estimates appear to be a better marker for conducting asthma surveillance. First, emergency room visits for asthma are greater in number than hospital admissions (1.8 million ED visits vs. 504,000 asthma hospitalizations), respectively. This represents (63.4 per 10,000) vs. (17.0 per 10,000) population, respectively or a 270% greater difference (US 1980-2004: MMWR, 2007). A rise in emergency service utilization is directly related to poorly controlled asthma (Camargo, 2007; Strek, 2006). A comparison of ED visits to asthma hospitalizations in our preliminary analysis (Section 3.2, Figure 10) has demonstrated that for 1 asthma hospital admission per

person, there are 4 ED visits (1116 to 4907). These visits result in a huge economic burden to the health care system (Beasley, 2006, Weiss, et al; 1992). Furthermore, by using asthma hospitalizations as a marker for asthma only touches upon patients who are severely ill. Asthma hospital admissions do not represent the total at risk population, but rather those who are at the highest risk with a more severe form of the disease.

Second, the short-term evaluations for asthma require more detailed data to include physician diagnosis and the hallmark signs and symptoms, which characterize this disease. These data can be found in the medical record with multiple supporting documents that provide a more comprehensive picture of asthma. The application of an automated feature to the medical record provides a stream of clinical data for conducting asthma surveillance. Moreover, the facilitation of these data through an electronic system would support national tracking efforts for testing hypotheses related to this chronic condition (L. Brink, personal communication, April 15, 2008).

2.2 RISK FACTORS

One of the difficulties in differentiating risk factors for asthma is the fact that we are unable to identify the cause (Anto, 2004; Pearce et al., 2000; von Mutius, 2000). To better address what may be causing asthma, research suggest an interaction between the environment and genetic factors (Etzel, 2003; Johnston & Sears, 2006; May, 1996; von Mutius, 2000). Migration studies support the role of the environment in the etiology of asthma. Individuals, who move from developing countries where prevalence rates are low to western countries where prevalence rates are high, see a sharp rise in asthma (Braman, 2006). A purely genetic component would take years to influence this change. However, changing influences in the environment are believed to play a role on disease development and exacerbation in susceptible individuals.

The risk of asthma exacerbation has been shown to differ significantly by age and sex with asthma being more dominant in young males up through puberty with a trend reversal in females continuing up through adulthood (Mandhane, Greene, Cowan, Taylor & Sears 2005; Merck Medicus Modules: Asthma-Epidemiology, 2001; Johnston & Sears, 2006; Schatz, Clark & Camargo, 2006). Hormonal differences may have likely contributed to this outcome (Johnston & Sears, 2006). Several studies account for the gender and age differences in asthma. One study reported that wheezing differed significantly in male children and female adolescents ($p = < 0.002$) (Mandhane et al., 2005). In Schatz et al (2006) male children under the age of 15 were reported to be two times more likely to be admitted for asthma than their female counterparts: A three-fold increase was reported in females over the age of fifteen. This noticeable age and gender differential in asthma was apparent for asthma severity, females were

found to be at greater risk (1.89) of having an asthma attack than their male counter parts between the ages of 18-70. For each yearly increase in age a 1% increase was noted for an asthma attack (OR 1.01[95% CI 1.001-1.023]) (Manfreda 2001 study cited in Johnston & Sears, 2006 p. 724).

Another environmental factor influencing asthma outcomes is respiratory infections, particularly, the rhinovirus, which has been associated with the majority of asthma exacerbations in children during the fall months (Dales, et al., 1996; Johnston & Sears, 2006; Wansoo & Schneider, 2005). One study reported seasonal variations in hospital admissions for a group of preschool children who were at a four times greater risk for being admitted for asthma vs. a non-respiratory admission, July-October. Fourteen percent of the variation was attributed to respiratory infections. Wansoo & Schneider (2005) observed a statistically significant rise in asthma hospitalizations in the autumn months which was linked to weed pollen ($p = < 0.001$). An epidemiological review on asthma exacerbation by Johnston & Sears (2006) provided documentation of this seasonal phenomenon and its presence in countries across the Northern Hemisphere. Evidence suggests that this seasonal pattern is coupled with a rise in respiratory infections. Thus, suggesting the role of respiratory infections on seasonal patterns of asthma and in asthma exacerbation.

Asthma is significantly impacted by indoor air quality. The Environmental Protection Agency (2002) reports that indoor air quality poses a health risk due to the emission of gases and particles from indoor pollutants such as oil, gas, coal, paints, carpets, furniture, and sanitizing and personal products. If these gases and or particles are not controlled or minimized through proper ventilation, they can invoke an allergic response triggering an asthma attack.

Indoor air exposures to house dust mites; cockroach and cat allergen concentrations also play a role in asthma morbidity (Etzel, 2003). Sensitivity to cockroach allergens demonstrated by a positive skin test correlated to a three-fold increase in asthma hospitalizations amongst inner- city children (Rosenstreich et al., 1997). Individuals exposed to indoor molds can also develop allergen sensitivities capable of triggering an asthma attack (Etzel, 2003)

Exposure to environmental tobacco smoke is another significant risk factor for asthma (Etzel, 2003; Chilmonczyk, 1993; Stick et al., 1996). According to a report, "*Clearing the Air*" published by the Institute of Medicine (IOM, 2000), there is sufficient causal evidence to link tobacco smoke to the development and exacerbation of asthma (IOM 2000 study cited in Etzel, 2003, p. 235). Children of parents who smoke are reported to have decreased pulmonary functions and more frequent asthma symptoms (Chilmonczyk et al., 1993). Maternal smoking has also been reported to increase the risk of wheezing and asthma development in children (Stick, Burton, Gurrin, Sly, & LeSouef, 1996).

Atopy or allergic asthma is a risk factor for asthma (Pearce et al., 2000; Johnston & Sears, 2006; Mandhane, et al., 2005; May, 1996; von Mutius, 2000). Allergen exposure is hypothesized to be a causal link to asthma though not enough population studies have been done to support this assumption (Pearce et al., 2000). Evidence does support a secondary cause in that sensitization from allergen exposure or prolonged exposure leads to the development of asthma symptoms (von Mutius, 2000; Wansoo & Schneider, 2005; Rosentreich et al., 1997; Pearce et al., 2000; May, 1996; Kelley et al., 2005; Etzel, 2003; Braman 2006; Anto, 2004; Johnston & Sears, 2006). In sensitized individuals, exposures to allergens such as pollens and molds can exacerbate asthma (Wansoo & Schneider, 2005; Etzel, 2003). Factors thought to impact allergen exposure include temperature, rainfall, relative humidity and thunderstorm activity. Anderson et.

al. (2001) found an increase in asthma admissions related to thunderstorms (p = < .001). The effect was greater in months with warm weather and not associated with rainfall. Similar results were reported by Marks et al. (2001) for thunderstorm activity. Asthma admissions were greater on days when (a) thunderstorms were present (OR= 15.0), (p = < .001), (b) during the warmer months, and (c) paralleled a 4-12 fold rise in grass pollens.

Socioeconomic status has been purported to be a risk factor for asthma (Camargo, et al., 2007; Volmer, 2001; Eisner et al., 2001; Wissow et al, 1988; DHHS, 2000; MMWR, 2007; Crain et al., 2002; Rosenstreich, et al., 1997). Higher prevalence rates have been observed in western cultures (Braman, 2006). Asthma prevalence rates have been shown to increase in individuals moving from rural to urban areas. However, when looking at inner city regions of the United States individuals of lower socioeconomic status have demonstrated an increase in asthma morbidity (Wissow et al, 1988; DHHS, 2000). Inner city children are exposed to a high level of household allergens (Crain et al., 2002). These allergens produce sensitivities, which precipitate asthma attacks (Rosenstreich, et al., 1997). Lower socioeconomic status correlates to (a) more frequent asthma emergency room visits (b) higher numbers of asthma hospitalizations, and (c) non compliance with asthma follow up care (Camargo, et al., 2007). Individual of a lower socioeconomic status are more likely to have a severe form of asthma and die prematurely from the consequences (Volmer, 2001). Lower socioeconomic status has been demonstrated to be a risk factor for asthma hospitalization based on a reduction in income by $10,000 decrements, hospitalization risk increased by 10% [OR= 1.10(95% CI, 0.9-1.3)]. After adjusting for demographic and asthma clinical factors the OR was significant 1.2 (95% CI 1.02-1.4) (Eisner, et al., 2001).

Air pollution is another factor considered to impact asthma. Criteria pollutants such as ozone, particulates, nitrogen oxide, carbon monoxide and sulfur dioxide mainly come from vehicle exhaust, power and industrial plants and factories (Clean Air Act, 1970). There has been a considerable amount of research done, which links the short term exposure to these pollutants with a rise in cardiovascular and pulmonary mortality and morbidity. Several of these studies (*) report directly on the relationship between air pollution and asthma (Barnett, et al., 2005*; Bell, McDermott, Zeger, Samet, & Dominici, 2004; Bell, Kim & Dominici, 2007; Bell, Dominici & Samet, 2005; Borja-Aburto et al., 2004; Burnett et al., 2001; Ito, De Leon, Lippmann, 2005; Lin, Burnett, Villeneuve & Krewski, 2002; Pope, 1989*; Luginaah, Fung, Gorey, Webster & Wills 2005; Villeneuve, Chen, Rowe & Coates, 2007*; Aburto-Borja, Castillejos, Gold, Bierzwinski, & Loomis, 1998; Paulu & Smith, 2008*).

2.3 RELEVANT RESEARCH ON AIR POLLUTION

There is a great deal of evidence that supports the biological plausibility of the induced adverse health effects from exposure to ambient air pollutants. Epidemiological studies suggest that inflammation or oxidative stress is produced when oxidants, metals, or reactive organic compounds found in these pollutants generate reactive oxygen species or free radicals, which damage DNA, thus promoting disease (Barry, 1991; Li et al., 2003, Risom, Moller & Loft, 2005; Bhattacharya, Alink, & Dopp, 2007; Rahman, Morrison, Donaldson, & MacNee, 1996). The development of free radicals and their effect on the body at the cellular level are associated with increased oxidative stress, which contributes to lung inflammation as seen in asthmatics (Rahman et al., 1996). The small fraction size particulates penetrate the lung tissue, damage the

mitochondria and create airway damage (Li et al., 2003). These particulates also have the ability to leave the pulmonary system and enter into the blood stream causing systemic inflammation responsible for sudden cardiac death, and interference in platelet activation that is capable of producing clots or ischemic heart disease (Nemar, Maskari, Ali & Al-Amri, 2007; Donaldson, Mills, MacNee, Robinson & Newby, 2005). These mechanisms provide evidence to support the conclusions drawn from epidemiological studies, which link air pollution exposure to cardiovascular and respiratory mortality and morbidity.

Dockery (1993) points out that the type of epidemiological study used to investigate air pollution exposure and its human health effects is particularly challenging due to the complexities surrounding air pollutant mixtures. First, individual exposure is universal and varies considerably from person and place. Although, National Ambient Air Quality Standards are in place to protect individuals from risk, including sensitive populations such as children, the elderly and asthmatics, the 2008, '*State of the Air Report*" published by the American Lung Association (ALA) indicates that nearly half of the US population (46.0%) live in areas where either ozone or particulate levels are at unhealthy levels.

The biological response resulting from these exposures is impacted by a time scale based on a pollutant threshold. In a chronic disease such as lung cancer, the exposure threshold is cumulative from the effects of smoking, while in acute disease the exposure threshold can be based on a less frequent or more severe form of an exposure. This is evident in the individual who is exposed to carbon monoxide and presents to the emergency room with carbon monoxide poisoning (Jaakkola, 2003; Committee on Advances in Assessing Human Exposure to Airborne Pollutants, 1991). In a disease such as asthma whose exposure pathway is not fully understood, a dose-response relationship is difficult to quantify. Furthermore, several environmental factors

are strongly linked to asthma development and exacerbation, thereby, making it difficult to measure the exposure relationships (Etzel, 2003; Rosenstreich et al., 1997; EPA, 2002: Chilmonczyk, 1993; Stick et al., 1996; Barnett, et al., 2005; Pope, 1989; Villeneuve et al., 2007; Wansoo & Schneider, 2005; Paulu & Smith, 2008). Therefore, finding the appropriate methodology that provides the structure to measure the short term effects of air pollution exposure on asthma requires careful consideration.

2.4 CASE CROSSOVER DESIGN

Epidemiologist have begun grappling with the inability to capture the onset "trigger" of an acute event such as asthma from the short term changes in air pollution using known retrospective review methods (Maclure, & Mittleman, 2008; Jaakkola, 2003). A new study design, the case crossover methodology was proposed "using cases only, and, for each individual, compares exposure just prior to the event with exposure at other control, or referent times" (Janes, Jones, Sheppard, & Lumley, 2005, p. 717). The same day of the week is usually chosen as the control or referent time to eliminate any confounding from weekly or and personal activity patterns (Janes et al., 2005; Maclure, & Mitttleman, 2008; Haley et al., 2006). Furthermore, the ability to make an intra-subject comparison between the cases and controls eliminates confounding in personal and exposure characteristics.

The exposure in the case crossover design is based on the exposure measure for the hazard period- the time period immediately prior to the event. A comparison is made to assess the exposure in the matched control for the same time period the exposure was taken for the case. Exposure can be measured in terms of the level of pollutant on the event day (Lag 0), cumulative

effect represented by (Lag 0+1+2), or an average over several days (Average 0,1, 2) (Haley et al., 2006). The strategies used to measure the effect should be modeled using a comparison of the different lag distributions in order to explore the relationship between different exposure levels and the event (Bell, et al., 2004). Solely using a single lag model would reduce the ability to explore these relationships.

The goal for using a case crossover methodology is to create a "matched set" between the exposure and respective control while eliminating all the temporal variations in the exposure and outcomes data, a no trend hypothesis (Janes et al., 2005). The proper assignment of a sampling strategy removes the variations in the data by having equal distance between the hazard event and control periods, matching for same day of the week and within the same calendar month. The use of several strategies has been proposed for the case crossover methodology each with their own particular strengths and weakness. The strategies include (a) unidirectional, (b) bi-directional and, (c) time-stratified The unidirectional approach is limited by a single index time. Therefore, exposure estimates are based solely on one measurement period. In the bi-directional approach, two control periods are present, one before and one after the event. However, the bi-directional strata do not always assure a fixed level of exposure across the referent periods. Hence, the goal for controlling the variations in the exposure and outcomes data can result in an overlap bias (Janes, Sheppard, Lumley, 2004). The time stratified is the preferably method of choice for air pollution studies (Lu, Symons, Geyh & Zeger, 2008). A time stratified approach selects a hazard period within a calendar month and selects corresponding control points to the case within the month, using the same day of week interval to appoint the control. The selection of the referent periods prior to initiating the experiment and within the same calendar month reduces the bias in the estimates of conditional logistic regression (Janes, et al., 2004).

The case crossover design is considered to be an alternative to the time series model, a most commonly used methodology to assess air pollution exposure and health events over time (Lu, et al., 2008; Maclure, & Mittleman, 2008; Haley et al., 2006). The case crossover methodology can not measure effects from long-term exposure found in such diseases like cancer. However, it can measure the trigger of an event or short term exposure (Jaakkola, 2003). This feature makes the case crossover methodology favorable for studying air pollution exposure and acute health outcomes.

The analytical measurement for the case crossover design follows the assumption of the matched case control study (Jaakkola, 2003; Lu et al., 2008; Janes et al, 2005). The estimator used to conduct the analysis in the case crossover methodology is conditional logistic regression. "The measure of the effect, called the odds ratio is calculated by dividing the number of subjects exposed only during the hazard period by those exposed during the control period. Conditional logistic regression analysis allows modeling using several or a varying number of control periods"(Jaakkola, 2003, p. 83)

Based on the statistical requirements for conditional logistic regression each event needs to be independent of one another (Lou & Sorock, 2007). This would not be the case in a dataset with recurrent events. One approach is to use only the single event in the analysis, however, this methodology results in a reduction of sample size and power estimate. To avoid the loss in power from a reduction in sample size, a "within-subject pair wise resampling technique or a weighted estimating equation" has been proposed (p. 2890). For the purpose of this research, multiple events will be included in the analysis. However, the use of a "wash out" period will be applied to assume an independence status. The wash out will be set to remove anyone who had a

repeat visit within a 28-day time period, thus, accounting for independence within the 28 day time strata (Haley et al., 2006).

2.5 SUMMARY OF LITERATURE REVIEW

A number of studies report on the relationship between the short term effects from ozone and cardiovascular and respiratory mortality. The effect of ozone is reported to be cumulative and associated with the accumulation of ozone concentrations over a few days prior to death (Bell et al., 2004). Although, the effects of ozone has also been observed on day-0 and days prior to an event, the relationship between the collective effect of ozone and death is lost when a single lag model is used. The distribution lag model, which takes into account the cumulative effect from multiple days of ozone exposure, "thus allows more flexibility for exploring the lag between exposure and death than single lag models" (p. 2373). In one study representing 95 US cities [approximately 120 million persons], Bell et al (2004) reported that for every 10-ppb increase in the previous week's ozone levels, total mortality was increased by 0.52%(95% CI 0.27%-0.77%). This effect was associated with an increase in cardiovascular and respiratory related deaths 0.64%(95% CI 0.31%-0.98%). While the effect of ozone on total mortality is evident, a comparison of these data may be limited due to the co-existence of other air pollutants with ozone (Bell et al., 2007). This relationship could potentially confound the short term effects of ozone mortality. In Bell's (2004) study the use of PM_{10} was selected to explore this relationship. A 1-day lag for PM_{10} was chosen as the effect estimate as Hoek (2007 study cited in Bell et al, 2004) reported this lag to have the largest point estimate. Concluding results were not sufficient to demonstrate that ozone mortality was impacted by the adjustment of the

particulate. However, a conclusion drawn from this research study suggests that the consequences from ozone on total mortality and cardiovascular and respiratory mortality are great. If these findings were applied to the US population, a significant number of premature deaths could be eliminated through proper interventions, including more stringent ozone regulations.

The ozone mortality relationship was further explored as it relates to cardiovascular and respiratory deaths using meta-analysis data from 39 time-series studies. The potential confounding relationships of particulates were not addressed in this investigation. Pooled lag estimates were compared to respective estimates from the National Morbidity, Mortality, and Air Pollution Study (NMMAPS). Two comparisons were made (a) the pooled effect estimates from the meta analysis, which includes all 9 cities, was compared to the pooled effect estimates from all 95 cities from the NMMAPS, and (b) the effect estimates from 8 cities, which were represented in both studies, were weighed against on another. For the first comparison, single lag and 2-day average lag effect estimates for a 10-ppb increase of ozone on total mortality was higher for the meta-analysis 0.87%(95% CI 0.55%-1.18%) than the baseline comparison using the NMMAPS study's day-0 lag [0.25%(95% CI, 0.12%-0.39%)]. The second comparison also resulted in a higher effect estimate for the meta analysis [0.83%(95% CI, 0.38%-1.29%)] than the NMMAPS estimate [0.48%(95% CI, 0.03% - 0.92%)]. Results from this study provide evidence that a relationship does exist between ozone and mortality. However, the authors call attention to the reader about a possible publication bias. Since effect estimates were different for 0 and 1-day lag vs. multiple lag days, Ito et al (2005, p. 447) reports that "when an association is found at multiple days, choosing only a single-day's result would underestimate the multi-day effects. Thus, using a risk estimate for a single day lag can result in bias in either direction".

Furthermore, the results in the meta-analyses were always greater than the baseline study results: perhaps the author wanted to publish positive findings from this study (Bell et al., 2005).

A second meta-analysis of 43 single pollutant model studies reported a pooled risk estimates for every 10-ppb increase in 1-hour daily maximum ozone concentrations for all-age non-accidental mortality 0.39%(95% CI 0.26-0.51%). The greatest effect was observed when PM was added to the model 0.40%(95% CI 0.27% -0.53%) vs. ozone alone 0.37%(95% CI 0.20-0.54%). The effects of PM with ozone require adjustment for temperature since the relationship with temperature is greatest for ozone. Including ozone in a model with PM without adjusting for the warmer to cooler temperatures may lead to particulates confounding the ozone mortality relationship. To explore this seasonal phenomenon, ozone and particulates, and (cooler and warmer) weather were examined. The all-age non-accidental mortality risk estimates for a 20-ppb increase in 24-hour average ozone was 2.2%(95% CI 0.8%-3.6%) for all-year data, and 3.5%(95% CI 2.1%-4.9%) for the summer months. This model demonstrated the effect temperature has on ozone, particularly during months where temperatures were higher. A second subset study included ozone with and without PM. Non-accidental mortality did not change with the all cities combined in the ozone model with PM vs. without PM, 1.5% vs. 1.6%, respectively. Further analysis was conducting by looking at the summer and winter months, with "ozone averaged for quintiles of PM " (p. 451) to determine if any correlation could be observed with PM and ozone. The results show a positive relationship between ozone in the summer months and a negative relationship in the winter. Thus, suggesting that PM does not affect ozone during the cooler temperatures (Ito et al., 2005).

Bell et al. (2007) continued to explore the relationship between particulates on ozone and mortality. Data from the 95 urban communities NMMAPS were tested using three different

methods. The first method attempted to determine whether any correlations existed between ozone and PM variables using different ozone concentrations accounting for geographical and seasonal differences. The Pearson correlation coefficient measurement for ozone by year, accounting for seasons were low for the US, PM_{10} (r = 0.25) and $PM_{2.5}$ (r = 0.22). Regional differences were observed with the highest correlations reported for the Industrial Midwest and Northeast for $PM_{10,}$ Southwest and Industrial Midwest for $PM_{2.5.}$ Winter months had the lowest levels of both particulates, with higher seasonal correlations observed in spring, fall and summer, respectively. The second method included the PM variables as covariates in a time series model, including the 1-day lag representation of the effect estimate. Corresponding PM_{10} and $PM_{2.5}$ risk estimates were 0.29(95% CI 0.03-0.55) and 0.22(95% CI 0.22 -0.65), respectively. Adjusting the model for PM resulted in an ozone mortality risk reduction per 10-ppb increase in the day-1 lag but the confidence intervals were wide and included negative numbers, PM_{10} [0.21(95% CI, -0.06 - 0.47)] and $PM_{2.5}$ [0.21(95% CI, -0.22- 0.65)]. Restricting the data in this model by adjusting for particulates limited the sample estimates, which reduced the power for finding a statistical association. In the third method, sample estimates were further restricted by using only a subset of the ozone data. An observation was made that the percent change in mortality was affected by PM at different ozone levels. However, one can not conclude that a significant relationship exists due to the reduced power in using a model with a limited sample.

Research studies have inferred an association between $PM_{2.5}$ and total mortality in a study conducted in Mexico City. A percent increase in total mortality was observed for every 10 ug/m^3 increase in PM_{25} , day -0 lag (1.34%) and (1.36%) on day-4 lag. The greatest association for total mortality (1.48%) was observed for the 5-day mean lag exposure for every 10 ug/m^3 increase in $PM_{25.}$ The elderly were the most sensitive population affected, with an increase of

1.6% total mortality noted in the individuals over 65 years of age (95% CI 0.04-3.12). The day-4 lag from $PM_{2.5}$ was also associated with an increase in respiratory and cardiovascular deaths, 2.50% and 2.19%, respectively. The percent increase in total mortality for ozone (mean day 1-2 lag) was 1.76% for cardiovascular causes. This same effect was not observed for ozone and respiratory causes -0.74%(95% CI -3.58-2.10) (Borja-Aburto et al., 1998).

In summary, these findings reflect the public health significance of the effects from short term exposure to ozone and particulates on total, cardiovascular, respiratory and total non-accidental mortality. The risk estimates are greatest for ozone at lag days 0-1 and 2, and a 2-day average lag for total mortality and cardiovascular and respiratory mortality, and 0-1 and 2 day single lag for non-accidental mortality. Effect estimates for particulates (PM_{10} and $PM_{2.5}$) have been observed with the greatest risk reported at 1-day lag and day-0 and day-4 for total mortality, 5-day average lag for $PM_{2.5}$ for total mortality, and day-4 lag for $PM_{2.5}$ and cardiovascular and respiratory mortality. Lag estimates represent the risk associated with short term exposure to these pollutants and risk of death. If these numbers were applied to a larger population, a great number of premature deaths could be prevented. Furthermore, these findings underscore the need to have more stringent regulations placed on ozone and particulates to protect the public's health.

There has also been a considerable amount of research done which links short term exposure to these pollutants with a rise in cardiovascular and pulmonary morbidity. Several of these studies report directly on the relationship between air pollution and asthma, and respiratory disease (Barnett, et al., 2005; Burnett et al., 2001; Lin et al., 2002; Pope, 1989; Luginaah et al., 2005; Villeneuve et al., 2007).

A correlation between elevated PM_{10} levels and increased hospital admissions for bronchitis, pneumonia, pleurisy and asthma were reported by Pope (1989). During time periods when the steel mill was open, hospital admission rates for children 0-17 years of age were 3 times greater during exceedance periods of the 24-hour standard (> 150 ug/m^3), and nearly doubled when the average PM_{10} level (50 ug/m^3) were equal to or greater than the standard. Hospital admissions during the same time period for adults had risen 44% and 47%, respectively for the same standards. The closing of the steel mill resulted in a reduction in levels of PM_{10}, which corresponded to a reduction in admissions for respiratory admissions. The reduction was more notable for asthma and bronchitis admissions and greater for children.

Lin and associates (2002) assessed the relationship between short term exposure to $PM_{2.5}$, $PM_{10,}$ and coarse particulates $_{10\text{-}2.5}$ averaged over 1-7 days for childhood asthma hospitalizations using bidirectional case-crossover, and time series analysis. No statistically significant relationship was observed for either $PM_{2.5}$ or $PM_{10.}$ However, a statistically significant relationship was observed for PM $_{10\text{-}2.5}$ with a consistent rise observed in asthma hospitalizations in both boys and girls 6-12 years of age. The rise continued up to day-6 average for girls and continued up to day-7 average for boys. The effect estimate for day- 6 average was (RR= 1.18) using both case crossover and time series analysis for girls, and (RR = 1.16) and (RR = 1.12) respectively, for boys. The resultant rise in asthma hospital admissions 16% and 12%, respectively, was based on the corresponding risk estimate for an increase of 8.4 ug/m^3 of $PM_{10\text{-}2.5.}$ Both methodologies yielded similar results for coarse particulates $_{10\text{-}2.5.}$

In another study, respiratory admissions differentiated by age and sex were reported as they relate to the exposure from multiple air pollutants. The study used a case-crossover methodology and time series analysis to conduct the experiment. For both methods, a higher

number of respiratory admissions were observed in female's ages 0-14 and 15-64, than in males. In the case-crossover analysis for NO_2,hospitalizations for females 0-14 years of age increased by 19%, (p = < 0.05) per 16-ppb increase in the 24-hour NO_2 levels. No significant association was observed within any other female age group for NO_2, but was present for CO in females in the same age group. Effects estimates from CO were statistically significant on lag day 0-1 and 2 using the case crossover analysis (RR = 1.15), (RR = 1.19), (RR = 1.22), respectively, and time series analysis day-1 lag (RR = 1.06), (p = < 0.05). The time series analysis was the only methodology which reported statistically significant findings for SO_2 on day-0 lag in the 0-14 year old age group (RR = 1.11) (p = < 0 .05). For all females, COH effects were present on day-0 lag (RR= 1.09) using case crossover, and day-2 lag (RR=1.06) using time series analysis (p = < 0.05). An effect for CO day-2 lag was also reported in the 15-64 year old age group (RR = 1.15), (p = < 0.05). The only effect observed for males was in the 15-64 year old age group for PM_{10} at day -1 lag (RR=1.17), (p = < 0.05). Females appear to be at a disadvantage for respiratory admissions from air pollution. However, further investigation into factors such as biological differences with age and sex, and air pollution exposure needs to be conducted before any conclusions can be drawn from these research findings (Luginaah et al., 2005).

A case-crossover study conducted by Barnett and associates (2005) examined the effects of short-term exposure to pollutants ($PM_{2.5}$, PM_{10}, O_3, SO_2, and NO_2) on daily counts of childhood respiratory hospital admissions in children <1, 1-4, and 15-14 years of age. Hospital admissions significantly differed across cities, for diagnosis and age group. A percent increase in respiratory hospitalizations was observed in the 0 age group 2.5%(95% CI 1.1-4.0), and 2.2% (95% CI 1.2-3.2) in the 1-4 age group for particulates 0.1-2 *u*m in diameter. For individuals in the 5-14 year old age group, a 6% increase was observed for asthma admissions, which

correlated to a 5.1 ppb increase in the 24-hour NO_2 level. When the data was split to explore the seasonal (cool and warm) effects on asthma hospital admissions, warmer months generally had higher admission rates than cooler months. A 10.2% increase was observed in asthma admissions for the 5-14 year old age in the warmer months compared to seasons with cooler temperatures (7.2%).

Villeneuve and associates (2007) conducted a case crossover study to examine the effects of ambient air pollution on ED visits for asthma. The analysis was conducted for all months and by season for each different pollutant (SO_2, NO_2, CO, $PM_{2.5}$, PM_{10}, and O3). This analysis differentiated results based on age. The analysis that included all months demonstrated that no significant effect was observed during the winter months for these pollutants. Data representing the summer months resulted in statistically significant findings for NO_2, CO, $PM_{2.5}$, PM_{10} and O_3. The greatest risk was found using the 5-day average lag for each pollutant. Children 2-4 years of age and the elderly were the most sensitive to CO [day-5 average], (OR= 1.48 and 1.54), respectively, and for NO_2 [day-5 average] in these same children, (OR= 1.50), ($p = < 0.05$). A significant relationship was observed for both particulates at 3 and 5-day average lags ($p= < 0.05$) for all individuals in the all seasons and warm temperature model. A significant relationship was also observed with ozone during the all seasons and summer month's model using the 3 and 5-day average lags. The risk was more evident in the summer months with asthma admissions increasing by 11% for all individuals ($p = < 0.05$). Surprisingly, the effect estimate for the youngest (2-4 years of age) and oldest age group (75±) did not result in any significant findings. Statistically significant differences were only observed for individuals 5-14 years of age for the all months and summer month model using the 3 and 5-day average and the 1-day lag for the summer month model ($p = < 0.05$). All ozone models were statistically

significant in individuals 15-44 years of age. Only the all seasons model for individuals 45-64 years of age at the 5-day average was significant ($p = < 0.05$).

The strong seasonal effect of ozone was also observed by Burnett et al (2001) in a group of children less than two years of age. A risk for being admitted to the hospital for a respiratory condition was increased by 35% from May through August. The increase in the number of admissions corresponded to an association of 45.2 ppb increase in the 1- hour daily maximum ozone concentrations. The 5-day moving average for May- August was 34.8, April-September - 4.9, and January- December 16.1. These effect estimates demonstrate that the correlation between temperature and ozone is great. No relationship was observed between April-September when summer months were not included in the model. The January through December model that included the summer months was high but was the effect was reduced by including the winter months in the analysis. The very young are vulnerable to ozone and this conclusion is reflected by the reported increase noted for hospital admissions (35%) during the summer months-May through August when children < 2 years of age are at greatest risk.

These studies provide evidence of the relationship between exposure to ambient pollutants, including ozone and fine particulates and respiratory morbidity, including asthma. A temporal relationship exists between exposure and outcome. The biological response resulting from these exposures is impacted by a time scale based on a pollutant threshold. The point estimates for lag days that are associated with a statistically significant risk ($p = < 0.05$) are highlighted in Table 2**.**

Table 2: Point estimates for lag days associated with respiratory and asthma morbidity

Reference	Pollutant	Lag estimates	Morbidity
Luginaah et al., 2005	$PM_{2.5-10}$	Day-6 average Day-7 average	Asthma admissions
Luginaah et al., 2005	PM10	Day-1	Respiratory admissions
Villeneuve et al., 2007		Day-3 average, Day-5 average	Asthma ED visits
Villeneuve et al., 2007	PM2.5	Day-3 average, Day-5 average	Asthma ED visits
Luginaah et al., 2005	CO	Day-0-1-2,	Respiratory admissions
Villeneuve et al., 2007		Day-5 average	Asthma admissions
Luginaah et al., 2005	COH	Day-0-1-2	Respiratory admissions
Luginaah et al., 2005	SO2	Day-0	Respiratory admissions
Villeneuve et al., 2007	NO2	Day-1 Day-5 average	Asthma admissions
Villeneuve et al., 2007	Ozone	Day-1 Day-3 average, Day-5 average	Asthma ED visits
Burnett et al., 2001		Day-5 (moving) average	Respiratory admissions
Paulu & Smith, 2008		Day-4 average	Asthma ED visits

Since Allegheny County exceeds the standard for the annual and daily $PM_{2.5}$ standard and 8-hour ozone standard, it would be important to examine the role of these pollutants either alone or in conjunction with each other as they relate to asthma events. This type of study would augment the knowledge of practitioners in the area by providing information about a potential link between exposure to air pollutants and asthma. Information as such can be used by the public health official in a health alert to reduce behavior in sensitive populations when risk may be elevated.

3.0 CHAPTER 3- CRITICAL REVIEW OF LITERATURE

3.1 SURVEILLANCE

Disease surveillance is an essential component of public health. Historically, several different methodologies have been used to detect and stop the spread of disease (Lombardo & Ross, 2007). Crude forms of surveillance lead to the development of vital statistic records that were used to collect information systematically for births and deaths (Gostin & Hodge, 1998). The registration of morbidity data and reliance on these data by public health officials, eventually lead to the mandatory collecting of data for infectious diseases, and the classification of reportable conditions, which were considered to be a threat to the public's health (Lombardo & Ross, 2007; Gostin & Hodge, 1998). Currently, public health surveillance has moved beyond collecting data solely for infectious diseases to include information on chronic diseases, which accounts for the majority of the mortality and morbidity in the United States (Pew Environmental Health Commission, 2000).

Surveillance is the first essential service of public health (CDC: National Public Health Performance Standards [NPHPS]: Ten Essential Services). In order to carry out this service, key data fundamental to public practice needs to be collected and analyzed on a routine basis. Without such data to perform these analyses, the core functions of public health, assessment, policy development and assurance can not be carried out (Glad, Kotchian, Barron, 2004).

Gaps in surveillance capacity highlight the need to have standardized information routinely collected on "human exposures, environmental hazards, chronic disease and other health related health endpoints" (IOM, 1988; Pew Environmental Health Commission Report, 2000 as cited in Glad et al, 2004, p. 9). An integrated public health tracking system at the national level would provide the infrastructure to have access to these types of data to gain a better understanding of the "trends, causes, and environmental factors influencing disease" Pew Environmental Health Commission Report, 2000, p. 6). As the Centers for Disease Control and Prevention, National Environmental Public Health Tracking Program (2006) continues to build on this national initiative, public health practitioners continue to look for new ways to collect key data beyond the traditional method of using data collected from passive reporting systems such as data from physicians and laboratories who are fulfilling their reporting requirements imposed by the law (Schuchter, 2003).

The electronic medical record is one methodology being recommended for comprehensive data collection. The electronic feature provides quick access to information captured when the patient arrives for care. The automated feature also creates a stream of clinical data, which can be integrated into existing data systems for routine analysis or monitored by health care professionals at off site locations. Furthermore, the application of data mining tools can be applied to extract data from these records for scientific interpretation and visualization. These features alone underscore the need to consider this methodology for asthma surveillance.

3.1.1 Electronic medical records

The EMR is a rich source of health information that is slowly making its way into modern day health care (Coiera, 2003, Chapter 10). Included in the medical record are the patient demographics, past medical history, admission and progress notes, laboratory results, pharmacy notes, and discharge summaries. The arrangement of these documents into one computerized system eliminates the need for the conventional paper based record. The EMR was described in the 1997 Institute of Medicine Report: *The Computer-Based Patient Record: An Essential Technology for Health Care* (IOM, 1997) as "a type of clinical information system, which is dedicated to collecting, storing, manipulating, and making available clinical information to the delivery of patient care. The central focus of such system is clinical data and not financial or billing information. Such a system may be limited in their scope to a single area of clinic information (e.g., dedicated to laboratory data), or they may be comprehensive and cover virtually every facet of clinical information pertinent to patient care".

The EMR provides several advantages over the paper based medical record which includes (a) ease of access to information in a timely, automated format; (b) improved decision-making ability for the practitioner due to the availability of comprehensive information; and (c) use of data to monitor and report on disease outcomes (Coiera, 2003, Chapter 10). Furthermore, data mining tools can be used to extract data from the electronic medical record for quantitative evaluation (Krauthammer & Hripcsak, 2001; Friedman et al., 2004; Hripcsak et al., 2002; Chuang, Friedman, & Hripcsak, 2002; Friedman, 1997; Melton & Hripcsak, 2005; Chen et al., 2008).

The use of the EMR for surveillance improves efficiency of reporting since the burden to report a diagnosis is removed from the physician and replaced by an automated process (Lazarus,

et al., 2001). A classification code can be used to extract a diagnosis from the EMR for disease ascertainment. The automation features further allows data to be culled from multiple data sources for integration into a customized data system and used by healthcare agencies to fulfill reporting requirements for disease and surveillance (Pennsylvania's National Electronic Disease Surveillance System [PA-NEDSS]: 33 Pa.B.2439).

Several epidemiological studies have demonstrated the utility of using the electronic medical record (EMR) as a surveillance tool for detecting outbreaks of infectious diseases (Lazarus, et al., 2001; Panackal et al., 2002), assessing potential bioterrorism events (Lombardo, et al., 2003; Barthell et al., 2002), and enumerating clinical outcomes (Donahue et al., 1997; Hung, et al., 1998; Vollmer, et al., 2004).

The University of Pittsburgh Medical Center study conducted by Panackal et al. (2002) demonstrated the accuracy of using electronic reporting over the paper based methodology for confirming cases of infectious diseases. Of the two methods tested, the electronic reporting system was 74%(95% CI 66%-81%) more accurate than the paper-based reporting system 65% (95% CI 57%-73%), ($p = < 0 .05$) for confirming disease. The use of an automatic reporting system not only increased timeliness, but outperformed the paper based reporting method. To improve the completeness of data for electronic reporting, the authors recommend minimizing the use of free text information.

Military and healthcare organizations have used automated systems to group clinical symptoms, which may reflect patterns indicative of a bioterrorism-associated condition (Barthell, et al., 2002). A system of this type could be used to alert the medical staff if the frequency of signs and symptoms reached detectable levels above a normal standard (Wagner et al., 2004). Moreover, this system could alert emergency personnel as to a timelier response and recovery

plan. Early warning syndromic systems can serve many purposes, including the benefit added to the public health practitioner. For example, a syndromic system can be used to alert the epidemiologist about seasonal events such as influenza and allergies to communicating daily surveillance data to local county health departments (Lombardo et al., 2003).

A prototype for a syndromic surveillance system used in Allegheny County is the Real Time Outbreak and Disease Surveillance (RODS) system. This system has been operational for approximately 9 years. The input in the system is collected from the patient as they enter the hospital system. As presented in the patient's words, the chief complaint captured in the emergency department is used to detect disease clusters formulated by the syndromic classification of the chief complaint (Wagner, et al., 2004). Once classified into a disease category, the RODS system uses a "signally out" process to measure data in time and space to predict disease. Evaluation of these data is based on the CDC recommended guidelines for evaluating a syndromic surveillance system (MMWR, 2004). The assumption with RODS is based on the likelihood that the system can predict that a problem exists above the normal limits, and detects the problem in a timely manner. The accuracy in meeting these assumptions is evident by the early detection of gastrointestinal outbreaks reported in both adults and children where RODS picked up the problem 10 and 11 days, respectively, before the event and in a most recent event alerted public health officials about a rise in the number of daily respiratory cases which correlated to an environmental exposure from carbon monoxide.

To improve ways for positively identifying a disease, Donahue et al. (1997) recommends reviewing multiple databases or record types to confirm a positive diagnosis. For example, individuals diagnosed with asthma are 3 times more likely to be on an asthma medication. An acute visit to the emergency department for asthma with an order for a nebulizer treatment and

no other asthma medication criteria has been observed to be 100% predictive of a diagnosis of asthma vs. an urgent visit for asthma without any medication criteria (80%) (Volmer et al., 2004). An order for a nebulizer treatment without any further evidence to substantiate the asthma diagnosis reduces the positive predictive value (27%). Solely relying on the pharmacy records alone to confirm a diagnosis of asthma without further clinical information to substantiate the diagnosis would result in a lower sensitivity of findings. Individuals with a less severe form of asthma would likely not have been identified by just reviewing the pharmacy record in this study since individuals who have a milder form of asthma do not use asthma medications on a regular basis. This was evident in the study conducted by Volmer et al. (2004) where a very low positive predictive value was observed (27%) for a nebulizer treatment order without any further substantiating factors to back up an asthma diagnosis. Thus, searching multiple data types for a longer time periods can substantially improve the validation of an asthma diagnosis.

The ability to identify a patient with asthma is difficult without a physician diagnosis. The establishment of one standardized case definition or combination of symptoms to predict asthma is being considered (Sistek et al., 2006; Sistek et al., 2001; Pekkanen, et al., 2005; Sanders et al., 2007; Yu, Wong, & Li, 2004).

Pekkanen and associates (2005) suggest that to obtain a high probability (OR >1.0) or threshold rating for asthma, grouping signs and symptoms or using them in combination with an objective measure such as bronchial hyperresponsiveness (BHR) may be useful for epidemiological research. However, the researcher has to consider the cut-off point to use when combining asthma symptoms. For example, sensitivity for asthma has been shown to decline when multiple symptoms are combined, wheeze alone (63%), wheeze with breathlessness (48%),

and wheeze, breathlessness and no cold (38%). Specificity on the other hand is high 82%, 93% and 92%, respectively. A more stringent definition has a higher specificity and a lower sensitivity. Thus, including a single definition for asthma with single symptoms improves both sensitivity of findings and specificity.

In Sistek et al., (2006) a combination of respiratory symptoms were analyzed to determine their performance at predicting a diagnosis of asthma, and whether the addition of BHR improved the predictive value (Table 3). Wheezing with dyspnea overall had the highest discriminatory value for predicting a diagnosis of asthma with a sensitivity of 82% and 90% specificity. The positive predictive value was 42.7%. Even though the sensitivity for isolated wheeze was higher (93.4%) than wheeze with dyspnea (82%), the specificity indices were lower 76.4% vs. 90%, respectively, and the positive predictive value was lower at 26.4%. Symptoms of nocturnal and rest dyspnea and nocturnal chest tightness had rather low to moderate sensitivity (41-75%) and specificity greater than 85%. The inclusion of BHR with symptoms resulted in a high specificity for all symptoms > 90%, ($p = < 0.0001$) and lower sensitivity (< 83%) for all symptoms except chronic phlegm, nocturnal dyspnea, and chronic bronchitis (not highlighted in Table 3). Broncho hyperresponsiveness was found in both persons with asthma (85.2%) and without (19.5%). The estimated risk based on the highest predictive values for the inclusion of BHR with symptoms was greatest for wheeze with dyspnea and (62.7%) and rest dyspnea (61.1)% (Table 3). However, the inclusion of BHR with symptoms reduces predicting who has disease and increases the knowledge of who does not.

Table 3: Diagnostic value of symptoms in the presence and absence of a physician diagnosis of asthma: Survey and chief complaint data

Study	Age	Physician diagnosis of asthma	Wheeze	Cough	Shortness of breath /dyspnea	Fever	Other	Sensitivity %	Specificity %	Positive predictive value %
1. Community Survey-self reported symptoms against doctor's diagnosis (Yu et al., 2004, Table 1, p. 456).	8-12	Yes		X				63.3	75.3	14.1
							X Phlegm	52.0	69.9	10.0
			X					59.2	92.7	34.1
			X	X			X Phlegm	33.3	97.6	46.4
			X	X				45.4	96.5	44.9
2. ED visits using chief complaint and ICD 9 codes: characteristics for five chief complaints (Sanders et al., 2007, Table 3, p. 531).	2-18	Yes with chief complaint	X					27	99	96
				X				4	99	67
					X			10	99	84
						X		3	95	33
			X	X	X	X		44.8	91.6	79.3
		Chief complaint alone	X					56	97	94
				X				16	82	39
					X			21	91	64
						X		6	29	6.2
3 Predictive values of respiratory symptoms in the clinical diagnosis of asthma (Sistek et al., 2006, Table 1, p. 2109).	20-44	Yes					X BHR	84.6	80.5	28.2
			X		X		X BHR	72.3	96.1	62.7
					X rest		X BHR	33.9	98.1	61.1
			X					93.9	76.4	26.4
			X		X			81.5	90.1	42.7
			X	X				76.9	86.8	34.5
				X				43.1	83.9	19.4
							X nocturnal chest tightness	75.4	85.1	31.4
					X nocturnal			41.5	95.8	47.4
					X rest			43.1	92.9	35.4
4. Clinical diagnosis of current asthma: predictive value of respiratory symptoms (Sistek et al., 2001, Table 4, p. 217).	18-60	Yes	X					75	87.3	12.4
			X in combination with one symptom		X nocturnal			37.5	98.6	39.3
					X rest			38.4	98.7	42
					X exercise			54.2	95.7	23.1
							X nocturnal chest tightness	40.9	97.5	28.1
			X alone with 2 nocturnal symptoms	X	X		X chest tightness	80	85.9	11.9

A cross sectional survey was conducted by Yu, Wong & Li (2004) to determine how well child reported respiratory symptoms with and without spirometry could predict a diagnosis of asthma. An observed difference was noted by gender with boys having over twice the prevalence estimates than girls (8.2% to 4.1%), respectively, and a higher estimate was found in the older children 10-12 years (6.5%) vs. the 8-9 year olds (5.5%). Gender difference were reported for isolated respiratory symptoms which showed that more boys than girls had cough (17.0% vs. 10.7%), wheeze (36.0% vs. 31.4%), phlegm (13.2% vs. 6.5%), and in combination, cough and wheeze (44.4% vs. 45.7%), and cough plus wheeze and phlegm (46.8% vs. 45.5%), respectively. Seventeen models were created to look at predictors of asthma based on the isolated respiratory symptoms with spirometry measurements. Wheezing was highly predictive for asthma 59.2% (Table 3). Less than 8% of the cohort who did not have asthma had wheezing. A combination of wheeze, plus cough and phlegm had a positive predictive value of 46.4%. Less than 2% of this combination group with phlegm had asthma. When phlegm was removed from the model, sensitivity increased to 45.4% and the positive predictive value slightly increased but not significantly (44.9%). Thus, phlegm is not a good predictor of asthma (Table 3). A cutoff for spirometry at FEV_1:FVC < 75 had the highest positive predictive value for asthma 22.9%. The three symptom model of wheeze with cough, and phlegm and FEV1:FVC < 75 had the highest predictive value at 83.3%. Removing phlegm from the model increased the positive predictive value from 88.3% to 88.9%. The same indices were observed with wheeze alone and FEV_1:FVC < 75, the positive predictive value was 88.9%.

