A COMPARISON IN A SASKATCHEWAN POPULATION BETWEEN CLINICAL DIAGNOSIS OF ASTHMA AND AMERICAN THORACIC SOCIETY FORCED EXPIRATORY VOLUME AT ONE SECOND (FEV1) POST-BRONCHODILATOR IMPROVEMENT CRITERIA

J.A. GJEVRE

A comparison in a Saskatchewan population between clinical diagnosis of asthma and American Thoracic Society Forced Expiratory Volume at one second (FEV1) post-bronchodilator improvement criteria

by

© J.A.Gjevre Jr. B.A., MD, LMCC, FRCP(C) Internal Medicine and Respirology, ABIM Internal Medicine and Pulmonary Disease, FCCP

A thesis submitted to the school of graduate studies in partial fulfillment of the requirements for the degree of Masters of Science

Department of Medicine
Division of Clinical Epidemiology

July 2008

St. John's Newfoundland and Labrador

Abstract:

Background: The diagnosis of asthma can be challenging and is based on clinical symptoms, physical examination, and pulmonary function tests. Most patients with asthma will have a significant post-bronchodilator response on spirometry indicating airway hyper-responsiveness. However, having a significant bronchodilator response by itself is not diagnostic of asthma. Also, the definition of a "significant" response is controversial. Many respirologists use the American Thoracic Society (ATS) post-bronchodilator response criteria of 12% (provided it is $\geq$ 200 ml) improvement in FEV1 (or FCV) from the baseline spirometry.

Methods: This study retrospectively reviewed 644 patients who met the ATS criteria for a significant post-bronchodilator spirometric response. The staff respirologist's diagnosis of asthma, based on all clinical and pulmonary function data, was used as the standard for the diagnosis of asthma.

Results: It was found that relying on spirometric criteria alone was inadequate in asthma diagnosis as only 54.7% of 310 patients meeting ATS bronchodilator response criteria were felt to have asthma clinically. Increasing the post-bronchodilator percent improvement from the ATS criteria only marginally improved diagnostic specificity and resulted in a decline in sensitivity.

Conclusions: This further emphasizes the need to use spirometric criteria as a guide but not as an unimpeachable gold standard by which to make a diagnosis of asthma. The diagnosis of asthma depends on a combination of expert physician correlation of history, physical examination, and pulmonary function test results.

Acknowledgements:

I should like to recognize the mentorship of Dr. Don Cockcroft and Dr. Tom Hurst in this research. Thanks are further extended to the members of my thesis committee for their support: Dr. George Fox, Dr. John Harnett, Dr. Mark Borgaonkar, and Dr. Roy West.

Table of Contents: **page no.**

Tables 5

Figures 6

Abbreviations 7

Chapter 1: Introduction 8
1.1: Background 8
1.2: History of asthma 8
1.3: Epidemiology of asthma: General Overview 9
1.4: Epidemiology of asthma: Major Risk Factors 11
1.5: Epidemiology of asthma: Economic Costs 14
1.6: Pathophysiology of asthma 14
1.7: Diagnosis of asthma 15
1.8: Summary 19
1.9: Research Question 21

Chapter 2: Methods 22
2.1: Study design 22
2.2: Study population 22
2.3: Selection process 23
2.4: Data collection 24
2.5: Sample size 24
2.6: Statistical analysis 25
2.7: Ethical/Administrative approvals 27

Chapter 3: Results 28
3.1: Characteristics of the Study Population 28
3.2 Comparison of PFT parameters between diagnostic groups 32
3.3 Examination of characteristics of gender subpopulations 38
3.4 Development of diagnostic predictive models utilizing binary logistic regression 40
3.5 Development of diagnostic predictive model in female subpopulation 43
3.6 Development of diagnostic predictive mode for the male subpopulation 47

Chapter 4: Conclusions 50

Chapter 5: Discussion 51

Appendices:
Appendix A: ATS criteria 56
Appendix B: Data collection form 57
Appendix C: Discussion of the ROC curve 58

References: 63

Tables: **page no.**

3.1.1	Diagnostic Category Characteristics by Study Population	29
3.1.2	Percentage Improvement in Spirometry Post-Bronchodilator by diagnosis	31
3.3.1	Comparison of PFT parameters between gender groups	38
3.3.2	Comparison of PFT parameters between Asthmatic & Non-asthmatic Male Patients	39
3.3.3	Comparison of PFT parameters between Asthmatic & Non-asthmatic Female Patients	40
3.4.1	Logistic Regression Classification Table for the Total Population	41
3.5.1	Logistic Regression Classification Table for the Female Population	44
3.6.1	Logistic Regression Classification Table for the Male Population	47

Figures: **page no.**

3.1.1 Diagnostic Distribution of Patient Populations 30

3.2.1a Mean Percentage Improvement in FVC Post-bronchodilator in Diagnostic Groups 33

3.2.1b Mean Percentage Improvement in FEV1 Post-bronchodilator in Diagnostic Groups 34

3.2.1c Mean Percentage Improvement in FEF 25-75% Post-bronchodilator in Diagnostic Groups 35

3.2.2 Receiver Operating Characteristic Curve 37

3.4.1 ROC Curve for Predictive Model for Total Population 43

3.5.1 ROC Curve for Predictive Model in Female Subpopulation 46

3.6.1 ROC Curve for Predictive Model in Male Subpopulation 49

Abbreviations:

ATS	American Thoracic Society
CDC	Centers for Disease Control
COPD	chronic obstructive pulmonary disease
ERS	European Respiratory Society
FEV1	forced expiratory volume in one second
FVC	forced vital capacity
FEF 25-75	see MMFR
L	liter
ml	milliliters
MMFR	maximum mid-expiratory flow rate also known as FEF 25-75
PFT	pulmonary function test
ROC	receiver operator characteristic
RUH	Royal University Hospital
SD	standard deviation
Sec	second
%	percent

1.0 INTRODUCTION:

1.1 Background:

Asthma is an inflammatory disease of the airways with generally reversible airflow obstruction and airway hyper-responsiveness causing episodic respiratory symptoms [1]. An expanded disease definition is that it is characterized by wide variations over short periods of time in airflow resistance of the lung airways, and clinically by recurrent attacks of cough or wheeze separated by symptom free intervals. The airflow obstruction and clinical symptoms are largely or completely reversed by treatment with bronchodilator drugs and/or steroids [2]. While these are somewhat unwieldy definitions, it is important to have a clear understanding what exactly encompasses asthma. This is because asthma is likely not a specific disease but rather a syndrome that occurs from multiple precipitating mechanisms. These precipitating factors result in a common clinical complex involving a classical triad of reversible obstruction, inflammation of the airways, and airway hyperresponsiveness. In contrast, chronic obstructive pulmonary disease (COPD) is defined by the presence of airflow limitation that is not fully reversible after inhaled bronchodilators [3].

1.2 History of Asthma:

Asthma type symptoms have been reported since ancient times. Hippocrates described asthma symptoms in children and related the prevalence of disease to the weather/wind, seasons of the year, and water supply [4]. Galen wrote about catarrh with rapid breathing and "a feeling of constriction" which likely represented asthma [5]. Asthma was first clearly described in the English language literature by Sir John Floyer

in 1698 [6]. He gave clear observations about acute asthma attacks and asthma in general. Numerous physicians and scientists furthered the understanding of asthma including Laennec as well as Henry Salter who, in the 1859, described the hereditary component of asthma [7]. William Osler taught his students the modern theory that asthma was a bronchial disease with obstructing mucus, mucosal edema, and bronchial wall muscular contraction [8]. In the twentieth century, there was a progressive improvement in the understanding of asthma pathophysiology, diagnostic criteria, and treatment methods. Two of the most significant advances included Tiffeneau's description of forced expiratory volume at one second (FEV1) in 1949, as well as the discovery of synthetic cortisone by Kendall & Hench in 1948 with the subsequent use in asthma therapy by cortisone injection in 1950 and by beclomethasone inhalation in 1971 [9, 10, 11].

1.3 Epidemiology of Asthma: General Overview

Asthma affects between 6-10% of the North American population and the prevalence has increased by nearly 75% over that last 20 years [12]. There are over 4000 deaths every year in the United States from asthma while data from Canada is less well defined [13]. While the prevalence of asthma is clearly increasing over the last 20 years the reasons behind this increase remain elusive. Some of the theories raised include increasing industrialization with increased pollution, obesity, sedentary lifestyle with increased indoor exposure to allergens, improvements in asthma reporting rates, and other factors. In the United States, asthma prevalence increased 73.9% from 1980 to 1996 with rates increasing from 31.4 per 1000 to 54.6 per 1000 population [14].

Asthma surveillance is at times challenging to determine in part due to the difficulties of accurate diagnosis. Surveys of asthma depend on the physician's diagnosis of asthma based on physical examination and the patient's report of symptoms as well as certain supplementary pulmonary function testing. Obviously, there may be issues with the accuracy of the physician's diagnosis, the patient's symptoms, and the pulmonary function testing. Thus, inaccuracies may be compounded and the actual prevalence of asthma over- or under-reported. In 1997, the U.S. National Center for Health Statistics changed the annual asthma questionnaire form to improve the measurement of asthma prevalence. Since that time, the data collected shows a stabilization in asthma prevalence compared to the preceding years [15]. Nevertheless, asthma prevalence rates in the United States and elsewhere have overall risen significantly over the past two decades.

One of the challenges in ascertaining asthma prevalence is the issues around asthma and wheezing. A landmark asthma prevalence study in Arizona looked at the relationship between wheezing and asthma in children. The Tucson Children's Respiratory Study followed 826 children in their first six years and found that 49% had problems with wheezing [16]. However, only 14% of those children had persistent wheezing and may be predisposed to asthma in later years. Another large study looked at older children in Australia. In studying 8 to 11 year olds, Salome and colleagues, found that 34% had "any respiratory symptom," 24% were found to wheeze, 18% had airway hyperresponsiveness on pulmonary function testing, and only 13% were felt to have asthma [17]. Thus, while wheezing and pulmonary function testing are very important in

assisting with the diagnosis of asthma, the actual physician diagnostic rates are less and based on a combination of reasons. In conclusion, asthma prevalence has increased substantially over the past twenty years and improvements in diagnostic methods are of particular relevance in this disease.

1.4 Epidemiology of Asthma: Major Risk Factors

There are a number of well-described risk factors for higher asthma prevalence. Among the most common are ethnicity, socioeconomic status, environmental exposures, and respiratory infections.

Although controversial, there is an increased risk for asthma in some ethnic groups. In the United States, the lifetime prevalence of asthma in children is clearly higher in the black population (12.5%) compared to the general Hispanic (7.1%) or white population (7.5%) [18] The same 2004 Centers for Disease Control survey showed a rather low overall prevalence of adult asthma of 4.7% in Hispanics but relatively similar prevalence rates in adult blacks (8.1%) and whites (7.4%). It must be noted that some of the ethnic risk factor component may be related to environmental exposures and socioeconomic access to health services and thus great care must be taken when determining any attributable risk on a racial basis [19]. Nevertheless, a recent study from Alberta, a jurisdiction with universal access to health care, shows that even after adjusting for socioeconomic status, there is still an increased risk for asthma in certain groups [20]. Studies in asthmatic children and women of same socioeconomic class but

different ethnic origin do show some physiologic differences and thus race appears to have at least some component of increased risk for asthma [21, 22].

Asthma has been definitively linked with low socioeconomic status. A survey of randomly selected children from Los Angeles in 2000 revealed asthma rates were highest in black children and children from low income groups [23]. In Canada, similar studies have shown linkage to low income groups. An elegant study from Saskatoon, Saskatchewan in 2006 by Muhajarine, showed a clear link between hospitalizations in children and other child health outcomes with the parent's socioeconomic status and the social context in which the child lives [24]. How much of the increased asthma prevalence is related to low income and what is the component of asthma risk from sub-standard housing, air quality/pollution, ability to afford asthma medications, and other factors is not clear [25, 26].

