ABSTRACT

Title of Dissertation: THE IMPACT OF DISEASE SEVERITY AND PHENOTYPE ON SMOKING AMONG ADULTS WITH CHRONIC OBSTRUCTIVE PULMONARY DISEASE

Timothy James Tilert, Doctor of Philosophy, 2019

Dissertation directed by: Professor Min Qi Wang
Department of Behavioral and Community Health

Chronic Obstructive Pulmonary Disease (COPD) is estimated to be the third leading cause of death in the US. The most significant risk factor for COPD is long-term cigarette smoking. In spite of the myriad benefits of cessation, the proportion of adults with COPD who currently smoke is still nearly 50%. Little is known, however, about the characteristics of, and subsequent differences between, smokers with COPD, particularly at differing lung obstruction severity levels. The goals of this dissertation were to examine and compare the characteristics of smokers with diagnosed COPD as well as to explore the impact of disease severity and disease phenotype on smoking status among persons with COPD.

This research utilized secondary data on 10,219 examined adults, aged 40-79 years, from the 2007-2012 National Health and Nutrition Examination Survey. In Study 1, adjusted logistic regression analyses revealed multiple factors that were associated with self-reported COPD diagnosis with those reporting three or more respiratory symptoms having the strongest association (AOR=22.1, 95% CI=12.0-40.5). In Study 2,

it was shown that smoking status proportions did not differ by lung obstruction severity among those reporting a COPD diagnosis. In adjusted logistic regression analyses, multiple factors were associated with current smoking status among those with self-reported COPD with the presence of other smokers in the household having the strongest association with being a current smoker (AOR=19.5, 95% CI=10.2-37.5). In Study 3, three distinct phenotypes were found among the COPD population analyzed. In adjusted logistic regression analyses, COPD phenotype was differentially associated with continued smoking, above and beyond other predictors, with the older, heavy-smoking males with emphysema phenotype showing a significant positive association with continued smoking (AOR=3.7, 95% CI=1.3-10.9).

Understanding how differences in disease severity and disease phenotypes impact smoking status among persons with diagnosed COPD could help inform more targeted, and effective, interventions to reduce smoking rates in this high-risk population. These findings potentially provide guidance for current smoking cessation interventions aimed at smokers with COPD as well as provide the foundation for further exploration of the association between COPD phenotype and continued smoking.

THE IMPACT OF DISEASE SEVERITY AND PHENOTYPE ON SMOKING AMONG ADULTS WITH CHRONIC OBSTRUCTIVE PULMONARY DISEASE

by

Timothy James Tilert

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2019

Advisory Committee:
Professor Dr. Min Qi Wang, Chair
Dr. Ryne Paulose-Ram, Acting Director, NHANES
Associate Professor Dr. Donna Howard
Associate Professor Dr. James Butler III
Professor Dr. Sunmin Lee

Table of Contents

List of Figures and Tables

CHAPTER I.

INTRODUCTION

1.1 Problem Statement

Chronic Obstructive Pulmonary Disease (COPD) is estimated to be the third leading cause of death both in the U.S. (Murphy et al., 2013) and around the globe (WHO, 2016) and is a growing cause of suffering and morbidity worldwide. In the U.S., nearly 15 million adults are estimated to be suffering from COPD (NHLBI, 2012). The economic impact of COPD on the health system in the U.S. is enormous with an estimated $30 billion in direct costs in 2010 and an additional $20 billion in indirect costs (GOLD, 2018; Guarascio, Ray, Finch, & Self, 2013).

In more economically developed countries, cigarette smoking is widely regarded as the primary risk factor for developing COPD (Molfino, 2004; Scanlon et al., 2000; Silverman & Speizer, 1996) with a number of reports estimating that smoking is the cause in roughly 75% of all COPD cases (Lindberg, Bjerg-Bäcklund, Rönmark, Larsson, & Lundbäck, 2006b; Løkke, Lange, Scharling, Fabricius, & Vestbo, 2006; Lopez, A.D., Mathers, C.D., Ezzati, M., Jamison, D.T., & Murray, C.J., 2006).

Though there is no known cure for COPD, quitting smoking has been shown to improve respiratory symptoms, reduce the rate of lung function decline, decrease the risk of exacerbations, and improve the health-related quality of life for those with the disease (Anthonisen et al., 2002; Au et al., 2009; Bolliger et al., 2002; Willemse et al., 2004). Accordingly, smoking cessation is widely regarded as the most effective therapeutic intervention in patients of COPD (GOLD, 2018; Hersh et al., 2004; Stratelis, Mölstad, Jakobsson, & Zetterström, 2006; Willemse et al., 2004).

In spite of the benefits of quitting, a recent study has shown that the smoking rate among the adult COPD population is still more than double that of adults with no chronic diseases, i.e., 47% vs. 20%, respectively (Schauer et al., 2014). Recent NHANES data support this estimate of smoking among the COPD population showing that approximately 46% of adults aged 40-79 years with COPD are current smokers (Paulose-Ram et al., 2015). The NHANES data further revealed that a significantly higher proportion of persons with moderate or more severe lung obstruction currently smoke (56%) than do persons with mild lung obstruction (41%).

Though one might intuitively think that increased severity of lung obstruction would increase motivation for people with COPD to quit smoking, actual smoking behavior seems to contradict this intuition. This begs the question, "What role, if any, does severity play in the smoking behavior of persons with COPD?" Further, given the complex and heterogeneous nature of the disease, are there specific combinations of characteristics in addition to, or in lieu of, disease severity that are more predictive of smoking behavior than just severity? To answer these questions, this study examined the roles that disease severity and phenotype play in smoking status among adults with COPD. Specifically, this research investigated whether disease severity predicted smoking status directly, or moderated the relationships between other predictors of smoking status, among ever smokers with COPD and actual smoking behavior and whether COPD phenotypes differentially predicted smoking status among ever smokers with COPD.

1.2 Justification for Current Study

1.2.1 Significance of Research

It is known that smoking rates in the COPD population are higher than smoking rates in the population without chronic diseases (Schauer, 2014). Very recent data suggests that, in addition to this difference, rates of smoking are also higher for those with more severe lung

obstruction than for those with milder lung obstruction (Paulose-Ram et al., 2015). As these data have only recently been published (2015), little is known about the characteristics of, and subsequent differences between, smokers with COPD at differing lung obstruction severity levels. This research will be the first to explore the characteristics of smokers with diagnosed COPD at different severity levels and will also be the first to assess the moderating effect of lung obstruction severity on the relationship between factors associated with smoking and smoking status among persons with COPD.

As phenotyping is still relatively new in the COPD literature, this research attempted to address a number of known limitations in the current studies as well as add to the body of knowledge regarding COPD phenotyping by being the first study to phenotype COPD sufferers using smoking status as the outcome measure of interest. This was accomplished using a highly regarded, nationally-representative dataset (NHANES) and it is hoped that the results will generate substantial clinical, behavioral and/or scholarly interest.

1.2.2 Public Health Implications

From a public health perspective, COPD is estimated to be the third leading cause of death in the United States (Murphy et al., 2013) and around the globe (WHO, 2016) and is a growing cause of suffering, morbidity and mortality worldwide. Smoking cessation is the single most therapeutic intervention for COPD sufferers at all severity levels and has been associated with a decrease in symptoms, reduction in number of exacerbations and improved overall health status for those with the disease (Anthonisen et al., 2002; Au et al., 2009; Bolliger et al., 2002; Willemse et al., 2004). Additionally, smoking cessation is the only therapy that has been clearly shown to slow the rate of lung function loss in COPD patients (Anthonisen et al., 1994).

To date, the majority of smoking cessation interventions targeted at those with COPD have focused on a specific range of disease severity and have left unexplored differences in severity level of disease on intervention success. Two studies were found that drew comparisons in intervention success rates between groups with differing levels of COPD severity and found inconsistent, and somewhat contradictory, results (Górecka et al., 2003; Tashkin et al., 2011). In their intervention assessing the impact of airflow limitation diagnosis combined with cessation advice, Górecka et al found that more severe COPD predicted higher smoking cessation rates (Górecka et al., 2003). In their pharmacologically-based intervention, however, Tashkin et al found more severe COPD predicted lower smoking cessation rates, though too few participants with severe disease were studied to allow any firm conclusions (Tashkin et al., 2011). Given the scarcity of both intervention and observational research comparing smokers, and former smokers, with COPD at different severity levels, and the inconsistent results in the few intervention studies that did differentiate between disease severity levels, clearly more work is required to understand how differences in disease severity level may influence continued smoking and smoking cessation. Expanding our focus beyond just disease severity and evaluating other characteristics, particularly in groups, could also prove beneficial in identifying and understanding differences in COPD patients, identifying those with increased risks for specific outcomes, and, perhaps most importantly, personalizing therapies including smoking cessation (Cho et al., 2010).

In their meta-analysis of smoking interventions for COPD patients, Coronini-Cronberg et al noted that interventions for smokers with COPD are often based on evidence from smokers in the general population, not specifically those with COPD (Coronini-Cronberg, Heffernan, & Robinson, 2011), alluding to the need for more tailored interventions for COPD patients. We

know, from multiple systematic analyses, that behavioral counseling combined with pharmacological therapy increases the chance of quitting smoking for those with COPD. More effective behavioral counseling strategies can only improve the effectiveness of these types of interventions. Developing a better understanding of the differences between smokers with COPD of different severity levels and phenotyping subgroups of persons with COPD with select characteristics could inform more targeted, and effective, behavioral counseling strategies to reduce smoking rates in this high-risk population.

1.3 Definition of Key Terms and Acronyms

ATS: American Thoracic Society

COPD: Chronic Obstructive Pulmonary Disease

ERS: European Respiratory Society

FEV1: Spirometric measure of Forced Expiratory Volume in 1 second

FVC: Spirometric measure of Forced Vital Capacity

GOLD: Global Initiative for Obstructive Lung Disease

PHENOTYPE: The observable properties of an organism that are produced by the interaction of the genotype and the environment (Merriam & Webster, 2019)

1.4 Theoretical Framework

The Self-Regulation Model of Illness (SRMI) and the Inverted U Theory (Yerkes-Dodson Law) provide the theoretical guidance for this research project.

1.4.1 Self-Regulating Model of Illness (SRMI)

The SRMI presents a framework for understanding how symptoms and emotions experienced during a health threat or diagnosis combine to influence one's perception of illness which, in turn, influences subsequent coping behavior (Buunk & Gibbons, 2013; Diefenbach & Leventhal, 1996; Leventhal, Meyer, & Nerenz, 1980; Petrie & Weinman, 1997). The model proposes that one's cognitive perception of an illness threat consists of five dimensions, or subconstructs, including Identity, Cause, Consequences, Timeline and Cure/Control. This multidimensional illness representation then combines with one's emotional response to influence one's chosen coping strategy for that illness. The coping strategies, as laid out in the theory, have two dimensions: approach coping such as taking medication, seeing a doctor, or quitting smoking and avoidance coping such as denial, wishful thinking and ignoring advice to quit smoking. It is these coping strategies which then influence outcome appraisals (Leventhal et al., 1980) with the continual back and forth between the illness representations, chosen coping strategies and self-appraisals of resulting outcomes creating the titular self-regulatory mechanism of the model.

The SRMI may be a useful tool for understanding why individuals diagnosed with COPD continue to smoke as it has been successfully used to model number of illness-related coping behaviors including returning to work after coronary heart disease (Petrie et al., 1996), human immunodeficiency syndrome (HIV) medication adherence (Reynolds, 2003), diabetes self-management (Lange & Piette, 2006), and smoking among lung cancer patients (Browning, Wewers, Ferketich, Otterson, & Reynolds, 2009). A meta-analysis of the relationships between components of the model found significant associations between a number of the cognitive

illness dimensions and certain coping strategies as well as between certain illness representations and health outcomes (Hagger & Orbell, 2003).

The guiding principle of the SRMI for this research is the multidimensional factoring of the five dimensions of illness representation that leads a person to choose a coping strategy for his/her illness. The model posits that the label one gives to one's illness, one's perceived cause of the illness, how long one believes the illness will last, the consequences or effects of the illness on one's life, and whether one believes the outcome of the illness is controllable will determines one's illness representation (Browning et al., 2009). It is this multidimensional illness representation combined with one's emotional response to the health threat of the illness that determines one's coping strategy for the illness and whether it is an approach coping strategy or an avoidance strategy. Specifically, as it relates to this research project, it is proposed that specific combinations of these dimensions of illness representation and emotional response will combine to form a number of COPD phenotypes and that these phenotypes will then differentially influence smoking cessation (one form of approach strategy), or continued smoking (one form of avoidance strategy), as a chosen coping strategy for smokers suffering from COPD.

1.4.2 Yerkes-Dodson Law (Inverted U Theory)

The Yerkes-Dodson Law is a psychological theory that suggests a quadratic, or inverse U-shaped, relationship between performance and arousal. The law posits that increased arousal can improve performance up to an inflection, or turning, point. At this inflection point, the arousal essentially becomes excessive and performance subsequently begins to diminish.

Two psychologists, Robert Yerkes and John Dillingham Dodson, first described the Inverted U theory in 1908 (Yerkes & Dodson, 1908). They discovered that the application of

mild electrical shocks improved the performance of rats navigating through a maze. However, beyond a certain level of electrical current, the rats' performance changed direction and actually decreased with additional current.

While originally developed as a psychological theory, Calabrese et al recently posited that the dose-response nature of the theory has broad applicability over a wide range of sub-disciplines (Calabrese et al., 2007). As they noted. "A remarkable feature is that this broad spectrum of sub-discipline specific biphasic dose–response relationships most likely represents a broadly generalizable and fundamental dose–response relationship that is independent of biological model, endpoint measure and chemical/physical stressor agent."

Demonstrating the applicability of the Yerkes-Dodson law to health behavior and psychosocial characteristics, a recent study examining weight dissatisfaction among children revealed significant non-linear differences in weight loss for different levels of weight dissatisfaction. Specifically, the children who were only moderately dissatisfied with their weight lost weight during the study while the children in both the low and the high weight dissatisfaction groups not only did not lose weight, they actually gained weight. As the authors of the study publication noted, their findings suggest that only "moderate levels of weight dissatisfaction are associated with improved outcomes in a weight management program." (Johnston, Moreno, Regas, Tyler, & Foreyt, 2012).

Applying this theoretical construct to smoking cessation among persons with COPD, it reasonable to believe that smoking cessation, or its inverse continued smoking, might follow this trajectory, as well. That is, as severity increases, motivation to quit smoking also increases but only to a point. As both the difference in smoking rates by disease severity and the qualitative research suggest ("it's not worth stopping now") , there may exist a point in the disease severity

continuum at which smokers with COPD are no longer motivated to quit smoking and smoking behavior subsequently increases for those who have crossed this severity threshold. At this hypothetical inflection point, they may perceive their disease to be too severe for quitting to have any real, or lasting, effects. This theory, when combined with the recent smoking prevalence data among those with lung obstruction and qualitative research findings, supports the hypothesis that disease severity will moderate the relationships between factors predictive of continued smoking and the actual behavior itself.

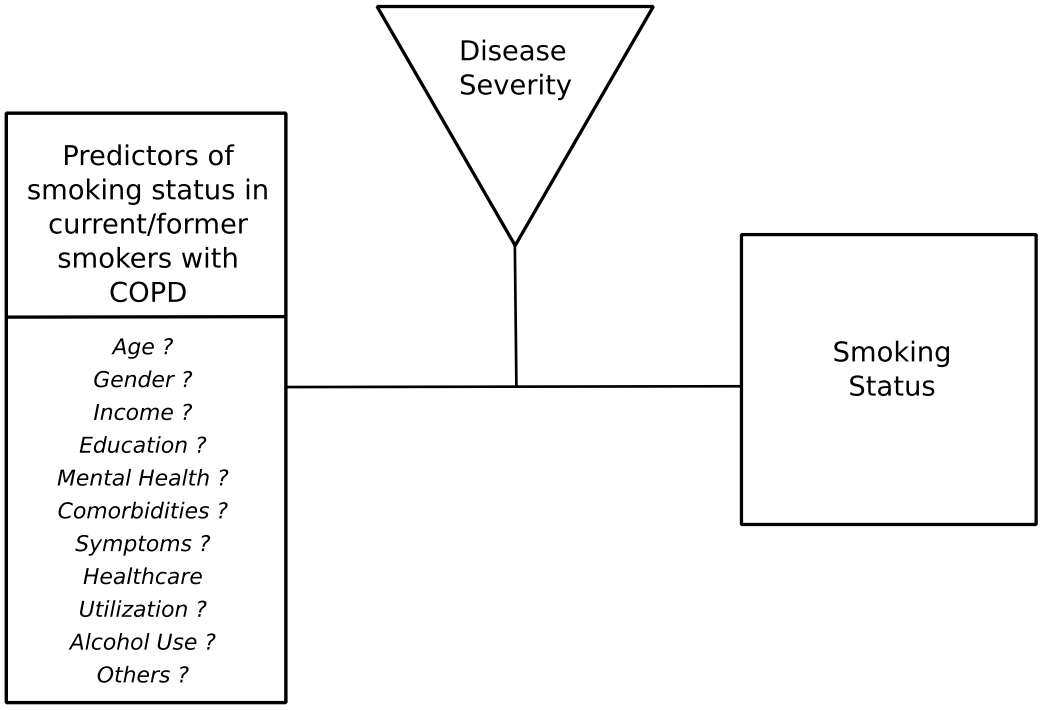
1.5 Overall Conceptual Framework

This research project essentially consisted of three parts, each with its own resulting journal manuscript.

The first part of this research project was atheoretical and was focused on determining rates of awareness (via diagnosis) of having the disease, rates of smoking among persons with diagnosed COPD, and characteristics of smokers and former smokers diagnosed with COPD both overall and, to help set up the predictor and moderation analyses, by severity of lung obstruction.

The second part of this research project, guided by the Yerkes-Dodson Law, consisted of modeling the likelihood of being a current smoker with COPD among all lifetime smokers with COPD (current and former) and assessing whether disease severity moderated the relationships between associated factors and being a current smoker. Figure 1.3.1 shows the conceptual model to test the hypothesis that lung obstruction severity will moderate the relationships between factors associated with smoking status and actual smoking status among persons with COPD.

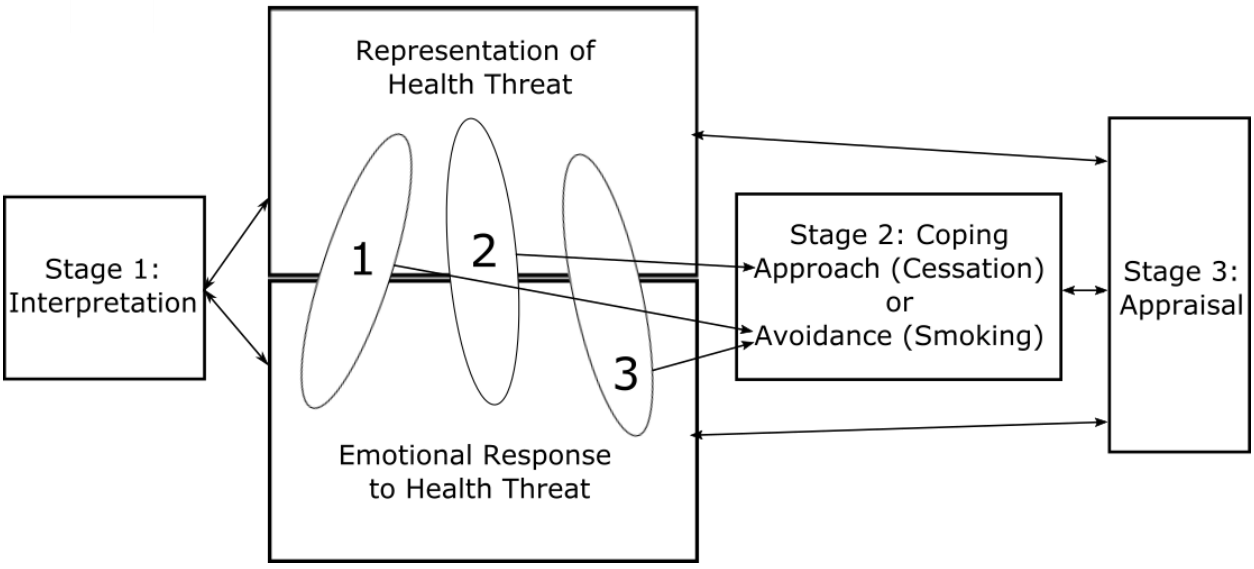
Figure 1.5.1 Theorized model depicting disease severity moderating the predictors of smoking status among current and former smokers with COPD



The third, and final, part of this research project, guided by the SRMI, consisted of determining the predominant phenotypes of COPD, comparing these known phenotypes with those determined in previous studies, and assessing whether one or more of these phenotypes are predictive of smoking status among current and former smokers with COPD, controlling for other factors. Figure 1.3.2 shows the conceptual model to test the hypothesis that COPD

phenotypes will differentially predict smoking status among current and former smokers with COPD, above and beyond other factors.

Figure 1.5.2 Theorized model depicting COPD phenotypes differentially predicting smoking status among current and former smokers with COPD (adapted from Leventhal's Self-Regulating Model of Illness)



1.6 Study Goal, Research Questions, and Specific Aims

The overarching goal of the research in this dissertation was to assess the roles that disease severity and COPD phenotype play in smoking status among persons with COPD. In pursuit of this goal, the studies presented here attempted to answer a number of research questions:

- What percent of the US adult population aged 40-79 years has been diagnosed with COPD?
- Among adults aged 40-79 years with diagnosed COPD, what percent are current smokers, former smokers and never smokers?

- How do the characteristics of adults reporting a COPD diagnosis differ from the characteristics of adults not reporting a COPD diagnosis?
- Among adults diagnosed with COPD, does smoking status differ by lung obstruction severity?
- Among ever smokers with diagnosed COPD, what factors are associated with current smoking status?
- Does lung obstruction severity moderate the relationships between the factors associated with smoking status and being a current smoker?
- What are the predominant phenotypes of COPD?
- Are the identified phenotypes consistent with, or different from, those previously found in other studies?
- Above and beyond other factors, are disease phenotypes predictive of smoking status among those with diagnosed COPD?

Predicated on these research questions, the specific aims of this investigation were to:

1. Determine the proportion of adults who have been diagnosed with COPD, smoking status proportions among those diagnosed, and factors predicting receipt of a COPD diagnosis.
2. Compare the biologic, behavioral, and psychosocial characteristics of adults reporting a diagnosis of COPD to those of adults not reporting a COPD diagnosis.
3. Assess whether smoking status differs by lung obstruction severity, determine the factors associated with current smoking among ever smokers with COPD and assess the moderating effect of disease severity on these predictive factors.

4. Determine phenotypes of COPD, compare/contrast findings to phenotypes found in prior studies, and assess whether one or more phenotypes are differentially associated with current smoking status among ever smokers with diagnosed COPD.

CHAPTER II.

LITERATURE REVIEW

2.1 Chronic Obstructive Pulmonary Disease

2.1.1 Definition

The definition of COPD has evolved substantially over the last 25 years. One of the earliest definitions was developed by the American Thoracic Society (ATS) in 1995 and formed the basis for the heuristic that COPD consists of emphysema and chronic bronchitis: "Chronic obstructive pulmonary disease is defined as a disease state characterized by the presence of airflow obstruction caused by chronic bronchitis or emphysema; the airflow obstruction is generally progressive, may be accompanied by airway hyper reactivity, and may be partially reversible." (American Thoracic Society, 1995).

In 2001, the Global Initiative for Chronic Obstructive Lung Diseases (GOLD) was formed and issued the first in their ongoing series of documents titled "Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease." In this first document, they define COPD as "a disease state characterized by airflow limitation that is not fully reversible. The airflow obstruction is usually both progressive and associated with an abnormal response of the lungs to noxious particles or gases." (Pauwels et al., 2001). The

definition was the first to reference the disease not being fully reversible and the first to declare that the disease is associated with exposure to noxious particles and gases, e.g., tobacco smoke.

Each of the most recent definitions by both the American Thoracic Society/European Respiratory Society (ATS/ERS) and GOLD committee references the usually progressive nature of the disease, inflammatory response in the lungs (and airways for GOLD), systemic consequences (exacerbations and comorbidities by GOLD), and the preventability and treatability of the disease. One renowned author in the field of COPD believes this statement about preventability and treatability was added by both respiratory groups in "an attempt to leave previous therapeutic nihilism regarding COPD behind". (Jørgen Vestbo, 2014). The most recent definitions of COPD by both groups are listed below:

ATS/ERS

"Chronic obstructive pulmonary disease (COPD) is a preventable and treatable disease state characterized by airflow limitation that is not fully reversible. The airflow limitation is usually progressive and is associated with an abnormal inflammatory response of the lungs to noxious particles or gases, primarily caused by cigarette smoking. Although COPD affects the lungs, it also produces significant systemic consequences." (Celli et al., 2004)

GOLD

"Chronic Obstructive Pulmonary Disease (COPD) is a common, preventable and treatable disease that is characterized by persistent respiratory symptoms and airflow limitation that is due to airway and/or alveolar abnormalities usually caused by significant exposure to noxious particles or gases." (GOLD, 2018)

2.1.2 Burden of Disease

COPD is estimated to be the third leading cause of death both in the U.S. (Murphy et al., 2013) and around the globe (WHO, 2016) and is a growing cause of suffering, morbidity and mortality worldwide. The prevalence in the U.S. has been estimated to range between 10% and 20% depending on the diagnostic criteria and methods used (Tilert, Dillon, Paulose-Ram, Hnizdo, & Doney, 2013). It is thought that the prevalence, and subsequent burden, of COPD will increase through the year 2030 as a result of continued exposures to COPD risk factors and increased life expectancy worldwide (Mathers & Loncar, 2006).

In addition to suffering, morbidity and mortality, there is also a massive economic burden associated with COPD. The health system costs of COPD in the U.S. are enormous with an estimated $30 billion in direct costs in 2010 and an additional $20 billion in indirect costs (GOLD, 2018; Guarascio et al., 2013). The burden is not limited to those who suffer from the disease but extends to family members or other caretakers, as well. Estimates of direct costs for home care are likely underrepresented as these estimates often do not include the care provided to COPD patients by family members and other unpaid caregivers. Indirect estimates are likely underreported, as well, as it is difficult to measure the true effects of lost income, reduced productivity, etc. on family members who quit their jobs or reduced their work hours to care for sick relatives. In developing countries, where human capital is frequently a critical national asset, it is thought that the indirect costs of COPD may pose a serious threat not just to individuals but to national economies, as well (GOLD, 2018).

2.1.3 Risk Factors

Tobacco Smoke

Many epidemiological studies conducted across the world have shown that cigarette smoking is the most important risk factor for COPD. Compared to non-smokers, cigarette smokers have more respiratory symptoms, a greater rate of decline in FEV1, higher utilization of health care and a greater death rate from COPD (Kohansal et al., 2009). In addition to cigarettes, other types of tobacco (e.g., water pipe, marijuana, cigar) have been shown to be risk factors for COPD, as well (Raad et al., 2011; Tan et al., 2009; Tetrault et al., 2007). The age one began smoking, number of pack-years smoked, and current smoking status are all factors predictive of COPD mortality (Pauwels et al., 2001). Passive cigarette smoke exposure (second-hand smoke) could also contribute to respiratory symptoms via increased exposure to inhaled particles and fumes (Eisner et al., 2005) though the effects on lung function are considered too small to have significant impact on the development of COPD (Pauwels et al., 2001).

