COPD MORTALITY IN COLORADO; THE ROLE OF ENVIRONMENTAL EXPOSURES, REGIONALITY AND ETHNICITY

by

GREGORY LOYD KINNEY

B.S./B.S., University of Massachusetts, 1997

M.P.H., Tulane University, 2006

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Epidemiology Program

2012
This thesis for the Doctor of Philosophy degree by

Gregory Loyd Kinney

has been approved for the

Epidemiology Program

by

Deborah SK Thomas, Chair

John E Hokanson, Advisor

Lee S Newman

Lisa Cicutto

Sharon Lutz

Date 12/12/12
Kinney, Gregory Loyd (Ph.D., Epidemiology)

COPD Mortality in Colorado: The Role of Environmental Exposures, Regionality and Ethnicity

Thesis directed by Associate Professor Deborah S.K. Thomas

ABSTRACT

Chronic Obstructive Pulmonary Disease (COPD) is a condition caused by damage to the airways and alveoli that results in loss of elasticity, increased inflammation, increased mucus production and alveolar wall damage. This alveolar damage and mucus overproduction results in irreversible decreased airflow for those affected. While most COPD is attributable to past or current smoking, COPD can also result from environmental or occupational exposures. We examined COPD mortality data at the Zip Code Tabulation Area (ZCTA), the smallest geographic level available from the Colorado Department of Public Health and Environment, identifying spatial clustering and assessing the relationship between these clusters and markers of interest to public health. We found that spatial clustering of COPD mortality in Colorado is primarily explained by the population age and smoking status within the state. One cluster in the northern part of the state in the Greeley area was independent after adjusting for the age and smoking. This cluster has increased poverty, proximity to farming and ranching and a large Hispanic population. We then explored the role of Hispanic ethnicity and COPD mortality in a well characterized cohort in the San Luis Valley region of the state. We found a protective effect of Hispanic ethnicity on COPD mortality that was associated with smoking behavior. Hispanic participants reported a smaller number of pack years of
smoking as non-Hispanic Whites. However this was achieved by accumulating more years of smoking with a lower number of cigarettes per day. Hispanics also reported less smoke inhalation compared to non-Hispanic whites. In summary, clustering of COPD mortality in the state of Colorado is associated with the age and smoking distribution across the state. Remaining clustering may be associated with Hispanic ethnicity, though in Hispanics living in the San Luis Valley ethnic differences in COPD mortality are explained by smoking behavior.

The form and content of this abstract are approved. I recommend its publication.

Approved: Deborah S.K. Thomas
# TABLE OF CONTENTS

## CHAPTER

I. INTRODUCTION.................................................................1
   Background and Significance........................................1
   Specific Aims and Hypotheses......................................2

II. LITERATURE REVIEW..........................................................4
   COPD and Tobacco......................................................4
   COPD and Environmental Exposures...................................5
   COPD and Hispanic Ethnicity..........................................9
   COPD in Colorado.......................................................11
   COPD and Smoking Exposure........................................13

III. RESEARCH DESIGN AND METHODS.................................14
   Overview of State Level Study Design............................14
   State Level COPD Cases...............................................14
   State Level COPD Denominator......................................15
   State Level Smoking..................................................15
   State Level Soil Composition.......................................17
   State Level Geospatial Data.........................................17
   State Level Clustering Comparison....................................18
   State Level Statistical Methods..................................20
      State Level Clustering.............................................20
      Power and Sample Size for Clustering Analysis..............22
      State Level Cluster Comparison................................23
CHAPTER I
INTRODUCTION

Background and Significance

Chronic Obstructive Pulmonary Disease (COPD) is a progressive, heterogeneous group of conditions causing limitation of airflow from the lung\(^1\). Chronic lower respiratory diseases, which include COPD, were the third leading cause of death in 2009 after diseases of the heart and malignant neoplasms\(^2\). Smoking is the primary risk factor for COPD and it is estimated that 75% of all COPD related deaths are due to smoking\(^3\). Other risk factors for COPD include occupational exposures\(^4\) and environmental exposure to indoor\(^5\) or outdoor\(^6\) air pollutants. Sociodemographic factors such as increased age, low family income, and \(\leq 12\) years of education have also been found to be associated with mortality due to COPD\(^7\).

The American Lung Association of Colorado took an initial step in describing COPD mortality in the state of Colorado in its 2007 “Colorado Chronic Obstructive Pulmonary Disease Surveillance Report” where it reported age adjusted mortality rates by county for the period 1990 through 2005. In this work, they found that COPD mortality rates were higher in counties designated 'frontier' or 'rural' and that age adjusting those rates removed this regional effect. This broad finding does not take advantage of all of the available data however. We propose expanding that work by aggregating COPD mortality data at the Zip Code Tabulation (ZCTA) area, identifying spatial clustering, identifying specific spatial clusters and assessing their relationship with environmental exposures using publicly available databases and spatial statistical
methodologies. We will also explore the role of ethnicity on COPD mortality at both the state level and in a well-characterized cohort followed since 1984 in southern Colorado.

**Specific Aims and Hypotheses**

**Specific Aim 1**

To examine the relationship between COPD mortality and environmental exposures and ethnicity within Colorado between 1990 and 2008.

_Hypothesis; COPD mortality in the state of Colorado exhibits important spatial clustering._

Is COPD mortality within Colorado between 1990 and 2008 randomly distributed over space after controlling for the age structure of the state?

_Hypothesis; COPD mortality in the state of Colorado exhibits important spatial clustering after controlling for age._

Is COPD mortality within Colorado between 1990 and 2008 randomly distributed over space after controlling for the age structure and the spatial distribution of smoking?

_Hypothesis; COPD mortality in the state of Colorado exhibits important spatial clustering after controlling for age and smoking._

Is COPD mortality within Colorado between 1990 and 2008 related to arsenic exposure measured in the soil after controlling for age and smoking?

_Hypothesis; There is a positive correlation between environmental exposure to arsenic in the soil and COPD mortality in the state of Colorado._

Is COPD mortality related to ethnicity across the state of Colorado?

_Hypothesis; Overall COPD mortality in the state will be related to ethnicity but that_
relationship will be mitigated by land use and environmental exposure.

Specific Aim 2

To examine important predictors of COPD related mortality in a well-characterized cohort (San Luis Valley Diabetes Study) nested within the state of Colorado.

Is ethnicity an important predictor of COPD related mortality in the San Luis Valley?

Hypothesis; Ethnicity is an important predictor of COPD mortality after controlling for smoking history, comorbid conditions and other important covariates.
CHAPTER II
LITERATURE REVIEW

COPD and Tobacco

COPD is a major cause of morbidity and mortality in the United States. Since 1980, COPD mortality has increased from 23.5/100,000 to 42.1/100,000 in 2002, while other causes of death such as heart disease have decreased and in 2002 COPD was the 4th leading cause of death in the US. Overall in the United States, there is a trend towards increased COPD mortality and that trend is strongest in the western mountain region. Colorado, Wyoming and Montana, all states in the “western mountain” region of the US, ranked within the top 6 in the country for COPD related mortality in 1999 and 2000.

COPD prevalence is unmeasured at the state level though estimates by the American Lung Association indicate that >150,000 Coloradoans are likely affected by it. Smoking and age are primary predictors of COPD and smoking is not well measured in Colorado at the ZCTA level. While the Colorado Tobacco Attitudes and Behaviors Survey (TABS) was conducted in both 2001 and repeated in 2005 using a well validated methodology, these data are only released at the county level (personal communication). The 2006 NHLBI report of state specific smoking rates indicates that Colorado ranks 41st with an overall smoking rate of 17.9 (95% C.I. 16.6-19.2).

Applying smoking rates aggregated at a larger level to a smaller level may decrease precision when alternate data sources are available. Lung cancer mortality has been shown to be strongly associated with cumulative smoking and has been used as a proxy measure for smoking related mortality by other groups and Colorado lung
cancer mortality is well reported and available at the ZCTA level of aggregation. Best and Hansell considered COPD mortality and lung cancer mortality jointly using Bayesian approach in Great Britain. They found that while COPD mortality clustered with lung cancer mortality, COPD mortality exhibited important spatial clustering that was not associated with lung cancer but occurred in areas known to have environmental exposures that are associated with COPD\textsuperscript{13}. Given Colorado's past and continued smoking habits and potential environmental exposures, COPD will likely remain an important consideration in Colorado public health for many years to come.

**COPD and Environmental Exposures**

Apart from smoking, environmental exposures whether ambient\textsuperscript{14} or through occupation\textsuperscript{15}, have been shown to be associated with increased COPD morbidity and mortality in a variety of populations\textsuperscript{16}. Chronic exposure to arsenic in food, water or air\textsuperscript{17} over months to years/lifetime, has been shown to be associated with an increased risk of cancers including lung cancer\textsuperscript{18}, epithelial changes, cardiovascular disease, diabetes, altered immune system function\textsuperscript{19} and bronchiectasis\textsuperscript{20}. Chronic exposure to water containing arsenic is associated with decreased lung function, chronic bronchitis\textsuperscript{21}, chronic cough\textsuperscript{22}, COPD, interstitial lung disease and bronchiectasis\textsuperscript{23} in non-smokers.

Arsenic is a metalloid that is a naturally occurring component of soil and human exposure can be through water, food or air\textsuperscript{17}. High soil concentrations of arsenic tend to arise in soils formed through the weathering of marine sediments (shale). Drever reports that while the levels of arsenic in granite, basalt, sandstone and limestone vary between 1 and 2 mg/kg, shale’s average 13mg/kg\textsuperscript{24}. Arsenic levels in the soil vary across the state
of Colorado from 1ppm to 126ppm and average 5.7ppm with the highest concentrations occurring in the “central Colorado mineral belt” and along the Pierre and Mancos shale formations (personal communication, David B Smith, USGS, Ft Collins). Arsenic in municipal tap water in Colorado is monitored under the federal Safe Drinking Water Act and a recent report by the non-profit Environmental Working Group reports Environmental Protection Agency data showing that arsenic levels in municipal water (between 1998 and 2003), vary widely in a sampling of 174 municipal water suppliers. Levels range from 75.5 ppb (95% CI; 63-88) to <0.01 ppb (95%CI; 0-0.05) and they report that 1,061,728 people are exposed to drinking water containing arsenic levels greater than 10 ppb, the current EPA guideline for maximum safe exposure [http://www.ewg.org/tapwater/]. Those that live in rural areas and use well water are not required to be monitored by the EPA and their arsenic levels are open to speculation.

Arsenic is used in industry in both organic and non-organic forms, both of which can be harmful to humans with the inorganic forms toxicities being the most thoroughly explored. Ingesting high levels of inorganic arsenic can cause death through multiple organ failure due to inhibition of sulphhydryl group-containing cellular enzymes and replacement of phosphate molecules in ‘high-energy’ compounds (‘arsenolysis’). Ingested low doses of arsenic are thought to exhibit carcinogenic activity by effecting redox status, transcriptional alterations and DNA repair. Epidemiologic studies have suggested that this effect is non-linear and varies between populations, doses and the overall health of the population being studied.

Studies examining cytotoxicity and cell irritation in the lung due to PM10
particle composition have shown that arsenic was one of the principal drivers of the observed negative effect of arsenic on WTHBF-6 (clonal bronchial fibroblast) cells in vitro. Kozul et al provide us with an important potential pathway leading from water born arsenic exposure to lung damage via an alteration in the immune system causing dendritic cells to show altered response to viral challenge. When their mouse model was exposed to 100 ppb arsenic in their drinking water and then challenged with flu virus, they were slow to respond and when a response did occur if was excessive and showed signs of inducing lung damage, increased inflammation and reduced blood O2 saturation. They suggest that this pathway may lead to excess bronchiectasis in individuals exposed to arsenic and then challenged with virus. Work by Olsen et al shows that airway epithelial cells in culture that are exposed to arsenic at doses ranging from 0-290 ppb healed at a slower rate in a dose response manner. As with many exposures there is disagreement as to the level of effect and several of these exposure/outcomes are areas of active research. Animal models suggest that arsenic exposure can lead to lung damage in those co-exposed to flu virus and cell culture models show that once damage is caused in the lung, epithelial cells exposed to low doses of arsenic are slower to heal than unexposed cells. Damage to the lung on a cellular level, altered immune response, and decreased healing of lung epithelium due to chronic low levels of exposure to arsenic are a logical pathway leading to a worse prognosis for those with COPD.
The USGS soil survey of the state of Colorado provides geocoded measures of soil levels of arsenic. There are visual correlations between high levels of arsenic in the soil and increased COPD mortality and to our knowledge the spatial relationship between these levels and COPD mortality has not been explored [Figure 1];

Figure 1: Soil Arsenic in Colorado. shows the soil levels of Arsenic with the lowest levels (1-5 ppm) in small, black circles and the highest levels (41-125 ppm) in large circles colored red. Each circle represents a sample point for the USGS survey.

Given that there is variance in arsenic exposure across Colorado in the soil, water and presumably the air, exploring the relationship between arsenic level and mortality associated with lung disease (COPD) is reasonable, possible given available data sources and of interest to public health in the state.

**COPD and Hispanic Ethnicity**

The literature regarding Hispanic ethnicity and COPD morbidity/mortality is mixed though generally indicates that Latino ethnicity is protective for COPD versus
non-Hispanic whites though these data have not been thoroughly explored in the United States. Another population characteristic that is potentially important when attempting to understand the distribution of COPD mortality in the state of Colorado is ethnicity. The United States was 12.5% Hispanic at the time of the 2000 census and Colorado’s population was 17.1% Hispanic. This was an increase from 12.9% in the state during the 1990 census. Latino ethnicity is an important aspect of Colorado demographics.

Proyecto LatinoAmerican de Investigación en Obstrucción Pulmonar (PLATINO) measured COPD prevalence in five major cities in Latin American countries (Mexico, Brazil, Chile, Uruguay and Venezuela) and found that COPD prevalence, measured by spirometry, ranged from 7.8% in Mexico City to 20% in Montevideo. This range was higher than expected when compared to other international sources (4% - 10%). Samet et al were able to compare COPD mortality rates in a Hispanic population to a white population in New Mexico where they found the Hispanics had lower rates than US whites in each age group [Table1].

On the other hand Lipton et al report that Hispanic ethnicity (measured by census report) was an important positive predictor of spatial “hot spots” of increased hospitalization charges related to COPD in California using 2 years of cross sectional data. Adams et al also found no protective effect of Hispanic ethnicity (Mexican American) by GOLD stage in the San Antonio Longitudinal Study of Aging (SALSA) despite lower exposure to tobacco in the Hispanic participants. This mixed literature suggests complicated interplay between ethnicity and COPD.
Little is known about genetic susceptibility to COPD mortality in specific Hispanic populations in part due to admixture. Brehm et al correctly point out that “Hispanic/Latino” is a heterogeneous group genetically and that earlier work has methodological flaws in its accurate assessment\(^3^9\). The mixed nature of these findings suggests that the effect of Hispanic ethnicity on COPD mortality may be complicated by heterogeneity in the measurement of ethnicity itself. Any protective effect of this specific ethnicity for COPD mortality could be masked by this heterogeneity.

The San Luis Valley Cohort will allow us to examine COPD mortality in a Hispanic population (measured by self reported Spanish ancestry) with little immigration or admixture\(^4^0\). This area of Colorado was settled by Spanish explorers in the early 1700’s and the population has remained relatively stable to the present day. This bi-ethnic cohort (Non-Hispanic White and Hispanic) was formed in a limited geographical area and has been followed for 25 years. Ancestrally Informative Markers (AIMs) were measured on a sample of the Hispanic participants of this cohort and admixture analysis finds 61.6% European, 32.8% Native American and 5.6% African components on
average\textsuperscript{41}. As such, this cohort offers a unique opportunity to understand COPD related mortality in this subpopulation of Hispanics.

The results of examining Hispanic ethnicity as it relates to COPD are mixed. Hispanic ethnicity appears to be protective for COPD mortality and may be predictive of clustering of COPD prevalence. Hispanic ethnicity does not appear to be associated with COPD prevalence overall but the studies of prevalence may not be comparable due to regional differences in Hispanic population.

This study will explore COPD mortality as it relates to Hispanic ethnicity overall in the state of Colorado and also in a well established cohort study nested within the state that has a long and well studied history and little immigration.