A review of emergency room visits for asthma in a pediatric population found that the chief complaint of wheezing, shortness of breath, fever, cough and dyspnea were recorded in greater than 95% of all asthma visits (Sanders et al., 2007). Fifty six percent of all asthma visits

were due to wheezing, 21% dyspnea, 16% cough, and 6% fever (Table 3). Each isolated chief complaint included in a model with a diagnosis of asthma was highly specific (99%). For all chief complaints combined (wheeze, cough, shortness of breath, and fever) with a diagnosis of asthma, a positive predictive value of 79.3% was reported. The most discriminating chief complaint for asthma was wheezing with a positive predictive value of (96%), followed by dyspnea (84%). Even when no diagnosis for asthma was reported, wheeze was the most frequent reported chief complaint (56%), specificity (97%) and a positive predictive value of 94%.

In Sistek et al. (2001) data was collected on 9,651 individuals 18-60 years of age. Data from a community questionnaire was analyzed to determine the positive predictive value of a single symptom and a combination of symptoms to predict a diagnosis of asthma. Wheezing was reported in 75% of the asthma cases (Sistek et al., 2001). Less than 4% of the population who did not have asthma had wheeze with one of the following symptoms, rest dyspnea, nocturnal dyspnea, exercise dyspnea and nocturnal chest tightness. Combining wheezing with any two of the nocturnal symptoms, chest tightness, shortness of breath and cough was found to have the highest sensitivity and specificity (80.5% and 85.9%), respectively (Table 3).

3.1.2 Classification of clinical free text data

The increased availability of the EMR subsequently increases the need to access the data using some form of knowledge model to transform the data for scientific interpretation (Manning & Schuetze, 1999; Krauthammer & Hripcsak, 2001; Friedman et al., 2004; Hripcsak et al., 2002; Chuang et al., 2002; Melton & Hripscak, 2005; Chen, Hripcsak, Xu, Markatou, & Friedman, 2008; Visweswaran, Hanbury, Saul & Cooper, 2003; Cooper, Buchman, Kayaalp, Saul & Vries, 1998). One approach to aid the researcher in the automated text classification of the data is to

use a form of machine learning to employ algorithms and techniques that, "build classifiers by learning from a pre-classified set of documents, the characteristics of the categories" (Sebastiani, 2007, p.1).

Visweswaran and associates (2003) use this approach to build a computer model by first classifying words, phrases and medical language concepts to determine which algorithms perform better at identifying adverse drug related events found in the EMR discharge summary. The standard method to identify an adverse drug related event is through a medical chart review. However, this method is costly and time consuming. By training a model using machine learning algorithms, the computer is able to pick up the relevant text data related to an adverse drug related incident. These data from the original set of records are characterized and reapplied to a subsequent set for classification. A continual "search and review" of the records is conducted to further locate words associated with an adverse drug related events (Cooper et al., 1998, p. 182).

A similar model was developed by Cooper and associations (1998) to identify patients in an intensive care unit who were diagnosed with deep vein thrombosis (DVT). A search of the EMR using criteria to identify patients of interest was the first step of the study. Once patients were classified during the "search and review" process it was hypothesized that new patients would be identified given a new search cycle was initiated. The use of this methodology demonstrated the effectiveness of using a machine learning model to identify the cases of DVT correctly in the study, and continued to outperform the model by identifying additional cases.

Natural Language Processing (NLP) operates through principles of machine learning to interpret data from clinical narratives and transforms it into a language which computers can understand (Manning & Schuetze, 1999). Natural Processing Language tools are used to apply a

text classifier to the clinical information through a set of classifying codes that extract clinical data from the EMR. The mechanism for translating the codes are established through an expert based classification system such as the Unified Medical Language System (UMLS). The classification system allows data found in the EMR to be linked through this web-based process and translated for understanding of the medical terminology (Coiera, 2003, Chapter 17). The UMLS is the infrastructure, which supports a network that is made up of a metathesaurus, semantic network, and lexicon (Lindberg et al. 1993 study cited in Coiera, 2003). The metathesaurus is a very large database made up of biomedical vocabulary and classifications that include the ICD-9 and ICD-10 coding systems. The semantic network is the input classifier that assigns concepts from the metathesaurus into certain categories (National Library of Medicine, Semantic Network fact sheet). The relationship between the categories and terminology concepts provide the structure of the semantic network. The lexicon is used to assist with NLP to code terminology into concepts for interpretation and meaning (Coiera, 2003, Chapter 17).

One natural language processor used by the scientific community is Medical Language Extraction and Encoding (MedLEE). This processor is used to extract text data from the medical record and translate the data into terms identified in the UMLS (Friedman et al., 2004; Chen et al., 2008; Coiera, 2003, Chapter 17). The text data is encoded and transformed into a variety of formats including, an XML output. Researchers have used MedLEE to extract information on the frequencies and the occurrence of clinical conditions from chest radiography reports (Hripcsak et al., 2002), identify patterns between drug - disease relationships (Chen et al., 2008), parse out a broad range of clinical information from discharge summaries (Krauthammer & Hripcsak, 2001) and, used as a screening tool to identify adverse medical outcomes reported in the New York Patient Occurrence Reporting Tracking System (Melton, & Hripcsak, 2005).

MedLEE has been in existence for over 20-years and serves as the industry standard for the medical community. Moreover, MedLee has been reported to be as accurate to that of physician review. An agreement of 95% was observed for MedLee's matching the physician review in accurately identifying the presence and absence of findings in the clinical narrative (Krauthammer & Hripcsak, 2001).

Another type of text processor, which can be used when large amounts of data need to be processed, is Practical Extraction and Reporting Language (PERL), a command line driven processor. This processor uses a regular search expression engine to encode pattern matching functions with automation (Wall, Schwartz, Christiansen, Potter, 1996). A powerful feature of PERL is its ability to conduct text processing with little manipulation of the input stream. When other programs require multiple lines of code to execute a simple function, PERL often times requires only one or two lines of code to complete the function. The programs ability to specifically match on any character, types or groups of characters, in any order found in the document suggests that this program may have utility in analyzing free text data from the electronic medical record (Breeding, 2002). Another feature of PERL is its ability to use a full complement of search expression for patterns for matching functionality, as well as logical and string manipulation capabilities of a procedural programming language. PERL compilers are available for UNIX, Lenox, Windows and Mac operating systems. However, the principal disadvantage of this application is that full knowledge of the codes and command structure within PERL is required in order to operate the text processor. Furthermore, PERL can not account for misspelled words, the plural form of the word, or double negative terms. The accuracy of findings may be further reduced by these weaknesses.

Another form of search agent, which can be used to analyze large amounts of text data, is the ATLAS.ti software. This application operates with a command line driven programming search that can analyzes data found in a textual, graphical, or video format (Muhr, 2003). This software agent uses an easy "drag and drop" approach to code, merge and link the data (ATLAS.ti-The Knowledge Workbench: Features). Searches are conducted using a modified version of Global Regular Expression Print (GREP), a form of line text coding where data outputs match each line. A major advantage of the software is its ability to add structure to the data, and transforms the data so relationships between different parts of the data can be visualized. Once the data has been extracted from the file it can be expressed as a single word, phrase, or classified into hierarchical trees or decision trees. Data can then be exported in a variety of formats, which include SPSS, HTML, XML, and CSV.

A disadvantage of using this type of software with the electronic medical records is the lack of ability to annotate the medical record in a way, which allows the software to search one record at a time. The electronic medical record is made up of a group of documents (a) narrative clinical reports, including free text history and physicals; (b) dictated physicals; (c) consultations; (d) clinical follow up visits; (e) emergency room notes; (f) laboratory results; (g) surgery reports, and; (h) discharge summaries. Each document contains a unique identifier number to represent the patient, a visit identifier to reference the visit, the date in which service was rendered and the record type, which classifies the type of visit through which a service was prescribed. An annotation "E_O_R" is found at the end of each record to indicate where the record ends. The limited line by line searching function in ATLAS.ti restricts the user from being able to use a full term document search by sentences, paragraph or sections within the record. The specificity of information captured is further reduced by the lack of understanding

how to retrieve information using a more complex search expression for a status, which can not be determined. Without a clear understanding of GREP, and knowledge on how to bypass the limited line by line searching feature of this software, the investigator is hesitant about suggesting the use of this software with EMRs based research.

There are several knowledge based applications suitable for mining data from the electronic medical record. However, each product comes with its own particular strengths and weaknesses. Each application should be tested and difficulties presented. The known successes and failures should be carefully weighed against the project under consideration.

3.1.3 Value of the ICD 9 Code

The systematic process of classifying diseases was first reported in the 18th century. The progress in preventive medicine has lead to the development of an international classification system, which not only consisted of a list of deaths, but included a corresponding one for disease (Coiera, 2003, Chapter 17, Section). The International Classification of Disease (ICD) system has become the standard way to collect and analyze mortality and morbidity data worldwide (World Health Organization [WHO], 1977) Since its inception, the classification system is in its tenth revision. Several countries including the U.S. have not fully adopted the 10th Revision. Proposed regulation will replace the 9th revision with ICD 10 code sets, effective October 1, 2011 (American Academy of Professional Coders, 2008). However, for the purpose of this research the disease under investigation has been classified using an ICD 9 code [ninth revision] terminology.

The ICD code is accepted as a "defacto reference point for many healthcare terminologies," (Coiera, 2003, Chapter 17, Section 17.1). Administrative classification of the

ICD-9 code has played a critical role in health care history, particularly for reimbursement (O' Malley, et al., 2005). Healthcare agencies use the ICD-9 codes as an indicator to monitor work volume, measure disease, and as a method to establish reimbursement for diagnosis (Lazarus, et al., 2001; Espino & Wagner, 2001; O'Malley, et al., 2005) More importantly for the epidemiologist, the use of the ICD-9 code can be used to study disease for a single diagnosis or grouping symptoms based on common diagnoses (Lazarus, et al., 2001; Lombardo et al., 2003; Sanders et al., 2007; Vollmer, et al., 2004; Donahue, et al., 1997; Hung et al., 1998). Medical researchers have successfully demonstrated the use of ICD- 9 coded chief complaint to detect the onset of gastrointestinal outbreaks (Wagner, et al., 2004) and respiratory related conditions (Espino & Wagner, 2001). Several studies have demonstrated the accuracy of using the ICD-9 coded physician diagnosis of asthma as a gold standard in conducting epidemiological research (Vollmer, et al., 2004, Sanders, et al., 2007). O'Malley (2005) provides a review of the limitations identified in the coding process, which may affect the reliability of the ICD 9 coded physician diagnosis. The limitations are defined in terms of a continuum from the time the patient presents to the emergency room until the time of discharge. The ICD 9 coded physician diagnosis is assigned following discharge from the hospital. The following information highlighted in Table 4 summarizes the potential sources of error along this continuum.

Table 4: ICD 9 Coding Process: Pathway for misclassification

Measures of Diagnostic Accuracy	Impacting factors
Certainty or accuracy of diagnosis	• Patient, physician and medical staff knowledge about the disease • Type of disease • Current state of medical technology • Translation of information into code
Communication	• A patient needs to be a good historian • Providers need to ask appropriate questions and obtain a full medical history. • Physicians need to understand the patient
Testing and diagnostic accuracy	• Presentations and etiologies of disease need to be understood. • Physician knowledge of best test and procedures • Availability of tests and procedures, and results interpreted correctly (sensitivity, specificity and predictive value).
Change in medical knowledge or testing	• Medical progress improves the ability to make diagnosis (i.e., stroke confirmed through non invasive testing using a brain MRI instead of a spinal tap or arteriography) (Provenzale et al. 2003 study cited in O'Malley et al., 2005 p.1629).
Medical record variance in terminology	• Omission of medical information from chart by staff and physician) • Non legible physician hand writing • Imprecision (use of synonyms)
Coder error variance	• Change in the coding classification system to include alphanumeric codes • One diagnosis code selected by the coder for a physician diagnosis. • Coder experience, and training [American Health Information Management Association-credentialed] • Miscoding, resequencing of the codes (diagnoses are reversed), and upcoding for higher reimbursement

Source: Journal of Health Service Research (October 2005); 40(5 Pt 2); p. 1625-1628

Although, the use of ICD coding is a complex process, awareness of coding accuracy needs to be taken into account when conducting research using the ICD 9 coded physician diagnosis. Particularly for a disease such as asthma whose differential diagnosis can greatly vary.

3.1.4 Diagnostic accuracy

Currently, there is no operational definition for asthma (Pekkanen et al., 2005). In order to conduct epidemiological research on asthma a standard case definition is needed to improve comparability across populations (CSTE, 1998). The (1998) Council for State and Territorial Epidemiologist (CSTE) case classification definition for asthma allows states to collect data on mortality and morbidity through the (1) Hospital Discharge, (2) Clinical and Laboratory case classification definition and (3) Prevalence survey. The classification status for definitions 1 and 2 allows one to determine the number of asthma cases by reviewing the health service utilization records, and the clinical and laboratory notes of the medical record (Table 5). The CSTE classification definitions for asthma are highly sensitivity and capture all levels of asthma from mild to severe. In case definition 3 the prevalence classification, a diagnosis of asthma is obtained through a patient's self report of the diagnosis, which relies on a patient's memory and understanding of the questions pertaining to ascertainment of a diagnosis. The CSTE recommends validating the accuracy of the self survey response through other reporting measures.

Another classification system widely used to measure performance of care and service is the National Continuous Quality Assurance (NCQA), Healthcare Effectiveness Data and Information Set (HEDIS). A classification definition exists for asthma, which is used by insurance companies to make comparisons across agencies. The HEDIS measure also relies on

pharmacy and medical utilization claims data and is more specific than the CSTE case classification definition for asthma (Lichter, 2004).

The CSTE case definition presents three methods for conducting asthma surveillance. Each method requires a baseline understanding of the health care utilization practices of individuals with asthma. These data could not be obtained unless a preliminary review of the medical record and secondary data sources were conducted.

Table 5: Council and State Territorial Epidemiologist: Asthma surveillance and case definition

Case Definition	Description

1. Hospitalization Discharge Classification

    - ICD-9 diagnosis codes 493.0 through 493.9 inclusive; (J45.0-J45.9) as primary diagnosis or secondary.

2. Clinical and Laboratory Classification

    Definitive Clinical Criteria

    - Wheezing lasting 2 consecutive days or more,
    - Chronic cough that responds to bronchodilatation that persists 3-6 weeks in the absence of allergic rhinitis sinusitis
    - Nocturnal awakening with dyspnea, cough, and or wheezing in the absence of other medical conditions known to cause these symptoms.

    Definitive Lab Criteria

    - PFT (spirometry, FEV1, FVC) demonstrating a 12% increment after the patient inhales a short acting bronchodilator

    - 20% decrement in FEV1 after a challenge by histamine, methacholine, exercise or cold air 20% diurnal variation in peak expiratory flow over 1 to 2 weeks

        I. Confirmed -Met any clinical symptom at least 3 times during the past year AND at least one of the laboratory criteria.

        II. Probable meets any of the following:
        - In the absence of supporting laboratory criteria, presence of any of the clinical symptoms which have been reversed by physician, treatment with asthma medications and have occurred at least 3 times during the past year.
        - In the absence of supporting clinical criteria met at least one of the laboratory criteria during the past year.
        - In the absence of supporting laboratory or clinical criteria, taken medications in the past year that was prescribed by a physician for asthma.

        III. Possible meets any of the following:
        - Shortness of breath (SOB) on exertion
        - Presence of wheezing or chronic cough in the absence of obvious respiratory infection
        - Presence of increased nasal secretion, mucosal swelling, nasal polyps or chronic sinusitis
        - Hyper expansion of the thorax
        - Wheezing during normal breathing
        - Prolonged phased o forced exhalation
        - Chest x-ray showing hyper expansion

(Table 5 Continued)

- FEV1 less than 80% of predicted value

Case Definition	Description
3. Prevalence Classification	
I. Confirmed	
• There is no confirmed case classification for self-report. Health departments are encouraged to validate the accuracy of survey self-response data.	
II. Probable:	
• A positive response to the survey question, "Did a doctor (or other health professional) ever tell you (or any household member) that you (they) had asthma?" AND	
A positive response to any of the following survey questions:	
"Do you (or the household member) still have asthma?"	
"Have you (or any household member) taken prescription medications for asthma (such as albuterol, inhaled steroids, cromolyn, theophylline, etc) during the past year?"	
"Have you had a wheeze episode in the past year?"	
III. Possible:	
• A suspect case meets any of the following: A positive response to survey question "Have you (or any household member) used over-the-counter medications for asthma during the past year?"	
Positive response to survey question, "Have you (or any household member) experienced episodes of wheezing during the past year?	

Table 5: Reproduced from the Council and State Territorial Epidemiologist. Asthma Surveillance Case Definition. CSTE position statement, 1998.

3.1.5 Healthcare utilization

The different types of health care services utilized by individuals with asthma are an important component to understanding this diagnosis. Particular to this work, an assessment of these health services for asthmatics in Allegheny County (AC) was needed to make meaningful interpretation of the results of the proposed research. This background review was based on estimates reported in a descriptive study conducted in 2004 by the PITT-PM Health Outcomes Project Team (Talbott, Zborowki, & Bilonick, 2007; [unpublished report]). A review of the (1994-2004) data from the Pennsylvania Health Care Cost Containment Council (PHC4)* was used to establish a baseline estimate of the proportion of persons living in AC who were admitted to the hospital for asthma. There was 364,244 admissions reported for circulatory (ICD-9 code 390-459.0), and respiratory (ICD9 460-519) diseases between 1999-2004. A total of 113,553 hospital admissions were related to respiratory conditions. The number of annual average daily admissions for respiratory diseases were 51.8 (range 20-167) per day. The highest daily number of admissions for a particular respiratory disease was as follows: pneumonia (17.6/day), chronic bronchitis (11.1/day), other respiratory diseases (7.5/day) and asthma (6.5/day). Hospital admission rates were highest in the oldest age groups 65-84 (445.7 per 10,000 population) and 85+ (995.0 per 10,000) followed by the younger children less than five years of age (164.5 per 10,000) (Table 6). A steady rise in hospitalization rates were noted, as age increased hospital admissions increased. These findings are consistent with national statistics on asthma hospitalizations (MMWR, 1980-2004; 2007, NCSH, 2007).

For all types of respiratory admissions, females had a higher admission rate than males (157.5 vs. 136.7), respectively (Talbott et al., 2007, [unpublished report]). However, males in

their early years less than five years of age and (5-24) were reported to have higher hospitalization rates (197.1 and 35.3 per 10,000 population) respectively, than females (130.5 and 28.4 per 10,000 population). This pattern in adolescent females after 25 years of age is reversed and continues up until age 65, when rates are higher for males. The gender difference may reflect the influence of hormones in adolescent females, and smoking and occupational exposure in older adult men. Seasonal effects were also evaluated by looking at hospitalizations for respiratory disease by month. The colder months, December through March yielded high numbers of admissions than months with warmer temperatures.

Table 6: Annual hospital admission rates Allegheny County residents 1999-2004 for respiratory disease by age

Hospital admissions for diseases of the respiratory system by disease category and age in Allegheny County residents 1999-2004: Annual average admission rates (per 10,000) population. First admissions and readmissions with primary discharge diagnosis (ICD9 460-519).

Diseases	**Age group**						
	<5	**5-24**	**25-44**	**4 5-64**	**65-84**	**85+**	**Total**
Acute bronchitis/bronchiolitis (466)	45.9	0.3	1.4	2.7	8.4	28.4	5.6
Other acute respiratory (460-465)	14.2	1.9	1.4	1.3	2.1	3.4	2.3
Other upper respiratory (470-478)	8.5	3.4	1.8	1.6	2.0	1.8	2.6
Pneumonia (480-486)	41.4	7.4	11.5	35.4	155.9	462.6	50.2
Influenza 487	1.3	0.2	0.3	0.5	1.5	4.3	0.6
Bronchitis (490-491)	0.2	0.2	2.5	30.7	129.9	150.8	31.5
Emphysema (492)	0.0	0.0	0.1	1.0	3.1	2.7	0.8
Asthma (493)	44.6	14.8	12.8	19.2	23.0	33.3	18.6
Other COPD (494-496)	0.0	0.0	0.1	1.2	6.9	9.2	1.6
Lung disease/other respiratory	8.5	3.6	7.6	26.0	113.1	298.4	33.8
Total	164.5	31.9	39.4	119.5	445.7	995.0	147.7

Reproduced from /projects/PITT-PM/Final_Report/CONSOL/FinalReport1_ver65.odt with approval from Dr. Evelyn Talbott (2007) Table 28, p. 187.

Data from emergency department (ED) visits in AC are reported by many health care agencies. A total of 14(70%) of the 20 hospitals surveyed in AC by the PITT-PM Health Outcomes Project Team (2007) indicates electronic data were available for emergency room visits. A total of 662,292 emergency room visits were reported in AC (2003-2004) of which 46% were associated with the University of Pittsburgh Medical Center (UPMC). Individuals living in AC reported their use of the larger (UPMC) emergency departments for care unless they lived on the outskirts of the city where smaller hospitals closer to their residence were used.

The data available on emergency room visits can be found in the medical record. The use of the medical record for data extraction can be timely and cost prohibitive for the health care agency. Currently, data are being stored electronically at many large health care agencies in the country. A total of the 14 hospitals surveyed in this analysis indicated they had the capacity to report data in an electronic format. These hospitals were inclusive under the UPMC network that has a large interoperable electronic medical record reporting system.

The UPMC Medical Archival Retrieval System (MARS) is used to provide de-identified electronic data on emergency room visits for the UPMC network of 10 hospitals. This system provides data from hospitals, physician offices and outpatient visits. MARS is useful as a tool for surveillance as it can provide information on demographics, past medical history, chief complaint, ICD 9 abstract, physical examination, discharge diagnosis and medication usage. The proposed research will receive de-identified data from the MARS database via an electronic medical record abstract where the demographic information is stored. Data for the emergency and admission intake, medication history, clinical and laboratory variables will be abstracted from the emergency note and history intake section of the medical record. The data to be

analyzed for this research study represents data from 6 hospitals (60%) of the UPMC network or 30% of the hospital network in Allegheny County.

3.2 PRELIMINARY ANALYSIS

In order to establish a baseline estimate of the burden of illness for the proposed study a preliminary analysis was conducted. The best estimate of asthma prevalence in AC has been reported from the 2002 Behavioral Risk Factor Surveillance Survey [BRFSS] that represents lifetime prevalence of asthma. Nine percent of AC residents responded "yes" to the question, "Have you ever been told by a doctor that you have asthma?" Based on this 9% estimate, approximately 113,933 (2002-2004 yearly average) persons in AC have asthma. Current asthma prevalence estimates from 2002-2005 for the Metropolitan Statistical Reporting Area of Pittsburgh which includes the U.S. Census-defined seven county region surrounding the city of Pittsburgh in Western Pennsylvania suggest that between (6.3-9.7%) of the population have asthma (BRFSS). Although, these results are inclusive of prevalence estimates for multiple counties they are based on the subjective recall of a patient's history. These findings are not the most robust estimates of asthma prevalence and require another methodology to validate the results. Hospital admission rates were determined to be a better indicator for measuring the impact of asthma morbidity for the proposed research.

To determine the rate of AC residents with a primary diagnosis of asthma who was being hospitalized due to asthma an analysis was conducted on the PHC4 and the MARS dataset. These datasets represented residents in AC with a first primary diagnosis of asthma who presented to the ED and were subsequently admitted for an asthma event from 2002-2005. Cases

were de-duplicated in both datasets, meaning that patients were only counted once as a primary diagnosis of asthma for the time period specified. The hospital admission rates are described in Table 7.

Table 7: Asthma hospital admission rates per 1,000 persons in Allegheny County
Comparing two data sources (2002-2005) for first admissions with primary discharge diagnosis ICD9 493.0-493.9. *(De duplicated)*

Column A	Column B	Column C	**Column D**	Column E	**Column F**
Year	Population in AC	Total Number inpatient admissions *(HC4) for asthma per year in AC	Asthma hospital admission rates in AC and 95% CI	Total number inpatient admissions ** (MARS dataset) for asthma per year in AC	Asthma hospital admission rates in UPMC dataset and 95% CI
2002	1,272,640	1,590	1.24 (1.18-1.31)	442	.347 (.316-.380)
2003	1,268,127	1,927	1.52 (1.45-1.58)	621	.489 (.452-.529)
2004	1,263,617	1,841	1.45 (1.39-1.52)	640	.506 (.468-.546)
2005	1,259,337	Not available		740	.587 (.546-.631)

** 98% compliance reporting of PHC4 data (not all admitted through ED for asthma)*
*** inclusive of 6 hospitals in UPMC dataset*

It was determined that for 2002-2004 (PCH4 dataset) that 1.2 per 1,000 persons with asthma were being hospitalized due to an asthma event (Table 7, Column D). Asthma hospital admission rates from the MARS dataset for the same time period were 300-400 times less than rates reported from the PCH4 data (Column F). However, the MARS dataset in this proposed research accounts for only 60% (6 out of 10 hospitals) of the UPMC network and only 30% of the AC hospital network. Both datasets represented children and adults. The UPMC MARS dataset did not include asthma visits from Children's Hospital but included individuals 0-103(M= 38.73, SD = 20.99). The PHC4 dataset age ranged from 0-103 (M= 44.72, SD= 26.36). The highest number of hospital admissions was reported for African Americans and females in

both datasets. The conclusion drawn from comparing the two datasets was that the proposed study dataset only captured a fraction of the asthma hospital admissions.

To estimate the number of AC residents with asthma who were utilizing the emergency room for asthma management, further analysis was conducted on the de-duplicated MARS dataset. Of the total 6,024 primary asthma cases, patients were seen in the ED and treated for asthma about 4-5 times more than those hospitalized for a diagnosis of asthma, respectively. This pattern was observed over a 4-year time period (Figure 10).

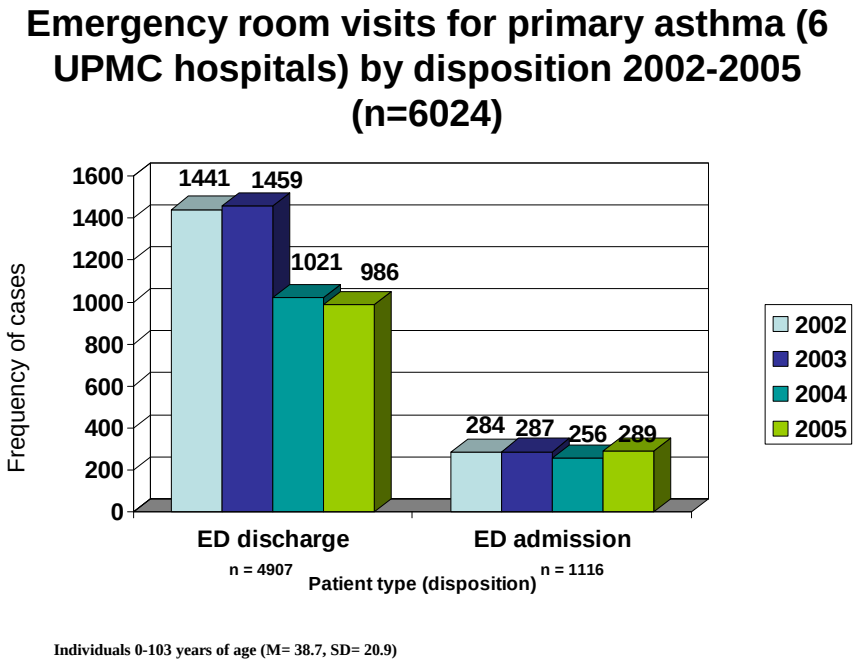


Figure 10: Preliminary analysis: Primary asthma diagnosis for emergency room visits (6 hospitals in UPMC network) by disposition

To further determine if individuals with a primary diagnosis of asthma had a prior visit for asthma which was not captured in the initial medical record abstract or for patients who may have been treated on an outpatient basis, a random sample (n=50) of the primary asthma dataset was conducted to look at the frequency of all types of visits for these 50 individuals. A total of

509 visits were identified (Table 8). Asthma related visits represented 102(20%) of all total number of visits. For individuals who presented to the ED for asthma, 75(73.5%) were being managed in the ED, 20(19.6%) admitted to the hospital for asthma, and 7(6.8%) treated on an outpatient basis. The preliminary analysis provided evidence that emergency room visits are greater in number and could be used to evaluate the short-term variations in asthma. Emergency room visits for asthma provide a better look at the "total" at risk population vs. asthma hospitalizations, which are reflective of those individuals who are at greatest risk with a more severe form of the disease. Moreover, further analysis is needed to determine the contributing factors leading to asthma exacerbations, which leads to utilization of services either in the emergency room or for those hospitalized.

Table 8: Healthcare utilization patterns for individuals with a primary asthma diagnosis 2002-2005

Random sample (n=50) primary asthma

Visit type	Frequency of all types of visits	Frequency of visits related to only asthma	Percent of visits related to asthma %
Emergency room	178	75	73.5
Inpatient hospitalizations	44	20	19.6
Outpatient	287	7	6.8
Total *	**509**	**102**	**100.0**

**Inclusive of 6 hospitals in UPMC network*

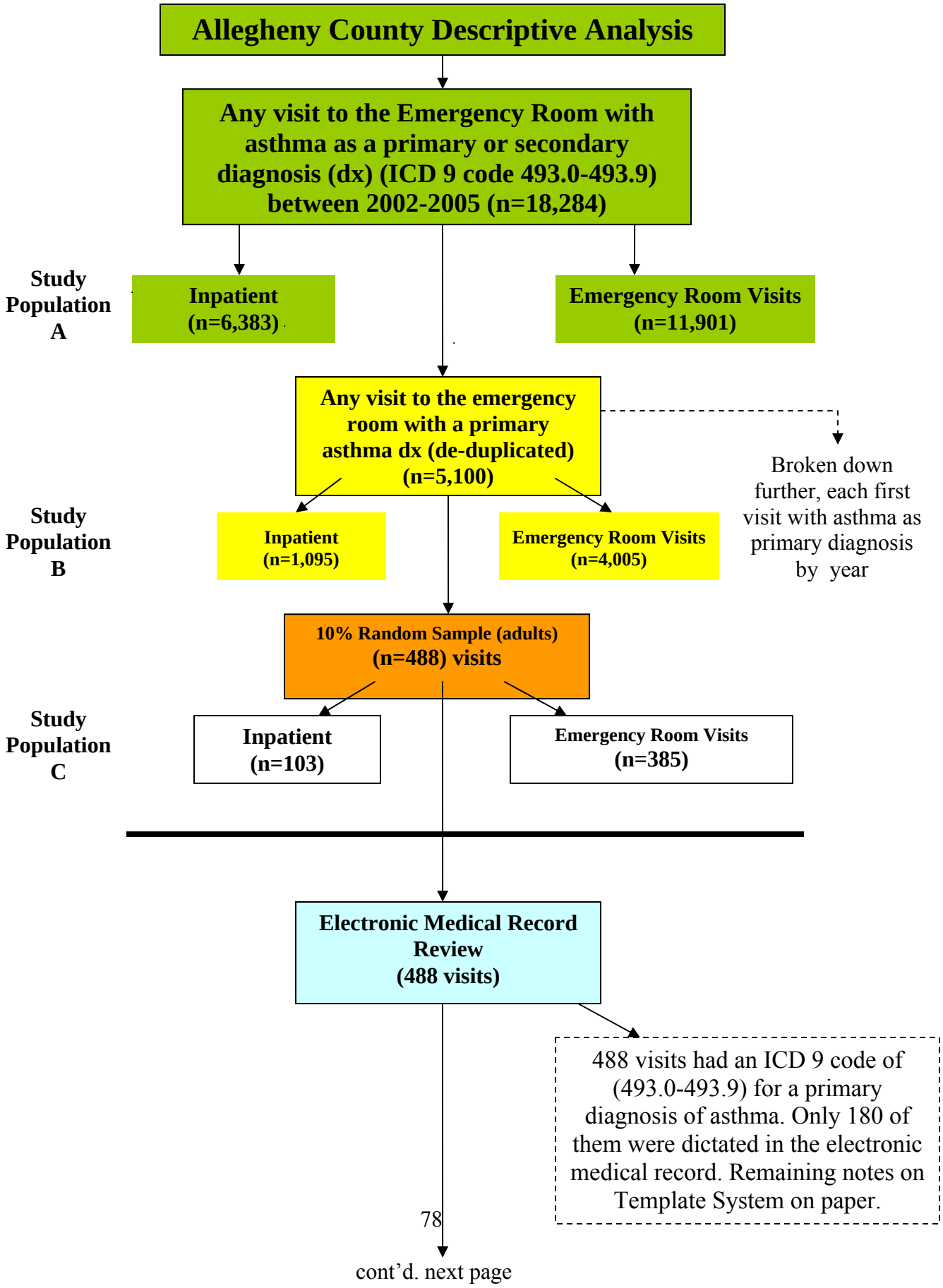
3.3 RESEARCH METHODS AND DESIGN

A cross sectional study was conducted for the period January 1, 2002 through December 31, 2005 using data abstracted from the electronic medical records for emergency room visits identified through the UPMC Medical Archival Retrieval System (MARS). The EMR abstract included data on emergency room visits from UPMC Braddock, Shadyside, Presbyterian, St. Margaret's, McKeesport and South Side Hospitals where electronic data was available for reporting. De-identified data was provided on the following variables: visit identifier, hospital, diagnostic related group, zip code, patient type, visit type, age, race, sex , insurance status, admission date, and discharge date, length of stay, chief complaint, and the admission and discharge diagnosis.

The medical record abstract was approved under the Exempt Status of the IRB# 0610013 (Appendix A). Four medical record abstracts were conducted which produced 4 different asthma datasets all with de-identified data for the variables mentioned above, and for the 6 hospitals included in the UPMC network. Each medical record abstract was conducted for the time period January 1, 2002 through December 31, 2005. The following criteria were used. Population A was produced through a medical record abstract using ICD-9-Code (493.0- 493.9; ICD-10-CM Codes: J45.0-J45.9) to identify all primary or secondary diagnoses of asthma, Population A (n=18,284). Population B included all first primary asthma ED visits with an ICD-9-CM code (493.0- 493.9; ICD-10-CM Codes: J45.0-J45.9). The records were de-duplicated and marked for the first asthma visit by year, and individuals 0-17 years of age were removed, Population B (n=5,100). Population C was obtained by conducting a 10% random sample of the all first primary asthma ED visits with an ICD-9-CM code (493.0- 493.9; ICD-10-CM Codes: J45.0-

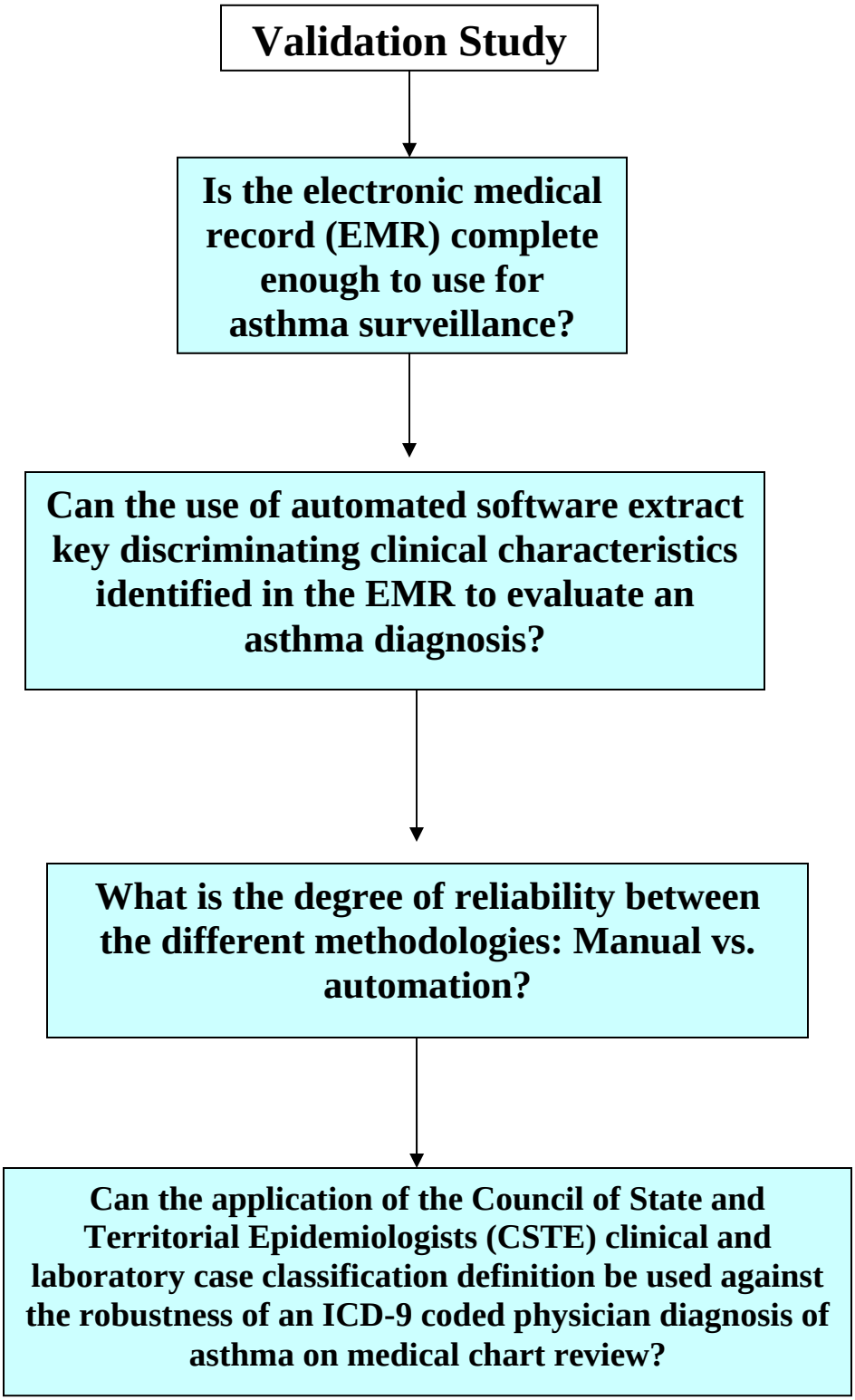
J45.9), Population C (n=488). Population D included any visit to the ED with a primary diagnosis of asthma with an ICD-9-CM code (493.0-493.9). These data for Population D represented recurrent visits for a primary asthma diagnosis in 6,979 individuals. For the purpose of the time stratified case, crossover study conducted under specific aim #3, the numbers of visits were reduced to fulfill the assumptions required under conditional logistic regression that requires control of the dependency factor for each asthma visit. A total of 10,183 visits were identified for the population of 6,979 persons, this was Population D (n= 10,183). The methods used to reduce the number of asthma visits is further explained under the Methods Section 3.3, Specific Aim #3. A diagram to outline the specific aims as they relate to each population of this research is illustrated in Figure 11.

Specific Aim #1



cont'd. next page

Specific Aim #2



cont'd. next page

Specific Aim # 3

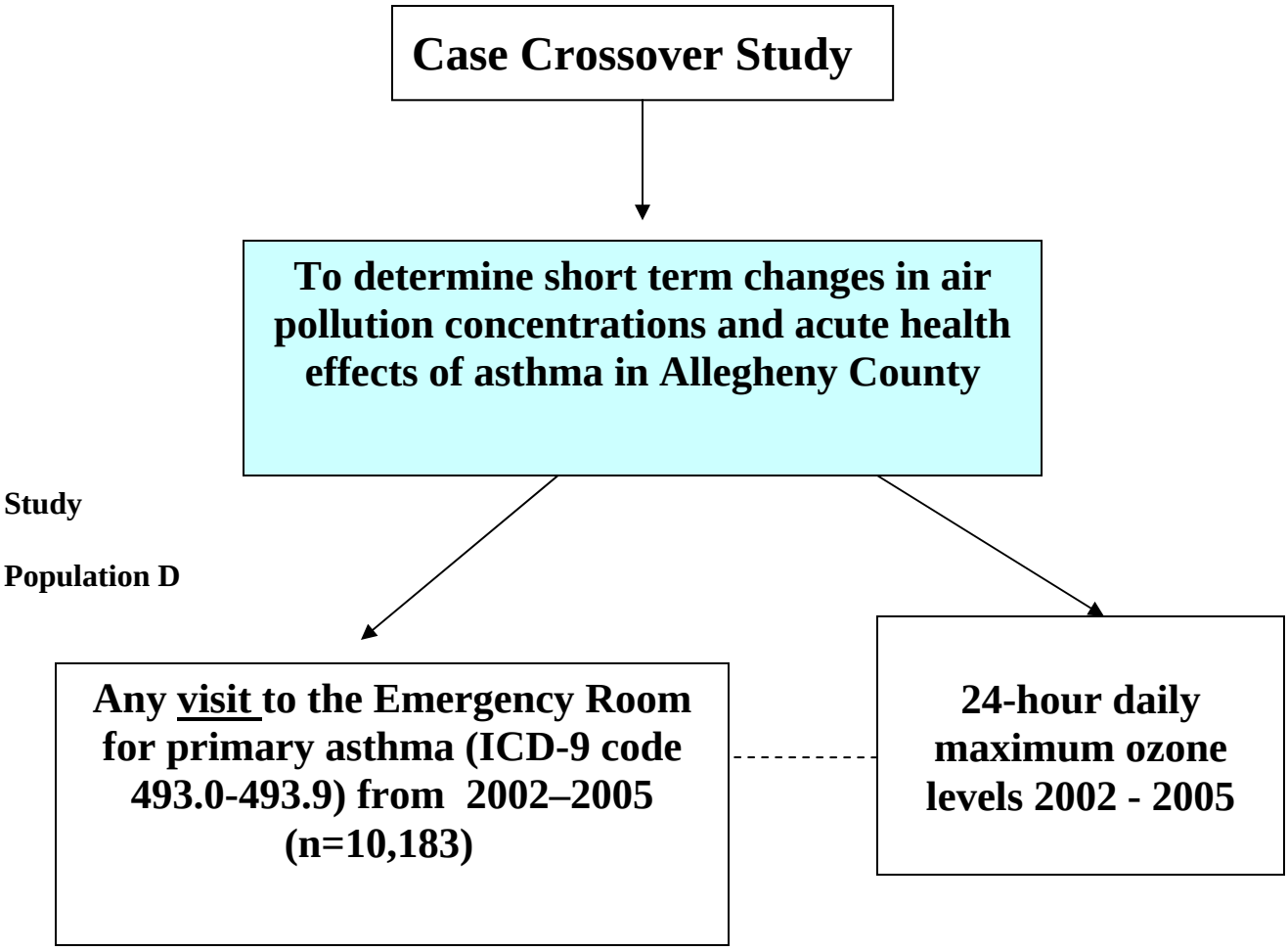


Figure 11: Specific Aims

The goal of this research is to evaluate different methodologies for asthma surveillance and demonstrate the sensitivity of their results. These specific aims are:

1. To describe asthma in a population of Allegheny County residents, a descriptive analysis was conducted using the CSTE Hospital Discharge Classification for all primary or secondary diagnoses of asthma who presented to the ED from January 1, 2002 through December 31, 2005 with an ICD-9-Code (493.0- 493.9; ICD-10-CM Codes: J45.0-J45.9) for asthma, and for all first primary asthma ED visits with an ICD-9-Code (493.0- 493.9; ICD-10-CM Codes: J45.0-J45.9). The records were de-duplicated and marked for the first primary asthma visit by year. The following variables were used in the descriptive analysis: age, sex, race, hospital, patient type, insurance type, zip code, chief complaint, and admitting and discharge diagnosis.

2. To describe whether the use of the electronic medical record was complete enough to use for asthma surveillance, a representative sample of the electronic medical records for individuals with a first primary asthma ED visit were obtained. A total of 488 visits had an ICD 9 code of (493.0-493.9) for a primary diagnosis of asthma. Only 180 of these ED visits were dictated in the electronic medical record. A summary describing the EMRs was based on the completeness of information found in the records when conducting the Validation Study and classify asthma cases according to the CSTE Clinical and Laboratory case classification definition. A secondary goal under specific aim #2 included a Validation Study to determine the following:

A. Whether the use of automated software could be used to extract key discriminating clinical characteristics from the EMR to validate a diagnosis of asthma?

B. What the degree of reliability is between the manual vs. automated methodologies?

C. Whether the use of the Council of State and Territorial Epidemiologists Clinical and Laboratory Case Classification definition could be used against the robustness of the ICD-9 coded physician diagnosis of asthma?

3. A case crossover study was conducted to determine the association between short-term changes in air pollution concentrations for the 24-hour maximum ozone concentration and daily 24 hour average $PM_{2.5}$ level) and the acute effects of any primary asthma emergency room visit for 6 hospitals in AC.

Specific Aim 1: In order to describe the asthma population used to conduct countywide asthma surveillance, the investigator had to find a methodology for properly identifying asthma cases. The CSTE Hospital Discharge Classification definition using both the ICD 9 coded primary and secondary asthma diagnosis (physician diagnosed) was used to conduct the first EMR abstract. To further explore other options for identifying asthma cases the CSTE Hospital Discharge Classification case definition criteria was modified to include only the ICD 9 coded primary asthma diagnosis. A descriptive analysis of each Population A-B created through the medical record abstract was conducted. The analysis included a review of the following variables, age, sex, race, hospital, patient type, insurance type, zip code, chief complaint, and admitting and discharge diagnosis. A comparison was made between the primary asthma admitting diagnosis vs. hospital discharge diagnosis to determine if one diagnosis was superior to the other for ascertaining cases of asthma.

Specific Aim 2: To explore whether the EMR was complete enough to conduct asthma surveillance, a sample of the electronic medical records representing individuals with a first

primary asthma ED visit was needed. A 10% random sample of (Population-B) was conducted. This step resulted in a third asthma dataset-Population C (n=488). Of the 488 visits, which had an ICD-9 code for a first primary diagnosis of asthma, only 180 of them were dictated in an electronic medical record. This is due to the large percentage of ED visits, which are done, via a Template system on paper and they do not use an electronic medical record system. This applies to UPMC hospitals, South Side, Braddock, McKeesport and Horizon. Only the electronic medical record that had an ED summary note available for review was included in the analysis (n=180). The descriptive summary of the electronic medical record was based on the completeness of data found in the records when conducting the validation study and classifying asthma cases according to the CSTE clinical and laboratory classification case definition.

