SOCIO-DEMOGRAPHIC FACTORS, SMOKING, SYMPTOMS, MORBIDITIES AND PULMONARY FUNCTION AND QUALITY OF LIFE IN INDIVIDUALS WITH A HEAVY SMOKING HISTORY

Matthew Ventresca, BA.

Submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Applied Health Sciences
(Health Science)

The Faculty of Applied Health Sciences
Brock University
St. Catharines, ON.

Abstract

Objective: To determine which socio-demographic, exposure, morbidity and symptom variables are associated with health-related quality of life among former and current heavy smokers.

Methods: Cross sectional data from 2537 participants were studied. All participants were at ≥2% risk of developing lung cancer within 6 years. Linear and logistic regression models utilizing a multivariable fractional polynomial selection process identified variables associated with health-related quality of life, measured by the EQ-5D.

Results: Upstream and downstream associations between smoking cessation and higher health-related quality of life were evident. Significant upstream associations, such as education level and current working status and were explained by the addition of morbidities and symptoms to regression models. Having arthritis, decreased forced expiratory volume in one second, fatigue, poor appetite or dyspnea were most highly and commonly associated with decreased HRQoL.

Discussion: Upstream factors such as educational attainment, employment status and smoking cessation should be targeted to prevent decreased health-related quality of life. Practitioners should focus treatment on downstream factors, especially symptoms, to improve health-related quality of life.

Keywords: Health-related quality of life (HRQoL), smoking, visual analog scale (VAS), EuroQol 5 Dimension (EQ-5D), multivariable fractional polynomial (MFP)

Acknowledgements

First and foremost, I would like to express my most sincere thanks to Dr. Tammemägi. Your courses in research methods and epidemiology have inspired me to pursue a career in health research. Your desire and achievement for excellence in teaching and research is evidenced by your many successes and is truly an inspiration. Thank you for your immeasurable efforts in aiding me while completing this thesis, it has been the most rewarding educational experience of my life. Most of all, thank you for leading a truly enjoyable and fun project which I will always look back upon fondly.

Thank you to the professors and administrative staff at Brock University especially Dr. Liu, Dr. Faught and Bev Minor. Since my first day at Brock, and throughout my graduate studies, you have provided inspiration and guidance. Thank you also to Dr. Cosby and Dr. Lawrance, your feedback and time are greatly appreciated.

Dad, thank you for teaching the most important life lessons; that nothing is impossible, to never quit when faced with failure and to remember to laugh. Lessons I will surely be teaching my nieces, nephews and children one day.

Babbo and Grandma, from my first days, to my first days of school and everyday thereafter, you have taught me to value hard work, to “be a good boy”, the importance of listening, and to take time to enjoy life. Grazie.

Aunt Rose and Uncle Joe, you are so much more than an Aunt and Uncle. Despite your distance you are somehow always present. Thank you for showing me how to be a good uncle and role-model, and for providing a home away from home.

To my siblings, cousins and friends, thank you for being there whenever a distraction was needed and for your support, however unconventional it may have been.

Last, but certainly not least, Monika. Nobody has had more fun writing a thesis than I. Thank you for happily tolerating the countless “dates” to Starbucks, the endless late-night/early-morning study sessions and most of all, your undying positivity.

Table of contents

List of Figures

List of Tables

List of abbreviations

β – Beta-coefficient

BMI – Body Mass Index

BIF – Bootstrap Inclusion Fraction

CHD – Coronary Heart Disease

CHF – Congestive Heart Failure

CI – Confidence interval

COPD – Chronic Obstructive Pulmonary Disease

EORTC – European Organization for Research and Quality of Treatment of Cancer

EQ-5D – EuroQoL Five Dimension

EQ-5D-3L – EuroQoL Five Dimension Three Level

EQ-VAS – EuroQol Visual Analog Scale

FEV_1 - Forced Expiratory Volume in one second

FVC – Forced Vital Capacity

GOLD – Global initiative for chronic Obstructive Lung Disease

HRQoL – Health-related quality of life

IQR – Interquartile range

LDCT – Low Dose Spiral Computed Tomography

MCS – Mental Component Summary score

MFP – Multivariable fractional polynomial

N – Sample Size

OR – Odds ratio

P – P-value

PVD – Peripheral Vascular Disease

PanCan – Pan Canadian Early Detection of Lung Cancer Study

PCS – Physical Component Summary score

QoL – Quality of Life

QUALEFFO 41 - Quality of Life Questionnaire of the European foundation for Osteoporosis 41

SD – Standard Deviation

SES – Socioeconomic Status

SF-6D – Short Form-6 Dimension

SF-36 – Short-Form-36

SGRQ – St. George's Respiratory Questionnaire

VAS – Visual Analog Scale

WHO – World Health Organization

Chapter 1: Introduction

1.1 What are quality of life and health-related quality of life?

Quality of Life (QoL) is a patient- or person- reported outcome measure of general well-being which focuses on an individual's own feelings and perceptions of present life circumstances.[1] However, there is no consensus between experts as to the exact list of factors, or "domains", which compose an individual's QoL.[1-5] Health-related quality of life (HRQoL) refers to the aspects of QoL which are affected by an individual's health. In spite of the broad definitions, researchers believe that individuals are familiar with the terms "quality of life" and "health-related quality of life".[1-4]

The domains of HRQoL can be analyzed specifically, or on a global level as a measure of an individual's overall health. Individually or globally, these domains can also be compared in groups of individuals with a specific disease. HRQoL research is useful in predicting disease symptoms and severity, survival time, medical decision-making of doctors and patients, compliance and evaluation of treatment, as well as cost-effectiveness analysis. This analytic versatility makes HRQoL an extremely useful and important outcome measure.[1-5]

1.2 Prevalence and effects of smoking on health and health-related quality of life

In 2010, approximately 17% (20% male, 14% female) of the Canadian population aged 15 years and older were current smokers and 26% were former smokers.[6-8] By the year 2025 there is expected to be over 1.6 billion smokers worldwide.[9] Tobacco use is the single most preventable cause of illness in North America, harming several organs and causing multiple chronic diseases including several cancers.[10,11] Multiple investigations conclude that smoking has a negative effect on domain specific and overall HRQoL.[12-21] While these studies have produced useful findings, several investigations have not included measures of morbidities, pulmonary function or

symptoms. Identifying factors contributing to poor HRQoL among former and current smokers would allow physicians and other healthcare workers to prioritize and tailor treatments which maximize HRQoL of those suffering from the adverse effects of smoking.

1.3 Measuring HRQoL

This investigation utilizes the EuroQol Visual Analog Scale (EQ-VAS) and the EuroQol 5 Dimensional 3 Level questionnaire (EQ-5D-3L), collectively referred to as the EQ-5D, to measure HRQoL.[22] The EQ-5D is a quick and straightforward test, designed for use in postal surveys, clinics and face-to-face interviews for clinical and economic investigations.[3,23] The EQ-VAS, is a vertical scale ranging in value from 0 at the bottom representing “worst imaginable health state” to 100 at the top, representing “best imaginable health state”. Respondents are asked to rate their current health state on the scale.[23] The EQ-5D-3L, or “descriptive system”, asks participants to report whether they experience none, some or severe problems in each of five domains; mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Given five dimensions and three possible scores for each dimension, a total of 243 combinations exist. Each combination represents a single health state. Domain level data can be analyzed by calculating odds ratios of reporting none, some or severe problems or by creating index values, ranging from 0 to 1, by applying standardized equations to a participant’s health state.[23]

1.4 Pan-Canadian Early Detection of Lung Cancer Study

To investigate the predictors of HRQoL in a population of former and current smokers with a heavy smoking history, data collected by the Pan-Canadian Early Detection of Lung Cancer Study (PanCan) were utilized. Individuals were recruited by newspaper, television and mailing advertisements as well as through physician and

dental offices between September 2008 and December 2010. Epidemiologic data were collected by detailed and structured in-person interview from 2537 participants.[24,25]

1.5 Gaps in the Literature

Research examining specific factors that affect HRQoL among the general population of individuals with a heavy smoking history is required. Many smoking-related HRQoL investigations took place in clinical settings where patients are more likely to present with severe health consequences related to smoking. Participants of these studies were often diagnosed with a specific smoking-related illness or were undergoing a specific treatment which smoking is suspected of interfering with.[26,27] Results of clinical studies are extremely useful to their respective study populations; however, there is a lack of external generalizability resulting from small sample sizes and specific inclusion criteria which most of the general public do not meet.

Whether clinical or population based, investigations into the factors affecting HRQoL commonly utilized data from studies which did not primarily focus on HRQoL.[14,16,28] Therefore, population, exposure, morbidity or symptom variables which significantly affect HRQoL, but are not included in the primary study, are excluded from analysis.[29-31] Furthermore, several previous investigations employed regression techniques which did not account for non-linear functions, or controlled for non-linear functions using categorical cut points which groups distinct observations into a single category.[32-37] With few studies including data from multiple variable groups and observing the continuous relationship between independent and dependent variables, an incomplete or incorrect understanding of HRQoL has been obtained.

1.6 Study Aims and Response to Research Needs

The PanCan enrolled 2537 current and former smokers who had a lung cancer risk ≥2% over 6 years. Due to their heavy smoking history, many of these study

participants had developed a range of health related issues that could decrease their HRQoL. The general aim of this study is to develop a better understanding of the risk factors associated with decreased HRQoL in individuals with a heavy smoking history.

Question #1. What socio-demographic, medical history, exposure, morbidity and symptom factors are associated with overall HRQoL as measured by the EQ VAS in individuals with a heavy smoking history?

Question #2. What socio-demographic, medical history, exposure, morbidity and symptom factors are associated with HRQoL domains (mobility, self-care, usual activities, pain/discomfort and anxiety/depression) as measured by the EQ-5D?

Researching factors that are associated with low HRQoL may help individuals and clinicians identify factors which might be modifiable, preventing reductions in HRQoL and leading to steps which may improve HRQoL through behavioural changes or treatment. Additionally, it is possible that knowledge of smoking associated-decline in HRQoL may be used to motivate current smokers to partake in smoking cessation programs and quit smoking.

1.7 Summary

Research incorporating data from socio-demographic, medical history, exposure, morbidity and symptom variables among large population-based samples is lacking. Linear and logistic regression models utilizing a multivariable fractional polynomial selection procedure identified significant independent predictors of HRQoL. Discovery of alterable factors associated with HRQoL identified priorities for preventions, interventions, or treatments to help optimize HRQoL.

Chapter 2: Review of Literature

2.0 Overview

This chapter reviews information required to understand the study rationale, beginning by defining and describing quality of life (QoL) and health-related quality of life (HRQoL). Tobacco related topics are then discussed, including prevalence of use, properties of tobacco smoke, health implications and previous research related to HRQoL. Chapter two concludes by stating the research needs this project addresses.

2.1 What is quality of life?

QoL is a patient, or person, reported outcome measuring general well-being by focusing on an individual's own feelings and perceptions of present life circumstances.[1-5] QoL depends on various factors which often include; happiness, general health, physical functioning, emotional functioning, cognitive functioning, social well-being, sexual functioning and existential issues.[1,2,5] Each of these factors is referred to as a "domain". While there is no widely accepted definition of QoL or the domains composing QoL, experts agree that it is a construct dependent on a combination of multiple domains.[1-5]

Regardless of the lack of definition, researchers believe people are familiar QoL and are able to decipher its general meaning.[1-4] However, individuals base their QoL on perceptions and values of personal experiences, all of which greatly vary throughout the general population.[2,4] For example, Quek (2005) reports age, pain, anxiety, and depression as significant QoL factors affecting patients with lower urinary tract infection symptoms.[38] Whereas a study of nurses working night shifts reports duration of sleep, family harmony status and diet, among other factors, as contributing to overall QoL.[39] This subjectivity and variability makes creating a single concrete definition and uniform measurement of QoL extremely difficult.[1-4,40] In order to develop accurate conclusions, it is essential that variables relevant to the target population are measured.

2.2 What is health-related quality of life?

HRQoL focuses on the aspects of health which affect QoL.[5] Even though HRQoL is not synonymous with QoL, the terms are often interchanged.[40] Domains often measured in HRQoL studies include: socioeconomic status (SES), social support, or presence of physical risk factors and illnesses.[5] By focusing on factors coinciding with the World Health Organization's (1948) definition of health, "a state of complete physical, mental and social well-being and not merely the absence of disease" one is assured that measures of HRQoL, appropriately measure overall health.[1-5,40, 41]

For scientific purposes, study investigators tend to define HRQoL by the items of the questionnaire being utilized to measure HRQoL.[2] To ensure study validity, various HRQoL questionnaires have been created, many designed for specific demographic and disease groups or to assess specific domains.[3] For example, the Short-Form-36 (SF-36) can be used to measure the health status in the general population. The European Organization for Research and Quality of Treatment of Cancer (EORTC) questionnaire is designed for use in cancer patients participating in clinical trials and also includes modules to be completed dependant on specific patient diagnoses. Questionnaires examining specific domains are often used in conjunction with other more general surveys to ensure that HRQoL, as a whole, is still measured.[2] According to reviews of HRQoL questionnaires, no gold standard exists. Instrument selection should depend on a variety of factors including study population and research needs.[42,43]

2.3 Utilizing quality of life research

QoL research has been popular and growing since the 1970's and is now a common component of clinical trials. As new treatments with varying effectiveness and side effects became available, focus shifted towards improving overall well-being as opposed to mere symptom relief.[1,2,4] Studies have shown HRQoL measures have the

ability to predict which patients are at greater risk of developing advanced stages of disease as well as survival time.[44-46] HRQoL also plays a significant role in shaping patient's and physician's treatment decisions, as well as cost-effectiveness analysis; influencing healthcare providers, insurance companies and drug agencies.[1,4,47,48] By researching which health conditions or interventions most effect HRQoL, the balance between QoL, survival time and healthcare spending can be efficiently managed.[1]

2.4 Researching HRQoL amongst smokers to improve care

Several research investigations conclude that smoking negatively affects HRQoL.[12-21] Tobacco use is the most preventable cause of morbidity and mortality in North America.[10,11] Tobacco harms nearly every bodily organ and causes multiple chronic diseases including cancers.[10] Compared to non-users, former and current smokers are more likely to report work absenteeism, utilize medical services, have longer hospitalizations and are more likely to experience adverse reactions to wound care.[11,49,50] Identifying factors contributing to poor HRQoL among smokers would allow physicians and other healthcare workers to prioritize and tailor treatments which maximize HRQoL of those suffering from the adverse effects of smoking.

2.5 Prevalence of smoking and tobacco use in Canada and globally

In 2014, approximately 18% (21.4% male, 14.8% female) of the Canadian population aged 12 years and older reported being a current smoker.[51] Approximately 74% reported smoking daily, averaging 13.9 cigarettes per day.[52] Cigarettes are the most common form of tobacco use.[10] However, 4% of Canadians have smoked a form of cigar within 30 days, and less than 1% report using smokeless tobacco products in the past 30 days.[7,10] As of 2012, 28% of the Canadian population aged 15 and over were former smokers.[52] Since the PanCan recruitment dates, September 2008 to December 2010, Canadian smoking prevalence appears to generally be declining (Figure 1).[53]

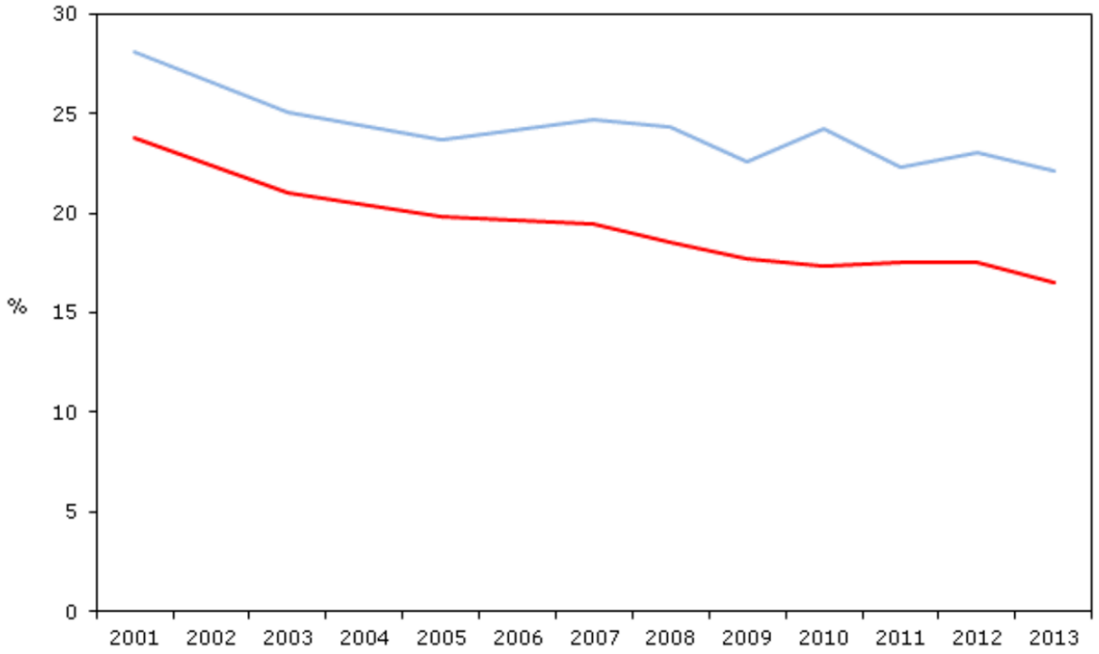

Figure 1. Percentage of smokers, Canadian population aged ≥12 by gender, 2001 to 2013, Canadian Community Health Survey.[53]

Worldwide, among low- and middle- income countries participating in the Global Adult Tobacco Survey, 48.6% of males and 11.5% of females were tobacco users and 40.7% of males and 5.0% of females were smokers.[54] By the year 2025 there is expected to be over 1.6 billion smokers.[9] Global trends indicate that the number of smokers is slowly declining in many industrialized countries yet overall prevalence remains stable or increasing among many middle- and low-income countries due to population growth.[55,56]

2.6 Why smoke?

Among adults who become daily smokers, 88% report first cigarette use by 18 years of age, and 99% by 26 years of age.[57] Factors contributing to young adults beginning and continuing to smoke are the influence of friends and siblings, and stress relief which is often associated with poor performance in school.[58,59] Furthermore, many young smokers are unaware of smoking's health implications, often downplaying the effects and equating them to other risky behaviours. Smoking is also often perceived to

be a temporary, immature behaviour that young smokers believe will pass as they age.[59] Throughout aging, continued use is primarily due to the biologically addictive properties of nicotine as well as the psychological associations individuals make with smoking.[60]

2.8 Cigarette smoke and tar

Over 600 chemicals are added to tobacco cigarettes. Among other purposes, many chemicals are added to improve taste and smell and increase addictiveness.[10,61] When lit, cigarette smoke contains between 4,000 and 7,000 chemicals.[10,62,63] Of these chemicals, 250 have harmful effects and 69 are known carcinogens.[10,64,65] Tar is the collection of total matter in cigarettes excluding nicotine and water. Nicotine, the addictive chemical in cigarettes, approximates 1.8% of a cigarette's total weight.

2.10 Developing an addiction

Within 10 to 20 seconds of puffing a cigarette nicotine travels to the brain, reaching peak blood concentration within minutes followed by a rapid decline.[10] Nicotine stimulates acetylcholine receptors in the brain which produces dopamine, creating a pleasure response. This positive reinforcement mechanism is partly responsible for the development of nicotine addiction.[10,66]

After an individual completes smoking a cigarette, nicotine and dopamine levels eventually decrease, leading to withdrawal. Symptoms of withdrawal may include irritability, anxiety, depressed mood, restlessness, sleep disturbance, and difficulty concentrating. Over time, acetylcholine receptors become desensitized to nicotine leading to decreased dopamine production. To compensate, smokers will gradually increase their nicotine intake by smoking stronger cigarettes or smoking more frequently. Withdrawal symptoms among long-term smokers begin to appear hours after smoking, peak within days of last cigarette and do not subside for 2 to 4 weeks. Experiencing withdrawal symptoms are the main reason why prolonged smoking cessation is

difficult.[10] Psychological, sensory and environmental cues, such as seeing someone smoke or smoking while drinking alcohol, can also prompt an individual to smoke.

2.11 Health outcomes associated with cigarette smoking

Worldwide, tobacco use is the leading preventable cause of morbidity and mortality (Figure 2).[56,67] Smokers are twice as likely to die prematurely and live approximately 10 fewer years than non-smokers.[68, 69] Smoking is causally linked to three of the top four causes of Canadian mortality (2011); cardiovascular disease (23%), stroke (6%) and chronic lower respiratory disease (4%). Furthermore, the Center for Disease Control estimates that smoking is the primary factor in 30% of cancer mortalities, which is the leading cause of death in Canada and the United States.[10,11,70,71]

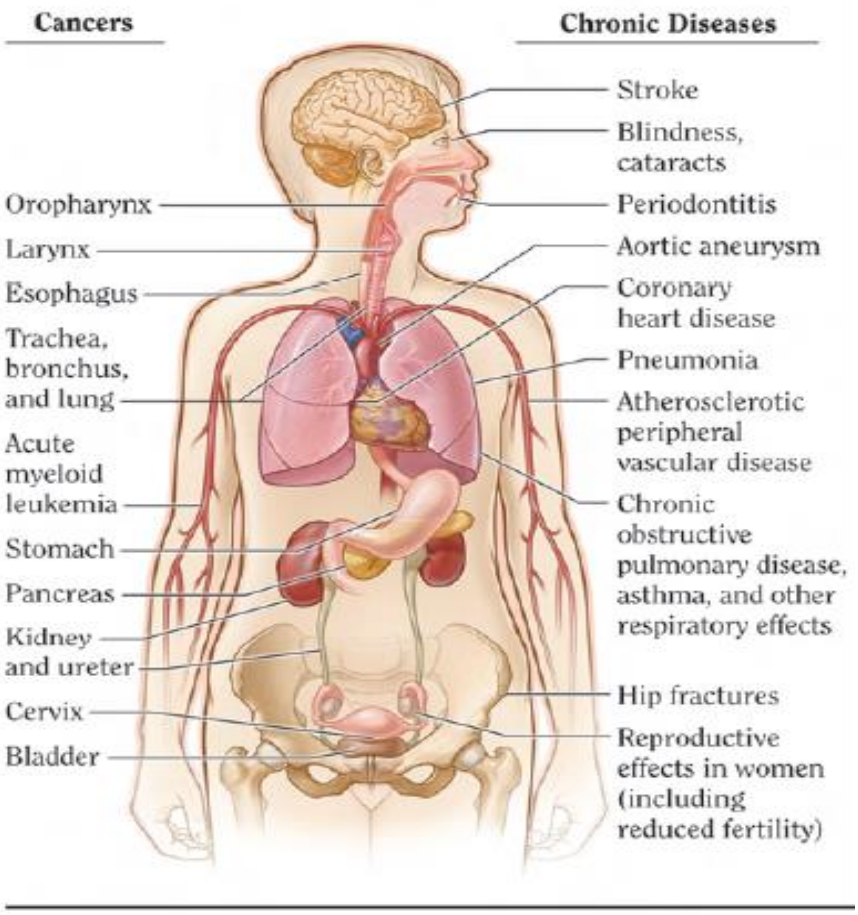

Figure 2. Diseases caused by smoking and exposure to second-hand smoke.[10]

2.11.1 The effect of smoking on cardiovascular disease and stroke

Nicotine, carbon monoxide and particulate matter increase risk of developing various cardiovascular diseases and stroke by elevating heart rate and lowering blood-oxygen levels. Oxidizing chemicals, such as free radicals, contribute to inflammation, endothelial dysfunction, oxidation of low density lipoprotein and platelet activation. These processes are known to contribute to the development of cardiovascular disease. This leads to the thickening of arterial walls due to plaque build-up and increases blood viscosity. As a result, when compared to never smokers, smokers are 2 to 4 times more likely to die from cardiovascular disease.[10,72,73] More specifically, smokers are at approximately 6 times greater risk of experiencing myocardial infarction, 2.7 times greater risk of peripheral vascular disease and 1.3 times greater risk for coronary artery disease.[10,74-76]

2.11.2 Pulmonary symptoms, morbidities and function

Upon inhalation, the chemicals of smoke bind together, slowly building tar along the trachea. As the airway narrows, oxygen inhalation and carbon dioxide exhalation are reduced, leading to wheezing, dyspnea, fatigue and physical activity impairment. Furthermore, tar build up will paralyze cilia movement, hindering their ability to remove toxins from the body, increasing likelihood of respiratory infection. Smokers may also experience phlegm secretion and coughing, the body's natural reaction to remove toxins.