**COPD in Colorado**

The American Lung Association of Colorado in conjunction with CDPHE used the 1990 – 2005 COPD mortality data to investigate the regional distribution of COPD mortality at the county level. This study found an overall association with regionality, defined as Urban (one city with 50,000 or more inhabitants or a census bureau defined urbanized area and a total metropolitan population of at least 100,000), Rural (neither urban nor frontier) and Frontier (6 or fewer people per square mile)\textsuperscript{42}

This work found that COPD related mortality at the county level was higher in the frontier and rural counties compared to the urban counties in the state [Table 2]. This difference was found to be due to the different age structure of the regions and the effect was mitigated when the rates were age adjusted\textsuperscript{43}.
Table 2: Crude and age adjusted mortality rates in Colorado, 1990 - 2005

<table>
<thead>
<tr>
<th>Region</th>
<th>Crude Rate*</th>
<th>Age-Adjusted Rate*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urban</td>
<td>39.9</td>
<td>54.5</td>
</tr>
<tr>
<td>Rural</td>
<td>51.4**</td>
<td>53.1</td>
</tr>
<tr>
<td>Frontier</td>
<td>57.1***</td>
<td>51.4</td>
</tr>
</tbody>
</table>

*Rates are COPD deaths per 100,000 population

** p <0.05 vs. urban, RR 1.29, 95% CI 1.24-1.33

*** p < 0.05 vs. urban, RR 1.43, 95% CI 1.34-1.52 and rural, RR 1.11, 95% CI 1.04-1.19

This study was performed using geographically large aggregate units (county) and many of these units may include heterogeneity in population density and demographics. Visual examination of the spatial distribution of the age adjusted rates suggests that there is spatial clustering of increased rates in the plains of Colorado and in certain mountainous areas. Addressing COPD mortality at a smaller aggregate level may reveal important spatial clustering beyond that detectable at the county or regional level and will allow us to examine important demographic features that are measured at the ZCTA level.

**COPD and Smoking Exposure**

Studies in the United Kingdom have modeled lung cancer mortality as a proxy for smoking and found that there was still significant spatial clustering of COPD mortality after adjusting for these effects. Best and Hansell\textsuperscript{13} interpreted the spatial clustering that was shared by both lung cancer mortality and COPD mortality to be primarily smoking
related and that spatial clustering attributable to COPD mortality alone to be due to exposures other than smoking, and speculated that these included environmental exposures. They found that COPD mortality not associated with lung cancer clustered in large towns and urban areas as well as industrial and mining areas of Britain.

Regions with clustering of increased COPD mortality that is not associated with age or smoking should be explored in order to understand whether risk factors in these areas are modifiable. These regions may become the target of public health efforts aimed at reducing COPD mortality due to specific regional factors.
CHAPTER III
RESEARCH DESIGN AND METHODS

Overview of State Level Study Design

This study will investigate COPD mortality in Colorado in two ways; ecologically by exploring regional geographic clustering of increased COPD mortality and longitudinally by exploring predictors of COPD mortality in an existing cohort study. Initially, geographic clustering of increased rates will be calculated without controlling for known risk factors. Subsequently, known risk factors will be controlled for and then clustering will be re-assessed. At each step, this study will compare demographic factors of interest to the public health community between clusters. This process will continue until all regional spatial clustering are explained. Those variables used will constitute a final model of COPD mortality in the state. The primary predictor of longitudinal COPD mortality will be Hispanic ethnicity and this study will control for known risk factors as well as known competing risks. Subsequent paragraphs will explore each of these approaches.

State Level COPD Cases

The Colorado Department of Public Health and Environment is able to provide access to state specific mortality data available for the years 1990 through 2008. This data source will be queried to obtain the broadest possible definition of COPD mortality including all primary and secondary causes of death using International Classification of Disease (ICD9 or ICD10); 490-496 (ICD-9), J40-J47 (ICD-10) (from: National Center for Healthcare Statistics). The switch from ICD9 to ICD10 occurred in 1999 and resulted
in a 6% increase in assignment of COPD as cause of death post 1999. ICD codes in the primary or any of 14 secondary causes of death fields will be stripped to 3 digits identifying these broad mortality classifications; 490: bronchitis NOS, 491: chronic bronchitis, 492: emphysemas, 493: asthma, 494: bronchiectasis, 495: extrinsic allergic alveolitis, 496: COPD. Any occurrence (primary or any of the 14 secondary codes) of ICD code 496 resulted in a classification of COPD. A primary code of asthma with no secondary code of emphysema, bronchiectasis or extrinsic allergic alveolitis was classified as not having COPD. A primary code of asthma with a secondary code matching any of the above was included as COPD. Primary bronchiectasis or primary extrinsic allergic alveolitis were included as COPD. Primary bronchitis or chronic bronchitis with a secondary code suggesting COPD was included as COPD.

Various age cutoffs have been used for research involving COPD; 25+ Mannino44, 35+ MMWR/NHLBI45 and 45+ Healthy People 201046. CDPHE data include age at death and will allow us to perform analysis using any of these cutoffs for comparison.

**State Level COPD Denominator**

US Census population counts from the 2000 census aggregated at the ZCTA level will constitute the denominator for these analyses.

**State Level Smoking**

Colorado Tobacco Attitudes and Behaviors Survey (TABS) data on smoking is generated by a CDPHE/Behavioral Risk Factor Surveillance System (BRFSS) smoking survey data is publicly available for 2000 at the county level. BRFSS is a collaborative project between the CDC and the states that measures behavioral risk factors such as
smoking as a part of its core module in all states\textsuperscript{47}.

Prevalence of current and past smoking was estimated using the 2000-2001 Colorado Tobacco Attitudes and Behaviors Surveys (TABS). This survey used a two phase, weighted random number dialing approach to survey the state and smoking rates are provided at the county level. TABS established a baseline for the evaluation of a program for tobacco use reduction and prevention based on tobacco Master Settlement Agreement funds\textsuperscript{48}. Respondents were English or Spanish speakers, 18 years or older living in a house with a telephone. After contact was established with a household an adult was randomly chosen for interview regardless of smoking status and a second household member was also chosen if they had smoked 100 or more cigarettes in their lives. The survey was based on the 1999 California Tobacco Survey\textsuperscript{49} and resulted in a representative statewide sample aggregated at the county level\textsuperscript{50}. Current and past smoking exposure is available through TABS at the county level.

Because county boundaries overlap ZCTA boundaries, we will estimate ZCTA level smoking counts using a spatial approach. County and ZCTA geographic boundaries will be obtained from the US Census TIGER files. Using ESRI’s ArcGIS, the areas of each county will be merged with each ZCTA area in order to calculate the percentage of each ZCTA falling into various counties. For all ZCTAs that cross the border of a county, the percent area included in each county will be calculated and subsequent analysis will assign population to each county based on this area. In the next step, the county level age specific smoking rates will be applied to the age specific population of the ZCTA based on the percent area in each county. These will be summed to create a ZCTA level
estimated count of smokers, which will then be used in the regression model to control for smoking.

**State Level Soil Composition**

The United States Geological Survey (USGS) reported soil composition for a spatially random sample of sites covering the state of Colorado conducted in 2006. The state was divided into 960 polygons and a random site was chosen in each. A field crew then went to each site and obtained a single soil sample from 0 to 12 cm depth from a location as close as possible to the randomly assigned sites. The actual sampling location was then geocoded for mapping. This methodology results in measures that are very comparable to the A-horizon (surface or topsoil) samples collected by USGS for its 2005 continent wide soil survey. Each sample was analyzed for 44 elements of environmental significance, including arsenic (ppm), using a combination of inductively coupled plasma-mass spectrometry for trace elements and inductively coupled plasma-atomic emission spectrometry for major elements following acid digestion. This approach is designed to assess material that humans may come into contact with through ingestion, inhalation or contact with the skin, during a normal day and should be seen as a proxy measure for chronic exposure to soil components.

**State Level Geospatial**

The Colorado Geographic Information Portal (CGIP) provides access to state specific (boundaries, roads, population center and regional populations) geospatial data for use in this study with ArcGIS.
State Level Clustering Comparisons

In 2008 Colorado adopted statewide public health improvement legislation\textsuperscript{52} which required assessment of health equity and social determinants of health. This resulted in a statewide effort (Colorado’s Health Assessment and Planning System, CHAPS) to identify a list of indicators that could be used to perform comparable community assessment. For this study, we used this list to determine variables available at the ZCTA level that could be used to compare clusters within domains of interest to public health; Community Description, Economic Opportunity (EO), Physical Environment (PE), Social Factors (SF), Health Behavior and Conditions (HBC), Mental Health (MH), Access to Care and Population Health Outcomes (PHO). Indicator variables, percent population Living Below Poverty Level (LBPL), percent of children(<18) below poverty level (LBPL<18), percent of workers that commute to work by biking, walking or public transportation (CommutingPTWB), from applicable domains, Economic Opportunity (EO) and Physical Environment (PE), that were available at the ZCTA level were selected for comparison in this study.

The CHAPS variable list was expanded to include other population aspects of interest, from the EO domain; percent of population <18 (P<18), percent of population >65 (P>65), percent of population >65 years below poverty level (LBPL>65), percent of population not US citizen (NUSC), NUSC living below poverty level (LBPLNUSC). From the PE domain; percent population working (Working), percent of population commuting to work (Commuting) and percent of population employed in farm/fish/forest (Working FFF).
The PHO domain includes measures of morbidity and mortality but is not specific for COPD and this domain was expanded using Colorado Hospital Association (CHI) discharge data. This data source was queried for the year 2000 using the same broad ICD criteria to identify primary discharge diagnosis of COPD for each hospitalization. Descriptive variables for each discharge were calculated using all hospitalizations within a cluster. Variables obtained from this source are; total count of COPD related hospitalizations, total days in hospital for each COPD hospitalization, total cost of COPD hospitalizations. Variables calculated at the cluster level are; hospitalizations/100,000, total cost per person, cost per day of hospitalization, total cost per hospitalization. Other CHAPS domains were not included because the data was not available at the ZCTA level or was not available for the year 2000.

**State Level Statistical Methods**

**State Level Clustering**

A spatial dataset will be constructed at the Zip Code Tabulation Area (ZCTA) aggregating COPD mortality counts between 1990 and 2008 as the numerator for the analysis and average population count from the 2000 census as the denominator. ZCTAs with extremely small event counts (<5 per ZCTA) will be excluded from the analysis. SaTScan (v8.0)[Martin Kulldorff and Information Management Services Inc.] software will be used to calculate a spatial scan statistic\(^{53}\) which is used to detect spatial clusters in count(discrete) data. This method requires that the number of events be Poisson distributed with a known, underlying population at risk. Census data provides the
underlying population in this case and COPD mortality is rare event count data and the probability of observing an event is Poisson distributed [APPENDIX A].

The spatial scan statistic creates a circular window of varying size, ranging from the smallest distance between a pair of region centroids, in this case zip code centroids, to a maximum value determined by the user, in this case one half the width of the study area. The software gradually moves that window across space recording the observed and expected events at each location and a region contributes all of its observations (numerators and denominators) if its centroid falls within the circle. This process creates many overlapping circles and allows the consideration of many circle sizes. The resulting clusters will be the shape of the aggregated zip codes as the term circle relates only to the ZCTA centroid distances. A maximum likelihood is then used to identify the most likely cluster in the data by comparing the risk within a region having a circle of a specific size to risk outside of that circle [APPENDIX B].

Significance is determined using Monte Carlo simulation assuming a constant risk testing whether the highest observed count is higher than we would anticipate given the null. This approach avoids the multiple testing risk inherent in testing all observed rates and provides both an overall test of clustering and a test of the significance of the most likely cluster. 999 datasets are generated assuming the null hypothesis and the maximum value for each iteration is stored resulting in a valid comparison between simulations as each iteration is independent. This p-value is then interpreted as the probability of a value higher than the maximum value observed being observed anywhere in the study area.
Significant spatial clustering of the data will be interpreted as increased risk for COPD mortality due to covariates that are not being controlled for in the raw data. Further hypotheses within this specific aim will attempt to address covariates potentially important in the observed spatial clustering. Adjusting for the age distribution of the population and observing changes in the location of significant spatial clusters will be interpreted as addressing heterogeneity in the age structure of the state.

Changes in the location of spatial clusters with this exposure adjusted data will be interpreted as being due to the influence of that exposure. For instance were a significant age adjusted cluster to cease to be significant after adjusting for smoking, that change would be interpreted as smoking having an important effect on COPD mortality in that spatial location. If a significant cluster were not affected by controlling for smoking, than the opposite would be true that smoking does not have an important effect on COPD mortality in that spatial location.

**Power and Sample Size for Clustering Analysis**

The spatial clustering analysis uses population count data and the sample size is essentially fixed overall and spatially for this interval. We are not interested in estimating over time changes in population or changes in rates of disease as we attempt to answer the ecological question “if there is spatial clustering, where does it occur”? In many but not all cases power is largely dependent on sample size and in the specific case where the analysis attempts to identify a spatial cluster, sample size will vary at each aggregate point, thus the ability to detect clusters has heterogeneous power. In other words; “there
is not a single summary power value for a test to detect clusters when applied to a study area with spatially heterogeneous population density.”

Monte Carlo methods can allow a comparison between your observations and a specific set of alternative hypotheses by simulating the test statistics distribution for the null hypothesis and then simulating the distribution of the studies specific alternative. As analysis of the geography of COPD mortality in Colorado is at an early stage defining a specific alternate hypothesis in order to attempt to assess power is likely premature.

**State Level Cluster Comparison**

Comparisons by cluster will be performed using Kruskal-Wallis analysis of variance which tests if the medians of the cluster are equal. This test requires that the distribution and scale of the values in a cluster have identical shape. This assumption was tested visually and clusters with markedly different distributions were removed for purposes of the overall test. A significant chi-Square p-value indicates that at least one of the clusters originates from a different population. We will use an overall test rather than individual comparisons by cluster in order to reduce the risk of reporting ecological fallacy. The ecological fallacy occurs when the risk for an individual is assumed using grouped data.

**Hypothesis Testing**

“*COPD mortality in the state of Colorado exhibits important spatial clustering*”

SaTScan will be used to identify clustering of crude COPD mortality rates as described above and the relative risk and p value for each cluster will be reported for each significant cluster. Clusters will be compared using CHAPS indicators.
“COPD mortality in the state of Colorado exhibits important spatial clustering after controlling for age.”

SAS will be used to adjust the crude COPD mortality rates for each ZCTA in the state to the age structure of the state obtained from the US Census 2000 report [APPENDIX C]. This model creates a table of predicted values for each aggregate area adjusted for current and past smoking. The results can then be read by SaTScan and an adjusted spatial scan statistic calculated and compared to the unadjusted clustering. Goodness of fit of the Poisson model will be examined using Pearson chi-squared and deviance tests and over/under dispersion of the data will be assessed. SaTScan will then be used to identify clusters of age adjusted COPD mortality rates and these clusters will be compared using CHAPS indicators.

“Is COPD mortality within Colorado between 1990 and 2005 randomly distributed over space after controlling for the spatial distribution of smoking?”

Building on clustering of crude COPD mortality rates we will adjust the expected values for each aggregate level (ZCTA) for smoking as assessed by TABS using SAS. SaTScan will then be used to identify clusters of age and smoking adjusted COPD mortality and clusters will be compared using CHAPS indicators.

“Is COPD mortality within Colorado between 1990 and 2005 related to environmental exposures measured in the soil?”

Building on earlier analyses this analysis will compare clustering of COPD mortality at the ZCTA aggregate to soil levels to arsenic. The USGS survey provides point level soil exposure data across the state as has been described previously. This point
level data will be aggregated to ZCTA level using block kriging\textsuperscript{57} with appropriate transformations of the data\textsuperscript{58}. Kriging is a geostatistical technique that predicts unmeasured interval values for spatial point data\textsuperscript{56}. This type of kriging goes beyond simply averaging all of the point values in an aggregate unit by including neighboring values and this type of interpolation is implemented in ArcGIS. The aggregate values will then be exported to SAS and included in the preceding model as a covariate and the adjusted rates will be used in SaTScan to identify significant clusters of COPD mortality.

**COPD mortality and Hispanic ethnicity**

**Overview of San Luis Valley Diabetes Cohort Study (SLVDS) Design**

This study will use an existing cohort of participants nested in the state of Colorado. This cohort was ascertained between 1980 and 1988 and was designed to assess the risk of type 2 diabetes in a Hispanic population. The study recruited roughly 50% Hispanic participants and 50% Non-Hispanic White (NHW) participants and has performed mortality follow-up until the present day. This mortality follow-up includes all primary and secondary causes of death. The study collected smoking assessments that include smoking status, duration, intensity and behavior for all participants at the baseline study visit. We will assess the risk of Hispanic ethnicity on COPD mortality accounting for smoking and the competing risk of cardiovascular disease mortality.

**San Luis Valley Diabetes Study (SLVDS) Participant Ascertainment**

The San Luis Valley Diabetes Study (SLVDS) was designed to determine first the prevalence of non-insulin dependent diabetes mellitus (NIDDM) in a clearly defined population composed of Hispanic and non Hispanic whites (NHW) and second the
incidence of NIDDM in that population. This location is ethnically 43.5% Hispanic as assessed by the 1980 census question “Are you of Spanish/Hispanic origin or descent?” and this population has experienced little in-migration or American Indian-Hispanic intermarriage. All participants were 20 – 74 years of age.