To begin the Validation Study a list of asthma symptoms or characteristics predictive of an asthma diagnosis was needed. A set of seven indicators deemed by the researcher and the University Of Pittsburgh Graduate School Of Public Health Environmental Public Health Tracking Team was selected as the criterion for identifying a case of asthma (Figure 12). The consensus was supported by evidence purported in the medical literature on asthma (Sanders et al., 2007; Hung et al., 1998; Sistek et al., 2001; Sistek et al., 2006; Sanders et al., 2007; Yu, et al., 2004). This step was necessary to test the diagnostic accuracy of using data abstracted from the EMR. A lexicon of modifiers was also created that represented alternative keywords for each indicator to reflect levels of certainty, negation, and double negatives (Table 9). A list of asthma medications adapted from the American Lung Association Asthma Medication Chart was included in the review (Table 10).

A trained medical technician conducted the medical chart review and recorded the findings onto the clinical questionnaire (Figure 12). Each record was marked for one code to

reflect the absence [no], presence [yes] or not mentioned status for each indicator. The use of these three classification concepts was needed for the automated application process by which these results would be compared for statistical validation.

Questionnaire A

Asthma Validation Study: Selected Indicators

Report ID: _____

Admission Date:_____mm/dd/yr _____

Age_____
Sex_____
Race_____
Chief Complaint_____

INDICATOR	ABSENCE (0)	PRESENCE (1)	NOT MENTIONED (2)
Wheeze/wheezing			
Cough			
Shortness of breath (SOB) Dyspnea-difficulty breathing			
Rhinitis (allergic rhinitis Runny nose)			
Fever			
Asthma history (hx) Past medical history of asthma (PMH)			
Medications/inhaler/nebulizer*			

* Included asthma medications (Table 10)

Figure 12: Questionnaire

As a quality control measure, a nurse reviewed the medical records for data recording accuracy. The clinical documentation found in the medical record was useful to the nurse for data interpretation. Alternative keywords such as the numeric component of the word fever were considered. The presence of a fever was recorded if a numeric value was found to be 100° F or greater. Contradictory statements as to the absence and or presence of an indicator were placed in the category which best reflected the acuteness of the sign or symptom as reported by the patient on presentation to the emergency room. For example, a medical record was marked as the presence of wheezing if the following documentation was found. A patient presented to the emergency room with coughing, wheezing and shortness of breath (SOB) which did not resolve in 20 minutes, thus promoting ED consultation. The documentation found in the Chest and Lung Section under the physical exam indicated the patient had no wheezing. This instance, although contradictory for wheezing was marked as a positive finding. Medical abbreviations and acronyms known by the review nurse were also considered while reviewing the EMR.

A secondary goal under Specific Aim 2 was to validate the findings from the medical record chart review by using some form of automated software. Three agents were employed, (a) ATLAS.ti software, (b) Practical Extraction and Reporting Language (PERL), and (c) Medical Language Extraction and Encoding (MedLEE) text processing. The level of agreement between the gold standard medical chart review, and computer based methodologies were assessed for statistical validation. ATLAS.ti software was used as a prelude to the more advanced agents employed in this research. Therefore, the researcher reports on the ATLAS.ti software features only in the review of literature.

In collaboration with CM, the programmer contracted through the University Of Pittsburgh Graduate School Of Public Health Department Of Epidemiology, a PERL program

was created to interact with the asthma medical record database. The PERL code used to create the program is listed in Appendix B. The list of words and phrases used to guide the programmer in developing the search expression are highlighted in Tables 9 & 10. The file containing the search text is listed in Figure 13. The search text was executed by PERL to match on the keywords found in the narrative of the EMR using these search terms.

Table 9: Keywords used for Practical Extraction and Report Language (PERL) matching

Selected Indicators	Coding scheme		
	Absence (0)	**Presence (1)**	**Not mentioned (2)**
Wheezing	(No) (Absence of), (denies) (without) wheeze, wheezing, wheezes	(Presence of) (patient with) Wheeze wheezing, Wheezes	No mention of the words wheeze wheezing, wheezes in review
Cough	(No), (Denies) (Absence of) Cough, coughing	(Productive), (chronic), (complains of) (present) cough , coughing	No mention of the word cough, coughing in review
Shortness of breath	(No) (denies) (Absence of) Shortness of breath, DOE, dyspnea, difficult breathing, can't breath, hard time breathing, trouble breathing, labored breathing, breathing labored	(Presence of) (patient with) (complains of) Shortness of breath, DOE, dyspnea, difficult breathing ,can't breath, hard time breathing, trouble breathing, breathing labored, labored breathing, can't catch breath	No mention of the word shortness of breath, DOE, dyspnea, difficult breathing
Rhinitis	(No) (Absence of), (without) rhinitis, allergic rhinitis	(Presence of) (patient with) (complains of) rhinitis, allergic rhinitis	No mention of the word rhinitis, allergic rhinitis
Fever	(No) (Absence of), (without), (denies) Fever, temperature, Afebrile	(Presence of) (patient with) Fever, temperature, Febrile	No mention of the word fever, temperature
History Asthma	(No), (Absence of), (denies) History of, hx, PMH, past medical history no asthma, no known asthma history	(Presence of) (patient with) (yes) (confirmed) all with History of asthma, hx of asthma, PMH asthma, past medical history asthma, asthma history	No mention of the word -History of asthma, hx of asthma, PMH asthma, past medical history asthma, asthma history
Medications/Inhaler	(No) inhaler nebulizer, (denies use of), inhaled steroids, beta agonist drugs, inhaler	(Yes) (use of), (taking) inhaler, nebulizer, theophyllene, steroids, beta agonist drugs, inhalers	No mention of the word inhaler, nebulizer, inhaled steroids, beta agonist drugs,

Table 10: Asthma medications*

Medicine	Generic Name	Brand Name
Short Acting Beta-2 Agonists (SABA) Inhaled or Oral Bronchodilator	Albuterol Albuterol sulfate	Accuneb® Proventil HFA® Ventolin HFA ® ProAir HFA ® Proventil Repetabs® (tablet) VoSpire ER®
	Pirbuterol acetate	Maxair®
	Terbutaline Sulfate	Brethine® (tabs only)
	Levalbuterol hydrochloride	Xopenex®
	Levalbuterol tartrate	Xopenex HFA®
Anticholinergics, inhaled bronchodilator	Ipratropium bromide	Atrovent®
Inhaled Corticosteroids	Beclomethasone diaproprionate	QVAR® QVAR® Beconase
	Budesonide	Pulmicort Respules® Pulmicort Turbuhaler® Rhinocort
	Flunisolide	AeroBid® Aerospan HFA , Nasarel
	Fluticasone proprionate	Flovent ® Flonase
	Mometasone furoate	Nasone Nasacort Asthmanex®
	Triamcinolone	Azmacort

*Adapted from the American Lung Association, December 2007: Asthma Medication Chart

(Table 10 Continued)

Long Acting Beta 2 Agonists	Formoterol fumarate	Foradil Aerolizer®
	Salmeterol Xinofate	Serevent Diskus®
Combined Medication: Inhaled Bronchodilator and Steroid	Budesonide + Formeterol fumerate	Symbicort®
	Flucticasone propionate + Salmeterol xinofate	Advair Diskus®
Leukotriene Modifiers: Oral Anti-inflammatory	Zafirlukast	Accolate®
	Zileuton	Zyflo®
	Montelukast	Singulair®

PERL search Text
Wheeze;wheeze,wheezing,wheezes;no,absence of,denies,without Cough;cough,coughing;no,absence of,denies,without Shortness of breath; shortness of breath,DOE,dyspnea,difficult breathing,can't breath,hard time breathing,trouble breathing,breathing labored,labored breathing,can't catch breath;no,denies,absence of Rhinitis;rhinitis,allergic rhinitis;no,absence of,denies,without Fever;Fever,temperature;no,absence of,denies,without History of Asthma;History of asthma,hx of asthma,PMH asthma,past medical history of asthma,asthma history;no,absence of,denies,without Meds/Inhaler;inhaler,nebulizer,Albuterol,Accuneb,Proventil,Ventolin,ProAir,VoSpire,Pirbuterol,Terbutaline,Brethine,Levalbuterol,Xopenex,Atrovent,QVAR,Beconase, Budesonide,Pulmicort,Rhinocort,Flunisolide,Nasarel,AeroBid,Flovent,Flonase,Nasone,Nasacort,Asthmanex,Azmacort,Foradil,Serevent,Symbicort,Advair,Accolate, Zyflo,Zafirlukast,Zileuton,Montelukast,Singulairinhaler,nebulizer,theophyllene,steroids, beta agonist drugs,inhalers; No meds:No medications;

Figure 13: PERL script

The following methods were written by the programmer CM to describe the logic of how the search expression was developed. Permission was given by the author to reproduce these methods in support of this research.

Regular Expression Methods

"Regular expression (pattern matching) was used to determine the presence or absence of various words and phrases in the electronic medical records (EMR). The presence or absence of these words or phrases were then used to determine whether a symptom was mentioned and whether that symptom was present in the patient or was not present in the patient.

To process the file containing the EMRs, a script was developed to read various parts of the header and extract information that was useful in identifying the record. Then, the rest of the record was read and accumulated into a single string with the elimination of any line break (line feeds or carriage returns) characters. The end of the record was determined by the E_O_R (end of record) line. Once the entire record was entered into the string variable without line break characters it was copied (one copy for each regular expression) and the copies were scanned for character sequences that matched the regular expressions of interest.

The regular expressions that were used to search the records can be divided into two categories, present or not present. In general, a regular expression for a symptom that was present was the name of the symptom or morphological equivalents (ex. wheeze and wheezing). The regular expression for this would look like:

(wheeze,wheezing)

This will match on any word in the positive list

A regular expression for a symptom that was not present was the name of the symptom with a not, no, or some other negating term in front of the name within the sentence (ex. no wheezing or not wheezing). The regular expression for this would look like:

(no,not) [^.]+(wheeze,wheezing)

This will match on any word in the negative list (no or not) preceded and followed by a space (there is an unseen space before the negative list), followed by a string of characters 0 or more long that does not include the period (the period is assumed to be the end of the sentence and this expression will only match inside the same sentence). Finally, a match from the positive list must follow before the end of sentence to produce a match.

After the scans of the record was completed using each of the regular expressions that were created, the final decision, as far as whether or not a symptom was present, not present, or not mentioned was arrived at by the following logic. First, if a symptom name was found with a negative modifier before it, then it was considered to be either 'not present' in that patient or denied by that patient. Second, for the rest of the cases, if a symptom name was present, then the symptom was considered to be 'present' in that patient. Finally, for all the rest of the cases where the symptom name was found, neither by itself nor with a negative modifier, then it was considered to be 'not mentioned'."

Instructions for Using the Tool

"To use the Medical Record Processing tool, first the Perl interpreter must be installed on the computer that is to be used. The installer may be downloaded at http://www.perl.org/get.html. Then the script file (AutoRegex.pl), the file containing the records that are to be scanned (input.txt), and the parameters file for the search (search.txt) are all placed into the same folder. Finally the Perl script is executed by double clicking on the icon for the AutoRegex.pl file. The output file will be created in the same folder that the other files are in and will be called regex.csv and will open in a spreadsheet program that reads comma separated value files.

The search file is organized with each line being a separate search. Each search being divided into 3 fields separated by semicolons. The first field is the title of the search. The second field is the positive search terms in a comma separated list with no spaces before or after the terms. This list is the words or phrases that are expected to denote the symptom that is being examined (wheeze, wheezing, wheezed). The final field is the negative list of words that are expected to negate the presence of the symptom (no, not, denies)."

The second methodology MedLee, was employed by our collaborator, biomedical expert RT. The goal of using MedLEE was to extract the set of indicators identified in the free text clinical documents using a lexicon library for translation and scientific interpretation. The UMLS concepts codes, which correspond to each asthma indicator, are highlighted in Table 11. If an indicator had more than one concept code, the results were collapsed at the end of the analysis to reflect one set of results for each indicator. The concept codes were used to match the words identified in the free text document along with the modifiers. These concepts are then linked to a semantic network, which classifies the words into a set of categories. The MedLee

lexicon provides a description of the findings, which is reported in an XML output file. The results were consolidated in this research by the programmer to reflect the presence or absence of a finding. If the indicator was not mentioned for the presence or absence of findings, a no mentioned status was assigned.

Table 11: UMLS key concept codes for MedLEE

Selected Indicators	Concepts	Assigned UMLS Concept Codes
Wheeze	Wheezing, Wheeze, Wheezes, Wheezy,Wheezings	C0043144
Cough	Coughing, Cough	C0010200
	Dry cough	C0850149
	Cough Non-productive	C0239134
	Cough Dry, Unproductive cough, on-productive cough	C0850149
Shortness of breath	Shortness of breath, SOB, Breath Shortness	C0392680
Rhinitis	Rhinitis	C0035455
	Runny Nose	C1260880
Fever	Fever, Hyperthermia, Febrile Pyrexia, Temperature elevation Body temperature increased, Temperature increase	C0015967
Asthma history	Asthma history	C0455544
	Asthma	C0004096
Medications	Inhaler, Inhalator, Inhalators	C0021461
	Nebulizers	C0027524

The next step of the Validation Study was to review the EMRs, which were marked as asthma by the ICD-9 coded physician diagnosis and classify the cases of asthma according to the criteria highlighted in the CSTE Clinical and Laboratory case definition (Table 5). In order to comply with the criteria highlighted in the CSTE Clinical and Laboratory Classification definition under the "confirmed" asthma category, medical records from the previous year had to be reviewed to determine if there was an occurrence of any asthma clinical symptom reported at least three times in the past year for asthma. A total of 838 clinical narrative reports, including free text history and physicals, dictated physicals, consultations, clinic follow up visits, ED notes, laboratory results, surgery reports and discharge summaries were obtained. Three possible classifications were presented: Confirmed, Probable, and Possible. If no determination could be made as to whether a patient met the criteria for any of the classification categories the record was marked as not classified. The criteria were modified from its original version to reflect the limited amount of laboratory data found in the EMR. The modified criteria are listed in Table 12.

Table 12: Modified CSTE Clinical and Laboratory case classification criteria

Classification	Criteria
Confirmed	Met any clinical symptom at least 3x during the past year • Wheezing lasting 2 consecutive days or more • Chronic cough responds to bronchdilatation that persists 3-6 weeks in the absence of allergic rhinitis sinusitis • Nocturnal awakening with dyspnea on exertion, cough, and or wheezing in the absence of other medical conditions known to cause these symptoms
Probable	Presence of any symptom reversed with a physician treatment or asthma medication 3 times in the past year
	Taken medications in the past year that were prescribed by a physician for asthma
Possible	Presence of any of the following symptoms during the past year • Shortness of breath • Wheezing or chronic cough in the absence of obvious respiratory infection • Presence of nasal secretions, mucosal swelling, nasal polyps or chronic sinusitis • Hyper expansion of the thorax • Prolonged phased to forced exhalation • Wheezing during normal breath • Chest X-ray noted to show hyper expansion

Upon review of the EMR, the free text chief complaint was assessed to determine how well it correlated with the medical chart review and with the ICD-9 coded primary asthma diagnosis for substantiating a case of asthma. The addition of reviewing the chief complaint in this portion of the analysis increased the researcher's ability to eliminate any misclassification of an asthma diagnosis.

Under Specific Aim # 3, a case crossover study was conducted to examine 10,135 emergency room visits and determine the association between short term changes in daily maximum 24-hour ozone concentrations and $PM_{2.5,}$ and the acute health effects for asthma. Ozone and $PM_{2.5}$ data from one monitoring station in Allegheny County, which is representative of the urban portion of the county was used in the analysis. Data was adjusted for temperature in the model.

The Case Crossover Analysis Tool (C-CAT), a software application developed by researchers from the New York State Department of Health was used to conduct the research. The software operates using a SAS based program. The statistical method used in this research was conditional logistic regression as the statistical estimator. Power and sample size estimates were conducted using PS Statistical Program (Dupont, 2004). The exposure variables were treated as a continuous variable using a matched study design to perform these calculations. A time stratified case crossover analysis was used in this research study. All referents (control periods) were based within the calendar month. This research was defined by the following hypothesis:

Test Ho: Odds ratio=1.00 Ha: Odd ratio $\neq$ 1.00

The null hypothesis assumes there is no change in asthma ED visits for every 10-ppb increase in the 24-hour maximum ozone concentration level.

3.3.1 Sample size and power estimate

The sample for the any primary asthma ED visit dataset originally had 11,286 visits for asthma. This represented a total of 6,979 persons. Individual's events are linked to a unique identifier (CASEID) in the C-CAT software. Individuals can have a repeat visit but only one per

day. Furthermore, the use of conditional logistic regression in a case crossover methodology requires control of the dependency factor for recurrent events. In order to assume an independence status for each visit, a "wash out" period was applied. The wash out was set to remove anyone who had a repeat visit within a 28-day time period, thus, accounting for independence within the 28-day time strata. A referent day was selected every 7 days from 21 days before to 21 days after the event day. A total of 1,103 ED asthma visits were washed out. This resulted in 10,183 visits represented by 30,405 referents: a 3:1 match (Haley et al., 2006). The estimated power to detect a 10% increase in the risk estimate with a sample size of 6,979 individuals is 74%, alpha = 0.05, two sided test; 3:1 matching, correlation =0.2, proportion of exposed controls =0.2 (Figure 14).

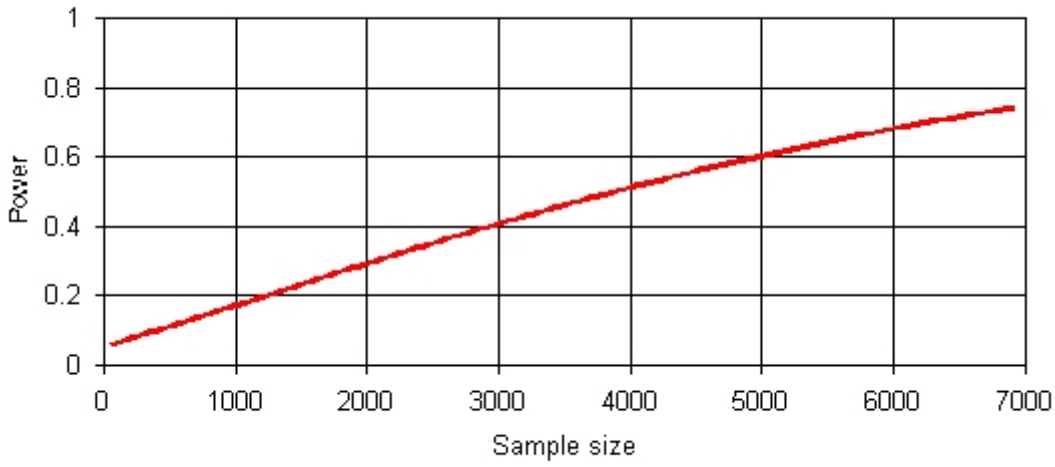


Figure 14: Sample size and power estimate

3.4 STATISTICAL ANALYSIS AND RESULTS

All data was analyzed using SPSS with the exception of the case crossover study, which employed the use of a SAS based program to conduct the analysis. The case crossover study relies on conditional logistics regression as the statistical estimator of a binary outcome with one or more predictors. The maximum likelihood values for the logistics model were used to report the best estimates or maximum probability of finding the results as reported by the odds ratios and 95% confidence limits.

All the descriptive analysis was conducted using summary statistics such as means, proportions, and standard deviations where appropriate. The Pearson's Chi Square statistic was used for categorical variables, and histogram and box plot for discrete and continuous variables. The Pearson's correlation coefficient was used to measure the strength of association between variables where appropriate. Ninety-five percent confidence intervals were constructed to determine whether statistically significant differences were noted between characteristics identified within each population.

3.4.1 Populations descriptive analysis

3.4.1.1 Population A

A total sample of 18,284 emergency room asthma visits were identified which represented any one who presented to the emergency department (ED) for the 6 hospitals represented in this research study with a primary or secondary diagnosis of asthma from 2002-2005. The visits in this MARS dataset accounts for only 60% (6 out of 10 hospitals) of the UPMC network and only 30% of the AC hospital network. Of the total number of ED visits,

11,901(65.0%) resulted in discharge and 6,395(35%) admitted. For the entire asthma cohort the age range was 0-103 (M= 43.3, SD= 22.3) (Table 13). The highest number of visits occurred in the 45-64 year old age group which represented 25.7%(4,700) of the population followed by individuals 30-44 years of age 23%(4,216). Individuals who were discharged from the hospital were younger 0-98 years of age (M= 35.99, SD= 19.8) vs. individuals who were admitted 0-103 years of age, (M= 56.97, SD = 20.14). Individuals (65±) years of age represent the highest proportion of individuals admitted to the hospital. This suggests that comorbidity plays a role in asthma hospitalizations for the oldest population (65±).

A review of the gender differences demonstrated that the number of ED visits was greatest for females (65%) and almost twice that of males (35.0%) (Table 13). When comparing gender differences by disposition, females were discharged from the ED more often than males, 63.9%(95% CI 63.0-64.8); 36.1%(95% CI 35.2-37.0), and admitted more frequently than their male counterparts 67.4%(95% CI 66.0-68.3); 32.6%(95% CI 31.4-33.7), respectively. This is consistent with the literature, which suggests that females are two times more likely than males to present to the ED for asthma care, to be admitted, and report a repeat asthma event (Singh et al, 1999). However, no conclusions can be made about the gender differences in this population since the higher proportion of females represent findings from only the 6 hospitals in the catchment area.

Analysis of the data on racial backgrounds demonstrated that more Whites 68.4%(95% CI 67.7-69.0) visited the emergency room than Blacks 29.6(95% CI 28.9-30.2), and Others 2.0%.(95% CI .018-.022. The same distribution patterns were observed when disposition was taken into account, Whites still had more ED visits and more admissions than their Black counter parts and Others. However, Blacks make up only 10% of the total population in AC, and 30%

of the population in the City of Pittsburgh (COP) (US Census, 2006). According to national statistics (NCHS, 2004), Blacks utilize the emergency room more often than Whites (195 per 10,000 vs. 43.6 per 10, 000) population, respectively; are hospitalized more frequently (33.3 vs. 10 per 10,000) population, respectively; and die at higher rates (3.1 vs. 1.7 per 10,000) population, respectively, from asthma. These national statistics suggest that asthma is not as well controlled in Blacks vs. Whites. In this population, the proportion of Blacks utilizing the ED for asthma is high when comparing to the small proportion of the total Black population in AC as a whole. However, no conclusions can be made from the findings in this research study since this was an abbreviated sample of 6 hospitals in the UPMC network.

Table 13: Primary and secondary asthma diagnosis for all emergency room visits (6 UPMC hospitals) by age, sex, race and disposition

	Total Emergency room visits with primary and secondary asthma diagnosis (n=18,284) by disposition =emergency room visits with discharge or emergency room visit with admission	
Total Emergency room visits with **primary and secondary asthma diagnosis (n=18,284)**	Emergency room visits with **discharge (n=11,901)**	Emergency room visits with **admission (n=6,383)**
Age n % CI	**Age n % CI**	**Age n % CI**
0-17 2,267 12.3 11.9-12.9	0-17 2,188 18.3 17.7-19.1	0-17 79 1.2 .99-1.53
18-29 3,439 19.0 18.2-19.3	18-29 2,856 24.0 22.0-24.7	18-29 583 9.6 8.4-9.8
30-44 4,216 23.0 22.4-23.6	30-44 3,028 25.4 24.6-26.2	30-44 1,188 18.5 17.6-19.5
45-64 4,700 25.7 25.0-26.3	45-64 2,678 22.5 21.7-23.2	45-64 2,022 31.6 30.4-32.7
65± 3,661 20.0 19.4-20.6	65± 1,150 9.6 9.1-10.2	65+ 2,511 39.33 38.0-40.4
Mean age= 43.3(SD 22.3) Range 0-103 years of age	Mean age = 35.99 (SD 19.8) Range 0-98 years of age	Mean age = 56.97 (20.14) Range 0-103 years of age
Total 18,284 100.0	**Total 11, 901 100.0**	**Total 6,383 100.0**
Sex n % CI	**Sex n % CI**	**Sex n % CI**
Male 6,379 34.9 34.2-35.5	Male 4,296 36.1 35.2-37.0	Male 2,083 32.6 31.4-33.7
Female 11,905 65.1 64.4-65.7	Female 7,605 63.9 63.0-64.8	Female 4,300 67.4 66.0-68.3
Total 1,8284 100.0	**Total 11, 901 100.0**	**Total 6,383 100.0**
Race n % CI	**Race n % CI**	**Race n % CI**
Black 5,414 29.6 8.9-30.2	Black 3,968 33.3 32.5-34.2	Black 1,446 22.7 21.0-23.6
White 12,503 68.4 67.7-69.0	White 7,750 65.1 64.3-66.0	White 4,753 74.5 73.5-75.3
Other 136 .7 .63-.88	Other 108 .9 .75-1.09	Other 28 .4 .30-.63
Unknown 231 1.3 1.11-1.43	Unknown 75 .6 .50-.79	Unknown 156 2.4 2.0-2.8
Total 18,284 100.0	**Total 11, 901 100.0**	**Total 6,383 100.0**

A review of the different insurance types for individuals who presented to the ED for asthma care demonstrated that the number of individuals who were discharged from the ED were younger (M= 35.99, SD= 19.8) than those admitted (M= 56.97, SD= 20.14). The proportion of individuals with managed care insurance was higher than the number of individuals without (Table 14). The exception to this was for individuals who had Medicare insurance. These individuals represent the oldest and sickest individuals in the entire cohort. Individuals with Medicare Managed Care (MMC) insurance vs. their Medicare counterpart were less likely to visit the ED and if they did visit the ED were more likely to be discharged than admitted. Although, no conclusions can be drawn from these findings, enrollees in MMC insurance programs are more likely to report using preventive services such as immunizations and smoking cessation than their traditional Medicare counterparts (Landon, Zaslavsky, Shulamit, Cioffi, & Clearly, 2004). Thus, it is possible that MMC individuals are healthier or more protected with these types of prevention initiatives. The high number of individuals identified in the Self-pay group 1,590(8.7%) suggest that the emergency room is used for asthma treatment even if someone does not have health care insurance. A more in-depth analysis is needed to evaluate the influences of different insurance types on the delivery and outcomes of asthma care.

Table 14: Insurance type for all primary and secondary asthma emergency room visits (6 UPMC hospitals) by disposition

				Total Emergency room visits with primary and secondary asthma diagnosis (n=18,284) <u>by disposition</u> =emergency room visits with <u>discharge</u> or emergency room visit with <u>admission</u>							
Total Emergency room visits with **primary and secondary asthma diagnosis (n=18,284)** Mean age= 43.3(SD 22.3) Range 0-103 years of age				Emergency room visits with **discharge (n=11,901)** Mean age = 35.99 (SD 19.8) Range 0-98 years of age				Emergency room visits with **admission (n=6,383)** Mean age = 56.97 (20.14) Range 0-103 years of age			
Financial	**n**	**%**	**CI**	**Financial**	**n**	**%**	**CI**	**Financial**	**n**	**%**	**CI**
Medicare	3,199	17.5	16.9-18.0	Medicare	1,108	9.3	8.8-9.8	Medicare	2,091	32.8	31.5-33.8
Medicare MC	2,014	11.0	10.5-11.4	Medicare MC	1,042	8.8	8.2-9.2	Medicare MC	972	15.2	14.3-16.1
Medicaid	713	3.9	3.62-4.19	Medicaid	283	2.4	2.1-2.6	Medicaid	430	6.7	6.1- 7.3
Medicaid MC	3,323	18.2	17.6-18.3	Medicaid MC	2,670	22.4	21.7-23.2	Medicaid MC	653	10.2	9.4- 10.9
Commercial	813	4.4	4.15-4.75	Commercial	494	4.2	3.8-4.5	Commercial	319	5.0	4.4- 5.5
Commercial MC	2,835	15.5	14.9 -16.0	Commercial MC	2,196	18.5	17.7-19.1	Commercial MC	639	10.0	9.2-10.7
Third PP	3,256	17.8	17.6 -18.3	Third PP	2,361	19.6	19.1-20.5	Third PP	895	14.0	13.1-14.8
Third PPMC	0			Third PPMC	0			Third PPMC	0		
Self pay	1,590	8.7	8.29-9.11	Self pay	1,385	11.6	11.0-12.2	Self pay	205	3.2	2.8-3.6
Other	535	2.9	2.69-3.18	Other	361	3.0	2.7-3.3	Other	174	2.7	2.3-3.1
VA	5	.02	.012-006	VA	0			VA	5	.1	.03-.16
Unknown	1			Unknown	1						
Total	**18,284**	**100.0**		**Total**	**11,901**	**100.0**		**Total**	**6,383**	**100.0**	

The annual daily average of emergency room visits for primary and secondary asthma from 2002-2005 were from 12.1-13.5 per day (de-duplicated). The highest number of visits was observed in 2005 with 13.5 per day. The average number of yearly visits for asthma were from 4,348(24.0%) - 4,943(27.0%) with the highest number of visits occurring in December-March with a second peak noted in October (Table 15). Similar findings were observed when disposition was taken into account. The highest number of asthma visits coincided with the fall and winter season where asthma morbidity has been shown to increase due to ragweed allergies and rhinovirus infection (Johnston & Sears, 2006; Dales, et al., 1996, Wansoo & Schneider, 2005). To determine whether the number of visits in this dataset were greater in the fall and winter months than spring and summer months, a Chi Square Goodness of Fit Test was used. The analysis showed that there was a statistically significant difference in the number of ED visits in the fall and winter months than spring and summer, $X^2(1, N=18,284)$, 33.4, $p < 0\ .05$ (Figure 15).

Table 15: Primary and secondary asthma emergency room visits (6 UPMC hospitals) for month and year by disposition

			Total Emergency room visits with primary and secondary asthma diagnosis (n=18,284) by disposition =emergency room visits with discharge or emergency room visit with admission					
Total emergency room visits with **primary and secondary asthma diagnosis (n=18,284) (De-Duplicated)**			Emergency room visits with **discharge (n=11,901)**			Emergency room visits with **admission (n=6,383)**		
Year	**n**	**%**	**Year**	**n**	**%**	**Year**	**n**	**%**
2002	4,422	24.0	2002	2,889	24.0	2002	1,533	24.0
2003	4,348	24.0	2003	2,865	24.0	2003	1,483	24.0
2004	4,570	25.0	2004	2,960	25.0	2004	1,610	25.0
2005	4,943	27.0	2005	3,186	27.0	2005	1,757	27.0
Total	**18,284**	**100.0**	**Total**	**11,901**	**100.0**	**Total**	**6,383**	**100.0**
Month	**n**	**%**	**Month**	**n**	**%**	**Month**	**n**	**%**
January	1,563	8.5	January	1,027	8.6	January	536	7.7
February	1,578	8.6	February	1,013	8.5	February	565	8.8
March	1,741	9.5	March	1101	9.2	March	640	10.0
April	1,536	8.4	April	967	8.1	April	569	8.8
May	1,424	7.7	May	934	7.8	May	490	7.6
June	1,469	8.0	June	996	8.3	June	473	7.3
July	1,484	8.1	July	979	8.2	July	505	7.8
August	1,364	7.4	August	854	7.1	August	510	7.9
September	1,474	8.0	September	1,005	8.4	September	469	7.3
October	1,576	8.6	October	1,039	8.7	October	537	8.3
November	1,432	7.8	November	961	8.0	November	471	7.3
December	1,643	8.9	December	1025	8.6	December	618	9.6
Total	**18,284**	**100.0**	**Total**	**11,901**	**100.0**	**Total**	**6,383**	**100.0**

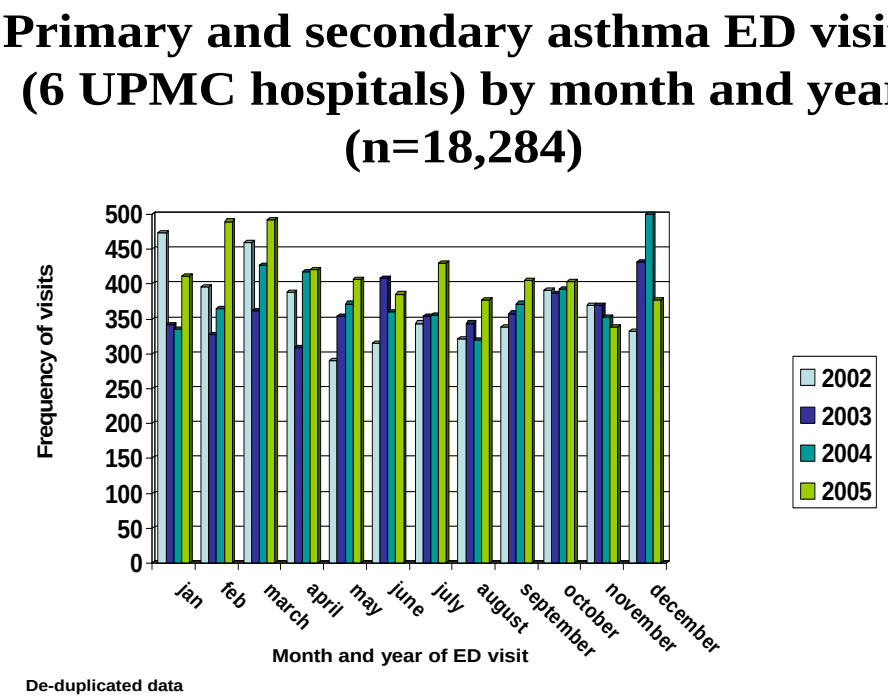


Figure 15: Primary and secondary asthma ED visits (6 UPMC hospitals) by month and year

A review of the frequency of ED visits to the different hospitals by year showed that the highest number of total ED visits were observed at St. Margaret's 4,279(23.4%) and Presbytarian Hospital 3,307(18.0%). Southside Hospital had the least number of visits 1710(9.3%). This pattern was consistent by year for the four-year period (Table 16).

Table 16: Primary and secondary asthma emergency room visits for 6 UPMC hospitals by year

Hospital	2002	2003	2004	2005	Total %
Braddock	782	779	755	795	3,111(17.0)
McKeesport	626	652	637	718	2,633(14.4)
Presbytarian	806	748	862	893	3,307(18.0)
Shadyside	745	752	824	921	3,242(17.7)
St Margaret's	1,025	1,030	1,064	1,160	4,279(23.4)
Southside	438	388	428	456	1,710(9.3)
Total* (%)	**4,422(24.1)**	**4,347(23.7)**	**4,570(24.9)**	**4,943(27.0)**	**18,284(100.0)**

The geographical distribution of asthma cases is another important feature in understanding the determinants of this disease. The data in the primary and secondary asthma ED visit dataset was available for 16,178 zip codes of which, 8,492 (52.4%) were identified within the COP boundaries, and the remaining 7,686(47.5%) zip codes were in Allegheny County, excluding the COP, respectively. 142 zip codes are included in the boundaries of Allegheny County. The total population, which resides in these zip codes is 1,259,337 persons. The population in the COP is 322,177 (US Census Data, 2000).

An analysis of the geographical distribution of these asthma cases showed that the highest number of cases was represented by individuals living in and around the City of Pittsburgh (COP) (Figure 16). The highest number of asthma ED visits observed by individuals living in the COP is of no surprise since this dataset represented 6 UPMC hospitals centered on the COP. Therefore, no conclusions can be made about the high proportion of asthma cases observed for the inner city residents and the possible role of environmental influences in these asthma cases.

The demographic subgroups for the total primary and secondary asthma emergency room visits for the population residing in the zip codes for the City of Pittsburgh vs. Allegheny County are highlighted in Tables 17 & 18. The highest frequency of ED visits by zip codes was observed in individuals 35-64 years of age, in Whites and females in both the COP and AC subgroups.

Table 17: Primary and secondary asthma emergency room visits for the City of Pittsburgh (zip codes) by age, sex and race 2002-2005 (only 6 hospitals of UPMC network)

Age			Sex		Total
			Male	Female	
0-17	Race	Black	190	146	336
		White	223	236	459
		Other	8	6	14
Total			421	388	809
18-34	Race	Black	342	574	916
		White	496	840	1,336
		Other	17	24	, 41
Total			855	1,438	2,293
35-64	Race	Black	427	1,061	1,488
		White	650	1,438	2,088
		Other	21	24	, 45
Total			1,098	2,523	3,621
65+	Race	Black	91	279	370
		White	426	951	1,377
		Other	16	35	51
Total			526	1,243	1,769
Total			**2,900**	**5,592**	**n= 8,492(52.4%)**

Table 18: Primary and secondary asthma emergency room visits for Allegheny County (zip codes) minus the City of Pittsburgh by age, sex and race (2002-2005) (only 6 hospitals in the UPMC network)

Age			Sex		Total
			Male	Female	
0-17	Race	Black	338	262	600
		White	405	281	686
		Other	14	7	21
Total			757	550	1,307
18-34	Race	Black	207	367	574
		White	421	820	1,247
		Other	6	11	17
Total			634	1,198	1,832
35-64	Race	Black	194	590	784
		White	685	1,607	2,292
		Other	1	20	21
Total			880	2,217	3,097
65+	Race	Black	44	117	161
		White	375	901	1,276
		Other	6	7	13
Total			425	1,025	1,450
Total			**2,696**	**4,990**	**n = 7,686 (47.5%)**

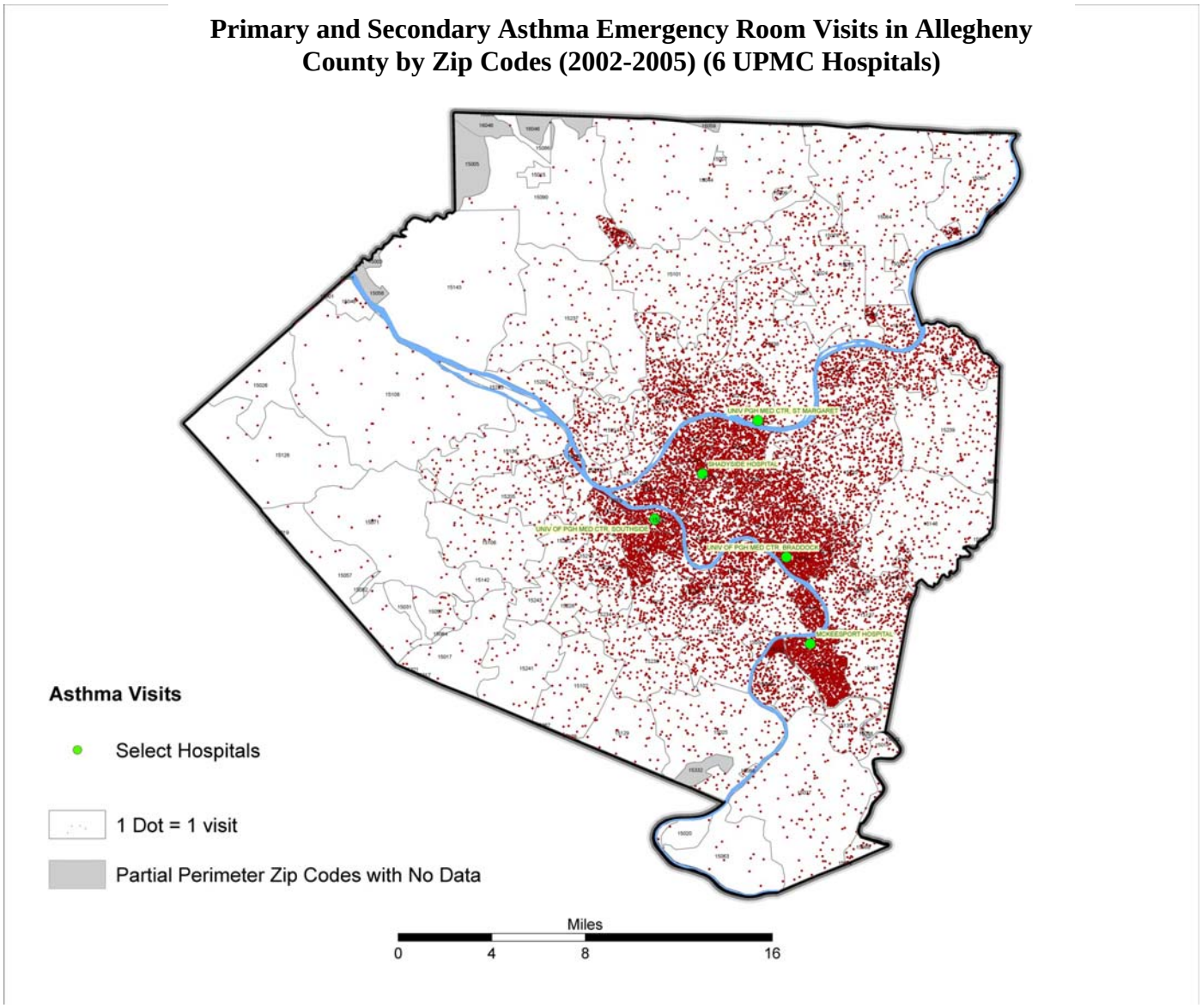


Figure 16: Primary and Secondary Asthma Emergency Room Visits in Allegheny County by Zip Code (2002-2005) for 6 UPMC Hospitals

** Available data on 16,178 zip codes. Presbytarian Hospital included in dataset, however, the hospital not listed on the map.*

3.4.1.2 Population B

A cohort of asthmatics with primary asthma were identified (n=5,100). This represented any one who presented to the ED from 2002-2005 with a first primary asthma visit. The age range was between 0-103 with the mean age distribution (M= 43.94, SD= 18.40) (Table 19). Individuals who were more likely to be admitted for asthma were older (M= 60.05, SD = 18.54) vs. individuals who were discharged from the ED (M = 39.53, SD= 15.72). In each age group, except for those 65 years of age and over, a visit to the emergency room was more likely to result in a discharge than an admission.

Gender differences were identical to those found in the primary and secondary asthma data set with females having higher number of ED visits 68.9%(95% CI 67.85-69.99) than males 31.1%(95% CI 29.79-32.33). Females presenting to the ED were more likely to be admitted than males 73.8% (95% CI 71.1-76.3) vs. 26.2%(95% CI 23.6-28.8), respectively.

An analysis of the data by race showed a higher proportion of ED visits in Whites than Blacks and Others, 63.2% (95% CI 61.9-64.5); 35.0%(95% CI 33.7-36.3); and 1.1%(95% CI .9-1.5), respectively. In this dataset, Whites were more likely to visit the ED and be admitted than Blacks (Table 19). No conclusions can be made about these findings since the White population in this sample represents a higher proportion of cases from the catchment area. The sample represents only 60% of the UPMC network, and 30% of all hospitals in AC. The hospitals represented in this sample were concentrated in and around the COP. The Black population in the COP is 30% and 10% in AC (US Census, 2006). Therefore, without an asthma population entirely representative of AC residents inclusive of the entire network of hospitals in AC, no conclusions can be made.

Table 19: First primary asthma emergency room visits (6 UPMC hospitals) by age, sex, and race by disposition

Total emergency room visits **first primary asthma diagnosis (n=5,100)**				Total emergency room visits **first primary asthma diagnosis with an emergency room discharge (n=4,005)**				Total emergency room visits **first primary asthma diagnosis with an admission (n=1,095)**			
Age	**n**	**%**	**CI**	**Age**	**n**	**%**	**CI**	**Age**	**n**	**%**	**CI**
0-17				0-17				0-17			
18-29	1338	26.2	25.0-27.4	18-29	1,268	31.6	30.2-33.1	18-29	70	6.3	5.0-8.0
30-44	1537	30.1	28.0-31.4	30-44	1,360	33.9	32.5-35.4	30-44	177	16.1	14.1-18.4
45-64	1415	27.7	26.5-28.9	45-64	1,052	26.2	24.9-27.6	45-64	363	33.1	30.4-35.9
65+	810	16.0	14.9-16.0	65+	325	33.0	31.6-34.5	65+	485	44.2	41.3-47.2
Mean age= 43.94(SD 18.40) Range=18-103				Mean age= 39.53(SD 15.72) Range=18-96				Mean age= 60.05(SD 18..54) Range= 18-103			
Total	**5,100**	**100.0**		**Total**	**4,005**	**100.0**		**Total**	**1,095**	**100.0**	
Sex	**n**	**%**	**CI**	**Sex**	**n**	**%**	**CI**	**Sex**	**n**	**%**	**CI**
Male	1,584	31.1	29.79-32.33	Male	1,297	32.4	30.9-33.8	Male	287	26.2	23.6-28.8
Females	3,516	68.9	67.85-69.99	Female	2,708	67.6	66.1-69.0	Female	808	73.8	71.1-76.3
Total	**5,100**	**100.0**		**Total**	**4,005**	**100.0**		**Total**	**1,095**	**100.0**	
Race	**n**	**%**	**CI**	**Race**	**n**	**%**	**CI**	**Race**	**n**	**%**	**CI**
Black	1,789	35.0	33.7-36.3	Black	1,512	37.8	36.2-39.2	Black	274	25.0	22.2-27.6
White	3,225	63.2	61.9-64.5	White	2,419	60.4	58.8-61.9	White	806	73.6	70.9-76.1
Other	61	1.1	.9-.1.5	Other	55	1.4	1.05-1.78	Other	6	.5	.25-1.18
Unknown	28	.5	.38-.77	Unknown	19	.5	.30- .73*	Unknown	9	.8	.43-1.55
Total	**5,100**	**100.0**		**Total**	**4,005**	**100.0**		**Total**	**1,095**	**100.0**	

The impact of insurance type on asthma ED visits was examined. Insurance types included: Medicare, Medicaid, Managed Care enrollees, and Self-pay (Table 20). Analysis by disposition for each insurance type demonstrated that a higher proportion of individuals had managed care insurance with the exception of the Medicare population. Individuals with traditional Medicare insurance were more likely to visit the ED, and be admitted than individuals with Medicare Managed Care insurance. Landon et al (2004) report that this phenomenon may likely represent a healthier cohort in the Medicare Managed Care group who report a higher use of prevention services than traditional Medicare enrollees. Furthermore, a correlation has been drawn between Medicaid Health Maintenance Organization (HMO) plan enrollment and a reduction in emergency room utilization and hospitalizations, and an increase use of outpatient services for asthma (Baker & Afendulis, 2005). A reduction in ED services and hospitalizations was not observed for the Medicaid HMO group in this dataset. This group had a higher proportion of ED visits than the Medicaid group. This higher proportion of visits could be related to lower co-pay for the Medicaid managed care population. Overall individuals who were discharged from the hospital were younger (M = 39.53, SD= 15.72) than individuals who were admitted (M= 60.05, SD= 18.54). An in-depth analysis of the different insurance types, by disposition and demographic subgroups is needed to gain a better understanding of how these factors may influence asthma outcomes.