Inhaled smoke will eventually reach and destroy the alveolar sac. In turn, oxygen transfer into the bloodstream is decreased and elasticity of the alveolar sac is reduced, further hindering physical activity and increasing likelihood of wheezing, dyspnea and fatigue. Damage to the alveolar sac is often permanent, resulting in long-term reduction of forced expiratory volume and forced vital capacity.[77,78] Over time smokers may develop chronic obstructive pulmonary disease (COPD) which is characterized by 3

factors; (1) airway thickening and narrowing (bronchitis); (2) destruction of alveolar walls (emphysema); and (3) coughing and phlegm build up as a result of chronic mucus secretion.[79] In its most severe cases, individuals suffering from COPD will be unable to perform light activity and will require oxygen tubes to breathe comfortably.[80]

2.12 What are the benefits of smoking cessation?

Smoking cessation leads to immediate and long-term improvements in physical health and HRQoL, even for older adults.[19, 81] Immediate effects of smoking cessation include decreased heart rate, blood pressure, blood carbon monoxide levels and improved breathing. Within one year of quitting, circulation improves and lung function increases, coughing and shortness of breath decrease, cilia regain normal function and the risk of coronary heart disease is halved. Within 5 years of smoking cessation, risk of experiencing a stroke returns to that of a never-smoker. After 5 to 10 years, risk of developing various cancers is halved. After 15 years, risk of developing coronary heart disease (CHD) returns to that of a non-smoker.[82] Quitting smoking by age 50 decreases mortality risk by half, and by age 30, almost completely.[69]

2.13 Previous research: socio-demographic, medical history, exposure, symptom, morbidity and pulmonary function factors and HRQoL

2.13.1 Socio-demographics and medical history

Gender

Collecting data from a nationally representative sample (N=3,010) of south Australians, Wilson and colleagues (2004) investigated the differences in HRQoL between male and female light, moderate and heavy smokers. T-tests were used to compare mean SF-36 scores for 733 smokers. Females reported lower mean scores in the domains of physical functioning (-5.1, P<0.001), role physical (-6.7, P=0.01), vitality (-7.8, P<0.001), social functioning (-4.6, P=0.005), role emotional (-10.0, P<0.001) and

mental health (5.7, P<0.001) than males. Stratifying analysis by smoking intensity, especially within “heavy smokers”, revealed similar results.[83] Other community-based investigations conducted in Korea, Taiwan, the United States and Canada report similar findings.[84-89] Clinical studies involving participants diagnosed with HIV/AIDS, asthma, fibromyalgia, diabetes, and individuals undergoing coronary bypass surgery also agree.[90-94]

Age

Asakawa and colleagues (2012) examine longitudinal data, collected bi-annually from 13,665 participants over 8 years (1996/97 – 2004/05) from the Statistics Canada National Population Health Survey (NPHS), with the purpose of identifying factors associated with HRQoL trajectories. Linear mixed-models were used to identify factors associated with HRQoL, measured using the Health Utility Index Mark 3 (HUI3). Regression models were stratified by age group (young 18 to 39; middle-aged 40 to 64; senior 65+) in order to identify variables significant at different life points. As a result of aging alone, adjusting for socio-demographics, number of chronic conditions, smoking and drinking status, and physical activity, HRQoL generally decreased in every increasing age group, especially after age 60.[32] These results are similar to other large scale national studies conducted in England, China, the United States and Australia which also investigate age and HRQoL in general populations.[29,95-97]

Education level

In a nationally representative American sample (N=3,663) aged 35 and older, Robert and colleagues (2009), explored differences in HRQoL between socio-economic subgroups. To verify consistency of findings and to allow for comparisons to other findings four questionnaires were used to measure HRQoL; (1) EQ-5D, (2) Short Form-6 Dimension, (3) Health Utilities Index Mark 3 and (4) asking participants to rate their

health as “excellent”, “very good”, “good”, “fair” or “poor”. Multivariable weighted least squares regression models reveal that compared to participants with at least a four year college degree, those with some post-secondary schooling, a high school degree, or less than a high school degree all had significantly lower HRQoL on each measure utilized.[29] These associations are weaker in the latest years of life which is similar to findings from Asakawa et al. (2012).[32] Other general population studies examining SES and HRQoL completed in Sweden, Poland, Greece and Canada agree with these findings.[88,98-100]

Employment Status

In a cross-sectional study exploring differences in HRQoL by working status, Hultman and colleagues analyzed data from 487 unemployed and 2917 employed study participants, aged 25 to 64, in northern Sweden. HRQoL was measured in 9 domains (entire life, somatic health, mental well-being, cognitive ability, social life, family life, activity, financial situation and meaning of life) from Hörnquist’s Quality of Life, change and status, assessment strategy. Unemployed participants reported lower mean HRQoL scores in every domain.[101] According to Ross and colleagues, lower income which is also a measure of SES, is associated with a lower HRQoL.[88] Results are similar to Zaninotto and colleagues who investigated HRQoL trajectories in England based on various socioeconomic, psychosocial and health conditions.[95]

2.13.2 Exposures

Smoking

Utilizing cross-sectional data from the Health Survey of England (2006) Vogl and colleagues explored the differences in HRQoL related to smoking status. Participants (N=13,241) included never smokers (46.9%), occasional smokers that have now quit (5.4%), ex-regular smokers (26.0%), light smokers (7.0%), moderate smokers (8.6%)

and heavy (5.9%) smokers. HRQoL was measured with the EQ-5D. Logistic regression modelling was used to calculate odds ratios of reporting some or severe problems in each EQ-5D dimension by smoking status, overall EQ-5D index scores were also calculated. Compared to never-smokers, heavy smokers reported significantly lower HRQoL in each EQ-5D domain. Furthermore, with the exception of ex-occasional smokers, all other smokers reported significantly lower mean EQ-5D index scores and were also more likely to experience lower HRQoL when controlling for other socio-economic, social capital, lifestyle and biologic factors compared to never smokers.[12] Other randomized controlled trials and population studies based in England, the Netherlands, Finland, Spain and the United States report similar findings. [12-17,19-21]

Marijuana use and HRQoL

Exploring gender differences in HRQoL among cannabis users in the United States, Lev-Ran and colleagues utilized data from 18,336 males and 24,378 females. The SF-12 was used to measure HRQoL. General linear models stratified by gender identified associations between marijuana use and HRQoL. Controlling only for age, male cannabis users reported significantly lower physical component summary score (PCS) (-0.78, $P<0.0001$) and mental component summary score (MCS) (-0.85, $P<0.0001$) than non-users. Female cannabis users also reported significantly lower PCS (-0.64, $P<0.05$) and MCS (-1.96, $P<0.0001$). When controlling for age, race/ethnicity, education level, household income, marital status, urbanicity and mood or anxiety disorders, daily cannabis use was significantly associated with reduced PCS in males (-0.35, $P<0.02$) and reduced MCS in females (-1.1, $P<0.0001$).[102] Conversely, among those suffering from fibromyalgia, marijuana use has been associated with increased HRQoL by reducing pain ($P<0.001$), stiffness ($P<0.001$), enhancing relaxation ($P<0.001$) and increasing a feeling of wellbeing ($P<0.001$).[103]

Alcohol use

Kaplan and colleagues (2012) examined alcohol use patterns over a 6 year span to identify changes in HRQoL among adults aged ≥50 years. A total of 5,404 participants were included in analysis. The dependent variable, HRQoL, was measured using the HUI3. Persistent moderate drinkers reported the highest quality of life at baseline compared to other drinking patterns; persistent nonusers (-0.059, P<0.01), persistent former drinkers (-0.078, P<0.001), decreasing users (-0.108, P<0.001), and those with unstable drinking patterns; U-shaped (-0.051, P<0.01) or inverted U shaped (-0.069, P<0.001). Throughout follow-up, the rate of HRQoL decline was equal between all groups.[104] Results from Kaplan and colleagues (2012) are consistent with findings from Byles and colleagues (2006) which utilized nationally representative data from Australia.[105] Volk and colleagues also discovered similar findings, adding that less frequent (P<0.05) and frequent high-quantity drinkers (P<0.05) report lower mental health scores than lifetime abstainers.[106]

2.13.3 Morbidities

The following are morbidities included in the PanCan study questionnaires pertaining to medical history. The morbidities discussed below are those with highest prevalence among study participants. Morbidities for which data were collected, but are not discussed here, are included in the thesis data analysis.

Cardiovascular risk factors and diseases

Data from 189,450 participants of the 2003 Behaviour Risk Factor Surveillance System were analyzed by Li and colleagues (2008) to examine the association between having multiple cardiovascular disease risk factors (diabetes, hypertension, hypercholesterolemia, BMI and smoking status) as well as cardiovascular diseases (myocardial infarction, angina and stroke) and HRQoL. HRQoL was measured using the

SF-36. Adjusting for age, gender, race/ethnicity, and education, odds ratios of experiencing poor or fair general health among those with and without cardiovascular disease were calculated. Individuals with cardiovascular disease were more likely to report poor or fair general health and experience ≥14 days of physical or mental illness or impaired activity.[30] Focusing on specific cardiovascular outcomes, Stafford and colleagues (2012) discovered that all precursors to cardiovascular disease (diabetes, hypertension and obesity) as well as cardiovascular diseases (myocardial infarction, stroke and angina) were associated with decreased HRQoL.[37] Soltoft and colleagues (2009) also analyzed data from the Health Survey for England to specifically compare the relationship between BMI and HRQoL by gender. After controlling for socio-economic status, psychosocial well-being and diagnosed morbidities, multiple linear regression models stratified by gender showed that those in normal BMI range reported the highest EQ-5D utility scores.[36,107]

Arthritis and HRQoL

Cross-sectional data from the National Health Measurement Study was analyzed by Khanna and colleagues (2011) to assess HRQoL in sample of American adults with and without self-reported arthritis. In total, 3844 adults aged 35 to 89 years participated. HRQoL was measured with six HRQoL instruments. Similar to national statistics, approximately 31% of participants reported having arthritis. HRQoL scores were stratified by gender, age group and presence of arthritis. In each gender and age group, for every HRQoL questionnaire, those with arthritis reported significantly lower mean HRQoL scores.[108] Conclusions made by Khanna et al. agree with other population studies conducted by Hill et al. (1999) in the United States, and Uhlig et al. (2007) in Norway, which also found HRQoL scores to be lower among those with arthritis compared to the general population.[109,110]

Osteoporosis and HRQoL

In their evaluation of the Quality of Life Questionnaire of the European Foundation for Osteoporosis 41 (QUALEFFO 41) Ferreira and colleagues (2009) recruited 110 post-menopausal females with and without diagnosis of osteoporosis. To verify accuracy of the QUALEFFO 41 participants also completed the SF-36. It is important to note that low scores on the QUALEFFO 41 indicate better HRQoL whereas high scores on the SF-36 represent better HRQoL. Mean HRQoL scores in all dimensions measured by the QUALEFFO 41 and SF-36 were worst among those with osteoporosis compared to age-matched controls without osteoporosis.[111] These results agree with findings reported by Adachi and colleagues (2010) who utilized data from 57,141 post-menopausal females as well as a Dutch population study (N=3,664) completed by Picavet and Hoeymans (2002).[31,112]

Chronic obstructive pulmonary disease

To determine if differences in HRQoL between former and current smokers can be explained by cough, phlegm, pulmonary function, or exercise capacity Heijdra and colleagues (2002) recruited 36 smokers, 21 ex-smokers, 19 never-smokers and 41 COPD patients. The St. George's Respiratory Questionnaire (SGRQ) was used to measure HRQoL. The SGRQ measures HRQoL on a scale from 0 to 100 with lower scores indicating better HRQoL. ANOVA results indicated that participants suffering from COPD reported the highest mean SGRQ scores compared to all other groups.[113]

Allergic rhinitis, asthma and HRQoL

Larsson and colleagues (2007) investigated the relationship between rhinitis and asthma on HRQoL with data from 5,918 Swedish study participants. The SF-36 was

used to measure HRQoL. In both males and females, non-cases reported highest HRQoL followed by individuals with rhinitis and without asthma. Asthmatics with or without rhinitis reported the lowest HRQoL scores in every SF-36 domain.[114] Similar findings are presented by Kalpaklioglu and Baccioglu (2008) as well as Chen and colleagues (2011) who researched Turkish and English populations, respectively.[115,116]

2.13.4 Pulmonary function

Utilizing data from a Norwegian general population survey (N=2,306), Voll-Aarneud and colleagues (2008) investigated the relationships between respiratory symptoms, COPD severity, pulmonary function and HRQoL. Among participants with COPD, the GOLD (Global initiative for chronic Obstructive Lung Disease) classification system was used to measure COPD severity: mild ($FEV_1 \geq 80\%$ predicted), moderate (50% to 80% predicted), severe (30 to 50% predicted) and very severe ($FEV_1 < 30\%$ predicted).[33,117] The SF-12 measured HRQoL. Linear regression modelling revealed a decrease in HRQoL among moderate (-3.1, $P<0.05$) and severe or very severe (-11.5, $P<0.001$) COPD participants, when compared to those without COPD.[33] Jones and colleagues (2011) added to these findings, reporting a significant decrease in HRQoL (measured by the SGRQ, the SF-12 and the Functional Assessment of Chronic Illness Therapy - Fatigue) with each increasing level of COPD severity.[118]

2.13.5 Symptoms

Voll-Aanerud and colleagues (2010) analyzed data from the European Community Respiratory Health Survey I and II to identify respiratory symptoms associated with HRQoL among those with and without COPD. HRQoL was measured using the SF-12. Linear regression models included data from 6,009 subjects, controlling for age, gender, smoking status, occupation, BMI, morbidities, and study center. Significant factors associated with HRQoL included wheezing, being woken up by chest

tightness, experiencing chronic cough and experiencing chronic phlegm production.[35] In a separate investigation of pulmonary function and HRQOL, Voll-Aanerud et al. (2008), also reported an negative association between dyspnea and HRQoL (-6.4, P<0.001).[33]

2.14 Overview of past research investigating factors affecting HRQoL among smokers

Given the causal link between smoking and several acute and chronic health outcomes, it is not surprising that multiple studies have shown an association between increased smoking exposure and reduced HRQoL. Most studies examined these effects by comparing established never, former and current smokers or they prospectively followed recent quitters to view the benefits of smoking cessation. Many smoking-related HRQoL investigations took place in clinical settings where patients were more likely to present with severe smoking effects. Participants of these studies were often diagnosed with a specific smoking-related illness or were undergoing a specific treatment which smoking is suspected of interfering with. For example, Garces and colleagues (2004) explored the relationship between smoking status and HRQoL among individuals diagnosed with lung cancer whereas Das and colleagues (2007) investigated the effect of smoking on short-term HRQoL among patients receiving sinus surgery.[26,27] According to a systematic review of general injured populations (studies including only those with a specific injury were excluded), hospitalized individuals were more likely to report lower HRQoL than non-hospitalized populations.[119] Results of clinical studies are extremely useful to their respective study populations; however, there is also a lack of external generalizability resulting from small sample sizes and specific inclusion criteria which the majority of the general public do not meet.

Whether clinical or population based, investigations of factors affecting HRQoL commonly utilized data from other studies which do not primarily focus on HRQoL.[14,16,28] Many ancillary studies exploring the impact of factors affecting HRQoL focused on the

effects of a specific variable (ie. gender), or variable group (ie. socio-economic indicators or cardiovascular risk factors), often excluding covariates known to affect HRQoL from analyses. This is also prevalent amongst ancillary studies investigating the impact of variables from multiple groupings as many factors unrelated to the primary study's objective, but known to affect HRQoL (ie. smoking status), were not measured.[29-31] Therefore, statistical analyses controlled for a selection of socio-economic factors, exposures, morbidities or functional abilities directly related to the study, but few studies accounted for several factors related to HRQoL.[83,95,120,121] Furthermore, several previous investigations did not attempt to identify non-linear functions between predictor variables and HRQoL.[32-37] With few studies including data from multiple variable groups and few analyses identifying non-linear relationships, an incomplete understanding of HRQoL exists.

This study builds on previous research by addressing the aforementioned issues. Ensuring strong study power, data from a large population sample of former and current smokers, the Pan-Canadian Early Detection of Lung Cancer Study (PanCan), representative of the Canadian population of smokers was used. Variation in HRQoL will be explained using an array of sociodemographic, medical history, exposure, symptom, morbidity, and pulmonary function variables. Linear and logistic regression models utilizing a multivariable fractional polynomial (MFP) selection procedure are used to explain variation in HRQoL.

Chapter 3: Methods

3.0 Overview

This chapter begins by describing the purpose of the PanCan, the primary study from which this project obtains its data. Following, the study design, patient recruitment, sample characteristics, and data collection techniques are discussed. Lastly, the statistical analysis strategies implemented to answer study questions are discussed.

3.1 Pan Canadian Early Detection of Lung Cancer Study

3.1.1 Purpose

The goal of the PanCan was to develop a new multi-modal screening strategy which would lead to the early detection of lung cancer. It is expected that screening high risk individuals for lung cancer would be effective in earlier detection of lung cancer, potentially improving cost effectiveness for the Canadian health care system if implemented.[24] Results of this study contribute to knowledge required to make an educated decision regarding the implementation of lung cancer screening in Canada.

3.1.2 Participant recruitment and sample size

PanCan recruitment took place in eight sites across Canada; Halifax, St. John's, Laval, Ottawa, Toronto, Hamilton, Calgary and Vancouver from September 2008 until December 2010. Participants responded to advertisements in newspapers, on television and through the offices of doctors and dentists. Sample size calculations estimated that 2500 participants were needed to draw definitive conclusions between biomarkers, spirometry and autofluorescence bronchoscopy use, and lung cancer.

3.1.3 Inclusion and exclusion criteria

To apply, potential participants were instructed to call a (1–800) phone number. Upon calling, a research assistant at the PanCan co-ordinating centre explained the study and collected lung cancer risk factor data which were entered into an Excel

spreadsheet to compute the prospective study subject's *lung cancer risk index*. The lung cancer risk was estimated according to a model prepared by Dr. Martin Tammemagi using Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial data.[122,123]

The purpose of the lung cancer risk index was to determine the probability potential study participants had of developing lung cancer. The prediction model probability is based on self-reported risk factors including age, education level, BMI, family history of lung cancer, history of COPD, chest x-ray in last three years, smoking duration and pack years smoked. Individuals were invited to enroll in the study if their risk of developing lung cancer within 5 years was ≥2%. In addition, a participant must have been between 50 and 75 years old and provided consent for screening procedures.

Potential participants were excluded if they had any medical condition such as an acute or chronic respiratory failure, bleeding disorder, or were unlikely to benefit from the screening being offered. Participants were also excluded if they were taking any oral anti-coagulants, had a known reaction to xylocaine, salbutamol, midazolam, and alfentanil, or were pregnant, unwilling to have a spiral chest computed tomography, had a chest CT within 2 years, were unwilling to sign consent or had been diagnosed with cancer (excluding non-melanomatous skin cancer, localized prostate cancer, carcinoma in situ of the cervix, or superficial bladder cancer) within 5 years.

3.1.4 Data collection techniques

In the event that a potential participant met inclusion criteria without presenting any exclusion criteria, he or she was registered as a study participant and administered four questionnaires. The first focused on multiple epidemiologic factors, including socio-demographic factors, smoking patterns, occupational exposures, family history of cancer and medical data. The second and third questionnaires; the short form 12 (SF-12) and the EuroQol Five Dimension (EQ-5D) were used to measure HRQoL. The final

questionnaire, the State-Trait Anxiety Inventory, measured anxiety intensity and distinguished between state anxiety and trait anxiety. Following the baseline questionnaires, a blood sample and spirometry measurements were taken. Participants were not financially rewarded for study participation.

Measuring quality of life: The European Quality of Life Scale (EQ-5D)

This thesis utilized the EQ-5D to measure HRQoL.[23] The EQ-5D consists of two parts. The first is the EQ-5D descriptive system, also referred to as the EQ-5D-3L (*Figure 3*). This system consists of five questions, each question targeting one specific dimension; mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each dimension has three levels of response; no problems, some problems and extreme problems. A total of 243 unique combinations, or "health profiles", exist.

Domain responses can be analyzed individually, by calculating odds ratios of reporting none, some or severe problems. Additionally, each health profile can also be transformed into a single index value ranging from 0 to 1. Transformation is completed by applying standardized formulas derived from the general population to respective health profiles. Therefore, index values are representative of the general population's opinions of a particular health state rather than the study participant's.[124]

The second part of the EQ-5D-3L is the Visual Analog Scale (VAS) which measures HRQoL using a utility value. Unlike index values which generate a HRQoL score based on question responses and mathematical formulas derived from the general population, a utility score, such as the EQ-5D VAS, is created by asking the user to rate their own HRQoL by providing a single numerical value *(Figure 4)*. Both components of the EQ-5D-3L were designed for use in postal surveys, clinics and face-to-face interviews, with various populations, and takes only a few minutes to complete.[23]

By placing a check-mark in one box in each group below, please indicate which statements best describe your own state of health today.

Mobility

I have no problems in walking about ❑

I have some problems in walking about ❑

I am confined to bed ❑

Self-Care

I have no problems with self-care ❑

I have some problems washing or dressing myself ❑

I am unable to wash or dress myself ❑

Usual Activities (*e.g. work, study, housework, family or leisure activities)*

I have no problems with performing my usual activities ❑

I have some problems with performing my usual activities ❑

I am unable to perform my usual activities ❑

Pain/Discomfort

I have no pain or discomfort ❑

I have moderate pain or discomfort ❑

I have extreme pain or discomfort ❑

Anxiety/Depression

I am not anxious or depressed ❑

I am moderately anxious or depressed ❑

I am extremely anxious or depressed ❑

Figure 3. EQ-5D-3L (Canadian English version)

To help people say how good or bad their state of health is, we have drawn a scale (rather like a thermometer) on which the best state you can imagine is marked 100 and the worst state you can imagine is marked 0.

We would like you to indicate on this scale how good or bad your own health is today, in your opinion. Please do this by drawing a line from the box below to whichever point on the scale indicates how good or bad your state of health is today.

Your own state of health today

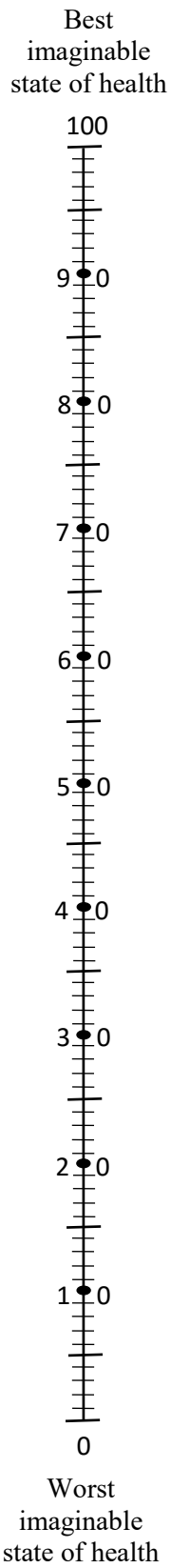

Figure 4. EuroQol Visual Analog Scale (Canadian English version)

EQ-5D: Validity and Reliability

The EQ-5D has been tested in various populations and has shown itself to be a valid and reliable measure of HRQoL. Investigations of over 1,500 participants from the general Canadian population residing in Alberta, and 423 Americans both concluded that the EQ-5D has strong construct, convergent and discriminant validity when compared to the 12 item Short Form Health Survey (SF-12).[125,126] Luo and colleagues (2005) used Spearman's correlations to conclude the EQ-5D to have strong discriminant, convergent and construct validity (R>0.5) when compared to the Health Utility Index 2 and Health Utility Index 3.[127] Kontodimopoulos et al. (2008) agreed with these findings and also reported that the EQ-5D's internal consistency reliability (α=0.743) exceeded the accepted level of 0.70 for group level comparisons.[128] In a general sample of 1,666 Israeli participants, Horowitz and colleagues (2010) concluded that the EQ-5D-3L was a responsive measure of HRQoL as it could adequately distinguish individuals with varying levels of disease as well as individuals with disease from those without. Pearson interclass correlation coefficients measuring test-retest reliability were also very high for both the EQ-5D (R=0.85) and the VAS (R=0.87).[129] Other validity and reliability investigations of the EQ-5D in a general Dutch population, those with COPD, asthma, musculoskeletal disorders, diabetes, cardiovascular disease and elderly populations reported similar findings.[8,130-134]

Spirometry Measurements

Spirometric measurements included in this study were Forced Expiratory Volume in one second (FEV_1) and Forced Vital Capacity (FVC). These measurements were used as an estimate of lung functioning and were analyzed as a percent of predicted (% predicted) using standardized prediction equations.[135] A flow-sensitive spirometer

(Presto Flash Portable Spirometer Version 1.2, Spacelab Burdick Inc., Deerfield WI) was utilized by trained technicians to collect these data.[136]

3.1.5 Lung cancer screening and study follow-up

Study participants were followed for 5 years. All participants were screened for lesions using low-dose computed tomography (LDCT) and half were also screened using autofluorescence bronchoscopy. If no nodules were detected, participants completed a follow-up questionnaire or telephone interview every 6 months and were required to be re-screened by LDCT one year after baseline. If nodules were detected, follow-up was determined by the size, shape, and growth of existing nodules. Information gathered throughout follow-up will not be discussed further as this investigation only utilizes baseline data.

3.2 Ethical considerations

Study investigators were responsible for clearly and thoroughly explaining all aspects of the PanCan (NCT00751660) as well as answering any questions potential participants may have had. Ethics approval was obtained from each participating study center, and all potential participants were given informed consent forms to sign. All data provided for analysis in the current study to the student, Matthew Ventresca, have been de-identified, so identification of individuals has not been possible. This thesis is a secondary data analysis and does not require ethical approval from Brock University.

3.3 Statistical analysis

All statistical analyses were completed using Stata version 13.[137] Statistics describing the study population and distributions of study variables were calculated. Chi-squared tests were used to compare frequencies for nominal and ordinal level data. All continuous variables were non-normally distributed and analyzed using non-parametric test of trend.[138] Major research questions were investigated using linear and logistic

regression techniques with a multivariable fractional polynomial (MFP) selection procedure. Selection of variables included in regression analyses were based on previous research and PanCan availability (Table 1).

Table 1. Variables assessed for potential inclusion in regression analyses.

Variable category	Possible factors associated with HRQoL
Independent variables	
Socio-demographic and medical history	Age, gender, race/ethnicity, education level, current work status, body mass index, familial cancer history
Smoking, Alcohol Consumption & Other Exposures	Smoking status, cigarettes smoked per day (30 days before study entry), cigarettes smoked per day (average throughout smoking period), pack years smoked, smoking duration, time since quitting smoking, weekly alcohol consumption, monthly marijuana use
Morbidities	Allergies, anemia, angina, arthritis, asthma, blood disease, COPD, diabetes, forced expiratory volume in one second % predicted, forced vital capacity, hypercholesterolemia, hypertension, kidney disease, liver disease, myocardial infarction, osteoporosis, peripheral vascular disease, pulmonary fibrosis, stroke
Symptoms	Dyspnea, cough, phlegm, hoarse voice, wheezing, chest pain, fatigued, poor appetite, weight loss
Dependent variables	
Quality of Life	EQ-5D VAS, EQ-5D domains (mobility, self-care, usual activities, pain/discomfort and anxiety/depression)

Question #1. What socio-economic, medical history, exposure, symptom, morbidity, and pulmonary function factors are associated with HRQoL?

A MFP selection procedure was used in conjunction with linear and logistic regression techniques to identify variables associated with HRQoL, measured by the EQ VAS and EQ-5D. Utilizing VAS response data allows for the identification of factors associated with HRQoL based on the opinion of the study participant, as opposed to index values which are often used in economic analysis and reflect opinions of the general population. Covariates included in regression models were selected from data collected by the PanCan and were based on previous research outlined in chapter 2.