Environmental Exposures

One COPD risk factor that is more prevalent in less developed countries is the exposure to ambient air pollution from cooking and heating with coal or biomass fuel indoors. It is estimated that nearly 3 billion people around the world use coal or biomass fuel as their primary energy source for both cooking and heating which places a very large population at risk globally (Orozco-Levi et al., 2006; Torres-Duque et al., 2008). High levels of environmental air pollution may also be harmful to one's respiratory health though the role outdoor air pollution plays in the development of COPD is still unclear (GOLD, 2018).

Occupational exposures

Occupational exposures to inorganic and organic dusts, noxious fumes and chemical agents may be an additional risk factor for COPD with both inorganic and organic dust exposure having been shown to accelerate the decline in FEV1 (Vogelzang et al., 1998). A 2002 analysis

of NHANES III data estimated the proportion of COPD attributable to occupational exposures to be 19.2% overall and 31.1% among never-smokers (Hnizdo, Sullivan, Bang, & Wagner, 2002). These estimates are in line with an American Thoracic Society statement concluding that "occupational exposures account for 10-20% of either symptoms or functional impairment consistent with COPD" (Balmes, Becklake, & Blanc, 2003). Workers in parts of the world with less occupational safety oversight likely have even higher risk for COPD from occupational exposures than workers in highly regulated areas like Europe and North America (GOLD, 2018).

Genetic factors

Though there is a strong association between cigarette smoke exposure and developing COPD, there remains significant variability in susceptibility to the disease with some heavy smokers never developing the disease. This may be explained, in part, by the complex gene-environment interaction. Alpha-1 antitrypsin deficiency is just one example of a known genetic risk factor for COPD (Stoller & Aboussouan, 2005). Though alpha-1 antitrypsin deficiency is rare, its association with COPD illustrates the interplay between genetics and environmental exposures in the development of COPD (GOLD, 2018). A number of studies have found associations between COPD and other genetic representations but results have not been consistent across populations (Pauwels et al., 2001).

2.1.4 Diagnosis

It is recommended that "a clinical diagnosis of COPD be considered in any patient aged 40 years or older with dyspnea, chronic cough or sputum production and a history of exposure to risk factors (smoke, dust, chemicals, air pollution, family history of COPD, etc.)" (GOLD, 2018). Spirometry is used to measure pulmonary function and is required for a COPD diagnosis.

GOLD suggests that a post-bronchodilator FEV1/FVC ratio less than 0.7 is indicative of airflow limitation (GOLD, 2018). Use of a fixed ratio, however, has been shown to overdiagnose in the elderly and underdiagnose in those under 45 years old when compared to using FEV1/FVC cutoffs based on predicted population values employing age, gender and height (Cerveri et al., 2008; Hansen, Sun, & Wasserman, 2007; Hardie et al., 2002). In addition to spirometric assessments, a diagnosis of COPD should be also be based on a patient's symptoms, risk of future exacerbations and comorbid conditions (GOLD, 2018).

Specifically as it relates to smoking cessation, we know that COPD is an underdiagnosed disease (Lindberg et al., 2006b; Mannino, Homa, Akinbami, Ford, & Redd, 2002; Viegi et al., 2007) but obtaining the COPD diagnosis is critical as smoking cessation has been shown to be more common among those with a diagnosis by itself (Stratelis et al., 2006) and when combined with physician advice to quit smoking (Górecka et al., 2003).

2.1.5 Heterogeneity of Disease

COPD is a multicomponent disease with effects on both the lungs and organs outside the lungs and can be of a primarily structural (including airway remodeling or parenchymal destruction) and/or functional nature (allergic reactions and inflammation). Though all COPD patients share some clinical characteristics, most notably airflow obstruction that is not fully reversible, COPD is a very heterogeneous disease with disease susceptibility, responses to treatment, and clinical presentation varying widely by patient, often within the same GOLD stage or level of airway obstruction.

Heterogeneity in Susceptibility

Though tobacco smoking is the primary risk factor for COPD, not all smokers go on to develop the disease (Løkke et al., 2006; Rennard & Vestbo, 2006). Even among smokers with diagnosed COPD, the relationship between self-reported smoking exposure and airflow limitation is poor, suggesting that susceptibility to COPD is not a simple, binary phenomenon (Agusti et al., 2010). In conjunction with smoking, there may be genetic differences in combination with other risk factors (e.g., illness, nutrition, other environmental exposures) that jointly determine susceptibility to COPD as opposed to smoking or genetic predisposition solely determining susceptibility to the disease.

Heterogeneity in Responses to treatment

Until recently, disease severity as described by airflow limitation has largely driven COPD treatment recommendations. The latest GOLD strategy document no longer recommends treatment solely based on disease severity and now recommends inclusion of symptoms, exacerbation risks and comorbidities when deciding on a treatment strategy (GOLD, 2018). This is supported by evidence of patients responding differently to a number of different treatment options (Decramer & Janssens, 2013). One example is roflumilast, an anti-inflammatory drug, which has been shown to be effective for COPD patients with chronic bronchitis but ineffective in individuals with more emphysematic COPD (Rennard, Calverley, Goehring, Bredenbröker, & Martinez, 2011). Response differences among COPD patients have been reported for other drugs, as well (Disantostefano, Li, Rubin, & Stempel, 2013; Han et al., 2010). As Miravitlles et al. noted, “the identification of patients who respond to different types of treatment is a first step toward more personalized treatment of COPD” (Marc Miravitlles, Calle, & Soler-Cataluña, 2012).

Heterogeneity in Presentation and Outcomes

COPD is often characterized by an assortment of concurrent lung and systemic indications. Though airflow limitation is the characteristic that defines the existence (and stage) of the disease, this measurement does not always correlate well with a number of other disease indicators such as radiographic features, prevalence of comorbid conditions, symptoms and exercise tolerance (Han et al., 2010; Vermylen & Kalhan, 2013). Describing results from the ECLIPSE study, Agusti et al. noted that within each stage of disease severity, prevalence of co-morbidities, symptoms, exercise tolerance and the number of reported exacerbations varied greatly between patients (Agusti et al., 2010). Additionally, they found that even among patients with severe disease assessed via airflow limitation, a substantial number did not report exacerbations, complain of symptoms or show impaired exercise tolerance.

These differences highlight the complexity and heterogeneity of COPD and support the latest GOLD recommendation that the treatment of patients with COPD should be based on multiple disease indicators (e.g., symptoms, exacerbation risks and comorbidities) rather than on spirometry results alone.

2.1.6 Disease Severity Based on Airflow Obstruction

Both the GOLD and the ATS/ERS guidelines have developed severity categories of airflow limitation based on the spirometrically assessed forced expiratory volume in 1 second (FEV1) values (Tables 2.1.6.1 and 2.1.6.2)

Table 2.1.6.1 Classification of Severity of Airflow Limitation (ATS/ERS guidelines)

Degree of Severity	***FEV1 %Predicted***
Mild	> 70
Moderate	60-69
Moderately severe	50-59
Severe	35-49
Very severe	< 35

Table 2.1.6.2 Classification of Severity of Airflow Limitation (GOLD guidelines)

Degree of Severity – Grading	***FEV1 %Predicted***
Mild	>= 80
Moderate	>= 50 and < 80
Severe	>= 30 and < 50
Very severe	< 30

One potential limitation of these spirometrically determined severity scales is that both of these classification systems assume lung obstruction (FEV1/FVC ratio below a predetermined threshold) before even grading severity. A number of studies refer to a group that is not classified per these systems but still has suboptimal lung performance; that is those who have a normal FEV1/FVC ratio but below normal FEV1 values (Köhler, Fischer, Raschke, & Schönhofer, 2003; Mannino et al., 2002) which is often referred to as "restrictive" lung disease.

The GOLD committee notes that these airflow limitation cutoffs referenced in their guidelines have not been clinically validated and they further acknowledge that COPD patients interactions with health care providers are largely driven by symptoms and other complications (GOLD, 2018). It has been suggested that management decisions in a clinical setting should be based on symptomatic responses to treatment rather than changes in spirometry that poorly predict symptomatic responsiveness (Jones, 2005; O'Donnell et al., 2007).

From a behavioral perspective, severity most certainly plays a role in one's decisions to cope with illness behaviors including seeking care, treatment adherence and, in the case of COPD, smoking cessation. It is possible that those who are in favor of airflow defined severity and those who are opposed to it are both right, each to an extent. That is, there may be certain outcomes, or specific situations, when airflow defined severity can be perfectly useful or predictive. In other cases, due to the heterogeneity and multi-systemic nature of the disease,

airflow limitation alone may not capture all the dimensions of disease severity and may be too simplistic a measure.

2.2 Smoking and COPD

2.2.1 Impact of Smoking Cessation on COPD

While there presently is no cure for COPD, smoking cessation has been shown to improve respiratory symptoms, reduce the rate of lung function decline, lower exacerbation risks, and improve the health-related quality of life for those with the disease (Anthonisen et al., 1994, 2002; de Granda-Orive & Martínez-Albiach, 2005; GOLD, 2018; Wagena, Knipschild, Huibers, Wouters, & van Schayck, 2005). In terms of lower exacerbation risks, results from Au et al. showed that reduction in risk for exacerbations was directly related to the duration of smoking abstinence (Au et al., 2009). If the smoking cessation occurs while the person is still in one of the early disease stages, the cessation could reduce the decline in FEV1 to that observed in healthy subjects (Anthonisen et al., 1994; Burchfiel et al., 1995).

In an analysis of the Lung Health Study, a 5-year, prospective randomized clinical trial of a smoking cessation intervention on nearly 4,000 smokers having mild to moderate airway obstruction, it was found that lung function, measured as percent of predicted FEV1, actually improved during the first year for quitters while lung function declined for those who continued smoking (Scanlon et al., 2000). Lung function declined for both groups from year 2 through year 5 but the rate of decline was twice as rapid for continuing smokers than it was for sustained quitters. Consistently, quitters who relapsed into smoking lost lung function and those who delayed quitting got a boost in lung function regardless of when they quit. In a separate analysis of Lung Health Study data, the smokers with early stage COPD in the smoking-cessation

intervention group also reported significantly fewer respiratory symptoms after 5 years than those assigned to the control group (Kanner, Connett, Williams, & Buist, 1999).

2.2.2 Smoking in Persons with COPD

In 2011, the Behavioral Risk Factor Surveillance System (BRFSS) administered an optional smoking cessation module, in addition to their main survey, to obtain additional information on smoking, chronic disease status and smoking cessation attempts. An analysis of the data collected from the five states where the optional module was administered showed significantly higher proportions of current smoking among those with COPD than among those with asthma, those with other chronic conditions and those with no chronic conditions (47.3%, 23.1%, 28.8% and 20.0%, respectively) (Schauer et al., 2014). The research also showed that current smokers with COPD were more likely older, female, of lower socio-economic status (SES), and unable to work than those with other chronic conditions and those with no chronic conditions. With respect to smoking quit attempts, about 40% of the smokers with COPD made no attempt to quit smoking in the prior year, which was statistically similar to the percentages of smokers who made no attempts to quit smoking in the other disease and non-disease populations.

Similar to the 47.3% estimate using BRFSS data, recent NHANES data showed that 46.2% of adults aged 40-79 years with lung obstruction are current smokers (Paulose-Ram et al., 2015). Though the population used had measured lung obstruction and were not confirmed with a COPD diagnosis, it is thought, due the advanced age of the sample population, that the bulk of this population is afflicted with COPD and not asthma. The smoking proportion of 46% found in the NHANES lung obstruction sample is similar to the smoking proportion of 47.3% found in the BRFSS sample by Schauer et al. The NHANES data further revealed that a significantly higher

proportion of persons with moderate or more severe lung obstruction currently smoke (56%) than do persons with mild lung obstruction (41%). These are the first known published rates of smoking that compared groups with differing levels of severity of lung obstruction. The results from these comparisons suggest that among adults aged 40-79 years, those with more severe lung obstruction are more likely to smoke than those with less severe lung obstruction. Additionally, this reported difference in smoking behavior by level of severity is one of the motivating factors for attempting to determine whether disease severity moderates relationships between factors predictive of smoking and current smoking status.

2.2.3 Racial and Gender Differences in Susceptibility to Smoke

In a retrospective record review of 160 patients with advanced COPD, Chatila et al found that among those with comparable lung function levels, African Americans presented with disease at younger ages and had lower lifetime cumulative smoking (measured in overall smoking pack-years) than whites (Chatila, Wynkoop, Vance, & Criner, 2004). Similarly, women with advanced COPD were younger at presentation and smoked less when compared to men, again without any differences in lung function or exercise performance. These differences in susceptibly by race and gender were echoed in a study conducted by Dransfield et al. on COPD patients over the age of 45, with at least 20 pack-years of smoking. Their results showed that African Americans had greater lung function loss per pack-year smoked than Caucasians and, similarly, women had greater lung function loss per pack-year smoked than men (Dransfield, Davis, Gerald, & Bailey, 2006). This research suggests that the damaging effects of smoking may not be equally distributed for all people. More specifically, these studies have shown that African Americans and women, in particular, may have more sensitivity to the deleterious effects

of tobacco smoke than their counterparts. The significance of these findings as they relate to the research conducted here is that persons of different races and genders may differentially be affected by smoking and, subsequently, may make decisions to continue or stop smoking as a result of the harmful effects imparted upon them.

2.3 Factors Associated with Smoking and Smoking Cessation in COPD

2.3.1 Clinical and biological factors

A number of clinical and biological factors have been identified as predictors of smoking cessation among patients with COPD. Symptoms and smoking levels have been shown to influence patient motivation to quit. Specifically, Hilberink et al. found that compared to patients who wanted to quit smoking in the next six months, patients who wanted to quit smoking in the next 30 days had experienced more severe symptoms (Hilberink, Jacobs, Schlösser, Grol, & Vries, 2006). Another study showed, however, that unpleasant respiratory symptoms alone were insufficient motivation for smokers to successfully achieve smoking cessation (J. S. Wilson, Elborn, & Fitzsimons, 2011).

Christenhusz et al. found that having high levels of salivary cotinine (a nicotine metabolite) was a significant predictor of continuous smoking abstinence in COPD patients after one year (Christenhusz, Pieterse, Seydel, & van der Palen, 2007). High levels of nicotine dependence are often used to explain why smoking cessation can be so difficult to achieve (Davila et al., 2009; Fagerström et al., 1996). However, nicotine dependence has not always been shown to reliably predict smoking cessation success. One study found that quitters with COPD with higher levels of nicotine dependence were less likely to remain abstinent for one year (Coronini-Cronberg et al., 2011) while another study found no association between nicotine

dependence and smoking cessation success among participants with mild to moderate COPD (Tashkin et al., 2011).

In a study on characteristics associated with sustained abstinence among male Veterans Administration (VA) enrollees with COPD, Adams et al showed that specific comorbidities were significantly associated with smoking cessation (Adams, Pugh, Kazis, Lee, & Anzueto, 2006). Their results showed that former smokers with COPD had more cardiac comorbidities, in particular, than current smokers. The authors pointed out that these differences might be explained in part by the lower mean age of the current smokers.

To date, the majority of smoking cessation interventions targeted at those with COPD have focused on a specific range of disease severity and have left unexplored differences in severity level of disease on intervention success. The Jiménez Ruiz et al. study was limited to those with severe or very severe COPD (Jiménez Ruiz et al., 2012); the Anthonisen et al. study was limited to those with mild COPD (Anthonisen et al., 1994); the Crowley et al. study was limited to those with severe COPD (Crowley, Macdonald, & Walter, 1995); two studies by Tashkin et al. were both limited to mild or moderate COPD (Tashkin et al., 2001, 2011). A number of other interventions and observational studies utilized participants with COPD without regard to severity and did not distinguish, or compare effectiveness of the intervention, between severity levels (Gratziou et al., 2014; Hilberink, Jacobs, Bottema, de Vries, & Grol, 2005; Pederson, Wanklin, & Lefcoe, 1991; Schauer et al., 2014; Tønnesen, Mikkelsen, & Bremann, 2006; Wagena et al., 2005). Two studies were found that drew comparisons in intervention success rates between groups with differing levels of COPD severity and found inconsistent, and somewhat contradictory, results (Górecka et al., 2003; Tashkin et al., 2011). In their intervention assessing the impact of airflow limitation diagnosis combined with cessation advice, Górecka et

al found that more severe COPD predicted higher smoking cessation rates (Górecka et al., 2003). In their pharmacologically-based intervention, however, Tashkin et al found more severe COPD predicted lower smoking cessation rates, though too few participants with severe disease were studied to allow any firm conclusions (Tashkin et al., 2011). One plausible explanation for the inconsistency is that there may be a point in the severity spectrum where sufferers of COPD feel that quitting smoking will offer no real benefit. Among a small set of semi-structured interviews of smokers with COPD of all severities conducted in the United Kingdom, one theme that emerged was that that participants felt that their disease was too advanced for them to quit smoking, "it's not worth stopping now" (J. S. Wilson et al., 2011).

2.3.2 Behavioral Factors

The impact of various aspects of mental health on smoking cessation among smokers with COPD has been highlighted in a number of studies. As Adams et al pointed out, a significant proportion of COPD patients have depression and anxiety (Adams et al., 2006). They also found that former male VA smokers with COPD had better perceived mental health than their counterparts who were current smokers. Anxiety or worries about future health problems have also been reported to motivate some smokers to achieve smoking cessation (Vangeli & West, 2008; Wells & de Lusignan, 2003). For some smokers, depression was associated with the re-initiation of smoking after a period of cessation (Tomson, Toftgård, Gilljam, & Helgason, 2006). Christenhusz et al. also showed inversely that a positive attitude towards smoking cessation was a significant predictor of continuous smoking abstinence after one year (Christenhusz et al., 2007). A qualitative study conducted by Hilberink et al. highlighted the importance of a smoker's level of motivation to quit. Their analysis showed that COPD patients

who were motivated to quit had identified more coping strategies to resist smoking under stressful situations (Hilberink et al., 2005).

2.3.3 Psychosocial Factors

Studies have shown that behavioral support directly from providers or as part the smoking cessation program can increase quit rates. Quist-Paulsen et al conducted a Medline search and found that patients with cardiovascular diseases and COPD had increased quit rates when their smoking cessation programs were combined with behavioral support (Quist-Paulsen, 2008). Adams et al found that former male VA smokers with COPD had better perceptions of their interactions with their health care providers than those of current male VA smokers (Adams et al., 2006). In their study exploring health beliefs related to smoking among older people with COPD, Schofield et al. found two common barriers to smoking cessation in this population to be patient misinformation and poor perceived communication with healthcare professionals (Schofield, Kerr, & Tolson, 2007).

2.3.4 Pharmacological Treatment combined with Behavioral Therapy

A number of recent meta-analyses have shown that smoking cessation is more effective when behavioral counseling is provided jointly with some form of pharmacological treatment. Hoogendoorn et al. recently conducted a systematic review of smoking cessation interventions in patients with COPD. Their review had the specific requirements that each intervention had to be a randomized controlled trial and each intervention also had to have a minimum 12-month follow-up period on abstinence rates (Hoogendoorn, Feenstra, Hoogenveen, & Rutten-van Mölken, 2010). Their analysis covered nine studies and showed that intensive counseling

combined with pharmacotherapy achieved a 12.3% continuous year-long abstinence rate and was the most effective smoking cessation intervention strategy when compared to usual care (1.4%), minimal counseling (2.6%) or intensive counseling alone (6.0%). Strassman et al. conducted a network meta-analysis of eight randomized trials of smoking cessation among COPD patients (Strassmann et al., 2009). They found the most effective intervention to be counseling in combination with nicotine replacement therapy (NRT) with the second most effective intervention being counseling in combination with an antidepressant. Though one method involved the use of NRT and the other an antidepressant, the most effective strategy when looking at the two methods combined was counseling combined with some form of pharmacotherapy. A systematic literature review conducted by Coronini-Cronberg et al. determined that the use of the pharmacological agent Bupropion combined with counseling was the most effective intervention of those studied and was approximately 19% more effective than placebo in achieving prolonged smoking abstinence (Coronini-Cronberg et al., 2011).

In each of these comprehensive reviews, it was determined that the most effective approach for smoking cessation in persons with COPD was counseling or behavioral therapy in combination with some form of pharmacological agent. In the absence of more effective pharmacologic interventions, it would seem then that more tailored and informed counseling and behavioral therapy programs may currently be the best way to increase rates of smoking cessation in this high-risk population.

2.4 Phenotyping and COPD

2.4.1 What is Phenotyping?

Per Merriam-Webster, a phenotype is defined as "the observable properties of an organism that are produced by the interaction of the genotype and the environment" (Merriam & Webster, 2019). Han et al proposed the following variation on this definition specifically as it relates to COPD: "a single or combination of disease attributes that describe differences between individuals with COPD as they relate to clinically meaningful outcomes (symptoms, exacerbations, response to therapy, rate of disease progression, or death)" (Han et al., 2010). As applied to this research, this more focused definition of a COPD phenotype suggests classifying COPD patients into different subgroups, or phenotypes, with a distinct focus on meaningful outcomes, one of which could be smoking cessation or its inverse, continued smoking.

The method for identifying phenotypes usually entails the use of statistical combinatorial techniques such as factor analysis or the technique used here, cluster analysis. The aim of cluster analysis is to assign observations to groups where, using distance measures, observations in the same group are closer to the other observations from the same group than they are to another group's observations. Applying Han's definition to the research here, these clusters (or phenotypes) should then allow us to differentiate between groups of individuals with COPD with respect to clinically meaningful outcomes. Specifically, for this research, the clinically meaningful outcome to be assessed is current smoking status.

2.4.2 Need for Phenotyping with COPD

Airflow obstruction severity, as measured by FEV1, is required for the staging of COPD with the rate of FEV1 decline serving as a useful measure of disease progression (Adis Medical Writers, 2015). However, FEV1 alone does not sufficiently reflect the myriad properties of COPD, and may particularly be inadequate for prediction of disease outcomes. A multitude of

studies have shown that FEV1 does not always correlate well with a number of things salient to patients such as quality of life (Jones, Quirk, Baveystock, & Littlejohns, 1992), exacerbation frequency (Alsaeedi, Sin, & McAlister, 2002), and exercise intolerance (O'Donnell, Lam, & Webb, 1999). FEV1 is also not a good surrogate for symptoms and, in some studies, has actually correlated weakly with symptoms (Mahler et al., 1992; Viegi et al., 2007). Another study has shown that respiratory symptoms alone are predictive of hospitalizations for both general respiratory disease and for COPD, even after controlling for FEV1 (J. Vestbo & Rasmussen, 1989). The consistent message across all these findings is that FEV1 alone may not be the best tool for predicting outcomes among those suffering from COPD and other factors should be considered. Support for such a multi-factored approach is alluded to in the conclusion of a recent official ATS-ERS workshop report on integrating multiple comorbidities in guideline development, "Furthermore, genetic, physiological, disease-related, and *other factors* that may influence the directness (applicability) of the evidence for the target population in clinical practice guidelines should be examined." (Fabbri et al., 2012).

Expanding on this premise that COPD is a complex, multifaceted disease and that severity as defined by FEV1 alone is insufficient to describe the heterogeneity of the disease, identifying phenotypes of COPD with the goal of individualized treatment has gained substantial research interest (Pinto et al., 2015). It is vital to classify patients using a multidimensional approach, not simply FEV1, as individuals sharing the same phenotype are "likely to have a similar progression of disease and response to treatments" (Jørgen Vestbo, 2014).

2.4.3 COPD Phenotyping Studies and Results

A systematic review was recently conducted of COPD phenotyping studies that used unsupervised statistical analyses to derive hypothesis-free phenotypes and that were associated with clinically relevant outcomes (Pinto et al., 2015). In addition to the criteria noted above, only studies with a minimum of 50 participants, 18 years and older, were reviewed. This combination of inclusion criteria yielded eight studies.

All of the eight studies were conducted in either Europe or the United States. The sample sizes used in the studies ranged from 213 to 1543 subjects with a median of 332 subjects. Seven of the eight studies described the research setting. One study recruited a subset of its participants from a lung cancer screening study conducted in local communities; the other six remaining studies were all conducted in university-based health care settings.

There was substantial heterogeneity in the participant characteristics across studies. Two of the studies only included severe COPD patients while four studies excluded mild (GOLD Stage I) COPD patients altogether. The study samples were also heavily biased towards male participants with the proportion of female participants ranging from a low of 7% to a high of 46%.

There was significant variability in the methods employed for the cluster analyses, as well. Half of the studies used a variable reduction strategy before clustering: two used principal components analysis alone; one used principal components analysis along with multiple component analysis; and one used factor analysis alone. For identification of phenotypes, five of the studies used the k-means clustering technique, one used the Wards method, one used tree-based supervised clustering, and one used self-organizing (or Kohonen) maps followed by the Wards method.

The number of phenotypes derived in the eight studies ranged from two to five. Five of the studies identified a phenotype of younger persons with severe respiratory disease along with poor health status, poor nutritional status and low cardiac co-morbidity prevalence. A similar phenotype was identified in two other studies with the singular exception that the individuals were not substantially younger. When validated against the selected outcomes, those with this phenotype had poor results throughout all the studies. In all eight studies, researchers also identified a phenotype of older people with mild respiratory disease and high obesity prevalence. Among subjects with this phenotype, the prevalence of cardiac and metabolic co-morbidities and inflammatory markers was increased in six of the eight studies covered in the systematic review.

As described, these studies varied widely with respect to the subject selection, phenotyping methods used and outcomes used for validation (Pinto et al., 2015). Additionally, a number of these studies appeared to have low generalizability as a disproportionate number of patients were males, had severe disease, and were recruited from health care settings. Despite these differences in study designs and inherent biases in participants, two consistent COPD phenotypes were identified in most of the studies with consistently similar outcomes associated with each of them, as well. This consistency across both phenotype results and associated outcomes, in spite of the methodological differences in the studies, suggests potential reliability and predictive utility in COPD phenotyping, particularly if given a more structured and standardized methodology.

2.4.4 Limitations of Current COPD Phenotyping Studies

COPD phenotyping poses a number of challenges, a large proportion of which are methodological. First, nearly all of the existing studies sampled their patients from respiratory

clinics or other tertiary care settings. These are essentially convenience samples and limit the generalizability of the results. A broader sampling strategy could help overcome this limitation and better reflect the full population of COPD patients, as a whole (Bourbeau, Pinto, & Benedetti, 2014).

A number of the studies included only persons with a select and limited disease severity (usually severe), again limiting the generalizability of the results. These studies have derived COPD phenotypes based only on individuals with severe disease and, thus, do not factor in the characteristics of less severe disease that exist in the complete, diseased population. Sampling across all disease severities would provide for more robust computations of phenotypes that better reflect the overall COPD population and not just those with severe disease.

With the exception of two studies noted in the systematic review, the existing studies only included patients with complete data when performing their statistical derivation of phenotypes. This almost certainly leads to biased results both by avoiding selection of variables that could be significant to the phenotype (Bourbeau et al., 2014) and by potentially excluding patients who differ from included patients on important characteristics (Pinto et al., 2015). A preferred method for handling missing data, and employed in only one of the exceptions noted above, is to use multiple imputation to impute the missing fields for each subject rather than exclude all subjects without a full complement of data.

It has been noted in critiques of current methodologies that the derivation of phenotypes is usually performed on cross-sectional data. The limitation of this strategy is that the temporal stability of the derived phenotypes cannot be assessed at a single point in time. A longitudinal cohort, with data collected at multiple time points, would be required in order to assess the

variability of the phenotypes over time and their evolution (Bourbeau et al., 2014; Han et al., 2010).

