PREDICTORS OF HOSPITALIZATION AMONG
CYSTIC FIBROSIS PATIENTS IN ONTARIO.

by

Anne Stephenson

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy, Graduate Department of Health Policy, Management and Evaluation, University of Toronto

Copyright © Anne Stephenson 2010
ABSTRACT

Predictors of hospitalization among cystic fibrosis patients in Ontario.

Doctor of Philosophy, Graduate Department of Health Policy, Management and Evaluation, University of Toronto

Anne Stephenson, 2010

This dissertation involved linking a clinical cystic fibrosis (CF) data registry with administrative databases to evaluate clinical, demographic, and geographical predictors of hospitalization in CF patients living in Ontario over a 10 year period. In addition, this work assessed the ability of administrative data to identify individuals with CF using the clinical registry as the reference standard.

Sex was an independent predictor of hospitalization rates for individuals with CF. Females had a significantly higher hospitalization rate compared to males even after adjusting for important clinical factors suggesting that this finding is not simply due to worse CF disease. In those between 7 and 19 years of age, the adjusted hospitalization rate was 38% higher in females (rate ratio [RR] 1.38, 95% confidence interval [CI] 1.11-1.73). Similarly in those over the age of 19, females had a 30% higher hospitalization rate compared to males (RR 1.30, 95% CI 1.06-1.59). Other significant predictors associated with higher hospitalization rates in both age groups were lower lung function, worse nutritional status, pancreatic insufficiency, and the presence of CF-related diabetes. The presence of *Burkholderia cepacia complex* in the sputum was a significant predictor in those over the age of 19 years (RR 1.54, 95% CI 1.26-1.89). Distance to CF centre, community size and socioeconomic status were not significant predictors of
hospitalization rates in either age group. There was no significant trend in hospitalization rates over time once rates were adjusted for markers of disease severity (p=0.08). Comparing administrative data with the CF registry data, administrative data captured hospitalizations more comprehensively. Despite CF being a specific diagnosis, health administrative databases alone were insufficient to reliably and accurately identify individuals with CF unless they had been hospitalized.

The reason for the gender disparity seen within this dissertation is likely multifactorial. There may be differences in outpatient management between the sexes, hormonal influences may modulate disease severity causing higher hospitalization rates, and patient and provider-level influences may affect the decision to hospitalize a patient. Further research is needed in this area to elucidate the factors contributing to this gender gap.
This dissertation is dedicated to my husband, Claudio, who has always been and continues to be my greatest supporter. His love, words of support and encouragement, and laughter have made this journey fun and this thesis possible. His selfless nature and ability to give to others is endless. He makes all aspects of my life easier and I am forever grateful. After 20 years of marriage, you are my best friend and life partner, I love you, and I look forward to the next 50+ years together!
ACKNOWLEDGEMENTS

There are many people who have supported me in my journey as a physician, researcher, and student who I would like to acknowledge.

I would like to recognize my wonderful committee members: Elizabeth Tullis, Jan Hux, Mary Corey, Joel Ray, and Peter Austin. Liz Tullis has been provided me with unflagging support, encouragement and guidance over the past 9 years. She has been a mentor and role model to me on many levels: as a physician, leader, researcher, and friend. She has inspired me from the early years of my training in medical school and has guided me through these past years to develop as a scientist and clinician. I appreciate her wisdom and leadership and most of all her friendship. Working and collaborating with Liz makes going to work a pleasure. I am also very grateful for the support of my co-supervisors, Jan Hux and Mary Corey. Jan’s ability to explain the most complex concepts in an understandable fashion is truly a gift. I am extremely grateful for her encouragement, positive attitude and characteristic warm laugh. Mary’s foresight created the CF clinical database that I used in this dissertation and I am thankful for her time, tutelage and guidance over the last several years. I hope to continue on her path in the years to come. I want to thank Joel for his commitment, support, and creativity. I could always count on him for prompt constructive feedback, his straightforward, no-nonsense manner and positive encouragement. Peter’s statistical expertise helped me understand complex methodologies and I look forward to future collaborations with all my committee members in future years.
I am also extremely grateful to Ruth Croxford who tolerated my never-ending questions and appearances at her door. Ruth volunteered hours of her time to help me with statistical concepts, methodological issues, SAS coding and cohort creation. It would have been impossible to complete this thesis without Ruth’s guidance, teaching, and expertise and I am truly grateful to her. To the many other individuals who generously donated their time to help me with this work and to whom I am very grateful: Ping Li, Jin Luo, Nelson Chong, Rahim Moineddin, Pamela Slaughter, Kinwah Fung, Sue Schultz, Marilu Joffe, Ian McIntosh, John Rudson, Lisa Strug, Laura Rosella, and Peter Durie. Finally, I would like to thank the Canadian Institute for Health Research for their support of my graduate training, the Canadian Cystic Fibrosis Foundation who provided the clinical registry data and funding for this research and Dr. Bob Hyland for his generous support of my graduate education.
# ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFP</td>
<td>alternative funding plan</td>
</tr>
<tr>
<td>BCC</td>
<td>Burkholderia cepacia</td>
</tr>
<tr>
<td>BMI</td>
<td>body mass index</td>
</tr>
<tr>
<td>CaCC</td>
<td>calcium-activated chloride channel</td>
</tr>
<tr>
<td>CCAC</td>
<td>community care access centres</td>
</tr>
<tr>
<td>CCFF</td>
<td>Canadian Cystic Fibrosis Foundation</td>
</tr>
<tr>
<td>CDC</td>
<td>Centre for Disease Control</td>
</tr>
<tr>
<td>CF</td>
<td>cystic fibrosis</td>
</tr>
<tr>
<td>CFRD</td>
<td>cystic fibrosis-related diabetes</td>
</tr>
<tr>
<td>CFTR</td>
<td>cystic fibrosis transmembrane conductance regulator</td>
</tr>
<tr>
<td>CIHI-DAD</td>
<td>Canadian Institute for Health Information-Discharge Abstracts Database</td>
</tr>
<tr>
<td>CPDR</td>
<td>Canadian patient data registry</td>
</tr>
<tr>
<td>ENaC</td>
<td>epithelial sodium channel</td>
</tr>
<tr>
<td>FEV1</td>
<td>forced expiratory volume in 1 second</td>
</tr>
<tr>
<td>FVC</td>
<td>forced vital capacity</td>
</tr>
<tr>
<td>GEE</td>
<td>generalized estimating equation</td>
</tr>
<tr>
<td>ICD-9</td>
<td>International classification of disease, 9th revision</td>
</tr>
<tr>
<td>ICES</td>
<td>Institute for Clinical Evaluative Sciences</td>
</tr>
<tr>
<td>IKN</td>
<td>unique encrypted health card number</td>
</tr>
<tr>
<td>IRT</td>
<td>immunoreactive trypsin</td>
</tr>
<tr>
<td>IV</td>
<td>intravenous</td>
</tr>
<tr>
<td>LOS</td>
<td>length of stay</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>MI</td>
<td>multiple imputation</td>
</tr>
<tr>
<td>NYSIIS</td>
<td>New York State Identification and Intelligence System</td>
</tr>
<tr>
<td>OCFD</td>
<td>Ontario cystic fibrosis database</td>
</tr>
<tr>
<td>OHCAS</td>
<td>Ontario home care administration system</td>
</tr>
<tr>
<td>OHIP</td>
<td>Ontario health insurance plan</td>
</tr>
<tr>
<td>PD</td>
<td>potential difference</td>
</tr>
<tr>
<td>PHIPA</td>
<td>Personal Health Information Protection Act</td>
</tr>
<tr>
<td>PI</td>
<td>pancreatic insufficiency</td>
</tr>
<tr>
<td>PS</td>
<td>pancreatic sufficiency</td>
</tr>
<tr>
<td>PsA</td>
<td>Pseudomonas aeruginosa</td>
</tr>
<tr>
<td>RPDB</td>
<td>registered persons database</td>
</tr>
<tr>
<td>SES</td>
<td>socioeconomic status</td>
</tr>
</tbody>
</table>
TABLE OF CONTENTS

ABSTRACT ........................................................................................................................ ii
DEDICATION ................................................................................................................... iv
ACKNOWLEDGEMENTS ................................................................................................. v
ABBREVIATIONS ........................................................................................................ vii
TABLE OF CONTENTS ................................................................................................... ix
1 INTRODUCTION ....................................................................................................... 1
  1.1 Overview of Cystic Fibrosis ..................................................................................... 1
    1.1.1 The Cystic Fibrosis Gene .................................................................................. 1
    1.1.2 Clinical Manifestations of CF: Gastrointestinal and Pulmonary ...................... 2
    1.1.3 Diagnosis of CF .............................................................................................. 4
    1.1.4 Treatment of CF ............................................................................................. 4
  1.2 Clinical Care for CF Patients .................................................................................... 8
  1.3 Prognosis and Survival in CF ................................................................................... 9
  1.4 Gender Differences in CF ....................................................................................... 10
  1.5 Socioeconomic Status and Health .......................................................................... 15
  1.6 Hospitalizations in CF ............................................................................................ 16
    1.6.1 Reasons for, and significance of, hospitalizations in CF ................................. 16
    1.6.2 Frequency of hospitalization: Literature review and critical appraisal .......... 18
  1.7 Proposed study ....................................................................................................... 25
  1.8 Hypotheses ............................................................................................................. 26
    1.8.1 General hypothesis ......................................................................................... 26
    1.8.2 Specific hypotheses ......................................................................................... 26
    1.8.3 Research questions ......................................................................................... 27
2 METHODS .................................................................................................................. 29
2.1 Data sources ............................................................................................................ 29
  2.1.1 Canadian CF Foundation Patient Data Registry.............................................. 29
  2.1.2 Health Care Administrative Data Sources....................................................... 31
2.2 Study Population ..................................................................................................... 33
2.3 Time Period ............................................................................................................. 33
2.4 Data Linkage ........................................................................................................... 33
  2.4.1 Identifying duplicate records within CPDR ..................................................... 33
  2.4.2 Identifying CPDR CF patients within RPDB ................................................... 34
2.5 Assessing “agreement” between the CPDR and ICES databases for the diagnosis of CF .............................................................................................................................. 35
2.6 Creating clinical cohort of CF subjects using the Ontario CF Database ................. 39
2.7 Outcome Definition: Hospitalization ..................................................................... 40
2.8 Predictor Variables .................................................................................................. 41
  2.8.1 Clinical data ...................................................................................................... 41
  2.8.2 Missing data ...................................................................................................... 42
  2.8.3 Socioeconomic Status ....................................................................................... 43
  2.8.4 Rural versus urban ............................................................................................ 43
  2.8.5 Distance from CF centre .................................................................................. 43
2.9 Statistical Analyses ................................................................................................. 44
  2.9.1 Sample Size Estimation .................................................................................... 44
  2.9.2 Analyses ........................................................................................................... 45
2.10 Ethics ..................................................................................................................... 48
3 RESULTS ...................................................................................................................... 49
  3.1 Data Linkage ........................................................................................................... 49
  3.2 Study Cohort .......................................................................................................... 49
3.3 Patient Characteristics ........................................................................................................ 50
   Ages 7-19 years ...................................................................................................................... 50
   Over 19 years of age ........................................................................................................ 51
3.4 Hospitalizations ............................................................................................................... 59
   3.4.1 Comparison between Health Care Administrative Data Sources and CPDR... 59
   3.4.2 Respiratory Hospitalizations ...................................................................................... 63
3.5 Analysis of predictors of hospitalization ......................................................................... 66
   3.5.1 Unadjusted analysis .................................................................................................. 66
   3.5.2 Adjusted analysis .................................................................................................... 71
3.6 Sensitivity Analyses ......................................................................................................... 78
3.7 Identifying individuals with CF: Comparison of administrative data and CPDR... 81
   3.7.1 OHIP and CIHI claims compared to CPDR ............................................................. 81
   3.7.2 CIHI-DAD claims only compared to the CPDR .................................................... 82
4 DISCUSSION ...................................................................................................................... 88
   4.1 Summary of research findings .................................................................................... 88
   4.2 Reasons for gender disparity in hospitalization rates .............................................. 90
      4.2.1 Outpatient Management .................................................................................... 90
      4.2.2 Hormonal Influences ....................................................................................... 92
      4.2.3 Patient- and Provider-level Influences ............................................................ 94
   4.3 Other predictors of hospitalization ........................................................................... 96
   4.4 Limitations of current study ...................................................................................... 101
      4.4.1 Limitations of the outcome variable ................................................................. 101
      4.4.2 Limitations of the predictor variables ............................................................... 102
      4.4.3. General limitations ......................................................................................... 104
   4.5 Future directions of research .................................................................................... 104
1 INTRODUCTION

1.1 Overview of Cystic Fibrosis

1.1.1 The Cystic Fibrosis Gene

Cystic fibrosis (CF) is the most common fatal genetic disorder in Caucasians, with an incidence of 1 in 3000 live births. It was not until 1989 that researchers discovered that CF, an autosomal recessive disease, was caused by mutations in a single gene located on the long arm of chromosome 7, termed the cystic fibrosis transmembrane regulator (CFTR) gene.\(^1\) The CFTR gene codes for a protein called the transmembrane conductance regulator which not only functions as a chloride channel, but also is involved in many other cellular functions. The CFTR chloride channel normally resides at the apical surface of epithelial cells including sweat glands, respiratory epithelium, pancreatic ducts, intestinal epithelium, biliary ducts and the vas deferens. The disease causes abnormal ion transport across epithelial membranes, resulting in thick tenacious secretions. To date, over 1600 different mutations in the CFTR gene have been identified.\(^2\) The most common mutation worldwide is a three base-pair deletion resulting in deletion of the phenylalanine residue at amino acid position 508, commonly referred to as ‘delta F508’. Mutations in the CFTR gene can be classified into one of five groups according to their predicted functional consequences. Class I – III are considered severe mutations producing no functioning CFTR protein while class IV and V mutations result in milder pancreatic disease as some functioning CFTR protein is made although decreased in quantity and/or quality.\(^3\) Typically severe mutations are associated with a more severe clinical phenotype; however the genotype-phenotype relationship is variable with significant heterogeneity in severity of lung disease.\(^4,5\) Overall,
there is a general deterioration of lung function with age, but severity of lung function at any given age can vary widely between individuals (even in those with severe mutations on both alleles). While \textit{CFTR} genotype clearly contributes to the pulmonary phenotype, other factors such as genetic modifiers and environmental influences have been postulated to explain the clinical variability in pulmonary CF disease.

\subsection*{1.1.2 Clinical Manifestations of CF: Gastrointestinal and Pulmonary}

The gastrointestinal tract and respiratory system cause the majority of morbidity and mortality in CF. Due to abnormal ion transport, secretions within the pancreatic ducts are thickened. Thick secretions block the ducts and result in destruction of the pancreatic tissue. Pancreatic exocrine insufficiency occurs in over 85\% of individuals with CF resulting in fat malabsorption, steatorrhea, weight loss, and malnutrition. With severe genetic mutations, damage to the pancreas begins in-utero and evidence of malabsorption is often seen at birth or within the first few years of life. The use of exogenous pancreatic enzyme replacement therapy ameliorates fat malabsorption improving overall nutritional status. Approximately, 15\% of individuals with CF have enough residual pancreatic function to permit normal digestion without enzyme supplementation. Such individuals are commonly referred to as “pancreatic sufficient” and are typically diagnosed later in life, have milder clinical manifestations, better nutritional status, and better overall prognosis.

In contrast to pancreatic disease, pathologic studies show that individuals with CF are born with essentially normal lungs. The lack of CFTR within the airway epithelium results in abnormal ion transport across the cell membrane. In addition to transporting chloride, the CFTR protein affects other channels within the cell membrane, most notably the epithelial
sodium channel (ENaC) which is responsible for salt and water re-absorption in the airway. ENaC is down-regulated by functional CFTR, therefore, the absence of CFTR results in over-activity of this channel. Excessive salt and water re-absorption ensues, depleting the airway surface liquid and impairing ciliary function.\textsuperscript{9,10} In addition, CF mucous itself has a reduced water content making the mucous difficult to clear. Such factors may contribute to the fact that early in life, patients with CF develop bacterial infections with organisms such as \textit{Staphylococcus aureus} and \textit{Hemophilus influenzae}.\textsuperscript{11} With age, \textit{Pseudomonas aeruginosa} (PsA) appears which can acquire a mucoid phenotype associated with a further decline in pulmonary function. Initially these organisms can be eradicated, but with time, they become permanent inhabitants of the CF airways regardless of aggressive chest physiotherapy and antibiotic treatment. The normal immune response is to recruit inflammatory cells in an attempt to eradicate infection but there is evidence to suggest that the inflammatory response in CF is excessive.\textsuperscript{12} Studies in infants with CF have higher number of neutrophils and interleukin-8 levels in bronchoalveolar lavage fluid compared with infants who do not have CF.\textsuperscript{13} Furthermore, some studies suggest that inflammation may actually precede infection. Persistent bacterial stimulation and the overzealous inflammatory response seen in the airways of those with CF, results in excess neutrophil products in the airway. Proteases released from neutrophils digest structural proteins within the lungs and the cycle of infection and inflammation results in airway damage ultimately causing bronchiectasis and cystic changes in the lung (Figure 1). Typically, individuals with CF experience a progressive decline in pulmonary function over time which is punctuated by intermittent exacerbations. Exacerbations are associated with increased clinical symptoms and/or a decline in lung function. Antibiotic therapy (oral and intravenous) is the mainstay of
treatment for pulmonary exacerbations in an effort to decrease bacterial density within the lungs, reduce inflammation, and restore pulmonary function. The cumulative effects of multiple exacerbations may contribute to worsening lung disease and individuals with more severe or frequent exacerbations may have worse outcomes. In addition to pulmonary infectious exacerbations, individuals with CF can experience other lung complications such as asthma, allergic bronchopulmonary aspergillosis, pneumothorax, atelectasis and massive hemoptysis. All of these conditions contribute to the morbidity and ultimately the mortality, of lung disease in CF.

1.1.3 Diagnosis of CF

The diagnosis of CF is made by clinical and laboratory testing. Evidence of chloride channel dysfunction must be present (for example, sweat chloride ≥ 60 mEq/L, genetic testing showing a disease causing CF mutation), plus a family history of CF or clinical features consistent with CF. Newborn screening is widely available and was instituted within Ontario in 2008 in an attempt to identify individuals with CF early in life. The initial screen used is a high immunoreactive trypsinogen level (IRT). In Ontario, positive tests are followed up with a limited CFTR mutation screen and sweat chloride testing. Prior to newborn screening, individuals with mild forms of CF could go undiagnosed until much later in life.

1.1.4 Treatment of CF

CF is a multisystem disease that requires a variety of therapeutic interventions depending on the specific CF-related complication in each organ. The mainstay of therapy for CF is the management of complications resulting from the primary genetic defect rather
than altering the gene defect itself. Gene therapy has had limited success due to difficulties with the vectors used, the frequency of administration and the inflammatory response that is elicited.\textsuperscript{16} Because malnutrition and pulmonary infections are the primary causes of morbidity and mortality in CF, subsequent discussions will be limited to therapeutic interventions in these areas.