Assumptions of linear regression were evaluated. To determine if VAS scores and potential transformations were normally distributed, histograms and tests for skewness and kurtosis were examined. A plot of residuals against fitted values of VAS

scores was created to assess homoscedasticity. Pearson correlation coefficients were calculated and examined for all independent variables. According to Doormann and colleagues, collinearity begins to severely distort model estimations and subsequent prediction when Pearson's R≥0.7.[139] When correlation coefficients exceeded 0.7, the variable which had a weaker relationship with VAS score was excluded from regression modelling provided that previous literature verified omission was scientifically valid.

The Stata command ***mfp****:* ***reg*** was utilized to complete regression analyses. The selection of polynomial exponents available for testing models was left at default values (-2, -1, 0.5, 1, 2, 3). Maximum number of cycles and degrees of freedom were kept at their default levels; 5 and 4, respectively. Cluster adjustment according to a participant's site of study participation was applied to all multivariable regression models.[140] Continuous variables were centered to their mean and rounded to the nearest whole number for ease of interpretation. Mickey (1989) reports that the use of a typical P-value set at 0.05 may result in important variables being excluded from the regression model.[141] Sauerbrei and colleagues (2005) also discuss the importance of this issue in regression with an MFP selection procedure.[142] To ensure that all potential covariates were considered for selection into the final regression model, the cut point for entry was set at α=0.2 in partially adjusted regression models. Four partially adjusted regression models were prepared prior to each fully adjusted linear or logistic regression model, one per variable group (socio-demographics and medical history, exposures, morbidities and symptoms). Independent variables with P≤0.2 in each partially adjusted regression model collectively underwent the MFP selection process to determine which were significantly associated with VAS score when adjusting for all other factors. Statistical significance in the fully adjusted model was set at P≤0.05. For continuous predictors, a significant non-linear relationship was identified if P≤0.05 when non-linear was compared to linear predictors in nested models. The variation in VAS score explained by models

was assessed using R^2 values. Monitoring changes in effect estimates, significance values and variable pathway position in unadjusted and adjusted models allowed for the identification of confounders as well as potential upstream and downstream associations (Figure 5). Upstream associations involve variables, distal in pathway position from HRQoL, which are associated with an intermediary variable which is proximal and associated with HRQoL. Downstream associations are typically associated with HRQoL after full adjustment with no intermediary associations between itself and HRQoL. Variables with a distal pathway position may have upstream and downstream associations with HRQoL, which is common among socio-demographic characteristics but is also possible among other variable groups.

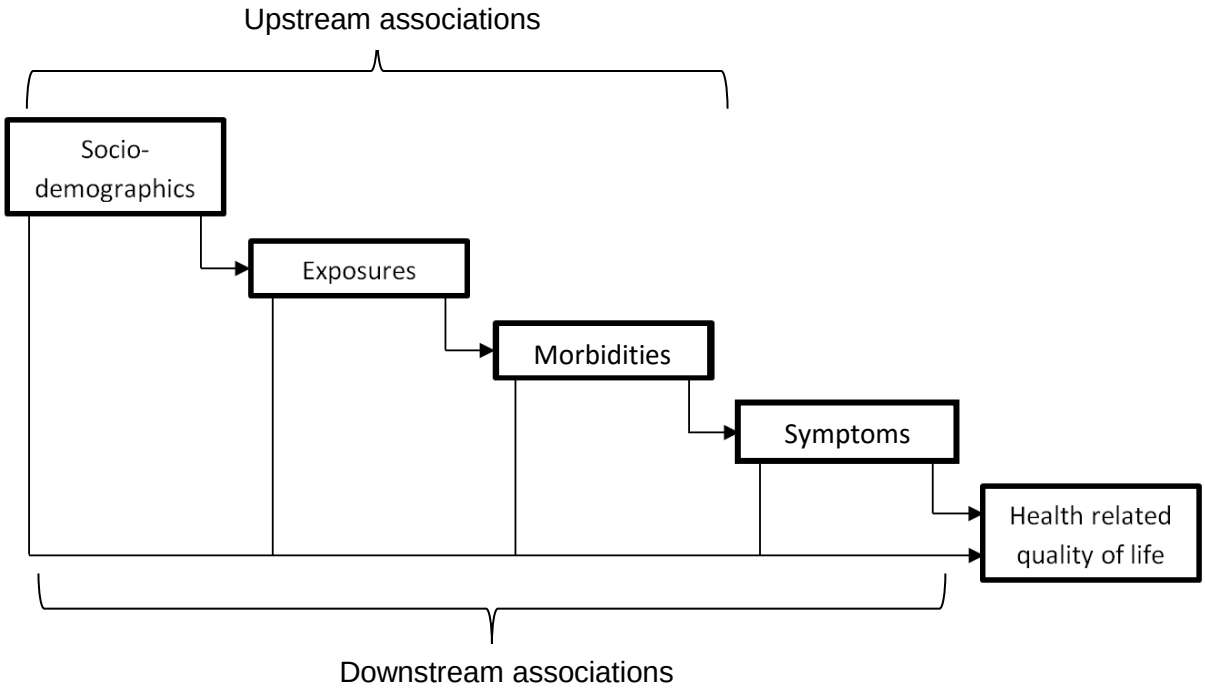

Figure 5. Pathway position, and upstream and downstream associations

3.3.1 MFP selection procedure

The traditional linear regression equation ($\beta_1 X$) is often suitable to modelling the relationship between predictor and outcome variables. However, a better non-linear fit

may be possible. The MFP selection procedure extends the linear regression equation ($\beta_1 X$) by adding an exponent and, potentially, additional polynomial terms ($\beta_1 X^{p1} + \beta_1 X^{p2} + \beta_2 X^{p1} + \beta_3 X^{p1} \ldots$).[142] Model complexity is controlled by two hypothesis tests, where $\alpha=0.05$ is typical for both. The first hypothesis test determines if an independent variable is significant to the regression model using up to two polynomials. The second test determines whether a non-linear fit is more appropriate.

Prior to starting the MFP selection process, significance levels for model inclusion ($\alpha_1=0.05$), non-linearity testing ($\alpha_2=0.05$) and the maximum degrees of freedom must be chosen (4). The full linear model is then fitted. The visiting order for subsequent comparisons is determined by the respective P-value for omitting each term from the model, beginning with the most significant predictors.

The MFP selection process begins. If the variable assessed is categorical, the joint significance of its dummy variables is tested at α_1 while other predictors are included as adjustment terms. If the result is insignificant the variable is removed from regression modelling, a significant result retains the variable. For continuous predictors, a two term polynomial is used to describe the relationship with the dependent variable. Assuming the term is significant and should be added to the model, it is then compared to the linear fit. An insignificant test result indicates that the linear fit is preferred. Significant results indicate that a non-linear fit is more appropriate. A following test comparing a single polynomial function to the dual polynomial function is completed. An insignificant result indicates that the single polynomial form is preferred to the dual polynomial function. Remaining variables are tested until none remain. This process will repeat up to five times, however, with a new list of predictors excluding those found to be previously insignificant.[142] This backwards selection process, which repeats until all independent variables significantly predict the dependent variable, reduces the likelihood for negative confounding variables to be excluded from the final regression model.[143]

Linear regression assumptions tested include; 1) linear relationship between continuous independent variables and the dependent variable; 2) independence of observations; 3) the dependent variable, VAS score, is normally distributed; 4) correlation of predictors; 5) reducing measurement error and 6) homoscedasticity.[144,145]

It is important to note that several comparisons are being made, increasing the probability of identifying erroneous significant relationships.[146] To ensure no significant associations are overlooked the significance level is not decreased, as would be done in a Bonferroni correction for multiple comparisons. However, bootstrap confidence intervals are also calculated to verify consistency of the regression model.

3.3.2 Bootstrapping stability analysis

The Stata command **mfpboot** was utilized to calculate 1000 bootstrap replications. Bootstrap samples were derived from the study sample and are the same size. Each observation was replaced after being selected. Therefore, some individuals were selected more than once and others not at all.[147] Since it is non-parametric, no statistical assumptions need to be satisfied.[147] Model stability was assessed by analyzing bootstrap inclusion frequencies (BIF). While not a hypothesis test, variables with a BIF≥50% are typically considered as having a $P<0.05$.[145]

To test for interaction effects new variables were created by multiplying independent variables suspected of having an interaction effect together (ie. gender * smoking duration). All interactions tested were chosen a priori, and based on literature review. Tested interactions focused on the relationship between socio-demographic variables in combination with each other and various exposure and morbidity factors. All interactions were tested univariately (in the presence of both independent variables) and in the fully adjusted linear regression model. Likelihood ratio tests were completed to determine if significant interactions existed.

Question #2. What socio-economic, medical history, exposure, morbidity and symptom variables are associated with each EQ-5D domain?

Logistic regression models, utilizing the MFP selection procedure, were used to identify socio-economic, medical history, exposure, morbidity and symptom variables significantly associated with reporting some or severe problems for each EQ-5D domain. Odds ratios of experiencing some or severe problems, in comparison to no problems, were calculated. Ideally, ordinal logistic regression would be used since each domain has three outcome levels. However, due to low frequencies of individuals reporting severe problems, which is common in population based data sets where participants tend to report higher HRQoL than clinical settings, individuals experiencing some or severe problems were combined into a single group.[23] Abreu and colleagues (2008) note that dichotomizing response variables, especially for HRQoL studies, may lead to loss of information and less appropriate or incorrect conclusions.[148] Ordinal logistic regression of three domain levels would be preferred, however, dichotomization is endorsed in the EQ-5D manual in this circumstance.[23]

Logistic regression requires that the dependent variable be dichotomous and that each subgroup be mutually exclusive. It is recommended that at least 50 participants be included in the regression model for each additional independent variable. The independent variables were not assumed to be linearly related to the dependent variable and are not required to be of equal variance within each group.[149] To complete logistic regression modelling the Stata command ***mfp: logistic*** was utilized with P=0.05. McFadden's pseudo R^2 was analyzed to assess model fit.[150] Similar to linear regression modelling, variables were included in univariate and partially adjusted models to determine which were included in each domain specific fully adjusted logistic regression model.

Chapter 4: Results

4.1 Population demographics and tobacco exposure

Statistics describing the socio-demographic, medical history, and smoking exposure of all PanCan participants are found in *Table 2, Table 3a and Table 3b.* A total of 2537 participants enrolled into the PanCan. The study participants' median age was 62 (mean=62.3; range 50 to 76; IQR: 58 to 67) years. The sample was 55% males and 45% females, and largely white (97.4%). At least some college education was completed by 47.1% of the sample. Over half of participants, 51.3%, were retired while 38.2% were employed. Median BMI of the sample was 25.8 kg/m^2 (mean=26.6; range 14.0 to 53.8; IQR: 23.6 to 28.9) with 935 (36.9%) participants falling within the normal BMI range.

Almost two thirds, 62.2%, of the sample defined themselves as current smokers. Current smokers were significantly younger than former smokers (mean=61.3 vs. mean=64.0, P<0.001) and smoked approximately 18 cigarettes daily, slightly less than one pack per day, within the 30 days prior to baseline. Collectively, the median smoking duration was 44 years (mean=44.1; SD= 5.9; IQR 40 to 48) with a median of 25 cigarettes per day throughout their smoking lifetime (mean=24.7; SD= 10.6; IQR 20 to 25). Approximately 1% of the study population smoked for fewer than 30 years. Compared to all other educational levels, participants with ≤8^{th} grade education had the longest smoking duration, approximately 47 years, and the greatest number of cigarettes smoked per day (smoking period), approximately 21. The median time since quitting smoking for former smokers was 5 years (mean=5.8; SD=4.3; IQR 2 to 10). Unadjusted mean time since quitting smoking did not significantly differ among socio-demographic subgroups. However, individuals with a familial history of cancer quit smoking, on average, almost 1 year prior to those who did not (mean=6.3 vs. mean=5.6, P=0.01).

In total, 2513 (99.1%) participants provided VAS baseline data (mean=76.8; SD=14.3; range: 1 to 100, IQR: 70 to 89) (Figure 6). Statistics describing VAS score

distribution by socio-demographic, medical history, exposure, symptom or morbidity variables are available in Table 4, Table 6, Table 8, and Table 10.

Table 2. Distributions of continuous variables

Variable	N	Median	Mean	SD	Range	IQR
Socio-demographic and medical history						
Age	2537	62	62.3	5.8	50 to 76	58 to 67
Body mass index	2537	25.8	26.6	4.4	14.0 to 53.8	23.6 to 28.9
Exposures						
Cigarettes per day (last 30 days)*	2531	18	17.8	10.8	0 to 80	10 to 25
Cigarettes per day (smoking period)	2537	25	24.7	10.7	1 to 100	20 to 25
Smoking duration	2537	44	44.1	5.9	11 to 69	40 to 48
Pack years†	2537	50	54.1	23.4	2.2 to 230	41 to 62.5
Time since quitting‡	2537	5	5.8	4.3	0.5 to 22	2 to 10
Average alcoholic drinks per week	2537	10	15.9	23.3	0 to 273	3 to 20
Marijuana use (joints per month)	2508	0	3.6	13.5	0 to 450	0 to 0
Symptoms						
Total symptoms†	2537	3	2.8	2.1	0 to 9	1 to 4
Morbidities						
FVC % predicted†	2517	0.9	0.92	0.17	0.35 to 1.76	0.81 to 1.03
FEV_1 % predicted	2517	0.8	0.82	0.19	0.15 to 1.68	0.70 to 0.94
Total morbidities†	2537	3	2.8	2.0	0 to 14	1 to 4

Abbreviations: FVC, forced vital capacity; FEV_1, forced expiratory volume in one second; IQR, interquartile range; N, sample size; SD, standard deviation.

* Current smokers only.

† Excluded from regression modelling.

‡ Former smokers only.

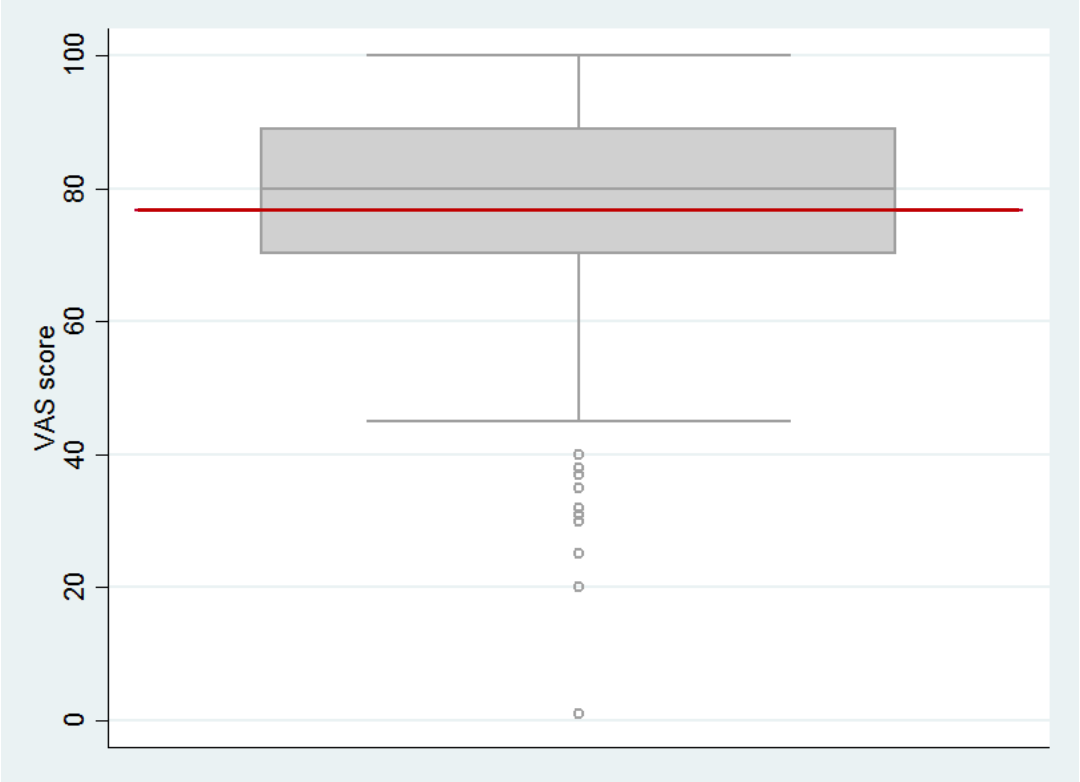

Figure 6. EQ VAS score boxplot with VAS score mean in red.

Table 3a. Smoking exposure by socio-demographic and medical history variables

Socio-demographic and medical history variables	%	Current (N=1579) N (%)	Former (N= 958) N (%)	P‡	Cigarettes per day (last 30 days)* Mean (SD)	P‡	Cigarettes per day (during smoking period) Mean (SD)	P‡	Smoking duration (years) Mean (SD)	P‡	Time since quitting (years)† Mean (SD)	P‡
Age, mean (SD)	**N=2537**	61.3 (5.8)	64.0 (5.6)	<0.001								
50 to 55	10.9	218 (79.0)	58 (21.0)	<0.001	21.2 (11.1)	<0.001	28.3 (11.5)	<0.001	37.5 (3.2)	<0.001	0.6 (2.0)	<0.001
56 to 60	19.8	369 (73.2)	135 (26.8)		20.1 (12.0)		27.2 (11.1)		41.3 (3.3)		0.9 (2.2)	
61 to 65	31.8	513 (63.5)	295 (36.5)		17.0 (10.4)		24.1 (10.4)		43.8 (4.7)		1.9 (3.5)	
66 to 70	25.5	345 (53.1)	305 (46.9)		15.5 (9.6)		23.2 (9.8)		46.5 (5.5)		3.3 (4.5)	
> 70 years	11.8	134 (44.8)	165 (55.2)		14.8 (8.6)		22.5 (10.1)		50.2 (6.0)		4.5 (5.0)	
Gender	**N=2537**											
Female	44.8	705 (62.1)	431 (37.9)	0.028	16.9 (10.2)	0.010	23.1 (9.6)	<0.001	43.5 (5.7)	<0.001	5.9 (4.2)	0.306
Male	55.2	874 (62.4)	527 (37.6)		18.5 (11.3)		26.1 (11.3)		44.5 (6.0)		5.7 (4.4)	
Race/ethnicity	**N=2526**											
White	97.0	1519 (61.8)	941 (38.3)	0.005	17.8 (10.8)	0.105	24.8 (10.7)	0.130	44.1 (5.9)	0.956	5.9 (4.3)	0.055
Non-white	3.0	52 (78.8)	14 (21.2)		16.1 (11.0)		22.2 (10.4)		44.3 (6.3)		4.0 (4.4)	
Education	**N=2537**											
≤ 8th grade	3.0	56 (74.7)	19 (25.3)	0.008	20.6 (14.8)	<0.001	29.4 (13.3)	<0.001	47.2 (7.0)	0.374	27.3 (12.7)	<0.001
9th to 11th grade	13.2	207 (61.8)	128 (38.2)		20.4 (11.9)		25.7 (12.2)		45.1 (6.1)		23.6 (9.4)	
High school	26.2	444 (66.9)	220 (33.1)		18.7 (10.8)		24.8 (10.0)		42.7 (5.9)		24.0 (9.0)	
Technical certificate	10.6	168 (62.2)	102 (37.8)		17.2 (10.2)		24.2 (10.7)		44.3 (6.2)		23.2 (9.7)	
Associate degree	19.2	291 (59.9)	195 (40.1)		18.0 (10.7)		24.7 (10.3)		44.1 (5.5)		23.3 (9.2)	
Bachelor's degree	17.0	246 (57.1)	185 (42.9)		15.8 (9.5)		24.3 (10.5)		44.0 (5.6)		22.0 (8.7)	
Advanced degree	10.9	167 (60.5)	109 (39.5)		14.1 (9.1)		23.4 (10.1)		44.9 (5.0)		21.9 (8.5)	

Abbreviations: BMI, body mass index; N, sample size; P, P-value; SD, standard deviation.

***** Data presented for current smokers only.

† Data presented for former smokers only.

‡ P-values were computed for categorical comparisons using chi-square tests. P-values for comparisons of continuous variables age and BMI, which are non-normally distributed, were calculated using non-parametric test for trend.

Table 3b. Smoking exposure by socio-demographic and medical history variables

Socio-demographic and medical history variables		Current (N=1579)	Former (N= 958)		Cigarettes per day (last 30 days)*		Cigarettes per day (during smoking period)		Smoking duration (years)		Time since quitting (years)†	
	%	N (%)	N (%)	*P*‡	Mean (SD)	*P*‡	Mean (SD)	*P*‡	Mean (SD)	*P*‡	Mean (SD)	*P*‡
Employment status	**N=2520**											
Employed	38.2	654 (67.9)	309 (32.1)	<0.001	17.9 (10.4)	0.087	25.4 (10.9)	0.007	42.2 (5.2)	<0.001	4.9 (4.4)	0.144
Retired	51.3	702 (54.3)	592 (45.8)		16.5 (9.8)		23.7 (9.9)		45.8 (6.0)		6.4 (4.3)	
Unemployed	2.9	54 (75.0)	18 (25.0)		22.0 (10.8)		28.3 (11.9)		41.7 (4.9)		4.3 (3.9)	
Disabled	3.5	79 (88.8)	10 (11.2)		23.4 (14.6)		29.0 (13.0)		42.5 (4.5)		3.5 (3.1)	
Sick leave	0.5	11 (84.6)	2 (15.4)		30.5 (25.9)		28.9 (14.1)		42.3 (3.8)		1.5 (0.7)	
Other	3.5	67 (75.3)	22 (24.7)		18.3 (12.6)		23.9 (10.7)		43.9 (5.7)		5.5 (4.6)	
BMI, kg/m^2, mean (SD)	**N=2537**	27.4 (4.3)	26.0 (4.3)	<0.001								
Underweight	1.1	22 (81.5)	5 (18.5)	<0.001	18.7 (8.8)	0.005	23.4 (10.0)	<0.001	45.8 (4.6)	0.162	5.0 (5.9)	0.970
Normal	36.9	659 (70.5)	276 (29.5)		17.0 (10.3)		21.8 (10.8)		44.0 (5.9)		5.7 (4.4)	
Overweight	45.2	682 (59.5)	464 (40.5)		17.8 (10.5)		24.5 (10.2)		43.8 (5.9)		6.0 (4.3)	
Obese	16.9	216 (50.4)	213 (49.7)		20.2 (12.9)		28.4 (12.3)		44.9 (5.7)		5.4 (4.8)	
Familial cancer history	**N=2495**											
No	66.4	1030 (62.2)	626 (37.8)	0.993	17.6 (10.5)	0.428	24.6 (10.6)	0.738	45.0 (5.5)	<0.001	5.6 (4.3)	0.023
Yes	33.6	522 (62.2)	317 (37.8)		18.1 (11.5)		25.1 (10.7)		42.1 (6.0)		6.3 (4.4)	

Abbreviations: BMI, body mass index; N, sample size; P, P-value; SD, standard deviation.

***** Data presented for current smokers only.

† Data presented for former smokers only.

‡ P-values were computed for categorical comparisons using chi-square tests. P-values for comparisons of continuous variables age and BMI, which are non-normally distributed, were calculated using non-parametric test for trend.

4.2 Linear regression modelling

4.2.1 Socio-demographic and medical history variables

Statistics describing the relationship between socio-demographic and medical history variables, and VAS score are outlined in Table 4. Despite increasing age, the proportion of participants reporting VAS scores ≥80, the median, increased in successive age groups. Non-whites reported significantly lower VAS scores than white participants (mean=71.1 vs. mean=77.0; P=0.014). Of all socio-demographic subgroups, advanced degree holders had the greatest proportion of participants above the median (66.7%) and also reported the highest mean VAS score, 80.0. Employed (mean=78.0) and retired (mean=77.8) participants reported highest VAS scores among working groups. Disabled participants reported the lowest mean VAS score (59.5) of any socio-demographic subgroup and had the highest proportion (80.9%) of participants reporting VAS scores below the median. Participants in the normal BMI range reported the highest VAS scores. No significant differences in proportion or mean VAS score are observed between genders or participants with a familial history of cancer.

Table 4. Frequencies and descriptive statistics of VAS score by socio-demographic and medical history variables

Socio-demographic and medical history variables	N (%)	EQ-5D VAS score dichotomized at median			EQ-5D VAS score	
		≤ 79 N (%)	≥ 80 N (%)	*P**	Mean (SD; IQR)	*P**
Age	**2537**	**1085**	**1452**			
50 to 54	276 (10.9)	140 (50.7)	136 (49.3)	<0.001	73.9 (15.7; 70 to 75)	<0.001
55 to 59	504 (19.9)	242 (48.0)	262 (52.0)		75.0 (15.8; 70 to 80)	
60 to 64	808 (31.8)	339 (42.0)	469 (58.0)		77.2 (13.5; 70 to 80)	
65 to 69	650 (25.6)	252 (38.8)	398 (61.2)		78.5 (13.1; 70 to 80)	
> 70	299 (11.8)	112 (37.5)	187 (62.5)		77.7 (13.9; 70 to 80)	
Gender	**2537**	**1085**	**1452**			
Female	1136 (44.8)	472 (41.6)	664 (58.5)	0.264	77.2 (14.3; 70 to 80)	0.167
Male	1401 (55.2)	613 (43.8)	788 (56.3)		76.4 (14.2; 70 to 80)	
Race/ethnicity	**2526**	**1079**	**1447**			
White	2460 (97.4)	1046 (42.5)	1414 (57.5)	0.225	77.0 (14.1; 70 to 80)	0.014
Non-White	66 (2.6)	33 (50.0)	33 (50.0)		71.1 (17.8; 60 to 79)	
Education	**2537**	**1085**	**1452**			
8^{th} grade or less	75 (3.0)	43 (57.3)	32 (42.7)	0.001	72.0 (18.1;60 to 72)	<0.001
9^{th} to 11^{th} grade	335 (13.2)	164 (49.0)	171 (51.0)		74.1 (15.1; 66.5 to 80)	
High school graduate	664 (26.2)	288 (43.4)	376 (56.6)		76.5 (14.32; 70 to 80)	
Technical certificate	270 (10.6)	111 (41.1)	159 (58.9)		77.3 (14.4; 70 to 80)	
Associate degree/ Some college	486 (19.2)	203 (41.8)	283 (58.2)		77.1 (15.0; 70 to 80)	
Bachelor's degree	431 (17.0)	184 (42.7)	247 (57.3)		77.3 (12.7; 70 to 80)	
Advanced degree	276 (10.9)	92 (33.3)	184 (66.7)		80.0 (11.8; 74.5 to 80)	
Employment status	**2520**	**1076**	**1444**			
Working	963 (38.0)	400 (41.5)	563 (58.5)	<0.001	78.0 (12.7; 70 to 80)	<0.001
Retired	1294 (51.0)	509 (39.3)	785 (60.7)		77.8 (13.7; 70 to 80)	
Unemployed	72 (2.8)	48 (66.7)	24 (33.3)		66.9 (17.9; 55 to 70)	
Disabled	89 (3.5)	72 (80.9)	17 (19.1)		59.5 (18.1; 50 to 60)	
Sick leave	13 (0.5)	9 (69.2)	4 (30.8)		60.8 (21.5; 40 to 60)	
Other	89 (3.5)	38 (42.7)	51 (57.3)		77.6 (14.2; 70 to 80)	
Body mass index	**2537**	**1085**	**1452**			
Underweight	27 (1.1)	18 (66.7)	9 (33.3)	<0.001	70.6 (15.7; 60 to 70)	<0.001
Normal	935 (36.9)	323 (34.5)	612 (65.5)		79.0 (13.2; 70 to 80)	
Overweight	1146 (45.2)	491 (42.8)	655 (57.2)		77.0 (13.8; 70 to 80)	
Obese	429 (16.9)	253 (59.0)	176 (41.0)		71.7 (16.1; 65 to 75)	
Familial cancer history	**2495**	**1063**	**1432**			
No	1643 (66.5)	704 (42.5)	952 (57.5)	0.895	76.9 (14.2; 70 to 80)	0.781
Yes	828 (33.5)	359 (42.8)	480 (57.2)		76.5 (14.5; 70 to 80)	

Abbreviations: IQR, interquartile range; N, sample size; P, P-value; SD, standard deviation.
* P-values calculated for proportional comparisons with chi-squared test and non-parametric test of trend for comparisons of VAS score.