In addition to the methodological limitations noted above, Barker and Brightling may have stumbled upon a potential philosophical limitation, "most approaches to phenotyping have considered disease at a single scale and have not integrated information from different scales to provide multi-dimensional phenotypes" (Barker & Brightling, 2013). In this comment, they were referring to multi-dimensionality in terms of the biological hierarchy within a person, i.e., an organ is part of a person; a tissue is part of an organ; a cell is part of a tissue; and a gene is part of a cell. In line with this thinking, many of the COPD phenotypes that have been derived to date are somewhat clinical in nature and have only been validated against clinical outcomes like morbidity, mortality, etc. But there's no mandate that they need to be, or even should be. The genetics and psychology disciplines have been discussing behavioral phenotypes for decades (Hand, 1981; Johnson, Ekman, Friesen, Nyhan, & Shear, 1976; Rovet & Ireland, 1994). In one 1976 article proposing the existence of a distinct behavioral phenotype in the de Lange syndrome, the authors speculated, "It seems clear that there is a distinct behavioral phenotype in the de Lange syndrome" and "The delineation of these phenotypes should provide important evidence on somatic determinants of behavior." (Johnson et al., 1976). Given the wealth of behavioral and psychosocial information available to some researchers, using only biological or clinical data when creating and/or validating phenotypes may be too narrow an approach, particularly when the outcome of interest may be behavioral in nature, e.g. medication adherence or smoking cessation.

CHAPTER III.

STUDY 1: PREVALENCE AND FACTORS ASSOCIATED WITH SELF-REPORTED CHRONIC OBSTRUCTIVE PULMONARY DISEASE

3.1 Introduction

Chronic obstructive pulmonary disease (COPD) is a lung disease characterized by persistent respiratory symptoms and chronic airflow obstruction that is usually progressive and often interferes with normal breathing. The airflow obstruction is caused by a mixture of inflammation in the large airways (chronic bronchitis) and alveolar destruction (emphysema) with the contributions and effect of each varying between persons (GOLD, 2018). While the more familiar terms "chronic bronchitis" and "emphysema" are no longer used in the formal definition of COPD, they are often used for the actual COPD diagnosis ("WHO | COPD," 2018). COPD develops gradually and is usually diagnosed in persons 40 years and older (GOLD, 2018). The most common symptoms of COPD include breathlessness (dyspnea), chronic cough, chronic sputum production and wheezing though the presence of the characteristic symptoms of COPD may predate the presence of airflow limitation by several years (GOLD, 2018). Conversely, significant airflow limitation can be present in the absence of overt symptoms. As a result, COPD is difficult to detect in its early stages and mild COPD often goes undiagnosed. COPD diagnosis is further complicated by the presence of other disorders, such as asthma and bronchiectasis, which may also present with similar symptoms and signs. Consequently, it is estimated that between 50 and 80% of all COPD cases go undiagnosed (Lamprecht et al., 2015; Lindberg, Bjerg-Bäcklund, Rönmark, Larsson, & Lundbäck, 2006a; Mannino et al., 2002).

The main objective of this study was to examine sociodemographic, general health, and COPD specific factors, including severity of lung obstruction, that are associated with healthcare provider-diagnosed COPD among U.S. adults.

3.2 Methods

Study and data

This study was based on analyses of the 2007–2012 National Health and Nutrition Examination Survey (NHANES) data, which were the years that spirometry testing was conducted on the survey. NHANES is a cross-sectional survey of the civilian, non-institutionalized U.S. population conducted by the National Center for Health Statistics ("NHANES - National Health and Nutrition Examination Survey Homepage," 2018). Data were collected via household interviews and standardized physical examinations in specially equipped mobile examination centers (MEC). The NHANES samples are selected through a complex, multistage, probability design. The 2007–2012 NHANES cycles oversampled major U.S. demographic subgroups including Hispanic and Black persons, low income white persons, and persons aged 80 years and older. Additionally, Asian persons were oversampled in 2011-2012. The procedures to select the sample and conduct the interview and examination have been specified elsewhere (Zipf, Chiappa, & Porter, 2013). Informed consent was obtained from all participants and the NCHS Research Ethics Review Board approved the protocol ("NHANES - NCHS Research Ethics Review Board Approval," 2017).

Analyses were conducted on data collected from adults aged 40 to 79 years old. The lower boundary of the analyzed age cohort (40 years) was based on guidelines for diagnosis which recommend considering spirometry in symptomatic patients over the age of 40 years (GOLD,

2018). The upper boundary of the analyzed age cohort (79 years) was selected as this was the maximum eligible age for the NHANES spirometry examination.

During 2007–2012, 14,825 persons aged 40–79 years were eligible for the survey, 10,531 (71%) were interviewed and 10,219 (69%) were examined in the NHANES mobile examination center (MEC).

Self-reported, healthcare-provider diagnosed COPD

During the in-home interview participants were asked detailed questions on specific medical conditions. A participant was classified as having healthcare provider-diagnosed COPD if they responded affirmatively to either of the following questions on emphysema or chronic bronchitis: "Has a doctor ever told you that you have emphysema?" or "Do you still have chronic bronchitis?" (Halldin, Doney, & Hnizdo, 2015; Mannino, Gan, Wurst, & Davis, 2017).

Demographic and other covariates

Demographic data were collected during the household interview. Age was categorized as 40–59 and 60–79 years. Self-reported race and Hispanic origin were categorized as non-Hispanic white, non-Hispanic black, and Hispanic. Due to small sample size and heterogeneity of group, participants who reported other race or Hispanic groups (including multiple races) were not reported separately but were included in total estimates. Self-reported education was categorized as less than 12th grade, high school graduate or equivalent, and some college to include anyone who had attended college for any length of time.

General health status was categorized as very good or excellent, good, and fair or poor. The number of times in the past year a person reported seeing a doctor or healthcare professional was categorized as collected: none, 1, 2 or 3, 4 to 9, and 10 or more with the exception of the two

highest categories (10-12 and 13 or more) which were combined for a more balanced data distribution. Presence of co-morbid conditions (coronary heart disease, stroke, cancer/malignancy, diabetes and hypertension) were identified based on affirmative responses to being told by a doctor or other health professional that he/she had the particular condition. Number of co-morbid medical conditions was derived based on the presence of these five conditions.

Cigarette smoking

A *current smoker* was a respondent who reported smoking at least 100 cigarettes in their lifetime and at the time of the interview reported smoking every day or some days OR by a serum cotinine value greater than 10ng/mL (Hukkanen, Jacob, & Benowitz, 2005). A *former smoker* was a respondent who reported smoking at least 100 cigarettes during their lifetime but currently did not smoke. A *never smoker* was a participant who reported smoking fewer than 100 cigarettes during their lifetime.

Respiratory symptoms and medication use

Presence of chronic cough was defined as affirmative responses to "Do you usually cough on most days for 3 consecutive months or more during the year?" and the follow-up question that the coughing had been for at least 2 years. Presence of chronic phlegm was defined as affirmative responses to "Do you bring up phlegm on most days for 3 consecutive months or more during the year?" and the follow-up question that bringing up of phlegm had been for at least 2 years. Presence of wheezing was defined by the question, "In the past 12 months, how many attacks of wheezing or whistling have you had?" Number of attacks was categorized as 0, 1-3, and 4 or more. Presence of dyspnea was based on an affirmative response to "Have you had

shortness of breath either when hurrying on the level or walking up a slight hill?" Number of respiratory symptoms was derived based on the presence of these four symptoms.

Participants were defined as using a respiratory medication if they reported past 30-day use of any of the following drug classes based on the Global Initiative for Chronic Obstructive Lung Disease (GOLD) treatment guidelines: long-acting bronchodilator, short-acting bronchodilator, inhaled corticosteroid, systemic corticosteroid, methylxanthine, and leukotriene modifier.

Spirometric assessment of COPD and lung obstruction severity

NHANES spirometry data were used to identify persons with lung obstruction and assess severity levels. Spirometry testing was performed on NHANES participants aged 6-79 years during 2007-2012. It was conducted in accordance with recommendations of the American Thoracic Society (Martin Raymond Miller et al., 2005) using Ohio 822/827 dry-rolling seal volume spirometers. The full exam protocol is detailed elsewhere ("NHANES - Respiratory Health—Bronchodilator Procedures Manual," 2011). Using pre-bronchodilator spirometry values, participants were defined as having evidence of COPD when the ratio of the forced expiratory volume in one second (FEV1) to the forced vital capacity (FVC) is less than the lower limit of normal (LLN) representing the lower 5_{th} percentile based on person's age, sex, height, and race/ethnicity [14]. LLN values were determined using normative reference equations developed from NHANES III data by Hankinson et al. [15].

Adults showing no spirometric evidence of COPD were categorized as having no measured lung obstruction. Those showing spirometric evidence of COPD were further categorized as having mild (FEV1>70% predicted) or moderate or worse (FEV1≤70% predicted) lung obstruction (Pellegrino et al., 2005). Percent of predicted FEV1 was defined as the observed

FEV1 value divided by the predicted FEV1 value estimated for a person of the same age, gender, race/ethnicity, and height using race-specific reference equations (Hankinson et al., 2010). Reference values are available for non-Hispanic white, non-Hispanic black, and Mexican Americans. For the Hispanic group, the reference values provided for Mexican Americans were used.

Prior to the spirometry examination, participants were asked if they currently had a breathing problem that required the use of supplemental oxygen during the daytime. Persons who reported a healthcare provider diagnosis of COPD and were excluded from the spirometry exam due to daily supplemental oxygen use, which is suggestive of advanced disease, were assigned a lung obstruction severity of moderate or worse (n=84).

Statistical methods

Statistical analyses were performed using STATA™ version 13.1 (StataCorp, College Station, TX). Examination sample weights were used to account for differential probabilities of selection, nonresponse, and non-coverage. Taylor series linearization was used to calculate standard errors to account for the complex sampling design. Differences in the prevalence of diagnosed COPD between subgroups was evaluated using T-tests for categorical variables and linear trend tests using linear regression and orthogonal contrasts for continuous variables. Chi-square tests were performed to examine the association between covariates and the presence of a COPD diagnosis. Multivariable logistic regression was used to assess predictors of diagnosis along with their associated odds ratios and statistical significance. All reported estimates have a relative standard error ≤30% ("NHANES - Analytic Guidelines, 1999-2010," 2013).

Missing data

The percentage of missing data at the individual variable level ranged from 0% (multiple variables) to approximately 20% for the spirometry measures with the overall fraction of missing information for the aggregated NHANES data being roughly 4%. To assess the impact of the missing data, sensitivity analyses were conducted on the proportions with diagnosed COPD and the characteristics of those with and without diagnosed COPD using 40 multiply imputed datasets. As estimates were similar between those based on the unmodified (complete case) data and those based on the multiply imputed data, all estimates presented here are based on the unmodified data for ease of reproducibility.

3.3 Results

Prevalence of diagnosed COPD among US adults 40-79 years

During 2007-2012, 5.2% of adults aged 40-79 years reported being diagnosed with COPD, which includes ever emphysema and current chronic bronchitis. Prevalence was similar by sex, higher among non-Hispanic white (5.9%) and non-Hispanic black (4.6%) adults than Hispanic (2.0%) adults, and higher among adults aged 60-79 (8.3%) than 40-59 (3.6%) years (Table 3.1). Prevalence of diagnosed COPD significantly decreased with increasing education going from 9.2% among adults with less than a high school education to 3.7% among those with any college education. Prevalence was 14.3% among adults who reported fair or poor health status compared to 5.0% who reported good health and 1.7% who reported very good/excellent health. Prevalence of diagnosed COPD increased as the number of health care visits in the past year increased going from 1.8% among those with 1 visit to 12.6% among those with 10 or more visits.

Prevalence of COPD diagnosis was greater among adults who were current smokers (10.2%) than past (6.6%) or never (2.1%) smokers. A third (33.7%) of adults with moderate or worse lung obstruction had been diagnosed with COPD compared to 4.1% of those with mild lung obstruction and 3.1% with no lung obstruction. As the number of respiratory symptoms increased the prevalence of diagnosed COPD also increased. Specifically, prevalence was less than 1% (0.9%) when no respiratory symptoms were reported, 5.8% with 1 symptom, 14.9% with 2, and 36.7% with 3 or more symptoms.

Characteristics of US adults aged 40-79 with and without diagnosed COPD

Table 3.2 provides the characteristics of those with diagnosed COPD and those without diagnosed COPD, which includes those with undiagnosed COPD as well as those without the condition. Adults aged 40-79 years with diagnosed COPD were significantly different from those without diagnosed COPD in all sociodemographic and health characteristics examined, except gender (Table 3.2). Specifically, among those with diagnosed COPD, a higher proportion were non-Hispanic white (81.2% vs 71.4%), aged 60-79 (54.5% vs. 33.4%), and less reported higher education (40.9% vs 58.4%).

Over one half (50.8%) of adults with diagnosed COPD reported fair or poor health while only 13.8% reported very good or excellent health compared to 17.3% reporting fair/poor health and 44.2% reporting very good or excellent health among those without diagnosed COPD. Slightly more than three-quarters of adults (75.4%) with diagnosed COPD reported one or more of the following co-morbid conditions: hypertension, diabetes, cancer, heart disease, or stroke compared to 50.3% reporting one or more of these conditions for those without diagnosed COPD. 72.6% of adults with COPD reported having 4 or more healthcare visits in the prior year compared to only 40.7% of adults without COPD.

Nearly half (47.1%) of adults with diagnosed COPD currently smoked cigarettes and 33.4% were former smokers. These proportions were significantly different from adults without diagnosed COPD where more than one-half (51.3%) were never cigarette smokers and 22.7% currently smoked.

Among adults with diagnosed COPD, 39.5% reported having been diagnosed with emphysema alone, 45.3% reported current chronic bronchitis alone and 15.2% reported both diseases. Nearly half (49.1%) reported using a prescribed respiratory drug. Of those with diagnosed COPD, 55.7% showed no spirometric evidence of lung obstruction while 7.9% presented with mild lung obstruction and 36.4% had moderate or worse lung obstruction. Among adults without diagnosed COPD, 87% had no measured lung obstruction. However, 9.3% had mild lung obstruction and 3.6% had moderate or worse lung obstruction.

Respiratory symptoms were prevalent among adults with diagnosed COPD with 76.5% reporting dyspnea or shortness of breath, 41.4% having chronic cough, 30.9% chronic phlegm, and 59.3% at least one wheezing episode in the past year. Nearly 90% (89.4%) reported having at least one of these four respiratory symptoms compared to 35.9% among those without diagnosed COPD.

Predictors of being diagnosed with COPD by a healthcare professional

Table 3.3 details the adjusted odds ratios, confidence intervals, and associated *p*-values from multivariable logistic regression modeling used to examine the association of various covariates with a COPD diagnosis outcome. After controlling for all the other variables in the model, non-Hispanic blacks were 49% less likely (AOR=0.51, 95% CI = 0.35 to 0.74) and Hispanics were 68% less likely (AOR=0.32, 95% CI = 0.19 to 0.55) than non-Hispanic whites to be diagnosed with COPD. Persons aged 60-79 years were 65% more likely (AOR=1.65, 95% CI

= 1.07 to 2.53) than persons aged 40-59 years to be diagnosed with COPD. Those reporting fair or poor health were nearly three times as likely to be diagnosed with COPD (AOR=2.91, 95% CI = 1.55 to 5.46) than those reporting being in very good or excellent health. Persons reporting 4 to 9 healthcare visits in the past year were 3.49 times as likely (AOR=3.49, 95% CI = 1.30 to 9.39) to be diagnosed with COPD than those with no healthcare visits in the prior year and those with 10 or more healthcare visits in the past year were over 5 times as likely (AOR=5.06, 95% CI = 1.62 to 15.77) to be diagnosed than those with no healthcare visits in the previous year. Former smokers and current smokers were 75% and 70% more likely (AOR=1.75, 95% CI = 1.23 to 2.48 and AOR=1.70, 95% CI = 1.17 to 2.48, respectively) than never smokers to be diagnosed with COPD.

In terms of lung obstruction severity, adults with moderate or worse lung obstruction were nearly five times as likely (AOR=4.90, 95% CI = 3.28 to 7.32) than those with no measured lung obstruction to be diagnosed with COPD. Having mild disease was not significantly associated with being diagnosed with COPD.

Finally, the likelihood of being diagnosed with COPD significantly increased as the number of reported respiratory symptoms increased. Specifically, adults reporting 1 respiratory symptom were nearly 4 times as likely (AOR=3.91, 95% CI = 2.11 to 7.24), those reporting 2 respiratory symptoms were nearly 10 times as likely (AOR=9.57, 95% CI = 5.43 to 16.88) and those reporting 3 or more respiratory symptoms were about 22 times as likely (AOR=22.07, 95% CI = 12.03 to 40.49) than those with no respiratory symptoms to be diagnosed with COPD.

3.4 Discussion

Approximately 5.2% of US adults aged 40-79 years were estimated to have been diagnosed with COPD based on NHANES 2007-2012. This estimate is based on persons

reporting ever having emphysema or currently having chronic bronchitis, which is similar to previous studies using NHANES [9,10]. Case-definitions using other surveys (e.g., BRFSS, BOLD, NHIS), however, have included different combinations of the variables on ever emphysema, ever chronic bronchitis, current chronic bronchitis, and COPD. Based on NHIS 2012-2014, Ward et al. found that the prevalence of self-reported diagnosed COPD increased when an explicit question on COPD was included with the ever emphysema and ever chronic bronchitis questions [17]. Such a question, however, was unavailable on NHANES during 2007-2012 when the lung function testing was conducted on the survey. In terms of bronchitis, our prevalence estimate would have been higher at 8.6% (SE 0.6) if we had used ever chronic bronchitis rather than current chronic bronchitis. However, the confirmation of current bronchitis after affirmation of ever bronchitis allows the exclusion of respondents who may have once had acute bronchitis but currently do not.

Univariate analyses showed significant differences between those with and without diagnosed COPD across a number of sociodemographic factors with disproportionate percentages of those with a COPD diagnosis being older, less educated, with more co-morbid chronic conditions and presenting more respiratory symptoms. While the recent BRFSS analyses by Kosacz et al. showed a higher prevalence of self-reported COPD in men than women (Kosacz, Punturieri, & Croxton, 2012), our results showed no appreciable difference in COPD diagnosis status by gender. A possible narrowing of the gender gap is alluded to in the most recent GOLD report which noted that many earlier studies showed COPD prevalence to be much greater in men than in women but recent data, particularly from developed countries, show the prevalence of the disease to now be nearly equal in men and women (GOLD, 2018). It was further suggested in the GOLD report that this narrowing of the COPD prevalence gap between

genders is likely due to changing patterns of tobacco smoking. Similar to our univariate results, the previously mentioned BRFSS analyses reported higher proportions of self-reported COPD among older persons and those who did not graduate from high-school (Kosacz et al., 2012). After controlling for other covariates, however, we found education level to no longer be associated with a COPD diagnosis.

Cigarette smoking is widely recognized as the most significant environmental risk factor for COPD. Accordingly, we found over 80% of adults with diagnosed COPD were ever-smokers versus less than 49% who ever smoked among those without diagnosed COPD. A number of studies have suggested that the receipt of a COPD diagnosis could contribute to reduced levels of smoking among those with disease (Danielsen, Løchen, Medbø, Vold, & Melbye, 2016; Stratelis et al., 2006). Yet, in spite of their COPD diagnosis, we found nearly 50% (47.1%) of those diagnosed with COPD were still currently smoking. This figure is higher than the 36.7% prevalence of current smoking among those reporting a COPD diagnosis in the 2011 BRFSS data (Cunningham, Ford, Rolle, Wheaton, & Croft, 2015) but virtually mirrors the 47.3% of adults with self-reported COPD who were current smokers in the optional smoking cessation module of the 2011 BRFSS (Schauer et al., 2014).

Although COPD is defined based on airflow limitation, the impact of symptoms usually determines the decision to seek medical help (GOLD, 2018). Consistent with this, we found nearly 90% of adults with diagnosed COPD reporting one or more respiratory symptoms. Further, adults with 3 or more respiratory symptoms were 22 times as likely to be diagnosed with COPD than those with no respiratory symptoms. Chronic dyspnea is the most characteristic symptom of COPD (GOLD, 2018) and our results showed that three-fourths of adults with a COPD diagnosis reported having dyspnea. The relationship between COPD and respiratory

symptoms is not always predictable, however, given that symptoms may precede the development of airflow limitations and, conversely, not all persons with lung obstruction exhibit symptoms (GOLD, 2018). We found that over half (55.7%) of adults with diagnosed COPD had no measured lung obstruction though almost all presented with one or more respiratory symptoms (results not shown). Lamprecht et al. found that only 36.4% of their aged 40+ study participants who reported having COPD had airway obstruction, the remaining 63.6% had a COPD diagnosis with no airway obstruction (Lamprecht et al., 2015). These cases where there is a reported COPD diagnosis but no measured lung obstruction may be due to a number of factors including, but not limited to, improved lung function resulting from treatment, diurnal variation in lung function measures, participants' erroneous reporting of a diagnosis or physician misdiagnosis. We also found that nearly 13% of adults 40-79 years who did not report a COPD diagnosis had spirometric evidence of lung obstruction. However, nearly three-quarters of them (9.3%) had mild lung obstruction. Based on the adjusted regression analysis, having mild lung obstruction was not significantly associated with diagnosis while adults with moderate or worse lung obstruction were four times as likely to be diagnosed than those with no lung obstruction. These findings relating symptoms, airflow limitation and diagnosis add support to the GOLD committee premise that while COPD is defined based on airflow limitation, diagnosis in the clinical setting may be driven by the presentation of respiratory symptoms.

Limitations to this study include those normally associated with cross-sectional studies, most notably it is impossible to determine cause-effect or temporal relationships. A number of items used in these analyses were also collected using questionnaires and are subject to recall bias and other limitations of self-reported data. Additionally, as NHANES does not include persons from institutionalized settings (e.g., nursing facilities and assisted-care facilities) and

because COPD is associated with older age, our results may reflect an underestimation of the true prevalence of diagnosed COPD. Finally, in order to include spirometry data, these analyses were limited to using data last collected in 2012. This is a limitation as there are newer data available for the estimation of self-reported COPD prevalence. The greatest strength of this study is its use of NHANES nationally representative data which included both self-reported diagnosis information as well as objective lung function measures to afford comparisons between self-reported COPD and measured lung obstruction.

3.5 Conclusions

Approximately 5.2% of US adults aged 40-79 years were estimated to have self-reported COPD during 2007-2012. Among those with diagnosed COPD, over one third had moderate or worse lung function, just under half were still currently smoking, and nearly 90% had at least one respiratory symptom. Multiple factors were found to be associated with the receipt of a COPD diagnosis with number of reported respiratory symptoms having the strongest association. After controlling for other factors, having mild lung obstruction was not associated with a COPD diagnosis.

Table 3.1. Prevalence of self-reported diagnosed COPD among US adults aged 40-79 years: NHANES, 2007-2012

Characteristics (n=10,219 persons)	n[a]	Weighted percent with diagnosed COPD[b] (SE)	*p*-value[c]
Total	10,219	5.2% (0.4)	-
Gender			
Male	5,016	4.8% (0.5)	Ref
Female	5,203	5.6% (0.5)	0.216
Race and ethnic origin			

Non-Hispanic white	4,435	5.9% (0.5)	Ref
Non-Hispanic black	2,353	4.6% (0.6)	0.074
Hispanic	2,632	2.0% (0.3)	< 0.001
Age in years			
40-59	5,593	3.6% (0.4)	Ref
60-79	4,626	8.3% (0.7)	< 0.001
Education			
Less than 12th grade	3,104	9.2% (1.1)	
HS Grad/GED/Equivalent	2,329	5.6% (0.7)	
Any College	4,774	3.7% (0.3)	
General health condition			< 0.001[d]
Very good or excellent	3,027	1.7% (0.3)	Ref
Good	3,641	5.0% (0.7)	< 0.001
Fair or poor	2,598	14.3% (1.2)	< 0.001
Other medical conditions[e]			
Coronary heart disease	507	15.6% (2.4)	-
Stroke	493	16.2% (2.1)	-
Cancer	1,171	9.2% (1.4)	-
Diabetes	2,024	10.0% (1.1)	-
Hypertension	4,710	7.3% (0.6)	-
Number of other medical conditions			< 0.001[d]
0	4,403	2.7% (0.4)	
1	3,420	6.0% (0.5)	
2	1,810	8.4% (1.1)	
3 or more	586	17.3% (2.4)	
Number of healthcare visits in past year			< 0.001[d]
0	1,367	1.9% (0.5)	
1	1,570	1.8% (0.4)	
2 or 3	2,774	3.1% (0.4)	
4 to 9	2,910	7.1% (0.8)	
10 or more	1,593	12.6% (1.5)	
Smoking status			
Never smoker	4,990	2.1% (0.2)	Ref
Former smoker	2,712	6.6% (0.6)	< 0.001
Current smoker	2,513	10.2% (0.9)	< 0.001
Measured lung obstruction			
No measured lung obstruction	7,247	3.1% (0.4)	Ref
Mild	722	4.1% (0.7)	0.147
Moderate or worse	466	33.7% (2.5)	< 0.001
Respiratory symptoms			
Chronic cough	893	23.5% (1.9)	-
Chronic phlegm	755	22.5% (2.1)	-
Wheezing episodes past year			< 0.001[d]

0	8,696	2.4% (0.2)	
1-3	798	17.2% (1.8)	
4 or more	662	28.5% (2.6)	
Dyspnea	3,318	13.2% (1.0)	-
Number of respiratory symptoms			< 0.001[d]
0	6,110	0.9% (0.2)	
1	2,575	5.8% (0.7)	
2	973	14.9% (1.5)	
3 or more	561	36.7% (2.9)	

[a] Categories may not sum to the total (10,219) due to missing data.

[b] Self-reported, health professional-diagnosed COPD was assessed by an affirmative response to either ever having emphysema (MCQ160G) or still having chronic bronchitis (MCQ170K).

[c] Values are based on two-sided ttests on the difference in prevalence between the designated subgroup and the reference subgroup.

[d] Values are based on linear trend test using linear regression and orthogonal contrasts.