Table 20: Insurance type for all first primary asthma emergency visits (6 UPMC hospitals) by disposition

Total emergency room visits **First primary asthma diagnosis (n=5,100)**				Total emergency room visits **first primary asthma diagnosis with an emergency room discharge (n=4,005)**				Total emergency room visits **first primary asthma diagnosis with an admission (n=1,095)**			
Mean age= 43.94(SD 18.40) Range=18-103				Mean age= 39.53(SD 15.72) Range=18-96				Mean age= 60.05(SD 18..54) Range= 18-103			
Financial	**n**	**%**	**CI**	**Financial**	**n**	**%**	**CI**	**Financial**	**n**	**%**	**CI**
Medicare	747	14.6	13.7-15.6	Medicare	368	9.1	8.3-10.1	Medicare	379	34.6	31.8-37.4
Medicare MC	474	9.2	8.5-10.1	Medicare MC	292	7.2	6.5- 8.1	Medicare MC	182	16.2	14.5-18.9
Medicaid	176	3.4	2.9-3.9	Medicaid	136	3.3	2.8- 4.0	Medicaid	40	3.6	2.69-4.93
Medicaid MC	872	17.0	16.0-18.1	Medicaid MC	743	18.5	17.3-19.7	Medicaid MC	251	22.9	20.5-25.5
Commercial	204	4.0	3.4-4.5	Commercial	188	4.6	4.0-5.3	Commercial	16	1.4	.90-2.35
Commercial MC	923	18.0	17.0-19.1	Commercial MC	796	19.8	18.6-21.1	Commercial MC	127	11.5	9.8-15.6
Third PP	977	19.1	18.1-20.2	Third PP	814	20.3	19.1-21.5	Third party payer	163	14.8	12.9-17. 1
Third PPMC	0			Third PPMC	0			Third PPMC	0		
Self pay	680	13.3	12.4-14.2	Self pay	638	15.9	14.8-17.0	Self pay	42	3.8	2.8-5.1
Other	45	.88	.66-1.1	Other	30	.7	.5-1.06	Other	15	1.3	.8.-2.24
VA	2	.03	.01-.14	VA	0			VA	11	1.00	.56-1.7
Total	**5,100**	**100.0**		**Total**	**4,005**	**100.0**		**Total**	**1,095**	**100.0**	

The annual daily average of ED visits for a primary asthma diagnosis was between 3.18 and 4.15 per day. Yearly ED visits for 2002-2005 were from 1,090(21.4%) to 1,455(28.5%) (Table 21). The highest reported number of visits occurred in December– March with a second peak beginning in September. A significant difference was observed between the number of ED visits in the fall and winter months vs. the spring and summer, $X^2(1, N = 5{,}100) = 70.5, p < .05$. These seasonal differences may be related to environmental influences from a rise in weed pollens and rhinovirus infections occurring during the fall and winter months (Dales, et al., 1996; Johnston & Sears, 2006; Wansoo & Schneider, 2005). Both ED visits with discharge and admissions during the fall and winter months are greater (Figure 17).

Table 21: All first primary asthma emergency room visits (6 UPMC hospitals) by month, year and disposition

Total emergency room visits **First primary asthma diagnosis (n=5,100)**			Total emergency room visits **first primary asthma diagnosis with an emergency room discharge (n=4,005)**			Total emergency room visits **first primary asthma diagnosis with an admission (n=1,095)**		
Year	**n**	**%**	**Year**	**n**	**%**	**Year**	**n**	**%**
2002	1,455	28.5	2002	1,182	30.0	2002	273	27.1
2003	1,480	29.0	2003	1,201	30.0	2003	279	25.4
2004	1,075	21.1	2004	821	20.0	2004	254	23.1
2005	1,090	21.4	2005	801	20.0	2005	289	26.3
Total	**5,100**	**100.0**	**Total**	**4,005**	**100.0**	**Total**	**1,095**	**100.0**
Month	**n**	**%**	**Month**	**n**	**%**	**Month**	**n**	**%**
January	488	9.6	January	367	9.1	January	121	11.0
February	507	9.6	February	368	9.1	February	129	11.7
March	489	9.6	March	364	9.0	March	125	11.4
April	422	8.3	April	327	8.1	April	95	8.6
May	432	8.5	May	345	8.6	May	87	7.9
June	403	7.9	June	333	8.3	June	70	6.3
July	287	5.6	July	235	5.8	July	52	4.7
August	303	5.9	August	244	6.0	August	59	5.3
September	424	8.3	September	364	9.0	September	60	5.4
October	461	9.0	October	370	9.2	October	91	8.3
November	416	8.2	November	336	8.3	November	80	7.3
December	468	9.2	December	352	8.7	December	116	10.5
Total	**5,100**	**100.0**	**Total**	**4, 005**	**100.0**	**Total**	**1,095**	**100.0**

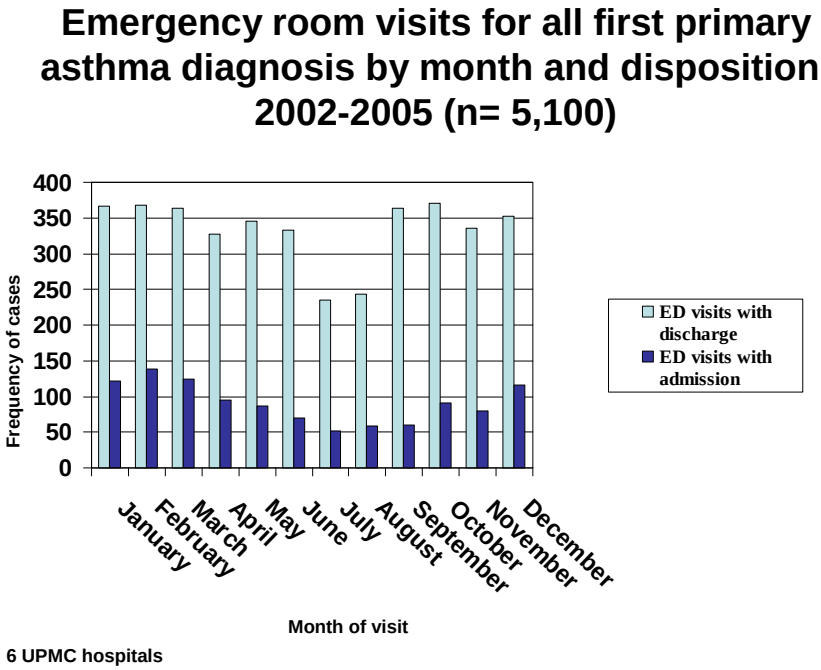


Figure 17: Emergency room visits for all first primary asthma by month and disposition

The highest number of visits was observed in and around the COP with St. Margaret's 1201(23.5%) and Shadyside Hospital's 917(18.0%) having the highest number of ED visits (Table 22). Southside Hospital had the least number of visits 489(9.6%). The observed differences by hospitals may be solely due to individuals choosing the larger emergency departments closest to the city for asthma care or a referral bias by practitioners treating asthma patients.

Table 22: First primary asthma emergency room visit for 6 UPMC hospitals

Hospital	Total %
Braddock	892(17.5)
McKeesport	732(14.4)
Presbytarian	869(17.0)
Shadyside	917(18.0)
St Margaret's	1,201(23.5)
Southside	489(9.6)
Total (%)	**5,100(100.0)**

The geographic distribution of primary asthma ED visits were similar to patterns seen with the primary and secondary asthma dataset with the highest number of cases concentrated around the city limits (Figure 18). This is no surprise since the 6 hospitals in this study were concentrated around the COP.

Emergency room visits for primary asthma (n=5,100) for 6 UPMC hospitals in Allegheny County

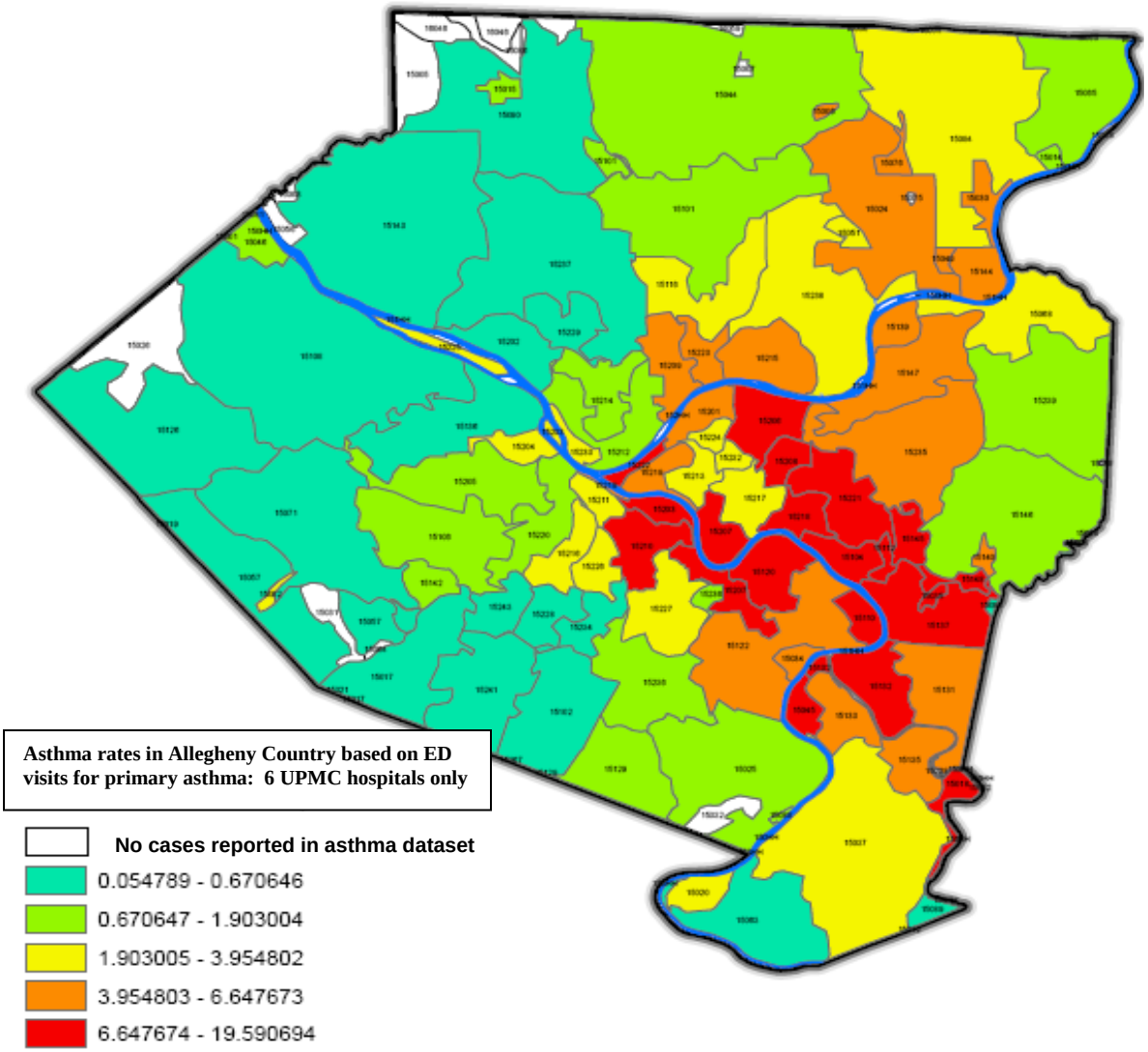


Figure 18: Emergency room visits for primary asthma (n=5,100) for 6 UPMC hospitals in Allegheny County

3.4.1.3 Population C

A 10% random sample was taken from the first primary asthma dataset. This random sample represented a total of (n=488) adults. The age range was from 18-103, (M=43.87, SD=19.08). The proportion of individuals within each age group was fairly similar with the exception of individuals 65± who were represented by the smallest proportion of individuals 82(16.8%) (Table 23). The mean age for individuals discharged from the hospital were much less than for those who were admitted (M= 38.91, SD= 16.22), (M= 62.38, SD= 17.61), respectively. A higher frequency of ED visits was observed for females than males 69.3%(95% CI 65-73.1) vs. 30.7%(95% CI 26.8-34.9), respectively. This same pattern was observed for gender by disposition.

Differences were observed by race with Whites (58.6%) having a higher proportion of emergency room visits than Blacks (39.8%) and Others, (1.6%). The differences were observed for individuals whether they were discharged or admitted. The higher proportion of whites reflects the higher number of Whites in the catchment area. The hospitals in this research study were drawn from the UPMC network, which represented 6 hospitals centered on the COP. The Black population in the COP is 30% and 10% in AC (US Census, 2006). Therefore, without an asthma population representative of AC residents inclusive of the entire network of hospitals in AC, no conclusions can be made about racial differences in this population.

Table 23: Random Sample- First Primary asthma emergency room visits (6 UPMC hospitals) by age, sex, and race by disposition

Total emergency room visits **first primary asthma diagnosis (n=488)**				Total emergency room visits **first primary asthma diagnosis with an emergency room discharge (n=385)**				Total emergency room visits **first primary asthma diagnosis with an admission (n=103)**			
Age	**n**	**%**	**CI**	**Age**	**n**	**%**	**CI**	**Age**	**n**	**%**	**CI**
0-17	0			0-17	0			0-17	0		
18-29	137	28.0	24.2-32.2	18-29	134	34.8	30.2-39.6 #	18-29	3	2.9	1.05-8.19*#
30-44	139	28.4	24.6-32.6	30-44	126	32.7	28.2-37.5 #	30-44	13	12.6	7.5-2.0 *#
45-64	130	26.6	22.9-30.7	45-64	94	24.4	20.3-28.9*	45-64	36	34.9	26.4-44.5
65	82	16.8	13.7-20.3	65+	31	8.0	5.7-11.2*#	65+	51	49.5	40.0-59.0 #
Mean age= 43.87(SD 19.08) Range=18-103				Mean age= 38.91(SD 16.22) Range=18-93				Mean age= 62.38(SD 17.61) Range=19-103			
Total	**488**	**100.0**		**Total**	**385**	**100.0**		**Total**	**103**	**100.0**	
Sex	**n**	**%**	**CI**	**Sex**	**n**	**%**	**CI**	**Sex**	**n**	**%**	**CI**
Male	150	30.7	26.8-34.9	Male	119	30.9	26.5-35.6	Male	31	30.0	22.0-39.5
Female	338	69.3	65-73.1	Female	266	69.1	64.3-73.4	Female	72	70.0	60.4-77.9
Total	**488**	**100.0**		**Total**	**385**	**100.0**		**Total**	**103**	**100.0**	
Race	**n**	**%**	**CI**	**Race**	**n**	**%**	**CI**	**Race**	**n**	**%**	**CI**
Black	194	39.8	35.5-44.1	Black	169	43.8	39.0-48.8	Black	25	24.2	17.0-33.4
White	286	58.6	54.1-62.8	White	209	54.2	49.2-59.1	White	77	74.7	65.5-82.1
Other	6	1.2	.57-2.65	Other	6	1.5	.73-3.35	Other	0		
Unknown	2	.4	.12-1.46	Unknown	1	.25	.06-1.43	Unknown	1	.25	.23-5.24
Total	**488**	**100.0**		**Total**	**385**	**100.0**		**Total**	**103**	**100.0**	

Insurance type was reported for all first primary asthma visits (Table 24). Enrollees in all insurance categories except Medicare were more likely to be discharged than admitted. This data was consistent with the primary asthma dataset. Significant differences were observed between the mean ages for individuals who were discharged vs. admitted, (M= 38.91, SD = 16.22) vs. (M= 62.38, SD = 17.61), respectively, $t(486)$, = 12.80, $p < 0.03$ (two-tailed). Older adults were more likely to be admitted. However, this is an abbreviated representation of ED visits of which no significant conclusions can be drawn. A larger sample of ED visits is need to conduct further analysis looking at the different insurance types, and factors influencing health service utilization.

Table 24: Random Sample- Insurance status for all first primary asthma emergency room visits by disposition

Total emergency room visits **first primary asthma diagnosis (n=488)**				Total emergency room visits **first primary asthma diagnosis with an emergency room discharge (n=385)**				Total emergency room visits **first primary asthma diagnosis with an admission (n=103)**			
Mean age= 43.87(SD 19.08) Range=18-103				Mean age= 38.91(SD 16.22) Range=18-93				Mean age= 62.38(SD 17.61) Range=19-103			
Financial	**n**	**%**	**CI**	**Financial**	**n**	**%**	**CI**	**Financial**	**n**	**%**	**CI**
Medicare	70	14.3	11.5-17.7	Medicare	31	8.0	5.7-11.2	Medicare	39	37.8	29.0-47.5
Medicare MC	48	9.8	7.5-12.8	Medicare MC	28	7.2	5.0-10.3	Medicare MC	20	19.4	12.9-28.1
Medicaid	14	2.8	1.7-4.7	Medicaid	14	3.6	2.1-6.0	Medicaid	1	.97	.23-5.24
Medicaid MC	78	15.9	13.0-19.5	Medicaid MC	69	17.9	14.4-22.0	Medicaid MC	7	6.7	3.3-13.3
Commercial	20	4.0	2.6-6.2	Commercial	17	4.4	2.7-6.9	Commercial	3	2.9	1.05-8.19
Commercial MC	92	18.8	15.6-22.5	Commercial MC	73	18.9	15.3-23.1	Commercial MC	20	19.4	12.9-28.1
Third PP	92	18.8	15.6-22.5	Third PP	86	22.3	18.4-26.7	Third party payer	6	5.8	2.7-12.1
Third PPMC	0			Third PPMC	0			Third PPMC	0		
Self pay	68	13.9	11.1-17.2	Self pay	65	16.8	13.4-20.9	Self pay	3	2.9	1.05-8.19
Other	7	1.4	.70-2.9	Other	5	1.27	.57-2.99	Other	2	1.0	.59-6.7
VA	0			VA	0			VA	0		
Total	**488**	**100.0**		**Total**	**385**	**100.0**		**Total**	**103**	**100.0**	

A review of the ED visits by month and year demonstrated that the highest numbers of ED visits were observed in October-December and March-June. No significant difference was observed between the number of ED visits which occurred in the fall and winter months vs. the spring and summer months, $X^2(1, N=488) = .03$, $p > .05$ (Figure 19).

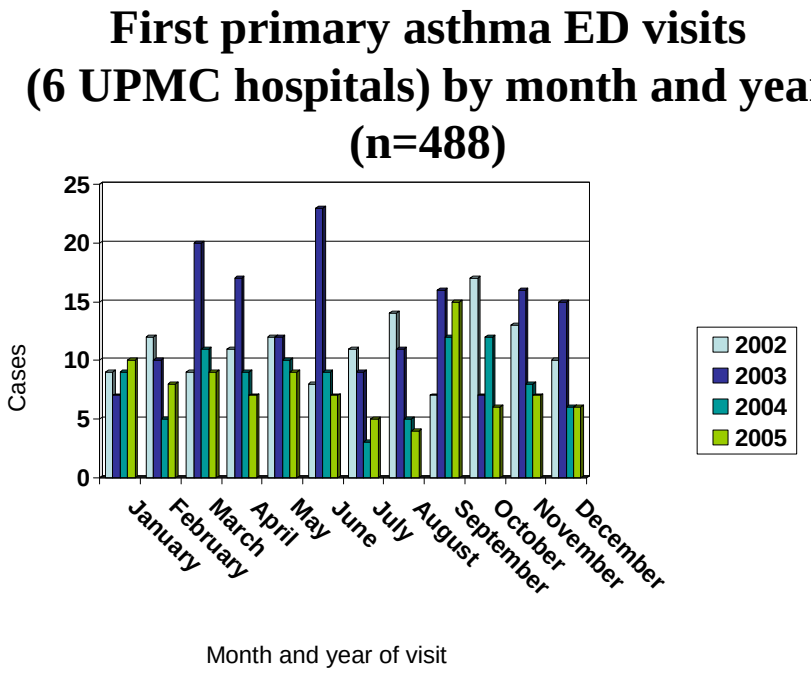


Figure 19: Random Sample: First primary asthma visits by month and year

An analysis of the data was conducted for each hospital by year. St. Margaret's Hospital had the highest number of ED visits 116(23.7%), followed by Presbytarian 90(18.4%), and Braddock 85(17.4%). Southside Hospital had the least number of visits 42(8.6%) (Table 25).

Table 25: Random Sample –First Primary asthma emergency room visits by hospital and year

Hospital	2002	2003	2004	2005	Total %
Braddock	27	27	13	18	85(17.4)
McKeesport	18	22	16	19	75(15.3)
Presbytarian	27	20	22	21	90(18.4)
Shadyside	23	33	16	8	80(16.3)
St Margaret's	28	46	21	21	116(23.7)
Southside	10	15	11	6	42(8.6)
Total*(%)	**133(27.5)**	**163(33.4)**	**99(20.2)**	**93(19.5)**	**488(100.0)**

*includes 6 UPMC hospitals

3.4.2 Comparison ICD-9 Coded chief complaint: Admitting vs. Hospital discharge diagnosis for asthma

Several studies have assessed the validity of using the ICD 9 coded chief complaint to study disease (Wagner, et al, 2004; Espino & Wagner, 2001). Available to the researcher in this dataset was the ICD 9 coded chief complaint admitting diagnosis and hospital discharge diagnosis. An analysis was conducted to determine which diagnosis improved the sensitivity of identifying a case of asthma. The admission diagnosis is sort of a working hypothesis of the physician's conclusion of the best explanation that fits the patient's symptoms as they enter the hospital. A discharge diagnosis is most often the diagnosis which best fits the patient's symptoms, history, examination and test results (S. McLinden, personal communication, February 8, 2007). Wagner et al (2004) report that as a patient moves through the admission process, the diagnostic accuracy for assigning a discharge (primary) diagnosis to the patient increases as more information becomes available. To determine which diagnosis was better at identifying a case of asthma, a comparison of the ICD-9 coded chief complaint admitting diagnosis was compared to the hospital discharge diagnosis (Table 26). The analysis which relied on the admitting diagnosis to classify an ED visit demonstrated that 1,479 visits received an ICD 9 coded diagnosis for asthma [493.0-493.9], 5,001 visits were grouped under the signs and symptoms category using the ICD 9 code [780.01-799.3] for wheezing, shortness of breath, tachypnea, respiratory distress, cough and orthopnea, and 1,251 visits were given an ICD 9 code [460-519.9] for a respiratory related condition. When the discharge diagnosis was used to classify the visits, 6,024 visits received an ICD 9 coded diagnosis for asthma [493.0-493.9]. The number in the sign and symptoms category decreased from 5001 to 1376, respectively, which suggests that by using the admission diagnosis to study asthma a great deal of primary asthma

cases were not being captured. However, many were being assigned into the sign and symptoms category for which described the patient's complaint or rationale for admission. These findings suggest that the admission diagnosis is more consistent with the chief complaint or the term or phrase, which describes the patient's complaint on admission. As the patient moves through the admission process and into the discharge phase of admission, the discharge diagnosis became more specific for a diagnosis of asthma.

Table 26: A comparison of the ICD-9 Coded chief complaint admitting diagnosis to the primary hospital discharge diagnosis

Grouped by Admitting Diagnosis Chief Complaint	**ICD-9 Code**	**n**	**Grouped by Primary Diagnosis Chief Complaint**	**ICD-9 Code**	**n**
Infection/parasite	(005.9-136.9)	71	Infection/parasite	(005.9-136.9)	362
Neoplasm's	(146.9-239.6)	35	Neoplasm's	(146.9-239.6)	143
Endocrine/Renal	(242.0-276.8)	115	Endocrine/Renal	(242.0-276.8)	306
Blood	(280.0-288.0)	36	Blood	(280.0-288.0)	81
Mental Health	(291.0-312.9)	533	Mental Health	(291.0-312.9)	878
Nervous System	(320-388.7)	334	Nervous System	(320-388.7)	395
Circulatory	(401.0-459.9)	622	Circulatory	(401.0-459.9	1,268
Respiratory	(460-519.9)	1,251	Respiratory	(460-519.9)	1,667
Asthma	(493.0-493.9)	1479	Asthma	(493.0-493.9)	6,024
Digestive	(520.6-579.3)	502	Digestive	(520.6-579.3)	1,006
Genitourinary	(584.5-628.9)	188	Genitourinary	(584.5-628.9)	452
Preg comps	(640.93-674.8)	6	Preg comps	(640.93-674.8)	51
Cutaneous/skin	(680.3-708.9)	146	Cutaneous/skin	(680.3-708.9)	307
Musculoskeletal	(710.1-733.99)	1127	Musculoskeletal	(710.1-733.99)	515
Perinatal	(760.9)	1	Perinatal	(760.9)	9
Sign/symptoms	(780.01-799.3)	4,706	Sign/symptoms	(780.01-799.3)	1,041
Wheezing, SOB, tachypnea, respiratory distress cough & orthopnea	(786.0-786.9)	5,001	Wheezing, SOB, tachypnea, respiratory distress cough & orthopnea	(786.0-786.9)	1,376
Injury/poisonings	(802.4-998.89)	1,643	Injury/poisonings	(802.4-998.89)	2,276
Supplement Vcodes	(V01.6-V72.9)	478	Supplement Vcodes	(V01.6-V72.9)	117
Missing		9	Missing		9
Total		**18,284**	**Total**		**18,284**

3.5 VALIDATION STUDY RESULTS

3.5.1 Electronic Medical Record: Composition and Completeness

A total of 180 EMRs were included this analysis. The EMRs represented adults who had a first primary asthma visit to the ED from 2002-2005. This primary asthma diagnosis was ascertained through an electronic medical record abstract using the corresponding ICD-9 code for primary asthma (493.0-493.9). The review assessed for completeness of findings in the EMR by identifying the selected indicators and CSTE clinical and laboratory case definition criteria used to classify a case of asthma. The following 838 documents were reviewed as part of the EMR: clinical narrative reports, including free text history and physicals, dictated physicals, consultations, clinic follow up visits, ED notes, laboratory results, surgery reports and discharge summaries. Age was not reported in 16(9%) of the records, gender was absent in 5(3%), and race was missing in 105(58.3%). Complete data was found for chief complaint, patient type, visit type, admission and discharge date, and type of record. No information was provided in the record about the hospital where the patient was being treated, resident zip code status, or insurance type. The identifying dataset used to locate the electronic medical record provided all the missing de-identified data. This feature increased the sensitivity of findings to 99% for each specific de-identified data point.

In this research project it was necessary to have the medical record data from the previous year to determine if there was an occurrence of any asthma symptom reported at least three times in the past year for asthma. This created one year of additional records for the reviewer to consider as supporting evidence to assist in classifying a case of asthma. Having the EMR dataset made it very easy to determine if a patient had three visits or more in the past year

simply by ordering the records by admission date. If a patient did not have three or more visits in the past year noted by the date of admission, the patient automatically did not meet the requirements to be classified into the confirmed asthma category. Therefore, all the remaining medical records were used to determine whether a case of asthma could be corroborated through existing documentation and to identify the respective probable and possible asthma classification category to which the patient belonged. This step took the most time, particularly when a case of asthma could not be substantiated by the documentation found in the ED visit summary. The automated feature of the medical record provided quick access to the record of interest, ease of navigation through the multiple record types, and the ability to revisit any supporting document to classify an asthma case. Since this process had to be repeated numerous times due to the overlap in classification criteria, and late addition of adding the comprehensive list of asthma medications to the study, the automated feature reduced the review time in half for each review cycle. However, the complexities surrounding the use of the EMRs were challenging. Since there is no standardized method for documenting in the medical record, good clinical judgment was necessary to sort out all contradictory statements found while reviewing the records. Follow-up consultations and progress notes associated with the respective ED visit, were not always in chronological order. This made it very confusing to the reviewer when an asthma sign and symptom or characteristic pertaining to the original ED visit were contradicted in another part of the medical record. The date of service rendered by the date marked on the record was used as the link to the initial asthma ED visit. If the date of the consultation or progress notes were not the same as the ED visit such as in the case of a visit originating on one day, and the corresponding progress note written after midnight, this situation required a great deal of time to sort out whether the record belonged to a new visit or an addendum to the existing visit. If an

asthma diagnosis was in question and no information pertaining to the diagnosis could be ascertained from the ED visit note, all the supporting records were used as evidence to classify the asthma diagnosis.

Overall, the use of the EMR provided the researcher the ability to have immediate access to key data. The automated feature made it easy to select records of interest and move freely about the dataset without a large paper trial of information. The use of multiple record types increased the sensitivity of findings for specific endpoints such as an asthma medication. Furthermore, the complex nature of this research study involving all previous 1-year visits coupled by the multiple records associated with each ED visit, and need to discriminate the broad range of criteria designated by the CSTE clinical and laboratory classification definition underscores the need to have access to data in an automated format. The use of the EMR greatly improved the ability to carryout this research in a timely, efficient manner. Moreover, the EMR served as a comprehensive tool to conduct epidemiological research on asthma.

3.5.1.1 Electronic Medical Record: Descriptive analysis

Of the total 180 ED visits, 146(81.1%) were confirmed discharges and 34(18.9%) admissions. Emergency department visits were reported for all six hospitals with Presbytarian 44(24.4%) and St. Margaret's Hospital 38(21.1%) having the highest number of ED visits and Southside Hospital 14(7.7%) the least (Table 27).

Table 27: Electronic Medical Record: First primary asthma ED visit by hospital and disposition

Hospital	Frequency of ED visit with discharge	Frequency of ED visit with admission	Total %
Braddock	24	5	29(16.1)
McKeesport	21	6	27(15.0)
Presbytarian	35	9	44(24.4)
Shadyside	22	6	28(15.5)
St Margaret's	32	6	38(21.1)
Shadyside	12	2	14(7.7)
Total*(%)	**146(81.0)**	**34(18.9)**	**180(100.0)**

** includes 6 UPMC hospitals*

A breakdown of the demographic characteristics of this cohort included individuals whose age ranged from 18-103 (M=46.71, SD=20.63). The frequency of visits was equally distributed across all age groups (Table 28). A gender difference was observed with female ED utilization being over twice that of males (70%) to (30%), respectively. Racial differences were observed for individuals presenting to the emergency department. Whites reported a higher frequency of ED visits than Blacks and Others, 56.1% to 43.3% and 0.5%, respectively. This is not a complete representation of the proportion of Whites and Blacks in Allegheny County as this represents only a small proportion of asthma cases from the six hospitals in this sample. These hospitals are located around the COP. The Black population in the COP is 30% and 10% in AC as a whole (US Census, 2006). Therefore, no conclusions can be made on these data based on the limited representation of hospitals in the UPMC network and county as a whole.

However, a comparison of the demographic characteristics of the population identified through the EMRs was made to confirm that this population was representative of the 488 primary asthma cases identified in the random sample, and the total primary asthma dataset

(n=5,100). The age distribution of the EMR population is almost identical to the random sample and primary asthma dataset (Appendix F). The proportion of cases by sex and race are equal across demographic subgroups.

Table 28: Electronic Medical Record: First primary asthma ED visit by age, sex and race

EMR emergency room visits with first primary asthma diagnosis (n=180)		
Age	**n**	**%**
0-17		
18-29	41	22.7
30-44	49	27.2
45-64	50	27.7
65+	40	22.2
Mean 46.71(SD 20.20)		
Range 18-103		
Total	**180**	**100.0**
Sex	**n**	**%**
Male	54	30.0
Female	126	70.0
Total	**180**	**100.0**
Race	**n**	**%**
Black	78	43.3
White	101	56.1
Other	1	.5
Unknown	0	
Total	**180**	**100.0**

3.6 ASTHMA SELECTED INDICATOR RESULTS

3.6.1 Manual Review

The manual review of the medical record was conducted to identify the selected indicators proposed to validate a case of asthma for this research study. The medical record review is considered to be the gold standard methodology. This process took approximately 80 hours of review time. The findings were validated using two automated applications by which the gold standard methodology results would be compared.

The following is a summary of the results of the medical chart review (Table 29). Of the clinical signs and symptoms and characteristics reported for asthma, shortness of breath was reported most frequently 128(71.1%), followed by wheezing 104(57.7%), and cough 102(56.7%). Surprisingly, wheezing the most frequently reported symptom for asthma was not present in 16(8.8%) of the visits and not mentioned in 60(33.3%). Rhinitis and fever were present almost equally 20(11.1%) to 19(10.5%), respectively. Rhinitis has been reported to result in low sensitivity for predicting asthma (Hung et al., 1998). Rhinitis was not mentioned in 148(82.2%) of the ED visits. Fever, a sign which is routinely assessed on every ED visit was not mentioned in 53(29.4%) of the cases. This finding could have resulted from an oversight on the reviewers behalf since there is a numeric component and several keywords that represent the word fever which may not have been picked up on initial chart review. The presence of an asthma history was found in 120(66.7%) records and was not mentioned in 58(32.2%). Two records were marked for no history of asthma.

The indicator, which had the highest level of sensitivity for predicting a diagnosis of asthma, was current use of an asthma medication. A total of 140(77.7%) of all patients presenting to the ED were on an asthma medication. Only eleven persons were not using any asthma medication prior to their ED visit and 29(16.1%) of the cases had no mention of an asthma medication. Individuals who were using their inhaler on presentation to the emergency department were counted as presence of being on an asthma medication. Individuals whose medical record did not contain any information other than given a nebulizer treatment in the emergency department with no evidence to substantiate an asthma diagnosis including, previous mention of an asthma medication was marked as absence of being on an asthma medication. Previous research has suggested that a nebulizer treatment without any further substantiate evidence to back up an asthma diagnosis including the past use of an asthma medication results in a low (27%) positive predictive value for asthma (Vollmer, 2004).

Table 29: Selected Indicators: Manual Review results

Selected Indicators	Total n=180
Wheezing	
0 absence	16(8.8)
1 presence	104(57.7)
2 not mentioned	60(33.3)
Total	180(100.0)
Cough	
0 absence	27(15.0)
1 presence	102(56.7)
2 not mentioned	51(28.3)
Total	180(100.0)
Shortness of Breath	
0 absence	15(8.3)
1 presence	128(71.1)
2 not mentioned	37(20.5)
Total	180(100.0)
Rhinitis	
0 absence	12(6.7)
1 presence	20(11.1)
2 not mentioned	148(82.2)
Total	180(100.0)
Fever	
0 absence	108(60.0)
1 presence	19(10.6)
2 not mentioned	53(29.4)
Total	180(100.0)
History of asthma	
	2(1.1)
0 absence	120(66.7)
1 presence	58(32.2)
2 not mentioned	180(100.0)
Total	
Asthma Medications	
0 absence	11 (6.1)
1 presence	140(77.7)
2 not mentioned	29(16.1)
Total	180(100.0)

3.6.2 Manual vs. Automation Review

Two automated methodologies were employed to validate the data extracted from the medical record on manual chart review. Each knowledge base methodology provided different approaches at structuring the data for scientific interpretation. However, the goal was to find a methodology equal or superior to the gold standard manual chart review. A comparison of the results from the three methodologies is provided in Table 30.

Table 30: Selected Indicators results: Manual vs. Automation

Indicators	Manual Chart Review	PERL	MedLEE*
Wheezing			
0 absence	16(8.8)	14(7.8)	18(10.0)
1 presence	104(57.7)	117(65.0)	84(46.7)
2 not mentioned	60(33.3)	49(27.2)	78(43.3)
Total	180(100.0)	180(100.0)	180(100.0)
Cough *			
0 absence	27(15.0)	29(16.1)	31(17.2)
1 presence	102(56.7)	103(57.2)	97(53.8)
2 not mentioned	51(28.3)	48(26.7)	52(28.9)
Total	180(100.0)	180(100.0)	180(100.0)
Shortness of Breath			
0 absence	15(8.3)	26(14.4)	13(7.2)
1 presence	128(71.1)	131(72.8)	112(62.2)
2 not mentioned	37(20.5)	23(12.8)	55(30.6
Total	180(100.0)	180(100.0)	180(100.0)
Rhinitis*			
0 absence	12(6.7)	0	14(7.8)
1 presence	20(11.1)	5(2.8)	21(11.7)
2 not mentioned	148(82.2)	175(97.2)	145(80.6)
Total	180(100.0)	180(100.0)	180(100.0)
Fever			
0 absence	108(60.0)	48(26.7)	126(70.0)
1 presence	19(10.6)	94(52.2)	25(13.9)
2 not mentioned	53(29.4)	38(21.1)	29(16.1)
Total	180(100.0)	180(100.0)	180(100.0)
History of asthma*			
0 absence	2(1.1)	3(1.7)	2(1.1)
1 presence	120(66.7)	53(29.4)	116(64.4)
2 not mentioned	58(32.2)	124(68.9)	62(34.4)
Total	180(100.0)	180(100.0)	180(100.0)
Asthma Medications *			
0 absence	11 (6.1)	2(1.1)	2(1.1)
1 presence	140(77.7)	141(78.3)	82(45.6)
2 not mentioned	29(16.1)	37(20.6)	96(53.3)
Total	180(100.0)	180(10.00)	180(100.0)

Combined UMLS Concept Codes for MedLee for selected indicators *
MedLee asthma medications results reported for Nebulizer/Inhalers only

The results for each classification concept for the presence; absence and not mentioned status for each indicator matched by PERL in comparison to the medical chart review were close (±10% difference) for all indicators except rhinitis, fever and past medical history (Table 30). Rhinitis was not a sensitive indicator for asthma as evident by the limited presence of this symptom 20(11.5%) reported in the medical record. PERL only matched on 5(2.8%) instances for the presence of rhinitis, zero matches in the absence category, and not mentioned in 175(97.2%) records. The percent difference observed between PERL and the medical chart review for rhinitis was 15% as noted in the not mentioned category 175(97.2%) vs. 148(82.2%), respectively. For the indicator fever, PERL matched on the presence of this finding in 94 records, whereas the manual chart review picked up nineteen-a difference of 80%. Furthermore, a fifty-five percent difference was reported between PERL and the medical chart review for the absent of fever 108(60.0%) vs. 48(26.7%), respectively. A past medical history of asthma, an indicator with a moderate to high degree of sensitivity was confirmed in 120(66.7%) records through manual chart review. PERL was only able to identify less than half of these cases 53(29.4%). PERL also reported that a history of asthma was not mentioned in 124(68.9%) of the EMRs. The results from the manual review showed that only 58(32.2%) of the cases had no past medical history of asthma. This represented a 53% discrepancy in finding. More favorably for PERL, the results for the asthma medication concept categories were almost identical to that of the medical chart review. PERL and the manual chart review results were nearly identical for the presence of an asthma medication 140(77.7%) vs. 141(78.3%) respectively. This finding is likely due to the highly specific asthma medication script created for PERL, which matched on the list of each asthma medication. A small difference was observed for instances where not

medications were found or were not mentioned, 2(1.1%) vs. 11(6.1%), and 37(20.6%) vs. 29(16.1%), respectively.

The findings from MedLee matched the manual chart review very closely for every indicator by category status with the exception of fever and asthma medications. A comparison of MedLee to the manual chart review for the absence of fever was 126(70.0%) vs. 108(60.0%), presence 25(13.9%) vs. 19(10.6%), and not mentioned status 29(16.1%) vs. 53(29.4%), respectively. PERL did not perform as well as MedLee against the medical chart review for fever; absence 48(26.7%), presence 94(52.2%), and not mentioned status 38(21.1%). The MedLee results for asthma medications were only reported for the mention of the word inhaler and nebulizer. This was a limitation for this study. Currently, there are 2 UMLS concept codes, which can be used to represent asthma medications, inhaler (C0021461), and nebulizer (C0027524). The use of an inhalers was present in 59(32.8%) of the records, absent in 2(1.1%) and not mentioned in 119(66.1%). The presence of a nebulizer was reported in 36(20%) of the records and not mentioned in 144(80%). The combined presence of these two concepts is reported in Table 31. The use of an inhaler and nebulizer were identified in 82(45.6%) of the EMRs. A limitation of using MedLee with the asthma project was the inability to have results included for the specific list of asthma medications. This was not discovered until a later phase of this research study. Since there are no UMLS concepts codes yet developed for the selected asthma medications, a proxy measure was reported for the combined results of an inhaler and nebulizer.

Table 31: MedLEE combined UMLS concept codes: Comparison to manual chart review and PERL

	Manual Chart Review	PERL	MEDLEE		
Cough			C0010200- All cough*	C0850149-dry	C0850149-productive
0 absence	27(15.0)	29(16.1)	3(17.2)	0	169(93.9)
1 presence	102(56.7)	103(57.2)	97(53.8)	20(11.1)	4(2.2)
2 not mentioned	51(28.3)	48(26.7)	52(28.9)	160(88.9)	7(3.9)
Total	180(100.0)	180(100.0)	180(100)	180(100.0)	180(100.0)
Rhinitis			Combined	C0035455-rhinitis	C1260880-runny nose
0 absence	12(6.7)	0	14(7.8)	1(.6)	13(7.2)
1 presence	20(11.1)	5(2.8)	21(11.7)	4(2.2)	19(10.6)
2 not mentioned	148(82.2)	175(97.2)	145(80.6)	175(97.2)	148(82.2)
Total	180(100.0)	180(100.0)	180(100.0)	180(100)	180(100.0)
History of asthma			Combined	C0455544-asthma hx	C0004096-asthma
0 absence	2(1.1)	3(1.7)	2(1.1)	3(1.1)	2(1.1)
1 presence	120(66.7)	53(29.4)	116(64.4)	65(36.1)	99(55.0)
2 not mentioned	58(32.2)	124(68.9)	62(34.4)	112(62.2)	79(43.9)
Total	180(100.0)	180(100.0)	180(100.0)	180(100.0)	180(100.0)
Asthma Medications			Combined	C0021461-inhaler	C0027524-nebulizer
0 absence	11 (6.1)	2(1.1)	2(1.1)	2(1.1)	0
1 presence	140(77.7)	141(78.3)	82(45.6)	59(32.8)	36(20.0)
2 not mentioned	29(16.1)	37(20.6)	96(53.3)	119(66.1)	144(80.0)
Total	180(100.0)	180(10.00)	180(100.0)	180(100.0)	180(100.0)

Several UMLS concept codes were identified for each selected indicator with a separate set of MedLee results for each concept code (Table 31). The data from the multiple concept categories were collapsed to make a comparison against the selected indicator results. For example, the UMLS concept code for rhinitis (C0035455) was represented by a second set of concept results for runny nose (C1260880). MedLee results for rhinitis were very similar to the PERL results, but very low in comparison to the manual chart review findings. Results for rhinitis were absent in 1(.6%) of the records, present in 4(2.2%) and not mentioned in 175(97.2%) of the records. When runny nose, the second UMLS concept code (C1260880) was used the MedLee findings were almost identical to that of the medical chart review; absent 13(7.2) vs. 12(6.7%), present 19(10.6%) vs. 20(11.1%) and 148(82.2%) vs. 148(82.2%), respectively. Thus, the concept "runny nose" was more sensitive for a diagnosis of asthma than the word rhinitis. Combined results for the two terms were almost identical to the manual chart review findings. The term past medical history for asthma was another indicator, which was represented by 2 UMLS concept codes. The concept code asthma (C0004096) was found to have higher levels of agreement against the findings observed from the manual review than the concept code for asthma history (C0455544). Furthermore, the concept code for asthma (C0455544) performed better at identifying asthma cases then reported by PERL. Combined results for the two concept code results were almost an identical match to the manual chart review findings. The selected indicator cough was reported using one concept code (C0010200) to represent all combined total cough certainties. Only one other UMLS concept code (C085149) was identified to represent productive and dry cough. These results are reflected in the all cough category* as highlighted in Table 31.

In summary, the goal was to determine which automated methodology, PERL or MedLee performed better at identify the selected indicators from the manual chart review. The degree of concordance was calculated to determine this estimate using the following methods. The manual chart review results for one indicator were matched against the findings from the automated method for the same indicator. The sum of all instances where agreement was found between both methodologies for the negative (absent), positive (present) and not mentioned status matches for a particular indicator were divided by the total number of reviews. Data reported in Table 32 were used to calculate a degree of concordance between the manual chart review and PERL for the indicator wheeze. The results of the manual chart review suggest that PERL agreed on 3 absent, 64 present and 12 not mentioned instances. This is a total of 79 agreed upon matches, which divided by the total of 180 reviews yields a degree of concordance estimate of 43.8%. This methodology was applied for each indicator for both PERL and MedLee. The results for the degree of concordance estimates are listed in Table 33 .

Table 32: Methodology used to establish degree of concordance

Manual Chart Review	PERL: Wheeze			Total
	Absent	Present	Not mentioned	
Absent	3	11	2	16
Present	5	64	35	104
Not mentioned	6	42	12	60
Total	14	117	49	N=180

Table 33: Degree of Concordance: Manual chart review vs. automation

Manual	Automation	
Medical Chart Review	**PERL**	**MedLEE**
Wheezing	43.8%.	52.7%
Cough	45.0%	62.2%
Shortness of Breath	56.1%	56.1%
Rhinitis*	80.0%	75.5%
Fever	52.0%	58.3%
History of asthma*	41.6%	50.0%
Asthma Medications	64.4%	** can not be determined

*Combined UMLS concept codes
**Complete data not available to make comparison

The degree of concordance was greater for MedLee in all categories with the exception of rhinitis. Overall, the performance of MedLee was more superior to PERL for validating data extracted from the EMR. Several limitations were identified with PERL and not corrected with multiple attempts to resolve the issues. These limitations reduced the level of sensitivity in findings in this research study. PERL was not able to distinguish between a past, present or future event. Therefore, if a diagnosis of asthma were suggested as a possibility, PERL would not be able to account for ruling out this likelihood. A double negative such as cannot rule out asthma further reduced the differential. The inability of PERL to section out portions of the electronic medical record also increased the likelihood that a history of asthma reported for a family member may also have been included as a positive finding in this asthma review. Since PERL was used to match on all regular search expressions for all possible categories, there was no way to adjust the script to account for misspelled words, plurals or for word variations. PERL

was not able to account for these differences. This potentially reduced the sensitivity of findings in question. Furthermore, the PERL expression did not include the acronym SOB for shortness of breath. This was an oversight that could be corrected in future search applications. The positive feature of PERL was the ability to match on highly specific words such as those represented for each asthma medication. The complex nature of the lexicon used to represent asthma in the EMR abstract made it difficult to create a script that could be applied to this study with a high degree of sensitivity. Moreover, based on the low levels in degree of concordance reported between the gold standard manual chart review and PERL with the except of the medication indicator (66.4%), PERL would not be recommended as a tool for validating asthma data extracted through the electronic medical record. Additional techniques would have to be applied to handle the various challenges for the disease specific terms needed to validate this diagnosis.