Univariate and adjusted relationships between socio-demographic and medical history variables, and VAS score are described in Table 5. Significant differences were present in all nonparametric tests for trend and univariate regression models with the exception of gender and familial cancer history. Unadjusted models revealed a

significant non-linear relationship between age and VAS score (P=0.013). These effects were explained away after further adjustment, indicating a possible upstream association between age and VAS score. The locally weighted scatter plot smoothed relationship, which more accurately depicts the univariate relationship between age and VAS score, is depicted in Figure 7.

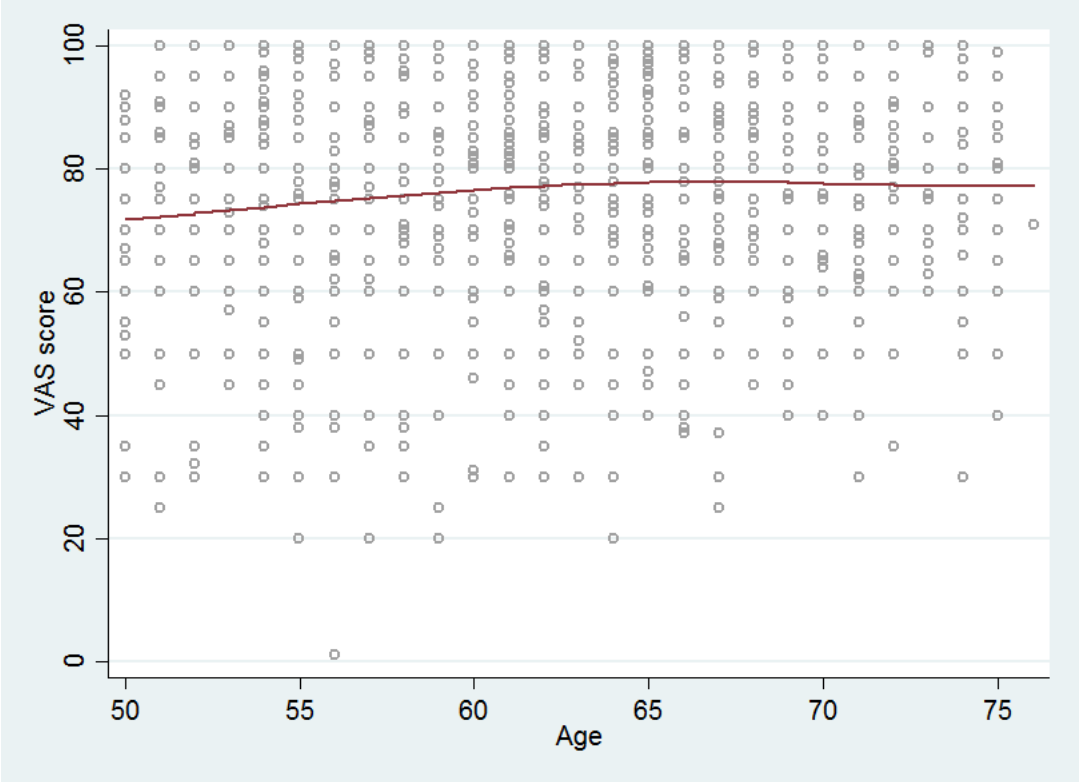

Figure 7. Locally weighted scatterplot smoothing of VAS score and participant age at baseline

Gender was statistically insignificant in all regression models. Race/ethnicity significantly predicted VAS score in all regression models. VAS scores tended to increase with education level, depicting a positive linear relationship. Compared to employed participants, the unemployed, disabled or persons on sick leave reported significantly lower VAS scores in all models. Unadjusted mean VAS scores for these groups ranged from approximately 11 to 18 points lower than employed or retired participants. VAS scores of retirees did not significantly different from the employed.

A significant non-linear relationship was observed between BMI and VAS score in univariate (P<0.001) and adjusted models (partially adjusted: P<0.001, fully adjusted: P=0.012, Figure 8). Familial cancer history was insignificant in all models tested. The fully adjusted model (to be described) which included all relevant variables (R^2=0.284), explained 0.163 more variation in VAS score than the preliminary model which included only socio-demographic variables (R^2=0.121).

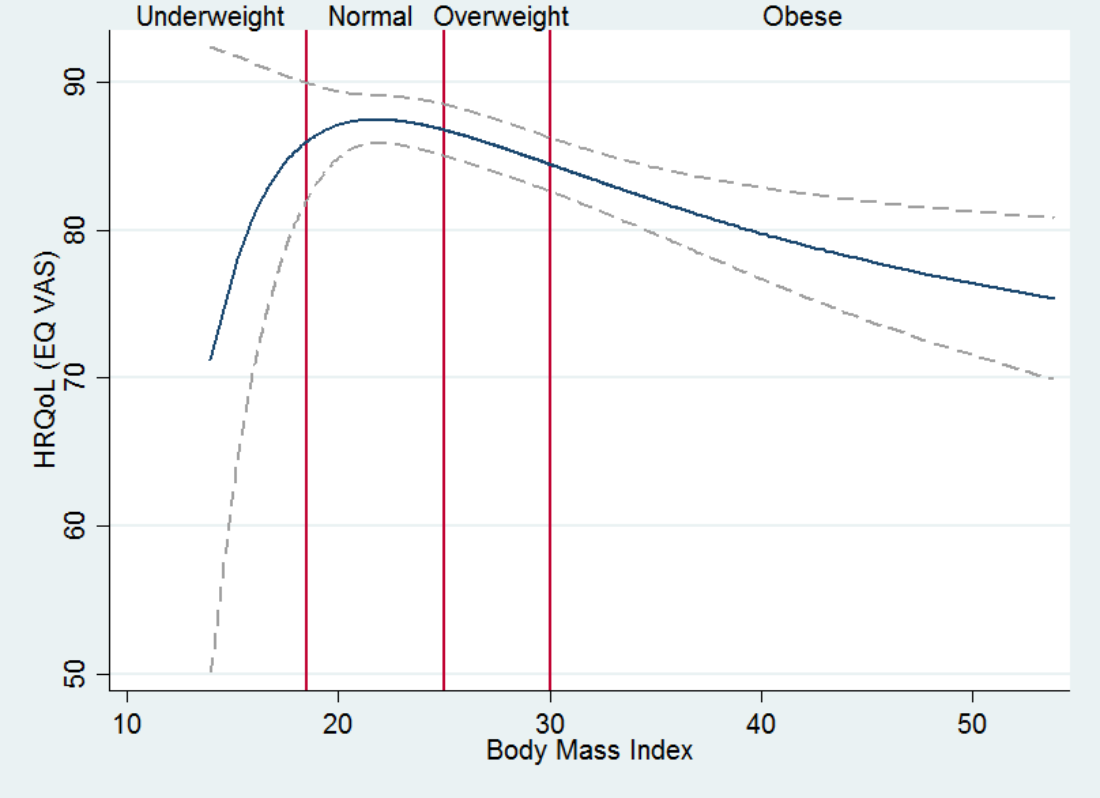

Figure 8. Fully adjusted non-linear relationship and confidence interval between BMI and EQ-5D VAS, N=2397

Table 5. Univariate and linear regression models between socio-demographic and medical history factors, and VAS scores

Socio-demographic and medical history variables	Univariate model	Partially adjusted model (N=2447, R^2=0.121)	Fully adjusted model (N=2397, R^2=0.284)*
	β (CI; P-value)†	β (CI; P-value)†	β (CI; P-value)†
Age			
1st term	0.66 (0.33 to 0.98; <0.001)‡	0.16 (0.09 to 0.24; 0.001)	(P=0.084)
2nd term	-0.29 (-0.44 to -0.14; <0.001) ‡		
Gender (male vs. female)	-0.79 (-1.91 to 0.33; 0.168)	-0.91 (-1.86 to 0.037; 0.057)	(P=0.054)
Race/ethnicity (non-white vs. white)	-2.06 (-3.26 to -0.85; 0.001)	-5.44 (-8.31 to -2.56; 0.003)	-3.88 (-7.59 to -0.18; 0.042)
Education	0.88 (0.55 to 1.21; <0.001)	0.49 (0.19 to 0.80; 0.007)	(P=0.939)
Employment status (vs. employed) §			
Retired	-0.15 (-1.30 to 1.00; 0.803)	-1.06 (-2.64 to 0.53; 0.159)	-0.67 (-1.85 to 0.51; 0.220)
Unemployed	-11.00 (-14.3 to -7.7; <0.001)	-10.07 (-14.26 to -5.87; 0.001)	-6.66 (-11.30 to 2.02; 0.012)
Disabled	-18.50 (-21.5 to -15.5; <0.001)	-17.24 (-21.1 to -13.4; <0.001)	-11.30 (-14.60 to -8.14; <0.001)
Sick leave	-17.20 (-24.7 to -9.69; <0.001)	-16.01 (-29.9 to -2.09; 0.030)	-12.15 (-20.68 to -3.62; 0.012)
Other	-0.37 (-3.36 to 2.62; 0.808)	-0.59 (-4.04 to 2.85; 0.632)	0.04 (-4.38 to 4.45; 0.985)
BMI‖			
1st term	-117.00 (-155.0 to -80.6<0.001)¶	-96.05 (-178.7 to -13.4; 0.029)**	-59.27 (-147.79 to 29.24; 0.157) ††
2nd term	373.00 (284.0 to 463.0; <0.001)¶	324.8 (146.8 to 502.8; 0.004)**	212.85 (12.41 to 413.28; 0.040) ††
Familial cancer history	-0.35 (-1.54 to 0.85; 0.570)	(P=0.365)	Excluded

Abbreviations: β, beta-coefficient; CI, confidence interval; N, sample size.

***** Final MFP linear regression model contains socio-demographic, smoking/alcohol exposure, morbidities and symptoms which were significantly associated with VAS score in partially adjusted MFP models. Only socio-demographic and medical history variables are displayed in this table.

† Beta-coefficients, confidence intervals and p-values calculated using MFP linear regression models.

‡ Age has a non-linear relationship with VAS score, and the following two terms in the regression model describes this relationship;

$$Term\ 1 = \left[\left(\frac{age}{10}\right)^3 - 242.101493\right] * 0.6559657$$

$$Term\ 2 = \left[\left(\frac{age}{10}\right)^3 * \ln\left(\frac{age}{10}\right) - 442.9938443\right] * (-0.2921975)$$

§ When testing the total importance of employment status in the fully adjusted model, p< 0.001.

‖ When testing the total importance of BMI in the fully adjusted model, p< 0.001.

¶ BMI has a non-linear relationship with VAS score, and the following two terms in the regression model describes this relationship;

$$Term\ 1 = \left[\left(\frac{BMI}{10}\right)^{-2} - 0.141899438\right] * (-117.8698)$$

$$Term\ 2 = \left[\left(\frac{BMI}{10}\right)^{-2} * \ln\left(\frac{BMI}{10}\right) - 0.138539022\right] * (373.8621)$$

** BMI has a non-linear relationship with VAS score, and the following two terms in the regression model describes this relationship;

$$Term\ 1 = \left[\left(\frac{BMI}{10}\right)^{-2} - 0.1371742112\right] * (-96.04738)$$

$$Term\ 2 = \left[\left(\frac{BMI}{10}\right)^{-2} * \ln\left(\frac{BMI}{10}\right) - 0.1362485285\right] * (324.8107)$$

†† BMI has a non-linear relationship with VAS score, and the following two terms in the regression model describes this relationship;

$$Term\ 1 = \left[\left(\frac{BMI}{10}\right)^{-2} - 0.1371742112\right] * -59.27362$$

$$Term\ 2 = \left[\left(\frac{BMI}{10}\right)^{-2} * \ln\left(\frac{BMI}{10}\right) - 0.1362485285\right] * 212.845$$

4.2.2 Smoking and alcohol exposure

Statistics describing the variation in EQ-5D VAS score by smoking and alcohol exposure variables are presented in Table 6. VAS score of current smokers was approximately 3 points lower than former smokers (mean=75.6 vs. mean=78.7; P<0.001). The proportion of participants reporting VAS scores below the median value decreased by approximately 10% between current and former smokers. Almost two thirds of the study sample reported smoking durations between 40 and 50 years, approximately 99% of participants reported smoking for at least 30 years. In the 30 days prior to baseline, slightly above 10% of current smokers smoked one pack of cigarettes (20 cigarettes) per day and approximately 50% smoked ≥ 1 pack of cigarettes per day. Of all PanCan participants, approximately 80% reported smoking at least one pack per day, on average, throughout the smoking period of their lifetime. Throughout the smoking period of their lifetime, former smokers reported smoking an average of almost 4 more cigarettes per day than current smokers (mean=27.14 vs. mean=23.3; P<0.001). For all smoking exposure variables VAS scores tended to decrease as smoking exposure increased.

Significant differences in VAS scores were apparent between PanCan participants based on alcohol consumption. Mean VAS score tended to decrease as alcohol consumption increased, however, participants consuming 0 to 2 drinks per week reported slightly lower VAS scores than participants consuming 3 to 10 alcoholic drinks per week. Regarding marijuana use, VAS score decreased with increased exposure to marijuana, mean VAS score was highest among non-users.

Table 6. Descriptive statistics of VAS score by exposure variables

		EQ-5D VAS score dichotomized at median			EQ-5D VAS score	
Exposure variables	**N (%)**	**≤ 79 N (%)**	**≥ 80 N (%)**	**P***	**Mean (SD; IQR)**	**P***
Smoking Status	**2537**	**1085**	**1452**			
Former	958 (37.8)	351 (36.6)	607 (63.4)	<0.001	78.7 (13.8; 70 to 80)	<0.001
Current	1579 (62.2)	734 (46.5)	845 (53.5)		75.6 (14.4; 70 to 80)	
Cigarettes per day						
(30 days before baseline)	**2531**	**1083**	**1448**			
Former smokers†	958 (37.8)	351 (36.6)	607 (63.4)	<0.001	78.7 (13.8; 70 to 80)	<0.001
1st quartile: 0 to 9	330 (13.0)	124 (37.6)	206 (62.4)		79.6 (12.3; 70 to 80)	
2nd quartile: 10 to 19	463 (18.2)	199 (43.0)	264 (57.0)		76.4 (14.3; 70 to 80)	
3rd quartile: 20 to 24	308 (12.1)	134 (43.5)	174 (56.5)		76.1 (14.3; 70 to 80)	
4th quartile: >25	472 (18.6)	275 (58.3)	197 (41.7)		71.9 (15.2; 61 to 75)	
Cigarettes per day						
(during smoking period)	**2537**	**1085**	**1452**			
1st quartile: 1 to 19	529 (23.9)	183 (34.6)	346 (65.4)	<0.001	79.4 (12.7; 70 to 80)	<0.001
2nd quartile: 20 to 24	641 (25.3)	264 (41.2)	377 (58.8)		77.7 (13.5; 70 to 80)	
3rd quartile: 25	763 (30.1)	348 (45.6)	415 (54.4)		75.9 (14.4; 70 to 80)	
4th quartile: > 25	604 (23.8)	290 (48.0)	314 (52.0)		74.5 (15.7; 70 to 80)	
Smoking duration, years	**2537**	**1085**	**1452**			
1st quartile: ≤ 40	690 (27.2)	299 (43.3)	391 (56.7)	0.475	76.4 (15.0; 70 to 80)	0.971
2nd quartile: 41 to 44	699 (27.6)	307 (43.9)	392 (56.1)		76.6 (14.4; 70 to 80)	
3rd quartile: 45 to 48	588 (23.2)	235 (40.0	353 (60.0)		77.8 (13.6; 70 to 80)	
4th quartile: > 48	560 (22.1)	244 (43.6)	316 (56.4)		76.4 (13.9; 70 to 80)	
Time since quitting, years	**2537**	**1085**	**1452**			
Current smokers‡	1579 (62.2)	734 (46.5)	845 (53.5)	<0.001	75.6 (14.4; 70 to 80)	<0.001
1st quartile: 0 to 2	300 (11.8)	128 (42.7)	172 (57.3)		77.2 (14.6; 70 to 80)	
2nd quartile: 3 to 5	203 (8.0)	76 (37.4)	127 (62.6)		78.3 (14.0; 70 to 80)	
3rd quartile: 6 to 10	289 (11.4)	89 (30.8)	200 (69.2)		80.1 (12.6; 70 to 80)	
4th quartile: 11 to 22	166 (6.5)	58 (34.9)	108 (65.1)		79.1 (13.8; 70 to 80)	
Alcoholic drinks per week	**2537**	**1085**	**1452**			
1st quartile: 0 to 2	708 (27.9)	282 (39.8)	426 (60.2)	<0.001	77.6 (14.6; 70 to 80)	<0.001
2nd quartile: 3 to 10	641 (25.3)	246 (38.4)	395 (61.6)		78.3 (13.7; 70 to 80)	
3rd quartile: 11 to 20	558 (22.0)	242 (43.4)	316 (56.6)		76.4(13.4; 70 to 80)	
4th quartile: > 20	630 (24.8)	315 (50.0)	315 (50.0)		74.6 (15.0; 70 to 75)	
Marijuana use						
(joints per month)	**2508**	**1073**	**1435**			
Non-users§	1891 (75.4)	784 (41.5)	1107 (58.5)	0.034	77.5 (13.7; 70 to 80)	0.014
1st quartile: 1 to 4	210 (8.4)	86 (40.9)	124 (59.1)		76.9 (13.2; 60 to 80)	
2nd quartile: 5 to 9	100 (4.0)	52 (52.0)	48 (48.0)		74.1 (14.9; 70 to 80)	
3rd quartile: 10 to 20	159 (6.3)	80 (50.3)	79 (49.7)		73.2 (16.3; 65 to 75)	
4th quartile: >20	148 (5.9)	71 (48.0)	77 (52.0)		73.1 (18.6; 68 to 77)	

Abbreviations: IQR, interquartile range; N, sample size; P, P-value; SD, standard deviation.

* P-values calculated for proportional comparisons with chi-squared test and non-parametric test of trend for comparisons of VAS score.

† Former smokers were excluded from quartile calculation. Mean EQ-5D VAS scores of former smokers were compared to current smokers of varying intensity.

‡ Current smokers were excluded from quartile calculation. Mean EQ-5D VAS scores of current smokers were compared to time since quitting of varying intensity.

§ Non-users were excluded in quartile calculation. Mean EQ-5D VAS scores of non-users were compared to current marijuana users of varying intensity.

Univariate and adjusted MFP linear regression data are presented in *Table 7*. Unadjusted regression models of all exposure variables, except for smoking duration, showed significant associations with VAS score. Average weekly alcohol consumption was the only exposure variable with a significant non-linear relationship with VAS score in unadjusted analysis. Exposure variables remaining significant after partial adjustment included cigarettes per day within the 30 days prior to baseline as well as cigarettes per day averaged throughout a participant's smoking period and weekly alcohol consumption. Further adjustment for socio-demographic, morbidity and symptom variables explained away the effects of cigarettes per day averaged throughout a participant's smoking period. Cigarettes consumed in the 30 days prior to baseline and weekly alcohol consumption were the only significant exposure variables of VAS score after full adjustment. The final model, which included all relevant variables (R^2=0.284), explained 0.228 more variation in VAS score than the preliminary smoking and alcohol exposure model (R^2=0.056).

Table 7. Univariate and multivariable fractional polynomial linear regression models between smoking and alcohol exposure variables and EQ-5D VAS scores

Smoking/Alcohol Exposure	Univariate Model	Partially Adjusted Model* (N=2481, R^2=0.056)	Final Adjusted Model† (N=2397, R^2=0.284)
	β (CI; P-value)‡	β (CI; P-value)‡	β (CI; P-value)‡
Cigarettes per day (last 30 days)	-0.22 (-0.26 to -0.17; <0.001)	-0.16 (-0.21 to -0.11; <0.001)	-0.09 (-0.15 to -.004; 0.005)
Cigarettes per day (during smoking period)	-0.18 (-0.23 to -0.13; <0.001)	-0.15 (-0.19 to -0.10; <0.001)	(P=0.580)
Smoking duration (years)	0.0003 (-0.095 to 0.095; 0.995)	(P=0.41)	Excluded
Time since quitting (years)	0.39 (0.24 to 0.53; <0.001)	0.14 (0.08 to 0.42; 0.157)	(P=0.674)
Alcoholic drinks per week			
1st term	$-4.11e^{-6}$ ($-5.89e^{-6}$ to $-2.33e^{-6}$; <0.001)§	-0.05 (-0.07 to -0.02; 0.002)	-0.03 (-0.05 to -0.01; 0.01)
2nd term	-10.5^{4} (-13.58 to -7.43; <0.001)§		
Marijuana use (joints per month)	-1.31 (-1.81 to -0.79; <0.001)‖	-0.04 (-0.09 to 0.02; 0.190)	(P=0.465)

Abbreviations: β, beta-coefficient; CI, confidence interval; N, sample size;
***** Independent variables included in MFP linear regression modelling contain only smoking and alcohol exposures.
† MFP linear regression model contains socio-demographic, smoking/alcohol exposure, morbidities and symptoms which were significantly associated with VAS score in partially adjusted MFP models. Only smoking and alcohol exposure variables are displayed in this table.
‡ Beta-coefficients, confidence intervals and p-values calculated using multivariable fractional polynomial regression models.
§ Average weekly alcohol consumption has a non-linear relationship with VAS score, and the following two terms in the regression model describe this relationship;

$$Term\ 1 = \left[\left(\frac{drinks\ per\ week + 0.099}{100}\right)^{-2} - 39.1\right] * (-4.11e^{-6})$$

$$Term\ 2 = \left[\left(\frac{drinks\ per\ week + 0.099}{100}\right)^{\wedge.5} - 0.399\right] * (-10.50326)$$

‖ Average monthly marijuana use has a non-linear relationship with VAS score, and the following term in the regression model describes this relationship;

$$Term\ 1 = \left[\frac{\ln(marijuana\ use + 1)}{100}\right] + 3.079201368$$

4.2.3 Morbidities and quality of life (VAS scores)

Descriptive statistics outlining the relationship between pulmonary, cardiovascular and other morbidities, and VAS scores are available in Tables 8a and 8b. The most prevalent morbidities among the PanCan population were hypercholesterolemia (42%), arthritis (40%), hypertension (36%), allergies (31%) and COPD (21%). Mean number of morbidities tended to be higher among older participants and those with higher educational attainment. Significantly fewer morbidities were reported by males (mean=2.53 vs. mean=3.21; P<0.001) and employed participants (P<0.001) than their counterparts.

Mean number of morbidities increased among participants smoking more cigarettes per day in the 30 days prior to baseline (P<0.001) and those with greater number of cigarettes smoked per day throughout the entire smoking period of their lifetime (P=0.003). However, former smokers reported significantly more morbidities than current smokers (mean=3.03 vs. mean=2.71; P<0.001). A larger proportion of participants without morbidities reported VAS scores greater than the median value, 80, compared to those with morbidities. As the number of total morbidities increased, VAS scores progressively decreased. Participants with increased pulmonary function, FEV_1 percent predicted or FVC percent predicted, more commonly reported VAS scores greater than the median value in addition to reporting higher mean VAS scores.