Table 3.2. Characteristics of US adults aged 40-79 years with and without self-reported diagnosed COPD by socio-demographic factors, health status and chronic conditions: NHANES, 2007-2012

Characteristics (n=10,219 persons)	With diagnosed COPD[a] (n=566)	Without diagnosed COPD[a] (n=9,653)	*p*-value[b]
Gender			0.219
Male	44.0% (3.3)	48.2% (0.6)	
Female	56.0% (3.3)	51.8% (0.6)	
Race and ethnic origin			< 0.001
Non-Hispanic white	81.2% (2.7)	71.4% (2.0)	
Non-Hispanic black	9.6% (1.6)	11.0% (1.1)	
Hispanic	4.2% (0.8)	11.1% (1.3)	
Age in years			< 0.001
40-59	45.5% (2.6)	66.6% (0.6)	
60-79	54.5% (2.6)	33.4% (0.6)	
Education			< 0.001
Less than 12th grade	33.8% (2.6)	18.3% (0.9)	
HS Grad/GED/Equivalent	25.3% (2.5)	23.3% (0.9)	
Any College	40.9% (3.2)	58.4% (1.4)	
General health condition			< 0.001
Very good or excellent	13.8% (2.2)	44.2% (1.0)	
Good	35.4% (3.2)	38.4% (0.9)	
Fair or poor	50.8% (3.3)	17.3% (0.7)	
Other medical conditions			
Coronary heart disease	12.7% (1.6)	3.8% (0.3)	< 0.001
Stroke	10.9% (1.4)	3.1% (0.2)	< 0.001
Cancer	21.9% (2.2)	11.9% (0.5)	< 0.001
Diabetes	28.5% (2.7)	14.2% (0.5)	< 0.001
Hypertension	57.2% (2.8)	40.0% (0.7)	< 0.001
Number of other medical conditions			< 0.001
0	24.6% (2.5)	49.7% (0.7)	
1	37.6% (2.2)	32.5% (0.6)	
2	23.2% (2.5)	14.0% (0.5)	
3 or more	14.6% (1.9)	3.8% (0.2)	
Number of healthcare visits in past year			< 0.001
0	4.6% (1.2)	13.1% (0.5)	
1	5.7% (1.3)	16.7% (0.4)	

2 or 3	17.1% (2.3)	29.6% (0.8)	
4 to 9	38.9% (2.8)	27.8% (0.5)	
10 or more	33.7% (2.9)	12.9% (0.6)	
Smoking status			< 0.001
Current smoker	47.1% (3.1)	22.7% (0.8)	
Former smoker	33.4% (2.8)	26.0% (0.8)	
Never smoker	19.6% (1.6)	51.3% (0.8)	
Disease reported			
None	-	100.0% (0.0)	-
Emphysema	39.5% (2.7)	N/A	
Chronic bronchitis	45.3% (2.6)	N/A	
Both	15.2% (1.9)	N/A	
Taking prescription respiratory medication	49.1% (3.0)	5.6% (0.4)	< 0.001
Measured lung obstruction severity			< 0.001
No measured lung obstruction	55.7% (3.0)	87.0% (0.6)	
Mild	7.9% (1.1)	9.3% (0.5)	
Moderate or worse	36.4% (2.9)	3.6% (0.3)	
Respiratory symptoms			
Chronic cough	41.4% (2.5)	7.3% (0.5)	< 0.001
Chronic phlegm	30.9% (2.4)	5.7% (0.3)	< 0.001
Wheezing episodes past year			< 0.001
0	40.7% (3.2)	88.8% (0.4)	
1-3	25.4% (2.5)	6.6% (0.4)	
4 or more	33.9% (3.0)	4.6% (0.3)	
Dyspnea	76.5% (2.7)	27.7% (0.8)	< 0.001
Number of respiratory symptoms			< 0.001
0	10.7% (1.9)	64.1% (0.9)	
1	27.2% (2.7)	24.3% (0.6)	
2	26.2% (2.4)	8.2% (0.4)	
3 or more	36.0% (2.2)	3.4% (0.3)	

[a] Self-reported, health professional-diagnosed COPD was assessed by an affirmative response to either ever having emphysema (MCQ160G) or still having chronic bronchitis (MCQ170K).

[b] Chi-square values are based on the difference in the percentage of those self-reporting/not self-reporting a COPD diagnosis between categories. For the condition and symptom variables, it is the difference in the percentage of those self-reporting/not self-reporting a COPD diagnosis between those reporting/not reporting the condition or symptom.

Table 3.3. Adjusted logistic regression analysis[a] of factors associated with self-reported diagnosed COPD[b] among US adults aged 40-79 years: NHANES, 2007-2012

Characteristics (n=7,839 persons in model)	Adjusted Odds Ratio (95% CI)	*p*-value
Gender		
Male	Ref	-
Female	0.76 (0.52, 1.12)	0.166
Race and ethnic origin		
Non-Hispanic white	Ref	-
Non-Hispanic black	0.51 (0.35, 0.74)	0.001
Hispanic	0.32 (0.19, 0.55)	< 0.001
Age in years		
40-59	Ref	-
60-79	1.65 (1.07, 2.53)	0.023
Education		
Less than 12th grade	Ref	-
HS Grad/GED/Equivalent	0.68 (0.39, 1.17)	0.159
Any College	0.71 (0.44, 1.16)	0.170
General health condition		
Very good or excellent	Ref	-
Good	1.54 (0.94, 2.53)	0.085
Fair or poor	2.91 (1.55, 5.46)	0.001
Number of other medical conditions [c]		
0	Ref	-
1	1.10 (0.68, 1.78)	0.695
2	1.12 (0.61, 2.05)	0.712
3 or more	1.32 (0.68, 2.54)	0.402
Number of healthcare visits in past year		
0	Ref	-
1	2.05 (0.62, 6.77)	0.233
2 or 3	1.88 (0.70, 5.08)	0.207
4 to 9	3.49 (1.30, 9.39)	0.014
10 or more	5.06 (1.62, 15.77)	0.006
Smoking status		-
Never smoker	Ref	
Former smoker	1.75 (1.23, 2.48)	0.002

Current smoker	1.70 (1.17, 2.48)	0.007
Measured lung obstruction severity		
No measured lung obstruction	Ref	-
Mild	0.81 (0.55, 1.20)	0.293
Moderate or worse	4.90 (3.28, 7.32)	< 0.001
Number of respiratory symptoms [d]		
0	Ref	-
1	3.91 (2.11, 7.24)	< 0.001
2	9.57 (5.43, 16.88)	< 0.001
3 or more	22.07 (12.03, 40.49)	< 0.001

[a] Multivariable logistic regression model included all variables shown in table above.
[b] Self-reported, health professional-diagnosed COPD was assessed by an affirmative response to either ever having emphysema (MCQ160G) or still having chronic bronchitis (MCQ170K).
[c] Medical conditions include self-reported healthcare provider diagnosed coronary heart disease, stroke, cancer, diabetes, or hypertension.
[d] Respiratory symptoms include self-reported chronic cough (3 consecutive months or more during year AND for at least 2 years), chronic phlegm (3 consecutive months or more during year AND for at least 2 years), wheezing episode (past 12 months), or dyspnea.

CHAPTER IV.

STUDY 2: FACTORS ASSOCIATED WITH CONTINUED SMOKING AMONG US ADULTS WITH CHRONIC OBSTRUCTIVE PULMONARY DISEASE: THE CRITICAL ROLE OF OTHER HOUSEHOLD SMOKERS

4.1 Introduction

Chronic Obstructive Pulmonary Disease (COPD) is estimated to be the third leading cause of death both in the U.S. (Murphy et al., 2013) and around the globe (WHO, 2016) and is a growing cause of suffering and morbidity worldwide. In the U.S., nearly 15 million adults are estimated to be suffering from COPD (NHLBI, 2012). The economic impact of COPD on the health system in the U.S. is enormous with an estimated $32 billion in direct costs in 2010 and an additional $20.4 billion in indirect costs (GOLD, 2018).

While there is no cure for COPD, smoking cessation has been shown to improve respiratory symptoms, reduce the rate of lung function decline, reduce exacerbation risk, and improve the health-related quality of life for those with the disease (Anthonisen et al., 2002; Au et al., 2009; Bolliger et al., 2002; Willemse et al., 2004). Accordingly, smoking cessation is widely regarded as the most effective therapeutic intervention in patients at all stages of COPD (GOLD, 2018; Hersh et al., 2004; Stratelis et al., 2006; Willemse et al., 2004).

In spite of the numerous benefits of quitting, a recent study showed the smoking rate among the adult COPD population is still more than double that of adults with no chronic diseases, i.e., 47% vs. 20%, respectively (Schauer et al., 2014). Data from the National Health and Nutrition Examination Survey (NHANES) support this prevalence of smoking among the COPD population with 46.2% of adults aged 40-79 years with COPD being current smokers (Paulose-Ram et al., 2015). The NHANES data further reveal differences in smoking rates by

lung obstruction severity with a significantly higher proportion of persons with moderate or more severe lung obstruction currently smoking (56%) than persons with mild lung obstruction (41%).

In the general population. predictors of smoking cessation success include being older, being male, having higher income, smoking fewer cigarettes daily, having a history of quit attempts, and having a strong desire to stop smoking (Hymowitz et al., 1997; Monsó, Campbell, Tønnesen, Gustavsson, & Morera, 2001). Whether these predictors of successful cessation of smoking for the general population are also predictive of smoking status for persons with diagnosed COPD is unknown. Additionally, it is not known whether the differences in smoking status by disease severity reported by Paulose-Ram et al. persist among adults who have been diagnosed with COPD and are aware of their disease status.

The objectives of this research were to: 1) describe the characteristics of ever-smokers (current and former) with self-reported, healthcare provider-diagnosed COPD; 2) describe differences in smoking status among those with and without diagnosed COPD, overall and by disease severity; and 3) identify factors associated with continued smoking among those with diagnosed COPD.

4.2 Methods

Study and data

This study was based on analyses of the 2007–2012 National Health and Nutrition Examination Survey (NHANES) data, which were the years spirometry testing was conducted and included in the survey. NHANES is a cross-sectional survey of the civilian, non-institutionalized U.S. population conducted by the National Center for Health Statistics ("NHANES - National Health and Nutrition Examination Survey Homepage," 2018). Data were collected via household interviews and standardized physical examinations in specially equipped

mobile examination centers (MEC). The NHANES samples are selected through a complex, multistage, probability design. The 2007–2012 NHANES cycles oversampled major U.S. demographic subgroups including Hispanic and black persons, low income white persons, and persons aged 80 years and older. Additionally, Asian persons were oversampled in 2011-2012. The procedures to select the sample and conduct the interview and examination have been specified elsewhere (Zipf et al., 2013). Informed consent was obtained from all participants and the NCHS Research Ethics Review Board approved the protocol ("NHANES - NCHS Research Ethics Review Board Approval," 2017).

Analyses were conducted on data collected from adults aged 40 to 79 years old. The lower boundary of the analyzed age cohort (40 years) was based on guidelines for diagnosis, which recommend considering spirometry in symptomatic patients over the age of 40 years (GOLD, 2018). The upper boundary of the analyzed age cohort (79 years) was selected as this was the maximum eligible age for the NHANES spirometry examination.

For the period of 2007–2012, 14,825 persons aged 40–79 years were eligible for the survey, 10,531 (71%) were interviewed, 10,219 (69%) were examined in the NHANES mobile examination center (MEC) and 566 (5.5%) reported having healthcare-provider diagnosed COPD.

Self-reported, healthcare-provider diagnosed COPD

During the in-home interview, participants were asked detailed questions on specific medical conditions. A participant was classified as having healthcare provider-diagnosed COPD if they responded affirmatively to either of the following questions on emphysema or chronic

bronchitis: "Has a doctor ever told you that you have emphysema?" or "Do you still have chronic bronchitis?" (Halldin et al., 2015; Mannino et al., 2017).

Demographic and other covariates

Demographic data were collected during the household interview. Age was categorized as 40–59 and 60–79 years. Self-reported race and Hispanic origin were categorized as non-Hispanic white, non-Hispanic black, and Hispanic. Due to small sample sizes, participants who reported other race or Hispanic groups (including multiple races) were not reported separately, but were included in total estimates. Self-reported education was categorized as less than 12th grade, high school graduate or equivalent, and some college to include anyone who had attended college for any length of time. Annual family income was categorized as making less than $20,000 per year (low), making between $20,000 and $99,999 per year (middle), and making $100,000 or more per year (high).

General health status was categorized as very good or excellent, good, and fair or poor. The number of times in the past year a person reported seeing a doctor or healthcare professional was categorized as collected: none, 1, 2 or 3, 4 to 9, and 10 or more with the exception of the two highest categories (10-12 and 13 or more) which were combined for a more balanced data distribution. Presence of co-morbid conditions (coronary heart disease, stroke, cancer/malignancy, diabetes and hypertension) were identified based on affirmative responses to being told by a doctor or other health professional that he/she had the particular condition. Number of co-morbid medical conditions was derived based on the presence of these five conditions.

As odds ratios in the predictive model were similar at all non-zero counts of other smokers in the household, other smokers in the household was dichotomized to increase group

sizes and minimize standard errors of estimates. If the participant reported no other smokers in the household, the recoded variable was set to 'no' (zero) whereas if the participant reported having 1, 2, or 3+ smokers in the household, the recoded variable was set to 'yes' (one). Healthcare visits in the past year was similarly separated into two categories to increase group sizes with zero to two visits in the past year comprising the first category and three or more visits comprising the second category.

Cigarette smoking

A *current smoker* was a respondent who reported smoking at least 100 cigarettes in their lifetime and at the time of the interview reported smoking every day or some days OR by a serum cotinine value greater than 10ng/mL (Hukkanen et al., 2005). A *former smoker* was a respondent who reported smoking at least 100 cigarettes during their lifetime, but currently did not smoke. A *never smoker* was a participant who reported smoking fewer than 100 cigarettes during their lifetime.

Respiratory symptoms and medication use

Presence of chronic cough was defined as affirmative responses to "Do you usually cough on most days for 3 consecutive months or more during the year?" and the follow-up question of coughing having lasted for at least 2 years. Presence of chronic phlegm was defined as affirmative responses to "Do you bring up phlegm on most days for 3 consecutive months or more during the year?" and the follow-up question regarding bringing up of phlegm had been for at least 2 years. Presence of wheezing was defined by the question, "In the past 12 months, how many attacks of wheezing or whistling have you had?" Number of attacks was categorized as 0, 1-3, and 4 or more. Presence of dyspnea was based on an affirmative response to "Have you had

shortness of breath either when hurrying on the level or walking up a slight hill?" Number of respiratory symptoms was derived based on the presence of these four symptoms.

Participants were defined as using a respiratory medication if they reported past 30-day use of any of the following drug classes based on the Global Initiative for Chronic Obstructive Lung Disease (GOLD) treatment guidelines: long-acting bronchodilator, short-acting bronchodilator, inhaled corticosteroid, systemic corticosteroid, methylxanthine, and leukotriene modifier.

Spirometric assessment of COPD and lung obstruction severity

NHANES spirometry data were used to identify persons with lung obstruction and assess severity levels. Spirometry testing was performed on NHANES participants aged 6-79 years during 2007-2012. It was conducted in accordance with recommendations of the American Thoracic Society (Martin Raymond Miller et al., 2005) using Ohio 822/827 dry-rolling seal volume spirometers. The full exam protocol is detailed elsewhere ("NHANES - Respiratory Health—Bronchodilator Procedures Manual," 2011). Using pre-bronchodilator spirometry values, participants were defined as having evidence of COPD when the ratio of the forced expiratory volume in one second (FEV1) to the forced vital capacity (FVC) is less than the lower limit of normal (LLN) representing the lower 5_{th} percentile based on person's age, sex, height, and race/ethnicity (Pellegrino et al., 2005). LLN values were determined using normative reference equations developed from NHANES III data by Hankinson et al. (Hankinson et al., 2010).

Adults showing no spirometric evidence of COPD were categorized as having no measured lung obstruction. Those showing spirometric evidence of COPD were further categorized as having mild (FEV1>70% predicted) or moderate or worse (FEV1≤70% predicted)

lung obstruction (Pellegrino et al., 2005). Percent of predicted FEV1 was defined as the observed FEV1 value divided by the predicted FEV1 value estimated for a person of the same age, gender, race/ethnicity, and height using race-specific reference equations (Hankinson et al., 2010). Reference values were? available for non-Hispanic white, non-Hispanic black, and Mexican Americans. For the Hispanic group, the reference values provided for Mexican Americans were used.

Prior to the spirometry examination, participants were asked if they currently had a breathing problem that required the use of supplemental oxygen during the daytime. Persons who reported a healthcare provider diagnosis of COPD and were excluded from the spirometry exam due to daily supplemental oxygen use, which is suggestive of advanced disease, were assigned a lung obstruction severity of moderate or worse (n=84).

Statistical methods

Statistical analyses were performed using STATA™ version 13.1 (StataCorp, College Station, TX). Examination sample weights were used to account for differential probabilities of selection, nonresponse, and non-coverage. Taylor series linearization was used to calculate standard errors to account for the complex sampling design. Chi-square tests were performed to examine the association between each of the covariates and smoking status. A preliminary multivariable logistic regression was fitted to the data using only those variables, which were deemed to have significant associations with smoking status from the chi-square tests. The final model considered only those variables which were associated with smoking status as determined by adjusted Wald tests from the preliminary model and all reported adjusted odds ratios were based on this final model.

4.3 Results

Characteristics of current and former smokers, aged 40-79 years and diagnosed with COPD

During the period 2007-2012, among ever-smokers aged 40-79 years reporting a diagnosis of COPD, gender and race proportions were similar for current and former smokers (Table 4.1). Compared to former smokers, a higher percentage of current smokers: were younger (62.7% vs 23.9%), were less educated (42.4% with less than 12th grade education vs 35.0% and 29.0% with some college vs. 44.8%) reported having chronic cough (55.0% vs. 32.1%), reported having chronic phlegm (43.6% to 27.1%), reported having 3 or more respiratory symptoms (50.3% vs. 30.0%), reported having mild or worse depression (55.5% to 33.1%), reported having fair or poor health (65.0% to 45.5%), reported having other smokers in the household (67.5% to 9.4%) and reported two or fewer healthcare visits in the past year (32.2% vs. 19.7%).

Conversely, when compared to current smokers, a higher percentage of former smokers: were older (76.1% to 37.3%), reported having cancer, diabetes and hypertension (29.8% vs. 17.5%, 37.0% vs. 24.8%, 69.0% vs. 46.7%, respectively), had one or more chronic medical health conditions (84.5% vs. 70.4%), reported minimal depression (66.9% vs. 44.5%), reported good or better general health (54.5% vs. 35.0%), reported having no other smokers in the household (90.6% vs. 32.5%) and reported having 3 or more healthcare visits in the past year (80.3% vs. 67.8%).

Smoking status proportions among US adults aged 40-79 by diagnosed status and disease severity

Table 4.2 provides the reported smoking status overall and by COPD diagnosis status. The table further reports smoking status by measured lung obstruction severity for those with

diagnosed COPD. There was a significant difference in smoking status ($p<0.001$) between those diagnosed with COPD and those not diagnosed with COPD. Among those reporting a COPD diagnosis, 47.1% were current smokers, 33.4% were former smokers, and 19.6% were never smokers while, among those not reporting a COPD diagnosis, 22.7% were current smokers, 26.0% were former smokers, and 51.3% were never smokers. There was no difference in smoking status proportions ($p=0.735$) between those with mild lung obstruction (50.4% current, smoker 37.1% former smoker, 12.4% never smoker) and those with moderate or worse lung obstruction (55.5% current smoker, 37.4% former smoker, 7.2% never smoker). When including the group with no measured lung obstruction, differences in smoking status were significant ($p=0.006$) with a lower proportion of this group being current smokers and a higher proportion being never smokers (37.3% current smoker, 32.1% former smoker, 30.6% never smoker).

Predictors of current smoking status among ever-smokers, aged 40-79 years and diagnosed with COPD

Table 4.3 details the adjusted odds ratios, 95% confidence intervals, and associated *p*-values from multivariable logistic regression modeling used to examine the association of various covariates with smoking status. After controlling for all the other variables in the model, persons aged 60-79 years were 78% less likely (AOR=0.22, 95% CI = 0.12 to 0.41) than persons aged 40-59 years to be current smokers. Persons reporting having hypertension were 72% less likely (AOR=0.28, 95% CI = 0.13 to 0.60) than persons not reporting hypertension to be current smokers. Persons reporting one or more comorbid chronic conditions were 2.2 times as likely (AOR=2.20, 95% CI = 1.01 to 4.75) as those reporting no comorbid conditions to be current smokers. Those reporting fair or poor health were nearly two times as likely to be current smokers (AOR=1.99, 95% CI = 1.07 to 3.70) than those reporting good or better health. Finally,

persons with other smokers in the household were 19.51 times as likely (AOR=19.51, 95% CI = 10.16 to 37.47) to be current smokers than those with no other smokers in the household.

4.4 Discussion

A growing body of scientific literature suggests that patients with COPD benefit greatly from smoking cessation including decreased respiratory symptoms, reduced lung function decline, fewer exacerbations, and improved health-related quality of life (Anthonisen et al., 2002; Au et al., 2009; Bolliger et al., 2002; Willemse et al., 2004). In spite of the numerous benefits of smoking cessation, the prevalence of smoking among individuals with COPD remains more than double that of the population as a whole (Schauer et al., 2014; Tilert et al., 2018; Vozoris & Stanbrook, 2011).

In the current study, univariate analyses showed former smokers with COPD to be older, have one or more chronic medical health conditions, have minimal depression, report good or better general health, report more healthcare visits in the past year, and have no other smokers in the household. A number of these findings were mirrored in a study by Adams et al. who found former smokers with COPD to be older, have better mental health including depression, and more actively participated in their health care than current smokers with COPD (Adams et al., 2006). In contrast to the findings of Adams et al., we found coronary heart disease was less frequent among former smokers than current smokers. In both of these studies, the use of cross-sectional data, where temporal precedence cannot be established, renders it impossible to tell whether the presence of other chronic medical health conditions was part of the causal path to smoking cessation for the former smokers. Thus, it is possible that the additional comorbidities and/or more active participation in health care may have triggered increased cessation advice or interventions from healthcare providers, which led to the respondents becoming former smokers.

Cigarette smoking is widely considered to be the most significant risk factor for COPD. Not surprisingly, we found slightly less than 20% of adults with diagnosed COPD were never smokers; while more than 80% of persons with diagnosed COPD were either current or former smokers. These findings are consistent with previous research where between 10 and 20 percent of all COPD cases are attributable to occupational exposures (Blanc, 2012; Hnizdo et al., 2002) while the population attributable risk for COPD among current and former smokers was nearly 80% using ERS diagnostic criteria (D. Wilson, Adams, Appleton, & Ruffin, 2005). In another study, we found a significantly higher proportion of smokers among those with moderate or more severe lung obstruction (56%) when compared to those with mild lung obstruction (41%), without consideration of diagnosis status (Paulose-Ram et al., 2015). In the current study, after taking diagnosis into account, we found no difference in smoking status between those with moderate or more severe lung obstruction (55.5% current smokers) and those with mild lung obstruction (50.4% current smokers). In their study focused on intention to quit smoking among patients with COPD, Melzer et al. found symptom burden to have a stronger association with intention to quit smoking than spirometry-based COPD severity (Melzer et al., 2016). Our findings, along with those of Melzer et al., suggest that respiratory symptoms, self-reported health status, and the presence of co-morbid conditions may be better predictors of motivation and success of smoking cessation interventions among persons with COPD than the presence, or severity, of spirometrically-measured lung obstruction.

Among persons with diagnosed COPD, we found having other smokers in the home was the most significant predictor of being a current smoker. We believe this finding is an important and novel outcome as we were unable to locate a similar result in any of the literature reviews we assessed. Related to this finding, Hilberink and colleagues conducted a study using the

Transtheoretical Model to categorize the Stages of Change of smokers with COPD. Their results indicated those individuals who were motivated to quit had considered coping strategies like requesting that guests in their home not smoke and constructing no-smoking agreements with other members of the household. These results suggest that motivated smokers recognized the impact other smokers in the household had on their successful smoking cessation attempts and of ultimately remaining smoke free (Hilberink et al., 2006). The findings further suggest that health-care providers and researchers should target smoking cessation interventions at the household level and address interpersonal influences, as well, at the individual level. By focusing on household members quitting smoking, interventions at this level may be more successful at achieving long-term, sustained smoking cessation in the individual smoker.

One limitation of the study reported here is the data from NHANES are cross-sectional, rendering it impossible to determine causal relationships. Additionally, the questionnaire data are subject to self-report limitations including, but not limited to, respondents being too embarrassed to reveal information, respondents not being able to remember critical details, and social desirability bias. There were factors associated with smoking status in the general population that were not available to this study; most notably history of quit attempts and desire to quit smoking. Lastly, in order to assess the impact of lung obstruction severity on smoking status, we utilized spirometry data, which was last collected by the NHANES in 2012. Though this lack of recent data is certainly a limitation, the ability to evaluate smoking status by severity is a definite strength of this research. Perhaps the greatest strength of this study is the breadth, and national representativeness, of the data. The wide range of sociodemographic, behavioral, smoking, and examination data available in the NHANES allowed us to find the substantial association between other smokers in the home and continued smoking among those with diagnosed COPD.

We believe the research as a whole, and this finding in particular, contributes to an increased understanding of the types of interventions that could improve smoking cessation rates in this high-risk population.

4.5 Conclusions

Current and former smoking status proportions did not differ by measured lung obstruction severity among those reporting a COPD diagnosis. In adjusted logistic regression analyses, multiple factors were found to be associated with current smoking status among those with self-reported COPD with the presence of other smokers in the household having the strongest association with being a current smoker.

Table 4.1. Characteristics of ever-smokers, aged 40-79 years, with self-reported, health professional-diagnosed COPD: NHANES, 2007-2012

Characteristics (n=465 persons)	Proportion* of current smokers with self-reported COPD[a]	Proportion* of former smokers with self-reported COPD[a]	Design-based chi-square[b] p-value
Gender			0.699
Male	50.3% (5.4)	53.4% (5.2)	
Female	49.6% (5.4)	46.6% (5.2)	
Race and ethnic origin			0.361
Non-Hispanic white	81.0% (3.4)	85.9% (2.3)	
Non-Hispanic black	10.6% (2.2)	7.4% (1.3)	
Hispanic	3.4% (0.8)	3.4% (0.9)	
Other non-Hispanic	5.0% (1.9)	3.3% (1.3)	
Age in years			< 0.001
40-59	62.7% (3.7)	23.9% (3.7)	
60-79	37.3% (3.7)	76.1% (3.7)	
Education			0.039
Less than 12th grade	42.4% (4.1)	35.0% (4.5)	
HS Grad/GED/Equivalent	28.6% (3.9)	20.2% (4.0)	
Some College or AA degree	29.0% (4.6)	44.8% (5.6)	
Annual family income			0.104
Less than $20,000	38.7% (4.3)	25.6% (3.9)	
$20,000 to $99,999	57.6% (3.8)	69.3% (4.1)	
$100,000 or more	3.8% (2.0)	5.0% (2.0)	
Disease reported			0.544
Emphysema	49.2% (4.3)	42.1% (5.4)	
Chronic bronchitis	34.0% (4.9)	37.9% (4.5)	
Both	16.8% (3.0)	20.0% (3.7)	
On prescription respiratory medication	56.4% (3.6)	49.9% (4.5)	0.297
Self-reported chronic medical conditions			
Coronary heart disease	17.0% (2.9)	10.0% (2.0)	0.061
Stroke	11.9% (2.6)	12.1% (2.9)	0.972
Cancer/malignancy	17.5% (2.9)	29.8% (3.7)	0.012
Diabetes	24.8% (4.1)	37.0% (4.3)	0.038
Hypertension	46.7% (3.6)	69.0% (3.7)	0.001
One or more conditions	70.4% (4.0)	84.5% (2.3)	0.003
Number of chronic medical conditions			0.011
0	29.6% (4.0)	15.5% (2.3)	
1	37.6% (2.9)	35.0% (3.8)	
2	19.5% (3.2)	30.6% (4.4)	

3 or more	13.3% (3.0)	18.9% (3.3)	
Respiratory symptoms			
Chronic cough	55.0% (5.6)	32.1% (4.1)	0.008
Chronic phlegm	43.6% (4.7)	27.1% (3.5)	0.002
Wheezing episodes past year			0.108
0	32.8% (4.1)	44.6% (5.1)	
1-3	22.5% (2.7)	22.3% (4.0)	
4 or more	44.7% (4.6)	33.2% (5.1)	
Dyspnea	85.5% (3.0)	75.0% (4.6)	0.067
One or more respiratory symptoms	93.4% (1.8)	91.5% (2.4)	0.460
Number of respiratory symptoms			< 0.001
0	6.6% (1.8)	8.5% (2.4)	
1	18.4% (3.3)	37.4% (4.5)	
2	24.7% (3.1)	24.1% (3.7)	
3 or more	50.3% (4.9)	30.0% (3.9)	
Depression			0.001
Minimal	44.5% (4.9)	66.9% (3.6)	
Mild or worse	55.5% (4.9)	33.1% (3.6)	
Self-reported general health condition			0.001
Good or better	35.0% (4.8)	54.5% (3.8)	
Fair or poor	65.0% (4.8)	45.5% (3.8)	
Other smokers in household			< 0.001
No	32.5% (3.7)	90.6% (2.5)	
Yes	67.5% (3.7)	9.4% (2.5)	
Healthcare visits in past year			< 0.001
0-2	32.2% (3.3)	19.7% (3.4)	
3 or more	67.8% (3.3)	80.3% (3.4)	

* Standard errors of the estimates are given in parentheses.

a Self-reported, health professional-diagnosed COPD was assessed by an affirmative response to either ever having emphysema (MCQ160G) or still having chronic bronchitis (MCQ170K).

b Chi-square values are based on the difference in the percentages between current smokers diagnosed with COPD and former smokers with COPD. For the condition and symptom variables, it is the difference in the percentage of current/former smokers with COPD between those reporting/not reporting the condition or symptom.