Poor nutrition has clearly been shown to be associated with negative outcomes such as worse pulmonary function and worse survival in CF.\textsuperscript{17} Assessment of age-related anthropomorphic measurements is a simple and reliable way to monitor the nutritional status of individuals with CF. Monitoring and evaluating the degree of fat malabsorption using 3-day fecal fat measurements, fecal elastase-1, and serum trypsinogen can identify those individuals who would benefit from exogenous enzyme replacement therapy. Normalizing weight and growth with aggressive nutritional supplementation and pancreatic enzymes has been a fundamental component of the recommended treatment in CF. The high-fat high-calorie diet that has been recommended in Toronto since the 1970s has been shown to improve survival by 9 years compared to centres using low-fat diets and is now the standard of care for CF worldwide.\textsuperscript{17}

With respect to pulmonary disease, therapy is aimed at treating infections and alleviating bronchial obstruction which is time consuming and labour intensive. Chronic therapy involves a combination of daily chest physiotherapy, inhaled antibiotics twice daily, inhaled medications to thin the mucous (mucolytics) once or twice daily, and occasionally anti-inflammatory medications. Acute infectious exacerbations in CF are managed with oral or intravenous antibiotic therapy. Intravenous antibiotics typically involve two different
antimicrobials for a minimum of 10 days of therapy.\textsuperscript{18} Inhaled antibiotics and mucolytic therapies have been shown to improve lung function, reduce exacerbation rates, increase weight and improve quality of life.\textsuperscript{19-21} Anti-inflammatory therapy in CF has been shown to reduce rate of decline of pulmonary function however the use of oral corticosteroids and non-steroidal anti-inflammatory medications are associated with unacceptable rates of associated side effects limiting their long-term use.\textsuperscript{22,23}
Figure 1: Pathophysiology of CF

Abnormal Gene

Abnormal CFTR

Abnormal Na⁺, Cl⁻, Water movement through the cell

Thick dry mucous

Bronchial airway obstruction

Thickened mucous

Infection

Release of proteases & DNA

Inflammation

Progressive Lung Disease and Respiratory Failure
1.2 Clinical Care for CF Patients

Due to the multi-system nature of CF disease, CF care has evolved to include a multidisciplinary team comprised of physicians, nurses, physiotherapists, dietitians, respiratory therapists, and social workers so that all aspects of care are addressed and optimized. Within Canada, the Canadian CF Foundation (CCFF) has established 39 specialized CF care centres (both pediatric and adult) across the country to provide comprehensive care to patients and families dealing with CF. Most CF centres are in major cities; however some provinces run satellite CF clinics in rural areas of the province. For instance, physicians travel to satellite CF clinics in northern Ontario (e.g. Thunder Bay) approximately 2-3 times per year in order to better serve the northern CF population. Given the large size of some provinces like Ontario, it is possible that individuals with CF living in remote areas may not receive the same level of care as those living closer to a CF centre although this has not been formally studied. Canadian CF centres are accredited by the CCFF every few years in order to ensure an approved standard of care across the country. It is felt that most Canadians with CF are followed at CF centres for their CF care because of certain incentives such as the multidisciplinary approach to care as well as provincial financial support for CF drugs if medications are distributed from an accredited CF centre. The specific drugs covered by provincial CF drug programs vary considerably by province. Some provincial programs are quite extensive while others have various levels of subsidy, but all provinces have some CF drug coverage.

Guidelines on patient services, evaluation and monitoring in CF have been put forth by the American and European CF Foundations in order to promote a uniform standard of
treatment so that patient care can be optimized. Recommendations include frequency of outpatient routine clinic visits, sputum cultures, lung function and radiography. The guidelines recommend at least four clinic visits per year to a specialized CF centre. Although care within a specialized CF centre is advocated, monitoring by a family physician is also suggested. These guidelines state that the primary care physician also has an important role in administering immunizations, evaluating and treating milder pulmonary exacerbations, and assessment of non-CF-related conditions. Whether or not these practice guidelines are associated with improved outcomes hasn’t been definitively proven; however Johnson et al found that CF centres reporting the highest values for median forced expiratory volume in the first second (FEV₁) monitored patients, and obtained lung function measurements and sputum cultures more frequently suggesting that the recommendations may result in improved outcomes.

1.3 Prognosis and Survival in CF

The face of CF has changed substantially over the last 40 years. In the 1960s, CF was almost exclusively a pediatric disease with the median survival being less than 5 years of age whereas today, most individuals live well into adulthood. Prognostic variables linked to survival include CFTR genotype, initial disease presentation, pulmonary function, nutritional status, sputum bacteriology, age, socioeconomic status, pulmonary exacerbations, and gender. Literature has shown that the development of specialized CF centres has also contributed to the improved care and survival seen in CF over the past 20 years. Since the development of antibiotic therapy in the post-war era, and the focus on adequate nutrition during the 1970s, survival of individuals with CF has improved dramatically. Corey et al compared survival, growth, and pulmonary function between two
large similar clinic populations, one in Boston and one in Toronto. Individuals with CF in Toronto had a staggering 9 year survival advantage with the median age of survival being 30 years in Toronto and 21 years in Boston. Although there were some components that differed between the two centres, the most dramatic difference was the approach to nutrition and pancreatic supplementation. Toronto advocated a high fat, high calorie diet with appropriate enzyme supplementation while Boston, like the rest of the CF community at the time, recommended a low fat, high calorie diet with fewer pancreatic enzymes. The authors concluded that Toronto’s focus on aggressive nutritional support contributed significantly to the survival advantage of Toronto patients. Today, high-fat, high-calorie diets with a goal of normalizing body mass index in CF patients is the standard of care worldwide.

With the number of CF deaths per year decreasing, it is more challenging to show an impact on survival for many new therapies. Surrogate markers closely linked to survival, such as lung function, have often been used to demonstrate a benefit from a new therapy. Over the last decade, therapies such as mucolytic treatments, anti-inflammatory medications, and antibiotics, have resulted in a more modest improvement in lung function; however effects on survival have been more difficult to prove.

1.4 Gender Differences in CF

Worse overall survival in female patients with CF, despite similar disease severity to males, has been documented in North America, Europe, New Zealand and Australia. Furthermore, if a woman with CF develops CF-related diabetes (CFRD), her survival is significantly worse when compared to females without CFRD and males, with or without diabetes, suggesting a possible gender-diabetes interaction. Several potential risk factors
have been postulated to explain worse overall survival in females including increased susceptibility to detrimental respiratory pathogens; increased inflammatory response; increased rate of decline in lung function; worse nutritional status; genetics; poor fitness; delayed diagnosis; differential outpatient therapy and hormonal influences although very few of these reasons have been formally studied.

Within Canada, overall survival for females with CF over a 20-year period (1970-1989) was significantly worse than males. Median survival for males and females was 29.5 and 21.3 years, respectively. Rosenfeld et al examined a large cohort of CF patients in the U.S. in an attempt to identify explanatory factors for gender-related differences in survival. The study cohort consisted of over 21,000 individuals with CF and the results showed that females were 60% more likely to die than males among the 1-20 year age group even after adjustment of all potential explanatory variables (relative risk 1.6, 95% confidence interval [CI] 1.4-1.8). The cause of death was similar between males and females. Female infants less than one year of age seemed to have a slight advantage over their male counterparts although this was not statistically significant. Beyond 20 years of age, no difference in survival between the genders was noted. Multiple variables were examined to explain this difference; however pulmonary function, specifically FEV1 and forced vital capacity (FVC), were the only factors that had any effect on gender-related risk of death. After adjustment for FEV1 and FVC, the relative risk for death in females decreased but sex continued to be an independent risk factor for death. None of the other factors (race, mode of presentation, age at diagnosis, pancreatic status, sputum bacteriology, nutritional status and sweat chloride), alone or in combination, altered this risk.
Chronic bacterial infection of the lower airways is a major cause of morbidity and mortality in CF. Recurrent respiratory infections ultimately damage the airways, result in declining lung function, respiratory failure and death. Typical CF organisms include *Pseudomonas aeruginosa* (PsA), *Staphylococcus aureus*, *Hemophilus influenza* as well as other less common gram negative organisms such as *Burkholderia cepacia complex* (BCC). Once these organisms are consistently cultured from the sputum of patients with CF, it is difficult to eradicate them despite antimicrobial therapy. From an epidemiological point of view, the acquisition of these organisms, particularly PsA and BCC, is associated with a worse prognosis and decreased survival compared to those who are not infected with these bacteria. One possible explanation for worse survival in females with CF may be that they acquire detrimental organisms more frequently or earlier than their male counterparts. Demko et al completed a retrospective analysis of 848 patients with CF to evaluate whether or not females acquire PsA at an earlier age compared to males. The results showed that females with CF acquired mucoid PsA at 9.5 years of age compared to 11.2 years of age for males; however the authors concluded that this was not enough to account for the 3 year difference in median survival between the genders. Those who grew mucoid PsA were further classified into “early” (< 6 years of age) and “late” (> 6 years of age) acquisition. Subjects who acquired this organism early had a significantly worse 10-year survival compared to subjects with late acquisition (75% vs. 94%; p<.0005). Within each group, survival in females was significantly worse (p=0.01). Although there was a clear association between mucoid PsA and deterioration in clinical parameters such as FEV\textsubscript{1} and chest X-ray score, the deterioration was similar for males and females; therefore worse survival was not solely due to worse lung disease. Allen et al also found a higher proportion of female
children with bacteria identified from sputum cultures as compared to males with CF; however specific organisms were not stated in the paper. Rosenfeld et al reported a higher prevalence of PsA within the sputum in females between the ages of 6-15 years; however this factor did not explain the worsened female survival. There has been some suggestion in the literature that females colonized with BCC have worse outcomes than males with the same organism. Lewin et al reported that within a year after acquisition of BCC, 33% of women died compared to 21% of males. However, when overall mortality was compared between infected men and women, no significant difference was noted. The rate of decline of lung function following infection seemed to be greater for females although, given the low prevalence of this organism worldwide, it is not likely to significantly explain the survival advantage in males with CF. Furthermore, this study was published before knowledge of the different species of BCC which may have affected their conclusions.

With increasing life expectancy, people with CF are experiencing complications from their disease not previously recognized. One such complication is diabetes; in 2007, the US CF Foundation Patient Registry reported 20.6% of patients of all ages had a diagnosis of CFRD. CFRD is associated with advancing age, affecting 35% of adults in their twenties and as many as 50% of adults over the age of 30. Risk factors for the development of CFRD have been reported in several published papers to be increasing age, female gender, and pancreatic insufficiency. Marshall et al found that females were disproportionately affected by CFRD with a prevalence of 17.1% as compared to 12% in males. Furthermore, the presence of CFRD is associated with worse pulmonary function, more pulmonary exacerbations requiring intravenous (IV) antibiotic therapy in the preceding year compared to those without CFRD (mean (SD) of 1.55 (1.84) vs. 0.78 (1.32) courses of IV antibiotics,
age adjusted p < .001), and a higher prevalence of PsA and BCC than the non-CFRD subjects. Other studies have suggested that CFRD is associated with increased mortality, particularly in females. Milla et al reported that females with CFRD had significantly higher risk of death compared to males with or without CFRD and females without CFRD. Lung function was the only other significant predictor of subsequent risk of death. A female diabetic subject with poor lung function (FEV$_1$ < 50% predicted at the time of CFRD diagnosis) was five times more likely to die than a diabetic female with mild lung disease (FEV$_1$ > 85% predicted). In contrast, the risk of death of diabetic males with an FEV$_1$ < 50% was only 2 time higher than males subjects with an FEV$_1$ > 85% predicted. The etiology of the gender differences noted in diabetes is unclear but hormonal influences or a pro-inflammatory state may contribute to this disparity. It is also possible that diabetes itself is simply a marker for some unknown biologic difference that has not been elucidated.

Another possible explanation for worse outcomes in females is that there may be a difference in age and/or presenting symptoms in females with CF. Early diagnosis in CF through newborn screening has been associated with better nutritional status and possibly less severe lung disease, both of which are associated with improved survival in CF. Lai et al found that if patients were diagnosed with CF because of symptoms other than meconium ileus, females were diagnosed at significantly older ages than males (12.7 vs. 8.7 months; p<.001). The delay in diagnosis for females compared to males was longest if the patients presented with respiratory symptoms only, with a median 18 month difference (p<.001). Females who presented with a combination of respiratory and gastrointestinal symptoms were diagnosed significantly later than the corresponding male group (p=.003). Interestingly, an opposite trend was seen when the presenting symptoms were categorized as “other”
(including nasal polyps, rectal prolapse, electrolyte abnormalities, liver disease and other).
Within this category, males were diagnosed later at 44.1 vs. 33.1 months for females (p=.02).
The authors argue that this delay in diagnosis occurs within the first years of life during a
critical stage of development which may in fact be the optimal time for effective preventative
interventions. Although the delay in diagnosis could possibly be explained by differential
onset and severity of symptoms between genders, their analysis did not show any significant
differences between clinical symptoms, radiography, genetics, or acquisition of PsA. Another
potential explanation for gender differences in diagnosis is that, in those who present with
respiratory symptoms only, males may be more physically active; therefore respiratory
symptoms were more evident earlier resulting in medical attention. This may reflect society
and cultural trends related to gender biases. Of course this would apply to older children who
are ambulatory and would not explain the delay in diagnosis within the first year of life.
Although this delay in diagnosis for females may partially explain the gender difference seen
in survival, it is not possible to say for certain what the impact has on long-term survival.

1.5 Socioeconomic Status and Health

Low socioeconomic status (SES) has been consistently linked to worse health
outcomes and individuals living in low income areas have higher rates of mortality and
morbidity related to chronic disease. In terms of CF, several published papers suggest
worse outcomes with lower SES status. O’Connor et al showed a strong monotonic
association between median household income and mortality rate in American CF subjects
representing a 44% increased risk of death in the lowest income category; however this was
not adjusted for insurance status. O’Connor’s study also demonstrated that individuals with
CF living in areas with lower SES had worse lung function and nutritional status when
compared to people in higher income areas. Although there is no literature in the Canadian CF population, worse health outcomes in Canadians with other medical conditions have been noted in those with lower income. Demissie et al found that in Montreal, Quebec healthy boys from the lowest SES category had an FEV$_1$ lower by 8.2% predicted (95% CI, -13.8 to -2.1) when compared to the most advantaged group.$^{62}$ Worse outcomes following myocardial infarction have been documented in those with lower SES; however this is due in part to being sicker at baseline and less likely or less able to engage in healthier lifestyle activities such as eating fresh fruits and vegetables, regular exercise, weight reduction and smoking cessation.$^{63}$ A study conducted in Toronto showed that low income was associated with 50% higher rates of hospitalizations (relative risk 1.50, 95% CI 1.43-1.58) compared to those in the highest income category.$^{64}$ Admissions were for conditions such as asthma, bacterial pneumonia, seizures, and dehydration. Subjects were followed from birth to 9 years of age and the socioeconomic disadvantage was seen across all years. SES and its relationship to outcomes in CF has not been studied in a Canadian cohort but given the evidence in other illnesses, it is plausible that SES may impact Canadian CF patients in a similar fashion.

### 1.6 Hospitalizations in CF

#### 1.6.1 Reasons for, and significance of, hospitalizations in CF

Individuals with CF can be hospitalized for a variety of reasons, both CF- and non-CF-related. Because the respiratory system causes the greatest morbidity and mortality in CF, the treatment of pulmonary infectious exacerbations is generally felt to be the most common reason for admission to hospital although this has not been formally studied. In North America, antibiotic therapy is initiated only when clinical signs and symptoms suggest a worsening infection.$^{18}$ When a patient has failed oral antibiotic therapy for infectious
symptoms, they are typically treated with IV antibiotic therapy either in hospital or at home using homecare services. This differs from some centres in Europe where CF patients are electively admitted to hospital approximately every 3 months for IV antibiotics regardless of clinical symptoms in an effort to minimize any deterioration in lung function from ongoing infection and inflammation although the evidence for this approach is controversial.65 66

Hospitalizations for treatment of CF pulmonary exacerbations have been shown to be negatively associated with survival, health related quality of life, neurocognitive function, sleep and contribute to significant health care costs.14;67-70 Liou et al showed that annual number of pulmonary exacerbations was a significant predictor of 5-year survival even after adjusting for important clinical markers of disease severity.14 In fact, each acute pulmonary exacerbation within the year had a significant negative impact on 5-year survival equal to subtracting 12% from the measured FEV$_1$ % predicted value. If the number of exacerbations within a year was 4, the corresponding equivalent in loss of lung function was similar to that seen in those infected with BCC, a significant negative prognostic indicator. Another study carried out in the U.S. found that being hospitalized was the primary risk factor for acquiring a new infection with *Burkholderia dolosa* in a cohort of CF patients.71 Furthermore, the probability of dying within 18 months of acquiring this organism was 13% compared to 3% for control subjects. A similar event occurred in Toronto, Canada where 4 out of 5 patients died following the acquisition of a virulent strain of *Burkholderia cenocepacia*. The common factor among all 5 newly infected cases was a hospitalization occurring during an overlapping time period.72

Britto et al compared quality of life scores between the general population and individuals with cystic fibrosis.67 Pulmonary exacerbations in the CF group had a profound
negative impact on physical and psychosocial quality of life that was not explained by differences in lung function or nutritional status. In another study, Dobbin et al found that during CF exacerbations, subjects had disruptions to sleep architecture, specifically reduced REM sleep, which improved with treatment. Exacerbations also impaired performance of tasks designed to assess reaction time, concentration, and reasoning. Therefore, studies that provide insight into those at highest risk for hospitalization have the potential to significantly affect both patient outcomes and health care costs.

1.6.2 Frequency of hospitalization: Literature review and critical appraisal

The frequency of hospitalizations overall, and for pulmonary exacerbations in particular, in CF has not been extensively described in the literature. FitzSimmons reviewed the changing epidemiology in CF using national U.S. CF Foundation registry data. This large study of over 17000 individuals with CF described the care provided as well as summarized complication rates and survival statistics during 1990. This paper only superficially discussed the frequency of hospitalizations during 1990 for the study population, as it was not the focus of the article. Thirty-seven percent of the study cohort was hospitalized one or more times during the year with a mean length of stay of 11.9 days. However, reasons for hospitalizations were not specified, total number of hospitalizations was not stated, and age nor gender differences in hospitalization rates were not evaluated. The author did note a growing trend toward home therapy for CF including intravenous antibiotics, supplemental feeding and oxygen therapy. However it is difficult to interpret this finding since rates from previous years were not given. Furthermore, the author did not state whether home IV antibiotics were initially started in hospital, and hence were captured in the total number of hospitalizations, or if they were initiated in the outpatient clinic with no
associated hospital admission. If they were started in the outpatient setting only, then the number of “pulmonary exacerbations”, if using number of hospitalizations as a surrogate marker for exacerbations, would be underestimated and would not accurately reflect health services utilized.

Five years later in a related study, Konstan et al reported similar results again utilizing CF Foundation registry data. However, this study included 11,837 U.S. and 794 Canadian subjects. The purpose of this prospective observational study was to characterize the natural history of disease in a large cohort of individuals with CF and to compare practices to published guidelines on standard of care for CF. Among 12,631 study subjects, 4375 (35%) had one or more hospitalization during 1995 corresponding to 8,561 hospitalizations in total. The study results suggested females were hospitalized more than males (36.9% vs. 32.6% respectively). The proportion of patients hospitalized increased with age and lower pulmonary function. In addition, patients were hospitalized more frequently and had a longer hospital length of stay if they were older or if they were colonized with PsA. However, none of these comparisons were tested for statistical significance. Furthermore, they were all described in isolation rather than adjusting for each factor in a multivariable statistical model to identify independent predictors of hospitalization. Details of the hospitalizations were notably absent in the paper. It is not clear if admissions represented only pulmonary exacerbations or captured all hospitalizations and no information was provided on home IV therapy. Although 794 Canadian subjects were included in this study, the data were amalgamated and specific information on Canadians with CF was not available. Given the differences in care and health care systems between Canada and the
U.S., it is possible that factors associated with hospitalization may differ between the two countries.

Limited Canadian data are available in a report on “Respiratory Diseases in Canada” produced by Health Canada.\textsuperscript{75} The unique aspect of this report was that administrative data from Canadian Institute for Health Information were utilized to quantify total number of hospitalizations for Canadians with CF. This report stated there were 1381 hospitalizations in the CF population in 1998. Unfortunately, the report failed to provide sufficient detail in order to interpret this result. Firstly, no information was given on the total number of CF patients hospitalized and no denominator was given in association with the total number of hospitalizations (i.e. # of patients hospitalized/ number of patients hospitalized + not hospitalized) making it very difficult to interpret this result. Secondly, the report failed to specify if the number of hospitalizations represented all hospitalizations or only acute pulmonary exacerbations. There were no details on how hospital admissions were identified or defined using administrative data. Descriptive data revealed that individuals in their teen years, females between 5-24 years of age, and those living in eastern provinces seemed to have the most hospitalizations although no statistical measures were performed to quantify these conclusions. Finally, because health administrative data were used in isolation, adjustments for important clinical parameters were not possible. Females and those in their teen years may have been hospitalized more simply because they had more severe lung disease, but patient level clinical data were not utilized for the analysis. This is a significant limitation of most research using administrative data only, namely the lack of associated patient-level clinical data.
The American CF Foundation Annual Report for 2007 found 38% of patients were hospitalized one or more times for pulmonary exacerbations.\textsuperscript{56} As expected, length of stay and proportion of patients hospitalized increased as lung function deteriorated. These data are based on reports provided by the respective CF clinics across the country. The annual report is simply descriptive, with no statistical comparison or adjusted analyses carried out. Furthermore, the U.S. registry data does not capture all individuals with CF living in the country. Individuals with milder disease or those without health insurance may not be seen at a CF centre. The only study to date that has systematically examined specific predictors of pulmonary exacerbations (requiring hospitalization for IV antibiotics) in cystic fibrosis was published by Block and colleagues in 2006.\textsuperscript{76} Patients in this study were originally recruited as part of a clinical trial evaluating combination antibiotic susceptibility testing for the treatment of individuals with CF who were infected with multi-resistant bacteria. Multidrug resistance was defined as resistance to all antibiotics in at least 2 of 3 major anti-pseudomonal antibiotic classes. Two hundred and forty-nine adult study subjects were from 10 Canadian and two Australian CF centres. Patients were followed for up to 4.5 years. After 12 months of observation, 50% of subjects had experienced an exacerbation requiring intravenous antibiotic therapy. Multivariable Cox proportional hazards model was used to examine time to first exacerbation for the entire study duration. This analysis showed that younger age (hazard ratio [HR] 0.98, 95% CI 0.96 to .99), female sex (HR 1.45, 95% CI 1.07 to 1.95), lower lung function (HR 0.98, 95% CI 0.97 to .99), and a previous history of pulmonary exacerbations (HR 3.16, 95% CI 1.93 to 5.17) were significant predictors of pulmonary exacerbations. Females had a shorter time interval to first exacerbation even adjusting for lung function however in the multivariable logistic model predicting pulmonary
exacerbation during the first year of follow-up, female sex was a non-significant predictor (OR 1.37, 95% CI 0.69 to 2.72). The limitations of this study include the fact that only adults with multi-drug resistance bacteria were included limiting the generalizability of the results. Patients with multi-drug resistant organisms may be substantially different than those infected with pan-sensitive bacteria and in fact, a study by Lechtzin et al showed that such patients have an accelerated rate of decline of lung function when compared to patients with non-resistant strains.77

The few American epidemiologic studies done have reported on the proportion of patients hospitalized over a brief time period, typically 1 year, which limits their ability to understand trends in CF care over time. In addition, published articles in this area are from the U.S. which has a very different health care system. Given the publicly funded health care system in Canada and differences in practices of CF care, the findings of previous studies may not be generalizable to Canadians with CF.