Besides ethnic and income risk factors for asthma, environmental exposures are a significant risk for increased asthma prevalence. There appears to be a clear link between parental tobacco smoking and the risk of childhood asthma [27]. In addition, second-hand smoke exposure in non-smoking asthmatic adults has been shown in a prospective cohort study to increase asthma severity [28]. The National Health and Nutrition Examination Survey III in 1994 evaluated 18,825 adults and showed an overall prevalence of asthma of 4.5% and an increased prevalence rate related to current or past smoking and pet ownership [29]. Besides tobacco smoke exposure and pet exposure, other environmental

exposures are risk factors for asthma include house mites, cockroach allergen, and rat/mouse exposures [30].

Finally, another area being evaluated as an asthma risk factor is that of respiratory infections and allergen exposures. There is a clear increased prevalence of asthma among the developed world [31]. Part of this increased prevalence may be on the basis of national reporting systems and strong surveillance programs in Western nations. However, other factors are involved. There is interesting evidence that improvements in hygiene can reduce the number of T- 1 helper lymphocytes and increase the number of T-2 helper lymphocytes which related to atopy and allergy. This so-called hygiene hypothesis was first suggested by Strachan [32] and proposes that individuals in less hygienic environments will be exposed to allergens at a young age and thus maturing T-1 helper lymphocytes rather than T-2 helper lymphocytes. The predominance of T-1 helper lymphocytes in those populations is felt to lead to less allergies and asthma but follow-up studies have had variable results [33]. It should be noted that there is some difficulty with the hygiene hypothesis due to potential contribution of asthma risk from ethnic and socioeconomic differences.

1.5 Epidemiology of Asthma: Economic Costs

Asthma has a substantial economic burden both on the individual patient and society at large. For the patient, the costs of typical puffer medications in Canada can range from $25 to $200 per month depending on asthma severity. In the United States, it is believed that patients spend an average of $1000 per annum on asthma medications [34]. For society as a whole, there are significant problems due to absenteeism and/or decreased productivity. A recent study from France looked at the association between asthma control in children and the loss of workdays by their caregivers. They found that over a 1 year period 13% of parents missed over a week of work because of their child's symptoms [35]. In Canada, the most detailed study evaluating the total health-care costs including hospitalizations, showed that Canadian asthma expenses were between $504 and $648 million/year in 1990 [36]. Thus, asthma is a very important disease from both the number of people affected and the total economic costs associated with it.

1.6 Pathophysiology of Asthma:

The pathophysiology of asthma is related to three major factors including inflammation/edema of the bronchial walls, mucus production, and airway smooth muscle contraction and hypertrophy. Inflammation of the airways occurs with an initial trigger causing the release of inflammatory mediators from mast cell and macrophages. The trigger often is a specific agent such as cat dander or ragweed pollen, but may be a non-specific trigger that can be difficult to isolate and avoid. Mucus production is produced both by the initial inflammatory response but also later migration and activation

of eosinophils and neutrophils. Histologic studies show mucus gland and goblet cell hyperplasia. There is also development of hypertrophy of the airway smooth muscle cells. Finally, individuals with asthma have an exaggerated bronchoconstrictor response to many triggers/agents. It is felt that airway inflammation is a key factor underlying the airway hyper-responsiveness which is one of the main hallmarks of asthma [37, 38].

1.7 Diagnosis of Asthma:

Clinical:

Clinical history and physical exam are important in making the diagnosis of asthma. In addition, this helps with the exclusion of asthma mimics, assessment of the severity of airflow obstruction, and the identification of risk factors and triggers. Classic asthma symptoms include episodic cough (dry or productive), wheezing, shortness of breath, and chest tightness. On physical examination, the physician often will find, by percussion of the chest, hyperinflated lungs and hear wheezes with a prolonged expiratory phase of respiration.

Laboratory:

Laboratory testing for asthma includes tests for atopy such as an eosinophil count or IgE level. Unfortunately, these tests are not specific for asthma. There are tests for the indirect assessment of allergic bronchial sensitivity such as allergy skin tests. However, it

is a minority of asthma patients who will clearly demonstrate a significant positive skin test.

Pulmonary Function Testing:

Finally, there are the pulmonary function tests (PFT) that help with making a diagnosis of asthma and do clearly define the degree of airflow obstruction and airway hyper-responsiveness. Basic spirometry was first described by Hutchinson in 1846 with the invention of the water-seal spirometer. He used an inverted glass bell held by a pulley in a column of water. As the subject exhaled through a tube into the glass bell, the bell would move up as air displaced water. This allowed for measurement of vital capacity (VC) and forced vital capacity (FVC) which are measures of how much one can blow out from the lungs. Hutchinson reported measurements on 2130 patients and found that VC was directly related to height and inversely related to age. [39]. A water-seal spirometer when coupled to a fixed speed rotating kymograph would allow for the display of a volume (of exhaled air) over time (in seconds) graph.

Spirometry was later refined by the concept of the timed vital capacity. This was described in French by Tiffeneau [40] in 1949 and in the English literature by Gaensler in 1951 [41]. This concept of a timed vital capacity later became known as forced expiratory volume at one second (FEV1). The inverted bell water displacement spirometry was brought into use in major medical centres in the 1940s and 1950s but was very heavy and not portable. In the 1960s, a new spirometer was developed based on dry displacement bellows allowing for more portability and office-based use. Finally, in the

1980s, pneumotachometers, electronic flow transducers, and computer circuits allowed for the development of hand-held spirometers that would record flow-volume loops in addition to volume-time curves.

The development of spirometry allowed researchers to accurately measure the amount of air in a subject's lungs and how fast that person could exhale the air. However, it is obvious that normative values are essential since a tall person will have larger lungs (and exhaled air) than a small person. Numerous spirometry studies have assessed the normal values for FEV1, FVC, and other lung measurements. It has been clearly demonstrated that there are differences based upon one's age, height, weight, gender, and ethnic origin. As one ages, there is a reduction in lung elasticity and function and thus predicted normal value FEV1 declines with age. Increased weight also will affect PFT values. In regards to gender, it has been shown that women have slightly smaller chest volume than men of the same age, height, weight, and ethnic origin. The etiology of the ethnic differences are unclear but are likely related to the three-dimensional shape, size of lung, and the trunk/height ratio.

Previous investigators have established normal values and developed prediction formula for FEV1 and FVC values in the different groups based on age, height, weight, gender, and ethnicity. Generally, the predictive formulae are based upon non-smoking normal individuals although large-scale data from other research studies is available on smokers as well [42]. The major respiratory medicine associations have released position papers on the variation in the normal population, reference values, and the

standardization of lung function testing [43, 44, 45]. In addition to the normative values in the general population, there has been research into the improvement in airflow rates following inhaled bronchodilators. The American Thoracic Society (ATS) has developed criteria which suggest a significant post-bronchodilator FEV1 response to be 200 milliliters (ml) and 12 % improvement from baseline [43]. Although other groups have suggested slightly different criteria, the generally accepted world-wide standards are the ATS criteria [46].

While FEV1 has been used most frequently to help define asthma severity, the correlation between FEV1 and asthma symptoms is not optimal [47]. Likely more important than the actual FEV1 is the degree of airway reactivity and a rapidly changing FEV1 [48]. Because of this issue, it is common to use bronchoprovocation testing to assist with the diagnosis of asthma in atypical cases. It is well known that all individuals will have bronchoconstriction of their airways in response to various inhaled irritants. Patients with asthma tend to have an excessive response and thus bronchoprovocation testing may help identify hyperreactive airways.

Histamine and methacholine are the two most commonly used agents for bronchoprovocation testing in North America. Methacholine is used predominantly because it has less systemic side effects. The method for methacholine/histamine testing involves a pre-test baseline spirometry, followed by nebulized inhalation of a small dose of the irritating agent, followed by repeat spirometry. This procedure is repeated using higher doses until either there is a 20% drop in FEV1 or the patient reaches the maximal

dose. There are specific protocols for testing and interpretation that have been developed by the ATS [49]. In the atypical or difficult-to-diagnosis patient, either bronchoprovocation testing or bronchodilator responsiveness on spirometry can be very helpful towards eventual diagnosis of asthma. Indeed, in one study of chronic dyspnea, almost 30% of patients required these ancillary tests to achieve an asthma diagnosis [50]. However, it must be noted that while bronchoprovocation testing is highly sensitive for current asthma symptoms, the positive predictive value for is poor [51]. This is because other conditions (e.g. recent bronchitis, pneumonia, etc.) may also give a positive bronchoprovocation test. Thus, bronchoprovocation testing is not recommended as a general screening tool for asthma in the overall population and clinical history and correlation is needed.

1.8 Summary:

The current standard for asthma diagnosis is based on the typical clinical features in conjunction with a significant change in FEV1 after bronchodilator administration. The level of bronchodilator response considered to be diagnostically significant has been the subject of controversy. Definition of the degree of airway reversibility varies widely [52, 53]. The American Thoracic Society (ATS) has developed criteria which suggest a significant post-bronchodilator FEV1 response to be 200 milliliters (ml) and 12 % improvement from baseline [43]. To our knowledge, these criteria have never been fully validated in a North American population by testing against the clinical diagnosis of

asthma. In this study of patients seen at a tertiary care centre, who meet ATS spirometric criteria, we examine bronchodilator responses in reference to clinical diagnosis.

1.9 Research Question

To evaluate the utility of the ATS post-bronchodilator criteria of airway hyperresponsiveness (12% improvement in FEV1 and 200 cc volume) versus expert clinical diagnosis of asthma.

2.0 METHODS:

2.1 Study Designs:

This was a single site, Royal University Hospital (RUH), University of Saskatchewan, retrospective study of clinical medical records and pulmonary function laboratory data.

2.2 Study population:

The province of Saskatchewan has a population of approximately 1 million people. This includes 14.8% at 65 or more years of age, and 25% at less than 20 years of age (2003 data). This population enjoys universal health insurance. The provincial population is a diverse mix of ethnic backgrounds, with relatively large proportions of first nations and Caucasian peoples. There are eight hospitals for complete pulmonary function testing located in the province.

Royal University Hospital tends to see patients from northern and central Saskatchewan and has a referral base in excess of 500,000 people. The RUH-based respirologists are the largest respiratory medicine group in the province with 10 of the 17 adult respiratory medicine physicians in Saskatchewan. The RUH PFT laboratory does approximately 3500 different pulmonary function studies per year [personal communication with Dr. T Hurst, RUH PFT lab director]. The population examined in this study consisted of patients referred for PFT's for various lung diseases and

symptoms and subsequently seen in the outpatient respiratory medicine clinic at Royal University Hospital.

2.3 Selection Process:

All spirometry tests done on patients referred to the Royal University Hospital adult pulmonary function laboratory from September 1999 to September 2004 were reviewed. A total of 15,385 spirometry tests were performed during this period. All pulmonary function testing was performed in accordance with standard ATS protocol with the patient seated, the use of nose clips, and the best of three spirometry flow readings recorded. Vmax model 22 PFT machines (SensorMedics Yorba Linda, California) were employed. A significant bronchodilator response was defined per ATS criteria as an increase in FEV1 of 200 ml and 12% and/or an increase in forced vital capacity (FVC) of 200 ml and 12% [43]. There were a total of 1862 (12.1%) tests meeting the FEV1 improvement criteria and 1916 (12.4%) tests meeting the FVC improvement criteria. A total of 644 individual patients were represented. When a patient had multiple spirometric tests, the most recent test results were recorded. As access to medical records was required for clinical correlation, only adult (age > 18) patients who had been seen in outpatient clinics by staff respirologists (n = 10) were included in this study. This comprised 310 of the 644 patients meeting ATS spirometric criteria for asthma.

2.4 Data Collection:

All medical records were reviewed by a single investigator employing a standard data collection form. Data collected included age, gender, height, weight, and clinical diagnosis (which included: asthma, copd, bronchiectasis, interstitial pulmonary fibrosis, bronchiolitis obliterans organizing pneumonia, pleural disease, non-specific cough, non-specific dyspnea, cardiac disease, sinus disease, vasculitis, and obstructive sleep apnea). Also collected was the PFT data including pre- and post-bronchodilator FEV1, FVC, and MMFR (FEF 25-75). For the purposes of this study, all patients who had either an isolated diagnosis of asthma or a concomitant diagnosis of asthma were considered to have asthma for data analysis. The standard for asthma diagnosis was the staff respirologist's recorded diagnosis based on an extensive consultation assessment (history, complete physical examination, and spirometry) in the outpatient respirology clinic. A standard questionnaire was not used by the 10 respirologists as this study was a retrospective real-life comparison of physician diagnosis in respect to ATS spirometry criteria. All ten staff respirologists consented to the review of their patient's data.