Phase 1 of SLVDS was begun in 1984 and captured all cases of NIDDM in a two county region of southern Colorado (Alamosa and Conejos counties) through review of medical records in all locations providing health care in the area. This approach identified 343 cases who attended a clinic session where they received a fasting oral glucose tolerance test, blood was collected and they were surveyed as to demographics and exposures.

A control group of 607 was then selected using a geographically based two-stage sampling procedure where all residential structures were identified in the two county region. A sample of structures was identified (20.9%) and staff enumerated 96.6% of the sample using a structured interview. In the second stage control subjects were selected from that sampling frame based on county, age, sex and ethnic group of the NIDDM cases. That sample was then assessed identically to the cases and this phase was completed in 1986.

Phase 2 of SLVDS was conducted between 1986 and 1988 and collected all incident cases of NIDDM in that period (293) and a second random sample of controls not already participating in SLVDS (744). Participants in Phase 2 were assessed using identical methodologies to Phase 1. Phase 3 identified a further 100 participants resulting in an overall sample size of 1890.
An intensive follow-up of vital status was conducted on this cohort in 1998. This follow-up consisted of telephone contact, death certificate matching and tracking of posted obituaries and it identified 316 deceased individuals. Coroners’ reports, autopsy results and medical records were obtained for 98% of those deceased at that time. A less intensive mortality follow-up was conducted in 2002 identifying a further 149 decedents and all International Classification of Disease (ICD9 if death occurred before 1999 or ICD10 beginning in 1999) codes (primary and secondary) were recorded for each death. The most recent death match was performed in August of 2007 in a sub-cohort of 480.

**SLVDS COPD Mortality**

The broadest possible definition of COPD will be used including all primary and secondary causes of death using International Classification of Disease (ICD9 or ICD10); 490-496 (ICD-9), J40-J47 (ICD-10) as in the state level study. ICD codes in the primary or any of the secondary causes of death fields will be stripped to 3 digits identifying these broad mortality classifications; 490: bronchitis NOS, 491: chronic bronchitis, 492: emphysemas, 493: asthma, 494: bronchiectasis, 495: extrinsic allergic alveolitis, 496: COPD. Any occurrence (primary or any of the secondary codes) of ICD code 496 resulted in a classification of COPD mortality. Other related ICD codes will be aggregated as in the earlier study.

**SLVDS Covariates**

Ethnicity was assessed at baseline and a thorough smoking history was collected on members of this cohort regardless of when they joined it. Detailed assessment of physical activity including work related productive activity was performed on 903
participants in the first follow-up phase of the SLVDS (1987-1992) using the 1-yr Physical Activity History (PAH)\textsuperscript{59}. A general set of questions assessing employment status and industry were asked of each participant. Smoking history was obtained on all participants including variables measuring start and stop dates and amount smoked for participants who no longer smoked. This methodology allows us to calculate pack years in all current and former smokers. Other questions relating to smoking behavior such as depth of draw, frequency of inhalation and amount of cigarette left unsmoked were also collected.

**SLVDS Statistical Methods**

Cox Proportional Hazards regression will be used to investigate the association between Hispanic ethnicity in this cohort study. Ethnicity in the SLVDS cohort is primarily Hispanic and non-Hispanic white due to the cultural structure of the area. As such this will be the primary predictor of COPD mortality. Smoking history is of the most significant predictor of COPD and as such will be explored as a categorical (current, former and never) variable. Smoking history was also obtained on participants and pack years (as of the last visit for current smokers) is available. At the time of the 2002 mortality follow-up the SLVDS cohort had accumulated 26868.7 person-years of observation in 1887 participants.

A Cox regression of the log hazard ratio on a covariate with a standard deviation of 1.5000 based on a sample of 1887 observations achieves 99% power at an 0.05000 significance level to detect a regression coefficient equal to 0.5940. The sample size was adjusted as a multiple regression of the variable of interest on the other covariates in the
Figure 2: Power and Sample Size Calculation. Power and sample size chosen to detect a difference at four values of B for the SLVDS sample size.

Cox regression is expected to have an R-Squared of 0.5000. The sample size was adjusted for an anticipated event rate of 0.0240. The size of the regression coefficient to be detected = 0.594 (Hazards Ratio = 1.812) and the event rate is currently 0.024 so if the association remains stable and the event rate is reduced to 0.015 the study will retain has 80% power to detect the same HR [APPENDIX D].
CHAPTER IV
SPATIAL CLUSTERING OF CHRONIC OBSTRUCTIVE PULMONARY DISEASE MORTALITY IN COLORADO

Background

Because of the public health burden inherent in COPD related hospitalizations and mortality, the spatial distribution of COPD related illness has been examined by several groups. Studies have observed geographic variability and spatial heterogeneity of admissions related to regional socio-economic effects, such as poverty in areas with a history of high smoking rates, as well as local effects, such as occupation in mining or exposure to pollution. Colorado and the Rocky Mountain West experience a disproportionate rate of mortality related to COPD, while simultaneously exhibiting average to below average smoking prevalence and lower Smoking Attributable Mortality rates compared to other states.

This study examines clustering of COPD related mortality in the state of Colorado controlling for known risk factors where we hypothesize that statistical clustering of increased rates of COPD mortality will be identified. We will then compare areas of high COPD related mortality in order to better understand possible socio-demographic factors related this clustering. We will also investigate whether spatial clustering remains after controlling for the effect of age and smoking and how socio-demographic factors are related to those clusters.
Methods

This study uses georeferenced mortality data collected by the Colorado Department of Public Health and the Environment (CDPHE) between 1990 and 2008 as the primary outcome variable. Census data assessing demographic variables of interest was obtained from the 2000 US Census. Smoking variables were obtained from the 2000 Tobacco Attitudes and Behaviors Survey. The Zip Code Tabulation Area (ZCTA) level geographic identifier was available for each mortality event during these years. Each record captured the ZCTA of the home residence of the decedent at the time of death, month and year of death, gender, primary cause of death, and all secondary causes of death as recorded on the decedent’s death certificate. Thus, data were aggregated at the ZCTA level for the spatial analysis.

Mortality data were queried to obtain a broad definition of COPD related mortality, including all primary and secondary causes of death in those greater than 25 years of age at time of death using International Classification of Disease (ICD9 or ICD10); 490-496 (ICD-9), J40-J47 (ICD-10) (from: National Center for Healthcare Statistics). The switch from ICD9 to ICD10 occurred in 1999 and resulted in a 6% increase in assignment of COPD as cause of death after that year. As all 18 years of data were pooled for the numerator this temporal association was not considered. ICD codes in the primary or any secondary cause of death fields were restricted to 3 digits identifying these broad mortality classifications; 490: bronchitis NOS, 491: chronic bronchitis, 492: emphysemas, 493: asthma, 494: bronchiectasis, 495: extrinsic allergic alveolitis, 496: COPD. Any occurrence (primary or any of the secondary codes) of ICD
code 496 resulted in a classification of COPD. A primary code of Asthma with no secondary code of emphysema, bronchiectasis or extrinsic allergic alveolitis was classified as not having COPD and a primary code of Asthma with a secondary code matching any of the above was included as COPD. Primary bronchiectasis or primary extrinsic allergic alveolitis were included as COPD. Primary bronchitis or chronic bronchitis with a secondary code suggesting COPD was included as COPD.

The number of individuals older than 25 years of age for each ZCTA from the US Census serves as the denominator for these analyses. Age adjustment was conducted using the same census count categorized by 10 year age groups beginning with 25-34 and continuing through an aggregate age group of >75. Age adjustment was performed using the 2000 CDC age adjustment table distributions #14 and #18.

Prevalence of current and past smoking was estimated using the 2000-2001 Colorado Tobacco Attitudes and Behaviors Surveys (TABS). This survey used random digit dialing in the state of Colorado to establish a baseline for the evaluation of a program for tobacco use reduction and prevention based on tobacco Master Settlement Agreement funds. Respondents were English or Spanish speakers, 18 years or older living in a house with a telephone. After contact was established with a household an adult was randomly chosen for interview regardless of smoking status and a second household member was also chosen if they had smoked 100 or more cigarettes in their lives. The survey was based on the 1999 California Tobacco Survey and resulted in a representative statewide sample aggregated at the county level. Current and past smoking exposure is available through TABS at the county level.
Because county boundaries overlap ZCTA boundaries, we estimated ZCTA level smoking counts using a spatial approach. County and ZCTA geographic boundaries were obtained from the US Census TIGER files. Using ESRI’s ArcGIS, the areas of each county were merged with each ZCTA area in order to calculate the percentage of each ZCTA falling into various counties. For all ZCTAs that crossed the border of a county, the percent area included in each county was calculated and subsequent analysis assigned population to each county based on this area. In the next step the county level age specific smoking rates were applied to the age specific population of the ZCTA based on the percent area in each county. These were then summed to create a ZCTA level estimated count of smokers, which was then used in the regression model to control for smoking.

In 2008, Colorado adopted statewide public health improvement legislation which required assessment of health equity and social determinants of health. This resulted in a statewide effort (Colorado’s Health Assessment and Planning System, CHAPS) to identify a list of indicators that could be used to perform comparable community assessment. For this study we used this list to determine variables, available at the ZCTA level, that could be used to compare clusters within domains of interest to public health; Community Description, Economic Opportunity (EO), Physical Environment (PE), Social Factors (SF), Health Behavior and Conditions (HBC), Mental Health (MH), Access to Care and Population Health Outcomes (PHO). Indicator variables, percent population Living Below Poverty Level (LBPL), percent of children(<18) below poverty level (LBPL<18), percent of workers that commute to work
by biking, walking or public transportation (CommutingPTWB), from applicable domains, Economic Opportunity (EO) and Physical Environment (PE), that were available at the ZCTA level were selected for comparison in this study. The CHAPS variable list was expanded to include other population aspects of interest, from the EO domain; percent of population <18 (P<18), percent of population >65 (P>65), percent of population >65 years below poverty level (LBPL>65), percent of population not US citizen (NUSC), NUSC living below poverty level (LBPLNUSC). From the PE domain; percent population working (Working), percent of population commuting to work (Commuting) and percent of population employed in farm/fish/forest (Working FFF).

The PHO domain includes measures of morbidity and mortality but is not specific for COPD. Because of this the PHO domain was expanded using Colorado Hospital Association (CHI) discharge data. This data source was queried for the year 2000 using the same broad ICD criteria to identify primary discharge diagnosis of COPD for each hospitalization. Descriptive variables for each discharge were calculated using all hospitalizations within a cluster. Variables obtained from this source are; total count of COPD related hospitalizations, total days in hospital for each COPD hospitalization, total cost of COPD hospitalizations. Variables calculated at the cluster level are; hospitalizations/100,000, total cost per person, cost per day of hospitalization, total cost per hospitalization. Other CHAPS domains were not included because the data was not available at the ZCTA level or was not available for the year 2000.

Data management and estimation of adjusted numerators was performed using SAS 9.3 and spatial clustering of COPD mortality was assessed using a spatial scan
statistic implemented by the SaTScanv9.1.153. We performed a purely spatial analysis using a Poisson probability model scanning for high rates of mortality. The scan statistic reports a risk ratio comparing the ratio of observed to expected events within the cluster to the ratio outside of that cluster. Significance for this statistic is based on the p value obtained from 999 Monte Carlo replications of random distributions across the region. Visualization was conducted using ARCMap65.

Comparisons by cluster were performed using Kruskal-Wallis one way analysis of variance which tests the medians by cluster. This test requires that the distribution and scale of the values in a cluster have identical shape. This assumption was tested visually and one cluster with markedly a different distribution was removed for purposes of the overall test. A significant chi-Square p value indicates that at least one of the clusters originates from a different population. We used an overall test rather than individual comparisons by cluster in order to reduce the risk of reporting ecological fallacy. The ecological fallacy occurs when the risk for an individual is assumed using grouped data. We advise the reader to consider this when reading the grouped values for each cluster.

Results

Between 1990 and 2008, 57,354 COPD related mortality events represented 11.8% of the mortality experience in the 2,770,315 persons ≥25 years old counted in the 2000 US Census. The estimated cumulative incidence of COPD related mortality in Colorado for this period of time was 207.4/100,000 [Table 3].

The distribution of COPD mortality in Colorado shows significant spatial clustering. In general there is more clustering of crude COPD mortality in the Eastern
Table 3: Statewide Demographic Measures

<table>
<thead>
<tr>
<th>Crude Clustering</th>
<th>Colorado</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZCTA count</td>
<td>501</td>
</tr>
<tr>
<td>Population &gt;= 25 (2000)</td>
<td>2,770,315</td>
</tr>
<tr>
<td>Cases (1990-2008)</td>
<td>57,354</td>
</tr>
<tr>
<td>Cases/100,000</td>
<td>207.03</td>
</tr>
<tr>
<td>Living Below Poverty Line(%)</td>
<td>8.9</td>
</tr>
<tr>
<td>Population Under 18 (%)</td>
<td>25.7</td>
</tr>
<tr>
<td>LBPL &lt; 18 (%)</td>
<td>9.3</td>
</tr>
<tr>
<td>Population &gt; 65 (%)</td>
<td>10.3</td>
</tr>
<tr>
<td>LBPL &gt; 65 (%)</td>
<td>6.2</td>
</tr>
<tr>
<td>Population Not US Citizen (%)</td>
<td>1.8</td>
</tr>
<tr>
<td>Not US citizens, LBPL (%)</td>
<td>9.6</td>
</tr>
<tr>
<td>Working (%)</td>
<td>68.4</td>
</tr>
<tr>
<td>Commuting to work (%)</td>
<td>94.4</td>
</tr>
<tr>
<td>Commute to work PT/W/Bike (%)*</td>
<td>5.4</td>
</tr>
<tr>
<td>Employed in farm/fish/forest (%)</td>
<td>0.9</td>
</tr>
<tr>
<td>Hispanic Ethnicity (%)</td>
<td>7.5</td>
</tr>
</tbody>
</table>

Statewide median values of each of the demographic variables reported in this study.

* Workers >16 that commute to work by Biking, Walking or using Public Transportation plains, followed by the western slopes. Figure 3a shows crude COPD mortality ratios for each ZCTA in the state with darker colors represent higher ratios. There are 6 distinct, statistically significant (p<0.001) clusters of COPD mortality in Colorado (Figure 3b). The south-eastern plains of the state (dark brown) are a cluster of the highest prevalence of COPD mortality and include 53 aggregate ZCTA. This cluster is associated with a greater than 2 fold increase in COPD mortality compared to the rest of the state. The 6 clusters of COPD mortality have unique demographic characteristics shown in Table 4 LBPL, P<18, LBPL<18, P>65, LBPLNUSC, Working, Commuting, CommutingPTWB, WorkingFFF and Hispanic were significantly different between the 6 clusters of crude COPD mortality though poverty in the elderly (LBPL>65) and citizenship status (NUSC)
Legend Figure 3a; COPD mortality ratios in Colorado. Figure 3a shows COPD mortality ratios for each ZCTA in the state with darker colors representing higher ratios.