Table 8a. Frequencies and descriptive statistics of VAS score by morbidities and pulmonary function variables

		EQ-5D VAS score dichotomized at median			EQ-5D VAS score	
Morbidity variables	**N (%)**	**≤ 79 N (%)**	**≥ 80 N (%)**	**P***	**Mean (SD; IQR)**	**P***
Cardiovascular morbidities						
Stroke	**2505**	**1080**	**1449**			
No	2459 (97.2)	1043 (42.4)	1416 (57.6)	0.082	76.9 (14.2; 70 to 80)	0.031
Yes	70 (2.8)	37 (52.9)	33 (47.1)		73.3 (14.4; 63 to 75)	
Angina	**2502**	**1078**	**1448**			
No	2358 (93.3)	977 (41.4)	1381 (58.6)	<0.001	77.2 (14.1; 70 to 80)	<0.001
Yes	168 (6.7)	101 (60.1)	67 (39.9)		71.6 (15.5; 60 to 75)	
Myocardial infarction	**2530**	**1080**	**1450**			
No	2370 (92.8)	992 (41.9)	1378 (58.1)	0.001	77.1 (14.0; 70 to 80)	0.001
Yes	160 (7.2)	88 (55.0)	72 (45.0)		72.2 (17.1; 65 to 75)	
Hypercholesterolemia	**2507**	**1080**	**1451**			
No	1465 (58.0)	583 (39.8)	882 (60.2)	0.001	77.9 (13.8; 70 to 80)	<0.001
Yes	1066 (42.0)	497 (46.6)	569 (53.4)		75.3 (14.7; 70 to 80)	
Hypertension	**2525**	**1075**	**1450**			
No	1614 (64.0)	629 (39.0)	985 (61.0)	<0.001	78.0 (13.8; 70 to 80)	<0.001
Yes	911 (36.0)	446 (49.0)	465 (51.0)		74.9 (14.7; 70 to 80)	
Congestive heart failure	**2528**	**1080**	**1448**			
No	2511 (99.3)	1074 (42.8)	1437 (57.2)	0.534	76.8 (14.2; 70 to 80)	0.229
Yes	17 (0.7)	6 (35.3)	11 (64.7)		79.6 (16.6; 70 to 85)	
Peripheral vascular disease	**2503**	**1078**	**1449**			
No	2422 (95.8)	1017 (42.0)	1405 (58.0)	0.001	77.1 (14.1; 70 to 80)	0.001
Yes	105 (4.2)	61 (58.1)	44 (41.9)		71.2 (17.2; 62.5 to 75)	
Pulmonary morbidities						
Asthma	**2503**	**1079**	**1448**			
No	2282 (90.4)	945 (41.4)	1337 (58.6)	<0.001	77.4 (13.8; 70 to 80)	<0.001
Yes	245 (9.6)	134 (54.7)	111 (45.3)		71.3 (17.1; 60 to 75)	
COPD	**2501**	**1077**	**1448**			
No	1993 (79.0)	896 (38.8)	1330 (61.2)	<0.001	78.1 (13.5; 70 to 80)	<0.001
Yes	532 (21.0)	181 (57.1)	118 (42.9)		72.2 (15.8; 65 to 75)	
Pulmonary fibrosis	**2532**	**1082**	**1450**			
No	2529 (99.9)	1080 (42.7)	1449 (57.3)	0.402	76.8 (14.2; 70 to 80)	0.161
Yes	3 (0.1)	2 (66.7)	1 (33.3)		56.7 (32.1; 20 to 70)	
Pulmonary function						
FEV_1 % predicted	**2517**	**1077**	**1440**			
1st quartile: 0.939 to 1.684	630 (25.0)	198 (31.5)	431 (68.5)	<0.001	80.1 (12.8; 60 to 75)	<0.001
2nd quartile: 0.825 to 0.938	629 (25.0)	233 (37.0)	396 (63.0)		78.7 (13.3; 70 to 80)	
3rd quartile: 0.703 to 0.824	629 (25.0)	279 (44.4)	350 (55.6)		76.6 (13.5; 70 to 80)	
4th quartile: 0.151 to 0.702	629 (25.0)	367 (58.2)	263 (41.8)		71.7 (16.1; 70 to 80)	
FVC percent predicted	**2517**	**1077**	**1440**			
1st quartile: 1.03 to 1.76	629 (25.0)	226 (35.9)	403 (64.1)	<0.001	78.6 (13.5; 65 to 75)	<0.001
2nd quartile: 0.919 to 1.028	629 (25.0)	230 (36.5)	400 (63.5)		78.5 (13.9; 70 to 80)	
3rd quartile: 0.811 to 0.919	630 (25.0)	281 (44.7)	348 (55.3)		77.3 (13.1; 70 to 80)	
4th quartile: 0.346 to 0.811	629 (25.0)	340 (54.0)	289 (46.0)		72.8 (15.8; 70 to 80)	

Abbreviations: COPD, chronic obstructive pulmonary disease; FEV_1, forced expiratory volume in one second; FVC, forced vital capacity; IQR, interquartile range; N, sample size; P, P-value; SD, standard deviation.
* P-values calculated using chi-square tests for categorical comparisons and non-parametric test of trend for comparisons involving continuous data.

Table 8b. Frequencies and descriptive statistics of VAS score by morbidities

Morbidity variables		EQ-5D VAS score dichotomized at median			EQ-5D VAS score	
	N (%)	≤ 79 N (%)	≥ 80 N (%)	P*	Mean (SD; IQR)	P*
Other morbidities						
Liver disease	**2529**	**1081**	**1448**			
No	2383 (94.3)	1008 (42.3)	1375 (57.7)	0.068	77.0 (14.1; 70 to 80)	0.036
Yes	146 (5.7)	73 (50.0)	73 (50.0)		74.0 (15.9; 70 to 75)	
Kidney disease	**2522**	**1078**	**1444**			
No	2479 (98.3)	1060 (42.8)	1419 (57.2)	0.906	76.8 (14.2; 70 to 80)	0.631
Yes	43 (1.7)	18 (41.9)	25 (58.1)		75.1 (15.5; 70 to 80)	
Osteoporosis	**2530**	**1082**	**1448**			
No	2135 (84.3)	910 (42.6)	1225 (57.4)	0.734	76.9 (14.2; 70 to 80)	0.313
Yes	395 (15.7)	172 (43.5)	223 (56.5)		76.1 (14.5; 70 to 80)	
Arthritis	**2529**	**1082**	**1447**			
No	1506 (59.6)	510 (38.0)	934 (62.0)	<0.001	78.3 (13.5; 70 to 80)	<0.001
Yes	1023 (40.4)	572 (49.8)	513 (50.2)		74.6 (15.1; 70 to 77)	
Allergy	**2530**	**1081**	**1449**			
No	1745 (69.0)	716 (41.0)	1029 (59.0)	0.010	77.3 (14.1; 70 to 80)	0.010
Yes	785 (31.0)	365 (53.5)	420 (46.5)		75.7 (14.5; 70 to 80)	
Diabetes	**2529**	**1078**	**1451**			
No	2301 (91.0)	949 (41.2)	1352 (58.8)	<0.001	77.2 (14.1; 70 to 80)	<0.001
Yes	228 (9.0)	129 (56.6)	99 (43.4)		72.6 (14.8; 65 to 75)	
Anemia	**2531**	**1080**	**1451**			
No	2319 (91.6)	977 (42.1)	1342 (57.9)	0.069	77.0 (14.2; 70 to 80)	0.008
Yes	212 (8.4)	103 (48.6)	109 (51.4)		74.4 (14.8; 66 to 80)	
Morbidities total	**2533**	**1082**	**1451**			
0	274 (11.0)	84 (31.3)	190 (68.7)	<0.001	81.1 (13.4; 75 to 80)	<0.001
1	413 (16.3)	135 (32.7)	278 (67.3)		79.7 (12.2;75 to 80)	
2	559 (22.0)	206 (36.9)	353 (63.2)		78.8 (12.6; 70 to 80)	
3	467 (18.4)	207 (44.3)	260 (55.7)		76.4 (13.5; 70 to 80)	
4	327 (12.9)	153 (46.8)	174 (53.2)		75.8 (14.4; 70 to 80)	
5	240 (9.5)	130 (54.2)	110 (45.8)		72.8 (15.5; 65 to 75)	
6	120 (4.7)	75 (62.5)	45 (37.5)		70.6 (17.4; 60 to 71)	
≥7	133 (5.2)	92 (69.2)	41 (30.8)		66.9 (16.7; 59 to 70)	

Abbreviations: IQR, interquartile range; N, sample size; P, P-value; SD, standard deviation.
* P-values calculated using chi-square tests for categorical comparisons and non-parametric test of trend for comparisons involving continuous data.

Statistics describing the univariate and adjusted relationships between cardiovascular, pulmonary, and other morbidities, and VAS scores from linear regression modelling are available in *Table 9*. All morbidities were significantly inversely associated with VAS score in univariate analysis except for osteoporosis and kidney disease. Many variables were no longer significant after adjusting for other morbidities. In the final model which adjusts for all socio-demographic, exposure, morbidity and symptom factors, high cholesterol (P=0.015), myocardial infarction (P=0.015), peripheral vascular

disease (P=0.018), arthritis (P=0.036) and low FEV_1 percent predicted (P=0.006) were significantly associated with lower VAS scores. In univariate and preliminary morbidity-adjusted regression models, FEV_1 percent predicted, had a significant linear relationship with VAS score. However, in the fully adjusted model, a significant non-linear relationship became apparent (P=0.027) (Figure 9). The partially adjusted morbidity regression model explains 0.106 of the variation in VAS score, 0.174 less than the fully adjusted regression model.

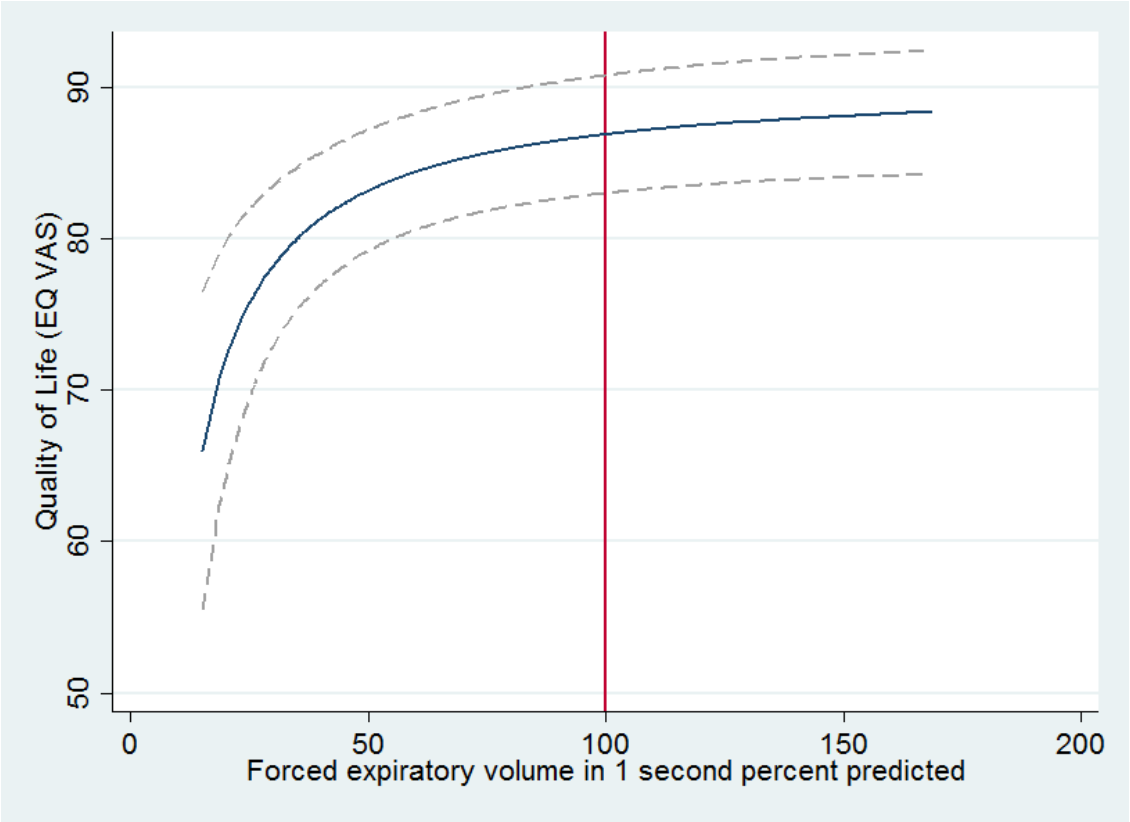

Figure 9. Adjusted non-linear relationship between forced expiratory volume in one second percent predicted and EQ-5D VAS score

Table 9. Univariate and linear regression models between morbidities variables and EQ-5D VAS scores

	Univariate Model	Preliminary Model* (N=2432, R^2=0.106)	Final Adjusted Model† (N=2397, Adj. R^2=0.284)
Morbidities	**β (CI; P-value)‡**	**β (CI; P-value)‡**	**β (CI; P-value)‡**
Cardiovascular			
Stroke	-3.63 (-7.02 to -0.25; 0.035)	(P=0.815)	Excluded
CHD	-5.12 (-7.24 to -2.99; <0.001)	(P=0.457)	Excluded
Angina	-5.61 (-7.84 to -3.38; <0.001)	-2.49 (-5.76 to 0.77; 0.114)	(P=0.500)
Myocardial infarction	-4.90 (-7.2 to -2.61; <0.001)	-3.37 (-6.14 to -0.60; 0.024)	-2.64 (-4.60 to -0.68; 0.015)
Hypercholesterolemia	-2.58 (-3.71 to -1.46; <0.001)	-1.32 (-2.22 to -0.42; 0.010)	-1.15 (-1.99 to -0.30; 0.015)
Hypertension	-3.11 (-4.26 to -1.95; <0.001)	(P=0.237)	Excluded
CHF	2.86 (-3.94 to 9.66; 0.410)	5.24 (-2.76 to 13.25; 0.165)	(P=0.216)
PVD	-5.87 (-8.66 to -3.08; <0.001)	-3.85 (-6.07 to -1.63; 0.005)	-2.61 (-4.63 to -0.59; 0.018)
Pulmonary			
Asthma	-6.15 (-8.03 to -4.27; <0.001)	-2.97 (-6.57 to 0.62; 0.091)	(P=0.182)
COPD	-5.91 (-7.26 to -4.56; <0.001)	-3.16 (-5.30 to -1.02; 0.010)	(P=0.635)
Pulmonary fibrosis	-20.16 (-36.29 to -4.02; 0.014)	(P=0.243)	Excluded
Pneumonia	-3.62 (-4.87 to -2.36; <0.001)	-1.68 (-3.09 to -0.26; 0.026)	Excluded
FEV_1 % predicted§	16.79 (13.83 to 19.75; <0.001)	12.2 (6.89 to 17.41; 0.001)	-3.70 (-5.95 to -1.45; 0.006)
Other			
Liver disease	-3.01 (-5.4 to -0.61; 0.014)	(P=0.135)	(P=0.933)
Kidney disease	-1.75 (-6.10 to 2.6; 0.430)	(P=0.780)	Excluded
Osteoporosis	-0.86 (-0.24 to 0.68; 0.274)	(P=0.491)	Excluded
Arthritis	-3.72 (-4.85 to -2.59; <0.001)	-3.01 (-4.55 to -1.47; 0.002)	-1.78 (-3.40 to -0.15; 0.036)
Allergy	-1.53 (-2.73 to -0.32; 0.013)	(P=0.593)	Excluded
Diabetes	-4.68 (-6.62 to -2.73; <0.001)	(P=0.052)	(P=0.279)
Anemia	-2.62 (-4.62 to -0.61; 0.011)	(P=0.283)	Excluded
Blood disease	-4.36 (-8.40, -0.34; 0.033)	(P=0.131)	(P=0.192)

Abbreviations: CHD, coronary heart disease; CHF; congestive heart failure; COPD; chronic obstructive pulmonary disease; FEV_1 % predicted, forced expiratory volume in one second percent predicted; N, sample size; PVD, peripheral vascular disease.

***** Variables eligible for inclusion in preliminary linear regression models are limited to the pulmonary, cardiovascular and 'other' morbidities variables listed in Table 1 as well as FEV_1 percent predicted.

† Linear regression model contains socio-demographic, smoking/alcohol exposures, morbidities and symptoms which were significantly associated with VAS score in partially adjusted MFP models.

‡ Beta-coefficients, confidence intervals and p-values calculated using MFP linear regression models.

§ FEV_1, percent predicted, has a non-linear relationship with VAS score and is described by the following term; $Term\ 1 = (FEV^{-1} - 1.25) * (-3.698979)$

4.2.4 Symptoms

Statistics describing VAS score by symptom are presented in Table 10 and regression associations are described in Table 11. The most prevalent symptoms were coughing (52.4%), phlegm (46.2%), dyspnea (45.3%), fatigue (39.2%) and wheezing (37.6%). Older participants reported significantly fewer symptoms (P<0.001); those aged 50 to 55 (mean=3.69) reported approximately one more symptom than those ≥70 years (mean=2.52). Participants with lower education or those who are not retired also tended to report greater symptoms. Former smokers experienced significantly fewer symptoms than current smokers (mean=2.15 vs. mean=3.21; P<0.001). Number of symptoms reported also significantly increased (P<0.001) as participants reported greater number of cigarettes smoked in the 30 days prior to baseline, and greater number of cigarettes smoked per day, throughout the smoking period of participant's lifetime.

Participants experiencing any symptom were more likely to report VAS scores below the median VAS value, 80, compared to those who were symptom free. Furthermore, mean VAS score of symptom-affected participants was significantly lower for all symptoms except hemoptysis which was also associated with lower VAS score but because of small numbers only approached significance. Mean VAS score significantly decreased as the number of reported symptoms increased.

All symptoms were significantly, inversely associated with VAS score in univariate analysis. After adjusting for all symptoms, dyspnea (P<0.001), phlegm (P=0.001), chest pain (P<0.007), having a poor appetite (P=0.017) and feeling fatigued (P<0.0001) remained significant. In comparison to all other partially adjusted regression models, symptoms explained the most variation in VAS score. The final model, which included all relevant variables (R^2=0.284), explained 0.094 more variation in VAS score than the preliminary symptoms model (R^2=0.190).

Table 10. Descriptive statistics of VAS score by symptoms variables

		EQ-5D VAS score dichotomized at median			EQ-5D VAS scores	
Symptom	**N (%)**	**≤ 79 N (%)**	**≥ 80 N (%)**	**P***	**Mean (SD; IQR)**	**P***
Dyspnea	**2531**	**1082**	**1449**			
No	1384 (54.7)	409 (29.5)	975 (70.5)	<0.001	80.1 (11.8; 75 to 80)	<0.001
Yes	1147 (45.3)	673 (58.7)	474 (41.3)		71.8 (15.3; 65 to 75)	
Cough	**2531**	**1082**	**1449**			
No	1204 (47.6)	428 (35.6)	776 (64.5)	<0.001	79.2 (12.9; 70 to 80)	<0.001
Yes	1327 (52.4)	654 (49.3)	673 (50.7)		74.6 (15.0; 70 to 80)	
Phlegm	**2530**	**1082**	**1448**			
No	1360 (53.8)	479 (35.2)	881 (64.8)	<0.001	79.2 (13.0; 70 to 80)	<0.001
Yes	1170 (46.2)	603 (51.5)	567 (48.5)		74.0 (15.1; 68 to 75)	
Hemoptysis	**2529**	**1082**	**1447**			
No	2485 (98.2)	1056 (32.5)	1429 (57.5)	0.027	76.9 (14.1; 70 to 80)	0.135
Yes	44 (1.8)	26 (59.1)	18 (40.9)		72.0 (20.2; 60.5 to 75)	
Hoarseness	**2530**	**1082**	**1448**			
No	1878 (74.3)	728 (38.8)	1150 (61.2)	<0.001	78.1 (13.5; 70 to 80)	<0.001
Yes	652 (25.7)	354 (54.3)	298 (45.7)		73.1 (15.6; 65 to 75)	
Wheeze	**2530**	**1082**	**1448**			
No	1577 (62.4)	543 (34.4)	1034 (65.6)	<0.001	79.5 (12.9; 70 to 80)	<0.001
Yes	953 (37.6)	539 (56.6)	414 (43.4)		72.3 (15.2; 65 to 75)	
Chest pain	**2524**	**1078**	**1446**			
No	2088 (82.8)	823 (39.4)	1265 (60.6)	<0.001	78.0 (13.5; 70 to 80)	<0.001
Yes	436 (17.2)	255 (58.5)	181 (41.5)		70.8 (16.2; 60 to 75)	
Poor appetite	**2529**	**1081**	**1448**			
No	2338 (92.5)	950 (40.6)	1388 (59.4)	<0.001	77.6 (13.9; 70 to 80)	<0.001
Yes	191 (7.5)	131 (68.6)	60 (31.4)		67.5 (15.7; 60 to 70)	
Fatigue	**2530**	**1082**	**1448**			
No	1536 (60.8)	492 (32.0)	1044 (68.0)	<0.001	80.7 (11.6; 75 to 80)	<0.001
Yes	994 (39.2)	590 (59.4)	404 (40.6)		70.7 (15.8; 60 to 70)	
Weight loss	**2527**	**1080**	**1447**			
No	2311 (91.4)	966 (41.8)	1345 (58.2)	0.002	77.1 (14.1; 70 to 80	0.001
Yes	216 (8.6)	114 (52.8)	102 (47.2)		73.6 (15.0; 66 to 75)	
Symptoms total	**2532**	**1450**	**1082**			
0	440 (17.6)	87 (19.8)	353 (80.2)	<0.001	83.9 (10.0; 80 to 85)	<0.001
1	377 (14.8)	122 (32.4)	255 (67.6)		80.6 (12.2; 70.5 to 80)	
2	408 (16.1)	135 (33.1)	273 (66.9)		79.7 (12.4;75 to 80)	
3	383 (15.1)	170 (44.4)	213 (55.6)		76.6 (13.1; 70 to 80)	
4	335 (13.2)	164 (49.0)	171 (51.0)		74.9 (12.7; 70 to 80)	
5	274 (10.8)	180 (65.7)	94 (34.3)		70.6 (15.0; 60 to 70)	
≥ 6	315 (12.4)	224 (71.1)	91 (28.9)		66.2 (17.1; 60 to 70)	

Abbreviations: IQR, interquartile range; N, sample size; P, P-value; SD, standard deviation.
* P-values calculated using chi-square test for categorical comparisons and non-parametric test of trend for continuous variable comparisons.

Table 11. Univariate and multivariate linear regression models between symptoms variables and EQ-5D VAS scores

	Univariate Model	Partially Adjusted Model* (N= 2497, R^2=0.190)	Final Adjusted Model† (N= 2397, R^2=0.284)
Symptoms	**β (CI; P-value)‡**	**β (CI; P-value)‡**	**β (CI; P-value)‡**
Dyspnea	-9.22 (-10.3 to -8.16; <0.001)	-4.77 (-6.42 to -3.12; <0.001)	-3.30 (-5.68 to -0.92; 0.013)
Cough	-4.59 (-5.7 to -3.49; <0.001)	(P=0.719)	Excluded
Phlegm	-5.23 (-6.33 to -4.13; <0.001)	-1.56 (-2.28 to -0.84; 0.001)	-1.35 (-2.18 to -0.52; 0.006)
Hemoptysis	-4.87 (-9.12 to -0.63; 0.025)	(P=0.811)	Excluded
Hoarseness	-5.02 (-6.28 to -3.75; <0.001)	-1.15 (-2.83 to 0.53; 0.150)	(P=0.053)
Wheeze	-7.13 (-8.25 to -6.02; <0.001)	-2.26 (-3.87 to -0.65; 0.014)	(P=0.096)
Chest pain	-7.21 (-8.67 to -5.76; <0.001)	-2.38 (-3.79 to -0.97; 0.005)	-2.59 (-3.70 to -1.49; 0.001)
Poor appetite	-10.06 (-12.1 to -7.99; <0.001)	-4.32 (-7.63 to -1.02; 0.018)	-4.14 (-7.31 to -0.96; 0.018)
Fatigued	-10.03 (-11.1 to -8.96; <0.001)	-6.43 (-7.69 to -5.17; <0.001)	-5.84 (-7.04 to -4.64; <0.001)
Weight loss	-3.53 (-5.51 to -1.54; 0.001)	(P=0.632)	Excluded

Abbreviations: β, beta-coefficient; CI, confidence interval; N, Sample Size.
* Linear regression model contains only symptoms variables.
† Linear regression model contains socio-demographic, smoking and alcohol exposure, morbidities and symptoms which were significantly associated with VAS score in partially adjusted MFP models. Only symptoms variables are displayed in this table.
‡ Beta-coefficients, confidence intervals and p-values calculated using linear regression models.

4.2.5 Summary of final linear regression model

All variables significant in the fully adjusted model are presented in Figure 10 and Table 12. In total, 15 socio-demographic, medical history, exposure, morbidity and symptom variables were associated with VAS score in the fully adjusted model. Upstream and downstream associations between all variables and VAS score are previously described. Several variables exhibited strong associations with VAS score, particularly employment status (unemployed, disabled or on sick leave) and symptoms variables dyspnea, poor appetite or fatigue. Non-linear relationships were observed between BMI and VAS score as well as FEV1 percent predicted and VAS score. The final regression model explained 0.284 of the variation in VAS scores.

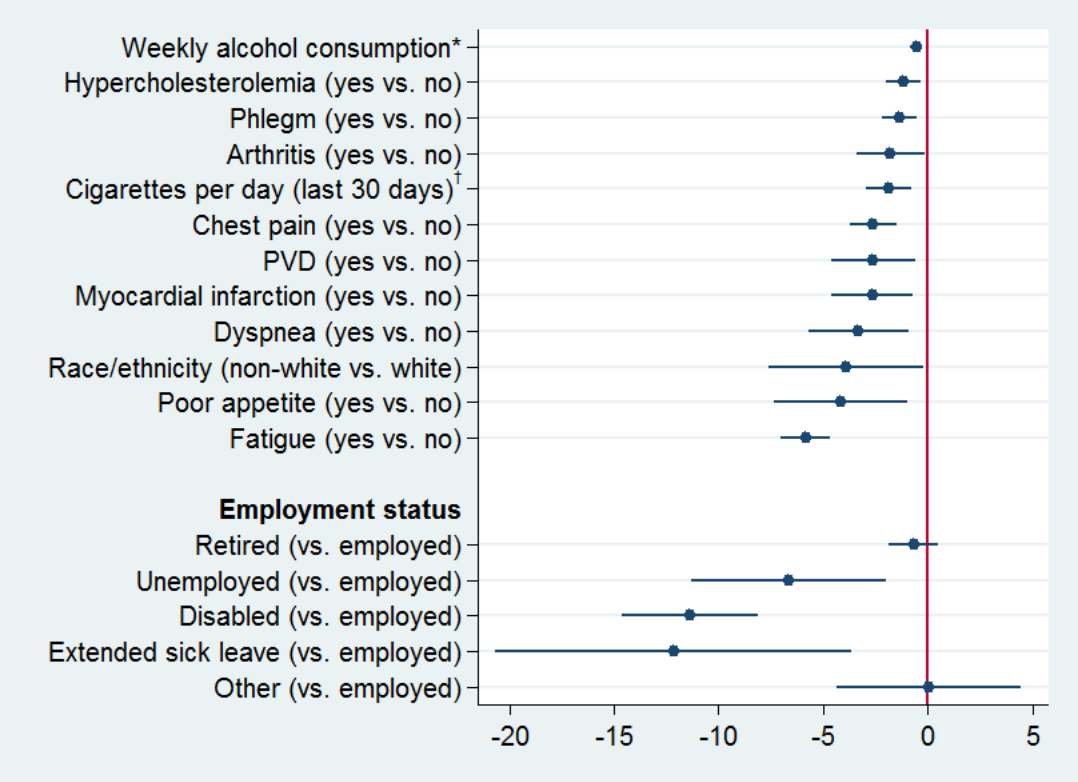

Figure 10. Adjusted effect estimates of variables significantly linearly associated with VAS score

***** To ease interpretation, estimate reflects the association between drinking 16 alcoholic drinks per week, the PanCan mean, and VAS score. **†** To ease interpretation, estimate reflects the association between smoking 20 cigarettes (one pack) per day in the 30 days prior to baseline.