2.5 Sample Size:

The minimum sample size for this study was based on the primary research question of the utility of spirometric variables, ATS post-bronchodilator criteria of airway hyperresponsiveness (12% improvement in FEV1 and 200 cc volume), in the diagnosis of asthmatic patients compared to other patients referred for PFTs.

As the primary outcome measure data is continuous in nature, the following equation was employed to calculate the minimum number of patients required in each group [54]:

Where m = minimum patient number/group.

$$m = \frac{2(z_{1-\alpha/2} + z_{1-\beta})^2}{d^2} + \frac{z^2_{1-\alpha/2}}{4}$$

where: $Z_{1-\alpha/2} = 2.58$, $Z_{1-\beta} = 1.28$, for a two sided $\alpha = 0.01$, and a $\beta = 0.1$, $d = \delta/\sigma$, and σ is 4.0% the percentage change in FEV1 post-bronchodilator standard deviation value [55]. The δ is the value of the clinically significant difference in the post-bronchodilator FEV1. This has been set at 12% as this represents current ATS criteria for a significant post-bronchodilator response. Equal sized groups were assumed.

$$m = \frac{2(2.58 + 1.28)^2}{(12/4)^2} + \frac{(2.58)^2}{4}$$

m = a minimum of 5 patients per group were required for an α of 0.01 and 90% power (1-β).

2.7 Statistical Analysis:

SPSS v.12.0 was employed for data entry and analysis. For the purpose of comparison, patients were divided into the categories of "asthma" and "not-asthma". The non-asthma group was further divided into "COPD", "bronchiectasis" and "other". The reason the non-asthma group was further subdivided was to try to isolate the "COPD" group in particular. Since COPD is a fairly common obstructive airflow disease in older individuals and some of these patients may have a partial reactive component, there is some clinical confusion at times between COPD and asthma. The remainder of the non-

asthma group is split between "bronchiectasis", a disease which often has reactive airways, and the "other" group. Two group comparisons of continuous data were performed using independent two-tailed t-tests. ANOVA was used for three or more group comparisons. The Bonferroni correction for multiple comparisons was utilized [56]. Frequency data was evaluated by chi-square testing. Comparison of varying levels of FEV1 and FVC response between diagnostic groups allowed us to plot a receiver-operator characteristic (ROC) curve [see Appendix C], which was used to further assess the value of these diagnostic criteria.

Data was entered into a computer spreadsheet. Continuous or numerical data were entered directly, categorical data was given a numerical code i.e. for gender, male was coded as 1, female was coded as 2. The numerical code was entered into the spreadsheet for categorical or frequency data. SPSS version 12.0 statistical software was utilized for data analysis.

Continuous data was available for the following variables: age on a ratio scale, FEV1, FVC, MMFR, percentage change in FEV1, FVC, MMFR post-bronchodilator on an interval scale. Continuous data when normally distributed permit use of parametric analytic techniques. Characteristics of the data were assessed; including measures of central tendency, dispersion, skewedness and kurtosis. Comparisons were primarily between two groups. This allowed use of t-tests for two group comparisons. Independent 2-tailed t-tests were used for comparisons of continuous ratio/interval data between groups Two-tailed tests were used rather than one-tailed in order to assess

differences between the two groups in either a positive or negative direction. When appropriate, 95% confidence intervals were reported. In order to correct for multiple testing and avoid an increased risk of type I error, the alpha value required before rejection of the null hypothesis was divided by k (the number of comparisons made).

Frequency data (categorical, binomial, or count data) was generated for gender and disease category. The chi-square distribution, which was used in this study is the most commonly utilized statistical method for analysis of frequency data [57]. However, the chi-square test is not an appropriate method of analysis if minimum expected frequency requirements are not met or if one of the expected frequencies is less than five. Accordingly, Fisher's exact test was used when one or more cell entries were less than five [57]. Multiple testing correction of α/k were also performed.

In development of a predictive model, binary-logistic regression was utilized with the dichotomous outcome of asthma or non-asthma. Variables were included by 'stepwise-enter' method. Variables were selected based on demonstrated association ($p<0.01$) with outcome in bivariate analysis. The Hosmer-Lemeshow goodness of fit test was employed [58]. Models were developed for the total population and separately for gender subgroups.

2.7 Ethical/Administrative Approvals:

Approval for this study was obtained from the University of Saskatchewan Research Ethics Board, and the Saskatoon District Health Research Services Unit.

3.0 RESULTS:

3.1 Characteristics of the Study Population:

Of the 310 adult respiratory clinic patients included in this study, 168 (54.2%) were male, and the remaining 142, female. The mean age of this population was 63.2 years (range: 22-89).

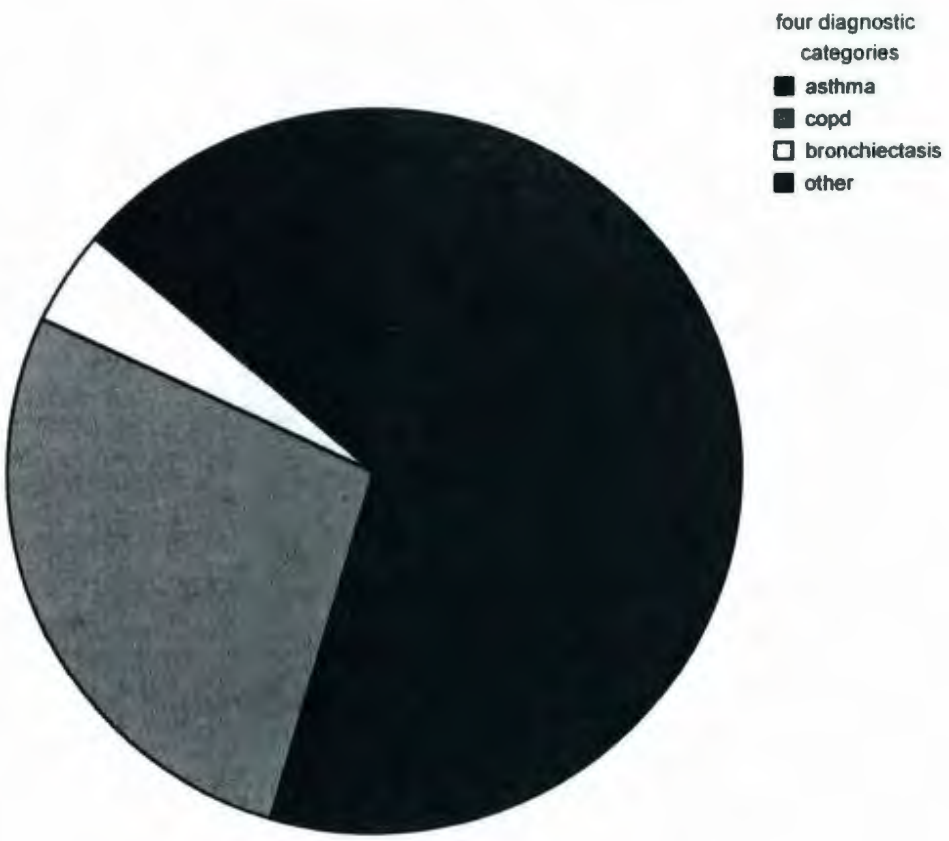
The clinical diagnoses represented in this population as extracted from the respiratory clinic medical record, were divided into four general categories: asthma, COPD, bronchiectasis, and 'other' (Figure 3.1.1.). The 'other' category included interstitial pulmonary fibrosis, obstructive sleep apnea, non-specific cough, non-specific dyspnea, pleural disease, sinus disorders, cardiac disease, vasculitis, and bronchiolitis obliterans organizing pneumonia. The representation of these diagnostic categories within the population is outlined in Table 3.1.1. Asthmatics made up the majority with 169/310 patients (54.5%).

Table 3.1.1.: Diagnostic category characteristics in study population

Diagnosis	Number of patients	Body Mass Index* (kg/m^2)	Age* (years)	% male
Asthma	169 (54.5%)	30.32 (7.48)	60.6 (7.5)	56.2% (95/169)
COPD	85 (27.4%)	27.16 (5.78)	67.7 (11.5)	51.8% (44/85)
Bronchiectasis	13 (4.2%)	26.10 (5.37)	55.7 (19.2)	61.5% (8/13)
'Other'	43 (13.9%)	29.95 (5.85)	66.7 (13.4)	48.8% (21/43)
Total:	310 (100%)			

*Mean values provided, standard deviations in brackets

Figure 3.1.1.: Diagnostic distribution of patient population



In the population as a whole, the mean FEV1 was 1.57 liters (L), with a median of 1.46 L and a standard deviation of 0.67 (range 0.49 to 3.99). The FVC was 2.91 L with a median of 2.78 L and a SD of 0.93 (range 1.35 to 6.15). The FEF 25-75% was 0.75 L/sec, with a median of 0.56 L/sec and a SD of 0.58 (range 0.13 to 3.12). The mean percent improvement in FEV1 post-bronchodilator for the total population was 18.66% (SD 6.35, median 16.67). The mean percentage improvement for FVC post-

bronchodilator was 13.25% (SD 8.00, median 12.03). The mean percentage improvement in FEF 25-75% post-bronchodilator was 28.42 (SD 25.51, median 25.53). The mean values for these variables in each of the diagnostic categories are outlined in Table 3.1.2..

Table 3.1.2.: Percentage improvement in spirometry post-bronchodilator by diagnosis

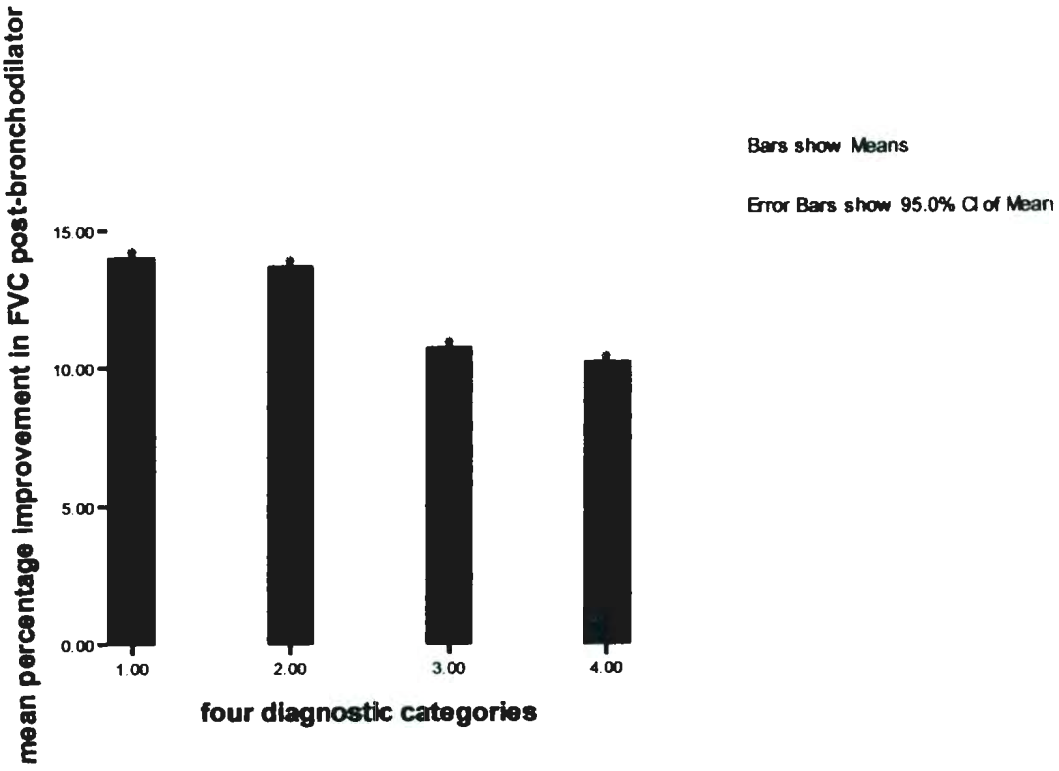
Variable	Diagnosis	Number of patients	Mean (standard deviation)	95% confidence intervals
Percentage improvement in FVC	Asthma	169	14.02 (8.09)	12.79, 15.24
	COPD	85	13.65 (8.59)	11.80, 15.50
	Bronchiectasis	13	10.75 (6.96)	6.55, 14.95
	'Other'	43	10.22 (5.84)	8.42, 12.02
	Total	310	13.25 (8.00)	12.36, 14.15
Percentage improvement in FEV1	Asthma	169	19.41 (6.89)	18.36, 20.45
	COPD	85	18.06 (5.59)	16.85, 19.27
	Bronchiectasis	13	18.58 (5.66)	15.15, 22.00
	'Other'	43	16.91 (5.42)	15.24, 18.58
	Total	310	18.66 (6.35)	17.95, 19.37
Percentage improvement in FEF 25-75%	Asthma	169	29.28 (25.27)	25.45, 33.12
	COPD	85	17.12 (19.88)	12.84, 21.41
	Bronchiectasis	13	39.79 (26.28)	23.91, 55.68
	'Other'	43	43.93 (26.45)	35.79, 52.07
	Total	310	28.42 (25.51)	25.57, 31.27