Legend Figure 3b: Clustering of Crude COPD Mortality. Figure 3b shows clustering of COPD mortality ratios in Colorado with darker colors representing more likely clusters. All clusters shown in color are significant at p=0.001.
Table 4: Crude Clustering Demographic Characteristics

<table>
<thead>
<tr>
<th>ZCTA count</th>
<th>Cluster 1 Southeastern Colorado</th>
<th>Cluster 2 Denver</th>
<th>Cluster 3 Grand Junction</th>
<th>Cluster 4 Colorado Springs</th>
<th>Cluster 5 Northeastern Colorado</th>
<th>Cluster 6 Salida</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>53</td>
<td>37</td>
<td>33</td>
<td>6</td>
<td>65</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Latitude</td>
<td>37.409787 N</td>
<td>39.749436 N</td>
<td>39.053689 N</td>
<td>38.834282 N</td>
<td>40.916929 N</td>
<td>38.533464 N</td>
<td></td>
</tr>
<tr>
<td>Longitude</td>
<td>102.617260 W</td>
<td>104.989480 W</td>
<td>108.539290 W</td>
<td>104.816540 W</td>
<td>102.218580 W</td>
<td>106.034680 W</td>
<td></td>
</tr>
<tr>
<td>Cluster Radius (km)</td>
<td>205.49</td>
<td>12.13</td>
<td>90.25</td>
<td>4.55</td>
<td>221.58</td>
<td>66.11</td>
<td></td>
</tr>
<tr>
<td>Total Population</td>
<td>129630</td>
<td>561611</td>
<td>126507</td>
<td>85032</td>
<td>123577</td>
<td>37280</td>
<td></td>
</tr>
<tr>
<td>Cases</td>
<td>5669</td>
<td>15278</td>
<td>4352</td>
<td>2961</td>
<td>3941</td>
<td>1411</td>
<td></td>
</tr>
<tr>
<td>Expected Cases</td>
<td>2683.54</td>
<td>11626.21</td>
<td>2618.89</td>
<td>1760.29</td>
<td>2558.23</td>
<td>771.75</td>
<td></td>
</tr>
<tr>
<td>Cases/100,000</td>
<td>242.5</td>
<td>150.8</td>
<td>190.7</td>
<td>193.1</td>
<td>176.8</td>
<td>209.8</td>
<td></td>
</tr>
<tr>
<td>Relative Risk</td>
<td>2.23</td>
<td>1.43</td>
<td>1.72</td>
<td>1.72</td>
<td>1.58</td>
<td>1.85</td>
<td></td>
</tr>
<tr>
<td>p (RR)</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>Living Below Poverty Line (%)</td>
<td>16.4</td>
<td>9.9</td>
<td>9.6</td>
<td>13.6</td>
<td>11.7</td>
<td>10.3</td>
<td>&lt;0.0001†</td>
</tr>
<tr>
<td>Population Under 18 (%)</td>
<td>26.2</td>
<td>22.6</td>
<td>24.4</td>
<td>22.7</td>
<td>28.9</td>
<td>21.4</td>
<td>&lt;0.0001†</td>
</tr>
<tr>
<td>LBPL &lt; 18 (%)</td>
<td>21.7</td>
<td>14.3</td>
<td>11.7</td>
<td>17.5</td>
<td>12.7</td>
<td>11.4</td>
<td>0.0004†</td>
</tr>
<tr>
<td>Population &gt; 65 (%)</td>
<td>15.4</td>
<td>10.2</td>
<td>15.5</td>
<td>11.9</td>
<td>14.0</td>
<td>16.7</td>
<td>0.0002†</td>
</tr>
<tr>
<td>LBPL &gt; 65 (%)</td>
<td>10.8</td>
<td>8.0</td>
<td>7.4</td>
<td>9.1</td>
<td>8.1</td>
<td>5.7</td>
<td>0.12†</td>
</tr>
<tr>
<td>Population Not US Citizen (%)</td>
<td>1.2</td>
<td>8.3</td>
<td>1.5</td>
<td>2.9</td>
<td>3.6</td>
<td>0.6</td>
<td>&lt;0.0001†</td>
</tr>
<tr>
<td>Not US citizens, LBPL (%)</td>
<td>26.4</td>
<td>25.4</td>
<td>8.3</td>
<td>17.9</td>
<td>4.3</td>
<td>0.0</td>
<td>0.07†</td>
</tr>
<tr>
<td>Working (%)</td>
<td>58.6</td>
<td>67.1</td>
<td>61.2</td>
<td>66.5</td>
<td>65.7</td>
<td>57.3</td>
<td>&lt;0.0001†</td>
</tr>
<tr>
<td>Commuting to work (%)</td>
<td>94.3</td>
<td>92.2</td>
<td>93.0</td>
<td>93.8</td>
<td>95.4</td>
<td>94.1</td>
<td>&lt;0.0001†</td>
</tr>
<tr>
<td></td>
<td>Cluster 1 Southeastern Colorado</td>
<td>Cluster 2 Denver</td>
<td>Cluster 3 Grand Junction</td>
<td>Cluster 4 Colorado Springs</td>
<td>Cluster 5 Northeastern Colorado</td>
<td>Cluster 6 Salida</td>
<td>p value</td>
</tr>
<tr>
<td>-----------------------</td>
<td>---------------------------------</td>
<td>------------------</td>
<td>--------------------------</td>
<td>----------------------------</td>
<td>--------------------------------</td>
<td>-----------------</td>
<td>---------</td>
</tr>
<tr>
<td>Commute with PT/W/Bike (%)*</td>
<td>5.9</td>
<td>10.6</td>
<td>5.2</td>
<td>6.4</td>
<td>5.1</td>
<td>5.5</td>
<td>0.0003†</td>
</tr>
<tr>
<td>Employed in farm/fish/forest (%)</td>
<td>4.7</td>
<td>0.2</td>
<td>2.5</td>
<td>0.1</td>
<td>5.1</td>
<td>0.9</td>
<td>&lt;0.0001†</td>
</tr>
<tr>
<td>Hispanic Ethnicity (%)</td>
<td>15.0</td>
<td>15.5</td>
<td>7.4</td>
<td>14.4</td>
<td>7.4</td>
<td>5.6</td>
<td>&lt;0.0001†</td>
</tr>
</tbody>
</table>

Median Values Shown
* Workers >16 that commute to work by Biking, Walking or using Public Transportation
† Kruskal-Wallis Chi-Square p value
Figure 4a

Legend Figure 4a: Age Adjusted COPD Mortality Ratios. Figure 4a shows age adjusted COPD mortality rates in Colorado with darker colors representing higher ratios.

Figure 4b

Legend Figure 4b: Clustering of Age Adjusted COPD Mortality. Figure 4b shows Clustering of age adjusted COPD mortality ratios in Colorado with darker colors representing more likely clusters. All clusters shown in color are significant at $p=0.001$. 
were not different between clusters. The southeastern cluster has the highest percent Living Below Poverty Line (LBPL) overall and also in at risk populations (<18, >65 and Population Not US Citizen (PNUSC)). Age adjusted COPD mortality by ZCTA are shown in Figure 4a.

There are 6 cluster of age adjusted COPD mortality with 40 ZCTAs in the Denver metropolitan areas representing the most likely cluster shown in Figure 4b. The most likely cluster is associated with a 1.28 times higher COPD mortality compared to the rest of the state. Similar geographic regions are represented in the age adjusted COPD mortality. The cluster in the southeast region of the state (the most likely cluster in the crude analysis) shifted slightly to the west. The risk of COPD mortality associated with this cluster has a Risk Ratio 1.34 and p-value 0.001.

Several demographic markers in Table 4 that contributed in the crude analysis are no longer significantly different after age adjustment in Table 5: percent LBPL, LBPL in vulnerable populations (<18, >65 and percent PNUSC) and total percent of the population >65. Demographic variables that now vary by cluster are percent total population <18, PNUSC, variables related to the work force (total percent working, commuting, commuting using public transportation and percent employed in farm/fish/forest) and percent Hispanic ethnicity. The primary cluster shows the lowest percent of population <18, and those employed in farm/fish/forest and the highest percent PNUSC, working, commuting to work using PT/W/Bike and Hispanic ethnicity.
<table>
<thead>
<tr>
<th>Cluster 1 Denver</th>
<th>Cluster 2 Southeast of Pueblo</th>
<th>Cluster 3 Northeastern Plains, Holyoke</th>
<th>Cluster 4 Northwest, Rangely</th>
<th>Cluster 5 Southwest, Dolores</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZCTA count</td>
<td>40</td>
<td>54</td>
<td>86</td>
<td>17</td>
<td>5</td>
</tr>
<tr>
<td>Latitude</td>
<td>39.784622 N</td>
<td>38.089719 N</td>
<td>40.582954 N</td>
<td>40.011447 N</td>
<td>37.515311 N</td>
</tr>
<tr>
<td>Longitude</td>
<td>104.962140 W</td>
<td>104.378070 W</td>
<td>102.304510 W</td>
<td>108.775280 W</td>
<td>108.452040 W</td>
</tr>
<tr>
<td>Cluster Radius (km)</td>
<td>13.18</td>
<td>94.05</td>
<td>224.51</td>
<td>106.68</td>
<td>23.17</td>
</tr>
<tr>
<td>Total Population</td>
<td>550954</td>
<td>304506</td>
<td>180616</td>
<td>67643</td>
<td>14578</td>
</tr>
<tr>
<td>Cases</td>
<td>17521</td>
<td>10314</td>
<td>5742</td>
<td>2268</td>
<td>482</td>
</tr>
<tr>
<td>Expected Cases</td>
<td>14454.27</td>
<td>7988.71</td>
<td>4738.46</td>
<td>1774.61</td>
<td>382.45</td>
</tr>
<tr>
<td>Cases/100,000</td>
<td>176.3</td>
<td>187.8</td>
<td>176.25</td>
<td>185.9</td>
<td>183.3</td>
</tr>
<tr>
<td>Risk Ratio</td>
<td>1.28</td>
<td>1.34</td>
<td>1.23</td>
<td>1.29</td>
<td>1.26</td>
</tr>
<tr>
<td>p (RR)</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>0.003</td>
</tr>
<tr>
<td>Living Below Poverty Line(%)</td>
<td>10.2</td>
<td>11.9</td>
<td>11.4</td>
<td>8.1</td>
<td>12.9</td>
</tr>
<tr>
<td>Population Under 18 (%)</td>
<td>25.4</td>
<td>26.2</td>
<td>29.1</td>
<td>26.5</td>
<td>26.9</td>
</tr>
<tr>
<td>LBPL &lt; 18 (%)</td>
<td>13.6</td>
<td>16.1</td>
<td>12.7</td>
<td>8.3</td>
<td>23.2</td>
</tr>
<tr>
<td>Population &gt; 65 (%)</td>
<td>9.2</td>
<td>13.5</td>
<td>13.2</td>
<td>13.7</td>
<td>14.1</td>
</tr>
<tr>
<td>LBPL &gt; 65 (%)</td>
<td>7.4</td>
<td>8.7</td>
<td>7.7</td>
<td>6.6</td>
<td>7.7</td>
</tr>
<tr>
<td>Population Not US Citizen (%)</td>
<td>7.5</td>
<td>1.4</td>
<td>1.6</td>
<td>1.6</td>
<td>1.3</td>
</tr>
<tr>
<td>Not US citizens, LBPL (%)</td>
<td>25.0</td>
<td>9.9</td>
<td>6.5</td>
<td>7.3</td>
<td>0.0</td>
</tr>
</tbody>
</table>
Table 5: Age Adjusted Clustering of Demographic Characteristics

<table>
<thead>
<tr>
<th></th>
<th>Cluster 1 Denver</th>
<th>Cluster 2 Southeast of Pueblo</th>
<th>Cluster 3 Northeastern Plains, Holyoke</th>
<th>Cluster 4 Northwest, Rangely</th>
<th>Cluster 5 Southwest, Dolores</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Working (%)</td>
<td>67.3</td>
<td>61.9</td>
<td>66.0</td>
<td>66.6</td>
<td>66.3</td>
<td>0.01†</td>
</tr>
<tr>
<td>Commuting to work (%)</td>
<td>92.1</td>
<td>92.8</td>
<td>95.6</td>
<td>92.8</td>
<td>92.0</td>
<td>&lt;0.0001†</td>
</tr>
<tr>
<td>Commute to work PT/W/Bike (%)</td>
<td>10.3</td>
<td>4.3</td>
<td>4.8</td>
<td>5.3</td>
<td>4.9</td>
<td>&lt;0.0001†</td>
</tr>
<tr>
<td>Employed in farm/fish/forest (%)</td>
<td>0.2</td>
<td>0.6</td>
<td>4.3</td>
<td>1.6</td>
<td>0.8</td>
<td>&lt;0.0001†</td>
</tr>
<tr>
<td>Hispanic Ethnicity (%)</td>
<td>16.4</td>
<td>14.0</td>
<td>7.6</td>
<td>9.7</td>
<td>7.9</td>
<td>0.0003†</td>
</tr>
</tbody>
</table>

* Workers >16 that commute to work by Biking, Walking or using Public Transportation

† Kruskal-Wallis Chi-Square p value
Figure 5a

Legend Figure 5a: Age and Smoking Adjusted COPD Mortality Ratios. Figure 5a shows age adjusted COPD mortality ratios in Colorado with darker colors representing higher ratio.

Figure 5b

Legend Figure 5b: Clustering of Age and Smoking Adjusted COPD Mortality. Figure 5b shows clustering of age and smoking adjusted COPD mortality rates in Colorado with darker colors representing more likely clusters. All clusters shown in color are significant at p=0.001.
Adjusting each ZCTA for current and past smoking removes all but one cluster from the state. Figure 5a shows the age and smoking adjusted COPD mortality ratio for each ZCTA. Figure 5b shows the single cluster remaining which is composed of 3 aggregates centered in an urban area in northern Colorado. Of all clusters identified this cluster includes the highest percent Hispanic (40.5%), highest percent working (70.2%) and is one of the highest LBPL (16.4%) and LBPL in the elderly (8.5%).

Table 6: Age and Smoking Adjusted Demographic Characteristics

<table>
<thead>
<tr>
<th></th>
<th>Cluster 1 Greeley</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZCTA count</td>
<td>3</td>
</tr>
<tr>
<td>Latitude</td>
<td>40.421845 N</td>
</tr>
<tr>
<td>Longitude</td>
<td>104.691750 W</td>
</tr>
<tr>
<td>Cluster Radius (km)</td>
<td>5.57 km</td>
</tr>
<tr>
<td>Total Population</td>
<td>53718</td>
</tr>
<tr>
<td>Cases</td>
<td>1792</td>
</tr>
<tr>
<td>Expected Cases</td>
<td>1424.57</td>
</tr>
<tr>
<td>Cases/100,000</td>
<td>184.9</td>
</tr>
<tr>
<td>Risk Ratio</td>
<td>1.26</td>
</tr>
<tr>
<td>p (RR)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Living Below Poverty Line(%)</td>
<td>16.4</td>
</tr>
<tr>
<td>Population Under 18 (%)</td>
<td>27.2</td>
</tr>
<tr>
<td>LBPL &lt; 18 (%)</td>
<td>21.7</td>
</tr>
<tr>
<td>Population &gt; 65 (%)</td>
<td>8.6</td>
</tr>
<tr>
<td>LBPL &gt; 65 (%)</td>
<td>9.5</td>
</tr>
<tr>
<td>Population Not US Citizen (%)</td>
<td>12.0</td>
</tr>
<tr>
<td>Not US citizens, LBPL (%)</td>
<td>24.0</td>
</tr>
<tr>
<td>Working (%)</td>
<td>70.2</td>
</tr>
<tr>
<td>Commuting to work(%)</td>
<td>91.8</td>
</tr>
<tr>
<td>Commute to work PT/W/Bike (%)*</td>
<td>2.5</td>
</tr>
<tr>
<td>Employed in farm/fish/forest</td>
<td>1.1</td>
</tr>
<tr>
<td>Hispanic Ethnicity (%)</td>
<td>40.5</td>
</tr>
</tbody>
</table>

* Workers >16 that commute to work by Biking, Walking or using Public Transportation
† Kruskal-Wallis Chi-Square p value
Discussion

Spatial clustering of mortality associated with COPD demonstrates regional differences within the state of Colorado, after controlling for known risk factors. Risk ratios of COPD mortality are increased in the Great Plains region of the US compared to other regions despite generally lower levels of smoking in this region. Eastern Colorado, approximately two fifths of the state, is a part of the Great Plains and this region includes the most likely cluster of crude COPD mortality in our analysis. The crude spatial distribution of COPD mortality (Figure 4a) provides evidence for the heterogeneity of the underlying disease process giving rise to the observed distribution. In a public health setting, disease clustering can indicate where disease risk is increased for unknown reasons or suggest additional areas where risk ratios of disease are increased but are associated with known risk factors. Controlling for the age structure and smoking prevalence removed statistical clustering on the eastern plains, leaving a single urban center as the remaining cluster. This suggests that the clustering observed on the eastern plains is attributable to current and past smoking in this region.

There are few studies investigating the spatial clustering of COPD mortality and none have been performed in Colorado at an aggregate smaller than the county level. Weinhold noted in 1997 that the observed link between COPD mortality and smoking that is evident in much of the United States does not hold in the Rocky Mountain region of the country where he observes high rates of COPD mortality drawn from National Center for Health Statistics with low Smoking Attributable Mortality from CDC. Within the state, the 2007 Colorado surveillance report identified the rural/frontier counties of
the state as having increased risk ratios of COPD mortality and that 11 of 16 counties with the highest rates occurred on the eastern plains. The current study shows increased clustering of COPD mortality in regions that are identified as rural/frontier and we find that that clustering is explained by the age structure and smoking history of these regions.

Several studies have investigated COPD related hospitalization spatially; Holt et al.\textsuperscript{61} used national Medicare claims data for the years 1995-2006 and a Bayesian hierarchical spatial modeling approach at the Health Service Area aggregate to investigate the spatial distribution of primary COPD diagnosis discharges. They found that 73\% of the variability in COPD hospitalization risk was attributable to spatial structure. This indicates that broad regional effects (socioeconomic and increased regional smoking rates) have more influence than local effects (specific occupational exposures, localized air pollution, local demographics) but cannot identify what those effects are. The report identifies the strongest regional effect stretching from east Texas across the deep south into northern Florida and then north through Appalachia to the Great Lakes States. This study then removes the regional clustering effects leaving localized effects which are essentially un-clustered, one of which exists on the south-eastern plains region of Colorado and extends into Kansas to the east and Oklahoma to the south.