Yet another limitation of previously published literature is the fact that the data used were from clinical registries or administrative databases alone rather than in combination. When trying to understand reasons for higher hospitalization rates in CF (for any reason), clinical variables need to be taken into account. Without this, it is very difficult to interpret the utilization data. One of the advantages of using CF clinical registries is the access to detailed individual patient-level data on important clinical variables. However registries often lack accurate, inclusive information on health care utilization, specifically hospitalization. The method for documenting clinic parameters in registries is directly inputting test results (such as lung function) into the registry whereas hospitalization data is a
combination of self-report (indirectly from patients through clinic staff) and supporting
documentation such as hospital discharge records. Although the validity of these methods
have not been formally studied in the setting of CF, many studies have shown that self-report
is not the gold-standard for defining health care utilization in other diseases. Raina et al
found that seniors underreported the number of contacts with medical specialists and general
practitioners. For major events such as surgery or hospitalization, level of agreement
between self-report and administrative data is higher however some researchers have found
that the inaccuracy of estimations increased as the volume of services increased. This
raises concerns for the CF population, who often have many encounters with the health care
system throughout their lifetime; therefore accurately reporting the number in a given year
could be challenging. Clinical registries also tend to capture specific hospitalizations rather
than all hospitalizations. For example, within the Canadian CF registry, all “CF-related”
hospitalizations are to be entered into the database yet the definition of a “CF-related”
admission can be obscure. For example, is an admission for a fracture considered CF-related
because of the possibility the patient may have CF-related bone disease? Is a hospitalization
for a suicide attempt CF-related because the individual was depressed about their disease?
The confusion in definitions may lead to inaccurate data related to this variable. Also, CF
registries tend to capture the primary reason for hospital admission which may underestimate
co-morbid conditions contributing to hospital admissions. Health care administrative
databases allow for multiple diagnostic codes to explain hospital admissions therefore
providing richer information on reasons for health care utilization. The treatment of
pulmonary infections at home using Community Care Access Centres services (CCAC),
more commonly known as homecare, is often utilized in CF. If a clinical registry tries to
capture frequency of pulmonary exacerbations by collecting detailed information about hospitalizations, home IV antibiotics may be missed, particularly if an individual was started on therapy directly from the outpatient clinic. Homecare records are available in administrative databases in Ontario thereby allowing more comprehensive capture of home IV therapy. Lastly, for CF registries, it is the CF clinic staff who gathers data on number of clinic visits, frequency of hospitalizations, and reasons for hospital admissions. Typically, clinic staff only has access to health care encounters within their institution. If an individual was hospitalized at a community hospital or went to a walk-in clinic for CF medical care, this information may not be available and hence inaccurately reported to the CF registry. One of the strengths of utilizing administrative databases is the ability to capture comprehensive health care utilization for Ontarians.

In summary, no study has systematically evaluated predictors of hospitalization in the general CF population and there are no published studies using a combination of health administrative data and clinical registry data. The most common reason for hospital admission is assumed to be due to pulmonary exacerbations but this has not been formally studied. Furthermore, there are no Canadian studies that have attempted to answer the questions: 1) How often are CF patients hospitalized? 2) What proportion of hospitalizations are related to the pulmonary system? 3) Have hospitalization rates changed over time with improved care? 4) What factors affect hospitalization rates in our Canadian CF population? and 5) Are there subgroups of individuals at higher risk for hospitalization?
1.7 Proposed study

This novel research involved creating a cohort of individuals with CF that incorporates both health care utilization data as well as patient-level CF clinical data over a 10-year period to examine health indicators in CF from a population perspective. Although survival might be argued to be the best outcome for such research, during the study period there were insufficient deaths in Ontario to make meaningful conclusions. Survival over the past decade has essentially reached a plateau, with few new therapies that have had the same dramatic impact as nutritional support did in the 1970s. Since pulmonary disease in CF is associated with significant morbidity and mortality, and hospitalizations for respiratory complications are common in CF, this was chosen as the outcome of interest for this study.

The creation of this body of work is an initial step towards a comprehensive program of health services research in Canadians with CF. The techniques used in this dissertation will be used to create the foundation upon which future epidemiologic health services studies will be built. Future research will expand upon this work allowing researchers to evaluate how care is being delivered to this patient population, identify gaps and disparities in medical care, evaluate how interventions impact health outcomes in CF, and help in the development of a comprehensive assessment of care in CF that could be used by physicians, advocacy groups and administration to improve health care delivery to this population. In addition, this work will provide important information about the comprehensiveness of the Canadian CF Patient Data Registry in capturing all Ontarians with CF. And finally, this work could serve as a model to study other chronic diseases which have similar database information.
The specific purpose of this study was to quantify longitudinal hospitalization rates for pulmonary exacerbations in Ontarians with CF over a 10 year period and to examine specific predictors (clinical, demographic, and geographic), in order to identify those at high risk for hospitalization. A secondary goal of this research was to evaluate the usefulness of administrative data to identify those individuals with CF using the CF clinical registry as the reference standard.

1.8 Hypotheses

1.8.1 General hypothesis

The general hypothesis of this dissertation was that patients with CF have frequent contact with the health care system, in particular, hospitalizations, and that there are clinical, demographic and social factors that are significantly associated with hospitalization.

1.8.2 Specific hypotheses

1.8.2.1 Primary Hypothesis

The primary hypothesis for this dissertation was that there exists a gender difference in hospitalization rates among CF patients.

1.8.2.2 Secondary Hypotheses

The secondary hypotheses are:

1) Predictors of hospitalization include age, lung function, nutritional status, bacteriology and CF-related co-morbidities. As lung function and nutritional status worsen, hospitalization rates increase. The presence of PsA, BCC, and CFRD increases hospitalization rates.
2) Demographic factors such as small community size or being farther from a CF centre are associated with higher hospitalization rates.

3) Those with lower SES have higher hospitalization rates.

4) Individuals with CF have frequent hospitalizations, primarily for respiratory-related complications.

5) Hospitalization rates change over time.

6) Newly developed algorithms applied to administrative data are able to identify individuals with CF accurately.

7) Administrative data sources capture hospitalizations more comprehensively than the CF clinical registry.

1.8.3 Research questions

The study questions that follow from the specific hypotheses are:

1) Is there a gender difference in hospitalization rates after adjusting for markers of disease severity?

2) What clinical, demographic, and social factors are significantly associated with hospitalization?

3) What proportion of hospitalizations are related to pulmonary complications?
4) Has the hospitalization rate for pulmonary complications changed over the 10 year study period?

5) How well do administrative data algorithms compare with the CF clinical registry for identifying individuals with CF in Ontario?

6) How do hospitalization counts compare between health care administrative data and the CF clinical registry?
2 METHODS

2.1 Data sources

This study utilizes two data sources, and capitalizes on the strengths of each. The first, the Canadian CF Patient Data Registry (CPDR), contains detailed clinical information on all CF patients receiving clinical care at one of 39 accredited CF centres across Canada, including 10 Ontario CF clinics. The second comprises linked administrative databases containing information on all publicly reimbursed health care services delivered in Ontario.

2.1.1. Canadian CF Foundation Patient Data Registry

The CPDR, established in the early 1970s, collects demographic and clinical data on an annual basis on all Canadians with CF attending one of 39 specialized CF clinics in Canada. It is an important resource that enables one to track epidemiologic trends in the CF population over time. In order for CF clinics to be accredited by the CCFF they must fulfill certain criteria. Accreditation involves agreeing to a site visit from the CCFF every 3-5 years, completion of a clinic incentive grant application annually which outlines in detail the clinic patient population, staff, costs and programs, and finally, agreeing to submit clinical data to the CPDR annually on all CF patients attending the clinic. In order for individuals to be included in the CPDR, they must meet strict published guidelines for CF to ensure that there are virtually no false positives included in the registry. Once the diagnosis is confirmed the patient’s data are entered into the registry. In the rare circumstance that a patient gets “undiagnosed” based on newer diagnostic tests, these individuals are identified within the CPDR as “deemed not to have CF” and can be excluded from any analysis. Most CF medications, when prescribed through a designated CF center, are reimbursed by a special provincial CF drug program. Because of this incentive, it is assumed that most of CF patients
are followed by one of these 39 programs within Canada and hence have data within
the CPDR.

Each CF clinic manually or electronically submits data on all patients attending the
respective clinics, and this information is then uploaded into the CPDR database (see
Appendix for CPDR data collection form). Data undergoes data validation checks to ensure
that they are of “reporting” quality prior to being used for reports or research. Unusual values
are flagged and then checked against the original source. Certain variables are documented at
baseline, and held constant, including patient name, date of birth, date of death, CF genotype,
symptom(s) at initial diagnosis, race and neonatal complications. Other variables vary by
year, such as the results of sputum cultures, pulmonary function testing, nutritional
assessment, and CF complications in that given year. The registry records clinical data on
variables such as height, weight, and lung function from the first clinic visit of the reporting
year. In an effort to better characterize sputum bacteriology of the CF population, clinics
reported any positive culture in the year for a variety of common CF bacteria rather than the
result of the 1st sputum culture. The CPDR currently contains over 81,000 records on over
5,500 individuals with CF – past and present – who have attended a CF clinic in Canada. In
2002, privacy legislation required that individuals with CF must give written consent for
their data to be entered into the CPDR. Obtaining ethics approval and patient consents has
been a prolonged process which was not complete at the time the CPDR data was requested
from the CCFF for this thesis. Because of this, only CPDR clinical data until 2002
(inclusive) were available for this thesis project.

In order to obtain access to the CPDR data for this dissertation, an application was
made to the Medical Scientific Advisory Committee at the CCFF outlining the study
proposal. This application underwent a review process and all queries and concerns were addressed prior to releasing the data. The CCFF agreed to release the identified CPDR data for this research project. Patient names were necessary in order for the registry to be linked to administrative databases outlined below. Once the data linkage was completed, all names were removed from the dataset before being released. In addition, representatives from the CCFF conducted a site visit to the Institute for Clinical Evaluative Sciences (ICES) where the CPDR data would be housed and the research would be conducted.

2.1.2 Health Care Administrative Data Sources

The Institute for Clinical Evaluative Sciences (ICES) is an independent, non-profit organization that conducts research on many important health-related issues for Ontarians. Since its inception in 1992, ICES maintains multiple administrative databases. Databases used for this dissertation include the Registered Persons Database (RPDB), Discharge Abstracts Database administered by the Canadian Institute for Health Information (CIHI-DAD), the Ontario Home Care Administrative System (OHCAS) Database and the Ontario Health Insurance Plan (OHIP) Database. ICES is named as a “prescribed entity” under Ontario’s privacy law in section 45 of the Personal Health Information Protection Act (PHIPA). Under this designation, ICES can utilize health information without individual consent for purposes of analysis and compiling statistical information about health care delivery. In order to have this privilege, ICES must have policies and practices which are approved by the Information and Privacy Commissioner of Ontario.

The RPDB from the Ministry of Health and Long-Term Care lists all people eligible for an OHIP card. This database includes name, date of birth, sex, and home postal code.
Because the Ministry is not always informed when someone moves, leaves Ontario or dies, the postal code information contained in the database may not always be up-to-date.

The CIHI-DAD data details all hospitalizations within Ontario. This information is collected by trained health records personnel at each hospital who code the data, including admission and discharge dates, diagnoses, as well as diagnostic and procedure information during the hospital stay. Until 2002, up to 16 discharge diagnoses according to the International Classification of Disease, Ninth Revision (ICD-9) were used. In addition to a most-responsible diagnosis accounting for the majority of the patient’s length of stay, diagnoses were classified by pre-admission co-morbidities, in-hospital complications, or secondary diagnoses not contributing to the length of stay. Same day surgery data are included between 1991 and 2002. The database is updated annually, with a lag time of up to 9 months after the end of the fiscal year.

The OHIP database contains records of all Ontario physician billings for fee-for-service reimbursement to the provincial insurance program for consultations, assessments, visits, as well as diagnostic and therapeutic procedures for both in-patient and out-patient claims. This database is available from 1991 onwards and has been used extensively to document health care utilization across the province for a variety of diseases. Data fields include date of service, a fee code, a diagnostic code, and the billing physician’s anonymized identity. Limited claims data may be available for those physicians who practice in a non-fee-for-service payment scheme such as an alternative funding plan (AFP). “Shadow billing” is expected in these circumstances; however, it may not be complete particularly if the physician’s practice is within a hospital setting. With respect to shadow billing, a registration form is completed which outlines the services rendered for outpatient and inpatient visits.
This is submitted to the Ministry of Health, such that there is a record of the care provided by physicians who are enrolled within an AFP. According to ICES data, approximately 20% of paediatricians are on a non-fee for service plan while this applies to only 3-7% of respirologists. The OHIP database is updated monthly, although claims may not be submitted for up to six months after the service was delivered.

The OHCAS database documents all home care claims for Ontarians. Data fields include service date, service type, and home care program.

2.2 Study Population

All persons diagnosed with CF who are captured within the CPDR, and who are successfully linked to the Ontario administrative databases, were eligible for the study.

2.3 Time Period

The study period was from 1993 to 2002.

2.4 Data Linkage

The CPDR was linked to ICES administrative data to newly create the Ontario CF Database (OCFD).

2.4.1 Identifying duplicate records within CPDR

The first step in the linkage process involved evaluating the CPDR registry for duplicates. A file un-duplication of the CPDR database was conducted to detect duplicate patient records using the AutoMatch probabilistic record linkage system on a Sun Unix computer platform. Linkage parameters were designed and coded in a format acceptable to AutoMatch. A seven-pass linkage was undertaken to maximize the search for duplicate patient records. The selected blocking variables (i.e. data fields which limit the number of
comparisons by comparing only records agreeing exactly on a given value of a blocking variable) were as follows: i) the combination of surname, first given name, date of birth and gender; ii) the combination of surname initial, first given name initial, date of birth and gender; iii) the combination of New York State Identification and Intelligence System (NYSIIS) phonetic code, date of birth and gender; iv) the combination of surname, first given name initial and gender; v) the combination of alternate surname, first given name initial, and gender; vi) the combination of surname initial, first given name initial, birth year, and gender and vii) date of birth. A visual inspection of all potential duplicate patient records was conducted to validate the results. Each unique individual was assigned a newly created linkage ID number.

2.4.2 Identifying CPDR CF patients within RPDB

Once the dataset was cleaned of duplicate records, the entire CPDR was probabilistically linked to the RPDB using AutoMatch matching software. The selected blocking variables (i.e. data fields which limit the number of comparisons by comparing only records agreeing exactly on a given value of a blocking variables) were as follows: i) the combination of surname, first given name, date of birth and gender; ii) the combination of the first 3 characters of the surname, first given name initial, date of birth and gender; iii) the combination of NYSIIS (surname phonetic) code and date of birth; iv) date of birth only; v) the combination of surname, birth month and birth day; vi) the combination of surname and first given name. This process identified all individuals in the CPDR who were eligible for an OHIP number and had a valid encrypted version of their health care number, also known as IKN number.
In order to assess the linkage rate, the number of positive matches to RPDB was compared to the number of individuals who were identified in the CPDR as being followed at an Ontario clinic at any point in their care. The CPDR contains information on all CF patients followed at a Canada CF clinic; therefore, individuals were identified as followed in Ontario if any records were from reporting centres located in the province of Ontario. The linkage rate was defined as:

\[
\text{% linkage} = \frac{\text{# individuals linked to RPDB}}{\text{# individuals followed at an Ontario CF centre}} \times 100
\]

Once the linkage was complete, patient names were removed from the newly created cohort, and individuals were identified using their IKN, thereafter. Individuals were then deterministically linked between all other administrative databases (i.e., OHIP, CIHI-DAD, OHCAS) using their IKN.

2.5 Assessing “agreement” between the CPDR and ICES databases for the diagnosis of CF

There are no validated algorithms using administrative data to identify individuals with CF. Completed work in diabetes mellitus research at ICES has previously validated an algorithm using administrative data for identifying persons with diabetes in Ontario and this algorithm has been applied to the National Diabetes Surveillance System. Variations on a similar algorithm were used in this study.
The diagnostic code for CF (277.x) also applies to other conditions (Table 1). Because of this, it was felt that using the “277” code alone for OHIP claims would lack specificity. In an attempt to increase specificity, an additional code for spirometry or oximetry was included. Because infants and young children cannot perform traditional spirometry, a chest x-ray was also included in the algorithm. Therefore, a diagnosis of CF was defined as "anyone with two or more OHIP claims bearing a diagnostic code for CF (277.x) AND a diagnostic code for spirometry/oximetry (J301, J304, J323) or chest x-ray (X090, X091, X092) within a 24 month period, OR one or more hospitalizations at any point utilizing CIHI-DAD code of ICD-9 277.0". Different algorithms were tested in order to identify that which maximized the accurate detection of individuals with CF within administrative data.

Typically when evaluating a new test, the outcome of the test is compared to the current gold standard using subjects who are representative of the intended population which contains individuals both with the condition and without. From these data, the sensitivity and specificity of the new test are calculated. There is no true gold standard to assess the ability of administrative data to identify individuals with CF. Validating algorithms using chart abstraction is not feasible for CF due to its low prevalence of the disease and the fact that CF patients are typically seen at specialized CF centres. Because the CPDR contains only subjects with CF, there is no reference standard for not having the disease. For these reasons, it was not possible to calculate the specificity of the administrative data using the CPDR as a reference. With respect to assessing the agreement between the two methods, a traditional kappa statistic could not be used to assess agreement between the CPDR and administrative data for several reasons. As mentioned, the CPDR only captures individuals with the disease
therefore there is no ‘gold standard’ for not having CF. As such, cell ‘d’ cannot be accurately populated (see Figure 2). Secondly, given the low prevalence of CF, the true negative cell will have an exceptionally large number of individuals within it, and hence, will overwhelm the kappa statistic, making it difficult to interpret. This may create a falsely high kappa value, and could conceal significant disagreement between methods in the identification of patients with the disease and thus was not used.