The high level of agreement observed between MedLee and the manual chart review suggests that MedLee could play a valuable role in asthma research. This tool did as well as a human for extracting and interpreting data from the EMR. A high degree of accuracy was observed for identifying both the presence and absence of key concept terms. The unique ability of MedLee to parse out text from the EMR through the mapping of the UMLS concept codes and modifiers, including the manual creation of a set of textual rules improved the ability to handle ambiguities. However, for terms where no decision could be made or contradictory statements were found for the absence or presence of the selected indicator term, results were placed in the unknown or not mentioned category. Based on the comparative results between the two methodologies, this limitation was not reflected in the findings. The ability of MedLee to account for past events was evident by the high level of agreement reported for past medical

history of asthma. Furthermore, MedLee did very well at identifying "acuteness" of illness. This is evident by the high level of concordance observed for the symptoms of asthma, which include shortness of breath, wheezing and cough. With each desired finding selected for the asthma Validation Study, MedLee performed very well at parsing out the information from the EMR. One challenge with using MedLee was identified at the end of the study when the reviewer was attempting to quantify the results. Three of the selected indicators had 2 or more concepts codes, which resulted in an overlap in findings. No method was described in the literature to handle this issue. A decision was made to collapse the data into one category for each selected indicator. This step may have resulted in an imprecise measure for rhinitis, past medical history of asthma and the proxy measure used to report on asthma medications. However, based on the degree of concordance observed against the manual chart review results this was not the case. Another issue found with MedLee was the inability to isolate asthma medications in the EMR. For this research study, a proxy measure could be reported for asthma medications using MedLee since UMLS concept codes have been designed for the word inhaler and nebulizer. The lack of the UMLS concepts codes for the specific asthma medications limited the ability to have meaningful results reported for asthma medication usage in this cohort, However, this issue has been brought to the attention of the developers of MedLee by the programmer. Thus, the implication of this finding has assisted in advancing practice in natural language research using MedLee. In conclusion, the process of a manual chart review is time consuming and costly. The use of MedLee over both PERL and the manual chart review substantially improved the efficacy and efficiency for conducting this type of research.

3.7 CLASSIFICATION OF ASTHMA POPULATION: CSTE CLINICAL AND LABORATORY CASE CLASSIFICATION DEFINITION RESULTS

Data was classified according to the modified version of the CSTE Clinical and Laboratory case classification definition (Table 12). The criterion was modified due to the limited amount of laboratory data found documented in the medical record. A summary of the classification results are highlighted in Table 34. The classification results are based on a review of the electronic medical record. A descriptive analysis was conducted to compare the demographic subgroups, and selected indicator characteristics for each asthma classification category.

3.7.1 Confirmed Classification

A confirmed case of asthma was based on individuals meeting the clinical symptoms identified in the CSTE case definition at least 3 times in the past year. A total of 14(7.7%) individuals met the criteria to be classified as a confirmed case of asthma. Wheezing lasting 2 consecutive days or more was the most frequent symptom reported in 7(50%) of the cases. Chronic cough treated with bronchodilators in the absence of an allergic sinusitis was reported in 4(28.5%) of the asthma cases. No individuals had nocturnal awakening with dyspnea on exertion, cough and or wheezing in the absence of another medical conditions known to cause the symptoms. Three cases were classified into the confirmed asthma category because documentation was found on three or more ED visits in the past year for an asthma attack, including an asthma history and verification of being on an asthma medication for treatment of the asthma attack symptoms.

Although this criterion was not considered part of the case definition for this classification category sufficient evidence was available to support a confirmed asthma diagnosis.

Table 34: CSTE Clinical and Laboratory case definition classification results (n=180)

Classification	Criteria	Frequency %
Confirmed (n=14)	Met any clinical symptom/event at least 3x during the past year: • Wheezing lasting 2 consecutive days or more • Chronic cough responds to bronchdilatation that persists 3-6 weeks in the absence of allergic rhinitis sinusitis • Nocturnal awakening with dyspnea on exertion, cough, and or wheezing in the absence of other medical conditions known to cause these symptoms • Asthma attack supported by documentation of a history of asthma and use of asthma medications*	7(50.0%) 4(28.5%) 0 3(21.4%) **Total =14(7.7%)**
Probable (n=126)	Presence of any symptom reversed with a physician treatment or asthma medication 3 times in the past year. **	
	Taken medications in the past year that was prescribed by a physician for asthma.	126(100.0%) **Total=126(70.0%)**
Possible (n=27)	Presence of any of the following symptoms during the past year: • Shortness of breath • Wheezing or chronic cough in the absence of obvious respiratory infection • Presence of nasal secretions, mucosal swelling, nasal polyps or chronic sinusitis • Hyper-expansion of the thorax • Prolonged phased to forced exhalation • Wheezing during normal breath • Chest X-ray noted to show hyper expansion • Supporting documentation (current or past asthma history)*	20(74.0%) 8(29.6%) 4(14.8%)0 0 0 11(40.7%) 1(3.7%) 5(18.5%) **Total=27(15.0%)**
Not Classified (n=13)	Not criteria to substantiate an asthma diagnosis	**Total=13(7.2%)**

**documentation to support confirmed asthma*

***Same asthma cases identified in confirmed category (CSTE recommends collapsing probable and confirmed due to overlap in criteria)*

3.7.2 Probable Classification

A probable case of asthma case was classified using the following criteria (a) presence of any symptom reversed with a physician treatment or asthma medication three times in the past year, (b) and taken medications in the past year that were prescribed by a physician for asthma. In order to be eligible for criteria (a) under the probable classification category individuals had to have three visits to the ED in the past year where their asthma symptom was treated by a physician or with an asthma medication. This step resulted in identifying the same 14 patients classified under the confirmed asthma category. Since the Council for State and Territorial Epidemiologists recommends combining both the confirmed and probable asthma categories due to similarities in the criteria, these 14 asthma patients were kept in the confirmed asthma category for the initial portion of the classification process. The two classification categories were later collapsed for further analysis.

A probable case of asthma was the easiest to validate by confirming that the individual was taking medications in the past year that were prescribed by a physician for asthma. The list of asthma medications previously reported in the methods section of this research was used to differentiate the presence of an asthma medication (Table 10).

126 (100%) persons met the eligibility criteria for probable asthma (Table 34). One person who presented to the emergency room for asthmatic bronchitis was non-complaint with using an inhaler but had previously been prescribed the inhaler in the past for asthma. This person was included in the probable asthma category.

3.7.3 Possible Classification

A total of 27(15.0%) of all cases were classified into the possible asthma category based on the criteria highlighted in Table 34. The criteria in this category were not mutually exclusive. Thus, the findings reflect whether each criterion was present for the ED visit. Shortness of breath on exertion or the notation of shortness of breath without exertion was the most common symptom reported 20(74.0%), followed by cough and wheezing in absence of an obvious respiratory infection 8(29.6%). The presence of rhinitis or increased nasal secretions, mucosal swelling, nasal polyps and or chronic sinusitis was reported in 4(14.8%) possible asthma cases. The presence of wheezing was reported in 11(40.7%) of the cases. No documentation was found to support the following criteria: hyper expansion of the thorax and prolonged phase for forced exhalation. Documentation was found in one record to suggest that a patient had hyperinflation of the lungs as identified on the chest X-ray. A total of 5(18.5%) records did not meet any of the criteria highlighted under this classification category. However, a review of supporting documents including previous 1-year ED visits and documentation of a past medical history of asthma, and or mention of asthma in the current ED note provided enough evidence to substantiate a possible asthma diagnosis.

3.7.4 Not Classified

A total of 13 cases that had been previously marked as asthma by the ICD 9 coded physician diagnose, had no supporting documentation in the EMR to corroborate this diagnosis. The total number of documents reviewed to substantiate the asthma diagnosis is highlighted in Table 35. A list of the presenting chief complaints found in the write up of the ED visit

summary notes were compiled for comparison. No records in the not classified group were found to have any of the signs and symptoms or asthma characteristics represented in the set of Selected Indicators. A non-productive cough was found in one case. A total of 13(7.2%) of the EMR were marked as not classified.

Table 35: ICD-9 coded physician diagnosis of asthma: Unable to classify

All ED visits with a primary asthma diagnosis [ICD 9 Code 493.0-493.9] (n=13) **Chief complaint on admission**	**Number of records reviewed to substantiate an asthma diagnosis**
Diarrhea	1
Diabetes-swelling back of neck	8
Five-day shoulder pain	1
Severe chest tightness	1
Bilateral temporal headache	1
Regulate anticoagulants	11
Right sided pain and weakness, status post stroke, hypokalemia	1
Right sided abdominal pain- Appendicitis	1
Muscle ache, joint pain, fibromyalgia	2
Back pain	6
Depression	1
Breast Biopsy	13
Tightness in chest-congestion	2

A comparison of the demographic subgroups by age, sex, and race were reviewed for each person represented in the CSTE clinical and laboratory asthma classification study (Table 36). This descriptive analysis was done solely to determine if any differences could be observed by classification category. Individuals classified as possible asthma were older in comparison to the other 3 classification groups (M =54.74, SD 20.63). More females were observed than males consistently across each classification category. Racial differences were also noted across classification categories with Whites having more visits than Blacks and Others. The higher proportion of Whites in the population represents the increased numbers of Whites in the

catchment area. These data reflect only 6(60%) hospitals out of the UPMC network and therefore can make no assumptions about the demographic breakdown as these data clearly reflect only a small portion of asthma visit.

Table 36: CSTE clinical and laboratory classification status by age, sex and race

Confirmed (n=14)			Probable (n=126)			Possible (n=27)			Not Classified (n=13)		
Age	**n**	**%**	**Age**	**n**	**%**	**Age**	**n**	**%**	**Age**	**n**	**%**
0-17	0		0-17	0		0-17	0		0-17	0	
18-29	4	28.5	18-29	32	25.3	18-29	3	11.1	18-29	2	15.4
30-44	4	28.5	30-44	34	26.9	30-44	6	22.2	30-44	5	38.4
45-64	4	28.5	45-64	34	26.9	45-64	8	29.6	45-64	4	30.7
65+	2	14.2	65+	26	20.6	65+	10	37.0	65+	2	15.3
Mean 43.0(SD 17.03)			Mean 45.61(SD 20.35)			Mean 54.74(SD 20.63)			Mean 45.6(17.65)		
Range 18-77			Range 18-103			Range 18-95			Range 18-86		
Total	**180**	**100.0**	**Total**	**180**	**100.0**	**Total**	**180**	**100.0**	**Total**	**180**	**100.0**
Sex	**n**	**%**	**Sex**	**n**	**%**	**Sex**	**n**	**%**	**Sex**	**n**	**%**
Male	6	42.8	Male	33	26.2	Male	11	40.7	Male	4	30.8
Female	8	57.1	Female	93	73.8	Female	16	59.3	Female	9	69.2
Total	**180**	**100.0**	**Total**	**180**	**100.0**	**Total**	**180**	**100.0**	**Total**	**180**	**100.0**
Race	**n**	**%**	**Race**	**n**	**%**	**Race**	**n**	**%**	**Race**	**n**	**%**
Black	5	35.7	Black	48	38.1	Black	10	37.0	Black	5	38.5
White	9	64.2	White	77	61.1	White	17	63.0	White	8	61.5
Other	0		Other	1	.8	Other	0		Other	0	
Unknown	0		Unknown	0		Unknown	0		Unknown	0	
Total	**180**	**100.0**	**Total**	**180**	**100.0**	**Total**	**180**	**100.0**	**Total**	**180**	**100.0**

The results from the selected indicators obtained from the gold standard medical chart review were compared to those observed in the CSTE clinical and laboratory classification categories. This step was conducted to determine how well the selected indicators were represented in individuals classified with asthma as defined by the CSTE clinical and laboratory case definition. Because neither methodology, the selected indicators nor the CSTE clinical and laboratory classification definition are considered to be the gold standard for conducting asthma research, a descriptive comparison of the results can only be made. The confirmed and probable asthma categories were collapsed into a probable asthma grouping as recommended by the Council for State and Territorial Epidemiologist (1998). The comparison was based on 2 classification statuses, probable and possible asthma (Table 37).

Table 37: Comparison of Selected Indicators to CSTE Clinical and Laboratory Case Classification Definition

Selected Indicators	CSTE Clinical and Laboratory Case Classification Definition			Manual Chart Review
Indicators	**Probable n=140**	**Possible n=27**	**Total asthma n=167**	**Total N=180**
Wheezing				
0 absence	11(7.8)	3(11.1)	14(8.3)	16(8.8)
1 presence	93(66.4)	11(40.7)	104(62.2)	104(57.7)
2 not mentioned	36(25.7)	13(48.1)	49(29.3)	60(33.3)
Total	140(100.0)	27(100.0)	167(100.0)	180(100.0)
Cough				
0 absence	24(17.1)	3(11.1)	27(16.1)	27(15.0)
1 presence	82(58.5)	19(70.4)	101(60.4)	102(56.7)
2 not mentioned	34(18.8)	5(18.5)	39(23.3)	51(28.3)
Total	140(100.0)	27(100.0)	167(100.0)	180(100.0)
Shortness of Breath				
0 absence	7(5.0)	2(7.4)	9(5.3)	15(8.3)
1 presence	108(77.1)	20(74.1)	128(76.6)	128(71.1)
2 not mentioned	25(17.8)	5(18.5)	30(17.9)	37(20.5)
Total	140(100.0)	27(100.0)	167(100.0)	180(100.0)
Rhinitis				
0 absence	9(6.4)	3(11.1)	12(7.1)	12(6.7)
1 presence	16(12.7)	4(14.8)	20(11.9)	20(11.1)
2 not mentioned	115(63.8)	20(74.1)	33(19.7)	148(82.2)
Total	140(100.0)	27(100.0)	167(100.0)	180(100.0)
Fever				
0 absence	86(61.4)	14(51.9)	100(59.8)	108(60.0)
1 presence	10(7.1)	9(33.3)	19(11.3)	19(10.6)
2 not mentioned	44 (31.4)	4(14.8)	48(28.7)	53(29.4)
Total	140(100.0)	27(100.0)	167(100.0)	180(100.0)

(Table 37 Continued)

History of asthma				
0 absence	0	2(7.4)	2(1.1)	2(1.1)
1 presence	115(82.1)	5(18.5)	120(71.8)	120(66.7)
2 not mentioned	25(17.8)	20(74.1)	45(26.9)	58(32.2)
Total	140(100.0)	27(100.0)	167(100.0)	180(100.0)
Asthma Medications				
0 absence	0	11(40.7)	11(6.5)	11 (6.1)
1 presence	140(100)	0	140(83.3)	140(77.7)
2 not mentioned	0	16(59.3)	16(9.5)	29(16.1)
Total	140(100)	27(100.0)	167(100.0)	180(100.0)

A total of 167(92.7%) cases of asthma were confirmed through a review of the EMR. One hundred and forty of these cases 140(77.7%) were probable asthma, and 27(15.0%) possible. The highest frequency of observed values for a probable asthma diagnosis matched the selected indicator for presence of shortness of breath in 108(77.1%) cases, followed by wheeze 93(66.4%) and cough 12(58.5%). A possible case of asthma matched on the presence of shortness of breath in 20(74.0%) of the cases followed by cough, 19(70.4%) and wheeze, 11(40.7%). Rhinitis was more frequently observed in the probable vs. possible asthma cases 16(12.7%) vs. 4(14.8%), respectively. Fever, a symptom always assessed on presentation to the ED was present in 10(7.1%) of the probable asthma cases and not mentioned in 44(31.4%). For the possible asthma category, fever was present in 9(33.0%) of the cases and not mentioned in 4(14.8%). The fact that fever was not mentioned in 44% of the probable asthma cases is most likely due to an oversight on medical chart review. The presence of a history of asthma was found in 115(82.1%) of the probable asthma cases, and only 5(18.5%) for the possible. An asthma history was not mentioned in 25(17.8%) of the probable asthma cases and 20(74.1%) of the possible. The absence of documentation to support a past medical history for asthma in 74.1% of the possible asthma cases made it difficult to validate the diagnosis. A case of asthma could not even be validated through the use of an asthma medication since none of these cases were on an asthma medication. The only criteria that could be used to classify a possible case of asthma were through the clinical signs or symptoms. In contrast, the probable asthma cases which had no documentation to support a past medical history 25(17.8%) of asthma were all on an asthma medications. The presence of being on an asthma medication was 100% predictive in the probable category and 0% in the possible.

Overall, the presence of being on an asthma medication was the most sensitive indicator for asthma 140(83.3.0%). The presence of the shortness of breath was the most prevalent clinical indicator observed 128(76.6.1%), followed by wheezing 104(62.2%) and cough 101(60.4%). The symptoms of rhinitis and fever were not highly sensitive for an asthma diagnosis. The presence of rhinitis and fever were found in less than 12% of all asthma cases. A past medical history of asthma was found in 120(71.8%) of all asthma cases.

3.8 VALIDATION STUDY DISCUSSION

The use of the CSTE Clinical and Laboratory case classification definition against the physician diagnosis of asthma on medical chart review provided similar estimates of asthma. Thirteen (7.2%) out of the 180 primary asthma cases could not be validated against the physician diagnosis. The CSTE case definition has a high level of sensitivity for identifying asthma that accounts for a wide range of asthma cases from mild to severe. Although each clinical and laboratory criteria are important indicators to substantiate an asthma diagnosis the lack of specificity in the criteria made it difficult to classify the asthma cases. Individuals in the confirmed asthma category are likely being treated by a physician and on an asthma medication to treat the asthma signs and symptoms. However, this criterion was not used to classify patients into the confirmed category, but instead the probable category. Moreover, the similarities in the criteria illustrated in the confirmed and probable classification categories introduce a misclassification bias for estimating asthma cases in one year by relying on asthma events from the previous year (Dombkowski, Wasilevich & Lyon-Callo, 2005). The CSTE Asthma Surveillance workgroup recommends combining both categories to enumerate asthma cases for

surveillance (CSTE, 1998). This step was done to make comparison with the selected indicators found on medical chart review and to determine the presence of the indicators within each asthma classification category. The criteria used to classify individuals within the possible asthma category were not mutually exclusive. Thus, individuals could have met one of several criteria. The symptoms described in this category such as shortness of breath is not specific, and fit within the range of signs and symptoms that characterize other disorders. Therefore, a review of all sections of the electronic medical record was needed to eliminate any misclassification of diagnoses. Individuals diagnosed with asthma may have persistent or intermittent exacerbation of disease. Even when individuals are symptom free, airway passages are still compromised (Marks, 2005). Therefore, depending on when along the course of illness a patient is clinically evaluated or enrolled as a study participant can impact whether a patient is considered for a diagnosis of asthma. Thus, a cross sectional study may not fully capture a cohort entirely representative of the asthma population. This is unlikely the case for the 13 individuals in this study whose asthma diagnosis could not be substantiated on medical chart review even though all records from the 1-year time period prior to the initial ED visit were assessed. The likely explanation in this instance was related to upcoding for financial reimbursement and not related to medical record search time error.

In conclusion, Vollmer and associates (2004) report that a combination of searching multiple databases and length of time searching increases the likelihood of positively identifying a case of asthma. A benefit to this study was the ability to have access to the medical record in an electronic format and multiple records for the 1-year study period prior to the initial ED visit. The automated feature provided a quick and easy solution to record selection, reduced the search and review time in half, and proved to be an efficient methodology for conducting this type of

research. The CSTE clinical and laboratory case classification definition lacks specificity and is very time consuming to validate using a medical chart review and through the use of the EMR. However, due to the overlap bias of symptoms for asthma, and lack of specificity in the CSTE criteria the selected indicator findings from the medical chart review were helpful to classify the asthma cases. After removing the 13 not classified asthma cases from the cohort, the presence of the selected indicators in the asthma classifications categories were not significantly different than what was reported on manual chart review. The use of selected indicators could be considered another adjunct tool to assist with classifying asthma cases if rhinitis and fever were removed.

A more comprehensive approach is needed to conduct asthma research using multiple methodologies or developing a gold standard that is statistically valid for this diagnosis. Neither the selected indicators nor the CSTE clinical and laboratory case classification definition are approved gold standard methodologies for diagnosing asthma.

4.0 CHAPTER 4- CASE CROSSOVER STUDY

4.1 POPULATION D

A total of 6,979 individuals were used in this case crossover analysis. This represented 10,183 asthma ED visits for individuals presenting to the ED for any primary case of asthma from 2002-2005. A total of 77.0%(5,431) of the any primary asthma population had one visit, and 22.2%(1,548) had multiple visits. A breakdown of the demographics for this population is highlighted in Table 38.

The age range for these individuals was from 0-103, (M=39.25, SD=20.82). Individuals 65 ± years were represented by the smallest proportion of individuals in this cohort 13.5%(948), followed by individuals 0-17 years of age who made up 14.3%(1004) of the population. These two groups have been reported to be the most compromised groups of individuals from air pollution exposure (ALA, 2008).

Gender differences were observed with females having a higher frequency of ED visits than males (65.9% vs. 34.1%), respectively. The higher female- to- male ratio observed in this cohort may due to the biological differences in gender or reflect a higher proportion of females in the catchment area represented by the 6 respective hospitals used in this analysis.

For the any primary asthma ED visit population racial differences were observed with Whites (60.1%) having the highest number of ED visits than Blacks (37.9%) and Others (2.0%).

The higher proportion of Whites is of no surprise considering this sample reflects only 6 hospitals from the UPMC network, which also reflects a higher proportion of Whites in the catchment area. The proportion of Blacks is high when compared to AC as a whole, which is made up by only 10% of the Black population and 30% in the COP.

Table 38: Individuals with any primary asthma emergency room visit (6 UPMC hospitals) by age, sex, and race

Any Primary asthma diagnosis for all individuals **(n= 6,979)**			
Age	**n**	**%**	**CI**
0-17	1,004	14.3	13.57-15.21
18-29	1,523	21.8	20.85-22.79
30-44	1,815	26.0	24.98-27.04
45-64	1,687	24.1	23.17-25.17
65+	948	13.5	12.7-14.38
Range 0-103			
Mean 39.25(SD 20.82)			
Total	**6,979**	**100.0**	
Sex	**n**	**%**	**CI**
Male	2,380	34.1	32.99-35.21
Female	4,599	65.9	58.99-61.29
Total	**6,979**	**100.0**	
Race	**n**	**%**	**CI**
Black	2,645	37.9	36.76-39.04
White	4,197	60.1	58.99-61.29
Other	137	2.0	1.63-2.29
Total	**6,979**	**100.0**	

The annual daily average of any primary asthma visit from 2002-2005 were from 6.06-7.83 per day. These data are not de-duplicated since they represent recurrent visits. The highest numbers of visits were observed for Shadyside 2,192(21.5%) and McKeesport Hospital's 2,003(19.6%). Southside had the least amount of visits 1,122(9.9%). The low number of visits

observed at Southside Hospital 1,055(10.3%) could be due to an individual preference for hospital selection or a lower referral pattern in physicians treating asthma at this hospital (Table 39).

Table 39: Any primary asthma ED visit by hospital

Hospital	2002	2003	2004	2005	Total %
Braddock	413	534	454	494	1,895(18.6)
McKeesport	485	556	453	509	2,003(19.6)
Presbytarian	377	380	313	366	1,436(14.1)
Shadyside	466	583	546	597	2,192(21.5)
St Margaret's	374	433	387	397	1,591(15.6)
Shadyside	95	372	317	271	1,055(10.3)
Unknown	1	2	1	2	6(1.0)
Total * %	**2,212 (21.7)**	**2,861(28.0)**	**2,471(24.2)**	**2,639(25.9)**	**10,138(100.0)**

*6 UPMC hospitals

The seasonal distribution for all visits was observed to be higher in September and October with a second peak beginning December through March. These findings are consistent with the literature, which suggests that the highest number of asthma visits coincided with the fall, and winter season where asthma morbidity has been shown to be high due to rag weed allergies and rhinovirus infections (Johnston & Sears, 2006; Dales, et al., 1996, Wansoo & Schneider, 2005). The data was analyzed to determine whether there were any statistically significant differences between the number of visits in the fall and winter months vs. spring and summer. A higher number of visits were found in the months with cooler temperatures. This difference was found to be statistically significant, $X^2(1, N = 10{,}183) = 37.5$, $p = < 0.05$. The sharp rise in 2003 for the month of December could be a result of the influenza season and its affect on asthma ED visits (Figure 20).

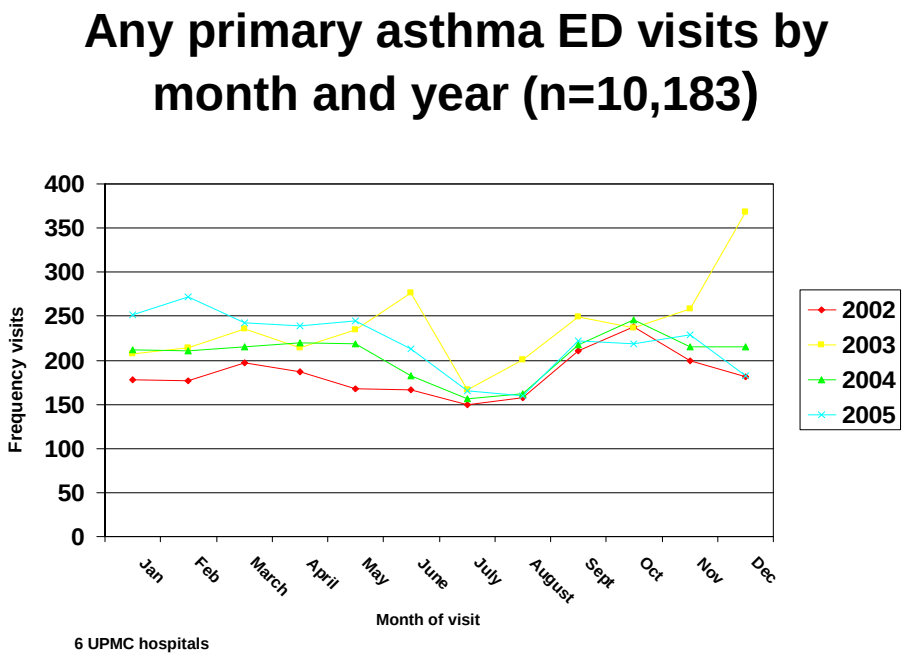


Figure 20: Any primary asthma ED visit by month and year

The geographical distribution of the population with asthma is highlighted in Figure 21. The highest numbers of individuals with asthma were concentrated in and around the COP. This is of no surprise since the 6 hospitals in this sample were all close to the inner city.

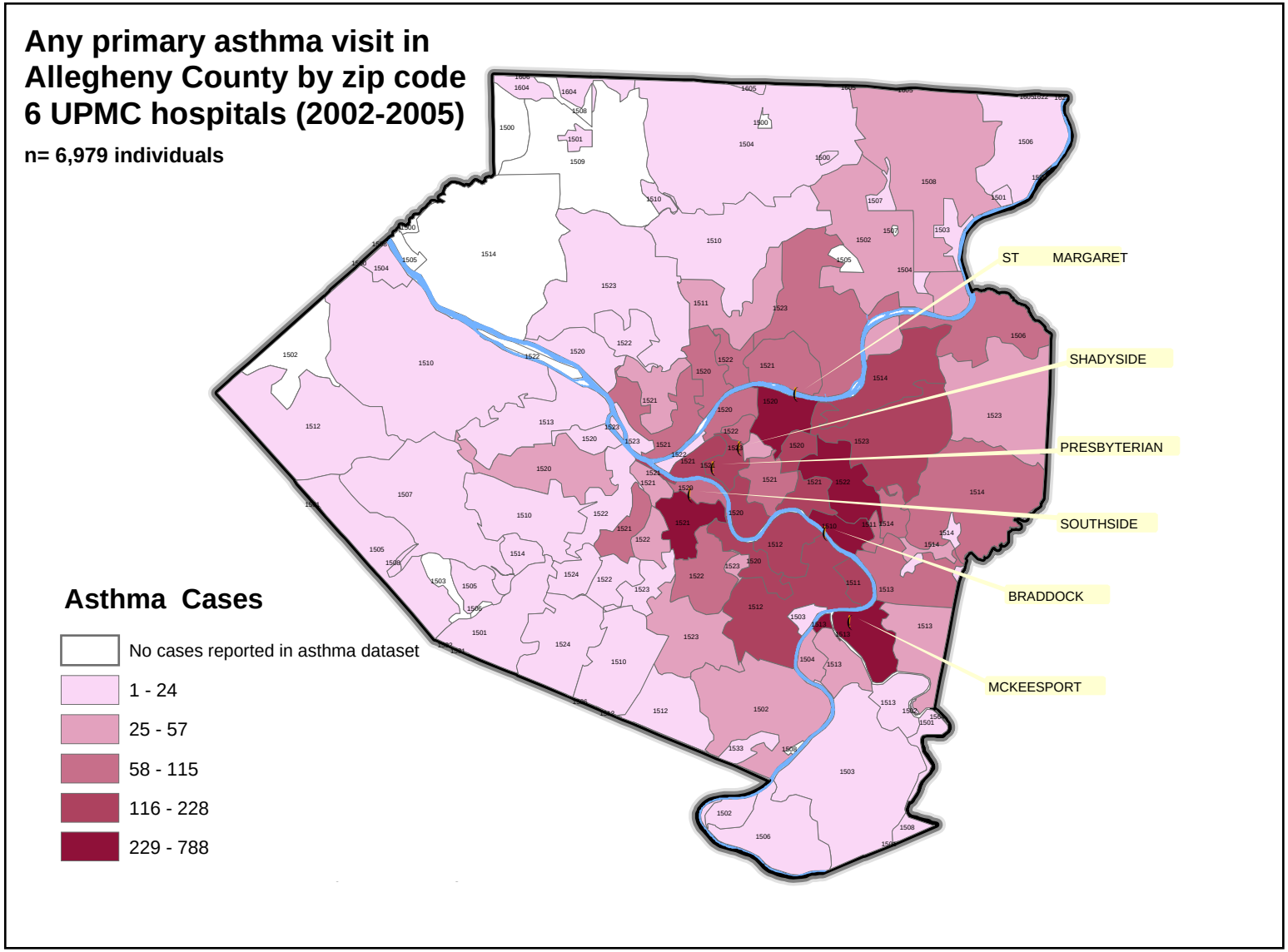


Figure 21: Any primary asthma visit for <u>6 UPMC hospitals</u> by zip code in Allegheny County 2002-2005

A case crossover analysis was conducted using ozone and $PM_{2.5}$. The summary statistics for the distribution of daily concentrations of the air pollutants and temperature for all months are reported in Table 40. A distinction was made by first reporting on each pollutant and temperature for all months, and the second set of summary statistics representing the summer months only (Table 41). The effects of ozone has been reported to be greatest in months with warmer temperatures (Villenueve, et al., 2007; Barnett et al., 2005; Ito et al., 2005; Bell, et al., 2007; Burnett et al., 2001; Bell et al., 2005; Paulu & Smith, 2008).

Table 40: Distribution of daily concentration of ozone, PM2.5, and temperature for all months (2002-2005)

Pollutant	Measure	Mean	Minimum	10th	25th	50th	75th	90th	Maximum
Ozone(ppb)	24 hour maximum	40.6	.00	16.0	25.0	37.0	55.0	68.0	130.0
$PM_{2.5}$	24 hour average	13.1	2.0	4.6	7.1	11.4	16.8	24.4	55.1
Temperature	24 hour average	51.6	6.5	26.2	36.9	54.7	67.8	69.9	83.1

The National Ambient Air Quality Standards (NAAQS) for ozone are based on the 1-hour and 8-hour standards. The 24-hour maximum ozone concentration level (ppb) was calculated for use in this study and agreed upon by the review committee for this analysis. The boxplots and histograms for all pollutants and temperature are reported in Appendix C.

The average 24-hour maximum ozone concentration level was 40.6-ppb with a range of .00-ppb to 130-ppb (Table 40). The spread of distribution of the 10th to the 90th percentile was 16.0-ppb to 68-ppb. Fifty percent of the 24-hour maximum ozone exposure levels were below the mean at 37-ppb.

The national standard for the 24-hour average $PM_{2.5}$ level is set at 65.0ug/m^3. The mean $PM_{2.5}$ concentration level in this dataset was 13.1ug/m^3, with a range of 2.0ug/m^3 to 55.1ug/m^3.

These levels are well below the standard. The average temperature for all months was 51.6° with a range of 6.5° to 83.1°.

Table 41: Table 41: Distribution of daily concentration of ozone, PM2.5, and temperature for summer months (2002-2005)

Pollutant	Measure	Mean	Minimum	10th	25th	50th	75th	90th	Maximum
Ozone(ppb)	24 hour maximum	55.5	16.0	34.0	43.0	55.0	66.0	78.0	130.0
$PM_{2.5}$	24 hour average	16.1	2.0	6.1	9.2	14.4	21.2	31.2	55.1
Temperature	24 hour average	65.4	32.2*	50.0	60.0	67.7	72.6	76.2	83.17

Summer months (April*- September)

The concentration levels for the summer months are reported in Table 41. Contrasting to all months, the 24-hour maximum ozone concentration in the summer months was higher at 55.5-ppb with a range of 16-ppb to 130-ppb. The spread of distribution from the 10th to the 90th percentile was 34-ppb to 78-ppb. Fifty percent of the exposure levels were equal to the mean at 55-ppb.

The 24-hour average $PM_{2.5}$ level for the summer months was 16.1ug/m^3, with a range of 2.0ug/m^3 to 55.1 ug/m^3. The distribution range for $PM_{2.5}$ concentration levels in the summer months were still below the national standard. Temperature levels were higher in the summer months with an average of 65.4°, range 32.2° to 83.1°. The lowest reported temperature of 32.2° was observed in April. The summary statistics for the summer months are also reported in Appendix C.

Data are reported for Pearson's correlation coefficients for air pollutants, temperature and humidity (Table 42). Both ozone and $PM_{2.5}$ were strongly correlated with temperature ($r = .72$), ($r = .529$), ($p < .01$), respectively. However, the correlation was greatest for ozone. A positive

correlation was also observed for PM and ozone (r = .573), (p < .01). Humidity was positively correlated with $PM_{2.5}$ (r = .08), (p < .01). This same relationship was not observed for ozone and humidity.

Table 42: Correlation coefficients for ozone and PM2.5 and weather variables

	$PM_{2.5}$	Temperature	Humidity	Ozone
$PM_{2.5}$	1	.529**	.084**	.573**
Temperature	.529**	1	.034**	.722**
Humidity	.084**	.034	1	-.042
Ozone	.573**	.722**	-.042	1

Correlation is significant at the p = < 0.01,(2-tailed)

A comparison of the different lag distributions were used in order to explore the relationship between different exposure levels and asthma events. The first step included the use of a single lag model for ozone and PM. No effect was observed for ozone alone 1.01(95% CI .999-1.03), or PM alone 1.03(95% CI .00-1.06) (Table 43). Unadjusted single lag model estimates for the 24-hour maximum ozone for 2-day lag and 5-day average lag were found to be significant (p = <.05); 2-day (1.03(95% CI 1.01-1.05)), and 5-day average (1.04(95% CI 1.01-1.07)), respectively. After adjusting for temperature, the effect estimate was only significant for the 2-day lag (1.02(95% CI 1.01-1.04), (p = < .05). This estimate represents a 2% increase in asthma admissions for every 10-ppb increase in 24-hour maximum ozone concentration level. No effect was observed for $PM_{2.5}$ after adjusting for temperature, 1.00(95% CI .997-1.04).

The combined effect for ozone and $PM_{2.5}$ were tested in a multiple lag model. The risk estimate was 1.01(95% CI .999-1.03). Adding temperature to the model did not result in any significant risk, OR= 1.00.

Table 43: Single and multiple lag models for 24-hour maximum ozone, PM2.5 and temperature: All seasons

Maximum Likelihood Estimates (Singe and Multiple lag Models) for 24 our Maximum Ozone, temperature and PM $_{2.5}$ All Months			
Single Models		**Multiple Models**	
	Odds Ratio (95%)		**Odds Ratio (95%)**
1. Ozone Alone (day 0)	1.01 (.999 – 1.03)	1. Ozone & $PM_{2.5}$	1.01 (.999 – 1.03)
2. PM $_{2.5}$ Alone (day 0)	1.03 (1.00 – 1.06)	2. Ozone & Temperature	1.00 (.998 – 1.02)
3. Temperature	1.05 (1.03 – 1.08)	3. PM $_{2.5}$ & Temperature	1.00 (.997 – 1.04)
		4. Ozone, Temperature & $PM_{2.5}$	1.00 (.997 – 1.04)
Unadjusted (ozone only)		**Adjusted for Temperature**	
Lags as separate model (Ozone only)		Lags adjusted for temperature	
1. Day One	1.02 (1.00 – 1.04)	Day One	1.02 (1.00 – 1.04)
2. Day Two	**1.03 (1.01 – 1.05)***	**Day Two**	**1.02 (1.01 – 1.04)***
3. Day Three	1.01 (.999 – 1.03)	Day Three	1.01 (.999 – 1.03)
4. Day Four	1.00 (.998 – 1.02)	Day Four	1.00 (.998 – 1.02)
5. Day Five	1.00 (.998 – 1.02)	Day Five	1.00 (.998 – 1.01)
5-Day Average	**1.04 (1.01 – 1.07)***	5-Day Average	1.02 (.999 – 1.05)

* indicates significance p = <0.05

The second step was to evaluate ozone levels during the summer months to determine if any association could be observed for asthma ED visits during months with warmer temperatures. Regardless of whether ozone was used in a single model, or in combination with $PM_{2.5,}$ no effect was observed, ozone alone 1.01 (95% CI .999 – 1.03) and ozone with $PM_{2.5}$ .999(95% CI .997-1.02), respectively (Table 44). $PM_{2.5}$ alone during the summer months was not significant 1.04(95% CI 1.00 – 1.08). Furthermore, single lag estimates for day-1 through 5- and the 5-day average lag for 24-hour maximum ozone concentration level was not significant during this time period. All confidence limits included one. The no ozone effect observed in

this analysis for all single and multiple models was believed to be a sample size issue since the dataset was reduced by half when conducting this portion of the analysis.

Table 44: Single and multiple lag models for 24-hour maximum ozone, $PM_{2.5}$: Summer months (April-September)

Models	Odds Ratio (95%)
1. Ozone Alone (day 0)	1.01 (.999 – 1.03)
2. $PM_{2.5}$ Alone (day 0)	1.04 (1.00 – 1.08)
3. Ozone & PM	.999(.997-1.02)
Lags as separate model (Ozone only)	**Odds Ratio(95%)**
1. Day One	1.01(.999-1.04)
2. Day Two	1.01(.999-1.03)
3. Day Three	1.00(.998-1.02)
4. Day Four	.999(.997-1.01)
5. Day Five	.998(.996-1.01)
5- Day Average	1.00(.997-1.04)

The results for the all months model was based on the full sample which represented individuals of all ages, including younger children and adolescents 0-17 years of age. Since the literature reports a differential in asthma prevalence by age, children (0-17 years of age) were removed from the dataset. This allowed us to determine whether the statistically significant results from the day-2 lag estimate observed in the all month's ozone model adjusted for temperature for all ages was influenced by the younger cohort. 1,007 individuals 0-17 years of age were removed from the dataset. The analysis was conducted using single lag models for ozone and a multiple lag model for ozone and $PM_{2.5}$ after adjusting for temperature.

Table 45: Single and multiple lag models for 24-hour maximum ozone, $PM_{2.5}$ and temperature for <u>adults only</u>: All months

Models	Odds Ratio (95%)	Odds Ratio (95%) Adjusted for temperature
1. Ozone Alone (day 0)	1.00(.998-1.02)	.999(.997-1.01)
2. $PM_{2.5}$ Alone (day 0)	1.02(.998-1.05)	.999(.996-1.03)
3. Ozone & PM	1.00(.996-1.04)	.999(.997-1.02)
Lags as separate model (Ozone only)	**Odds Ratio(95%)**	**Odds Ratio(95%) Adjusted for temperature**
1. Day One	1.02(1.00-1.04)	1.00(.999-1.03)
2. Day Two	**1.03(1.01-1.05)**	**1.02(1.00-1.04)***
3. Day Three	1.01(.999-1.03)	1.01(.999-1.03)
4. Day Four	1.00(.998-1.02)	1.00(.998-1.02)
5. Day Five	1.00(.999-1.02)	1.00(.998-1.02)
5- Day Average	1.03(1.00-1.07)	1.00(.998-1.05)

*Borderline significant

The single model lag estimate for ozone and $PM_{2.5}$ alone, or adjusted for temperature showed no risk (Table 45). The combined model of ozone and $PM_{2.5}$ adjusted for temperature was unchanged with no effect observed. The single lag model for days 0 through 5 and 5-day average showed no effect with the exception of the point estimate for the 2-day lag for ozone adjusted for temperature 1.02(95% CI 1.00-1.04). Although, the estimate is borderline significant this result suggests the need to conduct further analysis with the adult population using a larger sample size. The literature reports the influence of ambient pollutants on childhood asthma hospitalizations (Villeneuve, et al., 2007; Barnett, et al., 2005; Pope et al., 1991; Paulu & Smith, 2008). However, little research has been conducted which looked at the effects of ozone on asthma emergency room visits in adults. One study which adjusted for different age groups throughout found no effect from ozone in older adults in age groups 65-74 and 75+ (Villeneuve, et al., 2007). A significant effect was also observed for individuals 5-14

and 15-44 years of age for the 3-day average and 5-day average lag in the all month's and summer month's ozone model ($p = < 0.05$). In the same study, the 1-day lag was significant in both models for the 15-44 year old age group and the 5-day average lag in the all month's model for individuals 45-64 years of age.

Paulu & Smith (2008) also report an increase risk for the 4-day average lag estimate for ozone, which was associated with an 11% increase in asthma, related ED visits. The increase was more apparent among females than males, 12%(95% CI 6-18) vs. 11% (95% CI 4-18), respectively. An influence of age was observed with younger males 2-14 years of age having a 17% increase (95% CI 3-32) and females 15-34 years of age having a 20% increase (95% CI 10-31).

The influence of age has been reported as a significant factor that affects exposure levels. The body burden from exposure on children is greater due to a smaller body size and weight (Mathieu-Nolf's study cited in Villeneuve et al 2007, p. 10). Furthermore, the cells of children are not fully developed thus, placing them at increased risk for increased airway vulnerability (ALA, 2000). Contrasting to adults, particularly the elderly who are more susceptible due to comorbidity factors (Sandstrom et al. 2003 study cited Villeneuve et al 2007, p. 10). Thus, a misclassification of exposure in this case crossover study could have resulted from the differences in age.

4.2 DISCUSSION

The findings in this research study implicate ozone as a factor in asthma morbidity as evident by the 2% increase observed in asthma ED visits for a 10-ppb increase in the 24-hour

maximum ozone concentration at 2-day lag. This result was based on all individuals being included in the model. The all adults model, which removed children 0-17 years of age, was found to have a borderline risk estimate for the 2-day lag 1.02(1.00-1.04). A larger sample estimate may provide a different estimate of risk for adults.

Other studies have reported significant findings with different effect estimates using different ozone standards. Sample size estimates have been larger and smaller than what was used in this study. No study was identified in the literature, which neither used the 24-hour maximum ozone concentration level nor identified an effect estimate at 2-day lag for asthma morbidity. The only documentation found in the literature which reported a significant effect estimate for day-2 lag was observed with ozone and total mortality, cardiovascular and respiratory mortality and non-accidental mortality (Bell et al., 2004; Bell et al., 2005; Bell et al., 2007; Ito et al., 2005).

Paulu and Smith (2008) report an overall 11% increase in asthma emergency room visits associated with a 4-day average lag [0-3-days] for a 10-ppb increases in the 8-hour average ozone level. The differences reported for age were greater in the 15-34 year olds vs. the 2-14 year olds, 16(95% CI 8-24) vs. 11%(95% CI 1-23), respectively. However, the concentration of cases was greatest in males 2-14 years 17%(95% CI 3-32) and females 15-34 years of age 20%(95% CI 10-31).

A 34% increase in respiratory admissions was reported by Burnett et al (2001) in individuals less than 2 years of age for the 5-day average lag for the 1-hour maximum ozone concentration of 45-ppb, May through August. In this same study, an 18% increase in respiratory admissions was observed January through December for the 5-day average, but the

effect was not as great as that observed May through August. No effect was observed September through April.

Villeneuve (2007) reports that ozone exposure is the pollutant most frequently associated with asthma hospitalizations. An increase in asthma hospitalizations were observed for day-1 lag and 3 and 5-day averages for all months, and a greater effect observed in the summer months for the respective point estimates 1.04(95% CI 1.02-1.06); 1.07(95% CI 1.04-1.10); 1.08(95% CI 1.05-1.11), respectively, and 1.06(95% CI 1.04-1.09); 1.11(95% CI 1.07-1.16); 1.11(95% CI 1.06-1.15), respectively. The increases were statistically significant and observed for the 5-14 and 15-44 year old population for the 3 and 5-day lag average estimates. The effect was greatest in the summer months. An effect was also observed for the 45-64 year olds in all months for the 5-day lag average. No effect was observed with ozone for the oldest age groups 65-74, and 75+.

In summary, results from this research study were sufficient to demonstrate a significant 2-day lag effect, which corresponds to a 2% increase in all asthma ED visits related to a 10-ppb increase of the 24-hour maximum ozone concentration. The effect was not observed during the summer months when ozone has been reported to have its greatest effect on mortality and morbidity. However, due to a small sample size no risk was observed for ozone and asthma morbidity during that period. The 2% increase observed in asthma ED visits represents an addition of 203 asthma emergency room visits. These visits result in an increase use of emergency room services, which place a financial burden on the health care system. The total direct costs for asthma, which includes emergency room visits, asthma hospitalizations, physician services and medication use, are estimated at 14.7 billion dollars. Thus, a 2% increase observed for asthma ED visits noted from the 10ppb increase in the 24-hour maximum ozone concentration is significant when the financial burden is considered. If a larger sample estimate

were used the number of asthma ED visits would be significantly greater representing an even larger financial burden.

In conclusion, the findings from this research study are significant for several reasons. First, the use of emergency room visits is an extremely valuable methodology for conducting asthma surveillance. The larger number of visits per year in comparison to asthma hospitalizations represents a more definitive at risk population. Second, the short-term evaluation of asthma requires information on physician diagnosis and the hallmark signs and symptoms, which are the more detailed data that characterize this disorder. These data cannot be found in one location thus, by using the electronic medical record, data is available in one central location, which can be used to evaluate this diagnosis. The automated feature of the EMR improves the ability to apply knowledge-based models for scientific interpretation and validation of these data. The evaluation of these data can be done off site with an electronic transfer of the data. Furthermore, the facilitation of these data through an electronic system would support national tracking efforts for testing environmental hypotheses related to this chronic condition (L. Brink, personal communication, April 15, 2008). The linkage of the environmental data to the asthma ED visit in this research study generated significant findings regarding ozone exposure and asthma health risk. These findings are of great public health significance. The largest burden from asthma rests on the public at large. The 2008, '*State of the Air Report"* published by the American Lung Association (ALA) indicates that nearly half of the US population (46.0%) live in areas where either ozone or particulate levels are at unhealthy levels. Even at levels meeting the standard, Bell reports "that there is still a significant link between ozone and premature mortality" (Bell, Peng, & Dominici, 2006). In Allegheny County alone, the change in the national 8-hour ozone standard from 0.084-ppm to 0.075-ppm in May 2008 has

resulted in the identification of 9 days that attainment was exceeded. Had the standard not have been lowered, only 3 days would have been regarded as exceeding the standard (Allegheny County Health Department, Air Quality Program, May 2008) (Appendix E). In Allegheny County, a health alert is used to alert the public about exceedance days for ozone and particulates. This action not only serves as an educational intervention to alert the vulnerable populations with acute asthma, but also to inform public health policy for more stringent laws on ozone.