Table 12. Variables significantly associated with VAS score in the fully adjusted MFP linear regression model, N=2397, R^2=0.284

Significant independent variables	Beta-coefficient (confidence interval; P-value)*
Socio-demographics and medical history	
Race/ethnicity	
non-white vs. white	-3.88 (-7.59 to -0.18; 0.042)
Employment status	
Employed	Referent
Retired	-0.67 (-1.85 to 0.51; 0.220)
Unemployed	-6.66 (-11.30 to -2.02; 0.012)
Disabled	-11.37 (-14.59 to -8.14; <0.001)
Sick leave	-12.15 (-20.68 to -3.62; 0.012)
Other	-0.03 (-4.37 to 4.45; 0.985)
Body mass index	
1st term†	-59.27 (-147.79 to 29.24; 0.157)
2nd term†	212.85 (12.41 to 413.28; 0.040)
Smoking and alcohol exposures	
Cigarettes per day (last 30 days)	-0.09 (-0.15 to -0.04; 0.005)
Alcoholic drinks per week	-0.04 (-0.05 to -0.01; 0.004)
Morbidities	
Myocardial infarction	-2.64 (-4.60 to -0.68; 0.015)
Hypercholesterolemia	-1.15 (-1.99 to -0.30; 0.015)
Peripheral vascular disease	-2.61 (-4.63 to -0.59; 0.018)
Arthritis	-1.78 (-3.40 to -0.15; 0.036)
FEV_1 percent predicted‡	-0.04 (-0.06 to -0.01; 0.006)
Symptoms	
Dyspnea	-3.30 (-5.68 to -0.92; 0.013)
Phlegm	-1.35 (-2.18 to -0.52; 0.006)
Chest pain	-2.59 (-3.70 to -1.49; 0.001)
Poor appetite	-4.14 (-7.31 to -0.96; 0.018)
Fatigued	-5.84 (-7.04 to -4.64; <0.001)

Abbreviations: FEV_1, forced expiratory volume in one second; N, sample size;.
* Beta-coefficients, confidence intervals and P-values were calculated using linear regression with a MFP selection procedure.
† BMI has a non-linear relationship with VAS score, and the following two terms in the regression model describe this relationship;

$$Term\ 1 = \left[\left(\frac{BMI}{10}\right)^{-2} - 0.1371742112\right] * \text{-}59.27362)$$

$$Term\ 2 = \left[\left(\frac{BMI}{10}\right)^{-2} * \ln\left(\frac{BMI}{10}\right) - 0.1362485285\right] * (212.845)$$

‡ FEV_1, percent predicted has a non-linear relationship with VAS score, and the following term in the regression model describes this relationship;

$$Term\ 1 = [(FEV_1\ percent\ predicted)^{-1} - 1.25] * (\text{-}3.698979)$$

4.4 Interaction

A priori interactions between gender and smoking exposures (smoking duration, daily number of cigarettes consumed over the smoking lifetime and within the previous

30 days) as well as smoking exposure and morbidity subgroups were insignificant. Interactions between having diabetes and smoking exposures were also insignificant.

4.5 Bootstrapping

To assess stability of the final regression model, 1000 bootstrap replications were prepared. Although all variables were included in bootstrap analysis, bootstrap inclusion fractions (BIF) only for those variables significantly associated with VAS score in the final linear regression model or included in greater than 50% of bootstrap replications are presented in Table 13. BIF, mean effect estimates and mean exponents for all variables assessed in regression analysis are available in Appendix 1a and Appendix 1b. Of all variables assessed, 14 have BIF greater than 50%. Pulmonary fibrosis was not significant in the full regression model but did yield a high BIF, 58.2%. Alternatively, hypercholesterolemia (BIF=46.8%) and peripheral vascular disease (BIF=46.3%), which were significantly associated with VAS score in the full regression model, were included in less than 50% of bootstrap replications. Several significant variables in the full regression model produced BIF greater than 90%: BMI (99.9%), dyspnea (100%), chest pain (91.6%), poor appetite (96.6%), fatigue (100%), arthritis (92%) and FEV_1 percent predicted (99.5%).

A non-linear association between BMI and VAS score was present in 62.3% of all bootstrap replications (Appendix 2). Linear relationships most commonly described the relationships between cigarettes smoked per day in the last 30 days, weekly number of alcoholic drinks, and VAS score (Appendix 3 and Appendix 4). FEV_1 percent predicted was a significant predictor of VAS score in 99.5% of bootstrap replications. Non-linear functions (BIF=63.4%) more commonly described the relationship between FEV_1 percent predicted and VAS score (Appendix 5).

Table 13. Bootstrap inclusion fractions for variables significantly associated with VAS score in final regression model or with bootstrap inclusion fractions > 50% excluded from final model, 1000 replications, N=2397

Variable	Bootstrap Inclusion Fraction (%)
Socio-demographics and medical history	
Gender	77.7
Race/ethnicity	60.2
BMI	99.9
Current work status	100.0
Smoking exposure	
Cigarettes per day (last 30 days)	73.6
Average alcoholic drinks per week	65.4
Morbidities	
Myocardial infarction	51.7
Hypercholesterolemia†	46.8
Peripheral vascular disease†	46.3
Arthritis	92.0
Pulmonary fibrosis*	58.2
FEV_1 % predicted	99.5
Symptoms	
Dyspnea	100
Phlegm	55.6
Chest pain	91.6
Poor appetite	96.6
Fatigued	100.0

Abbreviations: BMI, body mass index; FEV_1, forced expiratory volume in one second
***** Pulmonary fibrosis was insignificant in the fully adjusted linear regression model but was included in > 50% of bootstrap replications, indicating that it may be an important variable.
† Hypercholesterolemia and peripheral vascular disease were both significantly associated with VAS score but included in < 50% of bootstrap replications

4.6 EQ-5D Quality of life dimensions

The proportion of PanCan participants reporting no problems, some problems or severe problems for the EQ-5D (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) are available in Table 14. While participants commonly reported experiencing some problems in any EQ-5D dimension, they were far less likely to report severe problems, typical of population based data sets.

Table 14. Participant distributions of EQ-5D quality of life dimensions

EQ-5D response	Mobility (N=2526)	Self-care (N=2527)	Usual activities (N=2525)	Pain and discomfort (N=2526)	Anxiety and depression (N=2526)
No problems	1941 (77%)	2461 (97%)	1973 (78%)	1211 (48%)	1716 (68%)
Some Problems	581 (23%)	58 (2%)	530 (21%)	1248 (49%)	768 (30%)
Severe problems	4 (0.2%)	8 (0.3%)	22 (0.9%)	67 (3%)	42 (2%)

4.6.1 Mobility

Almost 25% of the sample reported experiencing some or severe mobility problems (Table 14). Univariate and multivariate logistic regression models are presented in Tables 15a, 15b and 15c. Age, education, employment status and BMI were significantly associated with mobility in univariate and partially adjusted models. The effects of age and education were explained away when exposures, morbidities and symptoms were added to the regression model. However, being unemployed ($OR_{adjusted}$=4.52; CI: 1.31 to 15.63; P=0.017), disabled ($OR_{adjusted}$= 13.76; CI: 7.37 to 25.69; P<0.001) or on sick leave ($OR_{adjusted}$=36.03; CI: 18.28 to 71.01; P<0.001) and BMI remained significant predictors of mobility ($OR_{adjusted}$= 1.10; CI: 1.09 to 1.12; P<0.001). Multiple exposure variables were significantly associated with mobility in univariate and partially adjusted regression models. Only smoking duration remained significant when adjusting for all other factors. Increased smoking exposure to cigarettes and marijuana was generally associated with increased odds of having some or severe mobility problems. Similar to other models, several morbidities and symptoms were univariately associated with mobility problems. A total of 4 morbidities, congestive heart failure ($OR_{adjusted}$=3.41; CI: 1.36 to 8.53; P=0.009), hypertension ($OR_{adjusted}$=1.33; CI: 1.04 to 1.69; P=0.024), peripheral vascular disease ($OR_{adjusted}$=1.74; CI: 1.16 to 2.61; P=0.007) and arthritis ($OR_{adjusted}$=2.68, CI: 2.14 to 3.35; P<0.001) were significantly associated with experiencing mobility problems. Significant symptoms variables in the final model include dyspnea ($OR_{adjusted}$=1.74; CI: 1.56 to 1.94; P<0.001), poor appetite ($OR_{adjusted}$=1.65; CI: 1.01 to 2.69; P=0.046) and fatigue ($OR_{adjusted}$=1.68; CI: 1.36 to 2.09; P<0.001). Total variation in mobility problems explained by the model was 0.178. The Hosmer–Lemeshow goodness of fit test suggests suitable fit, P=0.294.

Table 15a. Univariate and adjusted logistic regression odds ratios of variables significantly associated with reporting some or severe problems with mobility

Variable group	Univariate Odds ratio (CI; P-value)‡	Partially adjusted* Odds ratio (CI; P-value)‡	Fully Adjusted (N=2406, Pseudo R^2=0.178)† Odds ratio (CI; P-value)‡
Socio-demographics and medical history		**N=2461, Pseudo R^2=0.106**	
Age (years)	1.02 (1.01 to 1.04; 0.006)	1.04 (1.02 to 1.06; <0.001)	(P=0.743)
Education level	0.87 (0.82 to 0.92; <0.001)	0.89 (0.83 to 0.95; <0.001)	(P=0.216)
Employment status			
Employed	Referent	Referent	Referent
Retired	1.52 (1.23 to 1.88; <0.001)	1.22 (0.97 to 1.53; 0.091)	1.29 (1.02 to 1.64; 0.033)
Unemployed	3.15 (1.91 to 5.22; <0.001)	3.20 (1.91 to 5.37; <0.001)	2.90 (1.82 to 4.63; <0.001)
Disabled	11.20 (6.90 to 18.16; <0.001)	12.26 (6.78 to 22.18; <0.001)	8.13 (4.42 to 14.95; <0.001)
Sick leave	7.93 (2.56 to 24.55; <0.001)	6.89 (1.20 to 39.57; 0.030)	6.71 (1.22 to 36.81; 0.029)
Other	1.10 (0.62 to 1.94; 0.739)	1.04 (0.45 to 2.39; 0.934)	1.17 (0.51 to 2.65; 0.715)
BMI			
1st term	$2.1e^{7}$ (3729.29 to $1.21e^{9}$)	$1.2e^{5}$ (48.11 to $3.42e^{8}$)	1.10 (1.09 to 1.12; <0.001)
2nd term	$4.24e^{-25}$ ($1.5e^{-31}$ to $1.19e^{-18}$)	$4.47e^{-23}$ ($1.17e^{-30}$ to $1.7e^{-15}$)	(P=0.111)
Exposures		**N=2493, Pseudo R^2=0.024**	
Cigarettes per day (last 30 days)	1.01 (1.01 to 1.02; <0.001)	1.02 (1.01 to 1.02; 0.001)	(P=0.585)
Cigarettes per day (smoking period)	1.02 (1.01 to 1.02; <0.001)	1.02 (1.00 to 1.03; 0.005)	(P=0.489)
Smoking duration (years)	1.04 (1.02 to 1.06; <0.001)	1.05 (1.03 to 1.08; <0.001)	1.04 (1.02 to 1.06; <0.001)
Time since quitting (years)	0.99 (0.97 to 1.01; 0.399)	1.04 (1.01 to 1.08; 0.021)	(P=0.351)
Marijuana use (joints per month)	1.01 (1.00 to 1.02; 0.008)	1.01 (1.00 to 1.03; 0.009)	(P=0.375)

Abbreviations: BMI, body mass index; CI, confidence interval; N, sample size.

***** Variables eligible for inclusion in preliminary linear regression models are limited to socio-demographics, medical history and exposure factors.

† Logistic regression model contains socio-demographic, smoking/alcohol exposures, morbidities and symptoms which were significantly associated with anxiety/depression in partially adjusted logistic regression models.

‡ Logistic regression models used to calculate odds ratios, confidence intervals and P-values.

Table 15b. Univariate and adjusted logistic regression odds ratios of variables significantly associated with reporting some or severe problems with mobility

	Univariate	Partially adjusted*	Fully Adjusted (N=2406, Pseudo R^2=0.178)†
Variable group	**Odds ratio (CI; P-value)‡**	**Odds ratio (CI; P-value)‡**	**Odds ratio (CI; P-value)‡**
Morbidities		**N=2439, Pseudo R^2=0.098**	
Stroke	2.15 (1.31 to 3.51; 0.002)	1.52 (0.99 to 2.34; 0.058)	(P=0.108)
Coronary heart disease	2.17 (1.59 to 2.97; <0.001)	1.32 (1.03 to 1.70; 0.027)	(P=0.452)
Angina	2.52 (1.82 to 3.48; <0.001)	1.83 (1.14 to 2.94; 0.013)	(P=0.051)
Myocardial infarction	1.92 (1.37 to 2.70; <0.001)	(P=0.498)	Excluded
Hypercholesterolemia	1.33 (1.10 to 1.60; 0.003)	(P=0.637)	Excluded
Hypertension	1.86 (1.54 to 2.24; <0.001)	1.37 (1.13 to 1.68; 0.002)	1.33 (1.04 to 1.69; 0.024)
Congestive heart failure	2.99 (1.15 to 7.78; 0.025)	2.66 (1.04 to 6.82; 0.041)	3.41 (1.36 to 8.53; 0.009)
Peripheral vascular disease	2.25 (1.50 to 3.37; <0.001)	1.58 (1.06 to 2.36; 0.026)	1.74 (1.16 to 2.61; 0.007)
Asthma	1.46 (1.09 to 1.95; 0.011)	(P=0.992)	Excluded
Pulmonary fibrosis	1.90 (-0.50 to 4.30; 0.122)	8.18 (1.06 to 63.34; 0.044)	(P=0.095)
Personal cancer history	0.41 (0.06 to 0.76; 0.022)	1.34 (0.91 to 1.97; 0.136)	(P=0.107)
Liver disease	1.53 (1.07 to 2.21; 0.021)	1.55 (1.08 to 2.22; 0.016)	(P=0.327)
Osteoporosis	1.53 (1.21 to 1.95; <0.001)	(P=0.270)	Excluded
Diabetes	1.86 (1.39 to 2.49; <0.001)	1.51 (1.24 to 1.83; <0.001)	(P=0.655)
Anemia	1.46 (1.07 to 1.99; 0.017)	(P=0.446)	Excluded
Asthma	1.46 (1.09 to 1.95; 0.011)	(P=0.992)	Excluded
COPD	1.65 (1.34 to 2.05; <0.001)	(P=0.252)	Excluded
Pneumonia	1.61 (1.31 to 1.96; <0.001)	1.30 (1.03 to 1.64; 0.030)	(P=0.084)
Arthritis	3.07 (2.54 to 3.72; <0.001)	3.04 (2.34 to 3.96; <0.001)	2.68 (2.14 to 3.35; <0.001)
Allergy	1.36 (1.16 to 1.65; 0.002)	(P=0.203)	Excluded
FEV_1 percent predicted	0.17 (0.10 to 0.28; <0.001)	0.24 (0.14 to 0.43; <0.001)	(P=0.054)

Abbreviations: CI, confidence interval; COPD, chronic obstructive pulmonary disease; FEV_1, forced expiratory volume in one second; N, sample size.

***** Variables eligible for inclusion in preliminary linear regression models are limited to morbidities and symptoms.

† Logistic regression model contains socio-demographic, smoking/alcohol exposures, morbidities and symptoms which were significantly associated with anxiety/depression in partially adjusted logistic regression models.

‡ Logistic regression models used to calculate odds ratios, confidence intervals and P-values.

Table 15c. Univariate and adjusted logistic regression odds ratios of variables significantly associated with reporting some or severe problems with mobility

	Univariate	Partially adjusted*	Fully Adjusted (N=2406, Pseudo R^2=0.178)†
Variable group	**Odds ratio (CI; P-value)‡**	**Odds ratio (CI; P-value)‡**	**Odds ratio (CI; P-value)‡**
Symptoms		**N=2507, Pseudo R^2=0.063**	
Dyspnea	2.76 (2.27 to 3.34; <0.001)	2.03 (1.86 to 2.22; <0.001)	1.74 (1.56 to 1.94; <0.001)
Cough	1.58 (1.31 to 1.91; <0.001)	(P=0.983)	Excluded
Phlegm	1.66 (1.38 to 2.00; <0.001)	1.17 (1.03 to 1.32; 0.012)	(P=0.433)
Hemoptysis	1.93 (1.03 to 3.59; 0.039)	(P=0.462)	Excluded
Hoarseness	1.46 (1.19 to 1.79; <0.001)	(P=0.758)	Excluded
Wheeze	2.02 (1.68 to 2.44; <0.001)	1.28 (1.00 to 1.64; 0.047)	(P=0.532)
Chest pain	1.37 (1.09 to 1.73; 0.008)	0.85 (0.64 to 1.06; 0.125)	(P=0.278)
Poor appetite	2.10 (1.54 to 2.87; <0.001)	1.35 (0.95 to 1.92; 0.092)	1.65 (1.01 to 2.69; 0.046)
Fatigued	2.41 (1.99 to 2.91; <0.001)	1.76 (1.32 to 2.34; <0.001)	1.68 (1.36 to 2.09; <0.001)

Abbreviations: CI, confidence interval; N, sample size.

***** Variables eligible for inclusion in preliminary linear regression models are limited to morbidities factors.

† Logistic regression model contains socio-demographic, smoking/alcohol exposures, morbidities and symptoms which were significantly associated with anxiety/depression in partially adjusted logistic regression models.

‡ Logistic regression models used to calculate odds ratios, confidence intervals and P-values.

4.6.2 Self-care

The large majority of participants (97%) reported no problems with self-care (Table 14). Univariate and multivariable logistic regression models for self-care are presented in Table 16a and Table 16b. Educational achievement and employment status were the only socio-demographic variables significantly associated with self-care. Increased exposure to both smoking and alcohol were also associated with higher odds of reporting some or severe problems with self-care. Their insignificance in the fully adjusted model indicates an upstream association may be present, with the effects of smoking and alcohol consumption mediating the relationship. Univariate analysis revealed several significant morbidities and symptoms. However only two morbidities, having an allergy ($OR_{adjusted}$=1.79; CI: 1.12 to 2.87; P=0.015) and FEV_1 percent predicted ($OR_{adjusted}$=0.15; CI: 0.04 to 0.51; P=0.002), and one symptom, fatigue ($OR_{adjusted}$=2.70; CI: 1.41 to 5.15; P=0.003) remained significant in the fully adjusted model. The independent variables explained 0.172 of the variation in self-care. The Hosmer–Lemeshow goodness-of-fit test suggests good model fit (P=0.909).

Table 16a. Univariate and adjusted logistic regression odds ratios of variables significantly associated with reporting some or severe problems with self-care

	Univariate	Partially adjusted*	Fully Adjusted (N=2464, Pseudo R^2=0.172)†
Variable group	**Odds ratio (CI; P-value)‡**	**Odds ratio (CI; P-value)‡**	**Odds ratio (CI; P-value)‡**
Socio-demographics and medical history		**N=2461, Pseudo R^2=0.131**	
Education level	0.82 (0.71 to 0.95; 0.010)	0.85 (0.72 to 1.01; 0.067)	(P=0.260)
Employment status			
Employed	Referent	Referent	Referent
Retired	1.43 (0.71 to 2.89; 0.316)	1.32 (0.73 to 2.42; 0.361)	1.46 (0.86 to 2.46; 0.157)
Unemployed	5.89 (2.02 to 17.21; 0.001)	5.46 (1.64 to 18.20; 0.006)	4.52 (1.31 to 15.63; 0.017)
Disabled	20.29 (9.40 to 43.82; <0.001)	18.35 (8.76 to 38.46; <0.001)	13.76 (7.37 to 25.68; <0.001)
Sick leave	35.07 (9.48 to 129.76; <0.001)	23.68 (9.38 to 59.79; <0.001)	36.03 (18.28 to 71.01; <0.001)
Other	2.79 (0.77 to 10.06; 0.118)	1.03 (-0.16 to 2.23; 0.090)	2.74 (0.77 to 9.76; 0.117)
Exposures		**N=2494, Pseudo R^2=0.027**	
Cigarettes per day (last 30 days)	1.03 (1.02 to 1.05; <0.001)	1.03 (1.01 to 1.04; 0.001)	(P=0.333)
Alcoholic drinks per week	1.01 (1.00 to 1.02; 0.002)	1.01 (1.00 to 1.02; 0.007)	(P=0.093)

Abbreviations: CI, confidence interval; N, sample size.

***** Variables eligible for inclusion in preliminary linear regression models are limited to socio-demographics, medical history and exposure factors.

† Logistic regression model contains socio-demographic, smoking/alcohol exposures, morbidities and symptoms which were significantly associated with anxiety/depression in partially adjusted logistic regression models.

‡ Logistic regression models used to calculate odds ratios, confidence intervals and P-values.

Table 16b. Univariate and adjusted logistic regression odds ratios of variables significantly associated with reporting some or severe problems with self-care

	Univariate	Partially adjusted*	Fully Adjusted (N=2464, Pseudo R^2=0.172)†
Variable group	**Odds ratio (CI; P-value)‡**	**Odds ratio (CI; P-value)‡**	**Odds ratio (CI; P-value)‡**
Morbidities		**N=2440, Pseudo R^2=0.060**	
Peripheral vascular disease	2.44 (1.03 to 5.79; 0.043)	2.02 (0.79 to 5.15; 0.142)	(P=0.319)
Asthma	2.66 (1.45 to 4.89; 0.002)	(P=0.642)	Excluded
COPD	2.76 (1.67 to 4.56; <0.001)	1.92 (0.80 to 4.62; 0.144)	(P=0.254)
Pneumonia	1.62 (0.97 to 2.70; 0.065)	(P=0.692)	Excluded
Arthritis	1.86 (1.13 to 3.05; 0.014)	(P=0.298)	Excluded
Allergy	2.07 (1.26 to 3.39; 0.004)	1.78 (0.98 to 3.21; 0.057)	1.79 (1.12 to 2.87; 0.015)
FEV_1 percent predicted	0.09 (0.03 to 0.32; <0.001)	(0.08 to 0.39; <0.001)	0.15 (0.04 to 0.51; 0.002)
Symptoms		**N=2508, Pseudo R^2=0.063**	
Dyspnea	2.42 (1.44 to 4.07; 0.001)	(P=0.413)	Excluded
Phlegm	1.77 (1.07 to 2.93; 0.026)	(P=0.685)	Excluded
Hemoptysis	3.96 (1.37 to 11.41; 0.011)	(P=0.254)	Excluded
Hoarseness	1.96 (1.19 to 3.25; 0.009)	(P=0.296)	Excluded
Wheeze	2.39 (1.45 to 3.95; 0.001)	1.70 (0.97 to 2.97; 0.062)	(P=0.219)
Chest pain	1.87 (1.08 to 3.26; 0.026)	(P=0.721)	Excluded
Poor appetite	3.20 (1.71 to 5.98; <0.001)	1.85 (0.80 to 4.24; 0.148)	(P=0.314)
Fatigue	3.88 (2.26 to 6.65; <0.001)	3.01 (1.36 to 6.69; 0.007)	2.69 (1.41 to 5.13; 0.003)

Abbreviations: CI, confidence interval; COPD, chronic obstructive pulmonary disease; FEV_1, forced expiratory volume in one second; N, sample size.

***** Variables eligible for inclusion in preliminary linear regression models are limited to socio-demographics, medical history and exposure factors.

† Logistic regression model contains socio-demographic, smoking/alcohol exposures, morbidities and symptoms which were significantly associated with anxiety/depression in partially adjusted logistic regression models.

‡ Logistic regression models used to calculate odds ratios, confidence intervals and P-values.

4.6.3 Usual activities

Almost 80% of the sample did not have any problems completing usual activities (Table 14). Odds ratios for variables associated with experiencing problems completing usual activities are reported in Tables 17a and 17b. With exception to being white and having a familial history of cancer, all socio-demographic variables were significantly associated with problems completing usual activities in univariate analysis. The effects of age, gender and education were explained away after further adjustment. In the fully adjusted model, compared to those who are employed, retired, unemployed and participants on sick leave were all more likely to report problems completing usual activities. A non-linear relationship between BMI and problems completing usual activities was also apparent in the fully adjusted model.

Multiple exposures were significantly associated with increased odds of reporting some or severe problems completing usual activities in univariate and partially adjusted analysis. However, after full adjustment no independent associations between exposure variables and usual activities were identified.

Several morbidities and symptoms were significantly associated with problems completing usual activities in all regression models. Despite 15 morbidities having a univariate association with problems completing usual activities, only arthritis remained significant in the fully adjusted model; ($OR_{adjusted}$=1.99; CI: 1.70 to 2.33; P<0.001). The partially adjusted symptoms model explained more variation in completing usual activities than any other domain (Pseudo R^2=0.1055). The fully adjusted model revealed three symptoms with downstream associations; dyspnea ($OR_{adjusted}$=2.13; CI: 1.88 to 2.42; P<0.001), poor appetite ($OR_{adjusted}$=2.17; CI: 1.26 to 3.73; P=0.005) and fatigue ($OR_{adjusted}$=2.35; CI: 1.76 to 3.16; P<0.001). The pseudo R^2 for the fully adjusted model is 0.190. Hosmer–Lemeshow goodness of fit test (P=0.663) suggests good model fit.