3.2 Comparison of PFT parameters between diagnostic groups.

Comparison of percentage improvement in these variables between these four diagnostic categories revealed significant differences between groups. For percentage improvement in FEV1 no significant differences were observed between the four diagnostic groups by ANOVA. For percentage improvement in FVC, differences were observed between the asthmatic group and the 'other' group ($p = 0.035$). No significant differences in FVC were observed between the asthmatic group in comparison to the COPD or bronchiectasis groups. Comparison of percentage improvement in FEF 25-75% revealed significant differences between the asthmatic group and both the COPD and 'other' groups ($p = 0.002$, $p = 0.004$). Mean plots are illustrated in Figure 3.2.1.

Figure 3.2.1a.

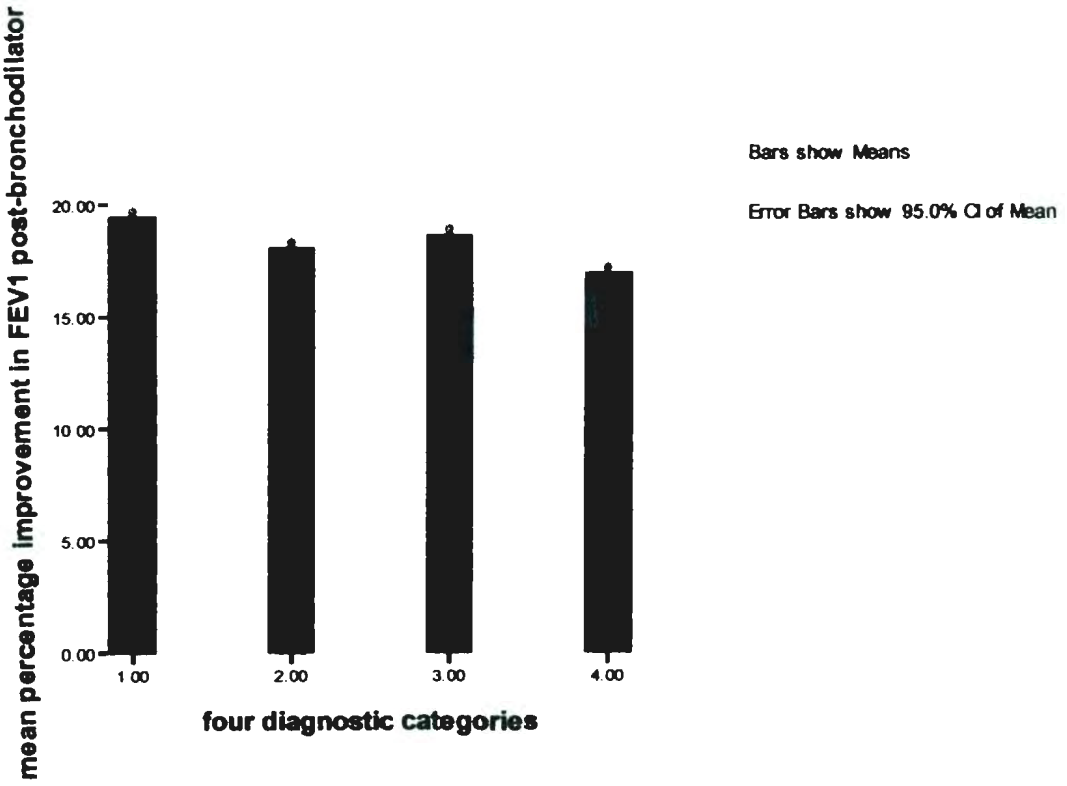
Mean percentage improvement in FVC post-bronchodilator in diagnostic groups



1 = asthma, 2 = COPD, 3 = bronchiectasis, 4 = other

Figure 3.2.1b.

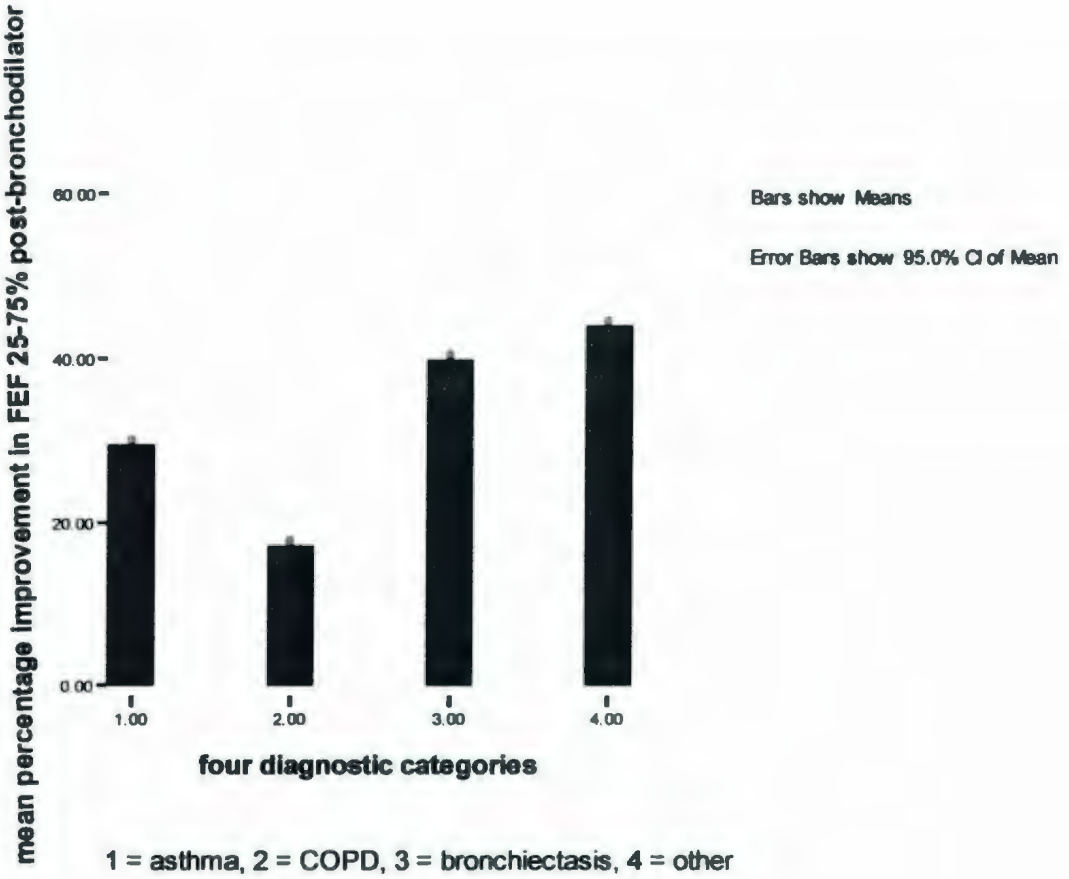
Mean percentage improvement in FEV1 post-bronchodilator in diagnostic groups



1 = asthma, 2 = COPD, 3 = bronchiectasis, 4 = other

Figure 3.2.1c.

Mean percentage improvement in FEF 25-75% post-bronchodilator in diagnostic groups



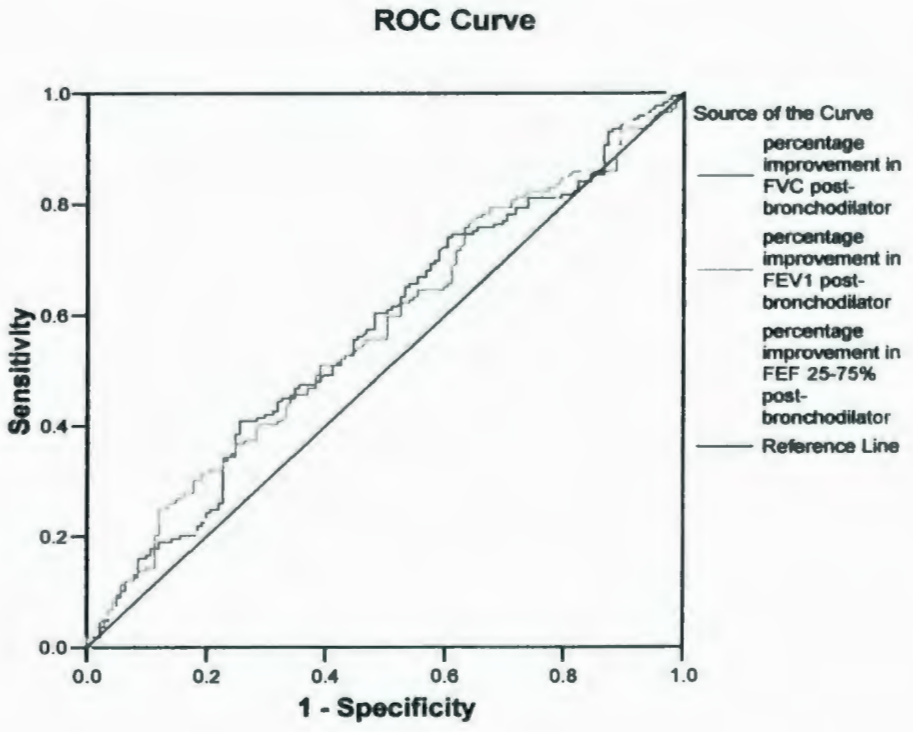
When the non-asthmatic diagnostic categories are collapsed into a single disease category the following is observed: a mean percentage FEV1 improvement post-bronchodilator of 17.82% (SD 5.51), mean percentage FVC improvement of 12.34% (SD 7.86), mean percentage FEF 25-75% improvement of 27.53% (SD 25.89). When these

values are compared to the asthma diagnostic group results, there is a difference between the percentage change in FEV1 ($p = 0.031$). However, this finding loses significance when corrected for multiple comparisons. No significant differences were observed between percentage changes in FVC or FEF 25-75% between these two groups. The population was then divided into those in whom **both** the FEV1 and the FVC criteria are met and those in whom just one of the two criteria are fulfilled. This comparison resulted in essentially equal numbers of patients in both the asthma and 'non-asthma' categories who met only one of the two ATS criteria. In the patients meeting both ATS criteria, 93/155 (60%) were diagnosed with asthma. This difference approached, but did not achieve significance with $p = 0.06$.

The diagnostic role of magnitude of bronchodilator response was then considered. Of the 310 patients meeting the 12% ATS criteria, 169 (54.5%) were clinically diagnosed with asthma. By increasing the requirement to 15% improvement in FEV1, there were 119/206 (57.8%) diagnosed with asthma. Of the 96 patients meeting a 20% FEV1 improvement, there were 61 (63.5%) with diagnoses of asthma. When requiring a percentage increase of 12% or greater in FVC and 15% or greater in FEV1, there were 74 asthmatics (60.2%) out of 123 patients. Within the confines of these arbitrary thresholds is apparent the highest positive predictive value at the 20% improvement in FEV1 level. However, in this study population alone, all of whom met the ATS criteria for a 12 % improvement post-bronchodilator, altering the diagnostic cut-off point to 20%, results in a positive predictive value of 63.5%, a negative predictive value of 49.5%, a sensitivity of 36.1% and specificity of 75.2%.