Table 1: Definition of ICD-9 codes for 277.x

<table>
<thead>
<tr>
<th>ICD-9 CODE</th>
<th>DEFINITION</th>
</tr>
</thead>
<tbody>
<tr>
<td>277.0</td>
<td>Cystic Fibrosis with or without ileus</td>
</tr>
<tr>
<td>277.1</td>
<td>Disease Porphyrin Metabolism</td>
</tr>
<tr>
<td>277.2</td>
<td>Purine/Pyrimid Disease NEC</td>
</tr>
<tr>
<td>277.3</td>
<td>Amyloidosis</td>
</tr>
<tr>
<td>277.4</td>
<td>Disease Bilirubin Excretion</td>
</tr>
<tr>
<td>277.5</td>
<td>Mucopolysaccharidosis</td>
</tr>
<tr>
<td>277.6</td>
<td>Deficiency Circulating Enzyme NEC</td>
</tr>
<tr>
<td>277.8</td>
<td>Metabolism Disorder NEC</td>
</tr>
<tr>
<td>277.9</td>
<td>Metabolism Disorder NOS</td>
</tr>
</tbody>
</table>

Instead, only the sensitivity of the administrative data was calculated and the level of agreement between the CPDR and the administrative databases in their ability to identify individuals with CF was assessed using the approach suggested by Gordis. This approach disregards persons labeled as negative by both methods (cell ‘d’) as this number may be extremely large and thereby may conceal significant disagreement between the observers in
identifying subjects whom at least one observer considers positive. Instead the positive percent agreement was calculated, limiting the denominator to those persons with a positive diagnosis in one or both diagnostic approaches (i.e. cells a, b and c in Figure 2). The following equation demonstrates Gordis’ proposed approach:

\[
\text{% agreement} = \frac{a}{a + b + c} \times 100
\]

**Figure 2**: 2 x 2 table comparing the Canadian Patient Data Registry and Healthcare Administrative Databases

<table>
<thead>
<tr>
<th>Healthcare Administrative Databases</th>
<th>CF Present</th>
<th>CF Absent</th>
</tr>
</thead>
<tbody>
<tr>
<td>CF Present</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>CF Absent</td>
<td>C</td>
<td>D</td>
</tr>
</tbody>
</table>

One must recognize, however, some individuals may exist within the administrative databases who truly have CF, but who are not followed in the CPDR, as they never enrolled at an accredited CF centre. Those at particular risk of this would be individuals with “milder” disease (i.e. pancreatic sufficient, minimal lung disease etc) or individuals who are diagnosed late in life. Where discrepancies were found between the two datasets, the individual cases
were examined in detail, specifically demographic variables (i.e. gender, age), details of administrative records, and clinical variables, where available.

All individuals found in both the CPDR and in the administrative data comprised the linked study cohort. The linked dataset is called the "Ontario CF Database" (OCFD). This dataset was used for subsequent analyses.

2.6 Creating clinical cohort of CF subjects using the Ontario CF Database

The CPDR typically records one observation per year on each individual with CF. Occasionally, individuals within the CPDR are seen at more than one CF centre within a given year. This is particularly true of students who may receive their primary CF care in one centre but attend a different CF centre for a temporary duration while in another city attending school. In these circumstances, there may be two or more records within the CPDR on an individual as each centre reports on all patients that attend the clinic in that year. In order to create one record per year for each subject, the programming was such that the first observation per annum for that individual was selected. However if there were missing clinical data, the newly created program would loop through the next observation in that year for that subject, filling in missing data for that individual until the clinical data was as complete as possible for that year.

The study period was 10 years. Some individuals were followed for the full 10 years while others entered and exited the dataset at some point over the study period. If individuals died, moved out of the province, or were diagnosed with CF part-way through the year, they may not have been “at risk” for the outcome (hospitalization). In order to address this issue, a “start year” and “end year” variable was created for all subjects. The “start year” was defined
as either 1993 or the year they were diagnosed with CF (if diagnosed after 1993). If date of
diagnosis was missing in the CPDR, then year of the birth date was used as the year of
diagnosis. The “end year” was defined as either 2002 or the year they died (if subject died
prior to 2002). Individuals in the cohort who died or were transplanted before the “start year”
were excluded. A “year” variable was created for each year the subjects should have been
followed between the start year and end year. In order to account for partial years, an “offset”
variable was created which represents the number of days for each year that the individual
was eligible for the outcome, i.e. hospitalization. The offset variable for the start and end
year may be less than 365 if the subject died or was diagnosed part way through a year. Once
a subject received a lung transplant, their data after that date were not included in the
analysis. This is due to the fact that after transplant, lung function dramatically changes and
the reason for hospitalization is significantly different than they are pre-transplant, thereby
potentially distorting any relation between clinical parameters and hospitalization.

2.7 Outcome Definition: Hospitalization

The main study outcome was respiratory-related hospital admissions. The subjects in
the OCFD were linked to CIHI-DAD to capture all hospitalizations. Respiratory-related
hospital admissions were defined as involving the respiratory system by looking at the ICD-9
code associated with the most responsible diagnosis. This is the diagnosis that contributes
most to the length of stay. A total of 110 ICD-9 codes were used to define respiratory
complications (see Appendix x). In order to exclude individuals who were not admitted with
a significant respiratory-related complication, individuals with a length of stay of less than 3
days were excluded unless there was an associated home care claim within 1 week of
discharge with no homecare claims within the preceding 4 weeks prior to the hospitalization
date. If however, there was a hospital admission of less than 3 days with an associated homecare claim within 7 days following discharge from hospital, this was assumed to represent initiation of home IV antibiotic therapy for a pulmonary exacerbation and was included in the total number of respiratory-related hospitalizations. In order to capture home IV antibiotic therapy initiated directly from the outpatient clinic, physician claims within the OHIP database were identified. An outpatient clinic visit was defined as: physician assessment (using A473, 474, 475, 478, 135, or 138) and spirometry or oximetry (J304 or 323) billed on the same service date. If a homecare claim was registered within 7 days of the clinic visit date, with no homecare claims prior to that clinic date, this was assumed to represent an individual who was started on home IV antibiotic therapy directly from clinic without a hospital admission. This was included in the total “hospitalization” count since it was assumed to represent treatment for a pulmonary exacerbation and prior to the availability of homecare, this would have represented a respiratory-related hospitalization. Hospitalizations that occurred prior to the diagnosis date for CF were not included in the total number of hospitalizations. The number of hospitalizations for respiratory-related complications per year was then calculated for each subject.

2.8 Predictor Variables

2.8.1 Clinical data

Clinical data on date of birth, sex, date of diagnosis, date of death, pancreatic status, pulmonary function, height, weight, and sputum bacteriology were available on study subjects. “Age” for a given year was calculated by subtracting the date of birth from January 1st of the given year. The CPDR dataset records lung function, specifically forced expiratory volume in the first second (FEV$_1$) and forced vital capacity (FVC) in litres. The Wang and
Hankinson pulmonary function equations for children and adolescent/adults were used to calculate the predicted norms for FEV$_1$ and FVC for each subject.$^{85,86}$ The percent predicted FEV$_1$ was calculated using the following equation: \[ \frac{\text{FEV}_1(L)}{\text{predicted FEV}_1(L)} \times 100. \] A similar process was carried out for FVC. Anthropomorphic measurements were used to assess nutritional status. Body mass index (BMI) was calculated for those over the age of 19 years using the following equation: \( \text{weight (kg)} / \text{height}^2 \text{ (m)}. \) Because BMI in children and adolescents varies with sex and age, growth charts from the Centres for Disease Control (CDC) were used to calculate the age and sex adjusted z-score BMI for subjects between 6 and 19 years of age.$^{87}$ In order to adjust for a cohort effect, a variable called “birth year” was created using the year of the subject’s date of birth. Calendar year was the reporting year. In terms of CF-related complications, specifically, pancreatic status and the presence of CFRD, these were documented in the CPDR. Once a subject was documented to have CFRD or pancreatic insufficiency, all subsequent years were categorized as such. With respect to microbiology, PsA and BCC were identified as present or absent in the CPDR dataset in two ways. A new variable was created documenting if subjects had “ever” or “never” grown the bacteria. Alternatively, the bacteria status (present or absent) for each year was recorded.

2.8.2 Missing data

Missing data were dealt with in various ways. If a patient was not seen in a given year, then the clinical parameters that change from year-to-year (pulmonary function, height, weight, bacteriology) were classified as missing for that given year. Individuals who had no recorded clinical parameters for lung function or nutritional status for any of the years they were followed were deleted from the analysis. For individuals who were 19 years of age or older, the subject’s mean height was calculated and this value was used to fill in missing
values for that individual. Regarding missing height data, for subjects who were less than 19 years of age, an age-adjusted z-score for height was calculated using the CDC growth curves for males and females. The mean z-height was calculated for each individual. If height was missing, the mean z-score for height was used and converted back to a height measurement in centimetres using CDC growth curves. For years that had missing weight and lung function measurements, multiple imputation, as first described by Rubin, was used to estimate the missing values prior to the final analysis. Bacteriology was limited to the presence or absence of PsA and BCC. If sputum results were missing for a given year, individuals were considered positive for that organism if they had ever grown it during the study period. Alternatively, the previous sputum culture result was used to replace missing values. The analyses were conducted using both ways to document microbiology.

2.8.3 Socioeconomic Status

SES was based on neighbourhood income quintiles using postal code information from the 1996 Canadian census and Statistics Canada data.

2.8.4 Rural versus urban

Each subject’s postal code was linked to census data. The definition of rural and urban area was dependent on the size of the community. Community size was defined in terms of the 2001 census population in each census metropolitan area or census agglomeration. If the population in a community was less than 10,000, it is considered rural. Otherwise, it was categorized as urban.

2.8.5 Distance from CF centre

Distance between the subjects and the CF centre involved calculating the shortest distance in kilometers using longitude and latitude measurements. These were calculated
using the subject’s postal code and the CF reporting centre postal code. This method doesn’t take into consideration roadways or local geography. It should also be noted that patient location may be misclassified in some rural areas because postal codes cover a large geography and latitude and longitude are assigned to the centre of the postal code. In an attempt to categorize distances in a meaningful way, four groups were chosen based on the type of commute that one would have to take in order to get from home to clinic. The first category of less than 50 km for a one-way trip was chosen as many individuals commute this distance routinely on a daily basis. An individual could travel this distance by taking public transit, a car, a bike or even on foot if they lived close enough. The second category, 50-199 km for a one-way trip, is a more substantial commute but a round trip could be done in a day trip, if required. Category 3 (200-400 km one-way trip) is unlikely to be done in one day but more likely requires an overnight stay. The farthest distance of over 400 km (one-way trip) may require an overnight stay or even a plane ride.

2.9 Statistical Analyses

2.9.1 Sample Size Estimation

Because there is a fixed sample size of CF patients within the CPDR, a power calculation could be performed for this study. Based on early CPDR national statistics, it was estimated that there would be approximately 1,200 individuals with CF followed in Ontario: 648 males and 552 females. Based on the literature, the estimated proportion of males hospitalized per year is 0.40, and for females 0.50. Using a two-tailed test, with an alpha = 0.05 and a beta = 0.10, one would need about 519 individuals per group to detect a 0.10 difference in hospitalization rates between males and females. Given the fact that there are
648 males and 552 females, there is a 91% chance of detecting as little as a 0.10 difference between males and females in the annual rate of hospitalization.

2.9.2 Analyses

2.9.2.1 Agreement Between Administrative Data and CPDR

Approach to assessing agreement by calculating sensitivity, positive predictive value and percent agreement previously outlined in section 2.5.

2.9.2.2 Multiple Imputation

As previously noted, missing weight and lung function data were estimated using multiple imputation (MI) in the analysis phase of the study as first described by Rubin. Several assumptions are made when using MI; specifically, the mechanism of missingness is either missing completely at random (MCAR) or missing at random (MAR) and the variables are normally distributed. MCAR requires the important assumption that there are no differences between subjects with complete data and subjects with missing data, such that the probability of having missing data for a given variable is the same for each subject, and is unrelated to any measured or unmeasured variables. When the mechanism of missingness is MAR, the probability of a variable being missing for a specific subject is related to other measured variables, but is independent of the variable itself.

MI uses the relationships between non-missing values to generate a predictive distribution for those missing values, using Bayesian methods. Typically, a non-informative prior distribution, implying that one has no strong preconceived ideas about the shape and location of the distribution, is used for a variable of interest. This is then combined with a likelihood function to create a posterior distribution for the missing variable. Next, multiple random draws are generated from the newly created posterior distribution, producing an
imputed value for each subject who originally had missing values. By creating a set of plausible values for each subject, this incorporates the uncertainty that is inherent in any observed values. One must specify the variables to use in the data model, selecting specific types of variables: 1) target variables, 2) auxiliary variables and 3) sample design variables. The target variables are those that will be included in the final statistical model, including the outcome variable. Auxiliary variables are those that are predictive of the target variables or associated with the missingness, but don’t necessarily need to be included in the final statistical model. Finally, sample design variables are those that may be unique to the dataset, such as variables denoting clusters within the data. By using an inclusive approach in the data modeling stage of MI, one can preserve the associations between variables which are essential to obtaining unbiased estimates with this technique. Once the imputed datasets are created, standard statistical procedures are used to analyze each dataset. Parameter estimates from each complete dataset are then averaged over “m” number of datasets. The “within imputation” variance, “between-imputation” variance, and “total variance” variance was calculated using information from all imputed datasets.92

2.9.2.3 Main Analysis of Hospitalization Rates

General descriptive statistics were used to characterize the study population. In particular, gender differences in the annual hospitalization rates were calculated. Baseline characteristics were evaluated with categorical variables assessed using a Chi Square test, and continuous variables were compared using an unpaired t-test.

In order to analyze data over multiple years on individual subjects, Poisson regression utilizing generalized estimating equations (GEEs) were used to account for the correlated nature of the data.93,94 This statistical method is an extension of generalized linear models to
account for situations whereby the observations are not independent of one another in the longitudinal nature of this study design. Using an iterative process, GEEs optimize the weights for individual clusters (in this case subjects followed over multiple years), by minimizing the weighted sum of the squared residuals.\textsuperscript{93} Weights and correlations are re-estimated after each iteration, and parameters are re-fitted correcting for this correlation. This process repeats itself until the parameter estimate stabilizes. The GEE method focuses on within-cluster rather than between cluster variation. With GEEs the correction for within-cluster correlations requires an a priori assumption of the correlation structure (independent, exchangeable, autoregressive, or unstructured). However, GEE analysis is robust against an incorrect selection of correlation matrix. For this analysis, an autoregressive correlation structure was specified which indicates that 2 observations taken close in time within an individual tend to be more highly correlated than 2 observations taken far apart in time from the same individual. In addition, Poisson regression assumes that the variance is equal to the mean; however, if there is a large degree of dispersion of the data this assumption no longer holds true. Over- and under- dispersion was evaluated by analyzing the data using both the Poisson distribution and the negative binomial distribution and assessing the goodness of fit of the models. Finally, because of the large number of zeros in the dataset, the zero-inflated Poisson regression model was also evaluated for goodness of fit.

Linearity of continuous variables was assessed using a quadratic term and/or categorization to assess for a threshold effect. A two-tailed p value < 0.05 was considered statistically significant. All analyses were done using SAS 9.1 statistical software.
2.10 Ethics

The research study received ethics approval from the Research Ethics Board of Sunnybrook Health Sciences Centre, St. Michael’s Hospital, SickKids Hospital, and the University of Toronto. Prior to 2002, written informed consent was not required for the CPDR. This study only uses data up until 2002 inclusive as the process of obtaining informed consent was not completed at this time and CPDR data beyond 2002 had not yet been collected. Individual informed consent was not required in order to use the ICES administrative databases as outlined in PHIPA.
3 RESULTS

3.1 Data Linkage

The un-duplication procedure of the CPDR yielded 6 duplicate records (< 1%) resulting in 4,238 unique subjects followed in the registry. All 4,238 individuals within the CPDR dataset were probabilistically linked to the Ontario RPDB, resulting in 1,497 successful matches. Separately, the CPDR was evaluated to calculate the number of subjects followed at an Ontario CF centre. There were 1,580 individuals identified within the CPDR who had been followed during the study period in an Ontario CF centre. Comparing the number of individuals at Ontario CF centres and the number of individuals found in RPDB corresponds to a linkage rate of 95% (Figure 3).

3.2 Study Cohort

Of 1,580 subjects identified as being followed in Ontario, 1,464 (93%) had a valid IKN and therefore had data regarding health services utilization over the study period. Details on how the study cohort was created can be seen in Figure 4. Seventy-nine subjects were excluded because either death or transplantation occurred prior to the start year (n=3), date of last contact in administrative data prior to the start year (n=13), no lung function or nutritional status data recorded for any of the years followed (n=47), subjects were subsequently found not to have CF and were recorded as such in the CPDR (n=8) and the sole CF reporting centre was outside of Ontario (n=8). Reliable measures of lung function are not available in individuals under the age of 6 years, therefore these subjects were excluded from the analysis (n=211), leaving 1,174 subjects for the final analysis.
3.3 Patient Characteristics

The cohort included 531 (45%) females and 643 (55%) males. Forty-three percent of the cohort was followed for the full 10 years and 74% were followed for 5 years or more. The dataset contained a total of 8,305 patient-years of data. Table 1a and 1b describes the baseline characteristics of the cohort by age, gender and year. Of note, the proportion of subjects with pancreatic insufficiency is higher in the younger cohort compared to the adults. This may represent a combination of a survivor effect and late-adult onset diagnoses of milder forms of CF. Other differences between the two age cohorts worth noting include the higher prevalence of CFRD and BCC infection in the older age group both of which are generally more common with increasing age.

Ages 7-19 years

No significant differences between males and females were noted in FEV₁ in any of the years followed (Table 1a). Growth for the cohort was normal with all years having a mean BMI z-score between 0 and -0.5. For all years studied, females had the same or slightly higher mean BMI z-scores compared to males with no statistically or clinically significant differences. CFRD, on the other hand was consistently found more frequently in females compared to males across the years. The rate of CFRD increased by 10% per year even after adjusting for age and sex (p=.001). Age and sex were independent predictors of CFRD (p<.0001 for both). There were no differences in microbiology between males and females in any of the years studied. As expected, annual proportion of subjects ever having a positive culture for PsA was high (range 55 to 90%) whereas annual percentage of patients with BCC was much lower (range 3-25%). The rate of BCC decreased by 25% per year after adjusting for age (p<.0001). The rates of PsA decreased by 2% per year (p<.0001) after adjusting for
age. Age was an independent predictor of both BCC and PsA (p<.0001). With respect to missing lung function and nutritional data, males were missing significantly more data than females for some of the years followed. No consistent pattern was seen across the study period.

**Over 19 years of age**

No differences were noted in lung function between males and females across the years studied (Table 1b). The mean BMI for both males and females indicated adequate nutritional status for all years studied. In contrast to the younger cohort, those over the age of 19 years did show differences in nutritional status with women having a lower BMI than males for all years. More females were noted to have CFRD across the study period however this was statistically significant only in the latter half of the study period (1999-2002). The rate of CFRD did not change significantly over time in the adult cohort (p=.12). Infection with PsA was seen more commonly in females but this only reached statistical significance in 4 of the 10 years followed and did not seem to follow a consistent pattern. Interestingly, the reverse was true of BCC with a higher proportion of males having positive sputum culture for this organism but this only reached statistical significance in the final two years of the study period perhaps representing a survival effect. Overall, the proportion of individuals with PsA ranged from 69% to 84% which was similar to younger age group while for BCC the percentage was higher in adults (range 15% to 44%). The rate of BCC decreased by 4% per year (p=.007). Age and sex were not significant predictors of rate of BCC. The rate of PsA increased over time (p<.005) after adjusting for sex and age. Age was a significant predictor of PsA (p=.002) while sex was of borderline statistical significance (p=.06). There were no significant gender differences in missing data with respect to lung function. Only in
2 of the 10 years was a there a statistically significant difference seen between the genders with respect to missing BMI. In 1993, females were missing more BMI data whereas the reverse was true in 1998.
Figure 3: Flowchart of Data linkage of Canadian Patient Data Registry (CPDR) with Healthcare Administrative Database

**Subjects linked to Administrative Database**

- **4238** unique subjects followed in the Canadian Patient Data Registry (CPDR)

  Probabilistic linkage using Name, DOB, gender

- **1497** CF subjects linked to Registered Persons Database which represents individuals with an **ONTARIO** health card

**Subjects identified as being seen in Ontario within CPDR**

- **4238** unique subjects followed in the Canadian Patient Data Registry (CPDR)

- **1580** subjects identified in CPDR as followed at an **ONTARIO** CF centre at some point during their care

**95% LINKAGE RATE**

\[
\frac{1497}{1580} \times 100
\]
Figure 4: Flowchart of creation of study cohort

4238 unique subjects identified in CPDR

1580 subjects identified in CPDR as followed at an ONTARIO CF centre at some point during their care

Excluded 116 subjects with no IKN:
- Out of province
- Name change
- Mismatch

1464 ONTARIO CPDR pts

Excluded 79 subjects:
- 1 Died before start year
- 2 transplanted prior to start year
- 8 excluded because primary CF reporting centre outside Ontario
- 8 excluded because date of last contact in admin data was prior to start year
- 47 subjects never had lung function data for any years followed
- 8 coded within CPDR as “no longer have CF”

1385 ONTARIO CPDR pts

Excluded 211 subjects:
- ≤ 6 years of age

1174 CF subjects Used in final analysis
### Table 1a: Descriptive data for participants between 7-19 years of age (inclusive)