Table 17a. Univariate and adjusted logistic regression odds ratios of variables significantly associated with experiencing some or severe problems with completing usual activities

Variable group	Univariate Odds ratio (CI; P-value)‡	Partially adjusted* Odds ratio (CI; P-value)‡	Fully Adjusted† (N= 2423, Pseudo R^2=0.196) Odds ratio (CI; P-value)‡
Socio-demographics and medical history		**N=2459, Pseudo R^2=0.106**	
Age (years)			
1st term	0.92 (0.87 to 0.97; 0.001)§	(P=0.227)	Excluded
2nd term	1.04 (1.02 to 1.07; 0.001)§		
Gender (male vs. female)	0.76 (0.63 to 0.92; 0.005)	0.78 (0.59 to 1.02; 0.070)	(P=0.521)
Education level	0.86 (0.82 to 0.91; <0.001)	0.91 (0.86 to 0.96; <0.001)	(P=0.477)
Employment status			
Employed	Referent	Referent	Referent
Retired	1.30 (1.04 to 1.61; 0.020)	1.26 (1.05 to 1.51; 0.012)	1.40 (1.12 to 1.76; 0.003)
Unemployed	4.14 (2.52 to 6.80; <0.001)	3.76 (2.43 to 5.81; <0.001)	3.41 (2.12 to 5.50; <0.001)
Disabled	21.64 (12.40 to 37.74; <0.001)	21.04 (11.02 to 40.19; <0.001)	16.62 (9.05 to 30.5; <0.001)
Sick leave	28.49 (6.25 to 129.81; <0.001)	25.33 (6.44 to 99.63; <0.001)	29.25 (8.45 to 101.28; <0.001)
Other	1.15 (0.65 to 2.03; 0.627)	1.14 (0.60 to 2.15; 0.688)	1.33 (0.63 to 2.77; 0.452)
BMI			
1st term	$1.72e^{15}$ ($9.20e^{10}$ to $3.23e^{19}$; <0.001)‖	$3.04e^{12}$ ($3.1e^{6}$ to $3.01e^{18}$; <0.001)¶	1.02 (1.01 to 1.03; <0.001)**
2nd term	$3.31e^{-17}$ ($2.09e^{-21}$ to $5.23e^{-13}$; <0.001)‖	$8.34e^{-15}$ ($3.1e^{-20}$ to $2.24e^{-9}$; <0.001)¶	
Exposures		**N=2493, Pseudo R^2=0.016**	
Cigarettes per day (last 30 days)	1.04 (1.02 to 1.06; <0.001)	1.01 (1.01 to 1.02; <0.001)	(P=0.517)
Cigarettes per day (smoking period)	1.02 (1.01 to 1.03; <0.001)	1.02 (1.00 to 1.02; <0.001)	(P=0.635)
Time since quitting (years)	0.97 (0.95 to 1.00; 0.047)	(P=0.361)	Excluded
Alcoholic drinks per week	1.004 (1.00 to 1.01; 0.031)	(P=0.711)	Excluded
Marijuana use (joints per month)	1.01 (1.00 to 1.02; 0.006)	1.01 (1.00 to 1.02; 0.030)	(P=0.593)

Abbreviations: BMI, body mass index; CI, confidence interval; N, sample size.

***** Variables eligible for inclusion in preliminary linear regression models are limited to socio-demographic and medical history, or exposure variables.

† Logistic regression model contains socio-demographic, smoking/alcohol exposures, morbidities and symptoms which were significantly associated with anxiety/depression in partially adjusted logistic regression models.

‡ Logistic regression models used to calculate odds ratios, confidence intervals and P-values.

§ Age has a non-linear association with usual activities and is described by the following terms;

$$Term\ 1 = \left[\left(\frac{age}{10}\right)^{3} - 238.328\right] * 0.9159611$$

$$Term\ 2 = \left[\left(\frac{age}{10}\right)^{3} * ln\left(\frac{age}{10}\right) - 434.8411837\right] * 1.040678$$

‖ BMI has a non-linear relationship with usual activities and is described by the following terms;

$$Term\ 1 = \left[\left(\frac{BMI}{10}\right)^{-2} - 0.1418960034\right] * 1.72e^{15}$$

$$Term\ 2 = \left[\left(\frac{BMI}{10}\right)^{-0.5} - 0.6137514637\right] * 3.31e^{-17}$$

¶ BMI has a non-linear relationship with usual activities and is described by the following terms;

$$Term\ 1 = \left[\left(\frac{BMI}{10}\right)^{-2} - .1371742112\right] * 3.04e^{12}$$

$$\text{Term } 2 = \left[\left(\frac{BMI}{10}\right)^{-0.5} - 0.6085806195\right] * 8.34e^{-15}$$

** BMI has a non-linear relationship with usual activities and is described by the following term;

$$Term\ 1 = \left[\left(\frac{\text{BMI}}{10}\right)^{3} - 18.712593543\right] * 1.020124$$

Table 17b. Univariate and adjusted logistic regression odds ratios of variables significantly associated with reporting some or severe problems with completing usual activities

	Univariate	Partially adjusted*	Fully Adjusted (N=2423, Pseudo R^2=0.196)†
Variable group	**Odds ratio (CI; P-value)‡**	**Odds ratio (CI; P-value)‡**	**Odds ratio (CI; P-value)‡**
Morbidities		**N=2444, Pseudo R^2= 0.08**	
Stroke	1.91 (1.16 to 3.17; 0.011)	1.46 (0.97 to 2.21; 0.073)	(P=0.173)
CHD	1.41 (1.01 to 1.98; 0.046)	(P=0.928)	Excluded
Angina	1.83 (1.31 to 2.57; <0.001)	1.58 (1.17 to 2.14; 0.003)	(P=0.160)
Hypertension	1.54 (1.27 to 1.87; <0.001)	1.22 (1.01 to 1.48; 0.037)	(P=0.078)
PVD	1.56 (1.01 to 2.40; 0.043)	(P=0.360)	Excluded
Asthma	1.81 (1.36 to 2.42; <0.001)	(P=0.960)	Excluded
COPD	2.35 (1.90 to 2.90; <0.001)	1.61 (1.35 to 1.91; <0.001)	(P=0.505)
Pneumonia	1.72 (1.41 to 2.11; <0.001)	1.27 (1.03 to 1.55; 0.024)	(P=0.212)
Liver disease	1.74 (1.21 to 2.51; 0.003)	1.60 (1.18 to 2.17; 0.002)	(P=0.209)
Osteoporosis	1.54 (1.20 to 1.96; 0.001)	(P=0.319)	Excluded
Arthritis	2.42 (2.00 to 2.93; <0.001)	2.18 (1.80 to 2.65; <0.001)	1.94 (1.65 to 2.29; <0.001)
Allergy	1.70 (1.39 to 2.06; <0.001)	1.41 (1.16 to 1.72; 0.001)	1.38 (1.11 to 1.73; 0.004)
Diabetes	1.54 (1.14 to 2.09; 0.005)	(P=0.215)	Excluded
Anemia	1.80 (1.32 to 2.44; <0.001)	1.42 (1.05 to 1.92; 0.024)	(P=0.156)
FEV_1 percent predicted	0.13 (0.08 to 0.22; <0.001)	0.22 (0.14 to 0.36; <0.001)	0.32 (0.15 to 0.68; 0.003)
Symptoms		**N=2506, Pseudo R^2= 0.113**	
Dyspnea	3.55 (2.90 to 4.35; <0.001)	2.19 (1.87 to 2.57; <0.001)	1.90 (1.60 to 2.24; <0.001)
Cough	1.59 (1.31 to 1.93; <0.001)	0.79 (0.60 to 1.05; 0.099)	(P=0.590)
Phlegm	1.95 (1.61 to 2.37; <0.001)	1.39 (1.11 to 1.75; 0.004)	(P=0.068)
Hoarseness	1.99 (1.62 to 2.44; <0.001)	1.30 (1.08 to 1.55; 0.006)	(P=0.149)
Wheeze	2.24 (1.85 to 2.72; <0.001)	1.21 (0.98 to 1.51; 0.078)	(P=0.612)
Chest pain	1.77 (1.41 to 2.23; <0.001)	(P=0.599)	Excluded
Poor appetite	3.71 (2.74 to 5.02; <0.001)	2.11 (1.41 to 3.17; <0.001)	2.37 (1.52 to 3.70; <0.001)
Fatigued	3.71 (3.04 to 4.52; <0.001)	2.50 (1.85 to 3.38; <0.001)	2.31 (1.71 to 3.11; <0.001)

Abbreviations: CHD, coronary heart disease; CI, confidence interval; COPD, chronic obstructive pulmonary disease; FEV_1, forced expiratory volume in one second; N, sample size; PVD, peripheral vascular disease.

***** Variables eligible for inclusion in preliminary linear regression models are limited to morbidities or symptoms.
† Logistic regression model contains socio-demographic, smoking/alcohol exposures, morbidities and symptoms which were significantly associated with anxiety/depression in partially adjusted logistic regression models.
‡ Logistic regression models used to calculate odds ratios, confidence intervals and P-values.

4.6.4 Pain and Discomfort

Of the 2,526 participants providing data for this dimension, 1,315 (52.1%) experienced some or severe pain or discomfort (Table 14). Table 18a and 18b detail regression results for variables significantly associated with experiencing pain and discomfort. Males were less likely to report pain and discomfort problems in preliminary regression models but the association was explained away after full adjustment. Participants with higher educational achievement were less likely to experience pain or discomfort in all models. In all models, retired, unemployed and disabled participants were significantly more likely to report experiencing pain or discomfort than the employed. A positive linear relationship between increased BMI and higher odds of experiencing pain and discomfort was present in univariate and adjusted analysis.

Individuals with a heavier smoking history, as measured by cigarettes smoked in the last 30 days or during the life-time smoking period and by smoking duration (Table 20), were significantly more likely to report some or severe pain. All significant exposure associations were explained away after full adjustment. Compared to all other domains, exposures explained the least variation in pain and discomfort (Pseudo R^2=0.0072). In univariate analysis, all measured morbidities except for anemia were associated with significantly greater odds of some or severe pain, however in the fully adjusted model, only arthritis ($OR_{adjusted}$=3.36, CI: 2.80 to 4.03; P<0.001) remained significant. In univariate analysis, all measured symptoms were associated with significantly increased odds of some or severe pain. Significantly associated symptoms in the full regression model include dyspnea ($OR_{adjusted}$=1.28; CI: 1.02 to 1.60; P=0.033), chest pain ($OR_{adjusted}$=1.63; CI: 1.18 to 2.25; P=0.003), poor appetite ($OR_{adjusted}$=1.60; CI: 1.21 to 2.11; P=0.001) and fatigue ($OR_{adjusted}$=1.68; CI: 1.39 to 2.04; P<0.001). The final model explained 0.142 of the variation in pain or discomfort. The Hosmer–Lemeshow goodness of fit test (P=0.501), suggests good model fit.

Table 18a. Univariate and adjusted logistic regression odds ratios of variables significantly associated with reporting some or severe problems with pain or discomfort

Variable group	Univariate Odds ratio (CI; P-value)‡	Partially adjusted* Odds ratio (CI; P-value)‡	Fully Adjusted (N=2406, Pseudo R^2= 0.139)† Odds ratio (CI; P-value)‡
Socio-demographics and medical history		**N=2460, Pseudo R^2= 0.041**	
Gender (male vs. female)	0.80 (0.68 to 0.93; 0.005)	0.78 (0.67 to 0.92; 0.003)	(P=0.297)
Education level	0.84 (0.80 to 0.88; <0.001)	0.86 (0.80 to 0.92; <0.001)	0.88 (0.82 to 0.95; 0.001)
Employment status			
Employed	Referent	Referent	Referent
Retired	1.27 (1.07 to 1.49; 0.006)	1.25 (1.10 to 1.41; <0.001)	1.26 (1.10 to 1.40; <0.001)
Unemployed	2.02 (1.23 to 3.33; 0.005)	1.82 (1.30 to 2.55; <0.001)	(P=0.058)
Disabled	8.02 (4.21 to 15.27; <0.001)	6.99 (2.91 to 16.75; <0.001)	4.57 (2.03 to 10.3; <0.001)
Sick leave	3.81 (1.04 to 14.0; 0.043)	3.24 (0.75 to 14.0; 0.116)	(P=0.169)
Other	1.09 (0.71 to 1.69; 0.685)	1.08 (0.77 to 1.51; 0.674)	(P=0.550)
BMI	1.05 (1.03 to 1.06; <0.001)	1.05 (1.04 to 1.06; <0.001)	1.03 (1.02 to 1.05; <0.001)
Exposures		**N=2493, Pseudo R^2= 0.009**	
Cigarettes per day (last 30 days)	1.01 (1.01 to 1.02; <0.001)	1.02 (1.01 to 1.02; <0.001)	(P=0.423)
Cigarettes per day (smoking period)	1.01 (1.00 to 1.02; 0.006)	(P=0.224)	Excluded
Smoking duration (years)	1.01 (1.00 to 1.03; 0.038)	1.02 (1.00 to 1.03; 0.025)	(P=0.561)
Time since quitting (years)	0.99 (0.97 to 1.01; 0.295)	1.03 (1.01 to 1.05; 0.008)	(P=0.750)
Alcoholic drinks per week	$7.27e^{-8}$ ($2.44e^{-8}$ to $1.21e^{-7}$; 0.003)§	1.0002 (1.0001 to 1.0004; <0.001)‖	(P=0.842)

Abbreviations: BMI, body mass index; CI, confidence interval; N, sample size.

***** Variables eligible for inclusion in preliminary linear regression models are limited to socio-demographic and medical history, or exposure factors.

† Logistic regression model contains socio-demographic, smoking/alcohol exposures, morbidities and symptoms which were significantly associated with anxiety/depression in partially adjusted logistic regression models.

‡ Logistic regression models used to calculate odds ratios, confidence intervals and P-values.

§ Alcoholic drinks per week has a non-linear association with experiencing pain or discomfort, and is described by the following term;

$$Term\ 1 = \{(alcohol + 0.0999999940395355)/100)^{-1} - 6.211180127\} * \ 1.000263$$

‖ Alcoholic drinks per week has a non-linear association with experiencing pain or discomfort, and is described by the following term;

$$Term\ 1 = \{(alcohol + 0.0999999940395355)/100)^{-2} - 39.07149013\} * \ 7.27e^{-8}$$

Table 18b. Univariate and adjusted logistic regression odds ratios of variables significantly associated with reporting some or severe problems with pain or discomfort

	Univariate	Partially adjusted*	Fully Adjusted (N=2406, Pseudo R^2= 0.139)[†]
Variable group	**Odds ratio (CI; P-value)[‡]**	**Odds ratio (CI; P-value)[‡]**	**Odds ratio (CI; P-value)[‡]**
Morbidities		**N=2444, Pseudo R^2= 0.091**	
CHD	1.46 (1.08 to 1.98; 0.015)	(P=0.424)	Excluded
Angina	1.68 (1.21 to 2.32; 0.002)	1.42 (0.92 to 2.18; 0.114)	(P=0.185)
Hypercholesterolemia	1.35 (1.15 to 1.58; <0.001)	1.20 (1.11 to 1.31; <0.001)	1.21 (1.08 to 1.36; 0.001)
Hypertension	1.46 (1.24 to 1.72; <0.001)	(P=0.207)	Excluded
Asthma	1.47 (1.12 to 1.92; 0.005)	(P=0.840)	Excluded
COPD	1.55 (1.27 to 1.88; <0.001)	1.22 (0.99 to 1.51; 0.067)	(P=0.781)
Pneumonia	1.57 (1.31 to 1.87; <0.001)	1.29 (1.06 to 1.56; 0.012)	(P=0.117)
Osteoporosis	1.62 (1.30 to 2.02; <0.001)	1.27 (0.99 to 1.63; 0.057)	(P=0.127)
Arthritis	3.83 (3.23 to 4.55; <0.001)	3.66 (2.95 to 4.53; <0.001)	3.32 (2.74 to 4.01; <0.001)
Allergy	1.35 (1.14 to 1.61; <0.001)	(P=0.307)	Excluded
Diabetes	1.65 (1.24 to 2.18; 0.001)	1.41 (0.95 to 2.09; 0.089)	(P=0.112)
Anemia	1.39 (1.05 to 1.85; 0.023)	(P=0.712)	Excluded
FEV_1 percent predicted	0.40 (0.26 to 0.62; <0.001)	0.58 (0.34 to 0.98; 0.033)	(P=0.632)
Symptoms		**N=2507, Pseudo R^2= 0.057**	
Dyspnea	2.24 (1.91 to 2.63; <0.001)	1.48 (1.18 to 1.85; 0.001)	1.31 (1.05 to 1.64; 0.016)
Cough	1.38 (1.18 to 1.62; <0.001)	0.87 (0.74 to 1.02; 0.080)	(P=0.503)
Phlegm	1.56 (1.34 to 1.83; <0.001)	1.24 (1.00 to 1.53; 0.047)	(P=0.521)
Hoarseness	1.62 (1.35 to 1.94; <0.001)	1.17 (1.02 to 1.35; 0.023)	(P=0.575)
Wheeze	1.89 (1.61 to 2.23; <0.001)	1.27 (1.05 to 1.53; 0.013)	(P=0.110)
Chest pain	2.19 (1.76 to 2.73; <0.001)	1.49 (1.17 to 1.91; 0.001)	1.61 (1.17 to 2.23; 0.004)
Poor appetite	2.25 (1.63 to 3.09; <0.001)	1.35 (1.14 to 1.61; 0.001)	1.49 (1.14 to 1.95; 0.004)
Fatigued	2.49 (2.11 to 2.94; <0.001)	1.84 (1.51 to 2.24; <0.001)	1.70 (1.41 to 2.03; <0.001)
Weight loss	1.44 (1.09 to 1.92; 0.011)	(P=0.524)	Excluded

Abbreviations: CHD, coronary heart disease; CI, confidence interval; COPD, chronic obstructive pulmonary disease; FEV_1, forced expiratory volume in one second; N, sample size.

***** Variables eligible for inclusion in preliminary linear regression models are limited to morbidities and symptoms factors.

† Logistic regression model contains socio-demographic, smoking/alcohol exposures, morbidities and symptoms which were significantly associated with anxiety/depression in partially adjusted logistic regression models.
‡ Logistic regression models used to calculate odds ratios, confidence intervals and P-values.

4.6.5 Anxiety or depression

A total of 2526 participants provided anxiety/depression domain level quality of life scores. Approximately two-thirds of the sample reported having no anxiety or depression problems (Table 14). Variables significantly associated with experiencing some or severe problems with anxiety or depression are described in Tables19a and Table 19b. Socio-demographic and symptom variables explained the greatest amount of variation in experiencing anxiety or depression. Odds of experiencing anxiety or depression decreased among older participants. However, these effects were explained away as other variables were added to the regression model. In all regression models females were almost twice as likely to experience problems with anxiety or depression. Participants with higher educational achievement were less likely to experience anxiety or depression. Retirees did not significantly differ compared to employed participants. However, unemployed ($OR_{adjusted}$=2.25; CI: 1.22 to 4.15; P=0.009), disabled ($OR_{adjusted}$=3.84; CI: 2.44 to 6.03; P<0.001) and individuals on sick leave ($OR_{adjusted}$=6.67; CI: 2.66 to 16.7; P<0.001) were all significantly more likely to experience some or severe problems.

Of all partially adjusted regression models from each variable group, exposures explained the least variation in experiencing anxiety or depression (Pseudo R^2=0.008). Adjusting for other exposure factors, increased time since quitting smoking and reduced alcohol consumption were both associated with increased odds of reporting no problems with anxiety or depression. After full adjustment no exposures were independently associated with anxiety or depression. Several morbidities were significantly associated with increased odds of reporting some or severe problems with anxiety or depression in univariate and partially adjusted regression models. However, only the effect of arthritis ($OR_{adjusted}$=1.24; CI: 1.04 to 1.46; P=0.014) remained significant after adjustment for socio-demographic, exposure and symptom variables. Almost all symptoms were

significantly associated with increased odds of anxiety or depression in univariate analysis. Half of the symptoms remained significant in the partially adjusted regression model and one third were associated with increased odds of experiencing anxiety or depression in the fully adjusted model. Having a poor appetite ($OR_{adjusted}$=2.51; CI: 1.73 to 3.63; P<0.001) and feeling fatigued ($OR_{adjusted}$=2.31; CI: 2.03 to 2.62; P<0.001) were symptoms most highly associated with experiencing anxiety or depression. The final regression model explained 0.114 of the total variation in experiencing anxiety or depression. Hosmer–Lemeshow goodness of fit test (P=0.490) suggests good model fit.

Table 19a. Univariate and adjusted logistic regression odds ratios of variables significantly associated with reporting some or severe problems with anxiety or depression

	Univariate	Partially adjusted*	Fully Adjusted (N=2493, Pseudo R^2=0.114)†
Variable group	**Odds ratio (CI; P-value)‡**	**Odds ratio (CI; P-value)‡**	**Odds ratio (CI; P-value)‡**
Socio-demographics and medical history		**N=2460, Pseudo R^2=0.051**	
Age (years)	0.97 (0.95 to 0.98; <0.001)	(P=0.234)	Excluded
Gender (male vs. female)	0.54 (0.46 to 0.64; <0.001)	0.52 (0.43 to 0.64; <0.001)	0.56 (0.45 to 0.71; <0.001)
Education	0.89 (0.85 to 0.94; <0.001)	0.92 (0.87 to 0.98; 0.010)	(P=0.136)
Employment status			
Employed	Referent	Referent	Referent
Retired	1.00 (0.83 to 1.20; 0.967)	0.96 (0.83 to 1.12; 0.616)	1.03 (0.88 to 1.20; 0.729)
Unemployed	3.18 (1.95 to 5.17; <0.001)	2.93 (1.78 to 4.83; <0.001)	2.25 (1.22 to 4.15; 0.009)
Disabled	5.42 (3.38 to 8.71; <0.001)	5.29 (3.23 to 8.65; <0.001)	3.84 (2.44 to 6.03; <0.001)
Sick leave	5.40 (1.65 to 17.68; 0.005)	7.81 (3.04 to 20.07; <0.001)	6.67 (2.66 to 16.7; <0.001)
Other	1.18 (0.74 to 1.88; 0.486)	1.09 (0.68 to 1.75; 0.714)	1.15 (0.57 to 2.31; 0.691)
Family history of cancer	1.36 (1.14 to 1.62; 0.001)	1.21 (0.91 to 1.60; 0.187)	(0.166)
Exposures		**N=2493, Pseudo R^2=0.008**	
Cigarettes per day (last 30 days)	1.01 (1.01 to 1.02; <0.001)	(P=0.267)	Excluded
Smoking duration (years)	0.98 (0.97 to 1.00; 0.013)	0.97 (0.96 to 0.98; <0.001)	(P=0.676)
Time since quitting (years)	0.97 (0.95 to 0.99; 0.004)	0.96 (0.92 to 0.99; 0.017)	(P=0.194)
Alcoholic drinks per week	1.00 (1.00 to 1.01; 0.061)	1.00 (1.00 to 1.01; 0.036)	(P=0.196)
Marijuana use (joints per month)	0.95 (0.92 to 0.98; <0.001)§	(P=0.234)	Excluded

Abbreviations: CI, confidence interval; N, sample size.

***** Variables eligible for inclusion in preliminary linear regression models are limited to socio-demographic and medical history, or exposure variables.

† Logistic regression model contains socio-demographic, smoking/alcohol exposures, morbidities and symptoms which were significantly associated with anxiety/depression in partially adjusted logistic regression models.

‡ Beta-coefficients, confidence intervals and p-values calculated using logistic regression models.

§ Average weekly marijuana use has a non-linear relationship with anxiety/depression and is described by the following term;

$$Term\ 1 = \left\{\left[\frac{marijuana\ use+1}{100}\right]^{-0.5} - 4.472135955\right\} * 0.9499725$$

Table 19b. Univariate and adjusted logistic regression odds ratios of variables significantly associated with reporting some or severe problems with anxiety or depression

	Univariate	Partially adjusted*	Fully Adjusted (N=2493, Pseudo R^2=0.114)†
Variable group	**Odds ratio (CI; P-value)‡**	**Odds ratio (CI; P-value)‡**	**Odds ratio (CI; P-value)‡**
Morbidities		**N=2444, Pseudo R^2=0.024**	
Asthma	1.25 (0.95 to 1.65; 0.110)	(P=0.876)	Excluded
COPD	1.59 (1.30 to 1.94; <0.001)	1.41 (1.09 to 1.82; 0.009)	(P=0.388)
Pneumonia	1.57 (1.30 to 1.88; <0.001)	1.39 (1.16 to 1.66; <0.001)	(P=0.150)
Osteoporosis	1.46 (1.17 to 1.82; 0.001)	1.22 (0.94 to 1.54; 0.110)	(P=0.260)
Arthritis	1.56 (1.32 to 1.85; <0.001)	1.44 (1.20 to 1.73; <0.001)	1.24 (1.04 to 1.46; 0.014)
Anemia	1.86 (1.40 to 2.47; <0.001)	1.63 (1.28 to 2.06; <0.001)	(P=0.110)
Blood disease	1.75 (0.99 to 3.09; 0.054)	1.57 (1.15 to 2.14; 0.004)	(P=0.522)
FEV_1 percent predicted	0.64 (0.40 to 1.00; 0.050)	(P=0.490)	Excluded
Symptoms		**N=2507, Pseudo R^2=0.086**	
Dyspnea	2.32 (1.95 to 1.75; <0.001)	1.63 (1.32 to 2.01; <0.001)	1.44 (1.21 to 1.72; <0.001)
Cough	1.15 (0.98 to 1.36; 0.094)	0.76 (0.61 to 0.95; 0.017)	(P=0.498)
Phlegm	1.27 (1.08 to 1.50; 0.005)	(P=0.774)	Excluded
Hemoptysis	1.78 (0.98 to 3.24; 0.059)	(P=0.706)	Excluded
Hoarseness	1.53 (1.27 to 1.84; <0.001)	(P=0.372)	Excluded
Wheeze	1.57 (1.32 to 1.86; <0.001)	(P=0.836)	Excluded
Chest pain	2.10 (1.70 to 2.59; <0.001)	1.36 (1.04 to 1.79; 0.026)	(P=0.104)
Poor appetite	4.44 (3.25 to 6.05; <0.001)	2.79 (1.92 to 4.06; <0.001)	2.51 (1.73 to 3.63; <0.001)
Fatigued	3.33 (2.79 to 3.96; <0.001)	2.53 (2.28 to 2.81; <0.001)	2.31 (2.03 to 2.62; <0.001)
Weight loss	1.48 (1.11 to 1.97; 0.007)	(P=0.588)	Excluded

Abbreviations: CI, confidence interval; COPD, chronic obstructive pulmonary disease; FEV_1, forced expiratory volume in one second; N, sample size.