A receiver operator characteristic curve (Figure 3.2.2) was then utilized to further evaluate the relationship between diagnostic sensitivity and specificity of the ATS criteria within this particular study population. Recalling that when a test contributes no diagnostic information the area under the curve would equal 0.5, and when a test has 100% sensitivity and specificity the area would then equal 1.0, values were obtained for area under the curve for percentage change of FEV1 of 0.569 ($p = 0.037$), of FVC equal to 0.569 ($p = 0.036$), and of FEF 25-75% of 0.517 ($p = 0.608$). As visualized in Figure 3.2.2, the ROC curve for each of these variables shadow the diagonal reference line.

Figure 3.2.2.: Receiver Operator Characteristic curve



3.3 Examination of Characteristics of Gender Subpopulations:

Dividing the population along gender lines, as in Table 3.3.1. the pulmonary function parameters for the total male and total female population are apparent. A comparison of the asthmatic subgroup with the non-asthmatic composite group is made for each gender in Tables 3.3.2. and 3.3.3..

Table 3.3.1.: Comparison of PFT parameters between gender groups.

Parameter	Male n = 168	Female n = 142
% asthmatics	56.5% (95/168)	52.1% (74/142)
Mean age (years)	63.62 (14.26)	62.70 (15.27)
Baseline FEV1:	1.76 (0.70)	1.34 (0.562)
Baseline FVC	3.34 (0.86)	2.40 (0.722)
Baseline FEF 25-75%	0.803 (0.61)	0.690 (.532)
Percentage improvement in FEV1 post-bronchodilator	18.88 (6.52)	18.39 (6.16)
Percentage improvement in FVC post- bronchodilator	13.65 (7.96)	12.78 (8.05)
Percentage improvement in FEF 25-75% post-bronchodilator	27.36 (23.44)	29.67 (27.80)
Body Mass Index kg/m^2	29.22 (5.79)	29.22 (8.04)

Standard deviations in brackets unless otherwise indicated

Table 3.3.2.: Comparison of PFT parameters between asthmatic and non-asthmatic male patients

Parameter	Diagnostic group: non-asthmatics n =73 asthmatics n = 95	mean value (SD)	Significance level	95% confidence intervals
Baseline FEV1	Non-asthmatic	1.65 (0.658)	p = 0.074	(-0.410, 0.020)
	Asthmatic	1.84 (0.728)		
Baseline FVC	Non-asthmatic	3.13 (0.742)	p = 0.004*	(-0.625, -0.120)
	Asthmatic	3.50 (0.916)		
Baseline FEF 25-75%	Non-asthmatic	0.796 (0.623)	p = 0.889	(-0.200, 0.174)
	Asthmatic	0.809 (0.598)		
% improvement in FEV1 post-bronchodilator	Non-asthmatic	17.80 (6.139)	p = 0.058	(-3.915, 0.063)
	Asthmatic	19.72 (6.717)		
% improvement in FVC post-bronchodilator	Non-asthmatic	13.03 (8.238)	p = 0.373	(-3.556, 1.342)
	Asthmatic	14.13 (7.756)		
% improvement in FEF 25-75% post-bronchodilator	Non-asthmatic	22.92 (2.682)	p = 0.407	(-10.249,4.172)
	Asthmatic	23.87 (2.449)		

* significant after correction for multiple comparisons

Table 3.3.3.: Comparison of PFT parameters between asthmatic and non-asthmatic female patients.

Parameter	Diagnostic group: non-asthmatic n = 68 asthmatics n = 74	Mean value (SD)	Significance level	95% confidence intervals
Baseline FEV1	Non-asthmatic	1.16 (0.491)	p < 0.001*	(-0.519, -0.162)
	Asthmatic	1.50 (0.577)		
Baseline FVC	Non-asthmatic	2.26 (0.656)	p = 0.025	(-0.507, -0.034)
	Asthmatic	2.53 (0.761)		
Baseline FEF 25-75%	Non-asthmatic	0.564 (0.482)	p = 0.006*	(-0.416, -0.070)
	Asthmatic	0.807 (0.552)		
Percentage improvement in FEV1 post bronchodilator	Non-asthmatic	17.72 (4.854)	p = 0.208	(-3.296, 0.726)
	Asthmatic	19.00 (7.128)		
Percentage improvement in FVC post-bronchodilator	Non-asthmatic	11.595 (7.360)	p = 0.092	(-4.937, 0.376)
	Asthmatic	13.876 (8.545)		
Percentage improvement in FEF 25-75% post-bronchodilator	Non-asthmatic	29.26 (28.730)	p = 0.865	(-10.061, 8.468)
	Asthmatic	30.06 (27.108)		

* significant after correction for multiple comparisons

3.4 Development of Diagnostic Predictive Models Utilizing Binary Logistic Regression:

A binary logistic regression was then employed to develop a predictive model for asthma diagnosis. The binomial dependant variable was asthmatic or non-asthmatic. The

independent variables providing the best model were: percentage improvement in FEV1, baseline FEV1, age, height (metres), baseline FVC, percentage improvement in FVC post-bronchodilator, baseline FEF 25-75%, and percentage improvement in FEF 25-75% post-bronchodilator. These eight variables provide a model with a chi squared of 48.598 with 8 degrees of freedom and $p < 0.001$. The Homer and Lemeshow goodness of fit test had $p = 0.346$. The classification table (Table 3.4.1) demonstrates 69.0% of predicted results to be correct.

Table 3.4.1.: Logistic Regression Classification Table for Total Population.

		Predicted Diagnosis		
		Non-asthmatic	Asthmatic	% Correct
Observed Diagnosis	Non-asthmatic	85	56	60.3
	Asthmatic	40	129	76.3
	Overall			69.0

The Exp(B) for each variable is as follows:

% improvement in FEV1 post-bronchodilator: Exp(B) = 1.054 (95% CI: 0.995, 1.116), $p = 0.076$

height: Exp(B) = 0.006 (95% CI: 0.000, 0.211), $p = 0.005$

Baseline FEV1: Exp(B) = 20.813 (95% CI: 4.031, 107.462), $p < 0.001$

age: Exp(B) = 0.990 (95% CI: 0.971, 1.010), $p = 0.346$

baseline FVC: Exp(B) = 0.831 (95% CI: 0.412, 1.675), $p = 0.604$

% improvement in FVC post-bronchodilator: Exp(B) = 1.037 (95% CI: 0.992, 1.085), p = 0.110

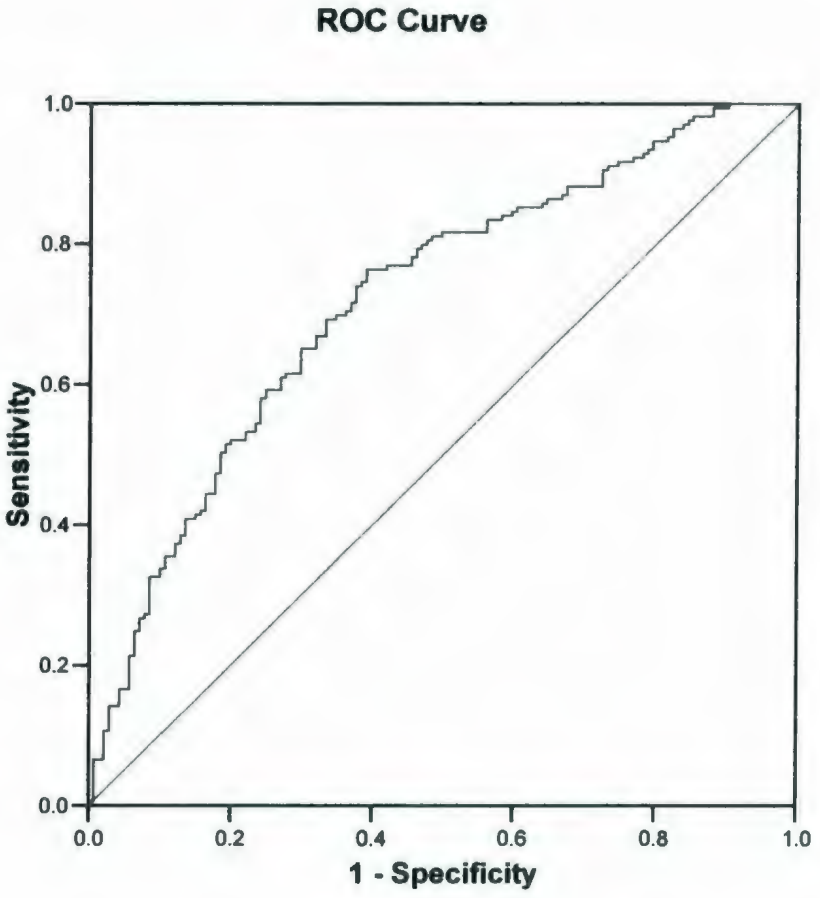
baseline FEF 25-75%: Exp(B) = 0.128 (95% CI: 0.035, 0.463), p = 0.002

% improvement in FEF 25-75% post-bronchodilator: Exp(B) = 0.989 (95% CI: 0.976, 1.002), p = 0.103

From this data it can reported that for this model from the total study population, the relative odds of a diagnosis of asthma increase by a factor of 1.054 for each unit increase in post-bronchodilator FEV1, increase by a factor of 20.813 for each unit increase in baseline FEV1, and increase by a factor of 1.037 for each unit increase in FVC post-bronchodilator. Conversely, the relative odds of a diagnosis of asthma decrease with a factor of 0.986 for each year of age, with a factor of 0.670 for each unit change in baseline FVC, decrease by a factor of 0.114 for each unit change in baseline FEF 25-75%, and by a factor of 0.992 for each percentage change in post-bronchodilator FEF 25-75.t

Further evaluation of the predictive capacity of this model was performed utilizing a ROC curve (Figure 3.4.1.), employing the predictive model probability as the test variable. The area under the curve was calculated to be 0.723, with 95% CI of 0.666 and 0.779 ($p < 0.001$).

Figure 3.4.1.: ROC curve for predictive model for total population



3.5. Development of Diagnostic Predictive Model in Female Subpopulation:

The differences evident between the genders were sufficient that separate predictive models employing binary logistic regression were examined. Again employed

was the dichotomous outcome of asthmatic or non-asthmatic as the dependant variable in both gender models. The independent variables found to make the best model for women included percentage improvement in FEV1 post-bronchodilator, baseline FEV1, age, baseline FVC, percentage improvement in FVC post-bronchodilator, baseline FEF 25-75%, percentage improvement in FEF 25-75% post-bronchodilator, height (cm), and weight (kg). These variables provide a model with a chi-squared of 44.953, with 9 degrees of freedom and $p < 0.001$. The Homer and Lemeshow goodness of fit test had $p = 0.486$. The classification table (Table 3.5.1.) demonstrates 73.9% of predicted results to be correct.