<table>
<thead>
<tr>
<th>Year</th>
<th>1993 (n=392)</th>
<th>1994 (n=411)</th>
<th>1995 (n=426)</th>
<th>1996 (n=435)</th>
<th>1997 (n=431)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>Female</td>
<td>Male</td>
<td>Female</td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>Subjects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>47</td>
<td>53</td>
<td>47</td>
<td>53</td>
<td>47</td>
</tr>
<tr>
<td>Age, years mean± SD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12.7±4</td>
<td>12.9±4</td>
<td>12.5±4</td>
<td>12.5±4</td>
<td>12.4±4</td>
</tr>
<tr>
<td>FEV1 % pred mean± SD (% missing)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>74.4±25 (5%)*</td>
<td>77.3±25 (12%)*</td>
<td>78.3±26 (7%)</td>
<td>80.4±27 (11%)</td>
<td>76.6±25 (5%)</td>
</tr>
<tr>
<td>z-score BMI mean± SD (% missing)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.32±.97 (4%)</td>
<td>-0.31±.98 (8%)</td>
<td>-0.20±.88 (5%)</td>
<td>-0.31±1.1 (8%)</td>
<td>-0.24±1.0 (3%)</td>
</tr>
<tr>
<td>Pancreatic insufficient</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>94</td>
<td>88</td>
<td>94</td>
<td>90</td>
<td>95</td>
</tr>
<tr>
<td>CF-Related Diabetes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>**</td>
<td>6†</td>
<td>**†</td>
<td>8†</td>
</tr>
<tr>
<td>P. aeruginosa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>83</td>
<td>89</td>
<td>83</td>
<td>87</td>
</tr>
<tr>
<td>B. cepacia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>25</td>
<td>15</td>
<td>21</td>
<td>13</td>
</tr>
</tbody>
</table>

*p<.05; † p<.01; ‡ p<.001 noted for differences between males and females for each year; **cells suppressed due to small numbers (≤5)

Note: For subjects, values are percentage of total subjects for that calendar year; for categorical variables, values are percentage of males/females for that calendar year; continuous variables are mean ± standard deviation.
Table 1a: Descriptive data for participants between 7 - 19 years of age (inclusive)

<table>
<thead>
<tr>
<th>Year</th>
<th>1998 (n=440)</th>
<th>1999 (n=451)</th>
<th>2000 (n=464)</th>
<th>2001 (n=475)</th>
<th>2002 (n=483)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Female</td>
<td>Male</td>
<td>Female</td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subjects</td>
<td>44</td>
<td>56</td>
<td>45</td>
<td>55</td>
<td>46</td>
</tr>
<tr>
<td>Age, years mean± SD</td>
<td>12.4±4</td>
<td>12.5±4</td>
<td>12.6±4</td>
<td>12.5±4</td>
<td>12.7±4</td>
</tr>
<tr>
<td>FEV1 % pred mean± SD (% missing)</td>
<td>75.7±24 (9%)</td>
<td>79.1±23 (13%)</td>
<td>80.7±23 (5%)*</td>
<td>82.0±23 (11%)*</td>
<td>79.8±23 (6%)</td>
</tr>
<tr>
<td>z-score BMI mean± SD (% missing)</td>
<td>-0.24±.95 (3%)*</td>
<td>-0.36±1.0 (8%)*</td>
<td>-0.14±.85 (4%)</td>
<td>-0.23±1.0 (7%)</td>
<td>-0.13±.92 (3%)</td>
</tr>
<tr>
<td>Pancreatic insufficient</td>
<td>94</td>
<td>92</td>
<td>91</td>
<td>92</td>
<td>90</td>
</tr>
<tr>
<td>CFRD</td>
<td>10†</td>
<td>**†</td>
<td>10†</td>
<td>**†</td>
<td>9†</td>
</tr>
<tr>
<td>P. aeruginosa</td>
<td>80</td>
<td>78</td>
<td>76</td>
<td>76</td>
<td>75</td>
</tr>
<tr>
<td>B. cepacia</td>
<td>7</td>
<td>9</td>
<td>6</td>
<td>7</td>
<td>5</td>
</tr>
</tbody>
</table>

*p<.05; † p<.01; ‡ p<.001 noted for differences between males and females for each year; **cells suppressed due to small numbers (≤5)

Note: For subjects, values are percentage of total subjects for that calendar year; for categorical variables, values are percentage of males/females for that calendar year; continuous variables are mean ± standard deviation.
Table 1b: Descriptive data for participants over the age of 19 years

<table>
<thead>
<tr>
<th>Year</th>
<th>1993  (n=335)</th>
<th>1994  (n=362)</th>
<th>1995  (n=381)</th>
<th>1996  (n=401)</th>
<th>1997  (n=413)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>Female</td>
<td>Male</td>
<td>Female</td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>Subjects</td>
<td>41</td>
<td>59</td>
<td>41</td>
<td>59</td>
<td>40</td>
</tr>
<tr>
<td>Age, years mean± SD</td>
<td>28.4±7</td>
<td>28.4±6</td>
<td>28.5±7</td>
<td>28.6±7</td>
<td>28.8±7</td>
</tr>
<tr>
<td>FEV1 % pred mean± SD (% missing)</td>
<td>58.8±23 (23%)</td>
<td>56.0±24 (15%)</td>
<td>59.0±24 (17%)</td>
<td>57.3±26 (19%)</td>
<td>57.3±23 (19%)</td>
</tr>
<tr>
<td>BMI mean± SD (% missing)</td>
<td>21.4±3.2† (19%)*</td>
<td>22.6±3.2† (10%)*</td>
<td>21.6±3.2† (15%)</td>
<td>22.6±3.2† (16%)</td>
<td>21.5±3.2‡ (19%)</td>
</tr>
<tr>
<td>Pancreatic insufficient</td>
<td>66†</td>
<td>79†</td>
<td>73</td>
<td>80</td>
<td>74</td>
</tr>
<tr>
<td>CFRD</td>
<td>16</td>
<td>12</td>
<td>18</td>
<td>13</td>
<td>18</td>
</tr>
<tr>
<td>P. aeruginosa</td>
<td>74</td>
<td>69</td>
<td>74</td>
<td>70</td>
<td>77</td>
</tr>
<tr>
<td>B. cepacia</td>
<td>38</td>
<td>44</td>
<td>40</td>
<td>43</td>
<td>37</td>
</tr>
</tbody>
</table>

*p<.05; † p<.01; ‡ p<.001 noted for differences between males and females for each year

Note: For subjects, values are percentage of total subjects for that calendar year; for categorical variables, values are percentage of males/females for that calendar year; continuous variables are mean ± standard deviation.
Table 1b: Descriptive data for participants over the age of 19 years

<table>
<thead>
<tr>
<th>Year</th>
<th>1998 (n=404)</th>
<th>1999 (n=418)</th>
<th>2000 (n=427)</th>
<th>2001 (n=442)</th>
<th>2002 (n=452)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sex</td>
<td>Female</td>
<td>Male</td>
<td>Female</td>
<td>Male</td>
</tr>
<tr>
<td>Subjects</td>
<td>42</td>
<td>58</td>
<td>44</td>
<td>56</td>
<td>44</td>
</tr>
<tr>
<td>Age, years mean±SD</td>
<td>30.1±8</td>
<td>29.9±8</td>
<td>30.5±8</td>
<td>30.0±8</td>
<td>31.0±9</td>
</tr>
<tr>
<td>FEV1 % pred mean±SD (% missing)</td>
<td>58.3±22 (16%)</td>
<td>57.0±25 (21%)</td>
<td>59.6±22 (19%)</td>
<td>58.6±25 (20%)</td>
<td>59.1±22 (20%)</td>
</tr>
<tr>
<td>BMI mean±SD (% missing)</td>
<td>22.0±3.5* (12%)*</td>
<td>22.8±3.5* (19%)*</td>
<td>21.8±3.3† (15%)</td>
<td>23.2±3.3‡ (18%)</td>
<td>22.3±4.1 (16%)</td>
</tr>
<tr>
<td>Pancreatic insufficient</td>
<td>75</td>
<td>80</td>
<td>74</td>
<td>79</td>
<td>74</td>
</tr>
<tr>
<td>CFRD</td>
<td>19</td>
<td>14</td>
<td>21*</td>
<td>14*</td>
<td>20</td>
</tr>
<tr>
<td>P. aeruginosa</td>
<td>84*</td>
<td>75*</td>
<td>82</td>
<td>74</td>
<td>83</td>
</tr>
<tr>
<td>B. cepacia</td>
<td>26</td>
<td>33</td>
<td>24</td>
<td>31</td>
<td>22</td>
</tr>
</tbody>
</table>

*p<.05; † p<.01; ‡ p<.001 noted for differences between males and females for each year

Note: For subjects, values are percentage of total subjects for that calendar year; for categorical variables, values are percentage of males/females for that calendar year; continuous variables are mean ± standard deviation.
3.4 Hospitalizations

3.4.1 Comparison between Health Care Administrative Data Sources and CPDR

Using CIHI-DAD, there were 7,712 hospitalizations for any reason over the 10-year study period (Figure 5). Eighty-two percent (6,311) of records were for inpatient care while 18% (1,401) were recorded in CIHI-DAD as same-day surgery. When limited to respiratory-related hospitalizations only, the number decreased to 3,901. The presumed initiation of IV therapy through homecare services identified a further 202 claims (119 new homecare claims from clinic and 83 new homecare claims started after a hospital admission of less than 3 days). These were included in the total number of respiratory-related “hospitalizations”, bringing the total number of respiratory related hospitalizations to 4,103. The diagnoses given for respiratory-related hospitalizations based on ICD-9 codes can be seen in Tables 2 & 3.

The total number of inpatient hospitalizations recorded in the CPDR for those subjects with a valid health care number was 4,534. This number includes admissions for several pre-specified conditions: pulmonary exacerbation, new diagnosis of CF, palliative care, transplant, liver disease, diabetes, other gastrointestinal disease, and other. Each admission recorded in the CPDR could be associated with 1 or more reason(s) explaining the hospitalization. A total of 811(18%) hospitalizations had no reason documented within the registry. Of those admissions that had a reason listed, pulmonary exacerbation accounted for 1,507/3,723 (40%) of the inpatient hospital admissions. An additional 219 courses of home intravenous antibiotic therapy for pulmonary exacerbations were recorded in the CPDR.
Compared to CIHI-DAD hospitalization records for inpatient care, the CPDR captured 72% (4,534/6,311) of all inpatient hospitalizations (Table 4).
Figure 5: Number of hospitalization identified using CIHI-DAD, all ages

Hospitalizations identified in CIHI-DAD (admitted for any reason)  
N=7712

- In-patient hospitalizations  
  N=6311 (82%)

  Limited to LOS ≥ 3 d and Respiratory-related

  Respiratory-related hospitalizations  
  N=3901

- Same day surgery admissions  
  N=1401 (18%)

  Limited to LOS < 3 d and Using homecare data

  Intravenous antibiotics initiated using homecare and CIHI-DAD databases  
  N=83

  Home intravenous antibiotics initiated from outpatient clinic setting using OHIP and homecare databases  
  N=119

  Respiratory-related hospitalizations  
  N=3984

- Respiratory-related hospitalizations  
  N=4103
Table 2: Respiratory-related hospitalizations identified using CIHI-DAD and Homecare Database

<table>
<thead>
<tr>
<th>Reason for Hospitalization (based on ICD-9 code used for most responsible diagnosis)</th>
<th>Number (%) (n=4103)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cystic Fibrosis</td>
<td>3213 (78.3)</td>
</tr>
<tr>
<td>Infections</td>
<td></td>
</tr>
<tr>
<td>Bacterial*</td>
<td>549 (13.4)</td>
</tr>
<tr>
<td>Viral</td>
<td>137 (3.3)</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td></td>
</tr>
<tr>
<td>Pneumothorax</td>
<td>56 (1.4)</td>
</tr>
<tr>
<td>Lung disease</td>
<td>59 (1.4)</td>
</tr>
<tr>
<td>Asthma</td>
<td>60 (1.4)</td>
</tr>
<tr>
<td>Other</td>
<td>29 (0.7)</td>
</tr>
</tbody>
</table>

* 119 events identified using the outpatient homecare algorithm were categorized as “bacterial infections” since these events were presumed to be initiation of home intravenous antibiotics for pulmonary infectious exacerbations.

Table 3: Additional diagnostic codes used to explain hospitalizations when “cystic fibrosis” was used as the most responsible diagnosis

<table>
<thead>
<tr>
<th>Reasons for Hospitalization (based on all ICD-9 diagnostic codes used)</th>
<th>Number (%) (n=3213)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cystic Fibrosis only</td>
<td>450 (14)</td>
</tr>
<tr>
<td>Cystic Fibrosis and diabetes only</td>
<td>64 (2)</td>
</tr>
<tr>
<td>Lung Disease*</td>
<td>1992 (62)</td>
</tr>
<tr>
<td>Other**</td>
<td>707 (22)</td>
</tr>
</tbody>
</table>

*includes infection, asthma, respiratory failure, pneumothorax, hemoptyis
**includes pancreatic disease, malnutrition, weight loss, cirrhosis
Table 4: Comparison of number of hospitalizations using CF Patient Data Registry versus CIHI-DAD

<table>
<thead>
<tr>
<th></th>
<th>All Hospitalizations</th>
<th>Respiratory-related hospitalizations</th>
<th>Initiation of home IV therapy for presumed pulmonary exacerbation</th>
<th>Respiratory admissions and home IV therapy</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIHI-DAD</td>
<td>7712</td>
<td>3901</td>
<td>202</td>
<td>4103</td>
</tr>
<tr>
<td>CF Patient Data Registry</td>
<td>4534*</td>
<td>1507*</td>
<td>219</td>
<td>1726</td>
</tr>
</tbody>
</table>

*No reason for hospitalization given in 811 (18%) of admissions therefore respiratory-related hospitalizations may be underestimated.

### 3.4.2 Respiratory Hospitalizations

Although 110 possible ICD-9 codes were available to capture respiratory-related admissions, only 65 (59%) were actually utilized as the most-responsible diagnosis (MRD) for the hospitalization using the study algorithm. The most common ICD-9 code used as the MRD for hospitalizations was ‘cystic fibrosis’, accounting for 78% of the admissions. ICD-9 codes related to specific infections accounted for a further 17% of admissions. Diagnostic codes pertaining to the respiratory system were found in 62% of those admissions where ‘cystic fibrosis’ was listed as the MDR (Table 3). Cystic fibrosis or cystic fibrosis and diabetes were the only diagnostic codes listed for 16% of the hospitalizations.

For those 7 years of age and older (n=1,174), a total of 6,194 hospitalizations were identified in CIHI-DAD, 4,931 (80%) were inpatient hospital admissions and 1,263 (20%)
were flagged as same-day surgery. Out of 4,931 hospitalizations, 3,676 (75%) were deemed to be respiratory related. Four hundred and ninety-five subjects had no hospitalizations while 679 subjects were hospitalized 1 or more times. The range of hospital counts per year was 0 to 15. Ninety-five percent of annual observations had between 0 and 2 hospital counts.

All analyses were carried out using Poisson GEE regression as there was no evidence of under- or over-dispersion (dispersion factor=0.94; values close to 1.0 suggest no evidence of under or over-dispersion). The binomial portion of the zero-inflated model was non-significant suggesting the dataset did not have an excessive number of zero counts.

Hospitalization rates were higher for females and older subjects (Figure 6). Over the study period, males had 1,733 hospitalizations while females had 1,943 hospitalizations. The unadjusted overall hospitalization rates were 0.42 hospitalizations per person-year for males compared with 0.60 for females (rate ratio [referred to as RR] 1.40, 95% CI 1.16-1.68; p=0.0004). Overall, hospitalization rates increased significantly over time even after adjusting for age and sex (p=0.04). However, there was no significant trend in rates over time (p=.08) once rates were adjusted for markers of disease severity. The gender-year interaction was not significant (p=.42). The annual predicted hospitalization rates for males and females were calculated separately and are shown in Figure 7.
Figure 6: Predicted hospitalization rate per person years by age and gender
(Model count=agegroup sex)

![Figure 6: Predicted hospitalization rate per person years by age and gender](image)

Figure 7: Predicted hospitalization rates per person year for males and females by year
(Model count=year; calculated for females and males separately)

![Figure 7: Predicted hospitalization rates per person year for males and females by year](image)
3.5 Analysis of predictors of hospitalization

3.5.1 Unadjusted analysis

Ages 7-19 years

For those subjects between the ages of 7-19 years, 786 subjects had 4,408 person years of data. Unadjusted Poisson GEE analysis revealed that sex was a significant predictor of hospitalization (Table 5). Hospitalization rate for females was 43% more than males (p=.01). All clinical predictor variables of hospitalization were statistically significant on univariate analysis (Table 5). The most influential predictors in this age group were the presence of diabetes, pancreatic insufficiency (PI) and ever having a positive culture for BCC (RRs 3.78 (95% CI 2.67-5.35), 2.65 (95% CI 1.61-4.37), and 2.12 (95% CI 1.51-2.98) respectively). Ever having cultured PsA was a significant predictor of hospitalization (p=.02). As expected, lung function and nutritional status were significant predictors of hospitalization with fewer hospital counts per year for those with higher lung function and higher z-score BMI (p<.0001 for both). Birth year was found to be a significant predictor of hospitalization (p=.004) with fewer hospital counts per year the later the birth year. Calendar year, distance to CF centre, SES, and community size were not significantly associated with hospitalization rates. Number of outpatient clinic visits was not included in the analysis of this age group because the definition used to identify outpatient visits did not produce acceptable measurements for this variable. With the algorithm used, 72% of the annual observations for the 7-19 year old age group had no clinic visits recorded which suggests a problem with the algorithm in this age category. This may be due to large pediatric centres
being on alternative funding plans where OHIP billing is not mandatory or necessary for financial remuneration and shadow billing is not enforced.

**Threshold Effect**

The linear effect was evaluated for all continuous predictor variables. A squared term for birth year was not significant (p=.41) and when categorized into 5 year blocks no significant threshold effect was identified therefore birth year was analyzed as a continuous variable in subsequent analyses. The quadratic term (FEV$_1^2$) for lung function was not a statistically significant predictor at the 0.05 level (p=.06) however it was further investigated using categories to assess its linearity. FEV$_1$ percent predicted was categorized in increments of 10-percentage points. FEV$_1$ percent predicted was a significant predictor throughout the lung function range reflecting its linear nature. Lastly, nutritional status using z-score BMI was evaluated and the quadratic term was non-significant (p=.27).

**Over 19 years of age**

In the over 19 age group, 629 subjects had 4,036 person years of data. The sex effect in older ages remained strongly significant as seen in Table 6. Hospitalization rate for females was 42% higher than males (p=.002). However age, birth year, and presence of PsA were no longer statistically significant predictors in adults. The most influential predictors in this older age group were low lung function (FEV$_1$ < 30% predicted), pancreatic insufficiency (PI), ever having a positive culture for BCC, and the presence of diabetes (RRs 6.17 (95% CI 4.39-8.67), 2.72 (95% CI 1.94-3.82), 2.47 (95% CI 1.98-3.71) and 2.11 (95% CI 1.69-2.72) respectively). Interestingly, SES was again a non-significant predictor although the p-value comparing the highest income quintile to lowest approached significance (p=.07). Unlike the younger cohort, 68% of annual observations had 1 or more clinic visits recorded.
Increased outpatient clinic visits were associated with a higher number of hospital counts (p<.0001). Distance to CF centre and community size continued to be non-significant predictors of hospitalization rates in those over the age of 19 years.