Table 4.2. Smoking status proportions* in the U.S. population, aged 40-79 Years, by self-reported COPD diagnosis status and measured lung obstruction severity: NHANES, 2007-2012

Diagnosis status / severity of lung obstruction	Current Smoker	Former Smoker	Never Smoker	COPD diagnosis chi-square[a] p-value	All lung obstruction severity chi-square[b] p-value	Mild to moderate or worse lung obstruction severity chi-square[c] p-value
Overall	24.0% (0.9)	26.4% (0.8)	49.5% (0.8)			
Diagnosed with COPD	47.1% (3.1)	33.4% (2.8)	19.6% (1.6)	< 0.001		
Mild lung obstruction	50.4% (10.0)	37.1% (11.9)	12.4% (8.4)		0.006	0.735
Moderate or worse lung obstruction	55.5% (5.9)	37.4% (6.3)	7.2% (2.9)			
No measured lung obstruction	37.3% (4.2)	32.1% (3.9)	30.6% (3.0)			
Not diagnosed with COPD	22.7% (0.8)	26.0% (0.8)	51.3% (0.8)			

* Standard errors of the estimates are given in parentheses.

[a] Chi-square values computed for differences in smoking proportions between those with/without self-reported, healthcare provider-diagnosed COPD.

[b] Chi-square values computed for differences in smoking proportions among those having mild, moderate or worse, and no measured lung obstruction.

[c] Chi-square values computed for differences in smoking proportions between those diagnosed with COPD and having mild lung obstruction and those diagnosed with COPD and having moderate or worse lung obstruction.

Table 4.3. Logistic regression results[a] predicting status of current smoker among ever-smokers, aged 40-79 years, with self-reported, health professional-diagnosed COPD[a]: NHANES, 2007-2012

Characteristic (n=430 persons in model)	Adjusted Odds Ratio (95% CI)	*p*-value
Age in years		
40-59	Ref	-
60-79	0.22[b] (0.12, 0.41)	< 0.001
Reported hypertension		
No	Ref	-
Yes	0.28[b] (0.13, 0.60)	0.001
One or more self-reported chronic conditions		
No	Ref	-
Yes	2.20[b] (1.01, 4.75)	0.046
Self-reported general health condition		
Good or better	Ref	-
Fair or poor	1.99[b] (1.07, 3.70)	0.031
Other smokers in the household		
No	Ref	-
Yes	19.51[b] (10.16, 37.47)	< 0.001

[a] Self-reported, health professional-diagnosed COPD was assessed by an affirmative response to either ever having emphysema (MCQ160G) or still having chronic bronchitis (MCQ170K).

CHAPTER V.

STUDY 3: CHRONIC OBSTRUCTIVE PULMONARY DISEASE (COPD) PHENOTYPES AND THEIR IMPACT ON CONTINUED SMOKING AFTER DIAGNOSIS

5.1 Introduction

Chronic obstructive pulmonary disease (COPD) is a leading cause of morbidity and mortality around the globe and is projected to become the third leading cause of mortality worldwide by the year 2020 (GOLD, 2018). COPD is characterized by airflow obstruction that is not fully reversible, does not change markedly over time, and that is usually progressive (GOLD, 2018). Cigarette smoke is widely recognized as the main risk factor for COPD although non-smokers with exposures to other noxious gases and particles (e.g. air pollution and biomass fuel exposure) may also develop the disease (Lamprecht et al., 2011; Liu et al., 2017).

Severity of airflow obstruction, assessed by forced expiratory volume in 1 second (FEV1), is required for the staging of COPD (GOLD) with the rate of decline of FEV1 often used as an indicator of disease progression (Adis Medical Writers, 2015). However, FEV1 does not sufficiently reflect all the markers of COPD, and may be inadequate for estimation of disease severity and prediction of outcomes. A multitude of studies have shown that FEV1 does not always correlate well with a variety of things important for patients such as quality of life (Jones et al., 1992; Mahler et al., 1992), exacerbation frequency (Alsaeedi et al., 2002), and exercise intolerance (O'Donnell et al., 1999).

FEV1 is also not a good surrogate for symptoms and, in some studies, has actually been shown to correlate poorly with symptoms (Viegi et al., 2000). With respect to continued smoking among those diagnosed with COPD, Tilert et al. found no association between disease severity as measured by FEV1 and smoking status (Tilert, under review). The consistent message across all these findings is that FEV1 alone may not be the best tool for predicting outcomes among those suffering from COPD and that other factors in addition to FEV1 should be considered.

Expanding on the ideas that COPD is a complex, multifaceted disease and that severity defined by FEV1 alone is insufficient to describe the heterogeneity of the disease, identifying phenotypes of COPD with the goal of targeted treatments has now garnered significant research interest (Pinto et al., 2015). A COPD phenotype has been defined as "a single or combination of disease attributes that describe differences between individuals with COPD as they relate to clinically meaningful outcomes (symptoms, exacerbations, response to treatment, speed of progression of the disease or death)" (Han et al., 2010). By using a combination of attributes,, not simply FEV1, to classify and group patients, individuals within the same classification, or phenotype, are more likely to have a similar disease progression and response to treatments (Jørgen Vestbo, 2014).

The objectives of this research were to: 1) determine phenotypes of COPD from the NHANES data; 2) compare these derived phenotypes with those from other studies and data sources; and 3) determine whether COPD phenotypes are differentially

associated with continued smoking among person diagnosed with COPD above and beyond other associated factors.

5.2 Methods

Study and data

This study was based on analyses of the 2007–2012 National Health and Nutrition Examination Survey (NHANES) data, which were the years spirometry testing was conducted and included in the survey. NHANES is a cross-sectional survey of the civilian, non-institutionalized U.S. population conducted by the National Center for Health Statistics ("NHANES - National Health and Nutrition Examination Survey Homepage," 2018). Data were collected via household interviews and standardized physical examinations in specially equipped mobile examination centers (MEC). The NHANES samples are selected through a complex, multistage, probability design. The 2007–2012 NHANES cycles oversampled major U.S. demographic subgroups including Hispanic and black persons, low income white persons, and persons aged 80 years and older. Additionally, Asian persons were oversampled in 2011-2012. The procedures to select the sample and conduct the interview and examination have been specified elsewhere (Zipf et al., 2013). Informed consent was obtained from all participants and the NCHS Research Ethics Review Board approved the protocol ("NHANES - NCHS Research Ethics Review Board Approval," 2017).

Analyses were conducted on data collected from adults aged 40 to 79 years old. The lower boundary of the analyzed age cohort (40 years) was based on guidelines for

diagnosis, which recommend considering spirometry in symptomatic patients over the age of 40 years (GOLD, 2018). The upper boundary of the analyzed age cohort (79 years) was selected as this was the maximum eligible age for the NHANES spirometry examination.

For the period of 2007–2012, 14,825 persons aged 40–79 years were eligible for the survey, 10,531 (71%) were interviewed, 10,219 (69%) were examined in the NHANES mobile examination center (MEC) and 566 (5.5%) reported having healthcare-provider diagnosed COPD.

Variables considered for feature selection and analyses

The target population for these analyses were persons reporting a diagnosis of COPD. A participant was classified as having healthcare provider-diagnosed COPD if they responded affirmatively to either of the following questions on emphysema or chronic bronchitis: "Has a doctor ever told you that you have emphysema?" or "Do you still have chronic bronchitis?" (Halldin et al., 2015; Mannino et al., 2017).

Feature selection for the cluster analyses considered those measures collected in the NHANES that were shown to be relevant in prior COPD phenotyping studies as well as a number of variables often associated with smoking, our outcome of interest. The list of measures included were: lung function (assessed by percent of predicted FEV1), number of respiratory symptoms reported, number of comorbid chronic conditions reported, dietary behavior, general health condition, age, body mass index (BMI), asthma status, sex, height, number of healthcare visits in past year, pack-years of smoking, PHQ9

depression index score, daily minutes of sedentary activity reported, presence of impaired activities of daily living, education level, dyspnea reported, taking prescribed respiratory medication, eosinophilia (airways inflammation measure), emphysema reported and chronic bronchitis reported (Basagaña, Barrera-Gómez, Benet, Antó, & Garcia-Aymerich, 2013; Burgel et al., 2010; Han et al., 2010; M. Miravitlles, Barrecheguren, & Román-Rodríguez, 2015; Pinto et al., 2015).

Combining levels of categorical (factor) variables

As cluster analysis requires the use of continuous or binary variables, a number of relevant categorical, or factor, variables were converted to binary variables. The cutoffs used in the data conversions to binary were those that had exhibited relevance in prior COPD phenotyping studies and/or resulted in sufficient cell sizes for analysis. The list of categorical/factor variables that were combined are as follows:

Dietary behavior ("How healthy is your diet?"): Fair or Poor = 1; Good, Very good or Excellent = 0 (Pinto et al., 2015; Vermylen & Kalhan, 2013)

General health condition ("Would you say your health in general is…?"): Fair or Poor = 1; Good, Very good or Excellent = 0 (M. Miravitlles et al., 2015; Pinto et al., 2015)

Number of healthcare visits in past year ("During the past 12 months, how many times have you seen a doctor or other health care professional about your health at a doctor's office, a clinic, hospital emergency room, at home or some other place?"): 4 to 9, 10 to 12 or 13 or more = 1; None, 1 or 2 to 3 = 0 (Burgel et al., 2010; Sundh et al., 2015)

Education level ("What is the highest grade or level of school you have completed or the highest degree you have received?"): Less than 9th grade or 9-11th grade (includes 12th grade with no diploma) = 1; High school graduate/GED or equivalent, Some college or AA degree or College graduate or above = 0 (Cokkinides, Halpern, Barbeau, Ward, & Thun, 2008; Mannino, Ford, & Redd, 2003)

Statistical methods

Statistical analyses were performed using R™ version 3.2.3 ("R: The R Project for Statistical Computing," 2015) and STATA™ version 13.1 (StataCorp, College Station, TX). For all regression analyses, examination sample weights were used to account for differential probabilities of selection, nonresponse, and non-coverage. Similarly, Taylor series linearization was used to calculate standard errors to account for the complex sampling design.

Given that the variables used were measured on different scales (e.g., BMI ranges from approximately 20-40kg/m2 while height ranges from approximately 120-200cm), the variables were first transformed to z-scores, prior to the cluster analyses, to avoid the overpowering of variables with larger scales on the resulting classifications (Garcia-Aymerich et al., 2011; Hancock et al., 2010).

As cluster analysis solutions can be sensitive to outliers (Milligan, 1980; Rocke & Woodruff, 1996), multivariate outliers were removed prior to conducting the analyses. Mahalanobis distances to the centroid of the dataset were computed for each participant, using all 21 variables, with the corresponding critical chi square value (Rousseeuw &

Zomeren, 1990) used as the cut-point for determining multivariate outliers. Nine participant records out of the 566 were removed from the cluster analyses based on their multivariate outlier status.

In prior phenotyping (cluster) analyses of COPD patients, the most prevalent range of solutions ranged from 2 to 5 clusters (Basagaña et al., 2013; Pinto et al., 2015). Based on these earlier findings, cluster analyses were conducted using all twenty combinations of k-means, hierarchical and partitioning around medoids (PAM) algorithms and 2 through 5 as number of clusters. Maximum silhouette width, a measure of how close each subject in one cluster is to subjects in adjacent clusters, was the criterion used to determine the optimal clustering approach. The resulting optimal solution was the *k*-means clustering algorithm with 3 clusters.

Once the feature selection was completed, the clustering algorithm was chosen, and the optimal number of clusters were determined; *k*-means clustering was used to identify the unique subgroups of COPD. *K*-means is an unsupervised approach to partitioning data into unique clusters using an iterative algorithmic approach to identify the optimal cluster (subgroup) number (*k)*. Briefly, each observation contains data for *m* continuous clustering variables (e.g., age, weight, and height, $m=3$). Each observation is placed into a *m*-dimensional field wherein each *m* is an axis. Initial centroids (*k* starting points) are placed in the data space at random and each observation is assigned to its nearest centroid. After all observations have been assigned to their nearest centroids, new centroids are computed which are the re-computed centers of this new grouping of

observations. With these new centroids, the process begins again, with each observation assigned to the nearest centroid and with the centroids recalculated to the centers of their new groupings. Once the solution converges, i.e., centroids no longer move after re-assigning observations, the clustering is complete and the groups (phenotypes) have been determined.

To properly evaluate the impact of phenotype in the final model, it was important that the same population sample be used in the final regression model as was used in the preliminary model. As such, it was required to have a cluster or phenotype assignment for each person. To ensure that each of the persons in the regression sample received a phenotype (cluster) assignment in the presence of missing data, multiple imputation using chained equations was used to create 40 imputed datasets for the analyses. As cluster assignments can vary between datasets, rendering the classic combination of imputed dataset results per Rubin's rules nearly impossible, the imputed datasets were stacked with the final cluster solution for each participant being the majority cluster membership from the stacked imputation dataset results (Basagaña et al., 2013).

After identifying the COPD subgroups (phenotypes), the potential predictive power of these phenotypes on continued smoking among those diagnosed with COPD was assessed. In a prior study (Tilert et al, under review), a multivariable logistic regression model was conducted that determined significant predictors of continued smoking among those diagnosed with COPD. This model was re-run, with the addition of the newly-computed COPD phenotype (factor with three levels) to assess the impact of

COPD phenotype on smoking status in this population after controlling for other known significant predictors.

5.3 Results

Characteristics of US adults aged 40-79 years diagnosed with COPD, by phenotype

Classification of 557 NHANES participants reporting a COPD diagnosis was conducted using *k*-means cluster analysis. Silhouette width statistics determined that the data could be optimally grouped into three clusters (phenotypes). Characteristics of the 557 COPD subjects according to these three phenotypes are presented in Table 5.1. Three unique phenotypes were identified with significant differences in selected characteristics found among the groups. The first phenotype (n=204) were primarily female (84.3%), younger (mean age 58.7 years) with chronic bronchitis (89.1%) and asthma (64.0%), with impaired activities of daily living (75.9%), frequently reported fair or poor health (76.8%) and reported more than 3 healthcare visits in the past year (85.3%). The second phenotype (n=121) was composed of more mixed gender participants (66.9% female), with chronic bronchitis (76.9%), low prevalence of dyspnea (19.8%), low prevalence of symptoms (0.9 reported symptoms on average), unimpaired activities of daily living (18.5%), lower levels of smoking among those that were current and former smokers (23.8 pack-years on average) and above average lung function (86.5 percent of predicted FEV1). Phenotype 3 was composed of mostly older (65.4 years on average), male (79.3%) participants, with emphysema (87.9%), high levels of smoking

among current and former smokers (49.5 pack-years on average) and higher percentage of subjects with less than a high school education (47.0%).

Predictors of current smoking status among ever-smokers, aged 40-79 years and diagnosed with COPD

Table 5.2 details the adjusted odds ratios, 95% confidence intervals, and associated *p*-values from multivariable logistic regression modeling used to examine the association of various covariates with smoking status. After adding COPD phenotype as a factor to the model predicting smoking status, the odds ratios for all the variables were of similar magnitudes to those from the model not including phenotype with two variables (one or more comorbid chronic conditions and self-reported general health condition) no longer meeting the predetermined .05 significance threshold in the model including phenotype. After controlling for the other variables in the model, persons aged 60-79 years were 85% less likely (model including phenotype AOR=0.15, 95% CI = 0.08 to 0.28; model without phenotype AOR=0.22, 95% CI = 0.12 to 0.41) than persons aged 40-59 years to be current smokers. Persons reporting having hypertension were 68% less likely (model including phenotype AOR=0.32, 95% CI = 0.16 to 0.64; model without phenotype AOR=0.28, 95% CI = 0.13 to 0.60) than persons not reporting hypertension to be current smokers. Persons reporting one or more comorbid chronic conditions were 2.01 times as likely (model including phenotype AOR=2.01, 95% CI = 0.90 to 4.48; model without phenotype AOR=2.20, 95% CI = 1.01 to 4.75) as those reporting no comorbid conditions to be current smokers. Note that while this result was significant at

the p=.05 significance level in the model without phenotype, this result was no longer statistically significant in the model including phenotype. Persons reporting fair or poor general health were 1.50 times as likely (model including phenotype AOR=1.50, 95% CI = 0.78 to 2.87; model without phenotype AOR=1.99, 95% CI = 1.07 to 3.70) as those reporting good or better general health to be current smokers. Similar to comorbid chronic conditions, the result for fair or poor self-reported health was significant at the p=.05 significance level in the model without phenotype but no longer statistically significant in the model including phenotype. Persons with other smokers in the household were 20.18 times as likely (model including phenotype AOR=20.18, 95% CI = 10.55 to 38.61; model without phenotype AOR=19.51, 95% CI = 10.16 to 37.47) to be current smokers than those with no other smokers in the household. Finally, those with the phenotype of younger, females with asthma and COPD overlap (ACOS) were 2.46 times as likely (AOR=2.46, 95% CI = 0.94 to 6.44; result not statistically significant at the p=.05 significance level) and those with the older, heavy-smoking, males with emphysema (OSME) phenotype were 3.71 times as likely (AOR=3.71, 95% CI = 1.26 to 10.88) to be smokers than those with the mixed-gender with chronic bronchitis and no dyspnea (CBND) phenotype after controlling for the other predictors.

5.4 Discussion

It has been shown in a multitude of studies that FEV1 alone may be insufficient for predicting outcomes among those suffering from COPD and for customizing treatments and interventions in this population (Alsaeedi et al., 2002; Jones et al., 1992;

Mahler et al., 1992; O'Donnell et al., 1999). In light of this, phenotyping has been proposed as a tool to potentially better predict outcomes and customize treatments and interventions for sufferers of COPD as phenotypes are based on a multidimensional set of factors rather than a single measure of disease severity (Pinto et al., 2015).

In the current study we identified three unique phenotypes among adults diagnosed with COPD: 1) younger, females, with chronic bronchitis and asthma, with impaired activities of daily living, who frequently reported fair or poor health and who reported more than 3 healthcare visits in the past year; 2) mixed gender persons, with chronic bronchitis, a low prevalence of dyspnea, a low prevalence of symptoms, with unimpaired activities of daily living, with lower levels of smoking among those that were current and former smokers and with above average lung function; and 3) older males, with emphysema, high levels of smoking among current and former smokers and a higher prevalence of having less than a high school education.

Our first phenotype (ACOS) closely resembles a phenotype described in a number of other studies (Gibson & McDonald, 2014; M. Miravitlles et al., 2015; Pinto et al., 2015). In their study utilizing data from primary care physicians and pneumologists, Miravitlles et al. described their ACOS phenotype as more frequently female (86% female in the current study's ACOS phenotype vs. the study average of 54%), smoked less (27 pack-years of smoking in the current study's ACOS phenotype vs. the study average of 37 pack-years) and had a higher FEV1% (80% in the current study's ACOS phenotype vs the study average of 77%) than other phenotypes. Additionally, they found

positive associations with their ACOS phenotype and both depression and health related quality of life (HRQoL). Both of these associations mirror similar relationships in the current study's ACOS phenotype which possessed depression scores roughly twice as high as the other two phenotypes combined and with nearly twice as many reporting fair or poor health in the ACOS phenotype as the other two phenotypes combined.

While we believe our measure of reporting fair or poor health is somewhat analogous to Miravitlles' measure of HRQoL, it is important to note that these measures may not, in fact be the same. Using similar but potentially different measures is one of several methodological challenges surrounding COPD phenotyping. One cause of these differences in measures is that many COPD studies are conducted in a clinical setting and are heavily reliant on imaging and other complex pulmonological measurements (Bon, Liao, Tseng, & Sciurba, 2013; Burgel et al., 2010; Han et al., 2010) whereas population-based studies like NHANES are more reliant on simpler biological measures and questionnaire data. An additional challenge with COPD phenotyping is that phenotype descriptions between studies often do not match *precisely*. That is, the phenotypes being compared may match on similar, but not identical, characteristics or they may match on many or most, but not all, of the group characteristics. Any mismatches in phenotype characteristics, or measures used, raises the possibility that the compared phenotypes may not be the same. Subtle mismatches in measures and characteristics also raise the possibility of additional, undiscovered granularity (i.e., subtypes) between broader phenotypes. For example, comparing our ACOS phenotype to one phenotype reported in

a study by Burgel et al (2010) revealed a precise match on age (58 years), an imprecise match on severe airflow limitation (they used GOLD stage 3 and 4 to define airflow limitation while we used PPFEV1), a mismatch on BMI (low in theirs vs higher in ours), an imprecise match on severe dyspnea (we had *prevalent* dyspnea we did not assess dyspnea *severity*), an imprecise match on frequent exacerbations (85% our phenotype had more than 3 healthcare visits in the past year though we could not ascertain if these were specifically exacerbations) and an imprecise match on severely impaired HRQoL (over 75% in our ACOS phenotype reported fair or poor health though we did not explicitly measure HRQoL). Given one precise match, four imprecise matches and one mismatch among the six characteristics presented, it is difficult to make a determination whether these two groups represent the same phenotype, different sub-phenotypes, or entirely different groups altogether. As the field of COPD phenotyping is still relatively new, further study and, more importantly, validation of phenotypes against multiple health outcomes will be required to clear up some of these methodological questions of how many characteristics are needed to define a phenotype, what specific characteristics and measures determine a phenotype, what criteria determines a match/mismatch between different studies and how to differentiate between sub-phenotypes, if any exist.

A number of researchers have suggested that, in addition to phenotypes being able to classify persons into distinct groups, they should also be able to predict clinically meaningful health outcomes and, ultimately, influence treatment decisions (Barker & Brightling, 2013; Bourbeau et al., 2014; Han et al., 2010). In terms of predicting

outcomes, the phenotypes we discovered in the NHANES data proved to be significant in models predicting current vs former smoking status among those diagnosed with COPD even after controlling for other factors. This illustrates the potential benefit of COPD phenotyping in that unique COPD phenotypes may help predict health outcomes above and beyond individual factors. Ideally, these phenotypes could subsequently be used for targeting treatments and interventions (smoking cessation, in our case) to individuals according to his/her phenotype. One example of how COPD phenotype information was used successfully to target treatments comes from results of the National Emphysema Treatment Trial where a select group of COPD patients with emphysema specifically of the upper lobes and low baseline exercise capacity obtained a survival benefit from lung volume reduction surgery (LVRS) and LVRS is subsequently directed at this COPD phenotype (Barker & Brightling, 2013; Fishman et al., 2003). In another example of targeted treatments for specific COPD phenotypes, a prospective clinical trial of the anti-inflammatory drug roflumilast showed a reduction in exacerbations for a subgroup of patients specifically with chronic bronchitis symptoms, frequent exacerbations and severe airflow obstruction and for whom roflumilast therapy is now targeted (Barker & Brightling, 2013; Calverley et al., 2009). Based on our results and the results from these and other studies, and in spite of its methodological challenges, COPD phenotyping shows great promise for predicting health outcomes above and beyond individual factors and may prove to be a useful tool in targeted treatments and interventions for individuals in this high-risk population.

One combination strength/limitation of the present study is that cluster analysis, an unsupervised technique, was used to determine the phenotypes. As Bourbeau et al noted, while clusters generated are less prone to bias from a priori hypotheses than are the results of analyses based on hypotheses (strength), they can be susceptible to spurious groupings, can be created from virtually any dataset, and could match selection biases (limitation) (Bourbeau et al., 2014). A second limitation, alluded to earlier in the discussion, is that there are no consensus rules on how to define or compare phenotypes so there remains a great deal of subjectivity in both the classifications of phenotypes as well as the comparisons to other studies. A great strength of our study is that the NHANES data covered a wide spectrum of persons both by demography and by disease status and severity. Many of the existing COPD phenotyping studies were performed in clinical settings and were limited to patients with severe disease thus limiting the generalizability of their results. When balancing the strengths and weaknesses of the study, we believe the research as a whole, contributes to an increased understanding of COPD phenotypes and their relationship with continued smoking status in this high-risk population.

5.5 Conclusions

Three distinct phenotypes exist among the COPD population analyzed. One phenotype (younger, females with asthma and COPD overlap), in particular, was similar to phenotypes found in a number of other studies. In adjusted logistic regression analyses, COPD phenotype was differentially associated with continued smoking with the older,

heavy-smoking, males with emphysema phenotype showing a significant positive association with continued smoking in this population.

Table 5.1. Characteristics of US adults aged 40-79 years diagnosed with COPD, by phenotype: NHANES, 2007-2012

Characteristic	Overall (n=557)	Cluster 1 (n=204)	Cluster 2 (n=121)	Cluster 3 (n=232)	*F* Value[a]	*p*-value (pr > *F*)
% With dyspnea (self-report)	79.0	94.6	19.8	96.1	391.5	< 0.001
% Female	54.0	84.3	66.9	20.7	140.3	< 0.001
% With emphysema	55.3	32.2	31.1	87.9	123.3	< 0.001
% With chronic bronchitis	63.3	89.1	76.9	33.0	109.2	< 0.001
Height (in cm)	166.5	161.0	165.3	172.0	102.5	< 0.001
Number of respiratory symptoms	2.1	2.6	0.9	2.3	99.3	< 0.001
% Impaired activities of daily living	50.0	75.9	18.5	44.2	63.0	< 0.001
Depression score (PHQ9)	6.2	9.4	3.8	4.8	49.5	< 0.001
Pack years of smoking [b]	37.0	27.4	23.8	49.5	41.4	< 0.001
% Reporting fair or poor health	59.1	76.8	33.3	57.9	31.3	< 0.001
% With asthma	43.6	64.0	33.9	30.6	30.3	< 0.001
Age (in years)	61.7	58.7	59.7	65.4	26.3	< 0.001
Body Mass Index	30.4	33.2	27.9	29.1	24.3	< 0.001
% Predicted FEV1	76.9	79.9	86.5	68.1	24.0	< 0.001
% Taking prescription respiratory medication	49.4	52.5	25.6	59.1	19.6	< 0.001
Number of chronic medical conditions	1.4	1.6	1.0	1.5	11.5	< 0.001
More than 3 healthcare visits in past year	74.1	85.3	67.8	67.7	10.8	< 0.001
% Reporting fair to poor diet	39.0	51.0	33.1	31.5	10.1	< 0.001
% Less than high school education	39.7	38.2	28.1	47.0	6.2	0.002
Daily minutes of sedentary activity	369.7	370.6	325.9	391.6	4.0	0.018
Eosinophils	0.23	0.22	0.22	0.26	3.2	0.040

[a] *F* values represent the ratio of the between-cluster variance to the within-cluster variance of the variable, with higher values representing higher relevance of the variable for separating cluster groups. *F* values were obtained with linear regression models using each variable as the outcome and the cluster group as the exposure.