Table 3.5.1.: Logistic Regression Classification Table for Female Population

		Predicted Diagnosis		
		Non-asthmatic	Asthmatic	% Correct
Observed Diagnosis	Non-asthmatic	46	22	67.6
	Asthmatic	15	59	79.7
	Overall			73.9

The Exp(B) for each variable is as follows:

% improvement in FEV1 post-bronchodilator: Exp(B) = 1.071 (95% CI: 0.0964, 1.190), p = 0.201

Weight: Exp(B) = 1.022 (95% CI:0.997, 1.047), p = 0.080

Height: Exp(B) = 0.906 (95% CI: 0.839, 0.978), p = 0.011

Baseline FEV1: Exp(B) = 262.027 (95% CI: 9.399, 7304.723), p = 0.001

age: Exp(B) = 0.977 (95% CI: 0.944, 1.011), p = 0.176

baseline FVC: Exp(B) = 0.260 (95% CI: 0.061, 1.107), p = 0.068

% improvement in FVC post-bronchodilator: Exp(B) = 1.027 (95% CI: 0.960, 1.100), p = 0.441

baseline FEF 25-75%: Exp(B) = 0.070 (95% CI: 0.007, 0.689), p = 0.023

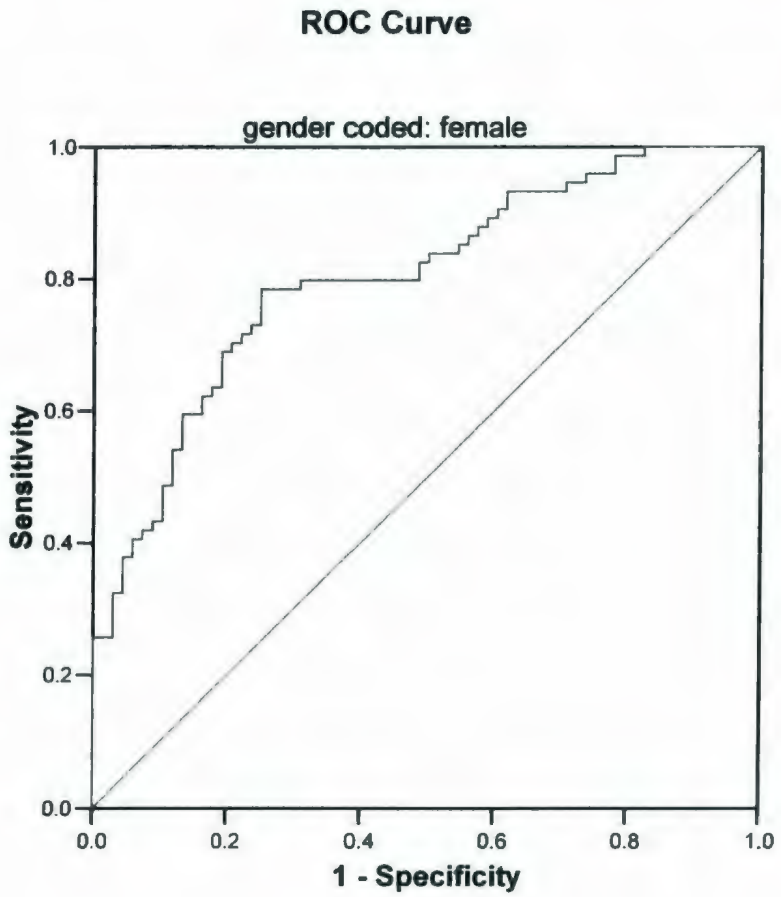
% improvement in FEF 25-75% post-bronchodilator: Exp(B) = 0.981 (95% CI: 0.962, 1.001), p = 0.061

From these results employing data for the female study population, it can be said that within this predictive model, the relative odds of a diagnosis of asthma increase by a factor of 1.071 for each percentage unit increase in post-bronchodilator FEV1, increase by a factor of 1.022 for each kg of weight, increase by a factor of 262.027 for each unit increase in baseline FEV1, and increase by a factor of 1.027 for each percentage unit increase in post-bronchodilator FVC. Conversely, the relative odds of a diagnosis of asthma decrease with a factor of 0.906 for each unit of height, decrease with a factor of

0.977 for each year of age, decrease with a factor of 0.260 for each unit change in baseline FVC, decrease with a factor of 0.070 for each unit change in baseline FEF 25-75%, and decrease with a factor of 0.981 for each percentage unit change in post-bronchodilator FEF 25-75%.

Further evaluation of the predictive capacity of this female subpopulation model was performed utilizing a ROC curve (Figure 3.5.1.), employing the predictive model probability as the test variable. The area under the curve was calculated to be 0.802, with 95% CI of 0.731 and 0.873 ($p < 0.001$).

Figure 3.5.1.: ROC curve for predictive model in female subpopulation



3.6. Development of a predictive model for the male subpopulation:

The independent variables found to make the best model for men included percentage improvement in FEV1 post-bronchodilator, baseline FEV1, age, baseline FVC, percentage improvement in FVC post-bronchodilator, baseline FEF 25-75%, percentage improvement in FEF 25-75% post-bronchodilator, and body mass index (BMI). These variables provide a model with a chi-squared of 25.14, with 8 degrees of freedom and $p = 0.001$. The Homer and Lemeshow goodness of fit test had $p = 0.927$. The classification table (Table 3.6.1.) demonstrates 66.7% of predicted results to be correct.

Table 3.6.1: Logistic Regression Classification Table for Male Population

		Predicted Diagnosis		
		Non-asthmatic	Asthmatic	% Correct
Observed Diagnosis	Non-asthmatic	37	36	50.7
	Asthmatic	20	75	78.9
	Overall			66.7

The Exp(B) for each variable is as follows:

% improvement in FEV1 post-bronchodilator: Exp(B) = 1.057 (95% CI: 0.981, 1.139), $p = 0.148$

BMI: Exp(B) = 1.035 (95% CI: 0.972, 1.102), p = 0.285

Baseline FEV1: Exp(B) = 9.214 (95% CI 1.074, 79.053), p = 0.043

age: Exp(B) = 1.003 (95% CI: 0.975, 1.032), p = 0.836

baseline FVC: Exp(B) = 1.330 (95% CI: 0.546, 3.239), p = 0.531

% improvement in FVC post-bronchodilator: Exp(B) = 1.032 (95% CI: 0.968, 1.099), p = 0.339

baseline FEF 25-75%: Exp(B) = 0.111 (95% CI: 0.019, 0.653), p =0.015

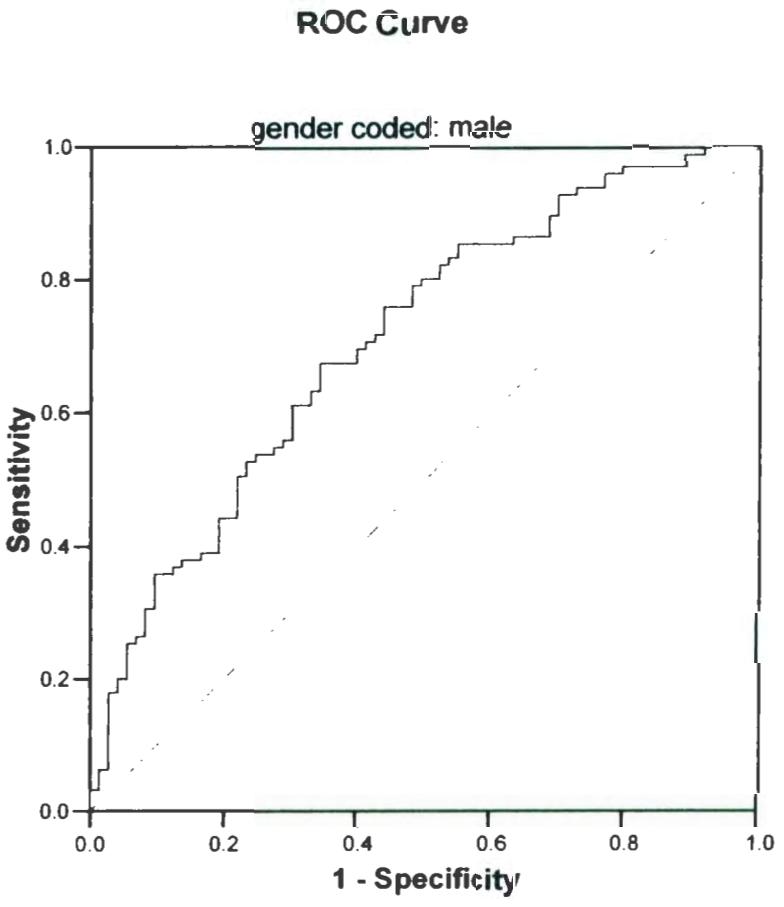
% improvement in FEF 25-75% post-bronchodilator: Exp(B) = 0.993 (95% CI:0.973, 1.013), p = 0.507.

From these results employing the male study population, it can be said that within this predictive model the relative odds of a diagnosis of asthma increase by a factor of 1.057 for each percentage change in post-bronchodilator FEV1, increase by a factor of 1.035 for each unit change in BMI, increase by a factor of 9.214 for each unit change in the baseline FEV1, increase by a factor of 1.003 for each year of age, increase by a factor of 1.33 for each unit change in the baseline FVC, and increase by a factor of 1.032 for each percentage change in post-bronchodilator FVC. Conversely, the relative odds of a diagnosis of asthma decrease with a factor of 0.111 for each unit change in baseline FEF 25-75%, and decrease with a factor of 0.993 for each percentage change in post-bronchodilator FEF 25-75%.

Further evaluation of the predictive capacity of this male subpopulation model was performed utilizing a ROC curve (Figure 3.6.1), employing the predictive model

probability as the test variable. The area under the curve was calculated to be 0.709, with 95% CI of 0.631 and 0.788 ($p < 0.001$).

Figure 3.6.1.: ROC curve for predictive model in male subpopulation



4.0 Conclusions:

1. No statistically significant difference in mean percentage improvement in FEV1 post-bronchodilator was observed between asthmatic patients and the non-asthmatic patients.

2. Patients meeting both FEV1 and FVC ATS criteria consisted of essentially equal numbers of asthmatics and non-asthmatics.

3. Significant differences in baseline spirometric data were evident between male asthmatics versus non-asthmatics and also between female asthmatics versus non-asthmatics. However, no persistence of significant post-bronchodilator changes were evident for either group. Thus, use of the ATS criteria for post-bronchodilator improvement is unhelpful in distinguishing asthma in this patient population.

5.0 Discussion:

There has long been controversy regarding the definition of airway obstruction and significant bronchodilator response. The threshold at which the spirometric criteria are set determine the prevalence of airways obstruction in the general population. In a recent large cross-sectional study, Viegi et al showed clear differences between "clinical criteria" and both European Respiratory Society (ERS) and ATS definitions of airways obstruction [59]. In that population, the ATS criteria had the highest sensitivity but lowest specificity for any respiratory symptom or disease. Likewise, there is variation in the clinical assessment of FEV1 bronchodilator response. In a recent Australian survey of laboratory testing practices for bronchodilator reversibility testing, Borg found significant variation in methods used to assess and interpret bronchodilator response [60].

Even with careful and consistent application of a single method for assessing bronchodilator response, there can be variation over time in the percentage of FEV1 improvement [61]. In the ISOLDE study population of moderate to severe COPD patients, investigators found that bronchodilator responsiveness was a continuous variable and over 52% of patients changed responder status between tri-monthly physician visits [62]. Finally, the ATS criteria have been found to inappropriately classify a high percentage of patients with lower heights as having a non-reproducible test [63]. This further emphasizes the need to use spirometric criteria as a guide but not as an unimpeachable gold standard by which to make a diagnosis of asthma.

In this study population, relying on spirometric criteria alone was found to be inadequate in asthma diagnosis as only 54.7% of 311 patients meeting ATS bronchodilator response criteria were felt to be asthmatic clinically. By raising the threshold of FEV1 bronchodilator response above the 12% required by ATS criteria, there was a little improvement in the positive predictive value of the spirometric test results, increasing from 54.5% to 63.5%. This increased positive predictive value was associated with a low sensitivity of 36.1%. When the spirometry results of the asthmatic patients were compared with those from the COPD, bronchiectasis and 'other' disease categories, there were no significant differences between groups for percentage improvement in FEV1, and only between the asthma group and 'other' group for percentage improvement in FVC.

The receiver operator characteristic areas under the curve values for these variables, although statistically significant, are uncomfortably close to the diagonal reference line which represents a test that contributes no information. These study results suggest the current ATS criteria are relatively non-specific for asthma. Sensitivity of these current criteria was outside the scope of this study as only patients actually meeting the PFT criteria were included in this study.