**Threshold Effect**

Continuous predictors were evaluated for linearity in the over 19 age group both by using a quadratic term and/or classifying continuous variables into categories. The quadratic term for FEV₁ (FEV₁²) was highly significant (p <.0001). This was further investigated by using categories of FEV₁ percent predicted increasing by increments of 10 percentage points. In contrast to those between 7-19 years of age, analysis of the parameter estimates in this age group revealed differential counts of hospitalization depending on the subject’s lung function. Based on this analysis, three FEV₁ (% predicted) categories were identified: < 40, 40-60, and > 60 and these were used in subsequent analyses to simplify the reporting of this effect. Nutritional status using BMI was interrogated in a similar fashion; however the effect of BMI on hospital counts was found to be a significant predictor throughout, and therefore remained continuous for multivariable modeling. The quadratic term for birth year was non-significant (p=.77).
Table 5: Unadjusted univariate analyses (ages 7 - 19 years)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter estimate (SE)</th>
<th>Rate ratio (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex (F vs. M)</td>
<td>.361 (.15)</td>
<td>1.43 (1.08-1.91)</td>
<td>.01</td>
</tr>
<tr>
<td>Age (years)</td>
<td>.065 (.017)</td>
<td>1.07 (1.03-1.10)</td>
<td>.0001</td>
</tr>
<tr>
<td>FEV1 (% pred)</td>
<td>-.032 (.003)</td>
<td>0.73 (0.69-0.77)*</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Z-score BMI</td>
<td>-.408 (.049)</td>
<td>0.66 (0.60-0.73)</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>CF-related Diabetes (present)</td>
<td>1.33 (.18)</td>
<td>3.78 (2.67-5.35)</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Pancreatic Insufficient</td>
<td>.977 (.25)</td>
<td>2.65 (1.61-4.37)</td>
<td>.0001</td>
</tr>
<tr>
<td>B. cepacia (ever)</td>
<td>.752 (.17)</td>
<td>2.12 (1.51-2.98)</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Pseudomonas (ever)</td>
<td>.451 (.19)</td>
<td>1.57 (1.08-2.29)</td>
<td>.02</td>
</tr>
<tr>
<td>Birth Year</td>
<td>-.059 (.014)</td>
<td>0.94 (0.91-0.97)</td>
<td>.004</td>
</tr>
<tr>
<td>Calendar Year</td>
<td>.014 (.02)</td>
<td>1.01 (0.98-1.05)</td>
<td>.45</td>
</tr>
<tr>
<td>Distance to CF centre (km)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 50</td>
<td>-.163 (.21)</td>
<td>0.85 (0.56-1.30)</td>
<td>.45</td>
</tr>
<tr>
<td>50-199</td>
<td>.108 (.22)</td>
<td>1.11 (0.72-1.73)</td>
<td>.63</td>
</tr>
<tr>
<td>200-399</td>
<td>-.062 (.29)</td>
<td>0.94 (0.53-1.66)</td>
<td>.83</td>
</tr>
<tr>
<td>≥ 400</td>
<td>0</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>Community Size (Rural vs. urban)</td>
<td>.082 (.15)</td>
<td>1.08 (0.80-1.46)</td>
<td>.59</td>
</tr>
<tr>
<td>SES Income Quintile</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Lowest</td>
<td>.193 (.17)</td>
<td>1.21 (0.86-1.70)</td>
<td>.27</td>
</tr>
<tr>
<td>2</td>
<td>.212 (.19)</td>
<td>1.24 (0.85-1.79)</td>
<td>.26</td>
</tr>
<tr>
<td>3</td>
<td>.086 (.19)</td>
<td>1.09 (0.75-1.58)</td>
<td>.65</td>
</tr>
<tr>
<td>4</td>
<td>.027 (.19)</td>
<td>1.03 (0.70-1.50)</td>
<td>.89</td>
</tr>
<tr>
<td>5 Highest</td>
<td>0</td>
<td>1.0</td>
<td></td>
</tr>
</tbody>
</table>

* for each Δ10 % increase in predicted in FEV1
Table 6: Unadjusted univariate analyses (> 19 years of age)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter estimate (SE)</th>
<th>Rate ratio (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex (F vs. M)</td>
<td>.354 (.12)</td>
<td>1.42 (1.13-1.79)</td>
<td>.002</td>
</tr>
<tr>
<td>Age group (yrs)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20-34</td>
<td>.057 (.11)</td>
<td>1.06 (0.85-1.30)</td>
<td>.59</td>
</tr>
<tr>
<td>≥ 35</td>
<td>0</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>FEV₁(% pred)</td>
<td>-.031 (.003)</td>
<td>0.74 (0.69-0.78)*</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>FEV₁ categories (%pred)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 30</td>
<td>1.82 (.17)</td>
<td>6.17 (4.39-8.67)</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>30-39</td>
<td>1.28 (.16)</td>
<td>3.60 (2.62-4.93)</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>40-49</td>
<td>.799 (.16)</td>
<td>2.22 (1.62-3.04)</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>50-59</td>
<td>.508 (.15)</td>
<td>1.66 (1.24-2.23)</td>
<td>.0008</td>
</tr>
<tr>
<td>60-69</td>
<td>.147 (.16)</td>
<td>1.16 (0.86-1.59)</td>
<td>.36</td>
</tr>
<tr>
<td>70-79</td>
<td>-.266 (.17)</td>
<td>.77 (0.55-1.08)</td>
<td>.12</td>
</tr>
<tr>
<td>≥ 80</td>
<td>0</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>-.112 (.017)</td>
<td>0.89 (0.86-0.93)</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>CF-related Diabetes (present)</td>
<td>.765 (.12)</td>
<td>2.11 (1.69-2.72)</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Pancreatic Insufficient</td>
<td>1.00 (.17)</td>
<td>2.72 (1.94-3.82)</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>B. cepacia (ever)</td>
<td>.907 (.12)</td>
<td>2.47 (1.98-3.71)</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Pseudomonas (ever)</td>
<td>.024 (.13)</td>
<td>1.02 (0.75-1.27)</td>
<td>.86</td>
</tr>
<tr>
<td>Birth Year</td>
<td>-.011 (.006)</td>
<td>0.99 (0.98-1.01)</td>
<td>.08</td>
</tr>
<tr>
<td>Calendar Year</td>
<td>.022 (.01)</td>
<td>1.02 (0.99-1.05)</td>
<td>.12</td>
</tr>
<tr>
<td># outpatient clinic visits/year</td>
<td>.117 (.02)</td>
<td>1.12 (1.08-1.16)</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Distance to CF Centre (km)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 50</td>
<td>.179 (.20)</td>
<td>1.20 (0.81-1.78)</td>
<td>.37</td>
</tr>
<tr>
<td>50-199</td>
<td>.348 (.21)</td>
<td>1.42 (0.93-2.15)</td>
<td>.10</td>
</tr>
<tr>
<td>200-399</td>
<td>.475 (.24)</td>
<td>1.60 (1.00-2.57)</td>
<td>.05</td>
</tr>
<tr>
<td>≥ 400</td>
<td>0</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>Community Size (urban vs rural)</td>
<td>0.026 (.16)</td>
<td>1.03 (.74-1.42)</td>
<td>.87</td>
</tr>
<tr>
<td>SES Income Quintile</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Lowest</td>
<td>.276 (.15)</td>
<td>1.32 (0.98-1.77)</td>
<td>.07</td>
</tr>
<tr>
<td>2</td>
<td>.111 (.16)</td>
<td>1.12 (.82-1.52)</td>
<td>.48</td>
</tr>
<tr>
<td>3</td>
<td>.079 (.14)</td>
<td>1.08 (.82-1.43)</td>
<td>.58</td>
</tr>
<tr>
<td>4</td>
<td>.040 (.14)</td>
<td>1.04 (.79-1.37)</td>
<td>.78</td>
</tr>
<tr>
<td>5 Highest</td>
<td>0</td>
<td>1.0</td>
<td></td>
</tr>
</tbody>
</table>

* for each Δ10 % increase in predicted in FEV₁
3.3.2 Adjusted analysis

The results of forward variable selection regression model building are shown in tables 7 and 8 for those between 7-19 years of age and tables 9 and 10 for those over the age of 19 years. The base model included predictor variables sex, age, FEV\textsubscript{1}, and z-score BMI or BMI. Only statistically significant predictor variables in univariate analyses were included in the multivariable model with the exception of PsA which was initially included because of its clinical significance in CF. The effect on the parameter estimates as each variable is added to the model is shown. No significant sex interactions were identified.

Ages 7-19 years

Sex was a significant predictor of hospitalization independent of all other clinical predictors. The hospitalization rate was 38% higher for females compared to males (p=.004) even after adjusting for other significant parameters. In addition, adjusted analyses revealed that lung function, z-score BMI, pancreatic status and CFRD were all significant predictors of hospitalization (Tables 7 & 8). The presence of BCC was of borderline significance with a p-value of 0.08. The hospitalization rate was 23% lower with each increase of 10% predicted FEV\textsubscript{1}. Each unit increase in z-score BMI resulted in hospitalization rates which decreased by 13%. The hospitalization rates were 71% higher in those subjects with pancreatic insufficiency while those with CFRD had 2.3 times higher the rate of hospitalizations compared to those without CFRD.

The effect on parameter estimates of adding each new covariate to the base model of sex, age, FEV\textsubscript{1} and BMI is shown in Table 7. If a parameter estimate changed by more than 10% with the addition of a new covariate, this suggested a confounding effect. When building the multivariable model, one can see that when CFRD was added to the model, the
parameter estimates for sex and pancreatic insufficiency significantly decreased suggesting that CFRD is confounded with these two variables. Further analyses revealed that the prevalence of CFRD per person years for females was significantly higher than for males (9 per 100 person years for females vs. 2 per 100 person years for males; p=.0001). Therefore part of the gender effect on hospitalization rate was due the fact that more females had CFRD; however, sex continued to be a significant predictor of hospitalization even after adjusting for CFRD status. The remaining parameter estimates were not significantly altered by the addition of any other variable. Although significant in the univariate analysis, birth year was not statistically significant in the full multivariable model.

Over 19 years of age

The sex effect on hospitalization rates persisted even after adjusting for important clinical markers of disease severity in the over 19 age group. The magnitude of the effect on hospitalization rates for each variable within the final multivariate model can be appreciated in Tables 9 & 10. Similar to the younger age group, females had 30% higher hospitalization rates compared to males (p=.01). Age group was a significant predictor in the final model with subjects over 35 years of age having 23% higher hospitalization rates per person years (p=.04). Lung function was an independent predictor of hospitalization rates. Hospitalization rates were 3 times higher in those with lowest lung function (FEV₁ < 40 % predicted; p<.0001) and 1.4 times higher in the mid-lung function range (FEV₁ 40-60 % predicted; p=.0003) when compared to those with FEV₁ ≥ 60 % predicted. Pancreatic insufficiency was of borderline significance in the final model (p=0.05) while better nutritional status, specifically BMI, was associated with a 7% decrease in hospitalization rate for each unit increase in BMI (p<.0001). The presence of CFRD and ever having a positive sputum culture
for BCC were both highly significant independent predictors of hospitalization rates associated with a p-value of < .0001 for both. The number of outpatient clinic visits in the year was a significant predictor of hospitalization rates.

The effect on parameter estimates of adding each new covariate to the base model of sex, age, FEV$_1$ and BMI is shown in Table 9. Covariates which affected the sex effect were interrogated further. The prevalence of pancreatic insufficiency was 81 per 100 person-years for males and 75 per 100 person-years for females (p=.05). The prevalence of CFRD was 20 per 100 person-years for females and 14 per 100 person-years for males (p=.03). Thus, the addition of diabetes status partially accounted for the sex effect on hospitalization rates because more females had diabetes; however sex remained a significant factor even after adjusting for this covariate. Number of outpatient clinic visits was a significant predictor of hospitalization and this variable decreased the effect of sex as well as other covariates.

Further analysis revealed that the rate of outpatient clinic visits for females was of borderline statistical significance with a p-value of 0.05.
Table 7: Model Building: parameter estimates (SE) for variables in various different models for subjects between 7-19 years of age

<table>
<thead>
<tr>
<th>Sex,age,FEV,z-bmi</th>
<th>+ panc insufficient</th>
<th>+ CFRD</th>
<th>+ cepacia</th>
<th>+ pseudo</th>
<th>+ birth year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex (F vs. M)</td>
<td>.403 (.12)</td>
<td>.399 (.12)</td>
<td>.302 (.12)</td>
<td>.314 (.12)</td>
<td>.318 (.12)</td>
</tr>
<tr>
<td>Age (years)</td>
<td>.013 (.02)</td>
<td>.013 (.01)</td>
<td>-.003 (.01)</td>
<td>-.008 (.01)</td>
<td>-.012 (.02)</td>
</tr>
<tr>
<td>FEV1(% pred)</td>
<td>-.028 (.003)</td>
<td>-.028 (.003)</td>
<td>-.027 (.003)</td>
<td>-.026 (.003)</td>
<td>-.026 (.003)</td>
</tr>
<tr>
<td>z-score BMI</td>
<td>-.163 (.05)</td>
<td>-.160 (.05)</td>
<td>-.153 (.05)</td>
<td>-.146 (.05)</td>
<td>-.144 (.05)</td>
</tr>
<tr>
<td>Pancreatic Insufficient</td>
<td>.698 (.26)</td>
<td>.698 (.26)</td>
<td>.698 (.26)</td>
<td>.698 (.26)</td>
<td>.698 (.26)</td>
</tr>
<tr>
<td>CF-Related Diabetes</td>
<td>.835 (.16)</td>
<td>.835 (.16)</td>
<td>.835 (.16)</td>
<td>.835 (.16)</td>
<td>.835 (.16)</td>
</tr>
<tr>
<td>B. cepacia (ever)</td>
<td>.225 (.16)</td>
<td>.225 (.16)</td>
<td>.225 (.16)</td>
<td>.225 (.16)</td>
<td>.225 (.16)</td>
</tr>
<tr>
<td>Pseudomonas (ever)</td>
<td>.187 (.15)</td>
<td>.187 (.15)</td>
<td>.187 (.15)</td>
<td>.187 (.15)</td>
<td>.187 (.15)</td>
</tr>
<tr>
<td>Birth Year</td>
<td>-.014 (.02)</td>
<td>-.014 (.02)</td>
<td>-.014 (.02)</td>
<td>-.014 (.02)</td>
<td>-.014 (.02)</td>
</tr>
</tbody>
</table>
Table 8: Rate ratio and associated 95% confidence intervals for subjects between 7-19 years of age using the FULL model

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Rate ratio (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex (F vs. M)</td>
<td>1.38 (1.11-1.73)</td>
<td>.004</td>
</tr>
<tr>
<td>Age (years)</td>
<td>1.00 (0.97-1.04)</td>
<td>.97</td>
</tr>
<tr>
<td>FEV₁ (Δ10% ↑ in % pred)</td>
<td>0.77 (0.73-0.81)</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>z-score BMI</td>
<td>0.87 (0.78-0.96)</td>
<td>.008</td>
</tr>
<tr>
<td>Pancreatic insufficient</td>
<td>1.71 (1.04-2.82)</td>
<td>.03</td>
</tr>
<tr>
<td>CF-related diabetes</td>
<td>2.30 (1.66-3.18)</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>B. cepacia (ever)</td>
<td>1.32 (0.97-1.79)</td>
<td>.08</td>
</tr>
<tr>
<td>Pseudomonas (ever)</td>
<td>1.22 (0.89-1.66)</td>
<td>.21</td>
</tr>
<tr>
<td>Birth Year</td>
<td>0.99 (0.98-1.05)</td>
<td>.45</td>
</tr>
</tbody>
</table>
Table 9: Model Building: parameter estimates (SE) for variables in various different models for subjects over 19 yrs of age

<table>
<thead>
<tr>
<th></th>
<th>Sex, age, fev, bmi</th>
<th>+ panc status</th>
<th>+ CFRD</th>
<th>+ cepacia</th>
<th>+ pseudo</th>
<th>+ clinic visits</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate (SE)</td>
<td>P value</td>
<td>Estimate (SE)</td>
<td>P value</td>
<td>Estimate (SE)</td>
<td>P value</td>
</tr>
<tr>
<td>Sex (F vs. M)</td>
<td>.313 (.11)</td>
<td>.004</td>
<td>.356 (.11)</td>
<td>.01</td>
<td>.292 (.10)</td>
<td>.005</td>
</tr>
<tr>
<td>Age group (year)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 35</td>
<td>.144 (.10)</td>
<td>.17</td>
<td>.224 (.10)</td>
<td>.03</td>
<td>.172 (.10)</td>
<td>.09</td>
</tr>
<tr>
<td>20-34 (ref)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FEV1(%) pred</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 40</td>
<td>1.32 (.11)</td>
<td>&lt;.0001</td>
<td>1.30 (.12)</td>
<td>&lt;.0001</td>
<td>1.27 (.12)</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>40-59</td>
<td>.482 (.09)</td>
<td>&lt;.0001</td>
<td>.456 (.10)</td>
<td>&lt;.0001</td>
<td>.434 (.10)</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>≥ 60 (ref)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMI (kg/m^2)</td>
<td>-.061 (.02)</td>
<td>.0002</td>
<td>-.052 (.02)</td>
<td>.002</td>
<td>-.061 (.02)</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Pancreatic insufficient</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.664 (.17)</td>
<td>&lt;.0001</td>
<td>.492 (.17)</td>
<td>.003</td>
<td>.388 (.17)</td>
<td>.02</td>
</tr>
<tr>
<td>CF-related diabetes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.601 (.11)</td>
<td>&lt;.0001</td>
<td>.554 (.11)</td>
<td>&lt;.0001</td>
<td>.552 (.11)</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>B. cepacia (ever)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.439 (.10)</td>
<td>&lt;.0001</td>
<td>.455 (.11)</td>
<td>&lt;.0001</td>
<td>.434 (.10)</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Pseudomonas (ever)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.073 (.11)</td>
<td>.52</td>
<td>.079 (.11)</td>
<td>.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinic Visits/year</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.098 (.01)</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 10: Rate ratio and associated 95% confidence intervals for subjects over 19 years of age for the FULL model

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Rate ratio (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex (F vs. M)</td>
<td>1.30 (1.06-1.59)</td>
<td>.01</td>
</tr>
<tr>
<td>Age group ≥ 35</td>
<td>1.23 (1.01-1.50)</td>
<td>.04</td>
</tr>
<tr>
<td>FEV1(%) pred &lt; 40</td>
<td>3.06 (2.43-3.87)</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>1.43 (1.18-1.74)</td>
<td>.0003</td>
</tr>
<tr>
<td>FEV1(%) pred 40-59</td>
<td>.93 (.90-.96)</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>FEV1(%) pred ≥ 60 (ref)</td>
<td>1.36 (1.00-1.86)</td>
<td>.05</td>
</tr>
<tr>
<td>Pancreatic Insufficient</td>
<td>1.62 (1.30-2.02)</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>1.08 (0.94-1.15)</td>
<td>.48</td>
</tr>
<tr>
<td>Clinic visits/year</td>
<td>1.10 (1.07-1.13)</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>
3.6 Sensitivity Analyses

Several sensitivity analyses were carried out. The multivariable analysis was repeated separately using 3 variations on the dataset: (1) dataset with only non-missing values (i.e. complete case analysis), (2) dataset with outlier observations (in outcome variable) removed (Note: outliers were defined as any outcomes that were within the top 5% of the data; analysis done using multiple imputation therefore no missing values) and (3) dataset using only CIHI-DAD data to calculate hospital counts (i.e. excluded outcomes identified using homecare algorithm; used multiple imputation therefore no missing values). The results are shown in Tables 11 and 12. Regardless of the method used to analyze the data, the conclusions, particularly with respect to the gender effect, remain constant throughout for both age cohorts.
Table 11: Sensitivity analyses: comparison of multivariable model results, ages 7-19 years

<table>
<thead>
<tr>
<th></th>
<th>Non-missing Data (Exclude missing data)</th>
<th>Multiple Imputation (Exclude Homecare)</th>
<th>Multiple Imputation (Exclude outliers)</th>
<th>Multiple Imputation (All data)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate (SE)</td>
<td>P value</td>
<td>Estimate (SE)</td>
<td>P value</td>
</tr>
<tr>
<td>Sex (F vs. M)</td>
<td>.271 (.11)</td>
<td>.02</td>
<td>.332 (.12)</td>
<td>.01</td>
</tr>
<tr>
<td>Age (years)</td>
<td>-.006 (.02)</td>
<td>.71</td>
<td>-.004 (.02)</td>
<td>.84</td>
</tr>
<tr>
<td>FEV1( % pred)</td>
<td>-.030 (.002)</td>
<td>&lt;.0001</td>
<td>-.029 (.003)</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>z-score BMI</td>
<td>-.129 (.06)</td>
<td>.02</td>
<td>-.148 (.06)</td>
<td>.02</td>
</tr>
<tr>
<td>Pancreatic Insufficient</td>
<td>.122 (.23)</td>
<td>.60</td>
<td>.399 (.26)</td>
<td>.12</td>
</tr>
<tr>
<td>CF-related diabetes</td>
<td>.823 (.16)</td>
<td>&lt;.0001</td>
<td>.886 (.18)</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>B. cepacia (ever)</td>
<td>.224 (.16)</td>
<td>.16</td>
<td>.313 (.17)</td>
<td>.06</td>
</tr>
<tr>
<td>Pseudomonas (ever)</td>
<td>.140 (.15)</td>
<td>.36</td>
<td>.198 (.17)</td>
<td>.24</td>
</tr>
<tr>
<td>Birth Year</td>
<td>-.010 (.02)</td>
<td>.61</td>
<td>-.013 (.02)</td>
<td>.52</td>
</tr>
</tbody>
</table>
Table 12: Sensitivity analyses: comparison of multivariable model results, ages > 19 years

<table>
<thead>
<tr>
<th></th>
<th>Non-missing data (Exclude missing data)</th>
<th>Multiple Imputation (Exclude HomeCare)</th>
<th>Multiple Imputation (Exclude outliers)</th>
<th>Multiple Imputation (All data)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate (SE)</td>
<td>P value</td>
<td>Estimate (SE)</td>
<td>P value</td>
</tr>
<tr>
<td>Sex (F vs. M)</td>
<td>.244 (.10)</td>
<td>.01</td>
<td>.251 (.11)</td>
<td>.02</td>
</tr>
<tr>
<td>Age group (yr)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 35</td>
<td>.114 (.10)</td>
<td>.24</td>
<td>.131 (.10)</td>
<td>.20</td>
</tr>
<tr>
<td>20-34 (ref)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FEV1(%) pred</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 40</td>
<td>1.37 (.14)</td>
<td>&lt;.0001</td>
<td>1.37 (.12)</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>40-59</td>
<td>.464 (.12)</td>
<td>&lt;.0001</td>
<td>.468 (.10)</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>≥ 60 (ref)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>-.081 (.02)</td>
<td>&lt;.0001</td>
<td>-.069 (.02)</td>
<td>.001</td>
</tr>
<tr>
<td>Pancreatic insufficient</td>
<td>.068 (.15)</td>
<td>.37</td>
<td>.276 (.16)</td>
<td>.09</td>
</tr>
<tr>
<td>CF-related diabetes</td>
<td>.422 (.11)</td>
<td>&lt;.0001</td>
<td>.429 (.12)</td>
<td>.0003</td>
</tr>
<tr>
<td>B. cepacia (ever)</td>
<td>.388 (.10)</td>
<td>&lt;.0001</td>
<td>.531 (.11)</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Pseudomonas (ever)</td>
<td>.071 (.10)</td>
<td>.49</td>
<td>.065 (.11)</td>
<td>.57</td>
</tr>
<tr>
<td>Clinic Visits/yr</td>
<td>.102 (.01)</td>
<td>&lt;.0001</td>
<td>.084 (.01)</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>
3.7 Identifying individuals with CF: Comparison of administrative data and CPDR

3.7.1 OHIP and CIHI claims compared to CPDR

Algorithm #1: Two or more OHIP claims using ICD 9 code ‘277’ plus either a claim for spirometry or chest radiography within a 24 month period or 1 or more CIHI-DAD claim using the ICD 9 code ‘2770’ represents an individual with CF.