Joo et al.\textsuperscript{60} examined geographic variation in hospital admissions for COPD exacerbations in the VA system and found that the Veterans Integrated Service Networks (VISN) that includes Colorado was at slight but not significant increased risk compared to the VISN with the baseline median rate (RR = 1.04, 95\%C.I. 0.93-1.18). They also
showed that Hispanic ethnicity was protective for exacerbations in crude (RR = 0.92, 95% C.I. 0.87-0.97) but not adjusted (RR = 0.94, 95% C.I. 0.87-1.01) models. Morris et al\textsuperscript{66} described hospitalizations for respiratory disease in the elderly at the county level where they observed significant geographic heterogeneity of COPD admission rates (Moran’s I, p<0.0001). They observed significant associations between low socioeconomic status, smoking and occupational exposure with increased rates of COPD hospitalizations. Jackson et al\textsuperscript{67} investigated the regional distribution of COPD related hospitalizations in Texas and found that COPD hospitalization rates were lower in non-metropolitan counties however they were unable to control for smoking status or history. They also noted the lowest rates of COPD hospitalization in Hispanics compared to NHW (24/100,000 versus 145/100,000).

The current study agrees with the observation that increased COPD mortality occurs in broad regions and adds to the current state of knowledge by showing that that regional effect is driven by the age structure and past smoking history. This study does not identify Hispanic ethnicity as a protective effect however. The final cluster that remains after age and smoking adjustment shows the highest percent Hispanic population of any cluster identified, suggesting that ethnicity may play a different role in COPD mortality than it does in hospitalizations for COPD.

COPD mortality is increased in specific regions of Colorado, but that increase is associated with known risk factors and controlling for the age structure and lifetime smoking history removes much of this clustering. This results in only a remaining, single cluster of increased rates in an urban center in the northern part of the state suggesting an
unmeasured marker of disease in this area. This cluster is characterized by high LBPL overall and in youth and the elderly, while the median percent employed is higher than the state overall. This set of cluster features along with a large Hispanic population and high non-citizen US citizens suggests a working poor population at increased risk for COPD mortality.

Each of these analyses, crude, age adjusted and smoking adjusted COPD mortality rates have public health implications. Clusters of increased crude COPD mortality suggest regions where public health resources should be available to address existing COPD and prevent exacerbations that result in hospitalization and worsening of the disease. Clusters that exist after age adjustment may indicate regions that experienced high rates of smoking in the past and increased attention should be paid to identifying and addressing subclinical COPD. The cluster that remains after adjustment for age and smoking may indicate a location experiencing increased mortality due to a unique local exposure which might be modifiable and should be investigated using a different study design.

Limitations

When we performed the spatial merge that allowed us to apply county level smoking rates to ZCTAs that merge was executed using the geographic centroid. This assumes that the population of the ZCTA was distributed uniformly across the area and this assumption is fallacious. In order to base this merge on the population density, point locations for all denominator values (individuals in homes) would have to be obtained and this was not possible at the time of this writing.
This study reports period prevalence of COPD related mortality in an 18-year time frame. The eastern plains of Colorado behave differently from a demographic perspective than the Denver urban center. There is a net out-migration between 2000 and 2010 in the age group 18-29 and a net in-migration after that reaches a maximum at age 64 of nearly 3.5% and the tapers off after age 85 to under 2%. The Denver area experiences net in-migration across most age groups older than 4 year old. Migration patterns may play a role in changes to the denominator of this study that are not accounted for well when treating mortality as a period measure. Small numbers of cases at the aggregate level examined in this study limit our ability to look at spatial clustering incorporating the dimension of time.

In order to compare clusters using demographic data, this study only considered clusters of more than one ZCTA. The spatial scan statistic also identifies as clusters singleton ZCTA’s that are unlike their neighbors. These singleton clusters may represent important local effects that result in increased COPD mortality and would benefit from studies designed to test for local effects.

**Conclusions**

COPD is a chronic disease associated with smoking and aging and these factors account for much of the geographic distribution of related mortality in the state of Colorado. This spatial heterogeneity can be of use to the public health community which needs to direct resources that are specific to the complicated aspects of this disease. An aging population represents challenges to public health as health care utilizations increases through this natural process. Better serving our seniors in regions that we have
shown to be associated with the age structure should be considered. Smoking prevention is clearly the best option, but smoking cessation efforts can be directed to regions that have known clustering that is effected by controlling for this risk factor. The role of ethnicity in COPD mortality is unclear, but social factors associated with Hispanic ethnicity may play a role in the cluster that remains after adjusting for age and smoking history. Other known risk factors such as occupational exposures and decreased air quality do not appear to play an important role in large scale clustering of COPD mortality but they likely represent important predictors on a local scale.
CHAPTER V
PUBLIC HEALTH REPORT OF SPATIAL CLUSTERING OF CHRONIC OBSTRUCTIVE PULMONARY DISEASE IN COLORADO, 1990-2008

Background

Chronic Lower Respiratory Disease, which includes Chronic Obstructive Pulmonary Disease (COPD), was the 3rd leading cause of death after heart disease and cancer in the United States and in Colorado in 2009. Seventy percent of the medical care costs of COPD are due to hospitalizations\(^6\) and COPD has been shown to be responsible for 23% of all readmissions occurring within 30 days of discharge in a large public health system\(^6\). The Patient Protection and Affordable Care Act places emphasis on reducing preventable hospital re-admissions\(^2\) making COPD an ideal disease to target in these efforts.

The purpose of this report is to describe the spatial distribution of COPD related mortality and to explore how this distribution is affected by the age structure and smoking behavior in the state. This information can be used by state and local public health agencies to identify regions with populations at risk for COPD related hospitalizations and hospital readmissions that may precede mortality events.

Methods

This report uses mortality data georeferenced to the Zip Code Tabulation Area (ZCTA), derived from death certificates between 1990 and 2008 obtained from the Vital Statistics Unit in the Health Statistics Section at the Colorado Departments of Public Health and the Environment. COPD related deaths were identified using primary and
underlying cause of death codes in those older than 25 using International Classification of Disease codes 490-496, J40-J47, excluding codes for asthma that do not accompany a code for COPD. COPD mortality rates were calculated based on US Census 2000 population counts of people older than 25 for each ZCTA. Smoking rates were derived from the 2000 Colorado Tobacco Attitudes and Behaviors Survey (TABS) for the same age group at the same aggregate. Spatial clustering of COPD mortality was performed using the SaTScan software and clusters were considered to be significant at the α=0.05 level. For the purposes of this report, only clusters showing regional effects (more than a single ZCTA) are reported. Hospital discharge data was obtained from the Colorado Hospital Association (CHI) for all hospitalizations with a primary discharge diagnosis of COPD during 2000. 95% Confidence Intervals (95% C.I.) using a large sample approximation were calculated for hospitalization rates within clusters in order to assess statistically significance differences. Significance is determined using 95% C.I. overlap for each comparison; non-overlapping 95% C.I. are significant and overlapping 95% C.I. are not.

Results

Figures 6 through 8 show maps of Colorado highlighting aspects of COPD mortality. The figures labeled “a” show mortality rates for each ZCTA with darker colors representing higher rates. The figures labeled “b” show significant regional clusters of increased COPD mortality with the darkest color representing the most likely cluster and lighter colors representing clusters that also reach statistical significance based on the clustering analysis. Tables 7 through 9 show variables related to COPD hospitalization and directly describe the clusters shown in the “b” figure above them.
Legend Figure 6: COPD related mortality (A) and clustering of mortality (B) in Colorado

Table 7: Crude COPD related hospitalizations in 2000

<table>
<thead>
<tr>
<th></th>
<th>Cluster 1 Southeastern Colorado</th>
<th>Cluster 2 Denver</th>
<th>Cluster 3 Grand Junction</th>
<th>Cluster 4 Colorado Springs</th>
<th>Cluster 5 Northeastern Colorado</th>
<th>Cluster 6 Salida</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Population &gt; 25</td>
<td>129,630</td>
<td>561,611</td>
<td>126,507</td>
<td>85,032</td>
<td>123,577</td>
<td>37,280</td>
</tr>
<tr>
<td>Total COPD Hospitalizations</td>
<td>366</td>
<td>1290</td>
<td>249</td>
<td>200</td>
<td>376</td>
<td>69</td>
</tr>
<tr>
<td>Total COPD Hospitalization days</td>
<td>1735</td>
<td>5529</td>
<td>1011</td>
<td>894</td>
<td>1759</td>
<td>361</td>
</tr>
<tr>
<td>COPD Hospitalizations/100,000</td>
<td>282.3</td>
<td>229.7</td>
<td>196.8</td>
<td>235.2</td>
<td>304.3</td>
<td>185.1</td>
</tr>
<tr>
<td>95% C.I. Hospitalization Incidence</td>
<td>253.1-311.3</td>
<td>217.2-242.2</td>
<td>172.4-221.3</td>
<td>202.3-267.8</td>
<td>273.5-335.0</td>
<td>141.4-228.8</td>
</tr>
<tr>
<td>Average $/COPD Hospitalization day</td>
<td>$2,217.50</td>
<td>$3,038.38</td>
<td>$2,035.74</td>
<td>$3,489.65</td>
<td>$2,481.11</td>
<td>$2,180.63</td>
</tr>
<tr>
<td>Average $/COPD Hospitalization</td>
<td>$10,511.91</td>
<td>$13,215.49</td>
<td>$8,265.59</td>
<td>$15,598.75</td>
<td>$11,587.29</td>
<td>$11,408.81</td>
</tr>
</tbody>
</table>
Table 8: COPD related hospitalizations in 2000 by age adjusted cluster

<table>
<thead>
<tr>
<th></th>
<th>Cluster 1 Denver</th>
<th>Cluster 2 Southeast of Pueblo</th>
<th>Cluster 3 Northeastern Plains, Holyoke</th>
<th>Cluster 4 Northwest, Rangely</th>
<th>Cluster 5 Southwest, Dolores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Population &gt;25</td>
<td>550,954</td>
<td>304,506</td>
<td>180,616</td>
<td>67,643</td>
<td>14,578</td>
</tr>
<tr>
<td>Total COPD Hospitalizations</td>
<td>1218</td>
<td>623</td>
<td>469</td>
<td>160</td>
<td>20</td>
</tr>
<tr>
<td>Total COPD Hospitalization days</td>
<td>5209</td>
<td>2975</td>
<td>2172</td>
<td>642</td>
<td>116</td>
</tr>
<tr>
<td>COPD Hospitalizations/100,000</td>
<td>221.1</td>
<td>204.6</td>
<td>259.7</td>
<td>236.5</td>
<td>137.2</td>
</tr>
<tr>
<td>95% C.I. Hospitalization Incidence</td>
<td>208.7-233.5</td>
<td>188.5-220.7</td>
<td>236.2-283.2</td>
<td>199.9-273.2</td>
<td>177.1-197.3</td>
</tr>
<tr>
<td>Average $/COPD Hospitalization day</td>
<td>$3,097.55</td>
<td>$2,592.27</td>
<td>$2,611.73</td>
<td>$2,264.96</td>
<td>$2,401.63</td>
</tr>
<tr>
<td>Average $/COPD Hospitalization</td>
<td>$13,247.25</td>
<td>$12,378.82</td>
<td>$12,095.26</td>
<td>$9,088.14</td>
<td>$13,929.45</td>
</tr>
</tbody>
</table>
Legend Figure 8: COPD related mortality adjusted for age and smoking (A) and clustering of mortality (B) in Colorado

Table 9: COPD Related Hospitalizations in 2000 by Age and Smoking Adjusted Cluster

<table>
<thead>
<tr>
<th>Category</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Population &gt; 25</td>
<td>53,718</td>
</tr>
<tr>
<td>Total COPD Hospitalizations</td>
<td>12</td>
</tr>
<tr>
<td>Total COPD Hospitalization days</td>
<td>56</td>
</tr>
<tr>
<td>COPD Hospitalizations/100,000</td>
<td>22.3</td>
</tr>
<tr>
<td>95% C.I. Hospitalization Incidence</td>
<td>9.7-35</td>
</tr>
<tr>
<td>$/COPD Hospitalization day</td>
<td>$2,551.38</td>
</tr>
<tr>
<td>$/COPD Hospitalization</td>
<td>$11,906.42</td>
</tr>
</tbody>
</table>
Unadjusted Clustering of COPD mortality

The distribution of COPD mortality in Colorado shows significant spatial clustering (Figure 6a). In general there is more clustering of crude COPD mortality in the Eastern plains, followed by the western slope. Figure 6a shows crude COPD mortality rates for each ZCTA in the state. There are 6 distinct, statistically significant (p<0.001) clusters of COPD mortality in Colorado (Figure 6b). The southeastern plains of the state are a cluster of the highest incidence of COPD mortality and include 53 aggregate ZCTAs. This cluster is associated with a greater than 2 fold increase in COPD mortality compared to the rest of the state (RR = 2.23, p=0.001).

Cluster 1 experiences a higher risk ratio of COPD related hospitalizations than clusters 2,3 and 6, clusters 2, 3,4 and 6 are significantly lower than cluster 5 and no other comparisons are significantly different in Table 7.

Age Adjusted Clustering of COPD mortality

Age adjusted COPD mortality by ZCTA is shown in Figure 7a. There are 6 cluster of age adjusted COPD mortality with 40 ZCTAs in the Denver metropolitan areas representing the most likely cluster (Figure 7b). This cluster is associated with a 1.28 times higher COPD mortality compared to the rest of the state (RR=1.28, p=0.001). Similar geographic regions are represented in the age adjusted COPD mortality. The cluster in the southeast region of the state (the most likely cluster in the crude analysis) shifted slightly to the west. The risk of COPD mortality associated with this cluster is an RR=1.34, p=0.001.
Cluster 1, 2 and 5 experience lower rates of COPD related hospitalization than cluster 3, cluster 4 experiences higher rates than cluster 5 and no other comparisons are significantly different in Table 8.

**Age and Smoking Adjusted Clustering of COPD mortality**

Adjusting each ZCTA for current and past smoking removes all but one regional cluster in the state. Figure 8a shows the age and smoking adjusted COPD mortality rates for each aggregate and the remaining mortality cluster. This cluster, shown in Figure 8b, is composed of 3 ZCTAs centered in an urban area in Greeley Colorado with an RR of 1.26, p=0.001. As no other clusters remain in the state after age and smoking adjustment no comparisons of hospitalization rates are included in Table 9.

**Discussion**

Spatial clustering of mortality associated with COPD demonstrates regional differences within the state of Colorado, after controlling for known risk factors. Rates of COPD mortality are increased in the Great Plains region of the US compared to other regions despite generally lower levels of smoking in this region. Eastern Colorado, approximately two fifths of the state, is a part of the Great Plains and this region includes the most likely cluster of crude COPD mortality in our analysis. The crude spatial distribution of COPD mortality (Figure 6) provides evidence for the heterogeneity of the underlying disease process giving rise to the observed distribution. In a public health setting disease clustering can indicate where disease risk is increased for unknown reasons or suggest additional areas where rates of disease are increased but are associated with known risk factors. Controlling for the age structure and smoking prevalence in Colorado removed statistical clustering on the eastern plains leaving a single urban center...
as the remaining cluster. This suggests that the COPD related mortality clustering observed on the eastern plains is attributable to current and past smoking in this region.

There are few studies investigating the spatial clustering of COPD mortality and none have been performed in Colorado at an aggregate smaller than the county level. Weinhold\textsuperscript{71} noted in 1997 that the observed link between COPD mortality and smoking that is evident in much of the United States does not hold in the Rocky Mountain region of the country where he observes high rates of COPD mortality drawn from National Center for Health Statistics in the same regions as low Smoking Attributable Mortality. Within the state the 2007 Colorado surveillance report\textsuperscript{72} identified the rural/frontier counties of the state as having increased rates of COPD mortality and that 11 of 16 counties with the highest rates occurred on the eastern plains.

Several studies have investigated COPD related hospitalization spatially; Holt et al\textsuperscript{73} used national Medicare claims data for the years 1995-2006 at the Health Service Area aggregate to investigate the spatial distribution of hospitalizations with primary COPD diagnosis discharges. They found that 73\% of the variability in COPD hospitalization risk was attributable to spatial structure. This indicates that broad regional effects (socioeconomic and increased regional smoking rates) have more influence than local effects (specific occupational exposures, localized air pollution, local demographics) but was unable to identify what those effects are. The report identifies the strongest regional effect stretching from east Texas across the Deep South into northern Florida and then north through Appalachia to the Great Lakes States. This study then removed the regional clustering effects leaving localized effects which are essentially un-
clustered, one of which exists on the south-eastern plains region of Colorado and extends into Kansas to the east and Oklahoma to the south.