As this study population was drawn from a tertiary care hospital PFT laboratory and respirology subspecialty clinical practice, these findings may not be generalizable to other populations. Of note, the mean age of asthmatics in this study population was higher than asthmatics in the general population. Furthermore, the patient population

included in this study was heterogeneous with both new and longstanding respiratory clinic patients. Accordingly, varying degrees of pharmacotherapeutic intervention would have been utilized. Clearly since the majority of follow-up clinic patients, who had been diagnosed with asthma, would be receiving appropriate medications, one may expect their spirometric data to reflect lesser baseline bronchoconstriction than new clinic patients. The goal in asthma therapy is to have normal or near normal lung function and symptom control. This anticipated superior control may result in a lesser degree of post-bronchodilator improvement in established asthma patients, which would influence the results. Thus, it is quite possible that some asthmatics were excluded from this study as their most recent PFTs (on ideal asthma control therapy) would not meet the ATS criteria. However, it should be pointed out that the research patient population was selected to include only those who, at the time of clinical assessment, met ATS PFT criteria. Recognized and treated asthmatics who no longer met the 12% improvement post-bronchodilator requirement would not have been included for study. However, the purpose of this study was to evaluate the utility of the ATS post-bronchodilator criteria of airway hyperresponsiveness versus expert clinical diagnosis in the patient population meeting the ATS criteria. Nevertheless, it is possible that the asthma patient study sample is not representative of the asthma population in general.

This study is limited in that it is a retrospective analysis comparing a recorded clinical diagnosis of asthma to spirometric results from the same date. Retrospective extraction of data from clinical records may be less rigorous than a prospective approach. The recorded clinical diagnoses had been arrived at by history and physical examination

of the patients by experienced respirologists at a single site. The spirometric data was usually available to these physicians at the time of the clinic visit and could potentially have influenced the clinical diagnosis. For example, in an unclear or borderline decision about asthma, a clinician may overly rely on spirometry or additional testing such as a bronchoprovocation test (e.g. methacholine challenge). Certainly in the comparison of spirometry with expert clinical diagnosis, the spirometry findings may affect the expert clinical diagnosis which is being used as the "gold standard." However, one would expect any bias to be in favour of a diagnosis of asthma given this research population who all met the ATS spirometry criteria for asthma.

Another potential issue with this study is that of multiple observers. There were ten different respirologists contributing to this data pool. Interobserver variation in diagnosis was not determined, as patients had been seen by a single respirologist only. It has been previously recognized that significant variation can exist between respirologists both in the classification of obstruction as well as determination of bronchodilator response. In an Argentinean study of 30 respirologists, the degree of disagreement for response to bronchodilator was 24% [64]. Thus, while the ATS criteria are not the "gold standard" for asthma diagnosis, it must also be acknowledged that even expert clinicians, reviewing the same patient and PFT data, may disagree about asthma diagnosis. This further illustrates the importance of a comprehensive evaluation of all potential asthma patients along with their PFT data. Nevertheless, expert clinical assessment along with the less well-defined "art of medicine", remains the standard for asthma diagnosis.

It is notable that many patients who did not have a clinical diagnosis of asthma had what is considered a significant bronchodilator response. In previous studies, it has been well recognized that approximately 10% of the COPD population has a component of bronchodilator response [65]. Also, a high percentage of patients with bronchiectasis have a bronchodilator response as previously shown as part of the clinical spectrum for that disease. However, there are other clinical findings that usually make the diagnosis of bronchiectasis fairly clear (e.g. significant purulent sputum, chest radiograph changes, etc). COPD can be more easily confused with asthma but also has certain classic findings that help to differentiate it. Overall, this study was to assess the usefulness in ATS bronchodilator criteria in the diagnosis of asthma. The findings that other diseases may have a significant bronchodilator response simply helps make the pure reliance on ATS criteria rather suspect and points to the importance of a comprehensive approach of history, physical examination, and PFT data analysis.

In summary, while the ATS FEV1 & FVC criteria are helpful in the diagnosis of asthma, relying on spirometric criteria alone is inadequate. In this study, only 54.7% of patients meeting the ATS bronchodilator improvement criteria were felt clinically to have asthma. Thus, use of the ATS criteria for post-bronchodilator improvement alone is unhelpful in distinguishing asthma in this patient population. The diagnosis of asthma depends upon a careful history, physical examination and consideration of complementary pulmonary function tests. Spirometric changes can be helpful in supporting the diagnosis, but the true significance of various levels of bronchodilator response remains unclear.

Appendix A:

ATS Criteria for a Significant Post-bronchodilator Response in FEV1 (or FCV)

A significant improvement is considered an increase in post-bronchodilator FEV1 of 200 milliliters and 12% improvement from baseline FEV1.

Reference:

American Thoracic Society. Standardization of spirometry. 1994 update. Am J Respir Crit Care Med 1995;152:1107-36.

Appendix B: Data Collection Form

Patient Study Number_____

Age____ Gender____ Height____ Weight____ Calculated BMI____

Referring Physician_____

Consulting Respirologist_____

Date seen in clinic_____

Clinical Diagnosis_____

Date of PFTs_____

Baseline FEV1_____ Baseline FVC_____ Baseline FEF25-75_____

Post-BD FEV1_____ Post-BD FVC_____ Post-BD FEF25-75_____

Appendix C: Discussion of the ROC curve

ROC is an abbreviation for receiver operator characteristic. This terminology reflects the origins of this form of analysis. ROC methodology was first developed to help distinguish radar signals from background noise and was the principal graphic device for signal detection theory. The motivation for more accurate radar data arose from World War II, following the Pearl Harbor bombings, when the importance of being able to distinguish enemy aircraft became obvious [66].

Soon after the end of World War II, the ROC had been taken up and used in experimental psychology and psychophysics [67].

Leo Lusted, a radiologist, first broached the concept of using ROC analysis in the medical decision making process. In 1969 he began employing this technique in studies of medical imaging devices [68].

Eventually, ROC analysis became widely used in clinical medicine. It is generally utilized in evaluating the effectiveness of one therapy or diagnostic tool in comparison to a previously established one. This method is extensively employed in epidemiologic medical research. A current medical literature search employing the term 'receiver operator characteristic curve' yields well over 1000 entries today.

ROC analysis is also used in the Social Sciences, where it is also known as ROC Accuracy Ratio. In this field it is a frequently employed method of evaluating accuracy of default probability models. ROC curves have also been increasingly employed in machine learning. In this setting Spackman first demonstrated the value of ROC curves in the comparison and evaluation of different classification algorithms [69].

In clinical medicine, the process of making a diagnosis may involve a number of different forms of examination assessments, ranging from physical examination parameters, laboratory measurements from various bodily fluids or tissues, radiographic features or measurements, or as in this study, physiologic function parameter measurements. The diagnostician will need to choose a point at which the parameter being examined will be decided to have changed from normal to abnormal. The establishment of this "cut-point" permits the evaluation method to categorize the patient's test as either showing or not showing the presence of disease. By creating this threshold value, a continuous outcome measure has been changed into the dichotomy of being positive or negative for disease.

Unfortunately, for the sake of simplicity, the 'cut-point' chosen inevitably will result in some 'normal' or disease-free patients having test results falling into the 'positive' for disease range. Others who are 'abnormal' or known to be carrying the disease will have test results in the 'negative' range. These groups obviously represent the false positive and false negative fractions of the population tested when compared to an established standard of diagnosis (often referred to as the 'gold standard').

With this knowledge derived by comparison to the gold standard, sensitivity and specificity of the diagnostic test under examination may be computed. Recalling that sensitivity is a reflection of the true positive fraction (true positive/true positive + false negative), and specificity a reflection of the true negative fraction (true negative/true negative + false positive) in the population being tested. Modification of the diagnostic threshold point or 'cut-point' in either an upward or downward direction will influence the proportion of patients who would fall into the false positive and false negative groups.

Thereby, every degree of alteration of the 'cut-point' results in change in the sensitivity and specificity for the diagnostic test. For every such test employed by diagnosticians, a multitude of sets of sensitivities and specificities exist. The clinical scientist needs to be able to choose the most appropriate or optimal set for application towards the disease or disorder in question. The prevalence of the disease in the population being tested and the implications of the disease itself need to be considered. For example, in the case of a diagnosis of a treatable cancer a high sensitivity would be highly desirable, even at the cost of a low specificity with many false positive individuals who would then be able to be eliminated by a further more definitive test. On the other hand, giving medications to a child wrongly diagnosed as having attention-deficit hyperactivity disorder exposes that child to short and long term risk of adverse events and the stigma of a label, whereas missing the diagnosis may simply delay the intervention. In that scenario a lower sensitivity with a high specificity would be preferable. No one ideal cut-point exists for all ROC curves. The diagnostician needs to weigh the benefits of high sensitivity with the potential risks of lower specificity and find an appropriate and acceptable middle ground for the specific disorder and population in question.

The ROC curve itself is a two dimensional graphic plot of sensitivity (the true positive rate) on the y axis versus 1-specificity (the false positive rate) on the x axis. This visual representation demonstrates the spectrum of sensitivity and specificity across the range of possible 'cut-points'. Once the curve is constructed, it is possible to evaluate the diagnostic accuracy of the test by measuring the 'area under the (ROC) curve', also termed the '*c*' statistic. The closer this area is to 0.5, the closer the morphology of the curve is to the diagonal and the less useful. The diagonal indicates the test is no better

than chance at discriminating between the presence and absence of disease. The diagonal therefore represents the null hypothesis. The more the ROC curve swings toward the upper left hand corner, the better the sensitivity and specificity of the test. An entirely perfect test would have an area of 1.0, with 100% sensitivity and 100% specificity. In this case the plotted line would run straight up the y axis until the top and then run horizontally. The existence of such a test is highly unlikely. However, the closer the area is to 1.0, the better the test is at distinguishing between affected and unaffected populations [70].

There are distinct advantages to employing ROC curves in clinical research. The visual presentation is easily appreciated and is highly complementary to sensitivity and specificity tabular data. It is a comprehensive representation of pure accuracy over the entire range of the test. The sensitivity and specificity for any chosen "cut-point" are readily available. The ROC curve is independent of prevalence, although prevalence of disease must be taken into account in choosing a 'cut-point'. In contrast to many analytic tools, it is not imperative to have normally distributed data. ROC curves can be used for both nonparametric and parametric data. In contrast to the smooth curve of the latter, the former has the disadvantage of a staircase appearance and can compare plots only at observed sensitivity and specificity. It is possible to calculate standard errors and go on to sample size estimation from ROC data [71].

Disadvantages of ROC plots include the absence of clearly identifiable features on the graph, including the decision thresholds and the number of subjects. Also, generally, construction of ROC plots requires utilization of computer software; however this is less of a concern with the 'user-friendly' resources which are now available. Other

concerns include background or random 'noise' or random variations which might affect the curve and degrade test performance. The largest issue appears to rest with the utilization of a gold standard for comparison. If the gold standard is not independent of the test under examination, the interdependence will give you a spuriously high area under the curve value. Additionally, if the test employed as the gold standard is poor at identifying disease, it is difficult to truly assess accuracy of any other form of discrimination whether by ROC curve or other method of comparison. Error can also arise from the presence of a co-morbidity which may influence the test, from verification bias on the part of the unblinded clinician, interobserver variation in situations where a subjective measurement is taken, 'test-review' bias, and 'incorporation' bias where the test result is incorporated into the evidence used to diagnose the disease. Finally, it should be recognized that tests are applied to specific populations and ROC curves reflect the sensitivity and specificity of the results in that specific patient population. Extrapolation to different populations may not be appropriate. For example, a test used to diagnose an advanced bulky malignancy may fail to diagnose or have a very low sensitivity for a small early form of the same disorder [72].

A receiver operator characteristic plot can be most useful in comparison of two or more measures. A test with a curve that lies completely above the curve of another will be obviously superior in accuracy. ROC curves which cross require more extensive comparison analysis than those entirely separate [73].

References:

1. National Asthma Education and Prevention Program. Guidelines for the diagnosis and management of asthma: expert panel report 2. Bethesda, Md.: U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, National Heart, Lung, and Blood Institute, 1997;NIH publication No. 97-4051.

2. Clark TJ, Godfrey S, eds. Asthma. 2nd Ed. Chapman and Hall, London, 1983.

3. Global initiative for chronic obstructive lung disease. Global strategy for the diagnosis, management and prevention of chronic obstructive pulmonary disease. NHLBI/WHO Workshop report. NIH Publication No. 2701. April 2002.