APPENDIX A

IRB

University of Pittsburgh
Institutional Review Board

Exempt and Expedited Reviews

University of Pittsburgh FWA: 00006790
University of Pittsburgh Medical Center: FWA 00006735
Children's Hospital of Pittsburgh: FWA 00000600

3500 Fifth Avenue
Suite 100
Pittsburgh, PA 15213
Phone: 412.383.1480
Fax: 412.383.1508

TO: Jo Ann Glad, BSN, MPH

FROM: Christopher M. Ryan, PhD, Vice Chair

DATE: December 12, 2006

PROTOCOL: The Feasibility of Using Electronic Medical Records to Determine Asthma Prevalence

IRB Number: 0610013

The above-referenced protocol has been reviewed by the University of Pittsburgh Institutional Review Board. Based on the information provided in the IRB protocol, this project meets all the necessary criteria for an exemption, and is hereby designated as "exempt" under section 45 CFR 46.101(b)(4).

- If any modifications are made to this project, please submit an 'exempt modification' form to the IRB.
- Please advise the IRB when your project has been completed so that it may be officially terminated in the IRB database.
- This research study may be audited by the University of Pittsburgh Research Conduct and Compliance Office.

Approval Date: December 12, 2006

CR:kh

APPENDIX B

PERL PROGRAM

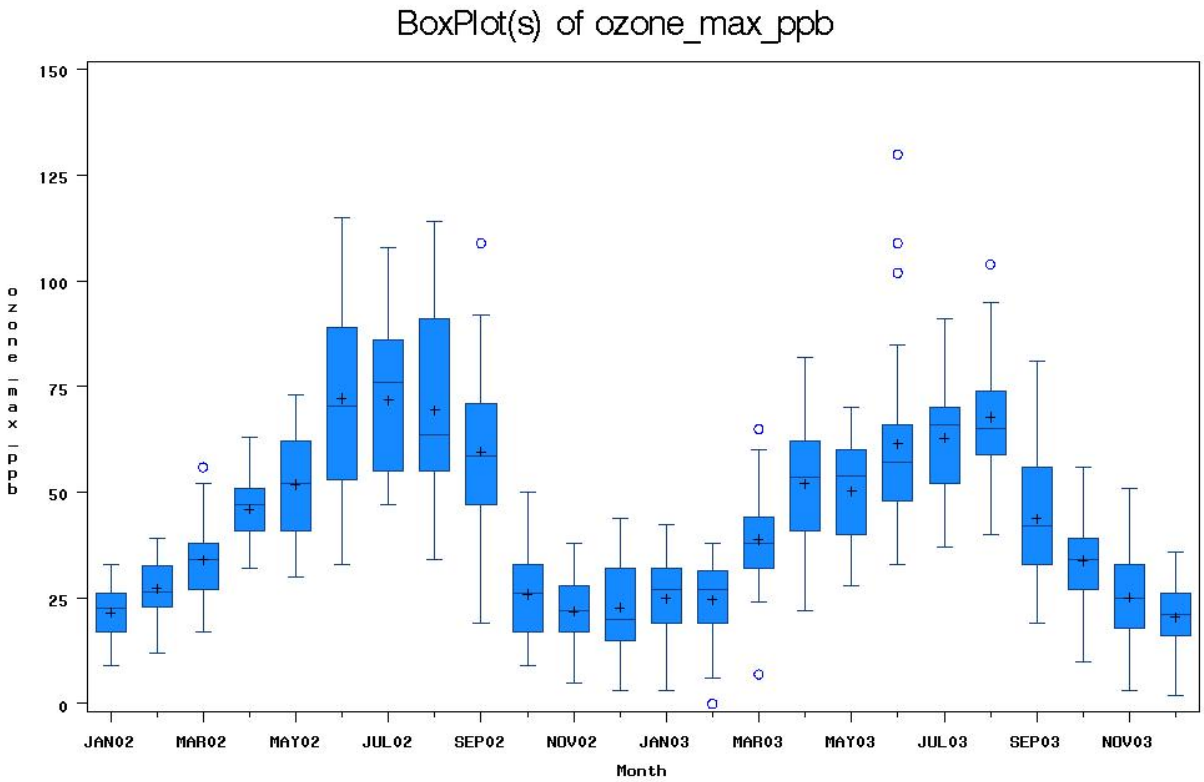
```
##################################################
#
#
#     Processing Medical Records          #
##################################################

open (INDAT, "<search.txt") || die "Error -  unable to open input file $!";

while (<INDAT>){
      chomp($_);
      $_ =~ /;/;
      $title = $`;
      $' =~ /;/;
      $positive = $`;
      $negitive = $';

      while ($positive =~ /,/){
            $positives{$`} = $`;
            $positive = $';
            $pos = $';
      }
      $positives{$pos} = $pos;

      while ($negitive =~ /,/){
            $negitives{$`} =$`;
            $negitive = $';
            $neg = $';
      }
      $negitives{$neg} =$neg;

      $search{title} = $title;
      $search{positives} = {%positives};
      $search{negitives} = {%negitives};

      $searches{$title} = {%search};

      %positives = ();
      %negitives = ();
      %search = ();
}

close (INDAT);

open (INDAT, "<input.txt") || die "Error -  unable to open input file $!";

while (<INDAT>){
      chomp($_);
      $record = $record.$_;               #Accumulate the text of the record
into a single string without line feeds
      if (/^Report ID/){
            $reportID = substr($_, 31);
      }elsif (/^Visit IDentifier/){
            $visitID = substr($_, 24, 39);
      }elsif (/^Principal Date/){
```

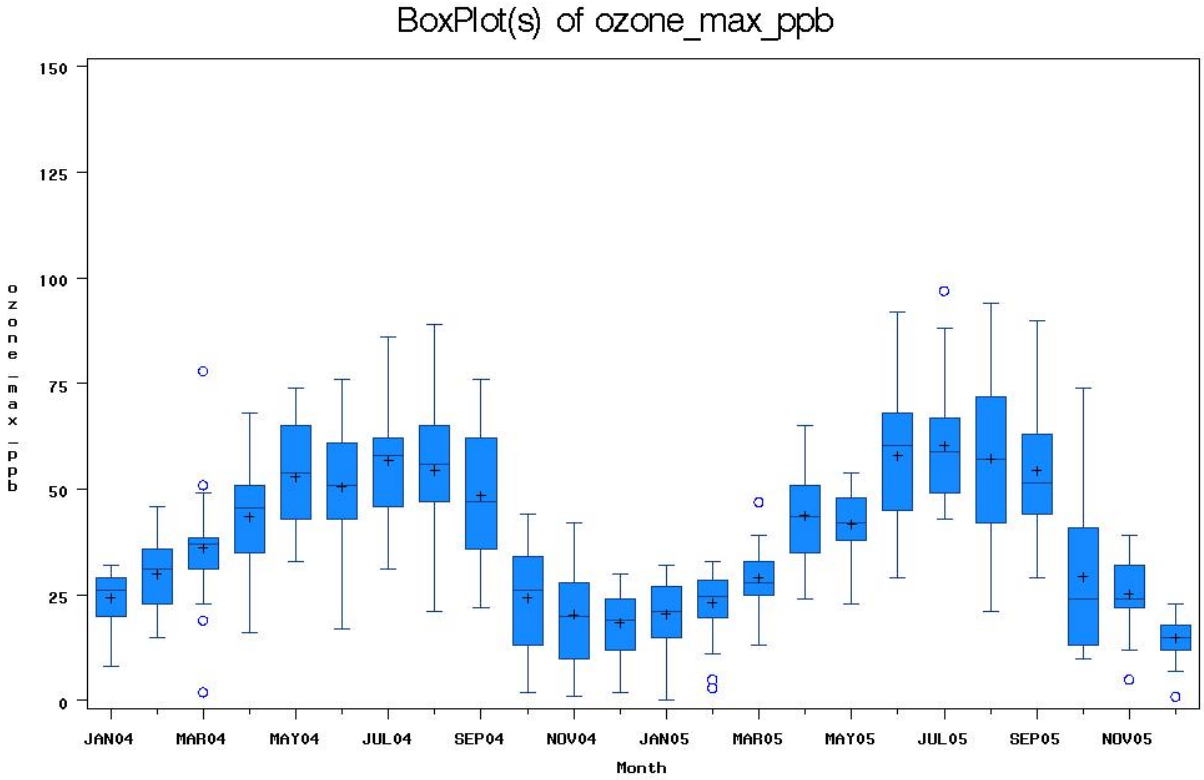
```
            $time = substr($_, 30, 8).substr($_, 39, 4);
      }elsif (/^Record Type/){
            $_ =~ /[A-Z]+$/;
            $recordtype = $&;
      }elsif (/^E_O_R/){                  #Create data structure for record
            $record{reportID} = $reportID;
            $record{visitID} = $visitID;
            $record{time} = $time;
            $record{recordtype} = $recordtype;
            $record{record} = $record;

            $records{$reportID} = {%record};

            $record = "";
      }
}

close (INDAT);

open (OUTDAT, ">regex.csv") || die "Error - unable to open output file $!";
print OUTDAT ",reportID,visitID,time,Record Type";

foreach (sort keys%searches) {
      $s = $searches{$_};
      %search = %$s;
      print OUTDAT ",$search{title},pos,neg";

}
print OUTDAT "\n";

foreach (sort keys%records) {
      $s = $records{$_};
      %record = %$s;

      print
OUTDAT"$record{reportID},$record{visitID},$record{time},$record{recordtype}";
print "$_\n";
      foreach (sort keys%searches) {
            $s = $searches{$_};
            %search = %$s;

            $s = $search{positives};
            %positives = %$s;

            foreach (sort keys%positives){
                  $pregex = $pregex . "$positives{$_}|";
            }
            chop($pregex);
            $pregex = "(" . $pregex . ")";

            $s = $search{negitives};
            %negitives = %$s;

            foreach (sort keys%negitives){
                  $nregex = $nregex . "$negitives{$_}|";
            }
            chop($nregex);
```

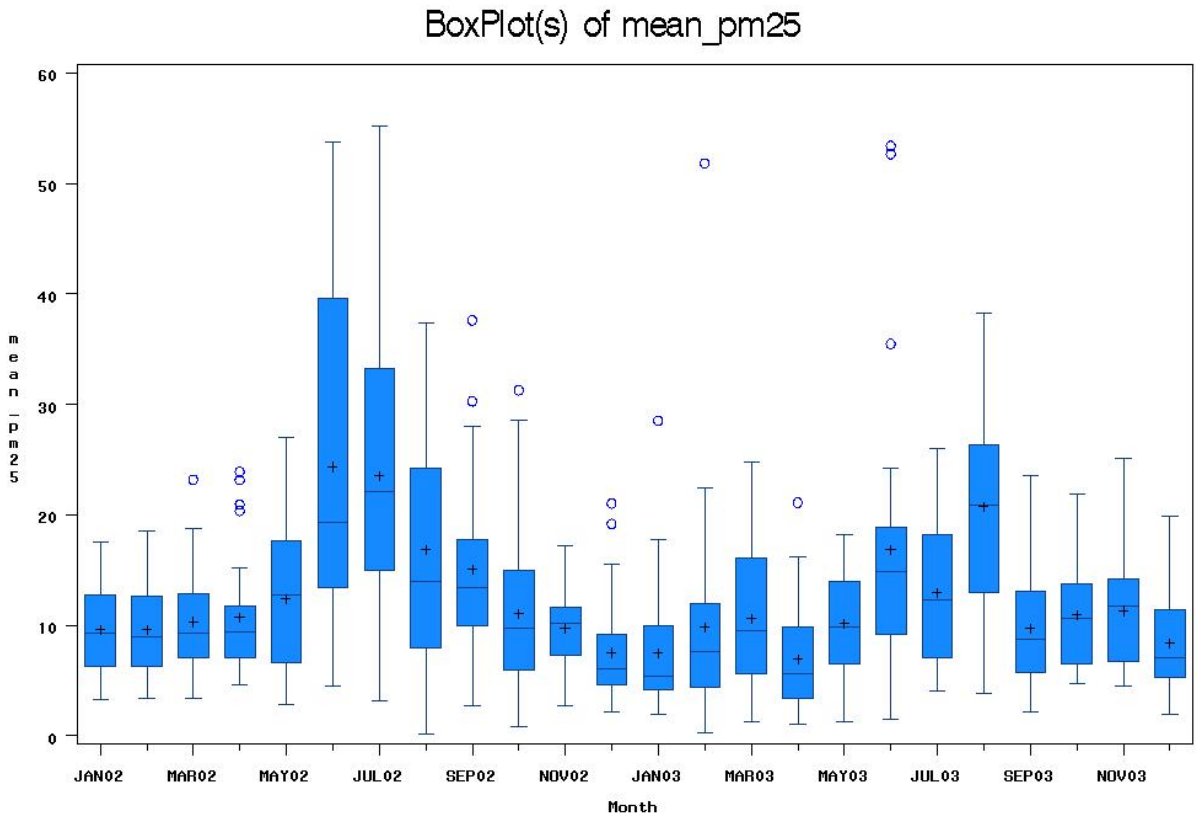
```
            $nregex = " (" . $nregex . ") [^\\.]+" . $pregex;

print "$pregex\n$nregex\n\n";

            $answer = 2;
            if ($record{record} =~ /$pregex/i){
                  $answer = 1;
$phit = $&;
            }
            if ($record{record} =~ /$nregex/i){
                  $answer = 0;
$nhit = $&;
            }
            print OUTDAT ",$answer,\"$phit\",\"$nhit\"";

$phit = "";
$nhit = "";
            $pregex = "";
            $nregex = "";
      }
      print OUTDAT "\n";
}

close(OUTDAT);
#EOF
```

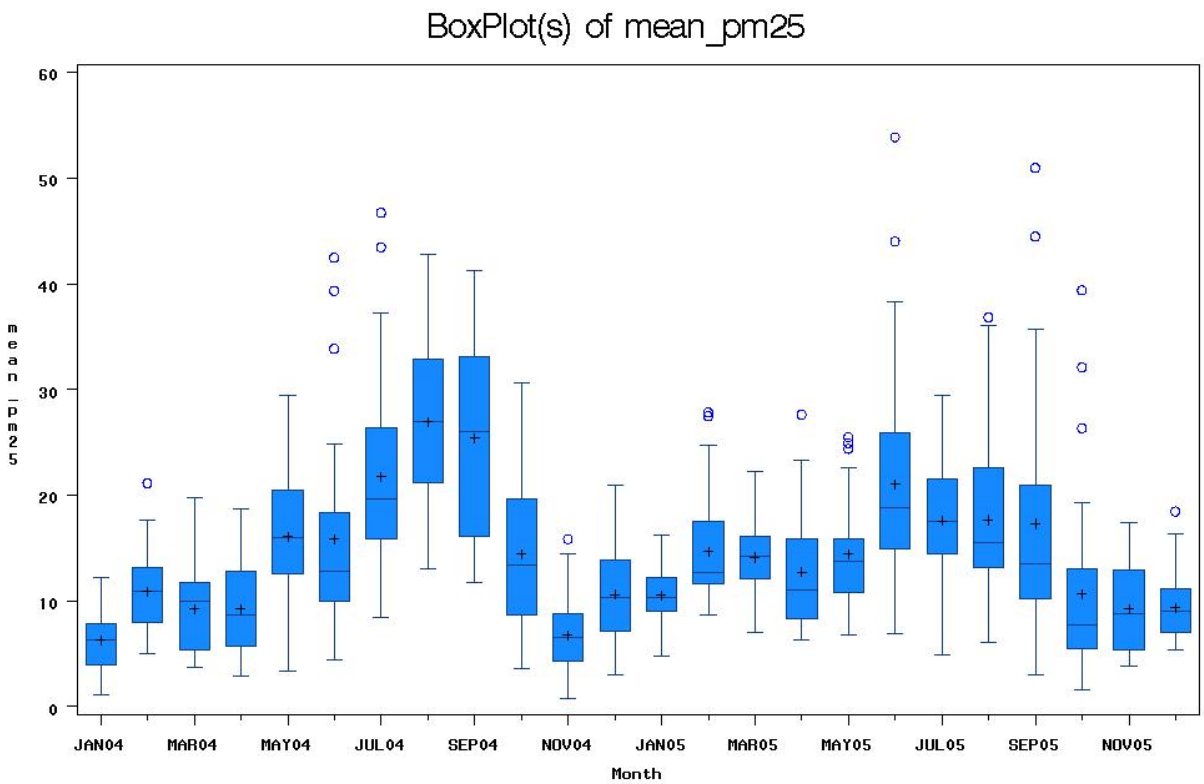
APPENDIX C

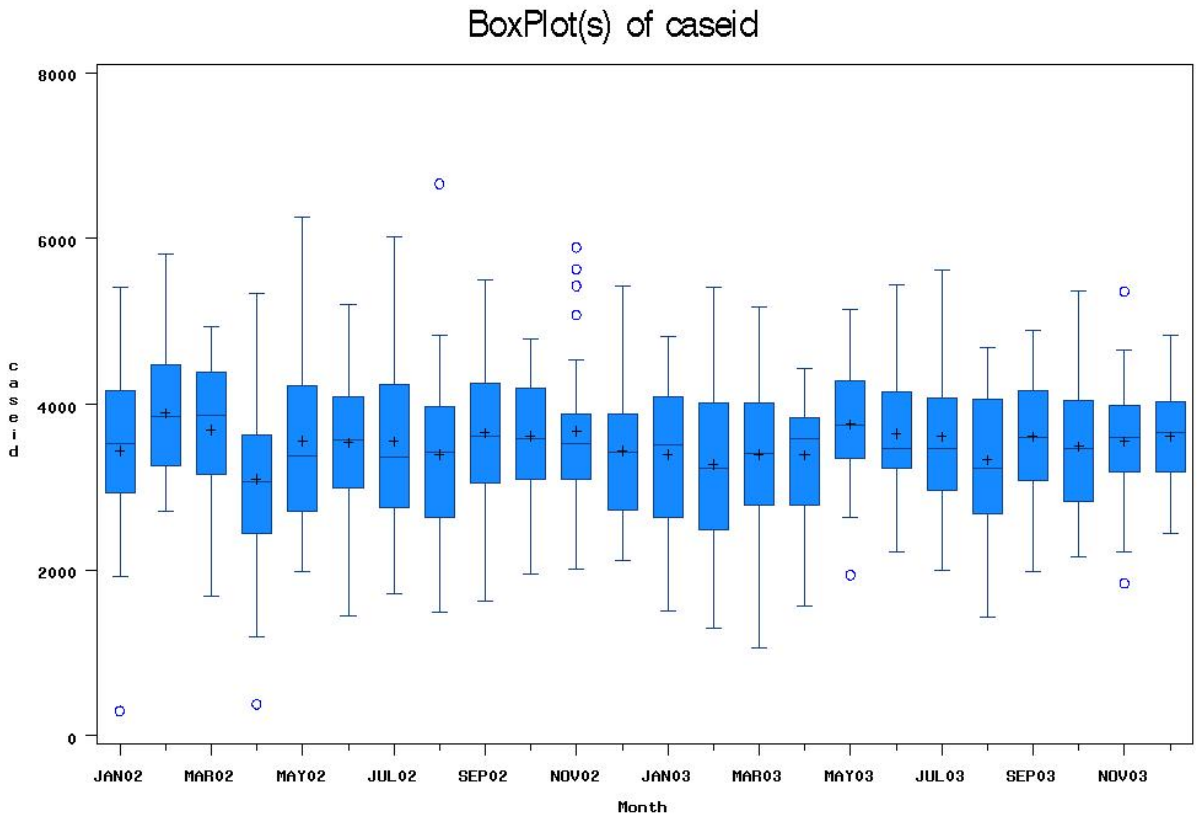
AIR POLLUTION DATA: ALL MONTHS

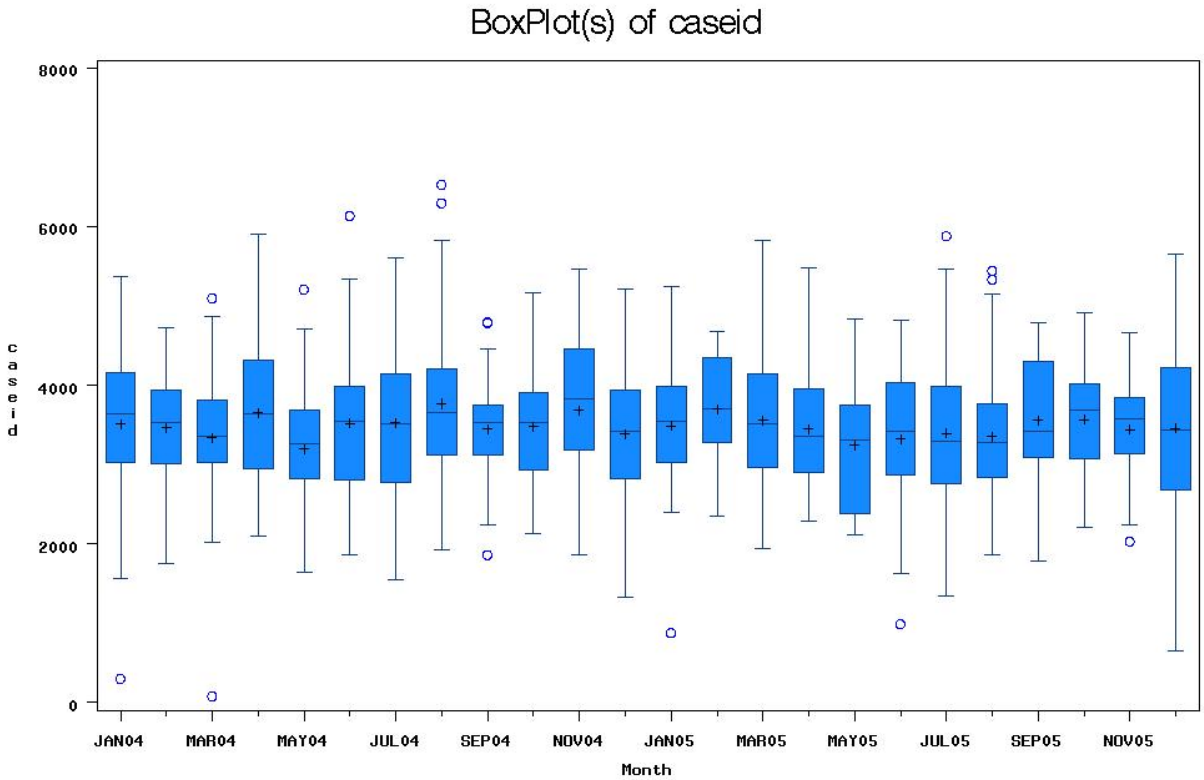

BoxPlot(s) of ozone_max_ppb
ozone_max_ppb
150
125
100
75
50
25
0
JAN02
MAR02
MAY02
JUL02
SEP02
NOV02
JAN03
MAR03
MAY03
JUL03
SEP03
NOV03
Month

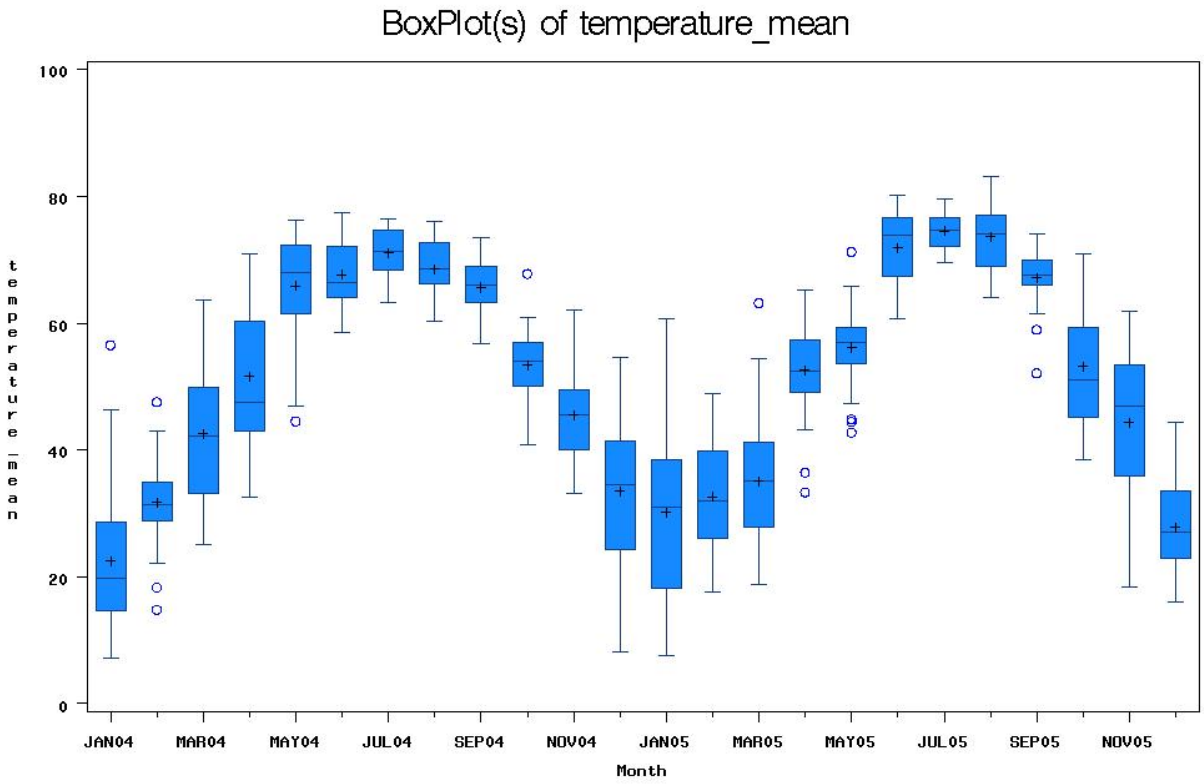

BoxPlot(s) of ozone_max_ppb
ozone_max_ppb
150
125
100
75
50
25
0
JAN04
MAR04
MAY04
JUL04
SEP04
NOV04
JAN05
MAR05
MAY05
JUL05
SEP05
NOV05
Month

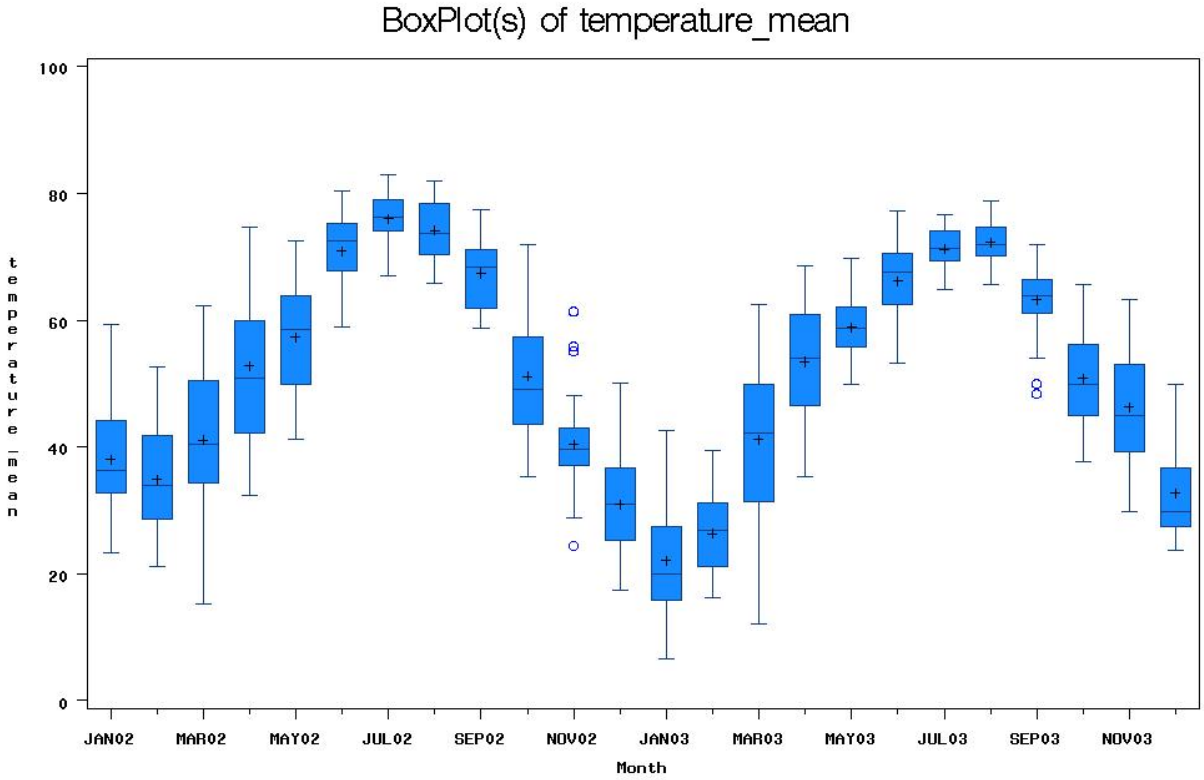

BoxPlot(s) of mean_pm25
mean_pm25
0
10
20
30
40
50
60
JAN02
MAR02
MAY02
JUL02
SEP02
NOV02
JAN03
MAR03
MAY03
JUL03
SEP03
NOV03
Month

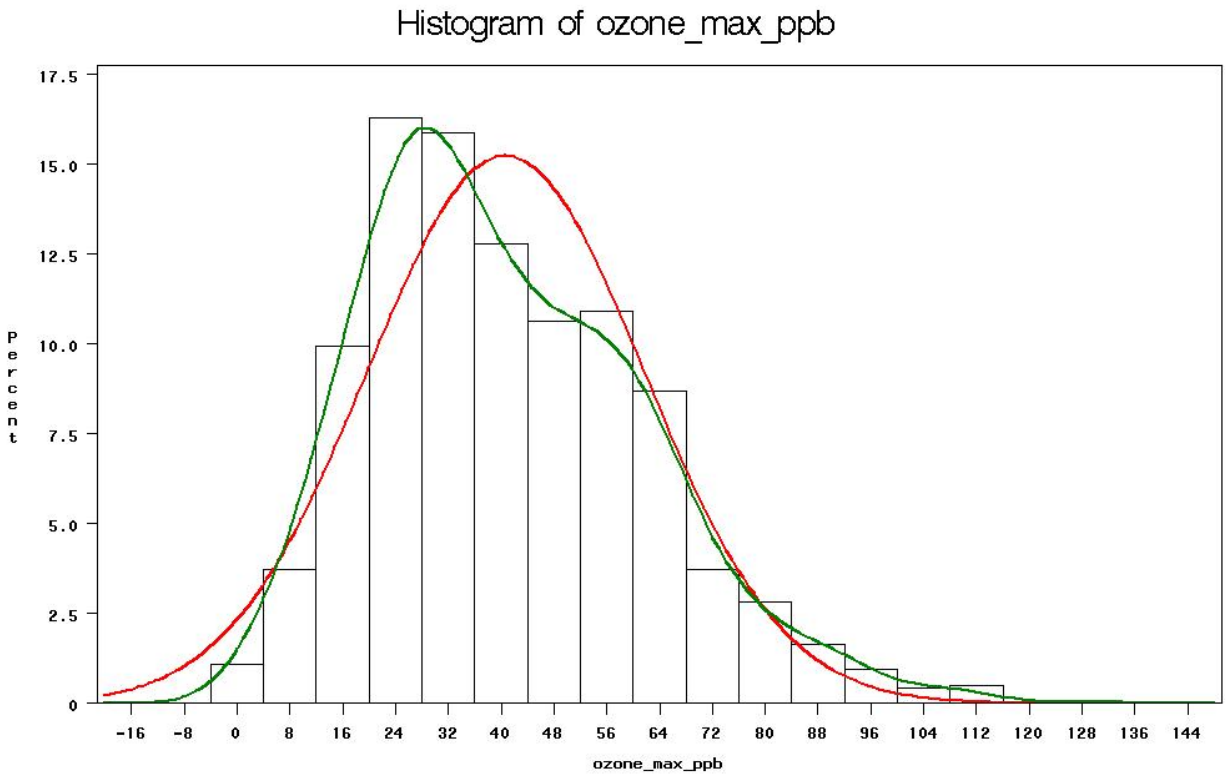

BoxPlot(s) of mean_pm25
mean_pm25
0
10
20
30
40
50
60
JAN04
MAR04
MAY04
JUL04
SEP04
NOV04
JAN05
MAR05
MAY05
JUL05
SEP05
NOV05
Month

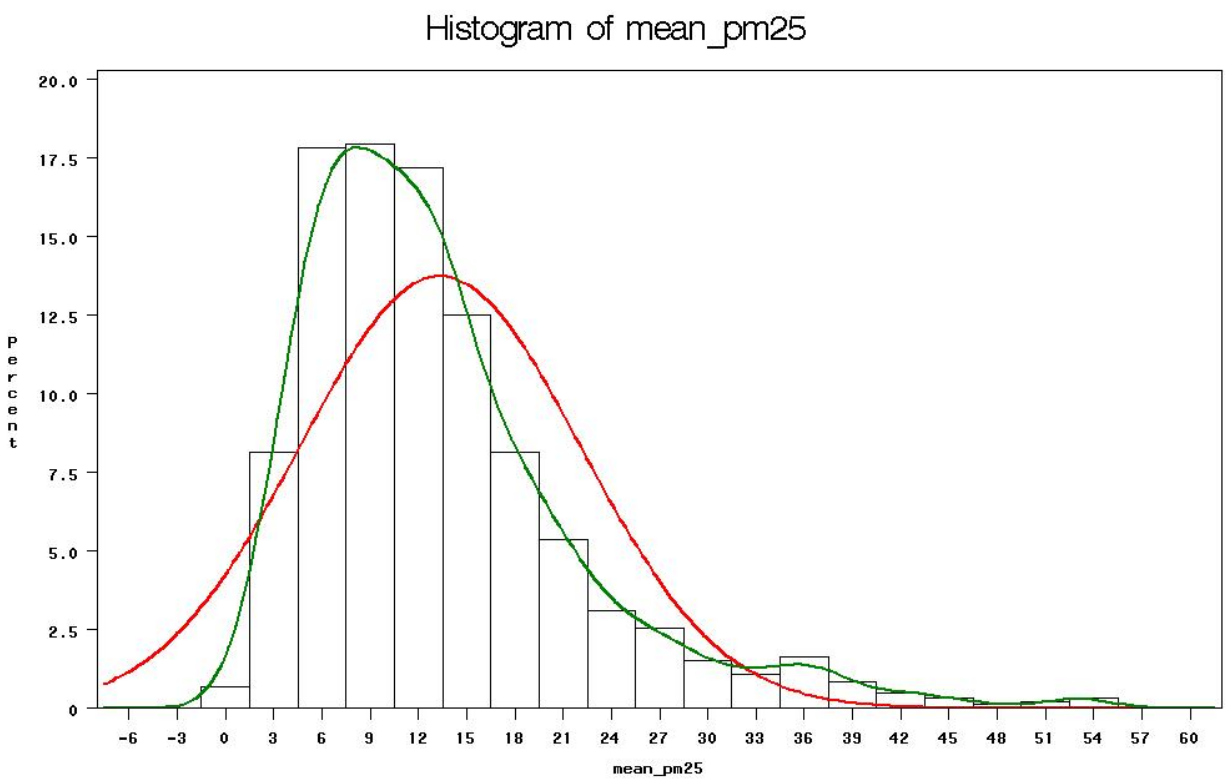

BoxPlot(s) of caseid
8000
6000
4000
2000
0
caseid
JAN02
MAR02
MAY02
JUL02
SEP02
NOV02
JAN03
MAR03
MAY03
JUL03
SEP03
NOV03
Month

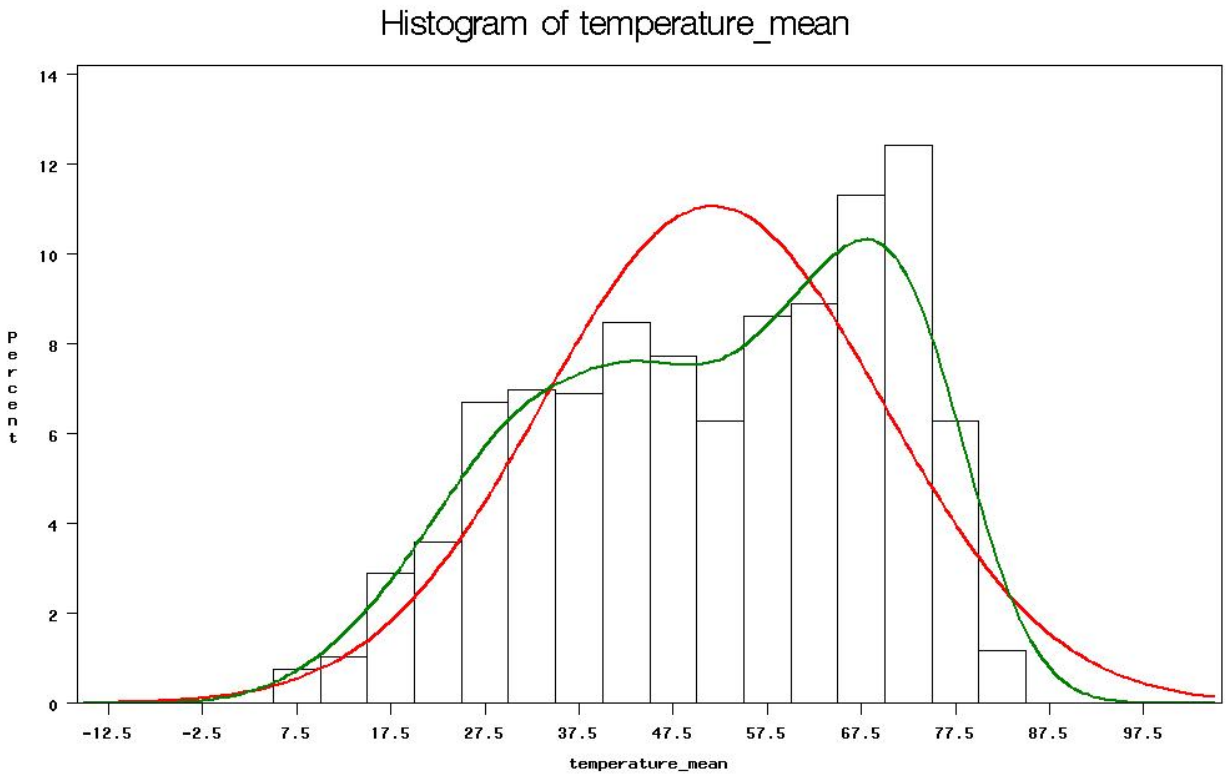

BoxPlot(s) of caseid
8000
6000
4000
2000
0
caseid
JAN04
MAR04
MAY04
JUL04
SEP04
NOV04
JAN05
MAR05
MAY05
JUL05
SEP05
NOV05
Month

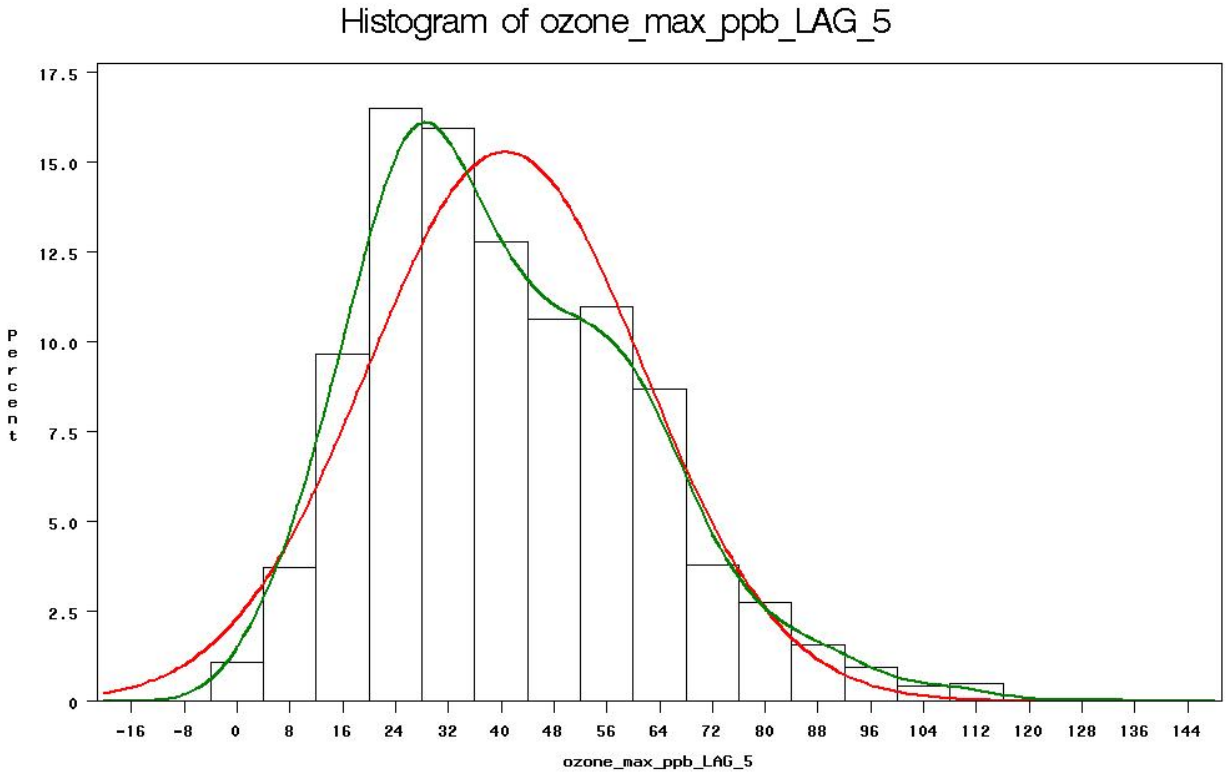

BoxPlot(s) of temperature_mean
temperature_mean
100
80
60
40
20
0
JAN04
MAR04
MAY04
JUL04
SEP04
NOV04
JAN05
MAR05
MAY05
JUL05
SEP05
NOV05
Month

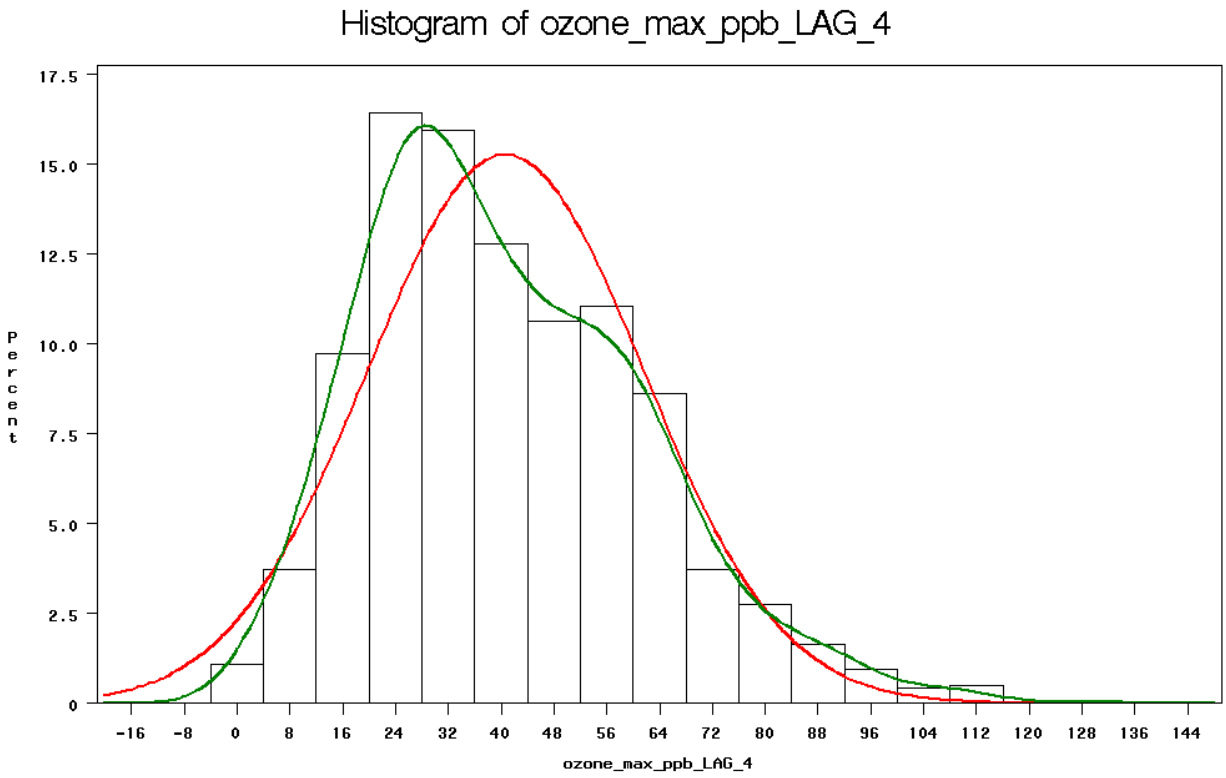

BoxPlot(s) of temperature_mean
temperature_mean
100
80
60
40
20
0
JAN02
MAR02
MAY02
JUL02
SEP02
NOV02
JAN03
MAR03
MAY03
JUL03
SEP03
NOV03
Month