[b] Pack years of smoking only computed for current and former smokers

Table 5.2. Logistic regression results predicting status of current smoker among ever-smokers, aged 40-79 years, with self-reported, health professional-diagnosed COPD: NHANES, 2007-2012

Characteristic (n=422 persons in model)	Model before Phenotype		Model including Phenotype	
	Adjusted Odds Ratio (95% CI)	*p*-value	Adjusted Odds Ratio (95% CI)	*p*-value
Age in years				
40-59	Ref	-	Ref	-
60-79	0.22 (0.12, 0.41)	< 0.001	0.15 (0.08, 0.28)	< 0.001
Reported hypertension				
No	Ref	-	Ref	-
Yes	0.28 (0.13, 0.60)	0.001	0.32 (0.16, 0.64)	0.002
One or more comorbid chronic conditions				
No	Ref	-	Ref	-
Yes	2.20 (1.01, 4.75)	0.046	2.01 (.90, 4.48)	0.087
Self-reported general health condition				
Good or better	Ref	-	Ref	-
Fair or poor	1.99 (1.07, 3.70)	0.031	1.50 (0.78, 2.87)	0.215
Other smokers in the household				
No	Ref	-	Ref	-
Yes	19.51 (10.16, 37.47)	< 0.001	20.18 (10.55, 38.61)	< 0.001
COPD Phenotype				
Mixed sex and disease without dyspnea (Cluster 2)	-	-	Ref	-
Females w/chronic bronchitis (Cluster 1)	-	-	2.46 (0.94, 6.44)	0.065
Males w/emphysema (Cluster 3)	-	-	3.71 (1.26, 10.88)	0.018

CHAPTER VI.

SUMMARY

6.1 Overview & Summary

Understanding how differences in disease severity and disease phenotypes impact continued smoking and smoking cessation among those with COPD could provide useful guidance for more targeted, and effective interventions aimed at reducing smoking rates in this high-risk group. With this in mind, the overarching goal of the research conducted for this dissertation was to assess the roles that disease severity and COPD phenotype play in smoking status among persons diagnosed with COPD.

In study 1, characteristics of adults aged 40-79 from a nationally representative sample of U.S. adults were examined to determine factors associated with healthcare provider diagnosed COPD. In the period 2007-2012, 5.2% of adults between the ages of 40 and 79 years reported receiving a COPD diagnosis. Of those diagnosed, 50.8% reported fair or poor health, 47.1% were still smoking cigarettes, 49.1% were taking prescription respiratory medicine, 36.4% had moderate or worse lung obstruction, and nearly 90% had at least one respiratory symptom. Compared to those without diagnosed COPD, more persons diagnosed with COPD reported fair or poor health (50.8% vs 17.3%), reported having four or more healthcare visits in the past year (72.6% vs. 40.7%) and were current or former smokers (80.5% vs. 48.7%). Notably, as the number of respiratory symptoms increased, the prevalence of diagnosed COPD increased

dramatically. Specifically, prevalence was less than 1% (0.9%) when no respiratory symptoms were reported, 5.8% with 1 symptom, 14.9% with 2, and 36.7% when 3 or more symptoms were reported. Compared to persons with no symptoms, persons reporting 3 or more symptoms were 22 times as likely to be diagnosed with COPD. This finding suggests that COPD diagnosis is largely driven by patient symptoms rather than from a clinically objective measure like spirometry. Finally, the National Heart, Lung and Blood Institute estimates that as many as 75% of persons with COPD smoke or used to smoke (NHLBI, 2019) while the COPD Foundation reports that 90% of those with COPD have smoked (COPD Foundation, 2019). The current study (study 1) splits the middle of these two estimates showing 80% of those diagnosed with COPD to be current or former smokers.

Study 2 explored the relationship between disease severity and smoking status among those diagnosed with COPD. Contrary to the proposed hypothesis, findings from study 2 did not support the idea that disease severity would have a direct, or moderating, effect on smoking status among those diagnosed with COPD. In fact, there was no statistically significant difference in smoking status between those with mild lung obstruction (50.4% current; 37.1% former; 12.4% never) and those with moderate or worse lung obstruction (55.5% current; 37.4% former; 7.2% never) among persons diagnosed with COPD. There were, however, significant differences in smoking status between those diagnosed with COPD (47.1% current smokers; 33.4% former smokers; 19.6% never smokers) and those not diagnosed with COPD (22.7% current smokers; 26.0% former smokers; 51.3% never

smokers). Additionally, logistic regression results showed significant associations between being a current smoker and reporting at least one co-morbid chronic condition (AOR: 2.2, 95% CI: 1.0-4.8); reporting fair or poor health (AOR: 2.0, 95% CI: 1.1-3.7); and having other smokers in the household (AOR: 19.5, 95% CI: 10.2-37.5).

Finally, in study 3, COPD phenotypes were derived using cluster analysis from the NHANES data, compared to those found in other studies, and added as a predictor to previously developed models of smoking status. Three unique COPD phenotypes were identified including: 1) younger, females with asthma and COPD overlap (ACOS); 2) mixed-gender persons with chronic bronchitis and no dyspnea (CBND); and 3) older, heavy-smoking males with emphysema (OSME). Of the 557 classified adults with COPD, 36.6% were ACOS, 21.7% were CBND, and 41.7% were OSME. The ACOS phenotype, in particular, was similar to phenotypes found in a number of other studies (Gibson & McDonald, 2014; M. Miravitlles et al., 2015; Pinto et al., 2015). When added to the previously developed models of smoking status, the derived COPD phenotypes were found to be differentially associated with continued smoking among those diagnosed with COPD with the OSME group being 3.71 (95% CI = 1.26, 10.88) times as likely to be a current smoker than the CBND group after controlling for age, hypertension, presence of chronic conditions, general health condition and presence of other smokers in the household. The ACOS group was also 2.46 times as likely to be a current smoker than the CBND group though this result was not statistically significant (95% CI = 0.94, 6.44).

6.2 Implications

There are two findings, in particular, from this dissertation that may have direct translational value specifically in terms of guiding smoking cessation interventions in this population. First, the second study revealed a very strong association between having other smokers in the household and continued smoking among those with a COPD diagnosis. The adjusted odds ratio of nearly 20 essentially means that an extremely high percentage of those with diagnosed COPD who continued to smoke had at least one other smoker in the household while, simultaneously, an extremely high percentage of those without a COPD diagnosis did not have another smoker in the household. Note that these are findings from just one study and would need to be validated by additional studies before this association could be taken as truth. Assuming that this association is true, it has obvious and direct implications for smoking cessation interventions in this population. Based on the results presented here, if there are no other smokers in the household, an individual diagnosed with COPD is much less likely to continue smoking than if there are other smokers in the household. Armed with this knowledge, smoking cessation interventions targeted at individuals who have been diagnosed with COPD might improve their likelihood of success by targeting the smoking cessation intervention at the entire household of smokers and not merely the individual.

The second finding with translational implications comes from the third study three where it was shown that there are certain combinations of characteristics, or

phenotypes, that are more likely than others to continue smoking after a COPD diagnosis. Again, these findings are from one study and would need to be validated by other studies before this association could be taken as truth. But, assuming the association to be true, if an interventionist knows the phenotype of a smoker who has been diagnosed with COPD, he or she has unique insight into the propensity for this person to continue smoking. There is some clear logistic, or implementation, utility for interventions with limited funds and/or resources. If these interventions know the phenotypes of the individuals in their program, resources could be directed at those phenotypes that are more likely to quit smoking in an attempt to maximize return on their limited resources. Tailored messaging could also be applied for the different groups during an intervention. For example, for members of the phenotype consisting of predominantly younger females with asthma and COPD overlap, messaging might include information from studies showing that women may be more sensitive to the harmful effects of smoke than men (Chatila et al., 2004; Dransfield et al., 2006) in an effort to target their perceived risk of continued smoking. Additionally, across all the phenotypes, nearly three fourths of the current and former smokers with COPD had more than 3 healthcare visits in the past year. This suggests a clear path of intervention for these individuals. In terms of reach, if smoking cessation interventions could be implemented at the point of healthcare, the vast majority of persons in this population would be a ready audience.

Perhaps the most useful implication of the phenotypic findings is the direction that it points to for future research. The phenotypes discovered here exhibited differences

in behavior (i.e., continued smoking) but it is not clear *why* these characteristic combinations are differentially associated with continued smoking in this population. From a translational perspective, if we can determine the reasons for these behavioral differences, we may be able to derive additional targets for intervention. That is to say, the reasons why people with certain combinations of characteristics continue to smoke, in spite of their compromised lung function, could inform smoking cessation interventions targeted at the root cause of the smoking and not merely the behavior itself.

6.3 Strengths & Limitations

One strength of this research is that it was conducted on a nationally representative sample of persons spanning six years of data collection from 2007 to 2012. In addition to this, most studies involving COPD phenotyping focused on a very specific, and limited, disease severity range while this study is one of very few studies to span the entire range of lung obstruction severities ranging from no lung obstruction to very severe lung obstruction. All of these factors (nationally representative sample, six years' worth of data, and the full range of disease severities) add support to the generalizability to the results obtained here. With regard to the phenotyping analyses, one of the common critiques of most phenotyping studies is the lack of attention paid to missing data. For these phenotyping analyses, this critique was addressed by employing multiple imputation to account for missing data. An additional strength of the research is that all of the analyses conducted in this dissertation considered variables across multiple

dimensions (demographic, socio-behavioral, physical, etc.) and were not limited to clinical data as was done in a number of other COPD phenotyping studies. The wide range of sociodemographic, behavioral, smoking, and examination data available in the NHANES allowed for the derivation of COPD phenotypes which were shown to be differentially associated with continued smoking after a COPD diagnosis. The diversity of the data further allowed for the discovery of a substantial association between other smokers in the home and continued smoking among those with diagnosed COPD.

A limitation of the research and results reported here is that the data used in the analyses (from the NHANES) were cross-sectional rendering it not possible to determine cause-effect or temporal relationships. As a result, the results presented here can only indicate associations and may not reflect a cause-effect relationship. Another limitation of the data used was that a number of data elements employed in these analyses were collected using questionnaires and are subject to the limitations of self-reported data including, but not limited to, respondents being too embarrassed to reveal information, respondents not being able to remember critical details and social desirability bias. Additionally, as NHANES does not include persons from institutionalized settings (e.g., nursing facilities and assisted-care facilities) and because COPD is associated with older age, the results presented here may reflect an underestimation of the true prevalence of diagnosed COPD and could reflect a biased sample of participants, particularly with respect to age. As it relates to the cluster analyses used to determine phenotypes, while clusters generated are less prone to bias from a priori hypotheses than are the results of

analyses based on hypotheses (strength), they can be susceptible to spurious groupings, can be created from virtually any dataset, and could match selection biases (weakness) (Bourbeau et al., 2014). Additionally, as there are currently no consensus rules on how to define or compare phenotypes, there remains a great deal of subjectivity in both the classifications of phenotypes as well as in the comparisons of the derived phenotypes to those derived in other studies. Despite the limitations, this study is believed to be the first to examine the relationships between both disease severity and COPD phenotype with continued smoking behavior in a nationally representative sample of adults diagnosed with COPD. The resulting findings could be informative for smoking cessation interventions targeted at persons in this high-risk population.

6.4 Future Research Directions

As noted above, the findings presented here are from a single study and would need to be validated using other populations, in different areas and conducted at different times. One somewhat obvious future research direction is that of conducting additional studies in an effort to validate some of the findings from this research. Notably, the finding of other smokers in the household being associated with continued smoking among those with COPD is something that could prove to be very useful to smoking cessation interventionists were this association to be validated through additional research.

With respect to phenotyping, it has been noted by several researchers that phenotypes require validation specifically against outcomes. A veritable lifetime of work could be done in this arena to include research surrounding the questions of: What are the true COPD phenotypes? Are there sub-phenotypes? Can there exist different phenotypes for different outcomes? For example, can characteristics be grouped one way for targeting medication adherence and grouped in a different way for targeting smoking cessation outcomes? The answers to these questions could not only prove useful for this specific population (smokers with COPD) but for phenotyping, as a whole.

From a methodological standpoint, a number of critiques of phenotyping have been that most researchers do not take into account missing data when conducting their analyses and simply use complete case analyses. This is not at all surprising given that there are no widely-accepted methods for implementing what is likely the most commonly used, or at least widely known, missing data technique - multiple imputation - when conducting cluster analyses. Surprisingly, there has been scant research done, or guidance supplied, in conducting cluster analyses using multiply imputed data even though the failure to address missing data is a common critique of COPD phenotyping studies (Bourbeau et al., 2014; Pinto et al., 2015).

As noted in the implications section above, there is a clear need to determine the reasons *why* there are differences in smoking status and behavior among persons diagnosed with COPD by phenotype. While the results presented here offer a unique insight into the relationship between COPD phenotype and smoking status, uncovering

the reasons behind the relationship could prove even more valuable in informing much-needed smoking cessations in this high-risk group. Uncovering these reasons will likely require qualitative evaluations of smokers with COPD in an effort to identify common themes, attitudes, beliefs, motivations and other behavioral precursors to smoking that exist in this population. Using small-group qualitative methods, like focus groups, affords researchers the opportunity to ask probing questions of the smokers with COPD followed by open-ended discussions which could get participants to open up about their smoking-related thoughts and beliefs in ways that could guide future research directions. A portion of this very dissertation, for example, sought to probe the relationship between disease severity and continued smoking in part on the premise that some smokers with COPD though that it was too late for them to quit, a theme derived from interviews conducted on smokers with severe COPD (J. S. Wilson et al., 2011).

A follow-up strategy to uncover the reasons behind continued smoking in this group might be to incorporate follow-up questions to participants diagnosed with COPD in large, population-based surveys like NHANES. In 2007-2012, along with the spirometry examination component, NHANES conducted a short respiratory questionnaire. Should NHANES or another population-based study conduct a respiratory component in the future, one worthwhile approach would be to incorporate a smoking-specific module into their respiratory questionnaire. Using the findings from small-group qualitative studies, in combination with guidance from health behavior theories, questions could be asked in the smoking module that could tease out the reasons, at a population

level and not simply anecdotally, why so many of those diagnosed with COPD continue to smoke. Equally as important, it may be possible to discover important motivators of smoking cessation among those with COPD by posing the same, or similar, questions that were asked of the smokers with COPD to those participants with COPD who have quit smoking. Ultimately, the goal of both the small-group and population-based strategies would be to develop an understanding of the reasons why people with such a progressively debilitating disease would persist with a behavior that not only caused the disease but poses ongoing and additional health threats. A better understanding of the reasons why many with COPD continue to smoke, and possible differences in reasons for distinct subgroups or phenotypes, is critical to informing the most tailored and effective smoking cessation interventions for this high-risk population.

APPENDICES

Appendix I: Methods

A1.1 Methods Overview

All aims for this study (1 through 4) were accomplished via secondary analyses of questionnaire, examination, and laboratory data from the 2007-2012 National Health and Nutrition Examination Survey (NHANES).

Aims 1 and 2 were accomplished by producing COPD diagnosis estimates, smoking status proportions, and smoking prevalence estimates, overall and by disease severity, that were weighted to be nationally representative of the civilian, non-institutionalized U.S. population aged 40-79 years and by producing variance estimates that account for the complex multi-stage survey design of the NHANES.

Aim 3 was accomplished by using multivariable logistic regression analyses, which included moderation analyses, to first determine demographic, clinical, and behavioral factors associated with current smoking among persons with diagnosed COPD, then to assess whether relationships between these factors and current smoking were moderated by disease severity.

Aim 4 was accomplished by cluster analyses to create subgroupings, or phenotypes, of COPD using variables derived both from previous phenotyping studies in addition to factors known to be associated with smoking from the COPD and smoking literature. The resulting phenotypes were then included in multivariable logistic

regression models, along with previously determined factors associated with current smoking status, to determine whether disease phenotypes were differentially associated with current smoking status among persons with diagnosed COPD, controlling for other factors.

A1.2 About NHANES

NHANES is a study conducted by the National Center for Health Statistics (NCHS) and was specifically designed to continually monitor the health and nutritional status of the US population. In 1999, NHANES became continuous, collecting data annually, with a focus on a variety of health indicators chosen to meet changing needs. Since 1999, the survey has examined a nationally representative sample of about 5,000 persons each year with participants located in approximately 15 different counties across the country. NHANES data are prepared and released in two-year release cycles.

The survey is unique in that it combines questionnaire, laboratory and physical examination components. The NHANES questionnaires cover a number of demographic, socioeconomic, dietary, and health-related topic areas. The physical examination components include medical, dental, and physiological measurements. Additionally, laboratory tests are administered in the survey by highly trained medical personnel. Findings from the survey are used to determine the prevalence of major diseases and risk factors for diseases in the United States. Findings from the NHANES form the basis for national standards on a number of measurements including height, weight, and blood pressure.

In collaboration with the National Heart, Lung, and Blood Institute (NHLBI) of the National Institute of Health, and the National Institute for Occupational Safety and Health (NIOSH), NHANES implemented spirometry testing in the study from 2007 through 2012 to assess pulmonary lung function on eligible participants aged 6 to 79 years. Exhaled Nitric Oxide (ENO), a marker of airway inflammation, was also measured. A detailed respiratory questionnaire, including respiratory symptoms, was an integral part of the spirometric examination. The final two years of data for these respiratory components were recently released to the public (2014) and the combined six-year (2007-2012) NHANES spirometry data present a unique opportunity to examine U.S. adults with lung obstruction at different severity levels. In addition to spirometry examination results, NHANES also collected medical conditions details and prescription drug data during this time period along with a wide variety of health-related predictors for smoking and smoking cessation. Additionally, NHANES has serum cotinine data which will supplement the self-reported smoking data to identify current smokers based on recent tobacco exposure. This breadth of bio-behavioral data from NHANES, when combined with the six years of spirometry data, provided a unique opportunity to phenotype persons with COPD in the U.S. using a large, nationally representative dataset.

There are multiple questionnaire, laboratory and examination components in the survey and each component may be targeted at a different age group, or subsample. As the primary focus of this research is on lung obstruction, the following sections will detail

the respiratory health and, more specifically, the spirometry examination component of NHANES.

A1.3 Study Measures

COPD Diagnosis Data

During the in-home interview participants were asked detailed questions on specific medical conditions A participant was classified as having diagnosed COPD if they responded affirmatively to either of the following questions on chronic bronchitis or emphysema: "Has a doctor ever told you that you have chronic bronchitis?" or "Has a doctor ever told you that you have emphysema?"

Spirometry Data and Disease Severity

All NHANES participants aged 6 to 79 years were eligible for the baseline spirometry component of the NHANES. Exclusion criteria for the exam included current chest pain or a physical problem with forceful expiration; taking daily supplemental oxygen; surgery of the eye, chest or the abdomen within last 3 months; stroke or heart attack within last 3 months; tuberculosis exposure; or coughing up blood in last month. Additionally, adults with a personal history of a detached retina or a collapsed lung and children (aged 6 to 15 years) with painful ear infections were also excluded.

The spirometry component was conducted in the NHANES mobile examination centers by trained health technicians. Spirometry testing was performed in accordance with recommendations of the ATS/ERS (M. R. Miller et al., 2005) using Ohio 822/827 dry-rolling seal volume spirometers with in-line biological filters (A-M Systems PFT

Filter Kit B) Spirometry was almost always performed in the standing position unless the participant had a physical limitation. Participants were instructed to perform a series of maximal forced expiratory maneuvers using standard technique until they achieved a reproducible spirogram, or until a maximum of eight spirometry curves had been obtained, or until the participant could not continue. The overall goal was for the participant to achieve three acceptable exhalation maneuvers by ATS criteria in which the two highest values for the FVC and the FEV1 (each taken from an acceptable forced expiratory maneuver) showed minimal variability: i.e. the two largest FVC values taken from 2 acceptable curves agreed within 150 ml, and similarly for the two largest values for the FEV1 (M. R. Miller et al., 2005).

Based on lung function values obtained from the pre-bronchodilator spirometry examination, lung obstruction was defined as a forced expiratory volume in the first second (FEV1) to forced vital capacity (FVC) ratio less than the lower limit of normal (11). FEV1% predicted was defined as the observed FEV1 value divided by the predicted FEV1 value estimated for a person of the same age, gender, race/ethnicity, and height using race-specific reference equations for the US population. The US population predicted values are available online and the method used to derive them described in the literature (Hankinson, Odencrantz, & Fedan, 1999). Reference values are available for non-Hispanic white, non-Hispanic black, and Mexican Americans. For the "other" race category, a correction factor of 0.88 will be applied to the corresponding values for non-Hispanic whites, which have been previously published as an adjustment factor for Asian

participants7. For the “other Hispanic” group, the predicted values for Mexican Americans will be applied.

Participants were categorized into the following disease severity stages: Mild lung obstruction - FEV1>70% predicted among those with lung obstruction.; Moderate or worse lung obstruction - FEV1<=70% predicted among those with lung obstruction (Paulose-Ram et al., 2015; Pellegrino et al., 2005).

Smoking Data

In the NHANES mobile examination center, venipuncture was performed on participants. Blood specimens were subsequently analyzed by the CDC’s National Center for Environmental Health to determine serum cotinine values, a nicotine metabolite with an average half of 16 to 19 hours that can be used to assess recent tobacco smoke exposure (Jarvis, Russell, Benowitz, & Feyerabend, 1988).

For this analysis, both the measured data and self-reported smoking data collected during the in-person household interviews were used to define the following smoking status categories: Current smoker - reported smoking at least 100 cigarettes in their lifetime and at the time of the household interview reported smoking every day or some days OR was defined by a measured serum cotinine value greater than 10 ng/mL, a value clinically determined to indicate active smoking (Hukkanen et al., 2005). Former smoker - reported smoking at least 100 cigarettes during their lifetime but currently did not smoke. Never smoker - reported smoking fewer than 100 cigarettes during their lifetime.

Persons who reported using only other forms of tobacco were not included in this definition.

A1.4 Sample Size Considerations

While the included ages for the NHANES spirometry data collection ranged from 6 to 79 years, the age range of the participants in the analytic sample used for this research is 40 to 79 years old. The lower limit of the sample age range, 40 years old, was chosen because COPD is a disease of older ages. It has a very low prevalence for those under the age of 40 and climbs steeply with age thereafter (GOLD, 2018). The upper limit of the sample age range was limited by the NHANES spirometry data collection protocol, which excluded participants above the age of 79 from the exam. For the period of 2007–2012, 14,825 persons aged 40–79 years were eligible for the survey, 10,531 (71%) were interviewed, 10,219 (69%) were examined in the NHANES mobile examination center (MEC) and 566 (5.5%) reported having healthcare-provider diagnosed COPD.

A1.5 Complex, Multi-stage Probability Sampling and Variance Estimation

The NHANES sample is designed to be nationally representative of the civilian, non-institutionalized U.S. population but NHANES study participants are not obtained using a simple random sample. Rather, a complex, multistage, probability sampling design is used to select participants. The NHANES sampling procedure consists of the four stages described below (Zipf et al., 2013):

Stage 1 - Selection of primary sampling units (PSUs), which are counties or small groups of contiguous counties.

Stage 2 - Selection of segments within PSUs that constitute a block or group of blocks containing a cluster of households.

Stage 3 - Selection of specific households within segments.

Stage 4 - Selection of individuals within a household.

Sample weights for NHANES participants incorporate adjustments for unequal selection probabilities and certain types of non-response, as well as an adjustment to independent estimates of population sizes for specific age, sex, and race/ethnicity categories.

In order to calculate the correct national estimates from the NHANES data, it is critical to use the sample weight of the smallest analytic subpopulation, construct sample weights for the combined analytic survey cycles, and create subsets of the data to reflect the subpopulation of interest before using the sample weights in the analyses.

Similar to using the proper statistical weights to obtain correct parameter estimates, it is also necessary to incorporate the attributes of the complex sample design to obtain unbiased standard error and variance estimates, which, in turn, affect test statistics and confidence intervals. In a complex sample survey setting such as NHANES, variance estimates computed using standard statistical software packages that assume simple random sampling are generally too low (i.e., significance levels are overstated) and biased because they do not account for the differential weighting and the correlation

among sample persons within a cluster. Software such as SUDAAN, Stata, R or SAS Survey Procedures that account for the sampling design effect must be, and were, used to calculate an asymptotically unbiased estimate of the variance and should be used for all statistical tests and the construction of confidence intervals.

A1.6 Specific Aims and Associated Analytic Approaches

A1.6.1 Aim 1: Determine the proportion of adults who have been diagnosed with COPD, smoking status proportions among those diagnosed, and factors predicting receipt of a COPD diagnosis.

RQ1: What percent of the US adult population aged 40-79 years has been diagnosed with COPD?

Research question 1 was answered by first determining those who were diagnosed with COPD. A positive diagnosis of COPD was defined as a positive response to physician or healthcare worker diagnosis of either chronic bronchitis or emphysema (Putcha, Puhan, Hansel, Drummond, & Boyd, 2013; Schnell et al., 2012). Prevalence estimates were computed using appropriate weights and study design parameters such that estimates are nationally representative and with asymptotically unbiased standard errors.

RQ2: Among adults aged 40-79 years with diagnosed COPD, what percent are current smokers, former smokers and never smokers?

The first step to answering research question 2 was to determine smoking status. Current smokers were defined as those reporting smoking at least 100 cigarettes in their lifetime and at the time of the household interview reporting smoking every day or some days or by a measured serum cotinine value greater than 10ng/mL. Former smokers were defined as those reporting smoking at least 100 cigarettes in their lifetime but at the time of the household interview no longer smoking along with having a measured serum cotinine value less than 10ng/mL. Never smokers were defined as those not reporting not having smoked at least 100 cigarettes in their lifetime. Percentages of current, former and never smokers were then computed and presented for all adults diagnosed with COPD. Percentages were computed using appropriate weights and study design parameters to ensure that estimates were nationally representative and with asymptotically unbiased standard errors.

A1.6.2 Aim 2: Compare the biologic, behavioral, and psychosocial characteristics of adults reporting a diagnosis of COPD to those of adults not reporting a COPD diagnosis

RQ3: How do the characteristics of adults reporting a COPD diagnosis differ from the characteristics of adults not reporting a COPD diagnosis?

For research question 3, percentages for each covariate/measure were presented separately for all those reporting a diagnosis of COPD and those not reporting a COPD diagnosis. Chi-square tests were conducted to assess for differences in covariates by

diagnosis status. Appropriate weights and study design parameters were employed to obtain estimates that were nationally representative and with asymptotically unbiased standard errors.

A1.6.3 Aim 3: Assess whether smoking status differs by lung obstruction severity, determine the factors associated with current smoking among ever smokers with COPD and assess the moderating effect of disease severity on these predictive factors

RQ4: Among adults diagnosed with COPD, does smoking status differ by lung obstruction severity?

The first step in answering research question 4 was to assess lung obstruction severity. For this, mild lung obstruction was defined as a measured FEV1 value > 70% predicted among those with diagnosed COPD. Moderate or worse lung obstruction was defined as an FEV1 value <= 70% predicted among those with diagnosed COPD (Paulose-Ram et al., 2015; Pellegrino et al., 2005). Using these determinations of lung obstruction severity along with the previous determinations of smoking status, chi-square tests were then conducted to assess for differences in smoking status by lung obstruction severity.

RQ5: Among ever smokers with diagnosed COPD, what factors are associated with current smoking status?

To answer research question 5, multivariable logistic regression analyses were conducted with the binary outcome of current smoking (yes = current smoker, no = former smoker) to determine significant factors associated with smoking among those

with diagnosed COPD. Estimates were computed using appropriate weights and study design parameters so that estimates were nationally representative and with asymptotically unbiased tests of significance for predictors.