Joo et al\textsuperscript{60} examined geographic variation in hospital admissions for COPD exacerbations in the VA system and found that the Veterans Integrated Service Networks (VISN), which includes Colorado, was at slight but not significant increased risk compared to the VISN with the baseline median rate (RR = 1.04, 95\%C.I. 0.93-1.18). They also showed that Hispanic ethnicity was protective for exacerbations in crude (RR = 0.92, 95\% C.I. 0.87-0.97) but not adjusted (RR = 0.94, 95\% C.I. 0.87-1.01) models.

Morris et al\textsuperscript{66} described hospitalizations for respiratory disease in the elderly at the county level where they observed significant geographic heterogeneity of COPD admission rates (Moran’s I, p<0.0001). They observed significant associations between low socioeconomic status, smoking and occupational exposure with increased rates of COPD hospitalizations. Jackson et al\textsuperscript{74} investigated the regional distribution of COPD related hospitalizations in Texas and found that COPD hospitalization rates were lower in non-metropolitan counties however they were unable to control for smoking status or history. They also noted the lowest rates of COPD hospitalization in Hispanics compared to NHW (24/100,000 versus 145/100,000).

COPD mortality is increased in specific regions of Colorado, but that increase is associated with known risk factors and controlling for the age structure and lifetime smoking history removes much of this clustering. This results in only a remaining, single cluster of increased rates in an urban center in the northern part of the state suggesting an unmeasured marker of disease in this area. This cluster is characterized by high rates of vulnerable populations living below the poverty level, while the median percent
employed is higher than the state overall. This set of cluster features along with a large Hispanic population and a high percentage of non-US citizens suggest a working poor population at increased risk for COPD mortality. This is in contrast to the “healthy worker effect” reported in some populations and indicates a distinct need for further study on the population living in this cluster.

Each of these analyses (crude, age adjusted and smoking adjusted COPD mortality rates) has public health implications. Clusters of increased crude COPD mortality suggest regions where public health resources might be targeted to address existing COPD and prevent exacerbations that result in primary and re-hospitalization for COPD. Clusters that exist after age adjustment may indicate regions that experienced high rates of smoking in the past and increased attention should be paid to identifying and addressing subclinical COPD. The cluster that remains after adjustment for age and smoking may indicate a location experiencing increased mortality due to a unique local exposure which might be modifiable and should be investigated using a different study design. An aging population represents challenges to public health as health care utilization increases through this natural process.

Drawing inferences from ecological data must be done with caution and in this case inferences are arrived at by comparing an unadjusted map to a map adjusted for other factors such as age or smoking. This report suggests that placing increased emphasis on tertiary prevention of existing COPD in our senior population in regions where we have shown clustering related to age (Figure 6b unadjusted clustering compared to figure 7b clustering adjusted for age) may ultimately contribute to reduction in COPD related hospitalizations and subsequent re-hospitalizations. Placing increased
emphasis on secondary prevention of COPD by identifying undiagnosed cases of COPD in regions where we have shown clustering related to smoking (Figure 7b, clustering adjusted for age vs 8b clustering adjusted for age and smoking) should have a larger effect still as treatment exists to stabilize COPD progression in early disease. Finally smaller areas with increased rates of COPD mortality that do not cluster but are higher than their neighbors should be explored with appropriate study designs in order to identify contributing local factors. These findings can be applied to current and future programming, future research and community collaboration opportunities that may contribute to decreasing hospitalization costs and mortality related to COPD in Colorado.
CHAPTER VI

THE PROTECTIVE EFFECT OF HISPANIC ETHNICITY ON CHRONIC OBSTRUCTIVE PULMONARY DISEASE MORTALITY AND THE ASSOCIATION WITH LESS INTENSE SMOKING BEHAVIOR

Introduction

Chronic Obstructive Pulmonary Disease (COPD) is currently the third leading cause of death in the United States and is strongly associated with smoking, which has causal linkages to both COPD and cardiovascular disease. Individuals with COPD and co-existing cardiovascular disease have higher clinical and economic burden compared to either condition alone. The combination of COPD and cardiovascular disease (CVD) represent roughly 50% of the smoking attributable deaths per year in the United States. Individuals living with COPD have reduced functional capacity and poorer quality of life and people living with COPD have frequent emergency department visits and overall utilization of the health care system than.

Hispanic/Latino ethnic populations experience differential rates of COPD and COPD mortality though the evidence remains conflicting as to whether the disparity is attributable to genetic susceptibility or variation in smoking rates and/or smoking behavior. The most recent National Vital Statistics Report indicates that chronic lower respiratory disease, which includes COPD, represent the 7th leading cause of death in Hispanics. Proyecto LatinoAmericano de Investigación en Obstrucción Pulmonar (PLATINO) found COPD prevalence rates in five major Latin American cities to be 8-20%, much higher than the 4% to 10% expected. Samet et al found Hispanic ethnicity to be protective compared to Non-Hispanic Whites (NHW) and attributed this to
differences in smoking\textsuperscript{36}; however, Adams et al notes no difference in COPD severity by GOLD stage in the San Antonio Longitudinal Study of Aging\textsuperscript{38} despite lower exposure to tobacco in Hispanic participants. Lipton et al report that Hispanic ethnicity (assessed by census report) was an important positive predictor of spatial “hot spots” of increased hospitalization charges related to COPD in California using 2 years of cross sectional data\textsuperscript{37}.

Recent work by Bruse et al found that Hispanic ethnicity was protective for COPD and reduced pulmonary function and attributed this difference to genetic susceptibility beyond the effect of smoking\textsuperscript{83}. There have been few reports describing ethnic differences in smoking behaviors such as smoking intensity and depth of inhalation. Strom et al performed a case control study comparing smoking behaviors in Hispanic and African-American lung cancer patients and found that Hispanic participants were less likely to smoke but more likely to smoke intensely than African-Americans\textsuperscript{84}. Brehm et al correctly point out that “Hispanic/Latino” is a heterogeneous group genetically and that existing work in Hispanic populations suffers from methodological flaws that failed to capture the diversity of this demographic group\textsuperscript{39}. Thus, the literature on the interplay between Hispanic ethnicity and cigarette smoking in developing COPD and its associated morbidity and mortality remains confusing and uncertain and demands additional work.

The conflicting and limited nature of the existing literature suggests that the effect of Hispanic ethnicity on COPD mortality may be complicated by a number of factors such as heterogeneity in the measurement of ethnicity and smoking behaviors such as smoking intensity and depth of inhalation that may differ based on ethnicity. Thus, the
objective of the current study was to determine the relationships between COPD mortality, Hispanic ethnicity and smoking behavior accounting for the competing risk of cardiovascular disease.

Methods

The San Luis Valley Health Studies are derived from a geographically based cohort of two counties (Alamosa and Conejos counties) in rural southern Colorado. The study was designed to measure the prevalence and incidence of Non-Insulin Dependent Diabetes (NIDDM) in a Hispanic population. Based on 1980 census data, 43.5% of the populations in these counties were of Hispanic descent. Recruitment of participants occurred from 1983-1988 and all participants were between 20 – 74 years of age at enrollment. The study region is the northern most extend of the Spanish Empire in the North America in 16th and 17th centuries and was also settled in the early 19th century due to land grants from the Mexican government during that time. After this early settlement there has been little in-migration and at the time of recruitment 2.8% of the Hispanic population reported being born in Mexico.

Intensive follow-up of vital status was conducted on this cohort in 1998, 2002, 2007 and 2010. This follow-up consisted of telephone contact, death certificate matching and tracking, posted obituaries, Coroners’ reports, autopsy results and medical records were obtained for 98% of those deceased. International Classification of Disease (ICD9 if death occurred before 1999 or ICD10 beginning in 1999) codes (primary and secondary) were recorded for each death. Overall, the death rate in the cohort was 39%. A further mortality follow-up was performed in August of 2007 in a sub-cohort of 480 and the
most recent mortality follow-up was completed in 2010 using a search of all remaining participant social security numbers within the state of Colorado.

A broad definition of COPD mortality was used that included all primary and secondary causes of death using International Classification of Disease (ICD9 or ICD10) codes; 490-496 (ICD-9), J40-J47 (ICD-10). ICD codes were stripped to 3 digits identifying these broad mortality classifications; 490: bronchitis NOS, 491: chronic bronchitis, 492: emphysemas, 493: asthma, 494: bronchiectasis, 495: extrinsic allergic alveolitis, 496: COPD. Any occurrence (primary or any of the secondary codes) of ICD code 496 resulted in a classification of COPD. A primary code of asthma with a secondary code for COPD, emphysemas, bronchiectasis or extrinsic allergic alveolitis was included as COPD and asthma with any other codes were identified as not having COPD. Primary bronchiectasis or primary extrinsic allergic alveolitis were included as COPD. Primary bronchitis or chronic bronchitis with a secondary code suggesting COPD was included as COPD.

Current and former smoking status and behavior were assessed from the participants baseline visit using a standard set of questions that recorded whether the participant had smoked more than 100 cigarettes over their life, whether they currently smoked, the age of initiation and the age at termination of smoking. A limited set of subjects provided data on whether the smoke was inhaled and, if so, how often inhalation occurred, how deeply the smoke was inhaled and how much of the cigarette was typically smoked. It was not possible to include these variables in multivariate models due to the study instruments skip pattern. Pack years were calculated using number of cigarettes smoked per day multiplied by the number of years smoked divided by 20 cigarettes per
pack and each of the component variables was available for analysis as continuous variables. Emphysema and chronic bronchitis was assessed using self-reported occurrence of a doctor’s diagnosis and the age of that diagnosis.

All demographic and univariate analyses were performed using SAS 9.2 (SAS Institute Inc, Cary NC, USA). Normally distributed continuous variables are expressed as mean and standard deviation and were compared using t-tests. Categorical variables are expressed as counts and percentages and were compared using a chi square test statistic. All p values are 2 tailed at a level of p<0.05.

Because of the strong, positive association between smoking and COPD, as well as between smoking and CVD, competing risks must be considered when assessing COPD mortality. To estimate the hazard of the subdistribution of cumulative incidence function, Fine and Gray’s approach was used and its variance estimated using the Delta method. Their approach allows for a person who experienced a competing risk to continue to contribute to the test statistic. The subdistribution hazard is defined as the hazard of failing from a given cause in the presence of competing events, given that a subject has survived or has already failed due to different causes. The Cumulative Incidence Functions were compared by smoking status, ethnicity and a combination of both by directly comparing the area between the curves of each CIF weighted for the occurrence of the competing risk as described by Pepe and Mori.

Competing risks regression was used to calculate the hazard for the risk of COPD mortality accounting for the competing risk of CVD mortality controlling for covariates. The relative risk reported in these analyses is the ratio of subdistribution hazards for the group of interest with respect to the baseline group, holding all other covariates equal.
We report the risk for Hispanic ethnicity compared to NHW ethnicity and its 95% confidence intervals controlled for age, gender, hypertension. We then modeled smoking as smoking status (Current Smoker or Former Smoker compared to Never Smoker), ATS packyears of smoking as a continuous variable and its components; years smoked and packs per day and finally for the combination of smoking status and the components of pack years.

Results

A total of 1887 participants were included in this analysis. They were divided between Hispanic and NHW (49% vs 51%) and by gender (47.5% Male Hispanic vs 52.8% Male NHW, p=0.1) and did not differ by age (54 vs 54.3, p=0.6). Hispanic participants had lower BMI (26.5 verses 27.2, p=0.0007) and were more likely to be hypertensive (41.3% vs 36.2%, p=0.025). Hispanic participants were more likely to be married, less likely to have completed greater than 12 years of school, less likely to be currently working and more likely to earn < $20,000/year (p<0.0001 in all cases). Smoking status was distributed differently by ethnicity (p<0.0001), where Hispanics were more likely to be current smokers (29% vs 20%) and less likely to be never smokers (41% vs 49%) as shown in Table 10.

Smoking behavior was compared between Hispanic and NHW participants in current and former smokers. The age of initiation of smoking was similar between Hispanic and NHW current smokers, as well as for the same comparison in former smokers. Hispanic former smokers stopped smoking at a later age (42% vs 39%, p=0.005). Hispanic current and former smokers accumulated less pack years (20 and 19 vs 33 and 24, p<0.0001 and p=0.004 respectively), and these pack years were
Table 10: Demographics and Other Study Population Characteristics by Ethnicity

<table>
<thead>
<tr>
<th></th>
<th>Hispanic N=929</th>
<th>NHW N=958</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years, SD)</strong></td>
<td>54.0</td>
<td>54.3</td>
<td>0.64</td>
</tr>
<tr>
<td><strong>Sex (n,%)</strong></td>
<td></td>
<td></td>
<td>0.104</td>
</tr>
<tr>
<td>Male</td>
<td>410</td>
<td>458</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>521</td>
<td>501</td>
<td></td>
</tr>
<tr>
<td><strong>BMI (kg/m², SD)</strong></td>
<td>26.5</td>
<td>27.2</td>
<td>0.0007</td>
</tr>
<tr>
<td><strong>Hypertension (n,%)</strong></td>
<td></td>
<td></td>
<td>0.025</td>
</tr>
<tr>
<td>Hypertension</td>
<td>355</td>
<td>336</td>
<td></td>
</tr>
<tr>
<td>No Hypertension</td>
<td>504</td>
<td>593</td>
<td></td>
</tr>
<tr>
<td><strong>Marital Status (n,%)</strong></td>
<td></td>
<td></td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Never Married</td>
<td>45</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>680</td>
<td>791</td>
<td></td>
</tr>
<tr>
<td>Divorced</td>
<td>67</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>Widowed</td>
<td>93</td>
<td>73</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>41</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td><strong>Education Completed (n,%) (years)</strong></td>
<td></td>
<td></td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>&lt;8</td>
<td>209</td>
<td>17</td>
<td>1.8</td>
</tr>
<tr>
<td>8 - 11</td>
<td>284</td>
<td>124</td>
<td>12.9</td>
</tr>
<tr>
<td>12</td>
<td>253</td>
<td>373</td>
<td>38.9</td>
</tr>
<tr>
<td>13+</td>
<td>185</td>
<td>445</td>
<td>46.4</td>
</tr>
<tr>
<td><strong>Working Status (n,%)</strong></td>
<td></td>
<td></td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Not Working</td>
<td>534</td>
<td>382</td>
<td>39.9</td>
</tr>
<tr>
<td>Working</td>
<td>393</td>
<td>576</td>
<td>60.1</td>
</tr>
<tr>
<td><strong>Income (n,%) ($/year)</strong></td>
<td></td>
<td></td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>&lt;20,000</td>
<td>697</td>
<td>414</td>
<td>43.2</td>
</tr>
<tr>
<td>20,000 - 50,000</td>
<td>200</td>
<td>447</td>
<td>46.6</td>
</tr>
<tr>
<td>&gt;50,000</td>
<td>34</td>
<td>98</td>
<td>10.2</td>
</tr>
<tr>
<td><strong>Cigarette Smoking (n,%)</strong></td>
<td></td>
<td></td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Current Smoker</td>
<td>267</td>
<td>193</td>
<td>20.2</td>
</tr>
<tr>
<td>ex-Smoker</td>
<td>277</td>
<td>298</td>
<td>31.1</td>
</tr>
<tr>
<td>Never Smoker</td>
<td>385</td>
<td>467</td>
<td>48.8</td>
</tr>
</tbody>
</table>

Table 10 compares the demographics of Hispanics and non-Hispanic Whites.
accumulated differently compared to NHW current and former smokers, with Hispanic participants smoking fewer packs per day (0.6 and 0.7 vs 1.0 and 1.0, p <0.0001 for both comparisons). Current Hispanic and NHW smokers did not differ in years smoked

Smoking behavior was compared between Hispanic and NHW participants in current and former smokers. The age of initiation of smoking was similar between Hispanic and NHW current smokers, as well as for the same comparison in former smokers. Hispanic former smokers stopped smoking at a later age (42% vs 39%, p=0.005). Hispanic current and former smokers accumulated less pack years (20 and 19 vs 33 and 24, p<0.0001 and p=0.004 respectively), and these pack years were accumulated differently compared to NHW current and former smokers, with Hispanic participants smoking fewer packs per day (0.6 and 0.7 vs 1.0 and 1.0, p <0.0001 for both comparisons). Current Hispanic and NHW smokers did not differ in years smoked (p=0.6), however Hispanic former smokers accumulated more years smoked compared to NHW former smokers (22 vs 20, p=0.047). Thus the decreased pack years in Hispanic participants does not fully describe the smoking exposure measured by pack years.