4. Hippocrates. Airs, Waters, Places. In: Hippocratic Writings. Chadwick J, Mann WN, eds. Oxford, Blackwell 1950.

5. Capel LH. A history of the pulmonary catarrhs and airways obstruction. Br J Dis Chest 1966;60:113.

6. Sakula A. Sir John Floyer's "A Treatise of the Asthma." Thorax 1984;39:248-254.

7. Sakula A. Henry Hyde Salter (1823-71): A biographical sketch. Thorax 1985;40:887.

8. Osler W. The principles and practice of medicine. New York, D Appleton 1892.

9. Tiffeneau R, Rousser J, Drutel R. Capacite vitale et capacite pulmonaire utilizable a l'effort: criteres statique et dynamique de la ventilation pulmonaire. Paris Med 1949;137:543-47.

10. Carey RA, Harvey AM, Howard HE, Winkenwerder WL. The effect of adrenocorticotrophic hormone (ACTH) and cortisone on the course of chronic bronchial asthma. Bull Johns Hopkins Hosp 1950;87:387-414.

11. Brown HM, Storey G, George WHS. Beclomethasone dipropionate: A new steroid aerosol for the treatment of allergic asthma. Br Med J 1971;1:585-590.

12. Akinbami L, Schoendorf K. Trends in childhood asthma: prevalence, health care utilization, and mortality. Pediatrics 2002;100(2 pt 1):315-22.

13. Centers for Disease Control. Asthma prevalence. Health care use and mortality. 200[illegible] Available at www.cdc.gov/nchs/products/pubs/pubd/hestats/asthma03-05.htm Accessed Nov 30, 2007.

14. Lugogo NL, Kraft M. Epidemiology of Asthma. Clin Chest Med 2006;27:1-15.

15. Centers for Disease Control. Surveillance for asthma – United States, 1980-1999. MMWR, March 29, 2002;51(ss01):1-13.

16. Martinez FD, Wright AL, Taussig LM, et al. New Eng J Med 1995;332:133-138.

17. Salome CM, Peat JK, Britton WJ, Woolcock AJ. Bronchial hyperresponsiveness in two populations of Australian school children. 1. Relation to respiratory symptoms and diagnosed asthma. Clin Allergy 1987;17(4):271-81.

18. Centers for Disease Control 2004 National Health Interview Survey, available at www.cdc.gov/nchs/data/series/sr_10/sr10_227.pdf Accessed Nov 30, 2007.

19. Smith LA, Hatcher-Ross JL, Wertheimer R, et al. Rethinking racial/ethnicity, income, and childhood asthma: racial/ethnic disparities concentrated among the very poor. Public Health Rep 2005;120:109.

20. Sin DD, Svenson LW, Cowie RL, et al. Can universal access to health care eliminate health inequities between children of poor and nonpoor families? A case study of childhood asthma in Alberta. Chest 2003;124:51.

21. Joseph CL, Ownby DR, Peterson EL, et al. Racial differences in physiologic parameters related to asthma among middle-class children. Chest 2000;117:1336.

22. Sherman CB, Tollerud DJ, Heffner LJ, et al. Airway responsiveness in young black and white women. Am Rev Respir Dis 1993;148:98.

23. Smith PA, Zeng Z, Wold CM, et al. Prevalence of childhood asthma and associated morbidity in Los Angeles County: impacts of race/ethnicity and income. J Asthma 2003;40:535-43.

24. Muhajarine N, Vu L, Labonte R. Social contexts and children's health outcomes: Researching across the boundaries. Critical Public Health 2006;16(3):205-218.

25. Rosenstreich DL, Eggleston P, Kattan M, et al. The role of cockroach allergy and exposure to cockroach allergen in causing morbidity among inner-city children with asthma. New Engl J Med 1997;336:1356.

26. Federico MJ, Liu AH. Overcoming childhood asthma disparities of the inner-city poor. Pediatr Clin North Am 2003;50:655.

27. Strachan DP, Cook DG. Parental smoking and childhood asthma: longitudinal and case control studies. Thorax 1998;53:204-12.

28. Eisner M, Yelin EH, Henke J, et al. Environmental tobacco smoke and adult asthma: the impact of changing exposure status on health outcomes. Am J Respir Crit Care Med 1998;158:170-5.

29. Arif AA, Delclos ES, Lee ES, et al. Prevalence and risk factors of asthma and wheezing among US adults: an analysis of the NHANES II data. Eur Respir J 2003;21:827-83.

30. Korsgaard J. Preventive measures in house-dust allergy. Am Rev Respir Dis 1982:125:80-84.

31. The International Study of Asthma and Allergies in Childhood (ISAAC) Steering Committee. Worldwide variation in prevalence of symptoms of asthma, allergic rhinoconjunctivitis and atopic eczema: ISAAC. Lancet 1998;351:1228.

32. Strachan DP. Hay fever, hygiene, and household size. BMJ 1989;299:1259-60.

33. Strachan DP, Butland BK, Anderson HR. Incidence and prognosis of asthma and wheezing illness from early childhood to age 33 in a national British cohort. BMJ 1996;312:1195-99.

34. Sullivan S, Elixhauser A, Buist AS, et al. National Asthma Education and Prevention Program working group report on the cost effectiveness of asthma care. Am J Respir Crit Care Med 1996:154:S84-95.

35. Laforest L, Yin D, Kocevar VS, et al. Association between asthma control in children and loss of workdays by caregivers. Ann Allergy Asthma Immunol 2004;93:265-71.

36. Krahn MD, Berka C, Langlois P, Detsky AS. Direct and indirect costs of asthma in Canada, 1990. CMAJ 1996:154:821-31.

37. Chung KF. Role of inflammation in the hyperactivity of the airways in asthma. Thorax 1986:41:657-662.

38. O'Byrne PM, Hargreave FE, Kirby JG. Airway inflammation and hyperresponsiveness. Am Rev Resp Dis 1987;136:S35-37.

39. Hutchinson J. On the capacity of the lungs, and on the respiratory functions with a view of establishing a precise and easy method of detecting disease by the spirometer. Trans Med Chir Soc London 1846, pp 29-137.

40. Tiffeneau R, Rousser J, Drutel R. Capacite vitale et capacite pulmonaire utilizable a l'effort: criteres statique et dynamique de la ventilation pulmonaire. Paris Med 1949;137:543-47.

41. Gaensler EA. Analysis of the ventilatory defect by timed vital capacity. Am Rev Tuberc 1951;69:256-278.

42. Enright PL, Johnson LR, Connettt JE, Voelker H, Buist AS. Spirometry in the Lung Health Study. 1. Methods and quality control. Am Rev Respir Dis 1991;143:1215-23.

43. American Thoracic Society. Standardization of spirometry. 1994 update. Am J Respir Crit Care Med 1995;152:1107-36.

44. European Respiratory Society. Standardized lung function testing. Lung volumes and forced ventilatory flows. 1993 update. Eur Respir J 1993;6(suppl 16):5-40.

45. American Thoracic Society. Lung function testing: selection of reference values and interpretative strategies. Am Rev Respir Dis 1991;144:1202-18.

46. Lindberg A, Jonsson AC, Ronmark E, et al. Prevalence of chronic obstructive pulmonary disease according to BTS, ERS, GOLD, and ATS criteria in relation to a doctor's diagnosis, symptoms, age, gender, and smoking habits. Respiration 2005;72(5):471-79.

47. Teeter JG, Bleecker ER. Relationship between airway obstruction and respiratory symptoms in adult asthmatics. Chest 1998;113:272-7.

48. Chan MT, Leung DY, Szefler SJ, et al. Difficult to control asthma: clinical characteristics of steroid-insensitive asthma. J Allergy Clin Immunol 1998;101:594-601.

49. Guidelines for methacholine and exercise challenge testing – 1999. American Thoracic Society. Am J Respir Crit Care Med 2000;161:309-329.

50. Pratter MR, Curley FJ, DuBois J, et al. Cause and evaluation of chronic dyspnea in a pulmonary disease clinic. Arch Intern Med 1989:149:2277-82.

51. Cockcroft DW, Murdock KY, Berscheid BA, et al. Sensitivity and specificity of histamine PC20 determination in a random selection of young college students. J Allergy Clin Immunol 1992;89:23-30.

52. Eliasson O, Degraff AC Jr. The use of criteria for reversibility and obstruction to define patient groups for bronchodilator trials. Influence of clinical diagnosis, spirometric and anthropometric variables. Am Rev Resp Dis 1985;132(4):858-64.

53. Brand PL, Quanjer PH, Postma DS, Kerstjens HA, Koeter GH, Dekhuijzen PN, et al. Interpretation of bronchodilator response in patients with obstructive airways disease. The Dutch Chronic Non-Specific Lung Disease (CNSLD) Study Group. Thorax 1992:47:429-36.

54. Campbell MJ, Julious SA, Altman DG. Estimating sample sizes for binary, ordered categorical, and continuous outcomes in two group comparisons. BMJ 1995; 311:1145-1148.

55. Dales RE, Spitzer WO, Tousignant P, Schechter M, Suissa S. Clinical interpretation of airway response to a bronchodilator. Epidemiologic considerations. Am Rev Respir Dis 1988:138:317-320.

56. Norman GR, Streiner DL. Biostatistics. The Bare Essentials. 2nd edition. B.C. Decker Inc., Hamilton 2000.

57. Daniel WW. Biostatistics: A Foundation for Analysis in the Health Sciences. 7th edition. Wiley & Sons Inc., New York 1999.

58. Hosmer DW, Lemeshow S. Applied Logistic Regression. Wiley & Sons Inc.. New York 1989

59. Viegi G, Pedreschi M, Pistelli R, Di Pede F, Baldacci S, Carrozzi L et al. Prevalence of airways obstruction in a general population: European Respiratory Society versus American Thoracic Society definition. Chest 2000;177(5S2):339S-45S.

60. Borg BM, Reid DW, Walters EH, Johns DP. Bronchodilator reversibility testing: laboratory practices in Australia and New Zealand. Med J Aust. 2004;180(12):610-613.

61. Tweeddale PM, Alexander F, McHardy GJ. Short term variability in FEV1 and bronchodilator responsiveness in patients with obstructive ventilatory defects. Thorax 1987:42:487-90.

62. Calverley PMA, Burge PS, Spencer S, Anderson JA, Jones PW, for the ISOLDE Study Investigators. Bronchodilator reversibility testing in chronic obstructive pulmonary disease. Thorax 2003;58:659-664.

63. Hankinson JL, Bang KM. Acceptability and reproducibility criteria of the American Thoracic Society as observed in a sample of the general population. Am Rev Respir Dis. 1991;143(3):516-21.

64. Quadrelli SA, Roncoroni AJ, Porcel G. Analysis of variability in interpretation of spirometric tests. Respiration 1996;63(3):131-6.

65. Nisar M, Earis JE, Pearson MG, Calverley PM. Acute bronchodilator trials in chronic obstructive pulmonary disease. Am Rev Resp Dis 1992;146:555-559.

66. Green DM, Swets JM. Signal detection theory and psychophysics. 1966, John Wiley and Sons Inc. New York. ISBN 0-471-32420-5.

67. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a ndamental evaluation tool in clinical medicine. Clin Chem 1993:39(4):561-77.

68. Lusted LB. Signal detectability and medical decision-making. Science 1971:171:1217-9.

69. Spackman KA. 'Signal detection theory: Valuable tools for evaluating inductive learning'. In Proceedings of the Sixth International Workshop on Machine Learning. Morgan Kaufman, San Mateo, CA, 1989, pp.160-163

70. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. Circulation 2007;115:928-935.

71. Altman DG. Bland JM. Diagnostic tests 3: Receiver operating characteristic plots. BMJ 1994;309:188.

72. Bewick V. Cheek L. Ball J. Statistics review 13: Receiver operating characteristic curves. Crit Care 2004:8(6):508-12.

73. Streiner DL. Cairney J. What's under the ROC? An introduction to receiver operating characteristics curves. Can J Psych 2007:52:121-128.