RQ6: Does lung obstruction severity moderate the relationships between the factors associated with smoking status and being a current smoker?

Research question 6 was addressed using moderation analyses to determine whether lung obstruction severity moderates the relationship between the significant associated factors discovered in research question 5 and the binary outcome of yes = current smoker or no = former smoker. To accomplish this, interaction terms for each significant factor and lung obstruction severity were created and multivariable logistic regressions were conducted for each one individually including these corresponding interaction terms.

A1.6.4 Aim 4: Determine phenotypes of COPD, compare/contrast findings to phenotypes found in prior studies, and assess whether one or more phenotypes are differentially associated with current smoking status among ever smokers with diagnosed COPD

RQ7: What are the predominant phenotypes of COPD?

A number of steps were taken to answer research question 7. First, feature selection for the cluster analyses considered those measures collected in the NHANES that were shown to be relevant in prior COPD phenotyping studies as well as a number of variables often associated with smoking, our outcome of interest. The chosen variables

were then transformed (i.e., normalized) to avoid the overpowering of variables with larger scales on the classification into subgroups. Mahalanobis distances were then computed for each observation to detect and remove outliers or extreme observations. Next, the *k*-means method of clustering was chosen to create the subgroups, or phenotypes, of COPD as this was the preferred method in five of the eight studies detailed in an earlier systematic analysis of COPD phenotyping (Pinto et al., 2015). Finally, the resulting clusters or phenotypes were characterized and assessed for interpretability in light of theory and previous published research.

RQ8: Are the identified phenotypes consistent with, or different from, those previously found in other studies?

Research question 8 could only be answered after a comprehensive search of the existing phenotyping literature. Using the findings from this comprehensive literature search, results from the current study were compared and contrasted to the results presented in other studies.

RQ9: Above and beyond other factors, are disease phenotypes predictive of smoking status among those with diagnosed COPD?

The method used to answer research question 8 was built upon the results found in research questions 5, 6 and 7. That is, a final multivariable logistic regression model was conducted again with the dependent variable of current smoking (yes = current smoker, no = former smoker). The independent variables included in this model were the significant factors determined from research questions 5 and 6 along with the categorical

variable of phenotype determined in the cluster analyses in research question 7. Results from this model indicated that lung obstruction phenotypes are indeed predictive of current smoking above and beyond (controlling for) other factors.

A1.7 Key Variables

Table 1.5 lists the NHANES measures that were explored in this study along with their data type and derivation. All were collected by NHANES during the period 2007-2012:

Table 1.5 Measure, data type and derivation for analysis

Lung Obstruction

Measure	Type	NHANES Variable or Derivation
Diagnosis (self-reported) of COPD	Binary (yes/no)	Positive response to physician or health professional ever told you that you had emphysema (mcq160g) or physician or health professional ever told you that you had chronic bronchitis (mcq160k)
Lung obstruction	Binary (yes/no)	Defined by the ATS/ERS standard of the ratio of Forced Expiratory Volume in 1 second (FEV1) to Forced Vital Capacity (FVC) being less than the lowest five percent of a normal, non-diseased population or lower limit of normal (LLN)
Lung obstruction severity	Categorical (none/mild/moderate or more severe)	Defined by the following ATS/ERS standards: • None - no lung obstruction detected • Mild - FEV1 > 70% predicted

		• Moderate or more severe - FEV1 <= 70% predicted
Length of time since diagnosis	Continuous (years)	Defined as the longer of current age (ridageyr) minus age at emphysema diagnosis (mcq180g) or current age minus age at chronic bronchitis diagnosis (mcq180k)
Ever told had asthma	Binary (yes/no)	mcq010

Smoking

Measure	Type	NHANES Variable or Derivation
Smoking status (Current, Former)	Binary (current/former)	Current smoker is defined as having smoked at least 100 cigarettes (smq020) and currently smoking cigarettes (smq040) or by a measured serum cotinine value greater than 10 ng/mL. Former smoker is defined as having smoked at least 100 cigarettes (smq020) and not currently smoking cigarettes (smq040).
Smoking pack years	Continuous (pack-years)	Packs of cigarettes smoked per day is calculated as either the average # of cigarettes smoked per day in the last 30 days (smd650), divided by 20, for current smokers or the # of cigarettes smoked per day when quit (smd057), divided by 20, for former smokers. Number of years smoked is calculated using current age (ridageyr) minus age when started smoking regularly (smq050q) for current smokers and age last smoked cigarettes regularly (smd055) minus age when started smoking regularly (smq050q) for former smokers. Smoking pack years is then calculated by multiplying

		the number of packs of cigarettes smoked per day by the number of years the person has smoked.
Time since quit smoking cigarettes	Continuous (years)	Time since quit smoking cigarettes is calculated using current age (ridageyr) minus age last smoked cigarettes regularly (smd055) for former smokers.
Time to first cigarette in the morning	Categorical	smq077
Nicotine dependency score (modified Fagerstrom)	Continuous	Nicotine dependency score is computed using the two most heavily weighted questions (how soon after you wake up do you smoke your first cigarette? (smq077) and how many cigarettes per day do you smoke? (smd650)) in the 6-question Fagerstrom Test for Nicotine Dependence (FTND). These two variables account for 6 of the 12 total points in the original FTND.
Number of smokers inside home	Categorical	smd415

Health/Medical

Measure	Type	NHANES Variable or Derivation
Number of times received healthcare in the last year	Categorical	huq050
Overnight hospital patient in the last year	Binary (yes/no)	huq071
Coronary heart disease	Binary (yes/no)	mcq160c
Stroke	Binary (yes/no)	mcq160f
Cancer	Binary (yes/no)	mcq220
Diabetes	Binary (yes/no)	diq010
Arthritis	Binary (yes/no)	mcq160a
Hypertension	Binary (yes/no)	bpq020
Chronic cough	Binary (yes/no)	Defined as coughing most days over a three month period (rdq031) for two years or more (rdd040)
Chronic phlegm	Binary (yes/no)	Defined as bringing up phlegm most days over a three month period (rdq050) for two years or more (rdd060)
Wheezing episodes past year[1,4]	Categorical	rdq080
Dyspnea	Binary (yes/no)	Dyspnea is assessed as a positive response to the question, "Have you had shortness of breath when hurrying up the level or walking up a slight hill?" (cdq010)
Take respiratory medication	Binary (yes/no)	Positive report to any of the following classes of medications: • Long-acting bronchodilators • Short-acting bronchodilators

		• Inhaled corticosteroids • Synthetic corticosteroids • Methylxanthines • Leukotriene modifiers
Exhaled nitric oxide (ENO) level	Categorical	enxmean
Eosinophilia	Binary (yes/no)	Eosinophilia is determined by having 350 or more eosinophil cells per microliter of blood (lbdeono)

Demographic

Measure	Type	NHANES Variable or Derivation
Age	Binary (40-59/60-79)	Categorized using age (ridageyr) as 40-59 or 60-79 years old
Race/Ethnicity	Categorical	Combined Mexican and Other Hispanic groups from race/ethnicity (ridreth1) to form: • Non-Hispanic white • Non-Hispanic black • Hispanic • Other non-Hispanic
Gender	Binary (male/female)	riagendr
Family Income	Categorical	indfmin2
Home owned or rented	Binary (owned/rented)	hoq065
Have health insurance	Binary (yes/no)	hiq011
Education	Categorical	Categorized using education (dmdeduc) as: < High School HS grad/GED or equivalent Any college
Language of interview	Binary (English/Spanish)	sialang

Physical and Nutritional

Measure	Type	NHANES Variable or Derivation
Physical activity – minutes sedentary activity on a typical day	Continuous	pad680
Physical activity – minutes moderate or more vigorous activity on a typical day	Continuous	Moderate or more vigorous activity will be calculated as the sum of the following NHANES variables: • pad615 - Minutes vigorous-intensity work • pad630 - Minutes moderate-intensity work • pad645 - Minutes walk/bicycle for transportation • pad660 - Minutes vigorous recreational activities • pad675 - Minutes moderate recreational activities
Body mass index (BMI)	Continuous	bmxbmi
Weight status (Perceived)	Categorical	whq030
Diet behavior (Perceived)	Categorical	dbq700
Impaired Activities of Daily Living (ADL)	Binary (yes/no)	Positive report to having any difficulty in one or more of the following daily household activities: • Walking between rooms on the same floor • Standing up from armless chair • Getting in and out of bed • Using a fork, a knife or drinking from a cup • Getting dressed

Psychosocial

Measure	Type	NHANES Variable or Derivation
Alcohol use (days drank alcohol in past year)	Categorical	alq120
Alcohol use (# days having 5+ drinks in past year)	Categorical	alq140q
Employment status	Binary (Working/Not)	ocd150
Trouble sleeping	Binary (yes/no)	slq050
Depression (PHQ9 depression score)	Continuous	The full PHQ9 depression diagnostic instrument was collected in the NHANES from 2007-2012. The score will be computed from responses to the following nine questions: 1. dpq010 - Little interest in doing things 2. dpq020 - Feeling down, depressed, or hopeless 3. dpq030 - Trouble sleeping or sleeping too much 4. dpq040 - Feeling tired or having little energy 5. dpq050 - Poor appetite or overeating 6. dpq060 - Feeling bad about yourself 7. dpq070 - Trouble concentrating on things 8. dpq080 - Moving or speaking slowly or too fast 9. dpq090 - Thought you would be better off dead
Seen mental health professional in past year	Binary (yes/no)	huq090

Number of days (in last 30) mental health was not good	Continuous	hsq480
General health condition (perceived)	Categorical	hsd010
Health now compared with one year ago (perceived)	Categorical	huq020

Appendix II: Human Subjects Protection

All of the research conducted as part of this dissertation employed publicly available, de-identified data. No interaction with human subjects was undertaken at any point. As a result, all of the research conducted herein qualifies for exempt status at the federal level in accordance with the U.S. Department of Health and Human Services' Code of Federal Regulations on Human Subjects Research (Title 45, Part 46). Similarly, at the institutional level, this research was deemed not subject to review by the University of Maryland's Institutional Review Board (IRB) as it did not meet the criteria for human subjects research.

IRB approval was originally obtained prior to data collection for the 2007- 2012 NHANES by the United States Centers for Disease Control and Prevention, National Center for Health Statistics as well as by Westat, the contract agency administering the survey.

REFERENCES

Adams, S. G., Pugh, J. A., Kazis, L. E., Lee, S., & Anzueto, A. (2006). Characteristics Associated with Sustained Abstinence from Smoking Among Patients with COPD. *The American Journal of Medicine*, *119*(5), 441–447. https://doi.org/10.1016/j.amjmed.2005.09.055

Adis Medical Writers. (2015). Phenotyping provides potential for a personalized approach in patients with chronic obstructive pulmonary disease (COPD). *Drugs & Therapy Perspectives*, *31*(3), 89–92. https://doi.org/10.1007/s40267-015-0183-3

Agusti, A., Calverley, P. M. A., Celli, B., Coxson, H. O., Edwards, L. D., Lomas, D. A., … Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints (ECLIPSE) investigators. (2010). Characterisation of COPD heterogeneity in the ECLIPSE cohort. *Respiratory Research*, *11*, 122. https://doi.org/10.1186/1465-9921-11-122

Alsaeedi, A., Sin, D. D., & McAlister, F. A. (2002). The effects of inhaled corticosteroids in chronic obstructive pulmonary disease: A systematic review of randomized placebo-controlled trials. *The American Journal of Medicine*, *113*(1), 59–65.

American Thoracic Society. (1995). Standards for the diagnosis and care of patients with chronic obstructive pulmonary disease. American Thoracic Society. *American Journal of Respiratory and Critical Care Medicine*, *152*(5 Pt 2), S77-121.

Anthonisen, N. R., Connett, J. E., Kiley, J. P., Altose, M. D., Bailey, W. C., Buist, A. S., … O'Hara, P. (1994). Effects of smoking intervention and the use of an inhaled anticholinergic bronchodilator on the rate of decline of FEV1. The Lung Health Study. *JAMA*, *272*(19), 1497–1505.

Anthonisen, N. R., Connett, J. E., & Murray, R. P. (2002). Smoking and Lung Function of Lung Health Study Participants after 11 Years. *American Journal of Respiratory and Critical Care Medicine*, *166*(5), 675–679. https://doi.org/10.1164/rccm.2112096

Au, D. H., Bryson, C. L., Chien, J. W., Sun, H., Udris, E. M., Evans, L. E., & Bradley, K. A. (2009). The Effects of Smoking Cessation on the Risk of Chronic Obstructive Pulmonary Disease Exacerbations. *Journal of General Internal Medicine*, *24*(4), 457–463. https://doi.org/10.1007/s11606-009-0907-y

Balmes, J., Becklake, M., & Blanc, P. (2003). Occupational contribution to the burden of airway disease (an official statement of the American Thoracic Society). *American Journal of Respiratory and Critical Care Medicine*, *167*, 787–797.

Barker, B. L., & Brightling, C. E. (2013). Phenotyping the heterogeneity of chronic obstructive pulmonary disease. *Clinical Science*, *124*(6), 371–387. https://doi.org/10.1042/CS20120340

Basagaña, X., Barrera-Gómez, J., Benet, M., Antó, J. M., & Garcia-Aymerich, J. (2013). A Framework for Multiple Imputation in Cluster Analysis. *American Journal of Epidemiology*, *177*(7), 718–725. https://doi.org/10.1093/aje/kws289

Blanc, P. D. (2012). Occupation and COPD: A Brief Review. *Journal of Asthma*, *49*(1), 2–4. https://doi.org/10.3109/02770903.2011.611957

Bolliger, C. T., Zellweger, J.-P., Danielsson, T., van Biljon, X., Robidou, A., Westin, Å., … Säwe, U. (2002). Influence of long-term smoking reduction on health risk markers and quality of life. *Nicotine & Tobacco Research*, *4*(4), 433–439. https://doi.org/10.1080/1462220021000018380

Bon, J., Liao, S., Tseng, G., & Sciurba, F. C. (2013). Considerations and pitfalls in phenotyping and reclassification of chronic obstructive pulmonary disease. *Translational Research*, *162*(4), 252–257. https://doi.org/10.1016/j.trsl.2013.07.006

Bourbeau, J., Pinto, L. M., & Benedetti, A. (2014). Phenotyping of COPD: Challenges and next steps. *The Lancet Respiratory Medicine*, *2*(3), 172–174. https://doi.org/10.1016/S2213-2600(14)70039-6

Browning, K. K., Wewers, M. E., Ferketich, A. K., Otterson, G. A., & Reynolds, N. R. (2009). The Self-regulation Model of Illness applied to smoking behavior in lung cancer. *Cancer Nursing*, *32*(4), E15-25. https://doi.org/10.1097/NCC.0b013e3181a0238f

Burchfiel, C. M., Marcus, E. B., Curb, J. D., Maclean, C. J., Vollmer, W. M., Johnson, L. R., … Buist, A. S. (1995). Effects of smoking and smoking cessation on longitudinal decline in pulmonary function. *American Journal of Respiratory and*

Critical Care Medicine, *151*(6), 1778–1785. https://doi.org/10.1164/ajrccm.151.6.7767520

Burgel, P.-R., Paillasseur, J.-L., Caillaud, D., Tillie-Leblond, I., Chanez, P., Escamilla, R., … Initiatives BPCO Scientific Committee. (2010). Clinical COPD phenotypes: A novel approach using principal component and cluster analyses. *The European Respiratory Journal*, *36*(3), 531–539. https://doi.org/10.1183/09031936.00175109

Buunk, B. P., & Gibbons, F. X. (2013). *Health, Coping, and Well-being: Perspectives From Social Comparison Theory*. https://doi.org/10.4324/9780203774106

Calabrese, E. J., Bachmann, K. A., Bailer, A. J., Bolger, P. M., Borak, J., Cai, L., … Mattson, M. P. (2007). Biological stress response terminology: Integrating the concepts of adaptive response and preconditioning stress within a hormetic dose-response framework. *Toxicology and Applied Pharmacology*, *222*(1), 122–128. https://doi.org/10.1016/j.taap.2007.02.015

Calverley, P. M., Rabe, K. F., Goehring, U.-M., Kristiansen, S., Fabbri, L. M., & Martinez, F. J. (2009). Roflumilast in symptomatic chronic obstructive pulmonary disease: Two randomised clinical trials. *The Lancet*, *374*(9691), 685–694. https://doi.org/10.1016/S0140-6736(09)61255-1

Celli, B. R., MacNee, W., Agusti, A., Anzueto, A., Berg, B., Buist, A. S., … ZuWallack, R. (2004). Standards for the diagnosis and treatment of patients with COPD: A

summary of the ATS/ERS position paper. *European Respiratory Journal*, *23*(6), 932–946. https://doi.org/10.1183/09031936.04.00014304

Cerveri, I., Corsico, A. G., Accordini, S., Niniano, R., Ansaldo, E., Antó, J. M., … de Marco, R. (2008). Underestimation of airflow obstruction among young adults using FEV1/FVC <70% as a fixed cut-off: A longitudinal evaluation of clinical and functional outcomes. *Thorax*, *63*(12), 1040–1045. https://doi.org/10.1136/thx.2008.095554

Chatila, W. M., Wynkoop, W. A., Vance, G., & Criner, G. J. (2004). Smoking patterns in African Americans and whites with advanced COPD. *Chest*, *125*(1), 15–21. https://doi.org/10.1378/chest.125.1.15

Cho, M. H., Washko, G. R., Hoffmann, T. J., Criner, G. J., Hoffman, E. A., Martinez, F. J., … Silverman, E. K. (2010). Cluster analysis in severe emphysema subjects using phenotype and genotype data: An exploratory investigation. *Respiratory Research*, *11*(1), 30. https://doi.org/10.1186/1465-9921-11-30

Christenhusz, L., Pieterse, M., Seydel, E., & van der Palen, J. (2007). Prospective determinants of smoking cessation in COPD patients within a high intensity or a brief counseling intervention. *Patient Education and Counseling*, *66*(2), 162–166. https://doi.org/10.1016/j.pec.2006.11.006

Cokkinides, V. E., Halpern, M. T., Barbeau, E. M., Ward, E., & Thun, M. J. (2008). Racial and ethnic disparities in smoking-cessation interventions: Analysis of the

2005 National Health Interview Survey. *American Journal of Preventive Medicine, 34*(5), 404–412. https://doi.org/10.1016/j.amepre.2008.02.003

COPD Foundation. (2019). What Causes COPD? Retrieved September 8, 2019, from https://www.copdfoundation.org/What-is-COPD/Understanding-COPD/What-Causes-COPD.aspx

Coronini-Cronberg, S., Heffernan, C., & Robinson, M. (2011). Effective smoking cessation interventions for COPD patients: A review of the evidence. *JRSM Short Reports*, *2*(10), 1–12. https://doi.org/10.1258/shorts.2011.011089

Crowley, T. J., Macdonald, M. J., & Walter, M. I. (1995). Behavioral anti-smoking trial in chronic obstructive pulmonary disease patients. *Psychopharmacology, 119*(2), 193–204.

Cunningham, T. J., Ford, E. S., Rolle, I. V., Wheaton, A. G., & Croft, J. B. (2015). Associations of self-reported cigarette smoking with chronic obstructive pulmonary disease and co-morbid chronic conditions in the United States. *COPD: Journal of Chronic Obstructive Pulmonary Disease, 12*(3), 281–291.

Danielsen, S. E., Løchen, M.-L., Medbø, A., Vold, M. L., & Melbye, H. (2016). A new diagnosis of asthma or COPD is linked to smoking cessation—The Tromsø study. *International Journal of Chronic Obstructive Pulmonary Disease*, *11*, 1453–1458. https://doi.org/10.2147/COPD.S108046

Davila, E. P., Zhao, W., Byrne, M., Webb, M., Huang, Y., Arheart, K., … Lee, D. J. (2009). Correlates of smoking quit attempts: Florida Tobacco Callback Survey, 2007. *Tobacco Induced Diseases*, *5*, 10. https://doi.org/10.1186/1617-9625-5-10

de Granda-Orive, J. I., & Martínez-Albiach, J. M. (2005). Smoking cessation in patients with chronic obstructive pulmonary disease. *Archivos De Bronconeumologia*, *41*(11), 625–633.

Decramer, M., & Janssens, W. (2013). Chronic obstructive pulmonary disease and comorbidities. *The Lancet. Respiratory Medicine*, *1*(1), 73–83. https://doi.org/10.1016/S2213-2600(12)70060-7

Diefenbach, M. A., & Leventhal, H. (1996). The Common-Sense Model of Illness Representation: Theoretical and Practical Considerations. *Journal of Social Distress and the Homeless*, *5*(1), 11–38. https://doi.org/10.1007/BF02090456

Disantostefano, R. L., Li, H., Rubin, D. B., & Stempel, D. A. (2013). Which patients with chronic obstructive pulmonary disease benefit from the addition of an inhaled corticosteroid to their bronchodilator? A cluster analysis. *BMJ Open*, *3*(4). https://doi.org/10.1136/bmjopen-2012-001838

Dransfield, M. T., Davis, J. J., Gerald, L. B., & Bailey, W. C. (2006). Racial and gender differences in susceptibility to tobacco smoke among patients with chronic obstructive pulmonary disease. *Respiratory Medicine*, *100*(6), 1110–1116. https://doi.org/10.1016/j.rmed.2005.09.019

Eisner, M. D., Balmes, J., Katz, P. P., Trupin, L., Yelin, E. H., & Blanc, P. D. (2005). Lifetime environmental tobacco smoke exposure and the risk of chronic obstructive pulmonary disease. *Environmental Health: A Global Access Science Source*, *4*(1), 7. https://doi.org/10.1186/1476-069X-4-7

Fabbri, L. M., Boyd, C., Boschetto, P., Rabe, K. F., Buist, A. S., Yawn, B., … Kent, D. M. (2012). ATS/ERS Ad Hoc Committee on Integrating and Coordinating Efforts in COPD Guideline Development. How to integrate multiple comorbidities in guideline development: Article 10 in Integrating and coordinating efforts in COPD guideline development. An official ATS/ERS workshop report. *Proc Am Thorac Soc*, *9*, 274–281.

Fagerström, K. O., Kunze, M., Schoberberger, R., Breslau, N., Hughes, J. R., Hurt, R. D., … Zatoński, W. (1996). Nicotine dependence versus smoking prevalence: Comparisons among countries and categories of smokers. *Tobacco Control*, *5*(1), 52–56. https://doi.org/10.1136/tc.5.1.52

Fishman, A., Martinez, F., Naunheim, K., Piantadosi, S., Wise, R. A., Ries, A., … Wood, D. E. (2003). A randomized trial comparing lung-volume-reduction surgery with medical therapy for severe emphysema. *New England Journal of Medicine*, *348*(21), 2059–2073. https://doi.org/10.1056/NEJMoa030287

Garcia-Aymerich, J., Gómez, F. P., Benet, M., Farrero, E., Basagaña, X., Gayete, À., … Antó, J. M. (2011). Identification and prospective validation of clinically relevant

chronic obstructive pulmonary disease (COPD) subtypes. *Thorax*, *66*(5), 430–437. https://doi.org/10.1136/thx.2010.154484

Gibson, P. G., & McDonald, V. M. (2014). Why is COPD phenotyping like sorting diamonds? *European Respiratory Journal*, *44*(2), 277–279. https://doi.org/10.1183/09031936.00091214

GOLD. (2018). *Global Strategy for the Diagnosis, Management and Prevention of COPD*. Retrieved from http://goldcopd.org/gold-2017-global-strategy-diagnosis-management-prevention-copd/

Górecka, D., Bednarek, M., Nowiński, A., Puścińska, E., Goljan-Geremek, A., & Zieliński, J. (2003). Diagnosis of airflow limitation combined with smoking cessation advice increases stop-smoking rate. *Chest*, *123*(6), 1916–1923. https://doi.org/10.1378/chest.123.6.1916

Gratziou, C., Florou, A., Ischaki, E., Eleftheriou, K., Sachlas, A., Bersimis, S., & Zakynthinos, S. (2014). Smoking cessation effectiveness in smokers with COPD and asthma under real life conditions. *Respiratory Medicine*, *108*(4), 577–583. https://doi.org/10.1016/j.rmed.2014.01.007

Guarascio, A. J., Ray, S. M., Finch, C. K., & Self, T. H. (2013). The clinical and economic burden of chronic obstructive pulmonary disease in the USA. *ClinicoEconomics and Outcomes Research: CEOR*, *5*, 235–245. https://doi.org/10.2147/CEOR.S34321

Hagger, M. S., & Orbell, S. (2003). A Meta-Analytic Review of the Common-Sense Model of Illness Representations. *Psychology & Health*, *18*(2), 141–184. https://doi.org/10.1080/0887044031000813321

Halldin, C. N., Doney, B. C., & Hnizdo, E. (2015). Changes in prevalence of chronic obstructive pulmonary disease and asthma in the US population and associated risk factors. *Chronic Respiratory Disease*, *12*(1), 47–60.

Han, M. K., Agusti, A., Calverley, P. M., Celli, B. R., Criner, G., Curtis, J. L., … Martinez, F. J. (2010). Chronic Obstructive Pulmonary Disease Phenotypes. *American Journal of Respiratory and Critical Care Medicine*, *182*(5), 598–604. https://doi.org/10.1164/rccm.200912-1843CC

Hancock, G. R., Mueller, R. O., Stapleton, L. M., Mueller, R. O., Mueller, R. O., Stapleton, L. M., & Mueller, R. O. (2010). *The Reviewer's Guide to Quantitative Methods in the Social Sciences*. https://doi.org/10.4324/9780203861554

Hand, J. L. (1981). Sociobiological implications of unusual sexual behaviors of gulls: The genotype/behavioral phenotype problem. *Ethology and Sociobiology*, *2*(3), 135–145. https://doi.org/10.1016/0162-3095(81)90026-1

Hankinson, J. L., Kawut, S. M., Shahar, E., Smith, L. J., Stukovsky, K. H., & Barr, R. G. (2010). Performance of American Thoracic Society-recommended spirometry reference values in a multiethnic sample of adults: The multi-ethnic study of atherosclerosis (MESA) lung study. *Chest*, *137*(1), 138–145.

Hankinson, J. L., Odencrantz, J. R., & Fedan, K. B. (1999). Spirometric Reference Values from a Sample of the General U.S. Population. *American Journal of Respiratory and Critical Care Medicine*, *159*(1), 179–187. https://doi.org/10.1164/ajrccm.159.1.9712108

Hansen, J. E., Sun, X.-G., & Wasserman, K. (2007). Spirometric criteria for airway obstruction: Use percentage of FEV1/FVC ratio below the fifth percentile, not < 70%. *Chest*, *131*(2), 349–355. https://doi.org/10.1378/chest.06-1349

Hardie, J. A., Buist, A. S., Vollmer, W. M., Ellingsen, I., Bakke, P. S., & Mørkve, O. (2002). Risk of over-diagnosis of COPD in asymptomatic elderly never-smokers. *The European Respiratory Journal*, *20*(5), 1117–1122.