Hispanic former smokers were less likely to report inhaling when they smoked (76% vs 83%, p=0.04) and of those reporting inhaling, current Hispanic smokers were less likely to inhale every puff than current NHW smokers (54% vs 74%, overall chi-square comparison for the three categories, p= <0.0001). Health outcomes related to smoking were also different by ethnicity. The percent reporting chronic bronchitis was similar between ethnicities, however the age at diagnosis of chronic bronchitis for Hispanic former smokers was significantly older compared to NHW (67 vs 33, p=0.004). The percent reporting emphysema was lower in Hispanic current smokers compared to NWH
Table 11: Smoking Related Variables

<table>
<thead>
<tr>
<th></th>
<th>Hispanic</th>
<th>NHW</th>
<th>p (cur HvNHS)</th>
<th>p (for HvNHW)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Current n=267</td>
<td>Former n=277</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Start Age (years ± SD)</td>
<td>21.1 ± 9.9</td>
<td>19.8 ± 8.0</td>
<td>0.088</td>
<td>0.073</td>
</tr>
<tr>
<td>Stop Age (former only) (years ± SD)</td>
<td>NA</td>
<td>42.1 ± 14.3</td>
<td></td>
<td>0.005</td>
</tr>
<tr>
<td>Pack Years (packs/year, SE)</td>
<td>19.6, 1.4</td>
<td>18.6, 1.4</td>
<td>&lt;0.0001</td>
<td>0.004</td>
</tr>
<tr>
<td>Years Smoked (years, SE)</td>
<td>31.7, 0.9</td>
<td>22.3, 0.9</td>
<td>0.59</td>
<td>0.047</td>
</tr>
<tr>
<td>Packs per Day (packs, SE)</td>
<td>0.6, 0.04</td>
<td>0.7, 0.04</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Inhaled when smoke (%Y)</td>
<td>82.1%</td>
<td>76.4%</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>Inhale puffs every/few/v-few (%)</td>
<td>54/27/18</td>
<td>65/26/9</td>
<td>&lt;0.0001</td>
<td>0.088</td>
</tr>
<tr>
<td>Chronic Bronchitis n(%)</td>
<td>13 (2.8)</td>
<td>9 (1.6)</td>
<td>0.17</td>
<td>0.95</td>
</tr>
<tr>
<td>Chronic Bronchitis age</td>
<td>51.6, 7.7</td>
<td>67.4, 8.7</td>
<td>0.09</td>
<td>0.004</td>
</tr>
<tr>
<td>Emphysema n(%)</td>
<td>4 (0.9)</td>
<td>9 (1.6)</td>
<td>0.01</td>
<td>0.72</td>
</tr>
<tr>
<td>Emphysema age</td>
<td>42.8, 4.7</td>
<td>63.0, 3.5</td>
<td>0.14</td>
<td>0.11</td>
</tr>
</tbody>
</table>

Table 11 compares smoking behavior variables in Hispanics to Non-Hispanic Whites by smoking status (Current Smokers verses Former Smokers). “cur HvNHS” compares Current Hispanic Smokers to Current Non-Hispanic White Smokers, “for HvNHW” compares Former Hispanic Smokers to Former Non-Hispanic White Smokers.
current smokers (1% vs 3%, p=0.01) however the age of diagnosis of emphysema did not
differ as shown in Table 11.

Figures 9a through 9c show the cumulative incidence functions for COPD
mortality over the course of the follow-up period, accounting for the effect of the
competing risk of CVD mortality, by smoking status, ethnicity, and the combined effect
of smoking and ethnicity.

As expected, current and former smokers have significantly higher COPD
mortality compared to never smokers (Cumulative Incidence of 7.9/100, 4.0/100, and
0.9/100, current, former, never smokers respectively, p<0.0001 current vs never,
p=0.0004 former vs never) (Figure 9a). Current and former smokers had similar COPD
mortality rates (p=0.224). Overall Hispanic participants had marginally lower cumulative
risk of COPD compared to NHW participants after accounting for the competing risk of
CVD mortality without considering smoking. Overall, ethnicity has no significant effect
on COPD mortality (Cumulative Incidence of 3.1/100 vs 3.9/100, Hispanic and NHW
respectively, p=0.071) (Figure 9b). Combining ethnicity and smoking after accounting
for the competing risk of CVD mortality shows that Hispanics who are current smokers
have significantly lower cumulative incidence of COPD mortality compared to NHW
(Cumulative Incidence of 6.1/100 vs 10/100, Hispanic and NHW respectively, p=0.037).
Among former and never smokers, there were no statistically significant differences in
cumulative incidence of COPD mortality between Hispanics and NHWs (Figure 9c).

Competing risk regression models predicting COPD mortality accounting for the
effect of CVD mortality are reported in Figure 10. Each row of Figure 10 shows the
Figure 9: Cumulative Incidence Function by Smoking. Figure 9 shows the Cumulative Incidence Function (CIF) for COPD mortality accounting for CVD mortality by smoking category. Current Smokers are plotted in black, Former Smokers are plotted in purple and Never Smokers are plotted in blue. Current Smokers are at increased risk for COPD mortality compared to Never Smokers (p=0.0001) but they have a similar CIF compared to Former Smokers (p=0.224). Former Smokers are at increased risk for COPD mortality compared to Never Smokers (0.00041).
Figure 10: Cumulative Incidence Function by Ethnicity. Figure 10 shows the Cumulative Incidence Function (CIF) for COPD mortality accounting for CVD mortality by ethnicity. Hispanic ethnicity is plotted in black and Non-Hispanic White ethnicity is plotted in Purple. Non-Hispanic White participants have a similar CIF for COPD mortality compared to Hispanic participants (p=0.071).
Figure 11: Cumulative Incidence Function by smoking and ethnicity. Figure 11 shows the Cumulative Incidence Function (CIF) for COPD mortality accounting for CVD mortality by smoking category and ethnicity. Solid lines indicate Hispanic ethnicity and dotted lines indicate Non-Hispanic White ethnicity. Hispanic Current Smokers are plotted in black, Hispanic Former Smokers are plotted in purple and Hispanic Never Smokers are plotted in glue. NHW Current Smokers are plotted in red, NWH Former Smokers are plotted in orange and NHW Never Smokers are plotted in yellow. NHW Current Smokers are at increased risk for COPD mortality compared to Hispanic Current Smokers (p<0.037). Hispanic Former Smokers have a similar CIF compared to NHW Former Smokers (p=0.392) and Hispanic Never Smokers are also similar to NHW Never Smokers (p=0.267).
Figure 12: Multivariable Modeling of COPD Related Mortality. Figure 12 shows the relative risk (sub-distribution hazards ratio) for Hispanic participants compared to Non-Hispanic Whites for each of five distinct competing risks regression models of COPD mortality accounting for Cardiovascular Disease mortality. Each model is controlled for age, gender and hypertension with smoking assessed using different measures.
relative risk and 95% confidence intervals for COPD mortality in Hispanics compared to NHW and all models are controlled for the effects of age, gender and hypertension. Model 1 (Not Smoking Controlled) provides a baseline model that is not controlled for smoking and each subsequent row shows smoking controlled for in a different way. Without controlling for smoking Hispanic ethnicity is not associated with COPD mortality (RR = 0.79, 95% C.I. 0.47-1.34, p=0.38). However Hispanic ethnicity is significantly protective for COPD mortality after controlling for current and former smoking status compared to never smokers (RR=0.58, 95% C.I. 0.34-0.99, p=0.047). Models controlled for smoking using pack years as a continuous variable Hispanic ethnicity is not significantly related to COPD mortality (RR = 0.92, 95% C.I. 0.54-1.56, p=0.7) or controlled for the individual elements used to calculate pack years, average packs per day (Ave Packs/Day; RR = 0.90, 95% C.I. 0.53-1.52, p=0.7) or years of smoking (Years of Smoking; RR= 0.61, 95% C.I. 0.35-1.1, p=0.07). As smoking status and years of smoking had the largest effect on the association between Hispanic ethnicity and COPD mortality, further models were run to test the effects of smoking status and accumulated smoking exposure. Controlling for smoking status and pack years of smoking (p=0.3) or its elements (Ave Packs/Day, p=0.2 and Years of Smoking, p=0.06) did not affect the significance of the association between Hispanic ethnicity and COPD mortality.

Discussion

Hispanics have lower COPD mortality compared to NHW after accounting for smoking status. Measures of smoking were assessed in this population and a significant protective effect of ethnicity was found only when controlling for smoking status, but that
effect was mitigated when controlling for differences in smoking behavior as measured by packyears, average packs per day smoked or years of smoking exposure. We found significant differences between Hispanic and NWH smokers that are associated with the intensity of the smoking experience; Hispanic ethnicity was associated with less intense smoking where smokers reported inhaling smoke at every puff less frequently. This difference in mortality may be mediated by differences between Hispanic and NHW smoking behaviors. Our study identified a significant protective effect of Hispanic ethnicity using competing risks regression that accounts for cardiovascular mortality. Our study also examined smoking behavior at a higher granularity than is typically reported, EG measuring smoking intensity, depth of inhalation and regularity of inhalation rather than reporting smoking status and pack years only.

We also observed that Hispanic former smokers who report having chronic bronchitis do so at a significantly older age compared to NHW former smokers. We speculate that this may be related to access to healthcare and the establishment of a subsequent diagnosis in Hispanics at an older age given their lower SES level. We also observe that several participants who quit smoking also reported physician diagnosed emphysema and chronic of those reporting emphysema, 70% reported quitting smoking within 3 years of that diagnosis and 50% reported quitting smoking within 3 years of a diagnosis of chronic bronchitis. The proximity of quitting smoking to a diagnosis may suggest that those participants identified as former smokers were experiencing more severe disease compared to people identified as current smokers.

Smoking is a complex behavior and may not be sufficiently described using smoking status or overall smoking exposure. Some aspects of this behavior that differ by
ethnicity are clear in this study; while overall exposure (pack years) is significantly lower in Hispanic participants, they were also exposed to less smoke due to a different, less intense pattern of exposure. They inhale less often and when they did inhale they inhale every puff less often. This would indicate that positive smoking status measures a less intense exposure in Hispanics compared to NHW. It also indicates that pack years, while lower in Hispanics, is lower than measured due to the pack year accumulation measuring less total exposure due to inhalation. Taken together, these findings suggests that the risk of COPD mortality in Hispanic compared to NHW subjects is confounded by these commonly unmeasured aspects of smoking behavior and that this might bias results in a protective direction. This bias may explain the protective effect of Hispanic ethnicity reported in other studies.

Other studies have found that Hispanics compared to NHWs report more current smoking, similar numbers of packyears, less packs per day and more years of smoking. In our study, Hispanic ethnicity is a risk factor for increased current smoking but decreased pack years due to decreased packs per day smoked over a similar number of years of smoking in current smokers and an increased number of years of smoking on former smokers. Consistent with these findings, Samet et al. compared COPD mortality rates between Hispanic and Non-Hispanic Whites in New Mexico and found that Hispanics had lower rates of COPD mortality between 1958 and 1982\(^{36}\), again suggesting a protective effect of Hispanic ethnicity. Bruse et al, also working in New Mexico, reported similar findings and identified Native American ancestry, measured using ancestrally informative markers (AIMs), as being independently protective for COPD and rapid decline in FEV\(_1\).\(^{83}\) They also observed that Hispanic smokers accumulated similar
numbers of pack years to NHWs but that differences were noted for an increased number of years of smoking with decreased packs smoked per day for Hispanics. Our study observed similar differences between Hispanic and NHW participants as they relate to smoking and applied those differences to COPD mortality. We are unable to assess genetic differences within Hispanic ethnicity at this time but Bruse’s observation that Native American ancestry was protective against COPD and pulmonary function decline was almost completely captured by self identified Hispanic ethnicity in that study. We observe a longer (more pack years) and less intense smoking exposure (less packs per day) in Hispanics in the San Luis Valley, which has a similar history of settlement and migration to the study in New Mexico. In our case, the duration of smoking appears to explain the protective effect of Hispanic ethnicity adjusting for smoking status. These similarities between Hispanic smokers in New Mexico and Colorado suggest that our findings may applicable to Hispanic smokers in New Mexico.

Our observation that the Hispanic population had lower income, less current employment and lower levels of high school education and decreased COPD mortality do not agree with other reports. Lipton et al. report that Hispanic ethnicity (assessed by census report) was an important positive predictor of spatial “hot spots” of increased hospitalization charges related to COPD in California using 2 years of cross sectional data. This study also identified low income and related variables as predictors of increased hospitalizations and high income and related variables as predictors of decreased hospitalization. Their work suggests that the environment as it relates to poverty may drive their findings due to increased poverty in areas with high Hispanic populations. Our work does not suggest that poverty is protective for COPD mortality,
rather that the potential protective effect of Hispanic ethnicity was not detectible in the Lipton study due to unmeasured factors such as smoking behavior. We did not assess healthcare utilization and associated costs in our study. Poverty may play a role in the cost of COPD in our population due to the availability of public health services in our study area. Accessing healthcare at a later stage of the disease when COPD generates greater costs may in fact be occurring in our population, but does not manifest as increased mortality. If Hispanic smokers in this region have access to public health services but access them later this would place unnecessary financial burden on this system. Treatments exist that slow or halt COPD progression once they are begun but treatment has not been shown to reverse COPD. Patients with more restrictive disease are more prone to exacerbations requiring more intense treatment and potential emergency room visits both of which are dangerous for the patient and increase costs. Improved secondary prevention of COPD progression and prevention of exacerbations improves the quality of life of the patient, reduces risk of death and decreases the financial burden of the disease.

This study has several strengths and limitations. Limitations include: 1) death certificate data were used for cause specific mortality rather than hospital records, however we considered any mention of COPD on the death certificate, not just primary cause of death. 2) This study used common questions related to smoking behavior but did not apply them to all participants for whom they may have been applicable. Future studies of smoking related disease should include more detailed surveys that capture current and past smoking behaviors. 3) Addiction to smoking was not assessed in this study, nor were known nicotine addiction genes. If addiction occurs at a different rate in
Hispanics compared to NHWs this could partially explain differences in smoking behavior between the groups. Pérez-Stable et al report that nicotine is not metabolized at a different rate in Hispanics compared to NHWs. 4) Ethnicity was self reported. Parra et al report average admixture for the Hispanic population of this study as 65% European, 35% Native American and less than 1% African and that the average time to an unadmixed ancestors is greater than seven generations. 5) Socioeconomic factors such as access to care were not evaluated in this study which limits our ability to discuss this important aspect of COPD mortality.

Strengths of this study include: 1) we used competing risks regression to adjust for the most common smoking related mortality risk, cardiovascular disease. Lung cancer did not play a large role in this study and the 17 decedents with report of lung cancer were evenly distributed between Hispanics and NHWs. This suggests that lung cancer is not acting as a competing risk that differs by ethnicity. 2) This study was conducted in a well characterized, population based sample of a region of the country that was 46.6% Hispanic at the time of baseline data collection. This reduces the effect of recall bias and allows us to explore the effects of current smoking behavior over time. 3) This study has very complete, long term (35,789 person years) follow-up of its participants and death certificates or coroners reports were obtained for 98% of decedent participants. 4) A thorough smoking assessment was performed that describes smoking behavior beyond smoking status and pack years.

The Hispanic population of the United States is increasing; between 2000 and 2010 U.S. Census reports an increase by 35.3 million to 50.5 million Hispanic people (13% to 16% of the total population of the US) and differences in specific disease
etiology would play in increasingly important role in the public health impact of smoking related disease. Hispanic ethnicity represents a diverse population in terms of ancestry, culture and behavior. In some cases this may result in true protection from COPD or COPD mortality, but this protective effect may be explained by smoking behavior which is not typically assessed on a population basis. Care should be taken in evaluating studies that describe Hispanic ethnicity as a factor related to COPD mortality as a perceived protective effect of ethnicity may result in decreased emphasis on smoking prevention and early disease detection in this large population due to a perceived decreased risk.

Conclusions

The low COPD mortality seen in Hispanic smokers may be due to smoking behavior. Hispanic smokers inhale less, possibly resulting in lower cumulative exposure to tobacco smoke. Thus, smoking behavior may play a key role in explaining differences in COPD mortality as they relate to Hispanic ethnicity.
CHAPTER VII

DISCUSSION AND FUTURE DIRECTIONS

Chronic Obstructive Pulmonary Disease (COPD) is a progressive disease that is not reversible and is primarily the result of smoking. COPD mortality has been shown to be increased in regions of the country that have an aging population, high current and former smoking rates, more rural population and report more poverty. Increased COPD mortality has also been shown to be associated with occupational and environmental exposures. This dissertation reports the geographic distribution of COPD related mortality in the state of Colorado where significant statistical clustering of crude and age adjusted COPD mortality occurred in the state. Controlling for the effects of age and smoking history explained regional clustering of increased mortality in all but one small region of the state. This study was interested in regional effects occurring in more than one contiguous ZCTA and not local effects where a single ZCTA was different than all of its neighbors. Because all regional clustering was explained by age and smoking history, the resulting cluster was the final model of clustering in Colorado. As expected, clusters differed demographically and these differences are likely due to local affects that influence COPD mortality.