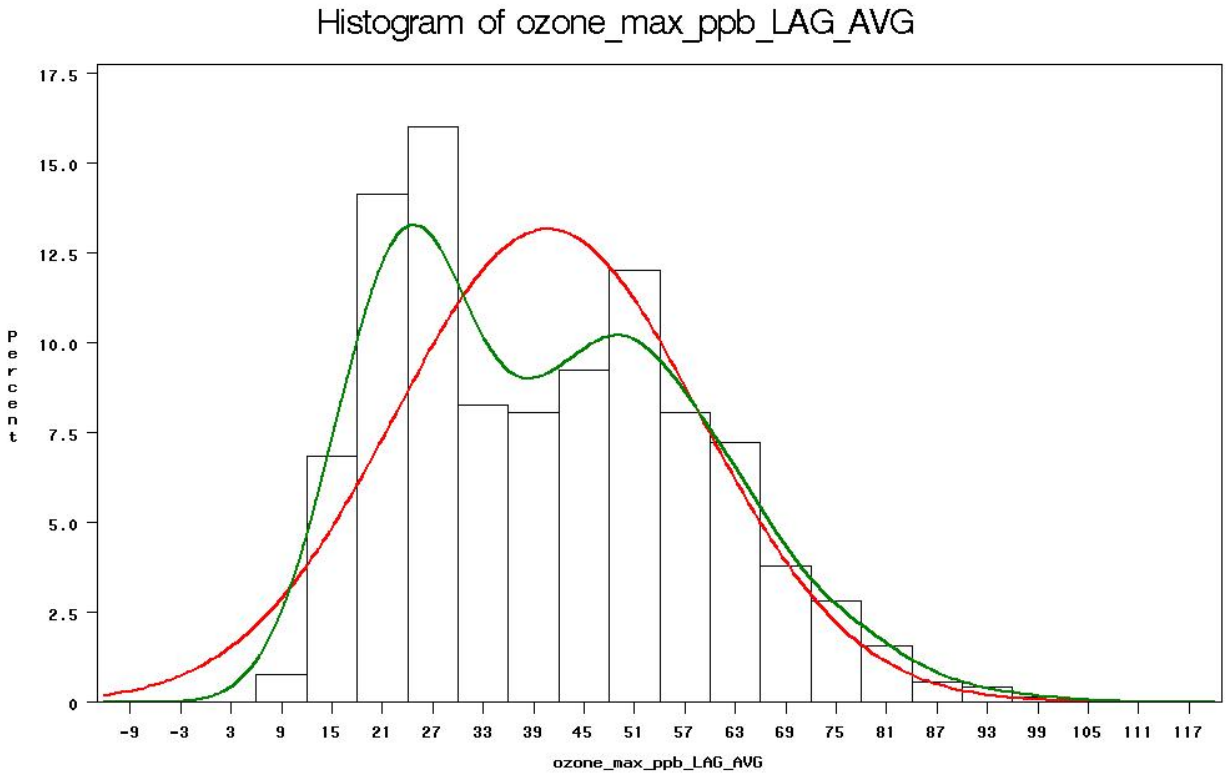

Histogram of ozone_max_ppb
Percent
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0
-16 -8 0 8 16 24 32 40 48 56 64 72 80 88 96 104 112 120 128 136 144
ozone_max_ppb

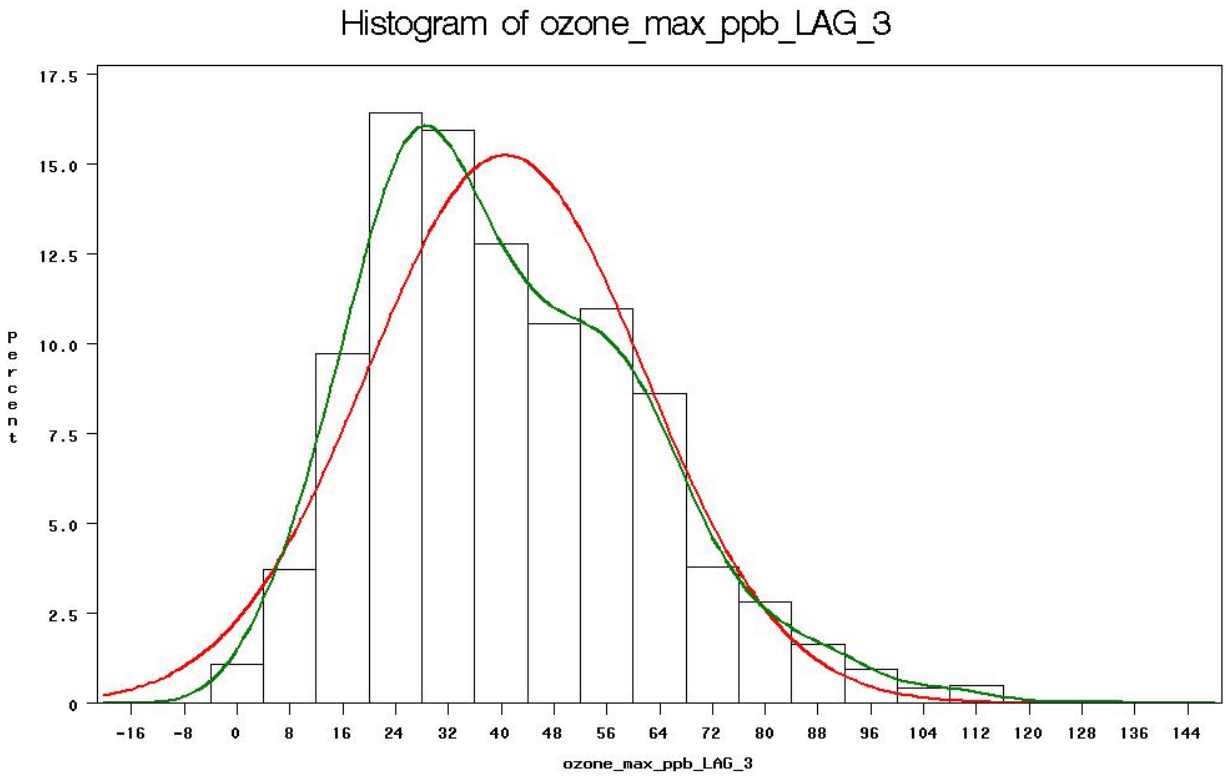

Histogram of mean_pm25
Percent
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0
-6 -3 0 3 6 9 12 15 18 21 24 27 30 33 36 39 42 45 48 51 54 57 60
mean_pm25

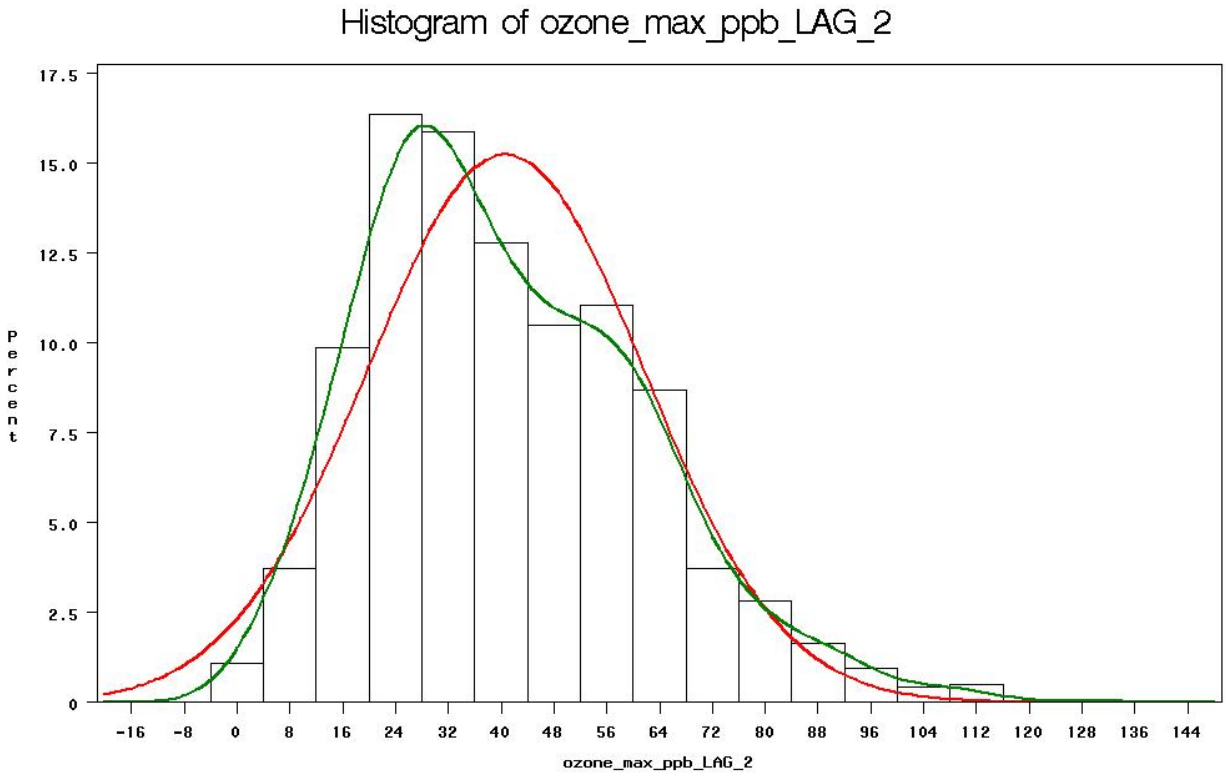

Histogram of temperature_mean
Percent
0
2
4
6
8
10
12
14
-12.5
-2.5
7.5
17.5
27.5
37.5
47.5
57.5
67.5
77.5
87.5
97.5
temperature_mean

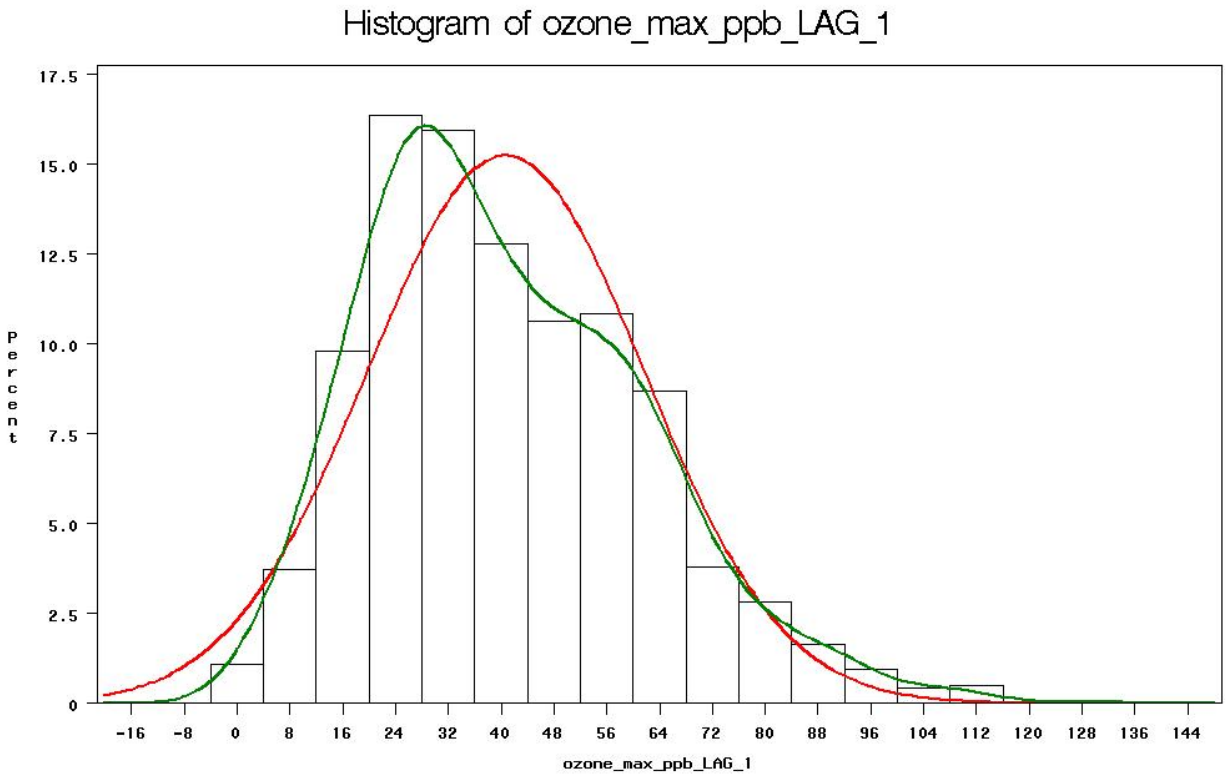

Histogram of ozone_max_ppb_LAG_5
Percent
0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
-16
-8
0
8
16
24
32
40
48
56
64
72
80
88
96
104
112
120
128
136
144
ozone_max_ppb_LAG_5

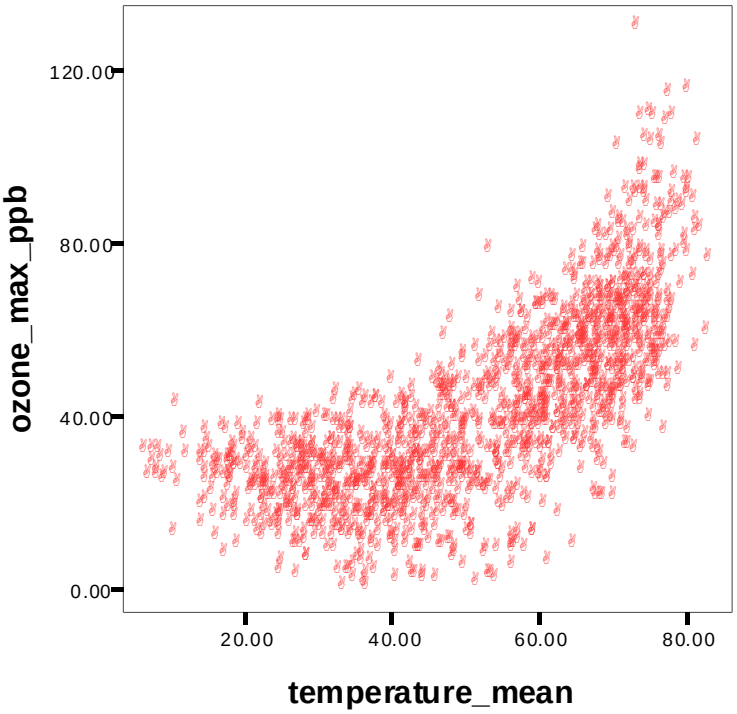

Histogram of ozone_max_ppb_LAG_4
Percent
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0
-16 -8 0 8 16 24 32 40 48 56 64 72 80 88 96 104 112 120 128 136 144
ozone_max_ppb_LAG_4


Histogram of ozone_max_ppb_LAG_AVG
Percent
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0
-9 -3 3 9 15 21 27 33 39 45 51 57 63 69 75 81 87 93 99 105 111 117
ozone_max_ppb_LAG_AVG

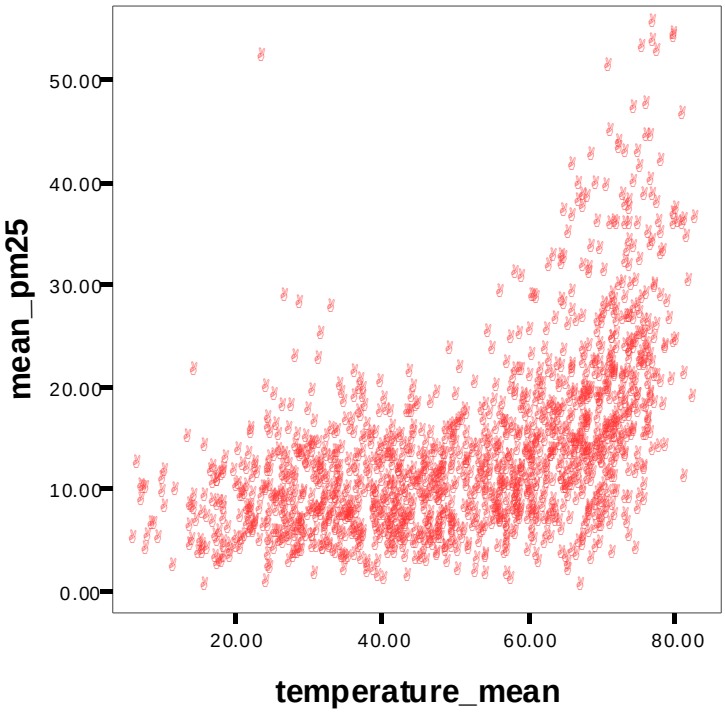

Histogram of ozone_max_ppb_LAG_3
Percent
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0
-16 -8 0 8 16 24 32 40 48 56 64 72 80 88 96 104 112 120 128 136 144
ozone_max_ppb_LAG_3


Histogram of ozone_max_ppb_LAG_2
Percent
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0
-16 -8 0 8 16 24 32 40 48 56 64 72 80 88 96 104 112 120 128 136 144
ozone_max_ppb_LAG_2

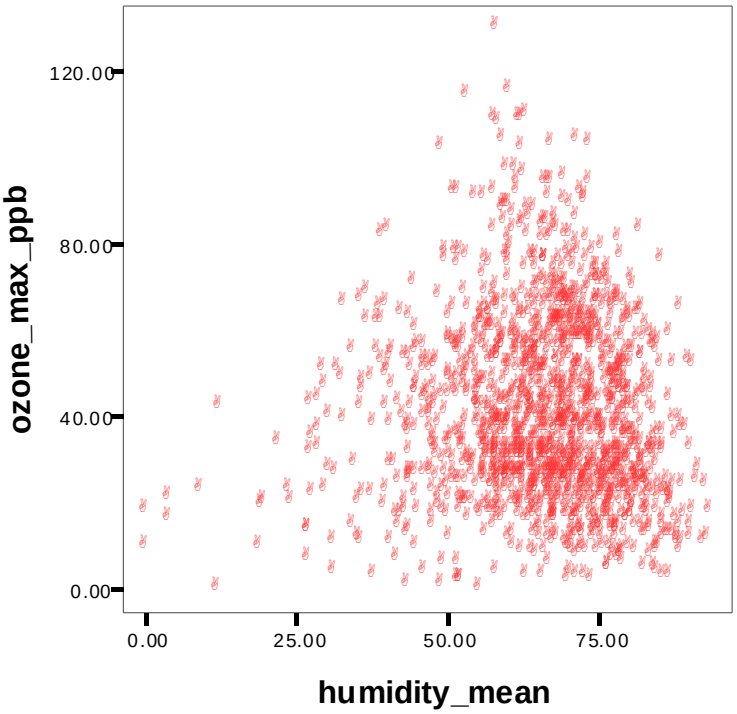

Histogram of ozone_max_ppb_LAG_1
Percent
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0
-16
-8
0
8
16
24
32
40
48
56
64
72
80
88
96
104
112
120
128
136
144
ozone_max_ppb_LAG_1


120.00
80.00
40.00
0.00
ozone_max_ppb
20.00
40.00
60.00
80.00
temperature_mean

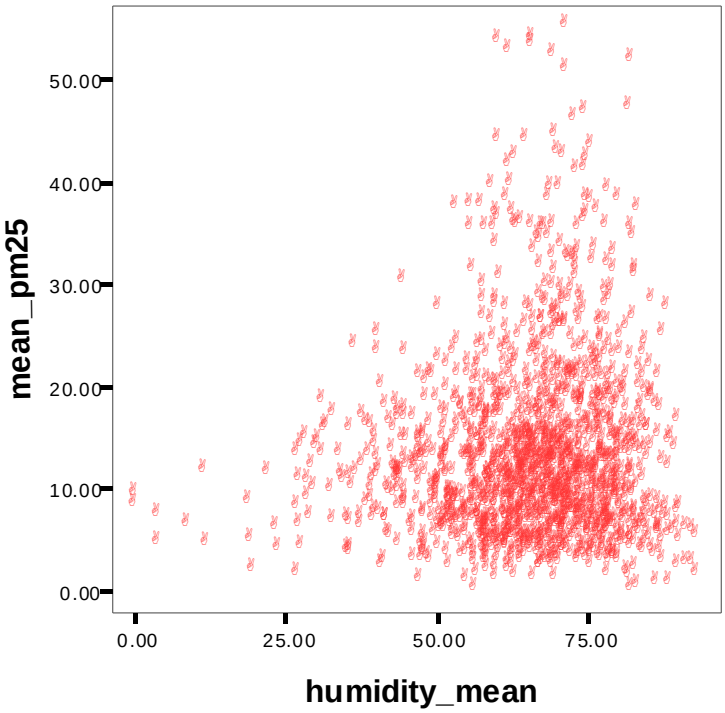
r =.72


mean_pm25
50.00
40.00
30.00
20.00
10.00
0.00
20.00
40.00
60.00
80.00
temperature_mean

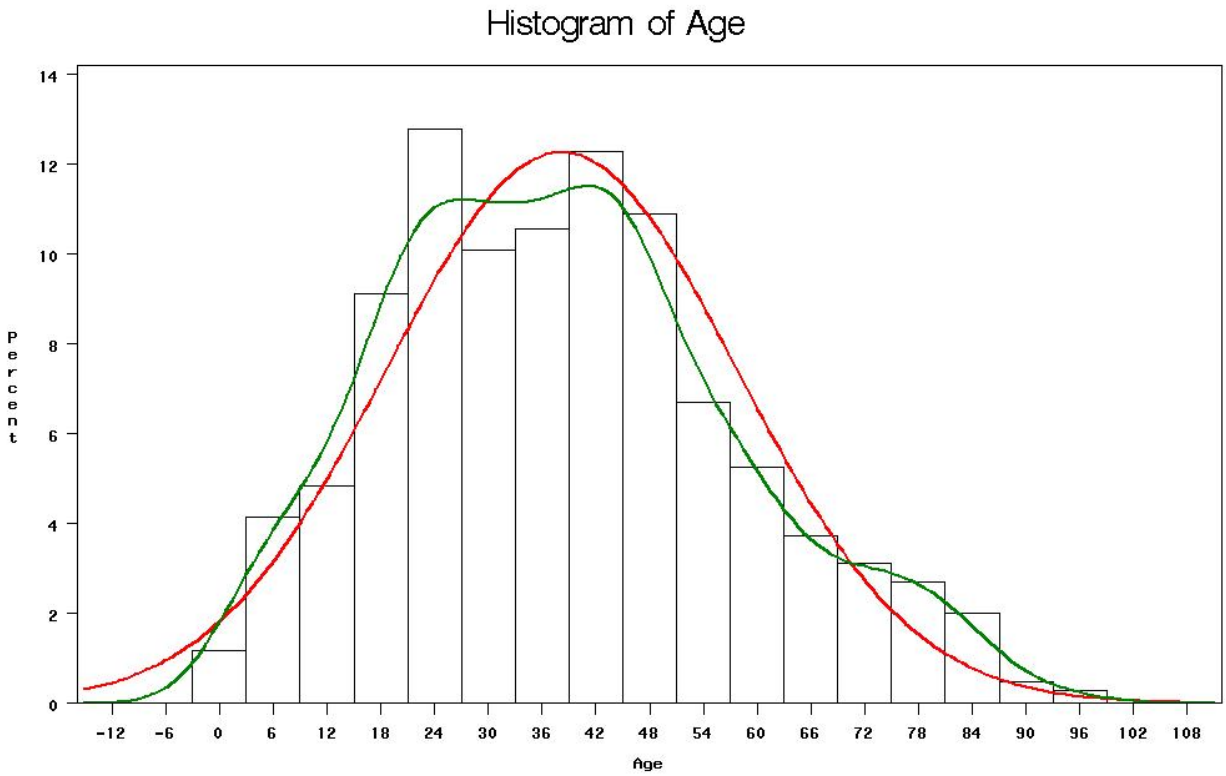
r = .52


120.00
80.00
40.00
0.00
ozone_max_ppb
0.00
25.00
50.00
75.00
humidity_mean

r = -.04

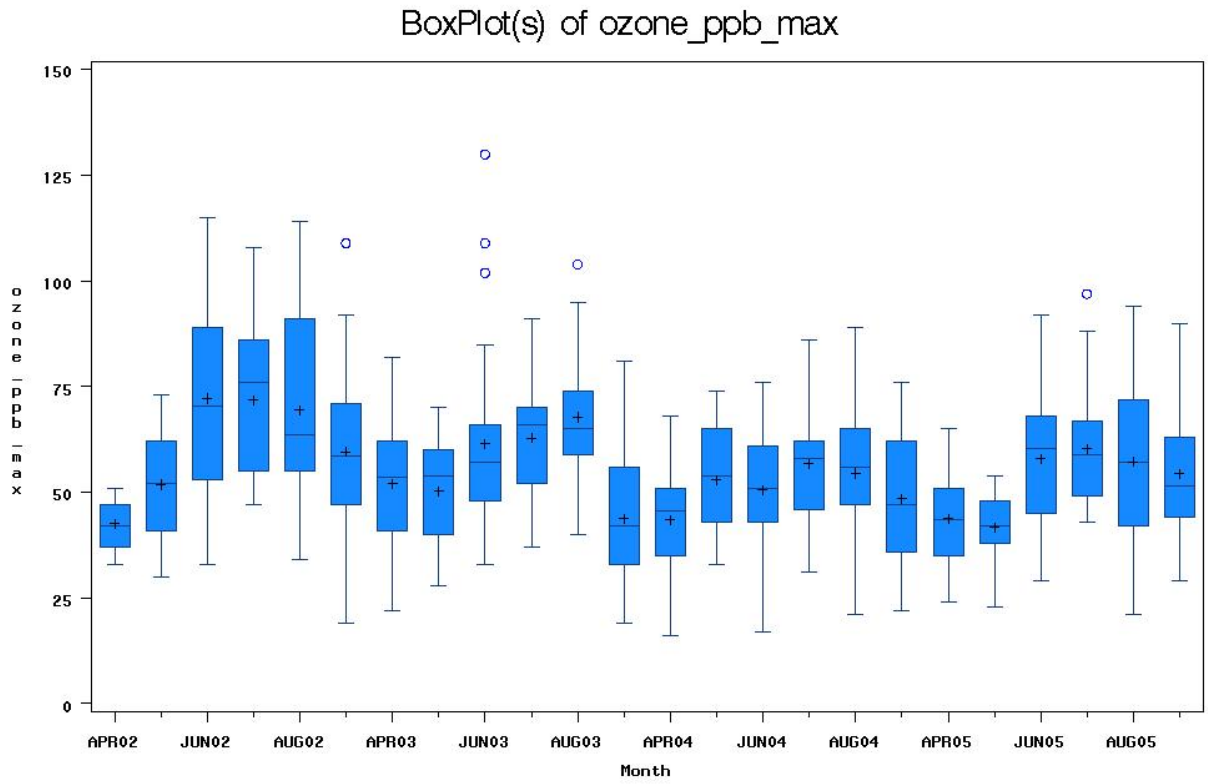

mean_pm25
50.00
40.00
30.00
20.00
10.00
0.00
0.00
25.00
50.00
75.00
humidity_mean

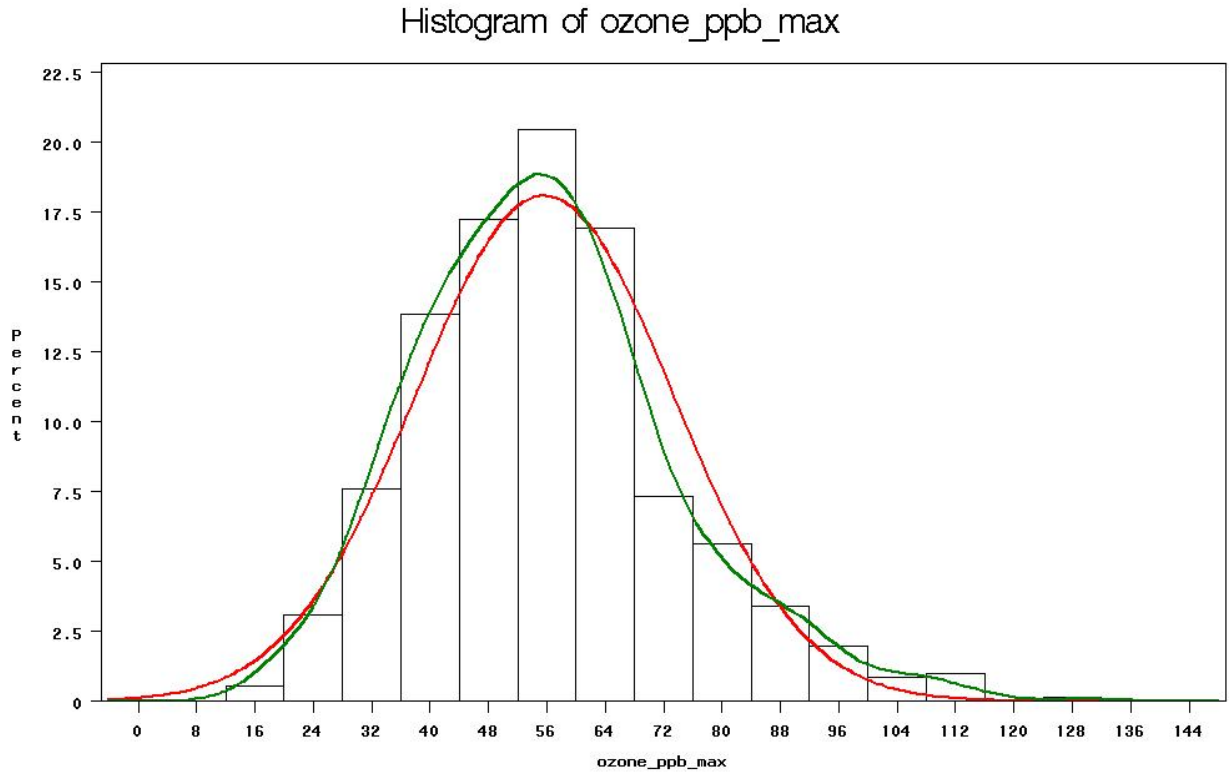
r = .08

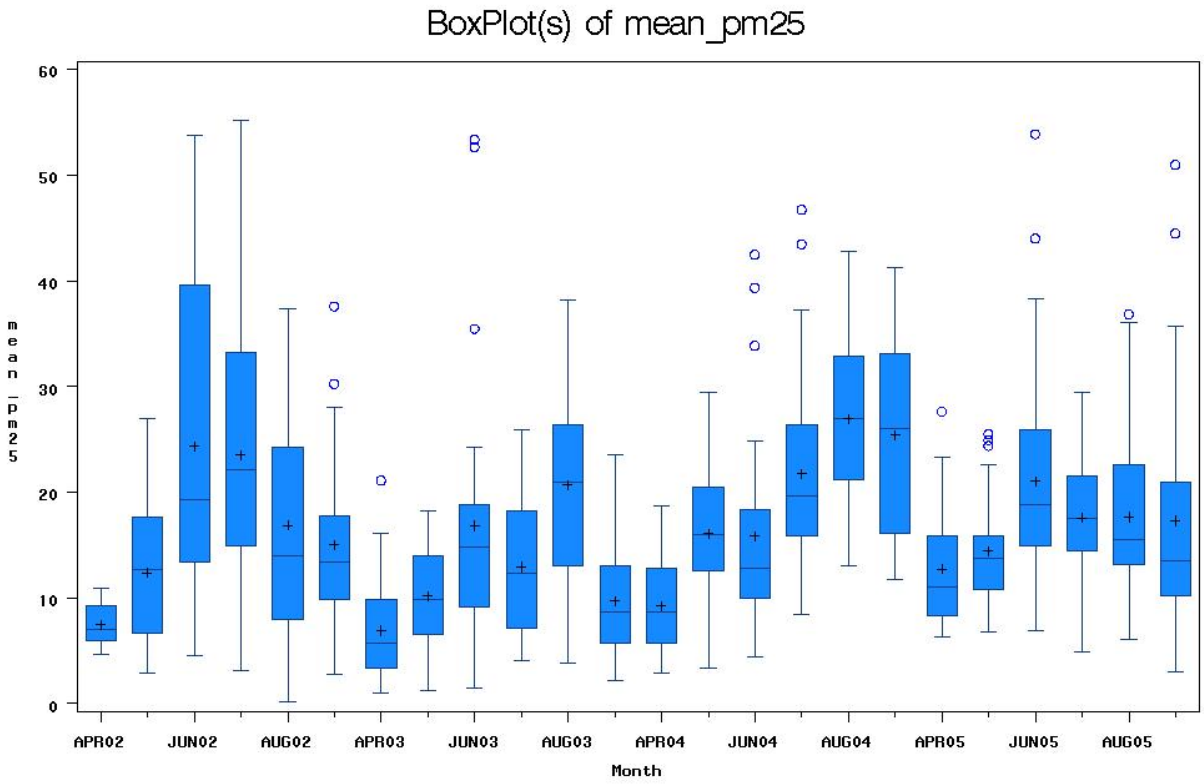

Histogram of Age
Percent
0
2
4
6
8
10
12
14
-12
-6
0
6
12
18
24
30
36
42
48
54
60
66
72
78
84
90
96
102
108
Age

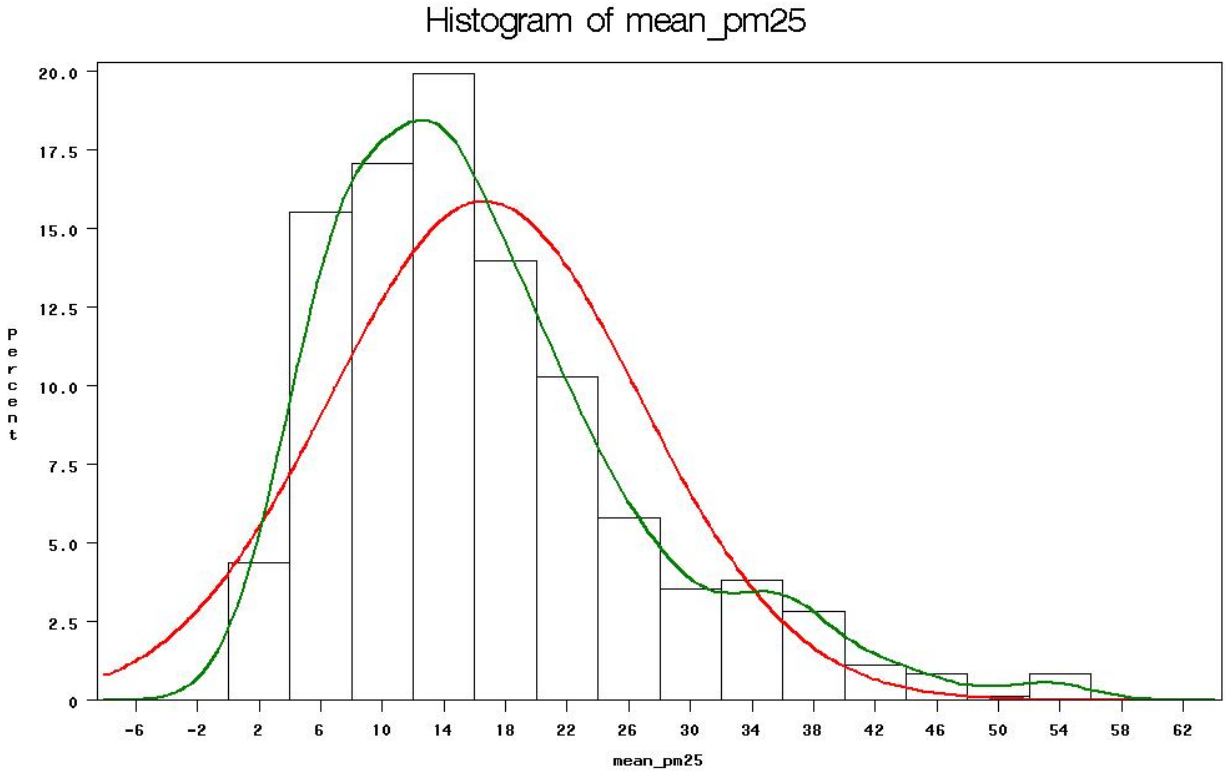
APPENDIX D

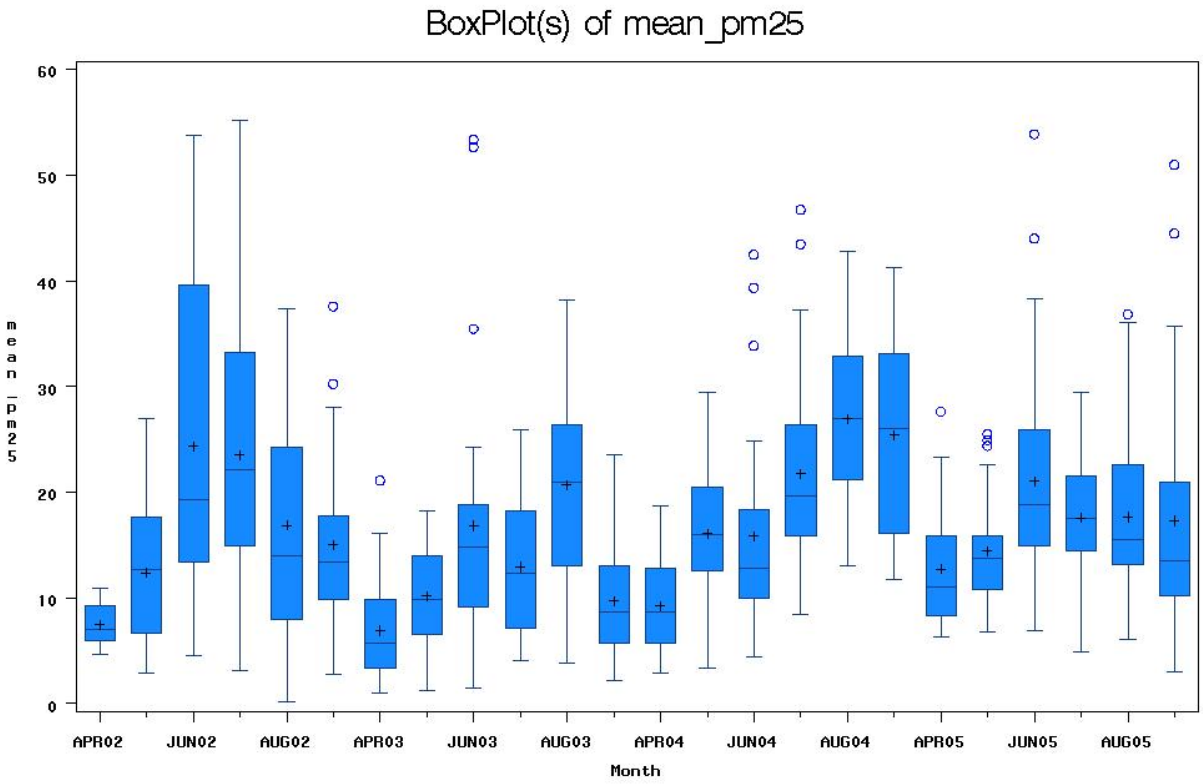
AIR POLLUTION DATA: SUMMER MONTHS ONLY

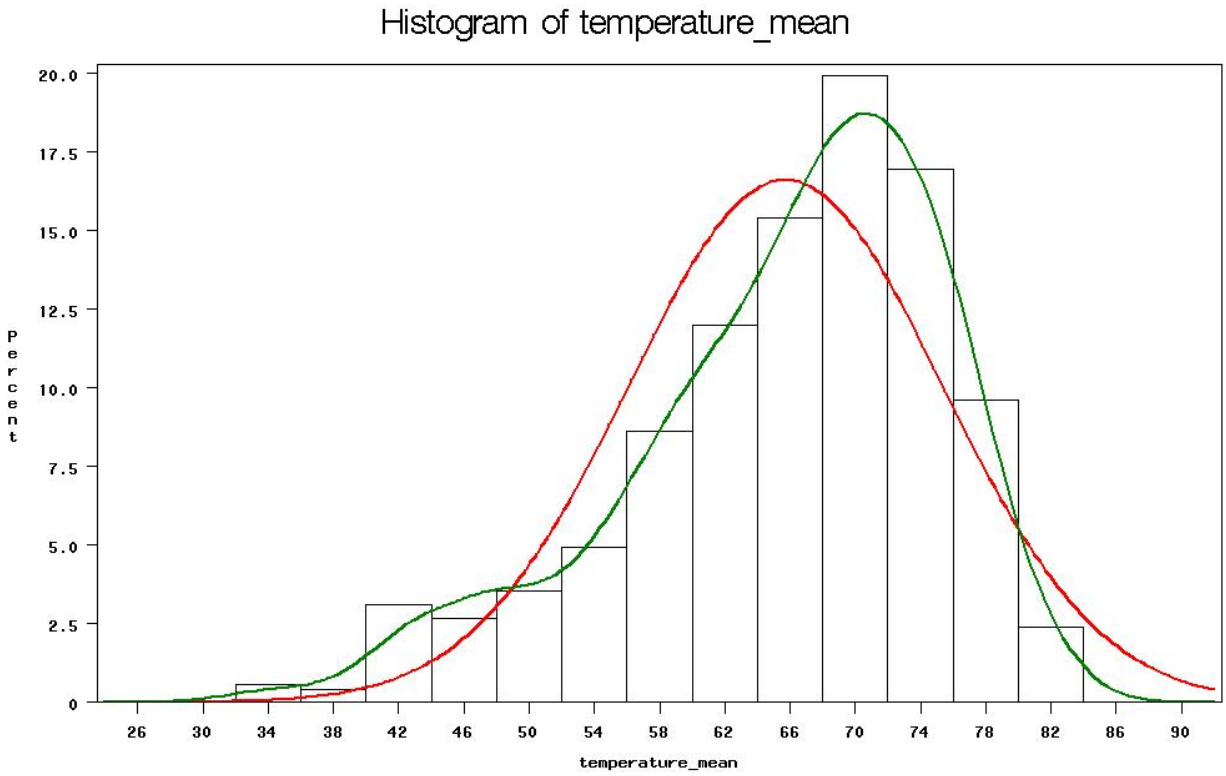

BoxPlot(s) of ozone_ppb_max
ozone_ppb_max
150
125
100
75
50
25
0
APR02
JUN02
AUG02
APR03
JUN03
AUG03
APR04
JUN04
AUG04
APR05
JUN05
AUG05
Month

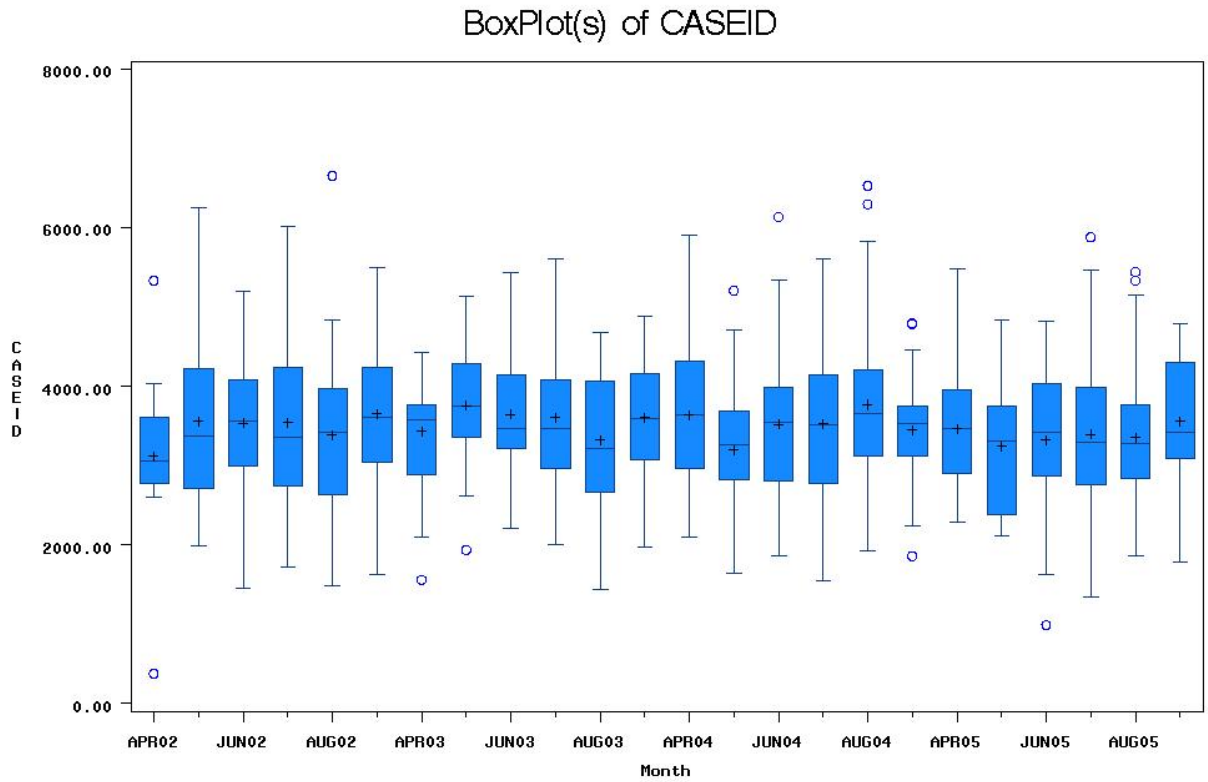

Histogram of ozone_ppb_max
Percent
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0
0
8
16
24
32
40
48
56
64
72
80
88
96
104
112
120
128
136
144
ozone_ppb_max

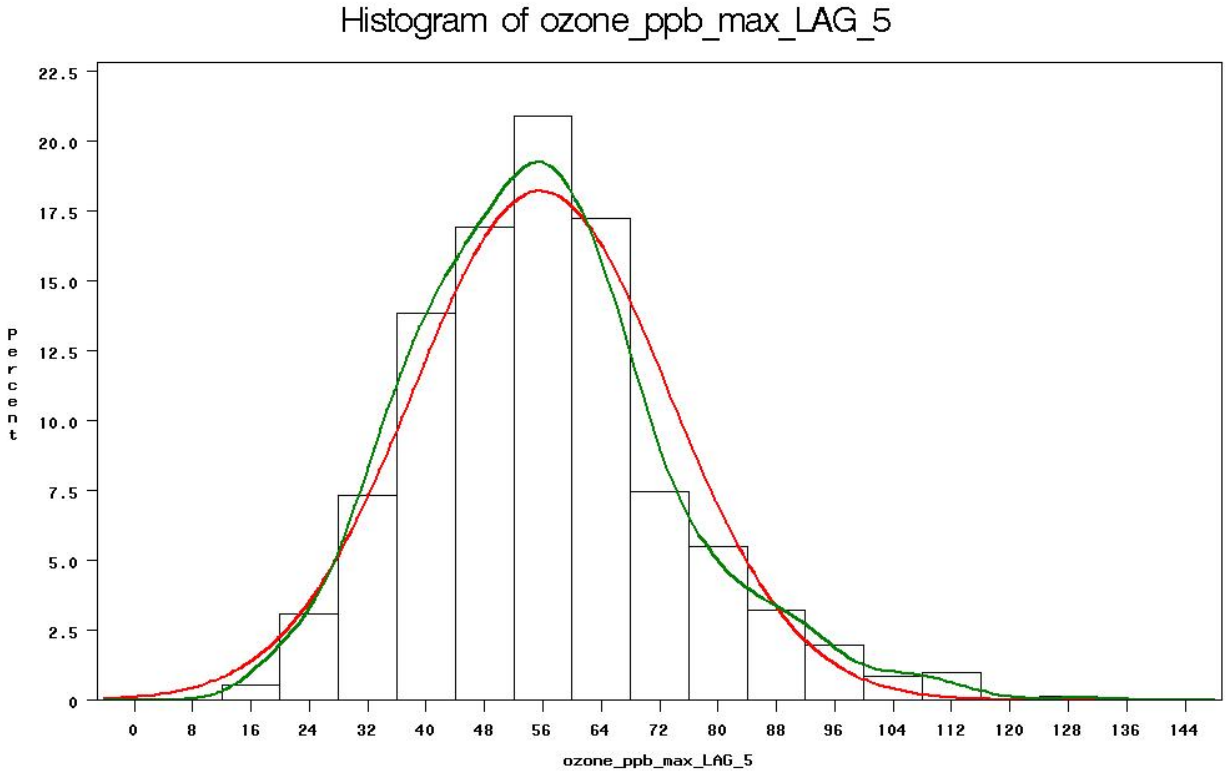

BoxPlot(s) of mean_pm25
mean_pm25
60
50
40
30
20
10
0
APR02
JUN02
AUG02
APR03
JUN03
AUG03
APR04
JUN04
AUG04
APR05
JUN05
AUG05
Month