Hersh, C. P., DeMeo, D. L., Al-Ansari, E., Carey, V. J., Reilly, J. J., Ginns, L. C., & Silverman, E. K. (2004). Predictors of Survival in Severe, Early Onset COPD. *Chest*, *126*(5), 1443–1451. https://doi.org/10.1378/chest.126.5.1443

Hilberink, S. R., Jacobs, J. E., Bottema, B. J. A. M., de Vries, H., & Grol, R. P. T. M. (2005). Smoking cessation in patients with COPD in daily general practice (SMOCC): Six months' results. *Preventive Medicine*, *41*(5–6), 822–827. https://doi.org/10.1016/j.ypmed.2005.08.003

Hilberink, S. R., Jacobs, J. E., Schlösser, M., Grol, R. P. T. M., & Vries, H. de. (2006). Characteristics of patients with COPD in three motivational stages related to smoking cessation. *Patient Education and Counseling*, *61*(3), 449–457. https://doi.org/10.1016/j.pec.2005.05.012

Hnizdo, E., Sullivan, P. A., Bang, K. M., & Wagner, G. (2002). Association between Chronic Obstructive Pulmonary Disease and Employment by Industry and Occupation in the US Population: A Study of Data from the Third National Health and Nutrition Examination Survey. *American Journal of Epidemiology*, *156*(8), 738–746. https://doi.org/10.1093/aje/kwf105

Hoogendoorn, M., Feenstra, T. L., Hoogenveen, R. T., & Rutten-van Mölken, M. P. M. H. (2010). Long-term effectiveness and cost-effectiveness of smoking cessation interventions in patients with COPD. *Thorax*, *65*(8), 711–718. https://doi.org/10.1136/thx.2009.131631

Hukkanen, J., Jacob, P., & Benowitz, N. L. (2005). Metabolism and disposition kinetics of nicotine. *Pharmacological Reviews*, *57*(1), 79–115.

Hymowitz, N., Cummings, K. M., Hyland, A., Lynn, W. R., Pechacek, T. F., & Hartwell, T. D. (1997). Predictors of smoking cessation in a cohort of adult smokers followed for five years. *Tobacco Control*, *6*(suppl 2), S57. https://doi.org/10.1136/tc.6.suppl_2.S57

Jarvis, M. J., Russell, M. A., Benowitz, N. L., & Feyerabend, C. (1988). Elimination of cotinine from body fluids: Implications for noninvasive measurement of tobacco smoke exposure. *American Journal of Public Health*, *78*(6), 696–698. https://doi.org/10.2105/AJPH.78.6.696

Jiménez Ruiz, C. A., Ramos Pinedo, A., Cicero Guerrero, A., Mayayo Ulibarri, M., Cristobal Fernández, M., & Lopez Gonzalez, G. (2012). Characteristics of COPD

smokers and effectiveness and safety of smoking cessation medications. *Nicotine & Tobacco Research: Official Journal of the Society for Research on Nicotine and Tobacco*, *14*(9), 1035–1039. https://doi.org/10.1093/ntr/nts001

Johnson, H. G., Ekman, P., Friesen, W., Nyhan, W. L., & Shear, C. (1976). A Behavioral Phenotype in the de Lange Syndrome. *Pediatric Research*, *10*(10), 843. https://doi.org/10.1203/00006450-197610000-00006

Johnston, C. A., Moreno, J. P., Regas, K., Tyler, C., & Foreyt, J. P. (2012). The Application of the Yerkes–Dodson Law in a Childhood Weight Management Program: Examining Weight Dissatisfaction. *Journal of Pediatric Psychology*, *37*(6), 674–679. https://doi.org/10.1093/jpepsy/jss040

Jones, P. W. (2005). St. George's Respiratory Questionnaire: MCID. *COPD*, *2*(1), 75–79.

Jones, P. W., Quirk, F. H., Baveystock, C. M., & Littlejohns, P. (1992). A Self-complete Measure of Health Status for Chronic Airflow Limitation: The St. George's Respiratory Questionnaire. *American Review of Respiratory Disease*, *145*(6), 1321–1327. https://doi.org/10.1164/ajrccm/145.6.1321

Kanner, R. E., Connett, J. E., Williams, D. E., & Buist, A. S. (1999). Effects of randomized assignment to a smoking cessation intervention and changes in smoking habits on respiratory symptoms in smokers with early chronic obstructive pulmonary disease: The Lung Health Study. *The American Journal of Medicine*, *106*(4), 410–416.

Kohansal, R., Martinez-Camblor, P., Agustí, A., Buist, A. S., Mannino, D. M., & Soriano, J. B. (2009). The natural history of chronic airflow obstruction revisited: An analysis of the Framingham offspring cohort. *American Journal of Respiratory and Critical Care Medicine*, *180*(1), 3–10. https://doi.org/10.1164/rccm.200901-0047OC

Köhler, D., Fischer, J., Raschke, F., & Schönhofer, B. (2003). Usefulness of GOLD classification of COPD severity. *Thorax*, *58*(9), 825. https://doi.org/10.1136/thorax.58.9.825

Kosacz, N., Punturieri, A., & Croxton, T. (2012). Chronic obstructive pulmonary disease among adults–United States, 2011. *MMWR. Morbidity and Mortality Weekly Report*, *61*(46), 938.

Lamprecht, B., McBurnie, M. A., Vollmer, W. M., Gudmundsson, G., Welte, T., Nizankowska-Mogilnicka, E., … Buist, S. A. (2011). COPD in Never Smokers: Results From the Population-Based Burden of Obstructive Lung Disease Study. *Chest*, *139*(4), 752–763. https://doi.org/10.1378/chest.10-1253

Lamprecht, B., Soriano, J., Studnicka, M., Kaiser, B., Vanfleteren, L., Gnatiuc, L., … Buist, S. (2015). Determinants of Underdiagnosis of COPD in National and International Surveys. *CHEST*, *148*(4), 971–985. https://doi.org/10.1378/chest.14-2535

Lange, L. J., & Piette, J. D. (2006). Personal models for diabetes in context and patients' health status. *Journal of Behavioral Medicine, 29*(3), 239–253. https://doi.org/10.1007/s10865-006-9049-4

Leventhal, H., Meyer, D., & Nerenz, D. (1980). The common sense model of illness danger. In *Medical Psychology* (Vol. 2). New York: Pergamon.

Lindberg, A., Bjerg-Bäcklund, A., Rönmark, E., Larsson, L.-G., & Lundbäck, B. (2006a). Prevalence and underdiagnosis of COPD by disease severity and the attributable fraction of smoking. *Respiratory Medicine*, *100*(2), 264–272.

Lindberg, A., Bjerg-Bäcklund, A., Rönmark, E., Larsson, L.-G., & Lundbäck, B. (2006b). Prevalence and underdiagnosis of COPD by disease severity and the attributable fraction of smoking: Report from the Obstructive Lung Disease in Northern Sweden Studies. *Respiratory Medicine, 100*(2), 264–272. https://doi.org/10.1016/j.rmed.2005.04.029

Liu, S., Zhou, Y., Liu, S., Chen, X., Zou, W., Zhao, D., … Ran, P. (2017). Association between exposure to ambient particulate matter and chronic obstructive pulmonary disease: Results from a cross-sectional study in China. *Thorax, 72*(9), 788–795. https://doi.org/10.1136/thoraxjnl-2016-208910

Løkke, A., Lange, P., Scharling, H., Fabricius, P., & Vestbo, J. (2006). Developing COPD: A 25 year follow up study of the general population. *Thorax, 61*(11), 935–939. https://doi.org/10.1136/thx.2006.062802

Lopez, A.D., Mathers, C.D., Ezzati, M., Jamison, D.T., & Murray, C.J. (2006). *Global Burden of Disease and Risk Factors*. https://doi.org/10.1596/978-0-8213-6262-4

Mahler, D. A., Faryniarz, K., Tomlinson, D., Colice, G. L., Robins, A. G., Olmstead, E. M., & O'Connor, G. T. (1992). Impact of Dyspnea and Physiologic Function on General Health Status in Patients with Chronic Obstructive Pulmonary Disease. *Chest*, *102*(2), 395–401. https://doi.org/10.1378/chest.102.2.395

Mannino, D. M., Ford, E. S., & Redd, S. C. (2003). Obstructive and restrictive lung disease and markers of inflammation: Data from the Third National Health and Nutrition Examination. *The American Journal of Medicine*, *114*(9), 758–762.

Mannino, D. M., Gan, W. O., Wurst, K., & Davis, K. J. (2017). Asthma and Chronic Obstructive Pulmonary Disease Overlap: The Effect of Definitions on Measures of Burden. *Chronic Obstructive Pulmonary Diseases*, *4*(2), 87.

Mannino, D. M., Homa, D. M., Akinbami, L. J., Ford, E. S., & Redd, S. C. (2002). Chronic obstructive pulmonary disease surveillance-United States, 1971-2000. *Respiratory Care*, *47*(10), 1184–1199.

Mathers, C. D., & Loncar, D. (2006). Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Medicine*, *3*(11), e442. https://doi.org/10.1371/journal.pmed.0030442

Melzer, A. C., Feemster, L. C., Crothers, K., Carson, S. S., Gillespie, S. E., Henderson, A. G., … Au, D. H. (2016). Respiratory and Bronchitic Symptoms Predict Intention to Quit Smoking among Current Smokers with, and at Risk for, Chronic

Obstructive Pulmonary Disease. *Annals of the American Thoracic Society*, *13*(9), 1490–1496. https://doi.org/10.1513/AnnalsATS.201601-075OC

Merriam, & Webster. (2019). Definition of Phenotype. Retrieved June 26, 2019, from https://www.merriam-webster.com/dictionary/phenotype

Miller, M. R., Hankinson, J., Brusasco, V., Burgos, F., Casaburi, R., Coates, A., … Wanger, J. (2005). Standardisation of spirometry. *European Respiratory Journal*, *26*(2), 319–338. https://doi.org/10.1183/09031936.05.00034805

Miller, Martin Raymond, Hankinson, J., Brusasco, V., Burgos, F., Casaburi, R., Coates, A., … Gustafsson, P. (2005). Standardisation of spirometry. *European Respiratory Journal*, *26*(2), 319–338.

Milligan, G. W. (1980). An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika*, *45*(3), 325–342. https://doi.org/10.1007/BF02293907

Miravitlles, M., Barrecheguren, M., & Román-Rodríguez, M. (2015). Frequency and characteristics of different clinical phenotypes of chronic obstructive pulmonary disease. *The International Journal of Tuberculosis and Lung Disease : The Official Journal of the International Union against Tuberculosis and Lung Disease*, *19*(8), 992–998. https://doi.org/10.5588/ijtld.15.0021

Miravitlles, Marc, Calle, M., & Soler-Cataluña, J. J. (2012). Clinical phenotypes of COPD: Identification, definition and implications for guidelines. *Archivos De Bronconeumologia*, *48*(3), 86–98. https://doi.org/10.1016/j.arbres.2011.10.007

Molfino, N. A. (2004). Genetics of COPD. *Chest*, *125*(5), 1929–1940. https://doi.org/10.1378/chest.125.5.1929

Monsó, E., Campbell, J., Tønnesen, P., Gustavsson, G., & Morera, J. (2001). Sociodemographic predictors of success in smoking intervention. *Tobacco Control*, *10*(2), 165–169. https://doi.org/10.1136/tc.10.2.165

Murphy, S. L., Xu, J., & Kochanek, K. D. (2013). Deaths: Final data for 2010. *National Vital Statistics Reports: From the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System*, *61*(4), 1–117.

NHANES - Analytic Guidelines, 1999-2010. (2013). Retrieved March 2, 2018, from https://wwwn.cdc.gov/nchs/data/nhanes/2011-2012/manuals/Spirometry_Procedures_Manual.pdf

NHANES - National Health and Nutrition Examination Survey Homepage. (2018, February 24). Retrieved March 2, 2018, from https://www.cdc.gov/nchs/nhanes/index.htm

NHANES - NCHS Research Ethics Review Board Approval. (2017, December 1). Retrieved April 12, 2018, from https://www.cdc.gov/nchs/nhanes/irba98.htm

NHANES - Respiratory Health—Bronchodilator Procedures Manual. (2011, January). Retrieved March 2, 2018, from https://wwwn.cdc.gov/nchs/data/nhanes/2011-2012/manuals/Spirometry_Procedures_Manual.pdf

NHLBI. (2012). *Morbidity & Mortality: 2012 Chart Book on Cardiovascular, Lung, and Blood Diseases*. 117.

NHLBI. (2019). COPD | National Heart, Lung, and Blood Institute (NHLBI). Retrieved September 8, 2019, from https://www.nhlbi.nih.gov/health-topics/copd

O'Donnell, D. E., Aaron, S., Bourbeau, J., Hernandez, P., Marciniuk, D. D., Balter, M., … Voduc, N. (2007). Canadian Thoracic Society recommendations for management of chronic obstructive pulmonary disease—2007 update. *Canadian Respiratory Journal, 14 Suppl B*, 5B-32B. https://doi.org/10.1155/2007/830570

O'Donnell, D. E., Lam, M., & Webb, K. A. (1999). Spirometric correlates of improvement in exercise performance after anticholinergic therapy in chronic obstructive pulmonary disease. *American Journal of Respiratory and Critical Care Medicine, 160*(2), 542–549. https://doi.org/10.1164/ajrccm.160.2.9901038

Orozco-Levi, M., Garcia-Aymerich, J., Villar, J., Ramírez-Sarmiento, A., Antó, J. M., & Gea, J. (2006). Wood smoke exposure and risk of chronic obstructive pulmonary disease. *The European Respiratory Journal*, *27*(3), 542–546. https://doi.org/10.1183/09031936.06.00052705

Paulose-Ram, R., Tilert, T., Dillon, C. F., & Brody, D. J. (2015). Cigarette smoking and lung obstruction among adults aged 40-79: United States, 2007-2012. *NCHS Data Brief*, (181), 1–8.

Pauwels, R. A., Buist, A. S., Calverley, P. M., Jenkins, C. R., Hurd, S. S., & GOLD Scientific Committee. (2001). Global strategy for the diagnosis, management, and

prevention of chronic obstructive pulmonary disease. NHLBI/WHO Global Initiative for Chronic Obstructive Lung Disease (GOLD) Workshop summary. *American Journal of Respiratory and Critical Care Medicine*, *163*(5), 1256–1276. https://doi.org/10.1164/ajrccm.163.5.2101039

Pederson, L. L., Wanklin, J. M., & Lefcoe, N. M. (1991). The effects of counseling on smoking cessation among patients hospitalized with chronic obstructive pulmonary disease: A randomized clinical trial. *The International Journal of the Addictions*, *26*(1), 107–119.

Pellegrino, R., Viegi, G., Brusasco, V., Crapo, R. O., Burgos, F., Casaburi, R. E. A., … Hankinson, J. (2005). Interpretative strategies for lung function tests. *European Respiratory Journal*, *26*(5), 948–968.

Petrie, K. J., & Weinman, J. (1997). *Perceptions of Health and Illness: Current Research and Applications*. Taylor & Francis.

Petrie, K. J., Weinman, J., Sharpe, N., & Buckley, J. (1996). Role of patients' view of their illness in predicting return to work and functioning after myocardial infarction: Longitudinal study. *BMJ (Clinical Research Ed.)*, *312*(7040), 1191–1194. https://doi.org/10.1136/bmj.312.7040.1191

Pinto, L. M., Alghamdi, M., Benedetti, A., Zaihra, T., Landry, T., & Bourbeau, J. (2015). Derivation and validation of clinical phenotypes for COPD: A systematic review. *Respiratory Research*, *16*(1), 50. https://doi.org/10.1186/s12931-015-0208-4

Putcha, N., Puhan, M. A., Hansel, N. N., Drummond, M. B., & Boyd, C. M. (2013). Impact of co-morbidities on self-rated health in self-reported COPD: An analysis of NHANES 2001–2008. *COPD: Journal of Chronic Obstructive Pulmonary Disease*, *10*(3), 324–332. https://doi.org/10.3109/15412555.2012.744963

Quist-Paulsen, P. (2008). Cessation in the use of tobacco—Pharmacologic and non-pharmacologic routines in patients. *The Clinical Respiratory Journal*, *2*(1), 4–10. https://doi.org/10.1111/j.1752-699X.2007.00038.x

R: The R Project for Statistical Computing. (2015). Retrieved May 1, 2019, from https://www.r-project.org/

Raad, D., Gaddam, S., Schunemann, H. J., Irani, J., Abou Jaoude, P., Honeine, R., & Akl, E. A. (2011). Effects of water-pipe smoking on lung function: A systematic review and meta-analysis. *Chest*, *139*(4), 764–774. https://doi.org/10.1378/chest.10-0991

Rennard, S. I., Calverley, P. M. A., Goehring, U. M., Bredenbröker, D., & Martinez, F. J. (2011). Reduction of exacerbations by the PDE4 inhibitor roflumilast—The importance of defining different subsets of patients with COPD. *Respiratory Research*, *12*, 18. https://doi.org/10.1186/1465-9921-12-18

Rennard, S. I., & Vestbo, J. (2006). COPD: The dangerous underestimate of 15%. *Lancet (London, England)*, *367*(9518), 1216–1219. https://doi.org/10.1016/S0140-6736(06)68516-4

Reynolds, N. R. (2003). The problem of antiretroviral adherence: A self-regulatory model for intervention. *AIDS Care, 15*(1), 117–124. https://doi.org/10.1080/0954012021000039815

Rocke, D. M., & Woodruff, D. L. (1996). Identification of Outliers in Multivariate Data. *Journal of the American Statistical Association*, *91*(435), 1047–1061. https://doi.org/10.1080/01621459.1996.10476975

Rousseeuw, P. J., & Zomeren, B. C. van. (1990). Unmasking Multivariate Outliers and Leverage Points. *Journal of the American Statistical Association*, *85*(411), 633–639. https://doi.org/10.1080/01621459.1990.10474920

Rovet, J., & Ireland, L. (1994). Behavioral Phenotype in Children with Turner Syndrome. *Journal of Pediatric Psychology*, *19*(6), 779–790. https://doi.org/10.1093/jpepsy/19.6.779

Scanlon, P. D., Connett, J. E., Waller, L. A., Altose, M. D., Bailey, W. C., Sonia Buist, A., & e Lung Health Study Research Group, D. P. T. for th. (2000). Smoking Cessation and Lung Function in Mild-to-Moderate Chronic Obstructive Pulmonary Disease. *American Journal of Respiratory and Critical Care Medicine*, *161*(2), 381–390. https://doi.org/10.1164/ajrccm.161.2.9901044

Schauer, G. L., Wheaton, A. G., Malarcher, A. M., & Croft, J. B. (2014). Smoking Prevalence and Cessation Characteristics among U.S. Adults With and Without COPD: Findings from the 2011 Behavioral Risk Factor Surveillance System.

COPD: Journal of Chronic Obstructive Pulmonary Disease, *11*(6), 697–704. https://doi.org/10.3109/15412555.2014.898049

Schnell, K., Weiss, C. O., Lee, T., Krishnan, J. A., Leff, B., Wolff, J. L., & Boyd, C. (2012). The prevalence of clinically-relevant comorbid conditions in patients with physician-diagnosed COPD: A cross-sectional study using data from NHANES 1999–2008. *BMC Pulmonary Medicine*, *12*(1), 26. https://doi.org/10.1186/1471-2466-12-26

Schofield, I., Kerr, S., & Tolson, D. (2007). An exploration of the smoking-related health beliefs of older people with chronic obstructive pulmonary disease. *Journal of Clinical Nursing*, *16*(9), 1726–1735. https://doi.org/10.1111/j.1365-2702.2007.01701.x

Silverman, E. K., & Speizer, F. E. (1996). RISK FACTORS FOR THE DEVELOPMENT OF CHRONIC OBSTRUCTIVE PULMONARY DISEASE. *Medical Clinics*, *80*(3), 501–522. https://doi.org/10.1016/S0025-7125(05)70451-X

Stoller, J. K., & Aboussouan, L. S. (2005). Alpha1-antitrypsin deficiency. *Lancet (London, England)*, *365*(9478), 2225–2236. https://doi.org/10.1016/S0140-6736(05)66781-5

Strassmann, R., Bausch, B., Spaar, A., Kleijnen, J., Braendli, O., & Puhan, M. A. (2009). Smoking cessation interventions in COPD: A network meta-analysis of

randomised trials. *European Respiratory Journal*, *34*(3), 634–640. https://doi.org/10.1183/09031936.00167708

Stratelis, G., Mölstad, S., Jakobsson, P., & Zetterström, O. (2006). The impact of repeated spirometry and smoking cessation advice on smokers with mild COPD. *Scandinavian Journal of Primary Health Care*, *24*(3), 133–139. https://doi.org/10.1080/02813430600819751

Sundh, J., Johansson, G., Larsson, K., Lindén, A., Löfdahl, C.-G., Sandström, T., & Janson, C. (2015). The phenotype of concurrent chronic bronchitis and frequent exacerbations in patients with severe COPD attending Swedish secondary care units. *International Journal of Chronic Obstructive Pulmonary Disease*, *10*, 2327–2334. https://doi.org/10.2147/COPD.S91362

Tan, W. C., Lo, C., Jong, A., Xing, L., FitzGerald, M. J., Vollmer, W. M., … Sin, D. D. (2009). Marijuana and chronic obstructive lung disease: A population-based study. *CMAJ : Canadian Medical Association Journal*, *180*(8), 814–820. https://doi.org/10.1503/cmaj.081040

Tashkin, D. P., Kanner, R., Bailey, W., Buist, S., Anderson, P., Nides, M., … Jamerson, B. (2001). Smoking cessation in patients with chronic obstructive pulmonary disease: A double-blind, placebo-controlled, randomised trial. *Lancet (London, England)*, *357*(9268), 1571–1575. https://doi.org/10.1016/s0140-6736(00)04724-3

Tashkin, D. P., Rennard, S., Hays, J. T., Ma, W., Lawrence, D., & Lee, T. C. (2011). Effects of varenicline on smoking cessation in patients with mild to moderate COPD: A randomized controlled trial. *Chest*, *139*(3), 591–599. https://doi.org/10.1378/chest.10-0865

Tetrault, J. M., Crothers, K., Moore, B. A., Mehra, R., Concato, J., & Fiellin, D. A. (2007). Effects of marijuana smoking on pulmonary function and respiratory complications: A systematic review. *Archives of Internal Medicine*, *167*(3), 221–228. https://doi.org/10.1001/archinte.167.3.221

Tilert, T., Dillon, C., Paulose-Ram, R., Hnizdo, E., & Doney, B. (2013). Estimating the U.S. prevalence of chronic obstructive pulmonary disease using pre- and post-bronchodilator spirometry: The National Health and Nutrition Examination Survey (NHANES) 2007–2010. *Respiratory Research*, *14*, 103. https://doi.org/10.1186/1465-9921-14-103

Tilert, T., Paulose-Ram, R., Howard, D., Butler, J., Lee, S., & Wang, M. Q. (2018). Prevalence and factors associated with self-reported chronic obstructive pulmonary disease among adults aged 40-79: The National Health and Nutrition Examination Survey (NHANES) 2007–2012. *EC Pulmonology and Respiratory Medicine*, *7*(9), 650–662.

Tomson, T., Toftgård, M., Gilljam, H., & Helgason, A. R. (2006). Symptoms in smokers trying to quit. *Tobacco Induced Diseases*, *3*(2), 44–58. https://doi.org/10.1186/1617-9625-3-2-44

Tønnesen, P., Mikkelsen, K., & Bremann, L. (2006). Nurse-conducted smoking cessation in patients with COPD using nicotine sublingual tablets and behavioral support. *Chest*, *130*(2), 334–342. https://doi.org/10.1378/chest.130.2.334

Torres-Duque, C., Maldonado, D., Pérez-Padilla, R., Ezzati, M., Viegi, G., & Forum of International Respiratory Studies (FIRS) Task Force on Health Effects of Biomass Exposure. (2008). Biomass fuels and respiratory diseases: A review of the evidence. *Proceedings of the American Thoracic Society*, *5*(5), 577–590. https://doi.org/10.1513/pats.200707-100RP

Vangeli, E., & West, R. (2008). Sociodemographic differences in triggers to quit smoking: Findings from a national survey. *Tobacco Control*, *17*(6), 410–415. https://doi.org/10.1136/tc.2008.025650

Vermylen, J. H., & Kalhan, R. (2013). Revealing the complexity of chronic obstructive pulmonary disease. *Translational Research : The Journal of Laboratory and Clinical Medicine*, *162*(4), 203–207. https://doi.org/10.1016/j.trsl.2013.07.008

Vestbo, J., & Rasmussen, F. V. (1989). Respiratory symptoms and FEV1 as predictors of hospitalization and medication in the following 12 years due to respiratory disease. *The European Respiratory Journal*, *2*(8), 710–715.

Vestbo, Jørgen. (2014). COPD: Definition and Phenotypes. *Clinics in Chest Medicine*, *35*(1), 1–6. https://doi.org/10.1016/j.ccm.2013.10.010

Viegi, G., Pedreschi, M., Pistelli, F., Di Pede, F., Baldacci, S., Carrozzi, L., & Giuntini, C. (2000). Prevalence of airways obstruction in a general population: European

Respiratory Society vs American Thoracic Society definition. *Chest*, *117*(5 Suppl 2), 339S-45S. https://doi.org/10.1378/chest.117.5_suppl_2.339s

Viegi, G., Pistelli, F., Sherrill, D. L., Maio, S., Baldacci, S., & Carrozzi, L. (2007). Definition, epidemiology and natural history of COPD. *The European Respiratory Journal*, *30*(5), 993–1013. https://doi.org/10.1183/09031936.00082507

Vogelzang, P. F., van der Gulden, J. W., Folgering, H., Kolk, J. J., Heederik, D., Preller, L., … van Schayck, C. P. (1998). Endotoxin exposure as a major determinant of lung function decline in pig farmers. *American Journal of Respiratory and Critical Care Medicine*, *157*(1), 15–18. https://doi.org/10.1164/ajrccm.157.1.9703087

Vozoris, N. T., & Stanbrook, M. B. (2011). Smoking prevalence, behaviours, and cessation among individuals with COPD or asthma. *Respiratory Medicine*, *105*(3), 477–484.

Wagena, E. J., Knipschild, P. G., Huibers, M. J. H., Wouters, E. F. M., & van Schayck, C. P. (2005). Efficacy of bupropion and nortriptyline for smoking cessation among people at risk for or with chronic obstructive pulmonary disease. *Archives of Internal Medicine*, *165*(19), 2286–2292. https://doi.org/10.1001/archinte.165.19.2286

Wells, S., & de Lusignan, S. (2003). Does screening for loss of lung function help smokers give up? *British Journal of Nursing (Mark Allen Publishing)*, *12*(12), 744–750. https://doi.org/10.12968/bjon.2003.12.12.11337

WHO. (2016). The top 10 causes of death. Retrieved November 27, 2018, from World Health Organization website: http://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death

WHO | COPD: Definition. (2018). Retrieved March 2, 2018, from WHO website: http://www.who.int/respiratory/copd/definition/en/

Willemse, B. W. M., Postma, D. S., Timens, W., & Hacken, N. H. T. ten. (2004). The impact of smoking cessation on respiratory symptoms, lung function, airway hyperresponsiveness and inflammation. *European Respiratory Journal*, *23*(3), 464–476. https://doi.org/10.1183/09031936.04.00012704

Wilson, D., Adams, R., Appleton, S., & Ruffin, R. (2005). Difficulties Identifying and Targeting COPD and Population-Attributable Risk of Smoking for COPD: A Population Study. *Chest*, *128*(4), 2035–2042. https://doi.org/10.1378/chest.128.4.2035

Wilson, J. S., Elborn, J. S., & Fitzsimons, D. (2011). 'It's not worth stopping now': Why do smokers with chronic obstructive pulmonary disease continue to smoke? A qualitative study. *Journal of Clinical Nursing*, *20*(5–6), 819–827. https://doi.org/10.1111/j.1365-2702.2010.03319.x

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, *18*(5), 459–482. https://doi.org/10.1002/cne.920180503

Zipf, G., Chiappa, M., & Porter, K. (2013). National Health and Nutrition Examination Survey: Plan and operations, 1999-2010. *Vital Health Statistics*, *1*(56).