The state of Colorado has a large Hispanic population and the association between Hispanic ethnicity and COPD is unclear, thus we investigated Hispanic ethnicity as a risk factor in COPD mortality. We examined longitudinal mortality in the San Luis Valley region of the state where we found that Hispanic ethnicity was significantly protective for COPD mortality when controlling for the effects of current and former smoking status. Further investigation of smoking behavior in Hispanic participants revealed that their
smoking behavior differs in significant ways that might indicate a less intense smoking experience. Hispanic participants experienced less pack years compared to NHW participants due to less packs per day smoked over an equivalent number of years smoked. Hispanic participants also report inhaling less often and inhaling less smoke at each inhalation.

The first goal of this dissertation was to examine the relationship between COPD mortality regional age structure, current and past smoking, environmental exposures and ethnicity in Colorado. It is important in studies of disease that involve geographic location to consider the aggregate unit of study carefully and select one that is reasonable for the pathway under investigation. Previous work I co-authored from Colorado showed increased COPD mortality at the county level clustered in rural and frontier regions of the State. Age structure and smoking prevalence are not likely to vary only at the country level so choosing a smaller aggregate such as ZCTA is reasonable. Examination of crude rates of COPD mortality at this smaller aggregate show similar geographic location of increased rates suggesting that the ZCTA aggregate captures similar information as the county aggregate in this case.

Controlling for the effects of age structure and smoking prevalence explained the statistical clustering of increased COPD mortality observed in the state leaving a single, small cluster in the northern region of the state. This indicates that other factors such as environmental exposure to soil arsenic or Hispanic ethnicity operate at a local level and understanding their effects will require a different study design to investigate. In our study soil arsenic was measured at the point level and these values could be averaged across the study aggregate to obtain an average exposure measure. It is likely, given our
results that the effects of soil exposure on COPD mortality do not extend across more than one ZCTA aggregate however. This indicates that the average value for an aggregate is not significantly associated with regional clustering of increased mortality. A more appropriate approach to this question would be to identify aggregates with locally increased rates and model the spatial features of COPD mortality within the smaller aggregate. This would require COPD mortality counts to be obtained at a smaller aggregate than was available at the time of this study but would allow local variation in soil arsenic to be modeled as it relates to local variation in COPD mortality. Caution should be used when moving an analysis from one areal unit to a smaller one as this increases the probability of creating a modifiable areal unit problem if the unit of analysis is not appropriate to the question being asked as we discussed previously.

Similarly regional effects of ethnicity do not appear to be associated with spatial clustering of COPD mortality in Colorado. Local variation in the distribution of the Hispanic population is evident in our findings describing the clusters and this could be important to explore with appropriate study designs. The final cluster remaining after adjustment for age and smoking is in a population with a median Hispanic ethnicity of 40.5% compared to a median 7.5% for the state overall indicating that Hispanic ethnicity may play an important local role in COPD mortality. This finding of increased COPD mortality in a region with a large Hispanic population appears to contradict this study’s findings in the SLVDS cohort of no association between ethnicity and COPD mortality. There are several possible explanations for this apparent inconsistency, immigration patterns are likely different between the two regions bringing with them different lifetime
exposures, risks and behaviors also environmental and occupational exposures are likely
different between the two regions.

These findings suggest directions for future studies that account for weakness in
ecological studies. This study identified statistically significant spatial clustering at a
regional level but was not designed to assess local effects that might increase COPD
mortality at a smaller aggregate. Regional clustering of aging populations and
populations with increased smoking exposure over their lives are potential targets for
public health and would benefit from increased awareness and programming attention
that focuses on existing COPD. This population is more likely to access health resources
including emergency room visits due to their disease and public health programming that
reduces exacerbations of COPD may have an impact on both limited public health
resources and deaths due to COPD.

Local effects that cause increased COPD mortality might include occupational
exposures that are specific to a small geographic region such as mining. Mining has been
carried out in Colorado for generations and these mines tend to be in hard rock in
mountainous areas due to the geology of the region. Hard rock mining primarily expose
workers engaged in direct mining operations and secondarily might exposed others due to
the disposal of tailings or expressed ground water. A study designed to assess this type of
local exposure might include an extensive employment history and a control group to
assess risk.

Other local effects such as feed lot operations or geologic soil composition may
be best understood using a spatial statistical approach. This approach may identify an
area proximal to an exposure area such as a feed lot or loess with components of interest
and establish a point location and outdoor exposure to dust arising from the operation for people living there. This approach is capable of accounting for environmental effects such as wind direction in order to assess whether people are more or less likely to be exposed and then identify clusters of people with increased COPD mortality accounting for the effects of spatial autocorrelation.

Ecological studies by their nature do not emphasize personal level data and mortality from COPD has been clearly shown to be influenced by personal variability. Pulmonary function as measured by Pulmonary Function Tests (PFTs) include Forced Expiratory Volume at 1 second (FEV$_1$) and Forced Vital Capacity (FVC) which are invaluable for assessing the severity and progression of COPD in an individual. While longitudinal measures are logistically impossible on a population level, measuring pulmonary function on a spatially appropriate sample of people could reveal regional effects that lead to future disease and disease mortality. Prospectively collecting PFT results in an appropriately designed study would also allow the assessment of incident COPD.

The public health community typically engages in primary prevention of disease as avoiding the occurrence of disease is preferable and more cost effective when compared to treating existing disease even at the early stages. In the case of COPD, primary prevention strategies have significantly decreased incident smoking and increased smoking cessation. Secondary prevention in COPD includes identifying people with sub-clinical disease and treating them before their pulmonary function becomes impaired to the point that they seek medical care. This effort would benefit from population studies of pulmonary function as COPD has been shown to exist in a
significant number of never smokers. Spatial clustering of reduced pulmonary function or sub-clinical COPD could help investigators identify non-smoking exposures that predispose individuals to COPD and its complications.

A hypothesis of this dissertation was that population exposure to geologic formations of arsenic containing soil would be associated with clustering of increased COPD mortality. This hypothesis was not born out at a regional level. The proposed effect of Arsenic exposure is that Arsenic containing airborne particulate matter exposure results in reduced lung function. Reduced lung function has been shown after acute airborne and drinking water based exposure and would result in more severe COPD symptoms. If these pulmonary effects manifest as increased COPD severity but not increased COPD mortality then we would expect them to manifest in measures such as hospitalizations for COPD exacerbations. We observe differences in measures of COPD related hospitalization in this study where admission count, cost and cost per day for COPD related hospitalization are different in different clusters of COPD mortality in the state. While this may represent differences in the cost of delivering health care in these areas it may also be an indication that regions experience differences in COPD severity. Studies that assessed personal exposure to As and its origins would be invaluable in pursuing the underlying cause of this observation. Methods exist for measuring inorganic As exposure in humans that use samples such as urine, serum or expressed breath condensate that could be collected in a population. Measuring personal exposure would greatly increase our ability to assess environmental exposures that result in local but not regional clustering of COPD mortality.
This study showed significant statistical clustering of COPD mortality at the borders of the state of Colorado. This restriction in data availability created artificial boundaries that may have influenced the analysis of clustering when it occurred in those regions. This effect can be clearly seen in the maps of crude clustering in the two large clusters at the north and south of the eastern border of the state. If an exposure is regional and operates to increase COPD mortality that exposure may not be effected by political boundaries, for example soil As due to geological formations would depend only on the geology, not on the location of the border with Kansas. Future studies of regional exposures would benefit from including data that crosses state borders. Measuring the outcome of interest identically in both states would be crucial in this case and care would have to be taken to evaluate procedural effects that could create artificial, state specific clustering. The US Census allows us to trust the comparability of denominator values for this type of analysis but differences in mortality coding, use of a medical examiner verses a coroner and data availability at comparable aggregates would require careful thought. Including data that is agnostic to state boundaries would allow an assessment of regional exposures that extend over those borders however and may show important associations that cannot be examined here due to these data limitations.

This study uses 18 years of data but did not consider longitudinal changes in demographics or exposures of interest. This was done due to constraints on reporting small sample size and Colorado has many frontier regions with small populations necessitating the aggregation of many years of outcomes. These analyses were also constrained by the association between smoking and COPD as smoking was well assessed in only one statewide study collected at only one time point. These longitudinal
data could be assessed for temporal trends across this study period and include control for the changing age structure of the state. Smoking incidence and prevalence have been declining for many years and these trends are known. Longitudinal changes in COPD mortality could be examined accounting for demographic changes and population changes in smoking and these trends would be of interest to the public health community as they may help in the assessment of the impact of public health programs aimed at decreasing COPD mortality.

The second goal of the study was to further explore Hispanic ethnicity as it relates to COPD mortality in a defined geographic region of the state in a well-described cohort with longitudinal mortality collection. We found that in this population the association between Hispanic ethnicity and death from COPD is partially explained by smoking behavior. It is possible that studies reporting a protective effect of Hispanic ethnicity for COPD mortality are missing a behavioral aspect of smoking that influences that association. In our cohort, Hispanic participants report more current smoking and less never smoking but significantly less pack years for current or former smokers. Additionally Hispanic participants accumulate these pack years differently compared to NHW participants in that they report equivalent years of smoking but less packs per day. Lastly Hispanic participants report a less intense smoking experience in that they inhale smoke from puffs less frequently than NHW current or former smokers. Taken together these differences in smoking behavior will drive an analysis of the effect of Hispanic ethnicity that only controls for smoking status towards a significant protective effect of ethnicity. Further, differences in the intensity of exposure likely remain after controlling for smoking status and pack years and in our study the only model that shows Hispanic
ethnicity to be significantly protective for COPD mortality was one that controlled for smoking status only. We were unable to directly control for specific smoking behaviors such as how many puffs the smoker typically inhaled and how much of the cigarette was allowed to burn in the ashtray as these variables were only capture for a subset of participants.

This analysis suggests several interesting directions for pursuing causes of COPD mortality. This study identified a significant impact of smoking behavior on COPD mortality and, while this aspect of smoking was measured, future studies could benefit from collecting more extensive smoking behavior and smoking behavior history. While smoking intensity (measured by packs per day) and smoking duration (measured by years of smoking) are known predictors of smoking related disease, this study indicates that other aspects of the smoking experience should also be considered. Smoking is not a homogeneous behavior; all cigarettes are not consumed entirely, all breaths that include smoke are not drawn to the same depth and all smoke is not held in the lungs for the same duration. Variables that measure these aspects of smoking are likely important drivers of the observed associations between smoking and disease. Further; there are likely cultural aspects of smoking that play an important role in smoking as an exposure. Other markers of smoking behavior that may play a role that varies due to cultural factors are exposure to second hand smoke and support for quitting smoking. In this study we show differences in smoking behavior that are ethnicity specific and failing to account for them muddles the association between ethnicity and COPD mortality. There reason to believe that other cultures perform the act of smoking in ways that alter its effect on an outcome and failure to consider this can result an inaccurate conclusion.
Identifying Hispanic ethnicity as a predictor of different smoking behaviors suggests that dependency and behavioral adaptation to smoking might play a role in altered behavior. Polymorphisms in the nicotinic acetylcholine receptor (nAChR) gene cluster CHRNA5-A3-B4 on chromosome 15 have been identified as playing a significant role in COPD severity in recent studies though their occurrence in Hispanics has not been assessed. Genetic changes in this region are associated with nicotine dependence, increased quantity of smoking, age at onset of smoking, cotanine levels in smokers and increased pleasure experienced with first cigarette. This behavioral heterogeneity could lead to the observed differences if the relevant polymorphisms occur with different frequency by ethnicity.

Competing risks for mortality play an important role in smoking related disease and we observed far more mortality related to cardiovascular disease (CVD) in this population than mortality related to COPD. CVD is primarily diagnosed at autopsy and COPD is primarily diagnosed in a patient between 50 and 60 years of age presenting to a clinician with a complaint of shortness of breath. It is likely that a participant surviving sub-clinical CVD until a diagnosis of COPD is made will also be assessed for the CVD risk at that point and treated. Lung cancer is also an outcome in smokers and it is typically diagnosed after they have spread to the point that symptoms are present and treatment is unlikely to be effective. Each of these smoking related diseases competes with the others in a smoker and it is unlikely that they act identically as competing risks in a patient. The current analytical model for competing risks regression allows risks to compete but it is blind to the censoring mechanism; the sub-distribution hazard function is modeled the same without considering prior knowledge of how the risks compete.
Future work in competing risks regression could include modeling the sub-distribution Hazard function for each competing risk considering their interplay.

Other competing risks include diabetes and this study was originally designed to assess type 2 diabetes prevalence and incidence in a population that allowed comparison by ethnicity. Type 2 diabetes has been shown to be associated with small but significant decreases in pulmonary function and these decreases may play a role in COPD mortality due to an altered disease course of COPD in an individual with the diabetes.

The current study builds on earlier work involving CVD in this cohort. That work collected data on CVD incidence and mortality and adjudicated each event and the current study was not able to update that work to include events after 1998. After that time period events were included based on ICD coding alone. No COPD related events were adjudicated and future work in this cohort could benefit from adjudicating all events of interest.

This study examined COPD mortality and patient report of emphysema and chronic bronchitis as the only measures of COPD. Ideally an assessment of pulmonary function and treatment for COPD would have been used to better understand COPD disease severity. The clinical course of COPD in an individual is an important predictor for mortality and including these measures in future studies is crucial for a better understanding of the life course of the disease. Measures of access to care, emergency room visits and treatment, and COPD related hospitalizations could serve as surrogates when these disease markers were not measured pro-actively.
REFERENCES

(1)  Rennard SI. COPD: overview of definitions, epidemiology, and factors influencing its development. Chest 1998; 113(4 Suppl):235S-241S.


(50) Levinson AH. Adult Survey Tehnical Report. 2011, Personal Communication


(52) BY SENATOR(S) Hagedorn BBGGIKMSSTTWaW, also REPRESENTATIVE(S) McGihon FGKLMRSaT. CONCERNING PUBLIC HEALTH, AND MAKING AN APPROPRIATION IN CONNECTION THEREWITH. 2008.


(64) SAS Institute Inc CNC. Statistical Analysis Software. 2010.


(68) Sullivan SD, Ramsey SD, Lee TA. The economic burden of COPD. Chest 2000; 117(2 Suppl):5S-9S.


APPENDIX A

Poisson Regression Equation

\[ f(k; \lambda) = \frac{\lambda^k e^{-\lambda}}{k!}, \]

Where \( e \) = the natural logarithm, \( k \) = the count of observed events, \( \lambda \) = the count of expected events.
APPENDIX B

SaTScan statistical methodology

\[ \max \left( \frac{Y_{in}}{E_{in}} \right)^{Y_{in}} \times \left( \frac{Y_{out}}{E_{out}} \right)^{Y_{out}} \]

Where;

\( Y_{in} \) = the observed number of events in the regions whose centroid is within the circle

\( E_{in} \) = the expected under the null hypothesis in the regions whose centroid is within the circle

\( Y_{out} \) = the observed number of events outside the regions whose centroid is within the circle

\( E_{out} \) = the expected under the null hypothesis outside the regions whose centroid is within the circle
APPENDIX C

Poisson Regression Methodology

SAS implements Poisson regression using Proc Genmod with /distribution = Poisson, link=log and offset=ZCTA population count.

\[ \log(Y/pop) = \alpha_0 + \alpha_1(\text{lung cancer deaths}) + \alpha_2(\text{ZCTA}) \]

where

- \( Y \) = COPD deaths in each aggregate area
- \( \text{pop} \) = the population at each aggregate level

which shows that the impact of a one unit change in a predictor variable results in a multiple of the mean value.

In SAS:

```
PROC GENMOD;
CLASS=ZCTA[categorical variable capturing aggregate region] ;
PREDICTED CLM;
RUN;
```
APPENDIX D

SLVDS Power Calculation Report Definitions

Power is the probability of rejecting a false null hypothesis. It should be close to one.

N is the size of the sample drawn from the population.

B is the size of the regression coefficient to be detected (in units of log HR)

SD is the standard deviation of X1.

P is the event rate.

R2 is the R-squared achieved when X1 is regressed on the other covariates.

Alpha is the probability of rejecting a true null hypothesis.

Beta is the probability of accepting a false null hypothesis.