This algorithm resulted in an exceptionally large number of apparent individuals with CF (Table 13). Comparing this method of identifying CF to the CPDR “gold-standard”, the sensitivity was 93% however the positive predictive value (PPV) was only 2% making it an unreliable and unusable method for identifying individuals with CF using only administrative data. The percent agreement was low at 2%.

Table 13: Comparison between CIHI-DAD and OHIP databases and the CF Patient Data Registry to identify individuals with CF using algorithm #1

<table>
<thead>
<tr>
<th></th>
<th>Canadian Patient Data Registry</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+</td>
</tr>
<tr>
<td>≥ 1 hospitalization using “277” in CIHI-DAD</td>
<td>1391</td>
</tr>
<tr>
<td>OR</td>
<td>-</td>
</tr>
<tr>
<td>OHIP claim using “277” AND 1 spirometry or chest x-ray within a 24 month period</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1497</td>
</tr>
</tbody>
</table>

Sensitivity 91%; PPV 2%; Percent agreement 2%
3.7.2 CIHI-DAD claims only compared to the CPDR

a) Algorithm #2: One or more hospitalization using the diagnostic ICD-9 code ‘277.0’

Using this algorithm, the number of individuals identified as having CF using administrative data decreases dramatically from over 65,000 to 1,484 (Table 14). The administrative data, using this method, had a sensitivity of 74%, PPV of 74% for identifying CF patients with a percent agreement of 59%.

Table 14: Comparison between CIHI-DAD and CPDR to identify individuals with CF using algorithm #2

<table>
<thead>
<tr>
<th></th>
<th>Canadian Patient Data Registry</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+</td>
</tr>
<tr>
<td>≥ 1 hospitalization using “277.0” in CIHI-DAD</td>
<td>1105</td>
</tr>
<tr>
<td></td>
<td>-</td>
</tr>
</tbody>
</table>

Total 1497

Sensitivity 74%; PPV 74%; Percent agreement 59%

b) Algorithm #3: One or more hospitalizations using ‘277.0’ and under the age of 65 years compared to hospitalized CPDR patients.

Because some CF patients within the CPDR have never been hospitalized, using the entire CPDR as the comparator to CIHI-DAD hospital administrative data means it will be
impossible to match some individuals who have never required a hospitalization. Therefore, the CPDR subjects were linked to CIHI-DAD prior to the comparison to determine those individuals with CF who had had at least one hospitalization regardless of reason. This group was then used as the comparator in this algorithm. The sensitivity was increased using this algorithm resulting in a sensitivity of 86% and a PPV of 78% and the percent agreement increased to 69% (Table 15).

Table 15: Comparison between CIHI-DAD and HOSPITALIZED CPDR subjects to identify individuals with CF using algorithm #3

<table>
<thead>
<tr>
<th>Canadian Patient Data Registry (excluding patients never hospitalized)</th>
<th>+</th>
<th>-</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>1105</td>
<td>319</td>
</tr>
<tr>
<td>≥ 1 hospitalization using “277.0” In CIHI-DAD, &lt; 65 years of age</td>
<td>1424</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>179</td>
<td>n/a</td>
</tr>
</tbody>
</table>

Total 1284

Sensitivity 86%; PPV 78%; Percent agreement 69%

c) Algorithm #4: Any hospitalization after diagnosis date of CF using ‘277.0’, under age 65 years compared to hospitalized CPDR patients.

Hospitalizations before the diagnosis of CF is made are unlikely to have a diagnostic code for CF (‘277.0’) given the specificity of this diagnosis. If the only hospitalization a patient had was prior to their CF diagnosis, then these patients would be missed using the
previous algorithm. Therefore, in order to better assess the agreement, any hospitalizations prior to the diagnosis of CF (obtained from the CPDR dataset) were eliminated. When the comparison is re-analyzed, the sensitivity increases further to 93% however the PPV decreased slightly to 75% with percent agreement of 71% (Table 16).

Table 16: Comparison between CIHI-DAD and CPDR patients HOSPITALIZED AFTER CF DIAGNOSIS DATE using algorithm #4

<table>
<thead>
<tr>
<th></th>
<th>Canadian Patient Data Registry (excluding patients never hospitalized)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+</td>
</tr>
<tr>
<td>≥ 1 hospitalization using “277.0” AFTER CF diagnosis date &lt; 65 years of age using CIHI-DAD</td>
<td>1069</td>
</tr>
<tr>
<td></td>
<td>84</td>
</tr>
</tbody>
</table>

Total: 1153

Sensitivity 93%; PPV 75%; Percent agreement 71%

d) Algorithm #5: Two or more hospitalizations after diagnosis date of CF using ‘277.0’, under age 65 years compared to hospitalized CPDR patients.

By increasing the number of hospitalizations required to identify individuals with CF, the algorithm is more specific (PPV increases to 91%) but sensitivity drops to 75% in identifying individuals with CF using administrative data only. The percent agreement is similar at 70% (Table 17).
Table 17: Comparison between CIHI-DAD and CPDR patients HOSPITALIZED after the diagnosis date to identify individuals with CF using algorithm #5

<table>
<thead>
<tr>
<th>Hospitalization Criteria</th>
<th>Canadian Patient Data Registry (excluding patients never hospitalized)</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥ 2 hospitalizations using “277.0” AFTER CF diagnosis date Under 65 years of age</td>
<td>+ 870</td>
</tr>
<tr>
<td></td>
<td>- 283</td>
</tr>
</tbody>
</table>

**Total** 1153

Sensitivity 75%; PPV 91%; Percent agreement 70%

Table 18: Summary of 5 algorithms used to identify individuals with CF using administrative data

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Reference Standard</th>
<th>Sensitivity</th>
<th>PPV</th>
<th>% agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥ 1 CIHI-DAD + 2 OHIP claims in 2 yrs</td>
<td>CPDR, all subjects (n=1497)</td>
<td>91%</td>
<td>2%</td>
<td>2%</td>
</tr>
<tr>
<td>≥ 1 CIHI-DAD claim in 2 yrs</td>
<td>CPDR, all subjects (n=1497)</td>
<td>74%</td>
<td>74%</td>
<td>59%</td>
</tr>
<tr>
<td>≥ 1 CIHI-DAD claim in 2 yrs and ≤ 65 yrs of age</td>
<td>CPDR, excluding subjects never hospitalized (n=1284)</td>
<td>83%</td>
<td>78%</td>
<td>69%</td>
</tr>
<tr>
<td>≥ 1 CIHI claim in 2 yrs after the diagnosis of CF, ≤ 65 yrs of age</td>
<td>CPDR, excluding subjects never hospitalized and hospitalizations prior to CF diagnosis date (n=1153)</td>
<td>93%</td>
<td>75%</td>
<td>71%</td>
</tr>
<tr>
<td>≥ 2 CIHI claim in 2 yrs after the diagnosis of CF, ≤ 65 yrs of age</td>
<td>CPDR, excluding subjects never hospitalized and hospitalizations prior to CF diagnosis date (n=1153)</td>
<td>75%</td>
<td>91%</td>
<td>70%</td>
</tr>
</tbody>
</table>
A summary of the algorithms used to identify individuals with CF in administrative data as well as the sensitivity and PPV for each are seen in Table 18. Using algorithm #4 above, 355 individuals identified by administrative data as having CF but who were not found in the CPDR were examined in detail in an effort to evaluate the strength of the administrative data evidence for the diagnosis of CF. This would allow an estimation of the number of patients with CF possibly missed by the CPDR. Firstly, patients were identified as possible CF if the individual had: a) between 3-5 hospitalizations, 100% of admissions associated using CF-related ICD-9 codes (2770, 486, 494, 0417, 4821, 5771, 5770, 4939, 9968) b) > 5 hospitalizations, at least 40% of their admissions were associated with a CF-related ICD-9 code as the primary reason for the hospitalization. After reviewing the CIHI-DAD data on these subjects, 39 (11%) subjects were highlighted as possible individuals with CF based on their number of hospitalizations using CF related diagnostic codes. The next step involved reviewing the hospitals at which the individuals were admitted. Hospitals that have a known CF program associated with them were categorized as a ‘CF’ hospital and the remaining hospitals were ‘non-CF’.

Of those 39 subjects, 35 had been admitted, at least once, to a CF hospital. Given the specialized nature of CF, it was felt that if an individual had CF and was admitted to a hospital with a known CF program, it was extremely unlikely that this individual would not be seen by the CF team and hence have data submitted to the CPDR. Therefore, it is assumed that these 35 individuals were likely followed in the CPDR but did not match to the RPDB during the initial linkage procedure (i.e. part of the group of 116 patients in the CPDR with no IKN). The remaining 4 individuals were admitted several times with CF-related diagnostic codes to a non-CF hospital. The likelihood of them not being captured within the
CPDR is much greater and may well have been missed by the registry. Based on this analysis and the assumptions specified, 1% of individuals identified by CIHI-DAD as having CF may have been missed by the CPDR.
4 DISCUSSION

4.1 Summary of research findings

This is the first body of work to systematically examine longitudinal hospitalization rates and predictors of hospitalization in a large Canadian CF population using linked administrative and clinical registry data. Individuals with CF were hospitalized frequently throughout their lives, most commonly for pulmonary complications. This work confirmed that there are significant differences in hospitalization rates between males and females, both in adolescents and adults. Higher hospitalization rates were found in females, even after adjusting for clinically important covariates suggesting that the increased hospitalization rates were not due to worse overall CF disease. This sex difference in hospitalization rates does not appear to be specific to CF. Sex disparities in hospitalization rates have been reported in asthma with females having higher hospitalization rates compared to males.\textsuperscript{95,96,97} Studies in asthma used only administrative data and therefore could not control for disease severity therefore authors could not address whether the higher hospitalization rates were simply due to more severe disease or if there was another factor playing a role in the gender disparity. In addition to gender, this research identified several important clinical predictors that help identify those at high risk for hospitalization.

This work also shows that hospitalization rates, adjusted for disease severity, have not significantly changed over the 10 year period. One might have expected hospitalization rates to decrease with the availability of more effective therapeutic strategies in CF. However, the reality is that there is no cure for CF which means that individuals with CF are living longer and their lung disease continues to progress over time. Unadjusted hospitalization rates are
increasing however this is simply due to the fact that there are more individuals with CF living with more progressive lung disease.

The work presented in this dissertation has also demonstrated that, despite CF being a specific diagnosis, health administrative databases alone were insufficient to reliably and accurately identify individuals with this condition unless they had been hospitalized. Conducting health services research on hospitalized subjects only reduces the potential sample size and limits the generalizability of any conclusions drawn. On the other hand, clinical registries alone significantly underestimate the number of hospitalizations for patients with CF. This highlights the importance of linking a clinical registry such as the CPDR with administrative data in order to accurately quantify health services utilization for individuals with CF. Furthermore, the underestimation of hospitalizations within the CPDR was extremely important information to provide to the CCFF for it highlights possible inadequacies of the registry definitions. This information prompted the creation of the “CPDR Instruction Manual” outlining data entry details and variable definitions which will be used by all CF clinics to improve the quality and consistency of data captured within the CPDR. The ideal situation is to combine clinical registry data with health administrative data so that a comprehensive picture of health care delivery can be obtained while having the ability to adjust for important clinical factors. Furthermore, the process of data linkage provides an opportunity to improve the data quality of participating clinical registries. In this study, 6 duplicate records were identified and subsequently corrected within the CPDR. The fact that less than 1% of records within the dataset were identified as duplicates suggests the data within the registry is of reasonably high quality. Other published studies have shown duplication rates as high as 8% which is likely an underestimate of the true duplication rate.
This study failed to identify a significant number of hospitalized individuals with strong administrative data evidence supporting a diagnosis of CF who were not linked to the CPDR. This supports the claim that most hospitalized individuals with CF receiving care in Ontario are followed within the registry and it seems reasonable to extrapolate that finding to the other provinces since the model of CF care and expectations for data submission to the CPDR is standard across Canada. However, this does not apply to CF patients who have never been hospitalized and who receive care on an outpatient bases. The OHIP data yielded an extremely large number of apparent individuals with CF (>64,000) who were not found within the CPDR. It was not possible to reasonably evaluate all such claims to look for supportive administrative evidence supporting a diagnosis of CF and therefore, it was impossible to quantify how many non-hospitalized individuals with CF were not captured within the CPDR. It is entirely possible that individuals with pancreatic sufficiency and minimal, if any, lung disease who would be considered to have ‘mild’ CF may not be followed at an accredited CF centre and hence not be captured within the CPDR. These individuals may not be on medications and if they are, they may have private drug insurance through work making the drug coverage incentive for many CF patients not relevant.

### 4.2 Reasons for gender disparity in hospitalization rates

#### 4.2.1 Outpatient Management

Sex-based differences may be the result of sociological, behavioural, hormonal or genetic influences. One sociological possibility relates to outpatient management of CF. “Ambulatory care sensitive conditions” have been defined as those diseases or complications that, when sub-optimally managed in the outpatient setting, result in avoidable
hospitalizations. An example of such a condition is diabetes and hospitalizations for acute diabetic emergencies. Improved outpatient management of one’s blood sugar will result in fewer hospitalizations for hyper- or hypoglycaemic complications. It is recognized that CF is a progressive, genetic disease for which there is no cure and one could argue that no amount of outpatient care can prevent all hospitalizations, unlike diabetes. However, it is possible that the rate of hospitalizations in CF could be diminished with more aggressive outpatient management. A study by Ramsey et al showed that intermittent administration of inhaled tobramycin over a 24 week study period in patients with CF resulted in a reduction in hospitalization rates. Those subjects on inhaled tobramycin were 26% (95% CI 2–43%) less likely to be hospitalized and 36% (95% CI 17–51%) less likely to require IV anti-pseudomonal antibiotic therapy than those receiving placebo. Another study by Fuchs et al showed that the administration of rhDNase, an inhaled mucolytic medication, for 24 weeks resulted in a modest reduction in the risk of pulmonary exacerbations required IV antibiotic therapy compared to those on placebo. The age-adjusted risk of an exacerbation requiring IV antibiotics was reduced 28% by the administration of rhDNase once daily (RR 0.72; 95% CI 0.52 to 0.98; p=0.04). Finally, treatment with oral azithromycin for 6 months in patients with CF resulted in fewer pulmonary exacerbations (hazard ratio 0.65; 95% CI 0.44 to 0.95; p=0.03) and 14% fewer patients requiring hospitalization (95% CI 1–25%; p=0.05). It is possible that the gender gap in hospitalization rates could be a result of differential prescribing patterns for medications that have been shown to reduce hospitalization rates. However medication use was not recorded in the CPDR clinical registry or the health administrative databases. Therefore, the impact of outpatient therapies on hospitalization
rates could not be specifically assessed in this dissertation. Furthermore, there are no published studies to date examining this question in CF.

4.2.2 Hormonal Influences

Whenever a gender disparity exists, hormonal influences come to mind as possible implicating factors. Sex hormones may modulate the severity of CF lung disease resulting in increased hospitalizations and ultimately worse survival. A proposed mechanism of CF lung disease is dessication of airway secretions due to a reduced volume of airway epithelial surface liquid. CFTR acts as a regulator of periciliary ion and water content. The absence of or a reduction in CFTR function leads to excessive epithelial Na+ channel mediated (ENaC) sodium re-absorption which depletes airway surface liquid further. A reduction in airway surface liquid results in impaired ciliary beats, reduced mucociliary clearance and potentially increased infections. Gender differences in respiratory transepithelial ion transport have been postulated. Female sex hormone stimulation of a component of respiratory epithelial Na+ and fluid absorption would be expected to aggravate CF airway pathology. Nasal potential difference (PD) measurements are used to assess the voltage across respiratory epithelium which is used as a surrogate for sodium and chloride transport in the lower airways. Sweezey et al compared nasal PD tracings throughout the menstrual cycle in women with and without CF. This study showed that amiloride-insensitive nasal PD measurements are increased in the luteal phase (when progesterone and estrogen levels are higher) compared to the follicular phase of the normal menstrual cycle in young women with CF. The authors concluded that the changes seen in nasal PD measurements most likely reflect altered transepithelial cation (Na+) absorption which ultimately may contribute to reduced airway surface liquid and worse lung disease in women with CF. In a similar vein, Coakley et al
postulated that high estrogen levels would reduce the ability of females with CF to regulate airway surface liquid volume homeostasis. In normal airways, active Cl- secretion via CFTR tempers ENaC-mediated Na+ reaborption. In CF airways, CFTR absence or dysfunction results in ENaC no longer being regulated. Increased intracellular calcium stimulates Cl- secretion through alternative pathways. One such alternative pathway of chloride conductance in the airway is through the calcium-activated chloride channel (CaCC). High estrogen levels reduce calcium-activated Cl- secretion through CaCC which may then worsen airway surface dehydration and decrease mucociliary clearance. This in turn, may result in worsening respiratory symptoms and consequently, more hospitalizations for pulmonary exacerbations. Further evidence for a hormonal effect in CF can be seen during pregnancy. Odegaard et al examined outcomes following completed pregnancies in CF. The change in lung function pre and post-pregnancy was non-significant however the researchers found that the number of IV antibiotic courses during pregnancy was nearly twice that of the year before or the year following pregnancy. There are several factors that may contribute to an increased need for antibiotic therapy during pregnancy including hormones causing reduced mucociliary clearance within the airway, development of gestational diabetes affecting ability to fight infection, or increased inflammation however these factors have not been extensively studied in women with CF. Finally, changes in symptomatology and lung function in relation to the menstrual cycle have been documented in CF. Parker et al reported a case of recurrent episodes of catamenial hemoptysis and pneumothoraces in a 32 year old women with CF. Johannesson et al documented lower lung function in women with CF during ovulation and menstruation felt to be related to changes in progesterone levels during the menstrual cycle. Having lower lung function, in
and of itself, can be a risk factor for pulmonary exacerbation. This dissertation showed that with lower lung function, hospitalization rates increased however, lung function did not explain the gender effect seen. Therefore it seems unlikely that hormonal effects that impact lung function directly are the reason for this gender disparity but rather, hormonal influences that affect presenting symptoms, mucous clearance, or inflammation may explain the higher hospitalization rates seen in females with CF.