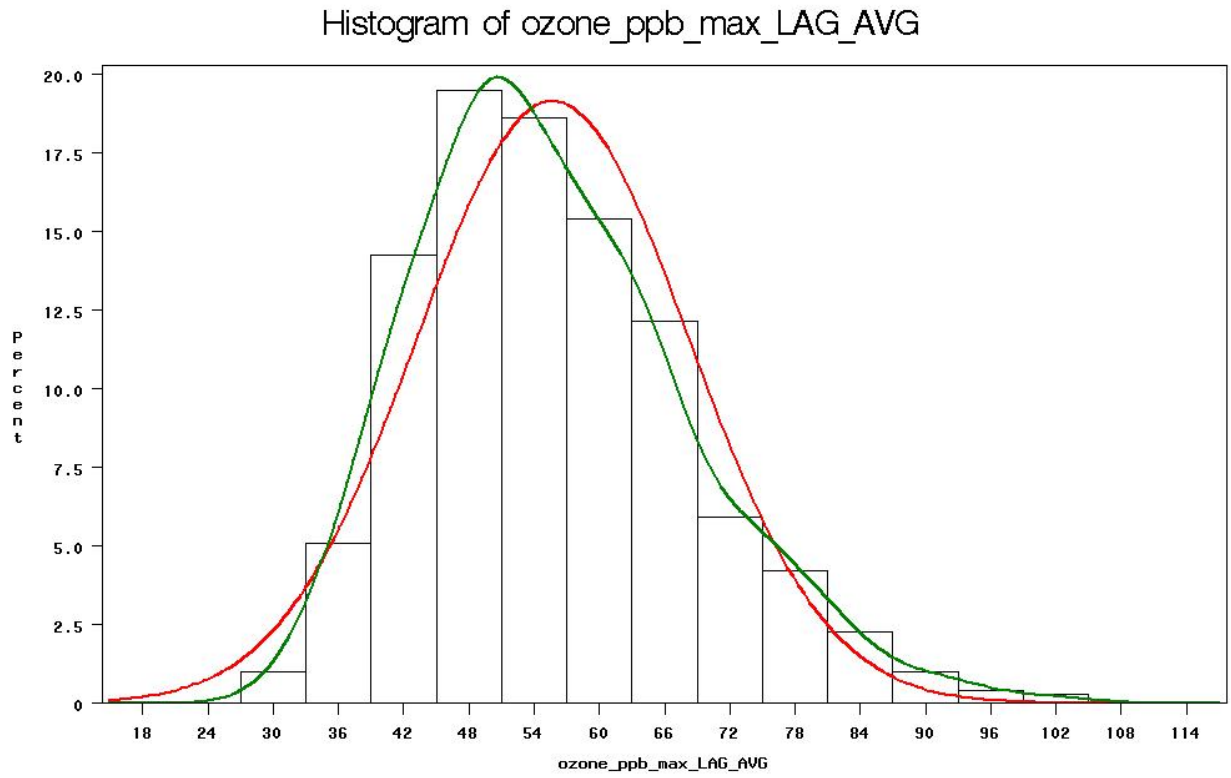

Histogram of mean_pm25
Percent
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0
-6
-2
2
6
10
14
18
22
26
30
34
38
42
46
50
54
58
62
mean_pm25

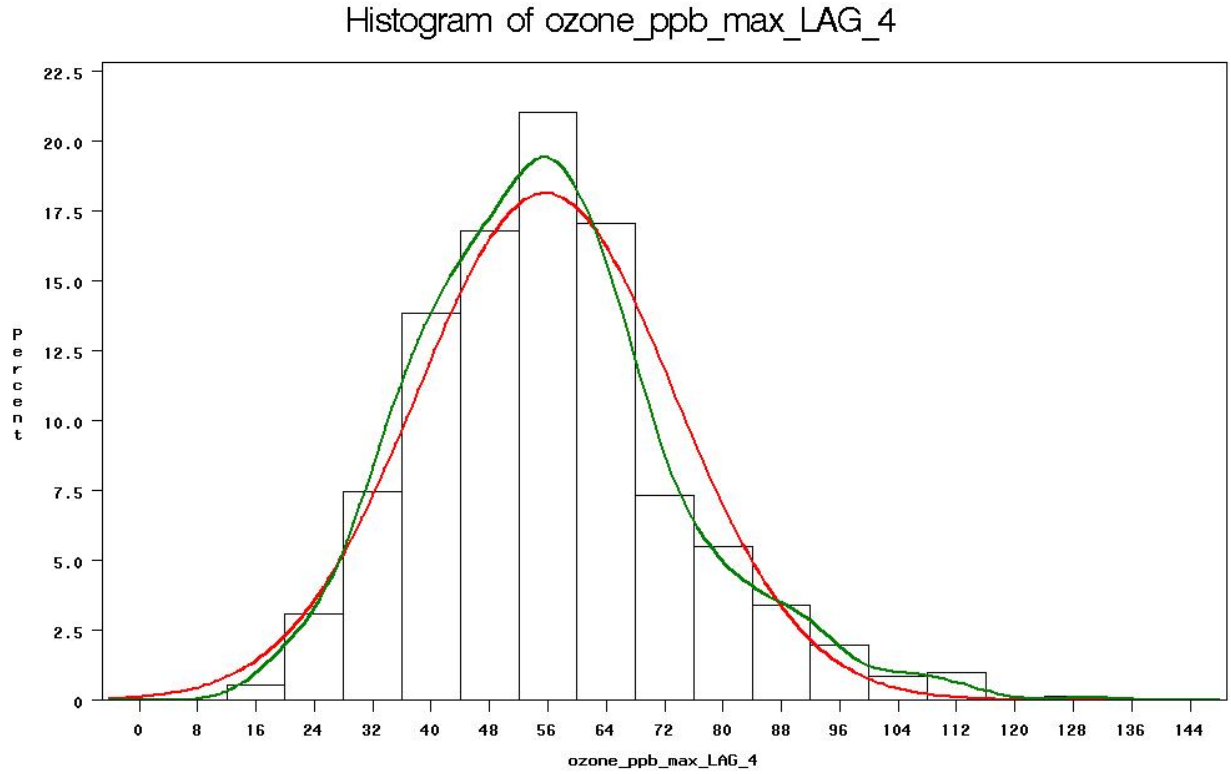

BoxPlot(s) of mean_pm25
mean_pm25
60
50
40
30
20
10
0
APR02
JUN02
AUG02
APR03
JUN03
AUG03
APR04
JUN04
AUG04
APR05
JUN05
AUG05
Month

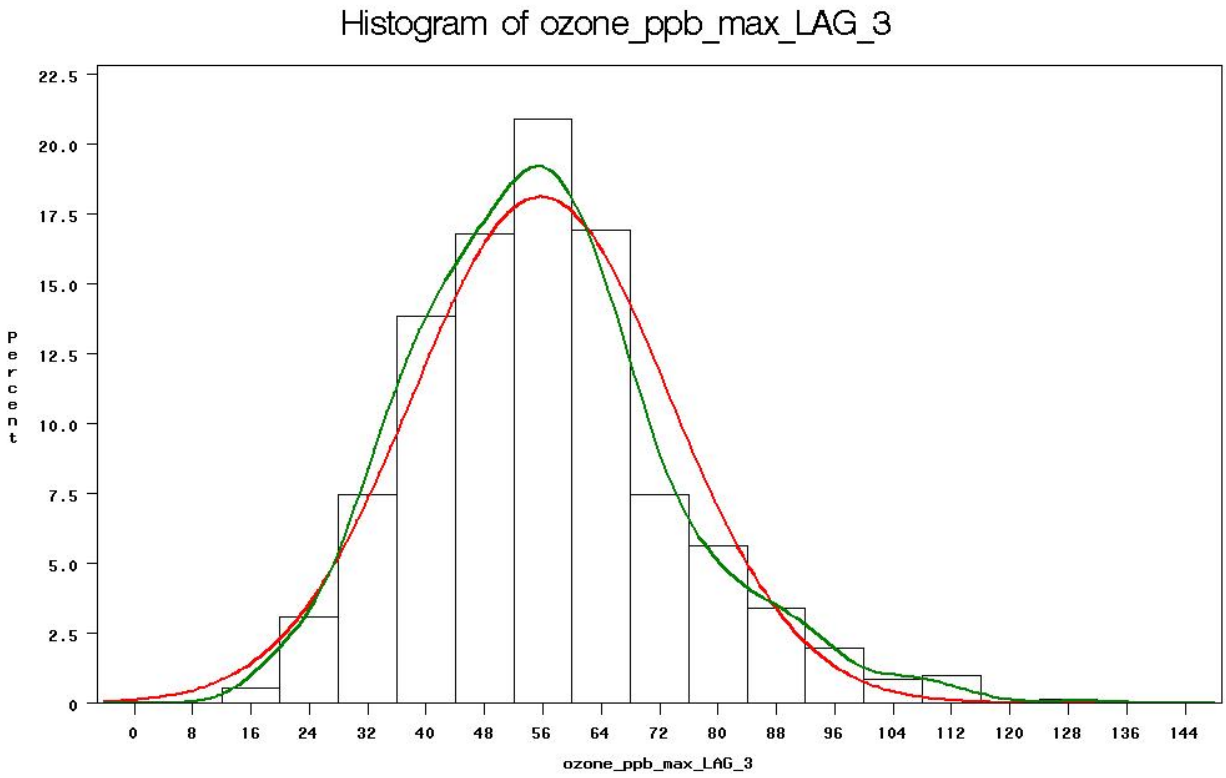

Histogram of temperature_mean
Percent
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0
26
30
34
38
42
46
50
54
58
62
66
70
74
78
82
86
90
temperature_mean

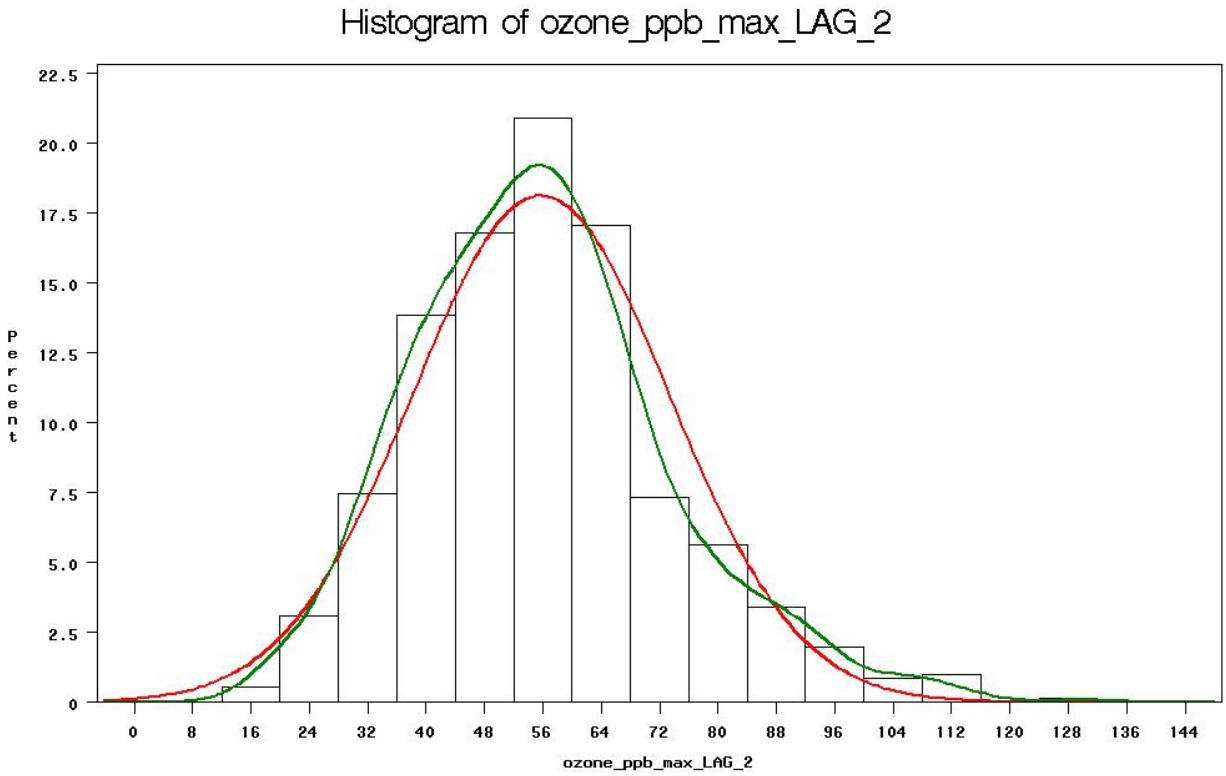

BoxPlot(s) of CASEID
8000.00
6000.00
4000.00
2000.00
0.00
CASEID
APR02
JUN02
AUG02
APR03
JUN03
AUG03
APR04
JUN04
AUG04
APR05
JUN05
AUG05
Month


Histogram of ozone_ppb_max_LAG_5
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0
Percent
0
8
16
24
32
40
48
56
64
72
80
88
96
104
112
120
128
136
144
ozone_ppb_max_LAG_5


Histogram of ozone_ppb_max_LAG_AVG
Percent
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0
18 24 30 36 42 48 54 60 66 72 78 84 90 96 102 108 114
ozone_ppb_max_LAG_AVG


Histogram of ozone_ppb_max_LAG_4
Percent
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0
0 8 16 24 32 40 48 56 64 72 80 88 96 104 112 120 128 136 144
ozone_ppb_max_LAG_4


Histogram of ozone_ppb_max_LAG_3
Percent
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0
0 8 16 24 32 40 48 56 64 72 80 88 96 104 112 120 128 136 144
ozone_ppb_max_LAG_3


Histogram of ozone_ppb_max_LAG_2
Percent
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0
0 8 16 24 32 40 48 56 64 72 80 88 96 104 112 120 128 136 144
ozone_ppb_max_LAG_2

APPENDIX E

AIR POLLUTION DATA: ALLEGHENY COUNTY HEALTH DEPARTMENT, 2008

Background information on ozone and PM 2.5 for Allegheny County (AC)

Air Pollutant data for Allegheny County: National Ambient Air Quality Standard (NAAQS)- Highest Concentration and Number of Exceedances 2006

Ozone	**NAAQS Standard**	**Highest Recorded Concentration**	**Number of NAAQS Exceedances**	**Monitoring Station**
1-hour average	0.12ppm	0.118ppm	0	3
8-hour average	0.08ppm	0.093ppm	7	3
PM 2.5				
24-hour average	65.0 ug/m^3	100.7 ug/m^3	3	12
Annual arithmetic mean	15.0 ug/m^3	19.0 ug/m^3	2	12

Number of Exceedances for 2003-2006

Ozone	**2003**	**2004**	**2005**	**2006**
1-hour Average	2	0	0	0
8-hour Average	11	1	11	7
PM-2.5				
24-hour average	9	7	10	3
Annual arithmetic mean	6	4	7	2

On May 27, 2008, the 8-hour standard was changed by EPA from 0.084 ppm to 0.075 ppm. Since that time, AC had 9 days that attainment was exceeded. Of those 9, only 3 would have exceeded the new standard.

Ozone Exceedance Days with level of exceedance above 8-hour standard (ppb)

	6/11/08	7/12/208	7/15/08	7/16/08	7/17/08	7/18/08	7/19/08	7/29/08	8/1/08
Harrison		76	76	91	83	82	88	86	81
Lawrenceville	76				76		79	84	
South Fayette									

APPENDIX F

COMPARISON OF RANDOM SAMPLE AND EMR TO PRIMARY ASTHMA DATASET BY DEMOGRAPHICS

Comparison of random sample and EMR to primary asthma dataset by demographics

	Age	Gender (M/F)	Race (B/W/O)
Primary (n=5,100)	M= 43.94 (SD 18.40)	31% 69%	35% 63% 20%
Random Sample (n=488)	M = 43.87 (SD 19.08)	31% 69%	40% 59% 1.0%
EMR (n=180)	M= 46.71 (SD 20.28)	30% 70%	43% 56% 1.0%

BIBLIOGRAPHY

Abraham, J., Bateson, T. (January 4, 2008). *Case Crossover Analysis Tool (C-CAT) Beta version 1.1 User's manual*. Copyright 2006. Health Research Inc. p 1-51.

American Academy of Professional Coders. (2008). ICD 10 Overview. Retrieved November 24, 2008 from, http://www.aapc.com/ICD-10/index.aspx.

American Association for Respiratory Care: AARC Practice Guidelines. Methacholine Challenge Testing (2001). Vol. 46 (5), 523-530.

American Lung Association. (2007). *Trends in Asthma Morbidity and Mortality*. American Lung Association. p. 1-32. Reprinted with permission ©2009 American Lung Association For more information about the American Lung Association or to support the work it does, call 1-800-LUNG-USA (1-800-586-4872) or log on to www.lungusa.org

American Lung Association. (2008*). State of the Air Report*. American Lung Association. p.1-204.

American Lung Association. Danger Zones. Air pollution and our children (2000). Retrieved November 20, 2008 from, http://mothersforcleanair.org/newsletter/2002-winter.html.

American Lung Association: Asthma Medications Chart. 2007. Retrieved July 31, 2008 from, http://www.lungusa.org/site/pp.aspx?c=dvLUK9O0E&b=263990.

Anderson, W., Prescott, G., Packham, S., Mullins, J., Brookes, M., & Seaton, A. (2001). Asthma admissions and thunderstorms: A study of pollen, fungal spores, rainfall and ozone. Quarterly Journal of Medicine. 94, 429-433.

Anto, J. (2004) The causes of asthma: Need to look at the data with different eyes. Allergy. 59,112-123.

ATLAS.ti-The Knowledge Workbench: Feature. Retrieved August 12, 2008 from, http://www.atlasti.com/features.html.

Baker, L.& Afendulis, C. (2005). Medicaid managed care and health care for children. Health Service Research. 40(5 Pt. 1), 1464-1488.

Balmes, J. (1993). The role of ozone exposure in the epidemiology of asthma. Environmental Health Perspective Supplements. 101(Suppl. 4), 219-224.

Barnes, P., Jonsson, B., & Kilm, J. (1996). The cost of asthma. European Respiratory Journal. 9, 636-642.

Barnett, A., Williams, G., Schwartz, J., Nellar, H., Best, T., Petroeschevsky, L., & Simpson, R. (2005). Air pollution and child respiratory health: A case crossover study in Australia and New Zealand. American Journal of Respiratory and Critical Care Medicine. 171, 1272-1278.

Barry, H. (1991). Reactive oxygen species in living systems: A thorough explanation of free radicals. American Journal of Medicine. V91 n3C, p14S(9).

Barthell, E., Cordell, W., Moorhead, J., Handler, J., Feied, C., Smith, M., Cochrane, D., Felton, C., Collins, M. (2002). The frontlines of medicine project: A proposal for the standardized communication of emergency department data for public health uses including syndromic surveillance for biological and chemical terrorism. Annals Emergency Medicine. 39(4), 422-429.

Beasley, R. (2002). The burden of asthma with specific reference to the United States. Journal Allergy Clinical Immunology. 100(5), 482-489.

Bell, M., Peng, R., Dominici, F. (2006). The exposure-response curve for ozone and risk of mortality and the adequacy of current ozone regulations. Environmental Health Perspectives. 114. Retrieved August 14, 2008 from, http://opa,yale.edu/news/article.aspx?id=2175.

Bell, M., Dominici, F., & Samet, J. (2005). A meta-analysis of time series studies of ozone and mortality with comparison to the National Morbidity, Mortality, and Air Pollution Study. Epidemiology. 16(4), 436-445.

Bell, M., Kim, J., & Dominici, F. (2007). Potential confounding of particulate matter on the short-term association between ozone and mortality in multisite time series studies. Environmental Health Perspectives. 15(11),1591-1595.

Bell, M., McDermott, A., Zeger, S., Dominici, F. (2004). Ozone and short-term mortality in 95 US Urban Communities, 1987-2000. JAMA. 292(19), 2372-2378.

Bhattacharya, K., Alink, G., Dopp, E. (2007). Oxidative stress and changed gene expression profiles in fiber-/particle-induced carcinogenesis. International Journal of Human Genetics. 7, 1-21.

Bora-Aburti, V., Castillejos, M., Gold, D., Bierzwinski, S., Loomis, D. (1998). Mortality and ambient fine particles in Southwest Mexico City, 1993-1995. Environmental Health Perspectives. 106(12), 849-855.

Boss, L., Kreutzer, R., Luttinger, D., Leighton, J., Wilcox, K., & Redd, S. (2001). The public health surveillance of asthma. Journal of Asthma. 38(1), 83-89.

Braman, S. (2006). The global burden of asthma. Chest. 130, 4S-12S.

Breeding, M. (2002). Expanding the systems librarian's toolkit: PERL. Retrieved March 4, 2008 from, http://www.librarytechnology.org/ltg-displaytext.pl?RC=9464.

Buist, A., & Vollmer, WM. (1990). Reflections on the rise in asthma morbidity and mortality. JAMA. 264, 1719-1720.

Burnett, R., Smith-Doiron, M., Stieb, D., Raizenne, M., Brook, J., Dales, R., Leech, J., Cakmak, S., & Krewski, D. (2001). Association between ozone and hospitalizations for acute respiratory diseases in children less than 2 years of age. Am J Epidemiology. 153(5), 444-452.

Camargo, C., Ramachandran, S., Ryskina, K., Lewis, B., Legorreta, A. (2007). Association between common asthma therapies and recurrent asthma exacerbations in children enrolled in a state Medicaid Plan. American Journal Health-System Pharmacy. 64, 1054-1061.

Centers for Disease Control and Prevention, National Centers for Environmental Health Tracking Program. (August 2006). *National network implementation plan* (NNIP). pp. 1-65.

Centers for Disease Control and Prevention. Asthma Data and Surveillance. Asthma Prevalence (1980-2004). Retrieved November 2, 2007 from, http://www.cdc.gov/asthma/asthmadata.htm.

Centers for Disease Control and Prevention. Behavioral Risk Factor Surveillance System. Selected Metropolitian/Micropolitian Area Risk Trends [SMART] 2002-2005. Retrieved November 21, 2007 from, http://apps.nccd.cdc.gov/brfss-smart/CompareMMSARiskchart.asp?MMSA=69&yr2=20.

Centers for Disease Control and Prevention. National Public Health Performance Standards. Ten Essential Services. Retrieved March 6, 2008 from, http://www.cdc.gov/od/ocphp/nphpsp/essentialphservices.htm.

Centers for Disease Control National Center for Health Statistics. (2005). Asthma surveillance data. National Health Interview Survey.

Chen, E., Hripcsak, G., Xu, H., Markatou, M, & Friedman, C. (2008) Automated acquisition of disease-drug knowledge from biomedical and clinical documents: An initial study. Journal of American Informatics Association. 15(1), 87-98.

Chen, T., Gokhale, J., Shoffer, S., Kuschner, W. (2007) Outdoor air pollution: Ozone health effects. American Journal of Medical Science. 333(4), 244-248.

Chilmonczyk, B., Salmun, L., Megathlin, K., Neveux, L., Palomaki, G., Knight, G., Pulkkinen, A., & Haddow, J. (June 10, 1993). Association between exposure to environmental tobacco smoke and exacerbations of asthma in children. New England Journal of Medicine. 328, 1665-1669.

Chuang, J., Friedman, C. & Hripcsak, G. (2002). A comparison of the Charlson comorbidities derived from medical language processing and administrative data. Proc AMIA Symposium. pp. 160-164.

Clean Air Act (1970). Background section on air pollution. Section 1. Retrieved December 20, 2007 from, http://www.nsc.org/EHC/mobile/acback.htm.

Coiera, E. Guide to Health Informatics (2nd Edition.). 2003. Oxford University Press.

College of Physicians, Surgeons of Alberta. Spirometry & Flow Volume Measurements: Standards and Guidelines. December 1998. Retrieved November 15, 2008 from, http://www.canahome.org/files/spirometry_guidelines.pdf

Committee on Advances in Assessing Human Exposure to Airborne Pollutants: Human exposure assessment for airborne pollutants: Advances and Opportunities, 1991. p. 37-52.

Cooper, G., Buchman, B., Kayaalp, M., Saul, M. & Vries, J. (1998). Using computer modeling to help identify subgroups in clinical data repositories. Proceedings American Medical Informatics Association Symposium. p. 180-184.

Council and State Territorial Epidemiologist. Asthma Surveillance and Case Definition. CSTE Position Statement 1998-EH. Retrieved May 4, 2006 from, http://www.cste.org/ps/1998/1998-eh-cd-01.htm.

Crain, E., Walter, M., O'Conner, G., Mitchell, R., Gruchalla, R., Kattan, M., Malindzak, G., Enright, P., Evans, R., Morgan, W., & Stout, J. (2002). Home and allergic characteristics of children with asthma in seven U.S. urban communities and design of an environmental intervention: The Inner City Asthma Study. Environmental Health Perspectives 110, 939-945.

Dales, R., Schweitzer, J., Toogood, J., Drouin, M, Yang, W., Dolovich, J. & Boulet, J. (1996). Respiratory infections and the autumn increase in asthma morbidity. European Respiratory Journal (1996). 9, 72-77.

Department of Health and Human Services. (May 2000). Action against asthma. A strategic plan for the Department of Health and Human Services (DHHS publication). Retrieved November 2, 2007 from, http://aspe.hhs.gov/sp/asthma/.

Dockery, D. (1993). Epidemiologic study design for investigating respiratory health effects of complex air pollution mixtures. Environmental Health Perspectives Vol. 101 (Suppl. 4.), 187-191.

Dombkowski, K., Wasilevich, E. & Lyon-Callo, S. (2005). Pediatric asthma surveillance using Medicaid claims. Public Health Report. 120, 515-524.

Donahue, J, Weiss, S., Goetsch, M, Livingston, J., Greinder, D. & Platt, R. (1997). Assessment of asthma using automated and full text medical records. Journal of Asthma. 34(4), 273-281.

Donaldson, K., Mills, N., MacNee., Robinson, S., Blanc, P. (2001). Role of inflammation in cardiopulmonary health effects of PM. Toxicology and Applied Pharmacology. Vol. 27(2), (Suppl. 1), p. 483-488.

DuPont, W., & Plummer, W. (2004). PS: Power and Sample Size. Retrieved March 6, 2008 from, http://biostat.mc.vanderbilt.edu/twiki/bin/view/Main/PowerSampleSize.

Eisner, M, Katz, P., Yellin, E., Shiboski, S. & Blanc, P. (2001) Risk factors for hospitalization among adults with asthma: The influence of socio demographic factors and asthma severity. Respiratory Research. 2, 53-60.

Environmental Protection Agency. (2000) Indoor Air Quality. Retrieved March 2, 2008 from, http://www.epa.gov/iaq/ia-intro-html.

Espino, J., & Wagner, M. (2001). Accuracy of ICD-9-coded chief complaints and diagnoses for the detection of acute respiratory illness. Proceedings American Medical Informatics Association, Fall Symposium. Retrieved November 2, 2007 from, http://rods.health.pitt.edu/LIBRARY/amia2001_final_reviedEspino.pdf.

Etzel, R. (2003). How environmental exposures influence the development and exacerbation of asthma. Pediatrics. 112(1), 233-239.

Ferris, T., Blumenthal , D., Woodruff, P., Clark, S., & Camargo, C. (2002). Insurance and quality of care for adults with acute asthma. Journal of General Internal Medicine. Vol. 17(12), 905-913.

Fleischauer, A., Silk, B., Schumacher, M., Komatsu, K., Santana, S., Vaz, V., Wolfe, M., Hutwagner, L., Cono, J., Berkelman, R., & Treadwell, T. The validity of chief complaint and discharge in emergency department –based syndromic surveillance. Academy Emergency Medicine. 11(12), 1262-1267.

Friedman, C., Shagina, L., Lussier, Y., Hripcsak, G. (2004). Automatic encoding of clinical documents based on natural language processing. Journal American Medical Informatics Association 11, 392-402.

Fung, K., Krewski, D., Chen, Y., Burnett, R., & Cakmak, S. (2003). Comparison of time series and case-crossover analyses of air pollution and hospital admission data. International Journal of Epidemiology. 32, 1064-1070.

Glad, J., Kotchian, S., & Barron, G. (2004). Developing a local comprehensive environmental and health tracking system: Using what we know to improve health and the environment. Journal of Environmental Health, pp. 9-14.

Gostin, L. & Hodge, J. (1998). The historical development of reporting as a public health practice [The "Names Debate:" the case for national HIV reporting in the United States, 61 Albany Law Review 679-743, 689-698]. Retrieved June 1, 2008 from, http://academic.udayton.edu/health/syllabi/Bioterrorism/5DiseaseReport/PHLaw11.htm.

Haley, V., Bateson, T., Boothe, V., Talbot, T., Dimmick, F., Paulu, C. & Warner, M. (2006). *Case crossover analysis of acute health effects and ambient air quality for environmental public health tracking.* (Draft) Technical Background Report. pp. 1-62.

Hripcsak, G., Austin, J., Alderson, P. & Freidman, C. (2002). Use of natural language processing to translate clinical information from a database of 889,921 chest radiographic reports. Radiology. 224, 157-163.

Hung, J., Posey, J., Freedman, R., Thorton, T. (1998) Electronic surveillance of disease states: A preliminary study in electronic detection of respiratory diseases in a primary care setting. Proceedings American Medical Informatics Association Annual Symposium, 688-92.

Institute of Medicine, Committee on the Assessment of Asthma and Indoor Air. (2002). *Clearing the Air. Asthma and Indoor Air Exposures.* National Academy Press: Washington DC.

Institute of Medicine, *The computer based patient report: An essential technology for health care* (1997). National Academy Press: Washington DC. Retrieved May 22, 2008 from, http://www.openclinical.org/emr.html.

Institute of Medicine. (1988). *The future of public health*. National Academy Press: Washington DC.

Ito, K., De Leon, S., Lippman, M. (2005). Associations between ozone and daily mortality: Analysis and meta-analysis. Epidemiology. 16:14;446-457.

Jaakkola, J. (2003). Case crossover design in air pollution epidemiology. European Respiratory J. 21(Suppl. 40), 81s-85.

Janes, H, Sheppard L., Lumley, T. (2004). Overlap bias in case-crossover design, with application to air pollution exposures. Statistics in Medicine. Vol. 24(2), 285-300.

Janes, H., Sheppard, Lumley, T. (2005). Case crossover analyses of air pollution exposure data: referent selection strategies and their implication for bias. Epidemiology. 16, 717-726.

Johnston, N, & Sears, M. (2006). Asthma exacerbation: Epidemiology. Thorax**.** 61, 722 - 728.

Kelley, C., Mannino, D., Homa, D., Savage-Brown, A., & Holguin, F. (2005). Asthma phenotypes: Risk factors, and measures of severity in a national sample of children. Pediatrics. 115(3), 726-731.

Krauthammer, M. & Hripcsak, G. (2001). A knowledge based model for the interpretation and visualization of NLP-parsed discharge summaries. Proceedings American Medical Informatics Association Symposium, pp. 339-343.

Kuschner, W. & Stark, P. (2003). Identifying work related asthma and other disorders. Postgraduate Medicine: Occupational Lung Disease. Vol. 113(4), 1-9.

Landon, B., Zaslavsky, A., Bernard, S., Cioffi, M., Clearly, P. (2004). Comparison of performance of traditional Medicare vs. Medicare managed care. JAMA. 291, 1744-1752.

Lanphear, B. & Gergen, P. (2003). Invited Commentary: Asthma surveillance in US children. American. Journal of Epidemiology. 158, 105-107.

Lazarus, R., Kleinman, K, Dashevsky, I., Demaria, A., & Platt, R. (2001). Using automated medical records for rapid identification of illness syndromes (syndromic surveillance): The example of lower respiratory infection. BMC Public Health. Retrieved November 2, 2007 from, http://www.biomedcentral/1471-2458/1/9.

Li, N., Sioutas, C., Cho, A., Schmitz, D., Misra, C., Sempf, J., Wang, M., Oberley, T., Froines, J., Nel, A. (2003). Ultra fine particulate pollutants induce oxidative stress and mitochondrial damage. Environmental Health Perspectives. Vol. 111, 14:455-460.

Lichter, E. (October 2004). *Using Medicaid data to track asthma in New England*. Asthma Regional Council of New England. pp. 1-31.

Lin, M., Chen, Y., Burnett, R., Villeneuve, P., & Krewski, D. (2002). The influence of ambient coarse particulate matter on asthma hospitalization in children: Case crossover and time series analysis. Environmental Health Perspectives. 110(6), 575-581.2

Lombardo, J. & Ross, D. (2007). Disease Surveillance, a public health priority. In J. Lombardo & D. Buckeridge, Disease Surveillance a public health informatics approach. (pp 1-39). John Wiley and Sons.

Lombardo, J., Burkom, H., Elbert, E., Magruder, S., Lewis, S., Loschen, W., Sari, J., Sniegoski, Wojcik, R., & Pavlin, J. (2003). A systems overview of the electronic surveillance system for the early notification of community based epidemics (ESSENCE II). Journal of Urban Health: Bulletin of the New York Academy of Medicine. 80(2), 32-42.

Lou, X. & Sorock, G (2007). Analysis of recurrent event data under the case crossover design with applications to elderly falls. Statistics in Medicine. 27, 2890-2901.

Lu, Y., Symons, J., Geyh., A., & Zeger, S. (2008). An approach to checking case crossover analyses based on equivalence with time series methods. Epidemiology. Vol. 19(2), 169-175.

Luginaah, I., Fung, K., Gorey, K., Webster, G., Wills, C. (2005). Associations of ambient air pollution with respiratory hospitalization in a government-designated “area of concern”: The case of Windsor Ontario. Environmental Health Perspectives. 113(3), 290-296.

Maclure, M., & Mittleman, M. (2008). Case crossover designs compared with dynamic follow up designs. Epidemiology. 19(2), 176-178.

Mandhane, P., Greene, J., Cowan, J., Taylor, R, & Sears, M. (2005). Sex differences in factors associated with childhood –and adolescent-onset wheeze. American Journal of Respiratory and Critical Care Medicine. 172, 45-54.

Manning, C., & Schutze. Foundations of Statistical Natural Language Processing. MIT Press:Cambridge, MA. May 1999.

Marks, G., Colquhoun, J., Girgis, S. Koski, M., Treloar, A., Hansen, P., Downs, S., & Car, N. (2006). Thunderstorm outflow preceding epidemics of asthma during spring and summer. Thorax. 56, 468-471.

Marks, GB. (2005). identifying asthma in population studies: from a single entity to a multi-component approach. European Respiratory Journal. 26, 3-5.

May, K. (1996). Atopic asthma and the immune system. Oral paper. Retrieved October 2, 2007 from, http://www.bio.davidson.edu/Biology/student/restricted/immunofinal.html.

Melton, G., Hripcsak, G. (2005). Automated detection of adverse events using natural language processing of discharge summaries. Journal of American Medical Informatics Association. Vol. 12(4), 448-457.

Merck Medicus Modules: Asthma Diagnosis, 2001. Retrieved October 2, 2007 from, http://www.merckmedicus.com/pp/us/hcp/diseasemodules/asthma/diagnosis.jsp.

Merck Medicus Modules: Asthma-Definition, 2001. Retrieved October 2, 2007 from, http://www.merckmedicus.com/pp/us/hcp/diseasemodules/asthma/default.jsp.

Merck Medicus Modules: Asthma-Epidemiology, 2001. Retrieved October 2, 2007 from, http://www.merckmedicus.com/pp/us/hcp/diseasemodules/asthma/epidemiology.jsp.

Merck Medicus Modules: Asthma-Pathophysiology, 2001. Retrieved October 2, 2007 from, http://www.merckmedicus.com/pp/us/hcp/diseasemodules/asthma/pathophysiology.jsp.

Mortality and Morbidity Weekly Report. Surveillance for Asthma-United States, 1960-1995. (April 24, 1998). 47(Suppl. 1), 1-28.

Mortality Morbidity Weekly Report. National Surveillance for Asthma-United States. 1980-2004. (October 19, 2007). 56(SS08), 1-14:18-54.

Mortality Morbidity Weekly Report. Surveillance for Asthma. 1980-1999. (March 29, 2002) 51(SS01), 1-13.

Mortality Morbidity Weekly Report. Syndromic Surveillance. September 24, 2004. Vol. 53. (Suppl), 1-268.

Muhr, T., (2004). *User's Manual for ATLAS.ti 5.0* (2nd Edition). Copyright © 2003-2004 by Thomas Muhr, Scientific Software Development, Berlin.

National Center for Health Statistics. Asthma Prevalence, Health Care Use and Mortality: United States, 2003-2005. National Health Interview Survey.

National Center for Health Statistics. Asthma Prevalence, Health Care Use and Mortality: United States, 2002. National Health Interview Survey, 2002.

National Center for Health Statistics. Changes to the National Health Interview Survey (1997), National Heart Lung and Blood Institute. Retrieved August 12, 2008 from, http://www.nhlbi.nih.gov/health/prof/lung/asthma/surveil.htm.

National Center for Health Statistics: NHIS. Early releases of selected estimates based on data from January-September 2007. Asthma episodes and current asthma.

National Heart, Lung and Blood Institute Chart book on cardiovascular, lung, and blood diseases, U.S. Department of Health and Human Services, National Institute of Health 2007. Unpublished data provided upon request to the NHLBI.

National Heart, Lung, and Blood Institute (NHLBI). National asthma education and prevention program. Expert panel report 2. Guidelines for the Diagnosis and Management of Asthma. Bethesda, MD: NIH, July 1997. NIH Pub. No. 97-4051.

Nemar, A., Al-Maskaria, S., Ali, B., Al-Amri, I (2006). Cardiovascular and lung inflammation effects induced by systemically administered diesel exhaust particles in rats. American Journal of Physiology Lung Cell Molecular Physiology. 292, L664-L670.

O'Donnell, R.,& Frew, A. (2002). Is there more than one inflammatory phenotype in asthma? Thorax. 57, 566-568.

O'Malley, K., Cook, K., Price, M., Raiford-Wildes, K., Hurdle, J & Ashton, C. (2005). Measuring Diagnoses: ICD Code Accuracy. Health Services Research. 40(5 Pt. 2), 1620-1639.

Panackal, A. M'ikanatha, N, Tsui, FC, McMahon, J., Wagner, M., Dixon, B., Zubieta, J., Phelan, M., Mirza, S., Morgan, J., Jernigan, D., Pasculle, AW, Rankin, J., Hajjeh, R., & Harrison, L. (2002). Automatic Electronic Laboratory-Based Reporting of Notifiable Infectious Diseases at a Large Health System. Emerging Infectious Diseases. 8(7), 685-691.

Paulu, C. & Smith, A. (2008). Tracking associations between ambient ozone and asthma-related emergency department visits using case crossover analysis. Journal of Public Health Management Practice. 14(6), 581-591.

Pearce, N., Douwes, J., & Beasley, R. (2000). Is allergen exposure the major primary cause of asthma. Thorax. 55, 424-431.

Peat , J., Toelle, B., Marks, G., & Mellis, C. (2001) Continuing the debate about measuring asthma in population studies. Thorax. 56, 406-411.

Pekkanen, J, & Pearce, N. (1999). Defining asthma in epidemiological studies. European Respiratory Journal. 14, 951-957.

Pekkanen, J., Sunyer, J., Anto, J., & Burney, P. (2005). Operational definitions of asthma in studies on its aetiology. European Respiratory Journal. 26, 28-35.

Pennsylvania Health Care Cost Containment Council. Pennsylvania Inpatient Data. Retrieved December 15, 2007 from, http://www.phc4.org/.

Pennsylvania National Electronic Disease Surveillance System [PA-NEDSS]. The PA Bulletin. Electronic Surveillance. [33 Pa. B. 2439]. Retrieved August 27, 2008 from, http://www.pabullentin.com/secure/data/vol33/33-20/941.htm

Pew Environmental Health Commission (2000). *Attack asthma: Why America needs a public health defense system to battle environmental health threats*. Baltimore, Maryland. Pew Foundation.

Pope, CA. (1989). Respiratory disease associated with community air pollution and a steel mill, Utah Valley. American Journal of Public Health. 79(5), 623-628.

Rahman, I., Morrison, D., Donaldson, K., MacNee, W. (1996). Systemic oxidative stress in asthma, COPD, and smokers. American Journal of Respiratory Critical Care Medicine. Vol. 154(4), 1055-1060.

Risom, L., Moller, P., Loft, S.(2005). Oxidative stress-induced DNA damage by particulate air pollution. Mutation Research. 592(1-2), 119-37.

Rosentreich, D., Eggleston, P., Kattan, M., & Baker, D. (May 8, 1997). The role of cockroach allergy and exposure to cockroach allergens in causing morbidity among inner-city children with asthma. New England Journal of Medicine. 336, 1356-1363.

Sanders, D., Gregg, W., Aronsky, D. (2007). Identifying asthma exacerbations in a pediatric emergency department: A feasibility study. International Journal of Medical Informatics. 76, 557-564.

Schatz, M., Clark, S., & Camargo, C. (2006). Sex differences in the presentation and course of asthma hospitalizations. Chest. 129, 50-55.

Schuchter, J. (2003). A new set of EARS: Surveillance for bioterrorism and beyond. Public Health Watch. Vol. 7(2), 1-6.

Schwartz, R., & Phoenix, T. Learning PERL (2nd Edition). 2001.

Sebastiani, F. (2002). Machine learning in text categorization. ACM Computing Surveys. Vol. 34(1). pp. 1-47.

Singh, A., Cydulka, R., Stahmer, S., Woodruff, P., & Carmargo, C. (1999). Sex differences among adults presenting to the emergency department with acute asthma. Archives Internal Medicine. 159:1237-1243.

Sistek, D., Tschopp, J., Schindler, C., Brutsche, M., Liebrich,U., Perruchoud, A., & Leuenberger, P. (2001). European Respiratory Journal. 17, 214-219.

Sistek, D., Wickens, K., Armstrong, R., D'Souza, W., Town, I. Crane, J. (2006). Respiratory Medicine. 100, 2107-2111.

Stick, S., Burton P., Gurrin, L., Sly, P., & LeSouef, P. (1996). Effects of maternal smoking during pregnancy and a family history of asthma on respiratory function in newborn infants. Lancet 348(9034), 1060-1064.

Strek, M. (2006). Difficult asthma. The Proceedings of the American Thoracic Society. 3, 116-123.

The Merck Manual of Diagnosis and Treatment. 17th ed. Whitehouse Station, NJ: Merck Research Laboratories; 1999. Retrieved October 3, 2007 from, http://www.merckmedicus.com/pp/us/hcp/diseasemodules/asthma/default.jsp.

U.S. Department of Health & Human Services (May 2000). *Action Against Asthma: A Strategic Plan for the Department of Health and Human Services.* Department of Health and Human Services. Retrieved October 20, 2007 from, http://aspe.hhs.gov/sp/asthma/overview.htm.

U.S. Department of Health and Human Services (USDHHS). National Heart, Lung and Blood Institute Chart book on cardiovascular, lung and blood diseases. National Institute of Health, 2007. Unpublished data.

U.S. National Library of Medicine: National Institute of Health. Semantic Network fact sheet. Retrieved March 20, 2008 from, http://www.nlm.nih.gov/pubs/factsheets/um;ssemn.html.

U.S. News & World Report: Allergy & Asthma (November, 2007). How does asthma in children differ from asthma in adults? Allergy and Asthma. Retrieved November 14, 2008 from, http://health.usnews.com/articles/health/allergy-and-asthma/2007/11/30/asthma-in-children.html.

US Census Data, 2000. Pennsylvania State & County Quick Facts. Retrieved February 21, 2008 from, http://quickfacts.census.gov/qfd/states/42/42003.html

US Census Data, 2006. Pennsylvania State & County Quick Facts. Retrieved November 24, 2008 from, http://quickfacts.census.gov/qfd/states/42/42003.html

Villeneuve, P., Chen, L., Rowe, B., & Coates, F. (2007). Outdoor air pollution and emergency department visits for asthma among children and adults. A case crossover study in Northern Alberta, Canada. Environmental Health. 6:40, 1-15.

Visweswaran, S., Hanbury, P., Saul, M., & Cooper, G. (2003). Detecting adverse drug events in discharge summaries using variations on the Bayes model. Proceedings American Medical Informatics Association Annual Symposium, 689-693.

Vollmer, W., O'Conner, E., Heumann, M., Frazier, A., Breen, V., Villnave, J., & Buist, S. (2004). Searching multiple clinical information systems for longer periods found more prevalent cases of asthma. Journal of Clinical Epidemiology. 57, 392-397.

Volmer T. (2001). The socio-economic of asthma. Pulmonary Pharmacology & Therapeutics. 14, 55-60.

von Mutius, E. (2000). The environmental predictors of allergic disease. Journal Allergy Clinical Immunology. 105(1 Pt.1), 9-19.

Wansoo, I, & Schneider, D. (2005). Effect of weed pollen on children's hospital admissions for asthma during the fall season. Archives of Environmental & Occupational Health. 60(5), 1-8.

Weiss, K, Gergen, P., Hodgson, T. (1992.) An economic evaluation of asthma in the United States. New England Journal of Medicine, 326(13), 862-866.

Wissow, L., Gittelsohn, A., Szklo, M., Starfield, B. & Mussman, M. (1998). Poverty, race, and hospitalization for childhood asthma. American Journal of Public Health. 78, 777-782.

World Health Organization [WHO]; International Classification for Disease [ICD]. Ninth Revision, Clinical Modification.

Yu, I., Wong, T., & Li. W. (2004). Using child reported respiratory symptoms to diagnose asthma in the community. Archives of Disease in Childhood. 89, 544-548.

Zbrowski, J., Bilonick, R., & Talbott, E. Projects/PITT-M/Final_Report/CONSOL/FinalReport1_ver65.odt. January 1, 2007. p. 1-183.