4.2.3 Patient- and Provider-level Influences

Many factors are considered when deciding about hospitalization for a pulmonary complication in CF. Factors affecting the decision to hospitalize a patient for an infective exacerbation includes a constellation of symptoms, some of which are objective measures (e.g. FEV1) while others are subjective complaints provided by the patient (e.g. worsening cough, darker sputum and hemoptysis). Decisions can be influenced by what is communicated to the health care provider and how and what is communicated can differ depending on the gender of the patient. As seen in a study of differences between males and females in the rate of knee and hip arthroplasty, women were found to have a greater need for joint replacement therapy, and despite equal willingness to undergo the procedure, they were less likely to discuss the option of surgery with their physician, which led to less frequent recommendation for surgery. Details on gender differences in communication in CF have not been studied. One can hypothesize however that if females with CF were more expressive about symptoms, this may influence the decisions made by the CF care team surrounding hospitalization. Another patient-level factor to explain the gender difference seen is the fact that females may be seen more frequently in the outpatient setting and therefore have more opportunity to be hospitalized. Although this may contribute to the
gender disparity, one would expect that if males had symptoms of a CF pulmonary exacerbation that went untreated, their condition would progress and deteriorate further increasing the likelihood that they would be hospitalized, even sicker than before.

Furthermore, in those over the age of 19, the analysis was adjusted for frequency of clinic visits and the gender disparity persisted. Sex differences in patient preferences for care, particularly for adults, may exist. If a male patient was the primary breadwinner, he may be more resistant to in-hospital care because of lost revenue from work for instance. The proportion of male and female employed CF patients was not available within the clinical registry therefore this specific question could not be addressed in this dissertation. From a provider-level perspective, a gender bias may exist simply because of the extensive literature showing worse survival in females with CF. It is possible that CF health care providers, consciously or sub-consciously, have a lower threshold for admitting a female with CF for intensive antibiotic therapy.

Female anatomy may impose increased vulnerabilities toward impaired mucociliary clearance which may in turn, predispose women with CF to develop respiratory symptoms, following a respiratory insult, earlier than males. Women tend to have smaller lungs than men relative to their height and females airways tend to be smaller overall. These sex-related differences in anatomy may predispose to increased particle deposition and reduced particle clearance.

Increased hospitalization rates for females are not explained by worse CF disease as measured by lung function, nutritional status, co-morbidities, and microbiology. Why then, do women with CF have worse overall survival despite seemingly being more aggressively treated? The answer to this is unclear. Being hospitalized can be associated with negative
consequences as was seen in the *B. dolosa* outbreak in the United States and the BCC outbreak in Toronto. It is possible that females are exposed to various negative consequences in hospital which may contribute to worse outcomes in general although this study did not address this issue specifically.

Gender disparities in outcomes of care are important to recognize and document as the findings have important implications for practice, research and perhaps policy. Differences between groups allow clinicians to identify those at highest risk for certain outcomes so that strategies can be implemented to prevent negative outcomes and maximize care for all patients. With respect to research, the fact that gender differences are noted in many areas of health care delivery and within a broad range of diseases, highlights the need for all researchers to consider gender as an important predictor in all research analyses. Although descriptive studies are informative, they often lack the explanatory power to understand and tease out the reasons for these gender differences. There needs to be more research examining potential reasons for difference in gender so our understanding of this phenomenon can increase and change can occur to close the gender gap.

### 4.3 Other predictors of hospitalization

Not surprisingly, lung function, regardless of age, was a significant independent predictor of hospitalization in the multivariate analysis. In those subjects between 7-19 years of age, this relationship was linear; however, this was not the case in adults. For children/adolescents, for each 10% increase in FEV₁ percent predicted, hospital counts were 28% lower. In adults, the risk of hospitalization differed depending on the severity of lung disease. For those with an FEV₁ below 40%, hospitalization counts were 3 times higher compared to those with lung function above 60% predicted. This is intuitive from a clinical
standpoint since those with severe lung disease have less pulmonary reserve and are more likely to be hospitalized when they suffer from a pulmonary exacerbation. With moderate lung disease, the increase in counts is 43% higher than those with lung function above 60% predicted. Said in another way, as FEV1 increases, up to around a value of 60% predicted, hospitalizations decrease. Above that level of lung function, there was little or no further decrease in hospitalization counts. The plateau seen in our adult cohort also fits with what is seen clinically. It is likely that patients with mild lung disease having an exacerbation will receive oral antibiotics as first-line therapy potentially averting a hospital admission for intravenous treatment. Whereas, those individuals with severe lung disease have less pulmonary reserve and are more likely to need more aggressive therapy during exacerbations. Although there is no cure for CF, developing strategies to slow the decline of lung function over time with current therapies may prevent hospitalizations. Also, prevention of viral infections with hand washing and influenza vaccinations may also reduce the number of hospitalizations due to worsening lung function.

Nutritional status, in both age cohorts, was an independent marker of hospitalization. For each unit increment in z-score BMI, hospital counts were 23% lower in subjects 7-19 years of age. For adults, for each unit increment in BMI, hospital counts were 6% lower. These results again reinforce the importance of nutrition in CF and that aiming for normal nutritional status improves outcomes. A plateau effect was not seen; however this may have been due to the fact that there are few patients in the obese category within this population. Aggressive nutritional supplementation should continue to be encouraged in this population.

Within the younger cohort (7-19 years of age), calendar year was a significant predictor of hospitalization in multivariate analysis. As the calendar year increased, hospital
counts decreased which may reflect more aggressive inhaled antibiotic therapy to eradicate PsA over the time period studied. This relationship persisted in the older cohort (> 19 years of age) which is not surprising. New available therapies for CF may explain this relationship.

Although health care providers may not be able to alter the development of certain clinical predictors of hospitalization such as pancreatic status or the presence of CFRD, they can optimize how these CF-related complications are managed. The current study identified these as risk factors for hospitalization. Knowing these factors are risks for hospitalization allows health care providers to target these groups in an effort to decrease hospitalization rates. Ensuring optimal enzyme therapy for individuals with PI and providing education and therapy for those with CFRD is an important goal for CF healthcare providers. By developing strategies to aggressively manage such CF–related complications, it may be possible to decrease the need for hospitalization although this needs to be confirmed in future prospective studies. Furthermore, individuals with PI are at risk for CFRD (rather than PS subjects) therefore it is possible that the effect size of CFRD on hospitalization may be even greater that the current results suggest if the analysis was restricted to PI subjects only. Another important finding of this research is that CFRD was found more commonly in women consistent with other literature in this area. It is possible that hormonal influences or the presence of gene modifiers play a role in this finding. The relationship between the female sex and the development of CFRD is unclear but warrants further investigation.

The lack of association between SES and hospitalization is somewhat surprising, since both the CF and non-CF literature has shown a significant relationship between SES and important health outcomes. Only one published study from the U.S. commented on SES and hospitalization rates specifically in CF, and there are no published studies from Canada.
Using Medicaid as a proxy for low SES, Schechter et al reported that 44.5% of Medicaid patients received intravenous antibiotics in hospital compared to 28.6% in the non-Medicaid group.\textsuperscript{61} Even after adjusting for age, race, sex, lung function and pancreatic status, the odds of a pulmonary exacerbation was 1.58 times higher (95% CI 1.27 to 1.96; p<.001) in the Medicaid group. Furthermore, utilization of ambulatory specialty care was similar in the two groups (Medicaid 4.5 vs. non-Medicaid 4.1; p=.40) making underutilization less likely as a cause for the increased hospital admissions. Outside of CF, a U.S. study of sickle cell disease, another chronic disease, showed that children with low SES had significantly higher healthcare utilization for inpatient care compared to children with higher SES characteristics.\textsuperscript{110} Another study assessing the impact of socioeconomic status on hospital use in New York City showed higher hospitalization rates in low-income areas.\textsuperscript{111} The authors suggested that this may be due to better outpatient care in higher income regions. Even in Canada where universal health care reduces economic barriers to accessing health services, individuals from lower-SES neighbourhoods experienced an excess number of hospitalizations for conditions that could be prevented by optimal outpatient care.\textsuperscript{100} What are the possible reasons that this study failed to show a relationship between hospitalization and SES? SES was characterized using neighbourhood income quintiles and postal code information rather than specific patient annual income data; therefore some individuals may be incorrectly classified. However, this method of categorizing SES has been used in other studies that have shown a positive association between SES and outcomes making it unlikely that is the reason for the lack of significant association in this study.\textsuperscript{100} A unique feature of general health care in Canada is the universal access to health services. Unique features of CF care specifically, are that the majority of Canadians with CF are followed at specialized
CF centres and there is a provincial CF drug program that covers the cost of most CF medications. It may be the distinctive combination of universal health care, the national network of specialty CF clinics, and drug coverage that results in similar utilization of health services regardless of SES. If true, this model of care may help to ameliorate any disparities between those with high and low income in other chronic diseases.

It was hypothesized that geographic location of an individual with CF may impact on hospitalization rates. Specifically, since all CF centres are in cities, individuals with CF living in rural areas, far from their primary CF centre, may receive sub-optimal care which may translate into more hospitalizations. However, the results of the study did not show a significant relationship between residence or distance to CF centre and hospitalization rates. Schechter et al reported that rural residence was not an independent predictor of FEV\textsubscript{1} although hospitalization rates were not examined. Geographical barriers to care have been documented in other diseases. For example, individuals living in rural areas within Canada had higher rates of readmission following acute myocardial infarction. Comprehensive access to a specialized CF centre, regardless of distance, may explain the findings in CF. Another contributing factor may be the Northern Health Travel Grant Program available in Ontario. This program provides funding for travel for residents of Northern Ontario who have been referred to a physician specialist or health facility. Distance travelled must be 100 km or more in order to be eligible. Furthermore, the CF community within Canada is relatively small and the Canadian CF Foundation serves as a resource both for CF education and clinic information. The fact that geography was not related to increased hospitalization, may be the result of a high level of awareness and education about this disease and the importance of ongoing chronic care in this patient population. This may not be the case in
other diseases that are not managed in such a cohesive and comprehensive way from birth to death.

**4.4 Limitations of current study**

Because the study is observational, the non-experimental study design raises the possibility of errors in sampling and measurement which may threaten the validity of the results. Furthermore, causation is difficult to establish in observational research. This type of research can describe what is seen but not necessarily why it happens.

**4.4.1 Limitations of the outcome variable**

Hospitalization count per year for pulmonary complications was the outcome of this study. The initiation of home IV antibiotic therapy for a pulmonary exacerbation was included in the number of “hospitalizations” since, without homecare, this event would have resulted in a hospitalization. Hospitalization for pulmonary complications was defined by listing ICD-9 codes used as the most responsible diagnosis (MRD) however this algorithm has not been validated by a chart review. The ICD-9 code for cystic fibrosis (“2770”) was included in the list of “respiratory” codes because of the concern that pulmonary CF exacerbations might be coded as CF rather than the particular infectious agent. In fact, the results showed that CF was listed as the MRD in 78% of claims identified. In an effort to further characterize these CF claims, the remaining diagnostic codes used to describe the admission were assessed. In 14% of hospital claims, the ICD-9 code for CF was the only code used to describe the admission. CF and diabetes was used in 2%, and 22% of claims used CF plus ICD-9 codes not related to the respiratory system. Therefore, although most of the admissions identified were associated with the pulmonary system, there were a small proportion of claims that may have been due to non-pulmonary causes. Although this may
have introduced some error in that not all admissions were secondary to the respiratory system, it doesn’t alter the conclusions drawn.

In order to identify the initiation of home IV antibiotic therapy, certain assumptions were made. Outpatient clinic visits were identified using OHIP claims and homecare claims before and after outpatient visits were evaluated. For institutions on alternative funding plans, where OHIP shadow billing may not be rigorously carried out, outpatient clinic visits may have been missed thereby underestimating home intravenous antibiotic therapy at those centres. However, there is no reason to suspect that these claims would be missed more frequently in one gender over the other therefore, the gender effect seen likely still holds true. Although the algorithm for identifying the initiation of home IV therapy has not been validated, it is reassuring that sensitivity analysis excluding home IV therapy resulted in the same conclusions supporting the notion that the results are robust.

Length of stay (LOS) was not considered in this study. One hospitalization in a given year with an associated LOS of 14 days would be given the same weight as a hospitalization with a LOS of 365 days. Clearly the severity of these two scenarios is not the same. It is possible that males had fewer hospitalizations per year but the duration of the hospital stay was longer than females suggesting more complicated and severe disease.

4.4.2 Limitations of the predictor variables

Missing clinical data, and the accuracy of non-missing data, could bias the results of the thesis. However, it should be noted that the primary clinical predictor of this research, namely sex, was unambiguous and was recorded in all study subjects. With respect to missing clinical data, assumptions about the mechanism of missingness were made and if these were not accurate, the results may be biased. However, literature has suggested that
multiple imputation strategies provide unbiased estimates, even if the missingness is not completely at random.\textsuperscript{113,114} Furthermore, the sensitivity analysis excluding missing data provided the same results as that obtained using imputation again suggesting the results are robust. Clinical data for pulmonary function and nutritional status markers were recorded during the first clinic visit of the calendar year which may or not have been at the time of a hospitalization. Without using a time-dependent covariate or obtaining clinical measures prior to hospitalization dates, it is impossible to say that the clinical predictors preceded the hospitalization. Furthermore, it is possible that females were hospitalized because of worse CF lung disease but this was obscured by the fact that lung function measurements were not done at the time of the hospitalization. Or it may be that the gender disparity in hospitalization rates was not explained by lung function because FEV\textsubscript{1} was not a sensitive enough marker of pulmonary disease. This possibility seems unlikely however since FEV\textsubscript{1} is widely used as a marker of pulmonary disease in CF and is strongly liked to outcomes like survival in CF.

Another limitation of this work was that outpatient clinic visits, as previously mentioned, were not reliably identified using administrative data in those between the ages of 7-19 years. This may be secondary to the lack of OHIP shadow billing by physicians on alternative funding plans. Because of this, the impact of outpatient visits on hospitalization rates could not be evaluated in the younger cohort. It is possible that the gender effect in this age group is explained by females being seen more frequently as an outpatient and hence, they have a greater opportunity to be hospitalized. Outpatient clinic visits did not explain the gender effect in those over the age of 19, but it is not clear if these findings can be generalized to those under the age of 19 years.
The fact that SES was not a significant predictor of hospitalization was surprising. Because SES was determined using postal code and census data rather than actual income, it is possible that individuals were misclassified which may obscure a relationship if it were present. However, there are several studies showing a relationship between outcomes and SES using the same technique making it less likely that the lack of this relationship in CF is simply due to an error in classification.

4.4.3. General limitations

This research was conducted using Ontario data, so inferences made are limited to the context of this province. Clinical care and practice patterns are similar across Canada, so although these results may be generalizable to Canadians with CF, there is substantial variation in other countries and the results may not apply in those circumstances. Furthermore, the study cohort excluded individuals younger than 7 years of age and hospitalizations after transplantation therefore, it is not possible to say if these findings apply to those groups.

4.5 Future directions of research

This body of work is the foundation upon which future research can be built. The study findings and the literature reviewed in the Introduction section of this dissertation confirm several gender disparities reported in CF outcomes. Such results highlight the importance of routinely conducting gender analyses in all CF research to understand the impact that gender has on outcomes in this population. By elucidating underlying pathways leading to gender differences in CF, targeted interventions can be developed to reduce documented disparities. For instance, evaluating differences in outpatient management in CF
may elucidate variations in care that explain some or all of the findings presented in this
dissertation.

From a larger perspective, it is important to note that there is a paucity of literature on
health services utilization in CF. A linked database for CF research, using clinical and
administrative data, is a valuable resource which could be used for important and novel
future projects even beyond the gender issue. Such a program of research could evaluate how
care is currently being delivered to the CF population in Ontario which is the initial step in
understanding health care utilization in this group. Such future work may identify gaps in
care which could be the focus of interventions to improve healthcare delivery in CF. An
example of such research is to quantify the involvement between primary care physicians
and the CF population. There is a sense that CF patients often view the CF team as their
‘primary’ healthcare providers although involvement of a family physician is recommended
in the American and European guidelines for CF care. No studies have documented the
proportion of CF patients who actually see a primary care physician. This is an important
question because routine tests or screening procedures, such as pap smears, may be missed if
a family physician is not involved in the care of this young patient population. Cervical
cancer can be prevented with routine pap smear screening making this is an important area of
research in this population. On the topic of how care is provided to this patient group, the
frequency of sub-specialty physician involvement when CF-complications arise, such as
diabetes, is unknown. Are outcomes for CFRD patients different if they are followed by both
their CF team and an endocrinologist compared to being followed by their CF team alone?
Or does endocrinology involvement result in worse outcomes because of the unique nature of
CFRD and lack of knowledge around the management of an individual with CF who
develops diabetes? And finally, the transition of patient care from pediatric to adult CF centres is yet another area which has never been studied using linked clinical and health administrative data. Transition is a potentially vulnerable time whereby patients may be at particular risk for poor outcomes or being lost to follow up. Using a linked database could provide important answers to questions such as: How well does transition of care work? Is there a delay from when patients leave a pediatric centre to when they have their first visit at an adult CF centre? Are there variations in outcomes around the time of transition across the province? Are individuals at higher risk for hospitalization around the time of transition? Having a linked dataset could provide rich information to better understand health care delivery in CF and more importantly, such research could help to elucidate gaps in care and identify those at risk for poor outcomes. Interventions can then be put in place in an effort to improve the care being delivered to this patient population.

It is crucial that future research not only characterize what is being done currently but evaluate how earlier intervention may affect outcomes. In 2008, Ontario initiated newborn screening for CF. The evidence suggests that by identifying these patients soon after birth, prior to the development of symptoms, outcomes are improved. But no researcher has examined whether or not this impacts on health care utilization. For example do patients diagnosed through the newborn screening program have fewer hospitalizations because of earlier and more aggressive care? Some literature suggests that infants identified through newborn screening actually acquire PsA infection at an earlier age since these infants are followed in a CF clinic after diagnosis and are exposed to potentially detrimental pathogens. Given the fact that acquisition of PsA is associated with negative outcomes, this may result in greater health care utilization. Furthermore, are fewer patients on disability
later in life if diagnosed through the newborn screening program? With respect to CF-related complications, there are no published studies showing that endocrinology involvement improves diabetic management and ultimately survival in CF. As physicians, we do our best to improve the lives of our patients; however, we must also ensure that we are not doing harm with our interventions. Screening strategies using chest radiographs and CT scans to monitor lung disease in CF have been recommended. However, there is no literature assessing radiation exposure and the risk of developing malignancies later in life? These are just some examples of the much needed research in this area. By pursuing health services research in CF, knowledge in this area can be greatly expanded and the results can be translated into improved care for all CF patients in Canada.
APPENDIX
## Appendix 1: Diagnostic and procedural codes use to identify the cohort, co-morbidity and outcome features

<table>
<thead>
<tr>
<th>Assessment</th>
<th>Disease or procedure</th>
<th>Period of Assessment</th>
<th>ICD-9 codes</th>
<th>OHIP codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study outcomes</td>
<td>Respiratory illness</td>
<td>Hospital admission with length of stay (LOS) ≥ 3 d</td>
<td>0310, 0318, 0319, 0338, 0339, 0340, 0380, 0381, 0382, 03840, 0388, 0411, 0412, 0415, 0417, 0418, 0419, 0798, 0799, 1173, 1363, 2770, 4644, 4650, 4658, 4659, 4660, 4661, 4789, 4800, 4801, 4802, 4808, 4809, 481, 4821, 4822, 4823, 4824, 4828, 4829, 483, 4846, 4847, 4848, 485, 486, 4870, 4871, 4878, 490, 4910, 4911, 49120, 49121, 4918, 4919, 49300, 49301, 49310, 49311, 49320, 49321, 49390, 49391, 496, 494, 5070, 5100, 5110, 5111, 5118-20, 5128, 5130, 5131, 515, 5168, 5178, 5180, 5183-5, 5188, 51881, 51882, 51889, 5190-4, 5198-99, 78600, 78601, 78602, 78609, 7861-64, 78650-52, 78659, 7869, 7907 (n=110 codes)</td>
<td>Same as above</td>
</tr>
<tr>
<td>Home intravenous therapy: started upon discharge from hospital</td>
<td></td>
<td>Hospital admission with LOS &lt; 3 d AND homecare claim within 7 days of discharge date with no homecare claims within 4 weeks of hospital admission</td>
<td>Same as above</td>
<td><img src="https://via.placeholder.com/150" alt="Image" /></td>
</tr>
<tr>
<td>Home intravenous therapy: started from the outpatient clinic with no hospital admission</td>
<td></td>
<td>Outpatient clinic visit AND homecare claim within 7 days of clinic visit with no homecare claims within 4 weeks of clinic visit date.</td>
<td>A475, 135, 473, 478, 138, AND J304 or J323</td>
<td><img src="https://via.placeholder.com/150" alt="Image" /></td>
</tr>
</tbody>
</table>
REFERENCES