***** Variables eligible for inclusion in preliminary linear regression models are limited to morbidities or symptoms variables.

† Logistic regression model contains socio-demographic, smoking/alcohol exposures, morbidities and symptoms which were significantly associated with anxiety/depression in partially adjusted logistic regression models.

‡ Logistic regression models used to calculate odds ratios, confidence intervals and P-values.

4.7 Background to the resultant models

Assumptions verified when conducting linear regression are described below.[144]

4.7.1 Assumption 1.Linear relationship between independent and dependent variables

As previously discussed, the MFP selection procedure allows for identification of independent variables which have a non-linear relationship with the dependent variable.

4.7.2 Assumption 2. Independence of observations

This assumption requires that all predictor variables be independent of one another, that the results of one participant do not affect the results of others. All participants completed study questionnaires independently and without knowledge of study results. However, modelling used cluster analysis to handle sampling by study sites, which is recommended when correlation of observations is expected within specific groups of the sample. Observations must still be independent between clusters.[153] Cluster analysis produces robust standard errors which compensate for clustering of data within subsets of the sample. Regarding PanCan data, it was possible that participants from the same study site, experiencing similar location specific exposures, would report VAS scores which were correlated with one another. Univariate and adjusted analyses revealed significant relationships between study site and VAS score. Cluster adjustment was applied to regression analyses to account for potential differences in HRQoL due to sampling of study participants in different study sites.

4.7.3 Assumption 3. Normality

VAS score, which should follow a normal distribution, is slightly skewed left with the upper tail being cut off since it is impossible for VAS scores to exceed 100 (Figure 12). Of multiple transformations, square and cubic transformations most closely resembled a normal distribution; however, negative kurtosis was present in both (Appendix 6). This assumption violation might typically be concerning. However,

according to Lumley et al. (2002), the normality assumption is unnecessary for large data sets, such as the PanCan, due to the central limit theorem.[154] Therefore, the original untransformed VAS score was utilized. Bootstrap analysis was conducted to assess stability of the final model.

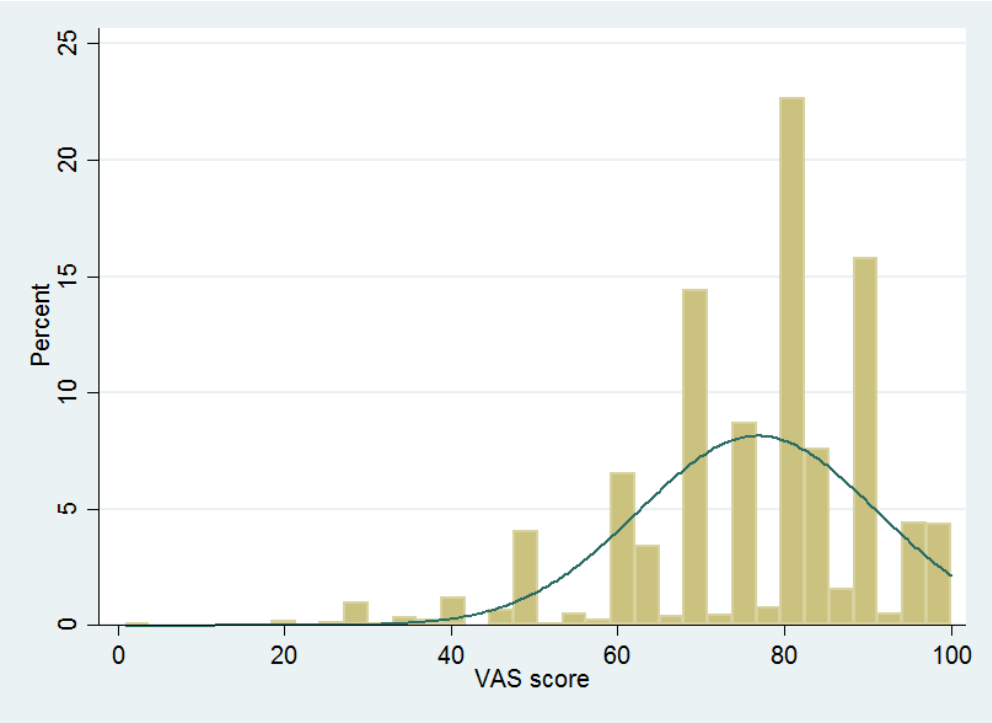

Figure 11. Histogram of PanCan Visual Analog Scale scores

4.7.4 Assumption 4. Correlation of predictors

Collinearity begins to severely distort model estimation and subsequent prediction when Pearson's r≥0.7.[139] The majority of independent variables were not highly correlated. However, high correlations were present between smoking status and number of cigarettes smoked per day (30 days prior to baseline) (r=0.712), smoking status and time since quitting (r=-0.727), pack years and average number of cigarettes smoked per day (r=0.889) as well as FEV_1 and FVC_1 (r=0.729). Only one variable from each pairing was selected for potential inclusion; cigarettes smoked per day (last 30 days), time since quitting, average number of cigarettes smoked per day (smoking

period) and FEV_1 percent predicted. Without a specific measure of smoking status, former smokers are still represented in the regression model, embedded as values of 0 in the variables cigarettes per day (30 days prior to baseline) with years since quitting smoking specified. Selected variables also exhibited greater univariate association with VAS scores. Average number of cigarettes smoked per day (smoking period) was selected for inclusion instead of pack-years. Peto (2012) noted that the effects of smoking should not be measured in pack-years and that a combination of smoking duration and average number of cigarettes per day is more scientifically helpful.[155] As a sensitivity analysis, collinear variables were forced into the final model and in all cases were insignificant and had no effect on the R^2 value.

4.7.5 Assumption 5. Reducing measurement error

To ensure accuracy of measurements qualified research assistants administered study questionnaires to all participants with appropriate tools when necessary. Errors in memory may have been present among variables measuring historical exposures, such as smoking duration. Additionally, participants may be inclined to under- or over-report various measurements for reasons of social desirability, such as number of cigarettes consumed within the previous 30 days or level of educational attainment.[156]

Distributions of all continuous variables are available in Table 2. Outlying observations were identified by identifying observations which exceeded three standard deviations from the mean.[156] Based on previous research, outlying observations were within a reasonable range of values and not representative of erroneous data. To ensure data naturality and interpretability, truncation or transformation was not completed as part of primary analysis.[145] As a sensitivity analysis, continuous variables were transformed to reduce the effect of outliers and also replaced their original in the final regression model.[145,157] In all cases, R^2 decreased and interpretation of the associations

between independent variables and VAS score did not change. The proportion of outliers for all continuous variables did not exceed reasonable limits, $\frac{\sqrt{N}}{N} \cong 1.9\%$.[158]

4.7.6 Assumption 6. Homoscedasticity

This assumption requires that error term in the dependent variable be consistent across the ranges of values of predictor variables. A plot of residuals against fitted values of VAS score reveals a relatively equally random distribution around 0, indicating minimal differences in error terms of VAS score across the sample (Figure 12).

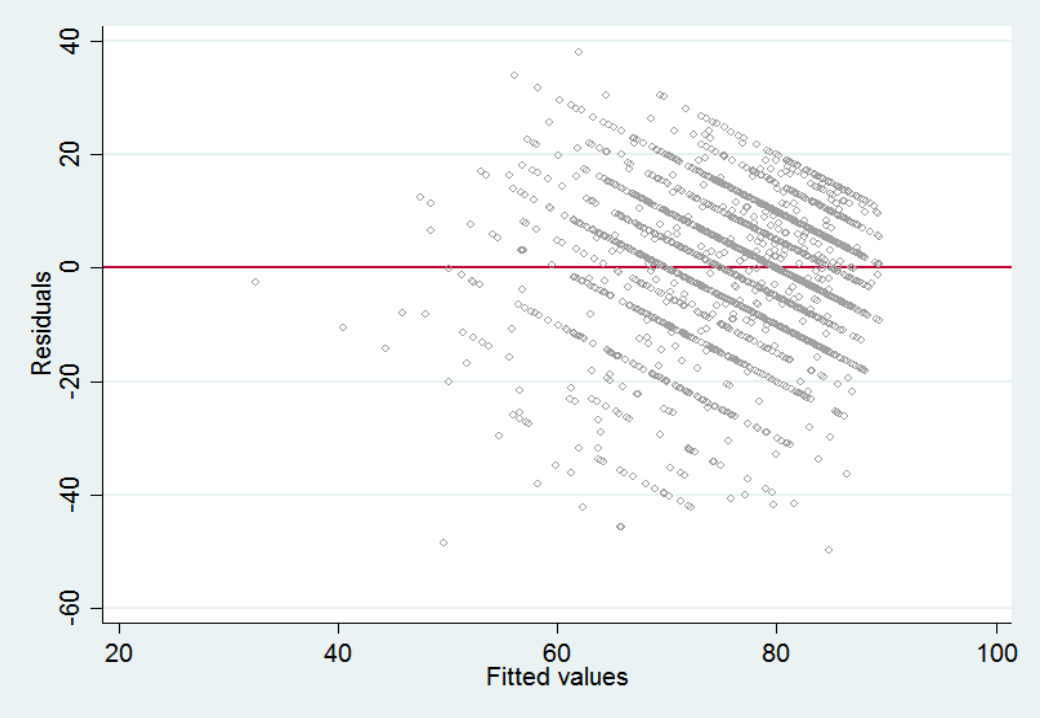

Figure 12. Plot of residuals against fitted values of VAS score

4.7.7 Background to the resultant models summary

Assumptions of linear regression modelling were analyzed. The MFP selection procedure was utilized to identify non-linear associations in addition to linear associations. The assumption of normality in the dependent variable was violated. Assumption specific issues are not believed to impact results interpretability.

The assumptions of logistic regression; that the dependent variable be dichotomous and that each subgroup be mutually exclusive, were met. Greater than 50 participants composed dependent variable subgroups, however, the categories of experiencing some problems and severe problems had to be combined to meet this criterion. Therefore, domain level data was analyzed using binary logistic regression rather than ordinal logistic regression.[149]

Chapter 5: Discussion and Conclusions

This investigation aimed to examine the associations between socio-demographic, medical history, exposure, morbidity and symptom variables, and health-related quality of life (HRQoL). Data from the Pan Canadian Early Detection of Lung Cancer (PanCan) study, a large nationally representative sample of former and current smokers, were analyzed. General HRQoL was measured using the EuroQol visual analog scale (EQ VAS) and EuroQol 5 Dimension (EQ-5D) descriptive system. The multivariable fractional polynomial (MFP) selection procedure was combined with linear and logistic regression modelling to identify factors associated with HRQoL.

5.1 Pertinent study findings

Multiple socio-demographic variables were found to have significant upstream and downstream associations with VAS scores as well as EQ-5D domains. A total of 15 variables were significantly associated with VAS score, after adjusting for all factors. Results suggest that increased smoking exposure has negative upstream and downstream associations with HRQoL. The fully adjusted relationship between cigarettes smoked in the previous 30 days and VAS score, for participants who smoke one pack per day, suggests an association which is stronger than experiencing phlegm, hypercholesterolemia, pneumonia or arthritis. Furthermore, smoking duration, intensity of smoking throughout the smoking period, and time since quitting smoking are believed to have upstream associations with VAS score and domain level problems, which primarily appeared to be mediated through morbidities and symptoms.

Of all morbidities, FEV_1 percent predicted was most often associated with reduced HRQoL. Symptoms variables explained the greatest variation in VAS score and several symptoms increased odds of reporting some or severe problems in multiple EQ-5D domains. Feeling fatigued or having a poor appetite had the greatest negative effect

on HRQoL when compared to gender, race/ethnicity and all other categorical symptoms or morbidities variables. Results of this study add to the limited number of population-based investigations which simultaneously examine the relationships between socio-demographic, medical history, exposure, morbidity and symptom variables and HRQoL.

5.2 Questions #1 and #2. *What socio-demographic, medical history, exposure, morbidity, and symptom variables are associated with a single estimator of overall HRQoL as measured by the EQ VAS in individuals with a heavy smoking history? What socio-demographic, medical history, exposure, morbidity, and symptom variables are associated with HRQoL domains (mobility, self-care, usual activities, pain/discomfort and anxiety/depression) as measured by the EQ-5D?*

5.2.1 Socio-demographic and medical history variables

Linear regression modelling suggests that socio-demographic and medical history variables may indirectly and directly affect HRQoL. All socio-demographic and medical history variables were significantly associated HRQoL in unadjusted models, with exception to familial cancer history. After full adjustment, gender, race/ethnicity, employment status and BMI had significant associations with HRQoL.

Aging

Univariate analysis shows that older participants report higher VAS scores than younger participants. Increase in VAS score levels off at approximately 70 years of age. Lower quality of life scores amongst younger participants are thought to be due to increased smoking exposure among younger participants. Furthermore, participants below the median age, 80, report approximately one more symptom than those above. Increased age was also significantly associated with increased likelihood of reporting some or severe problems in the domains of usual activities and mobility, and decreased risk of reporting problems with anxiety/depression. Similar to VAS score, these associations were no longer present in the fully adjusted model. It is important to note, potential participants with heaviest smoking exposure would not have survived as long

as those with less exposure. This may give the appearance that older participants have improved HRQoL when it may be that those with lowest HRQoL and greatest smoking exposure were excluded from analysis.

The insignificance of age in all fully adjusted models and bootstrap analysis indicates that changes in VAS score or dimensional ratings as one ages are due to changes in other socio-demographic, medical history, smoking or alcohol exposures, or changes of morbidities and symptoms diagnosed. Multiple population based studies report a decline in HRQoL throughout the aging process.[29,32] This generally holds true for smokers, however, several studies have concluded that HRQoL will increase, regardless of age, due to smoking cessation.[159-161]

Gender

Chi-square tests as well as unadjusted and adjusted regression models suggest no significant differences in VAS score by gender. However, a large proportion of bootstrap replications indicate that males consistently report lower VAS scores than females. Gender was also a significant factor in predicting HRQoL among EQ-5D domains. Males reporting lower HRQoL scores runs contrary to many studies which conclude that females tend to report lower HRQoL than males.[83,89,162-164] It may be that important predictors of HRQoL, specifically those more common among females, were not included in our analysis. Mental health measures are of particular interest since logistic regression analysis reveals that females were almost twice as likely to report some or severe problems with depression and anxiety than males. A recent investigation conducted by Coste and colleagues (2014) note the effect of many significant predictors of SF-36 dimensions were mitigated or insignificant after adjusting for depression.[165] Other investigations recognize the importance of other socio-demographic

characteristics such as marital status, social relationships, or family income, which were not available for inclusion in regression analysis.[83,163,166]

Race/ethnicity

Results from the fully adjusted VAS regression model and bootstrap replications indicate non-Whites report lower VAS scores than White participants. However, race/ethnicity was insignificant when assessing all domain level logistic regression models. This suggests that race/ethnicity is an important predictor of HRQoL outside of the domains included in the EQ-5D. Previous research agrees that non-white populations report lower HRQoL than white populations.[167] A review of multiple generic HRQoL instruments and race/ethnicity finds that non-white populations report poorer cognitive and social functioning, which are domains not measured by the EQ-5D.[168]

Education

Results indicate education to be an important upstream factor associated with VAS and HRQoL domains. PanCan participants with lower education reported greater exposure to harmful behaviours, such as smoking or consuming alcohol, in addition to increased total number of morbidities and symptoms. Maralani (2014) points out the advantages of being in a school environment in terms of reducing likelihood of smoking in later life. In addition to providing education and skills associated with reduced smoking, many schools enforce policies against smoking which are effective in reducing smoking later in life.[169] Given that 88% of smokers begin smoking by age 18 and 99% by age 26, it is imperative that social policies encourage any level of further education. School policies must not only inform students of the harmful effects of smoking, but also aim to incorporate all students in activities. Even activities unrelated to physical health have shown to decrease likelihood of smoking in later life.[57] Furthermore, higher education may result in increased salary, also associated with improved HRQoL.[163]

Employment status

Compared to employed participants, retired individuals showed little HRQoL difference in univariate, and partially or fully adjusted linear regression analysis. Despite their unemployed job status, and likely lower income, unemployed participants reported greater weekly alcoholic drinks, smoked more cigarettes within the 30 days prior to baseline and more cigarettes per day throughout their smoking lifetime than employed participants. Unemployed participants consistently reported lower HRQoL than employed participants. This finding agrees with previous research which adds that unemployed participants report lower HRQoL in addition to more inadequate prevention behaviours.[170] Extremera and Rey (2014), address the importance of cognitive emotion regulation strategies for those experiencing reduced HRQoL due to the social and financial stresses of being unemployed.[171]

Disabled participants or those on sick leave reported the lowest HRQoL of any subgroups measured. This pattern was also present in each of the EQ-5D dimensional measures. Even though several morbidities have been adjusted for, some physical or mental health conditions among the disabled or those on sick leave may be more impactful than the morbidities adjusted for in this study.

Body mass index

Unsurprisingly, a non-linear relationship between BMI and VAS score was identified after full adjustment. BMI was also a significantly associated with experiencing problems in multiple domains. Furthermore, hypercholesterolemia and myocardial infarction, both associated with decreased HRQoL, are more common among the overweight and obese.[172,173] Overweight and obese smokers must be mindful of the harmful cardiovascular effects caused by smoking as it increases risk for serious

physical consequences in addition to further reduction in HRQoL. The importance of BMI to general and domain specific HRQoL supports findings from previous research.[174-176]

5.2.2 Smoking and alcohol exposure

As previously discussed, multiple studies have confirmed the importance of smoking status, duration and intensity to the development of chronic cardiovascular or pulmonary diseases. Average number of cigarettes smoked per day (previous 30 days) was the only smoking exposure variable independently associated with HRQoL. However, smoking duration, smoking intensity throughout smoking period and time since quitting smoking appear to be upstream factors associated with HRQoL. Fully adjusted regression analysis suggests that smoking 20 cigarettes per day may be approximately equal to a 3 point reduction in VAS score. Compared to smoking status, or average number of cigarettes consumed throughout a participant's smoking period, cigarettes smoked per day in the 30 days prior to baseline is thought to be more meaningful to the variation in VAS scores because it more accurately describes individual participants' current smoking status and intensity, and is less prone to measurement error than smoking intensity during the lifetime smoking duration. Fully aware that smoking may be damaging to their body, study participants may also report lower VAS scores simply because they associate present smoking intensity with poor HRQoL. Multiple studies agree with findings that increased recent daily tobacco exposure is associated with decreased HRQoL.[19,21,166] It is also possible that important predictors relevant to HRQoL were not included in regression modelling.

Pertaining to domain level data, smoking duration was the only smoking exposure variable associated with increased odds of experiencing any domain level problem, mobility. In their study of maintaining and restoring lower body mobility, Ostbye and colleagues (2002) discover that not smoking or smoking cessation is related to

improvements in mobility. Risk of experiencing mobility impairments was equal to that of never-smokers after 15 years of cessation.[177]

Alcohol exposure

Aside from cigarettes consumed within 30 days prior to baseline, weekly alcohol intake is the only exposure variable significantly associated with VAS in the univariate, partially adjusted and fully adjusted regression models. Weekly alcohol consumption was significantly associated with experiencing anxiety or depression, problems in completing usual activities and problems with self-care in univariate and partially adjusted logistic regression models.

Non-parametric test of trend show a slight increase in VAS score among participants consuming 3 to 10 alcoholic drinks per week in comparison to those consuming less than 3 drinks per week. Participants consuming greater than 10 and greater than 20 alcoholic drinks per week showed successive decreases in VAS score. A review of potential health benefits arising from moderate alcohol consumption reports that all-cause mortality, cardiovascular disease, inflammation, the immune system and cancer, among other health conditions, may be positively affected by moderate alcohol intake.[178] These results correspond with results from a recent meta-analysis which concludes that moderate alcohol intake is associated with a decrease in cardiovascular disease as well as all-cause mortality.[179] However, pertaining to HRQoL, results from Strandberg (2004) [222], Byles (2006)[159], and Volk (1997)[160] indicate that moderate alcohol exposure offers little to no improvement over drinking less. Further research investigating effects of alcohol consumption and its potential indirect benefits on HRQoL by improving cardiovascular health are required.[178]

5.2.3 Morbidities

Cardiovascular morbidities

Despite all eight cardiovascular morbidities being associated with decreased HRQoL, only three, as well as BMI, were significantly associated with HRQoL in the fully adjusted MFP linear regression model; myocardial infarction, hypercholesterolemia and peripheral vascular disease (PVD). In addition to these variables, hypertension was included in slightly below 50% of bootstrap replications, indicating that it may be a meaningful predictor. Among domain level data, participants with hypertension, congestive heart failure or PVD reported higher odds of experiencing some or severe problems with mobility. Several studies agree with these conclusions.[30,36,37]

It may be possible that cardiovascular morbidities not included in final regression models are still important HRQoL factors. Royston and Sauerbrei (2008) note that collinearity may affect a model even though a correlation coefficient between variables does not exceed $r>0.7$. In such a case, only one of two potentially important variables would be included in the final regression model or bootstrap replication. However, various sensitivity analyses which included different combinations of cardiovascular morbidities forced into the final regression model showed no improvement in R^2 values.

Pulmonary function and pulmonary morbidities

Mean FEV_1 percent predicted, of the PanCan sample was approximately 80%. Slightly greater than 10% of participants provided FEV_1 percent predicted value greater than 100%. Therefore, approximately 10% of participants have FEV_1 percent predicted greater than the reference population used to compute FEV_1 percent predicted. These participants may be ‘healthy smokers’, those that smoke regularly but are unaffected by the common health effects of smoking. Non-parametric test of trend reveals that participants whose FEV_1 is below 100%, have a significantly longer smoking history,

greater number of average cigarettes smoked per day throughout the smoking period of their lifetime, and significantly more cigarettes smoked per day within the 30 days prior to baseline, than those above 100%. These findings agree with data from longitudinal studies which have linked smoking to quicker declines in FEV_1 percent predicted.[182-185]

PanCan results indicate that FEV_1 percent predicted is an important predictor of VAS score and multiple domain outcomes. However, FEV_1 percent predicted or pulmonary function variables are seldom analyzed in population samples. These results warrant further examination of pulmonary function variables among population samples, especially those including smokers. Multiple studies agree that FEV_1 percent predicted and other pulmonary function variables significantly predict HRQoL.[186-188]

A non-linear relationship was present in the fully adjusted regression model as well as over 60% of bootstrap replications. Researchers should be sure to adapt analysis involving FEV_1 and HRQoL to handle non-linearity.

Other pulmonary morbidities were rarely associated with VAS score or domain level outcomes in the presence of other covariates. However, pulmonary fibrosis was included among 58.2% of bootstrap replications, identifying it as a potentially important variable. As a population study, rather than clinical, prevalence and severity of chronic conditions were low. Therefore, some morbidities anticipated to be meaningful were not statistically significant. It is expected that pulmonary fibrosis is importantly associated with decreased HRQoL, but that the low number of participants with pulmonary fibrosis, 3 individuals, did not provide sufficient study power. Previous research published by Tomioka and colleagues agrees that pulmonary fibrosis is a significant predictor of HRQoL.[189]

Other morbidities

Arthritis is the only 'other morbidity' significantly associated with VAS score after full adjustment. Arthritis sufferers were also more likely to report problems with mobility, usual activities and pain or discomfort. Ibn Yacoub and colleagues (2012) agree, discovering that arthritis sufferers reported significantly lower HRQoL in all domains of the SF-36 and EQ VAS compared to the general population.[190]

Results from this thesis confirm findings from previous population based investigations and clinical studies, adding that the population of former and current smokers at elevated risk of developing lung cancer is also affected. Furthermore, this investigation adds that factors significantly associated with HRQoL are significant in the presence of other socio-demographic, exposure, symptom, and morbidity factors not commonly simultaneously included in previous studies. Additionally, symptoms factors and pulmonary function measures, such as FEV_1 percent predicted, were found to be highly significant with VAS and domain level HRQoL measurements. Inclusion of these variables, which are rarely incorporated in HRQoL analysis, should be considered.

5.2.4 Symptoms

Of all variable groups, symptoms variables explained the greatest variation in VAS score. Former smokers reported approximately one fewer symptom than current smokers. After full adjustment, five symptoms were highly associated with VAS score and four were included in over 90% of bootstrap replications. Multiple symptoms were significantly associated with each of the EQ-5D domains, in particular, dyspnea, fatigue and poor appetite. In their investigation of symptom distress in COPD patients Blinderman and colleagues (2009) discover dyspnea, chest pain, poor appetite and fatigue to be among the most frequently experienced, severe or stressful symptoms.[180] Furthermore, poor quality of life was associated with higher symptom distress.

Investigations by Voll-Aanerud (2008) and (2010), of participants diagnosed with COPD agree with our findings, concluding that dyspnea, chest pain and phlegm production significantly affect HRQoL.[33,35] Leander and colleagues (2009) agree, adding that symptoms are common among individuals without diagnosed morbidities.[181] Of PanCan participants with no respiratory morbidities, almost 80% reported experiencing one or more symptoms.

5.3 How does the VAS score of PanCan participants compare to other samples in Canada and elsewhere?

Compared to other Canadian and international populations, PanCan participants generally reported lower VAS scores than general populations and similar scores compared to smokers. In a comparison of the EQ-5D to the SF-12, questionnaires from 1,490 residents of Calgary, Alberta were analyzed. Mean VAS score of participants aged 55 to 64, 78.8, was greater than PanCan participants of the same age group from Alberta, 74.9. This is unexpected considering the smoking history of PanCan participants. However, participants aged 65 to 74 reported similar mean VAS scores (Alberta sample: mean=76.0 vs. PanCan sample: mean=76.6).[125]

Luo and colleagues examined data from a general population sample. Americans aged 45 to 64 reported a mean VAS score of 84.[127] A similar PanCan age group, 50 to 65 years old, had a mean VAS score of 75.9. In Luo's sample, participants aged ≥65, mean VAS=81 also reported greater mean VAS score than the same age group of the PanCan, mean VAS=77.9. The closing gap in VAS score between samples may be the result of older PanCan participants smoking less intensely than younger participants, thereby increasing their HRQoL.

Another American general population sample with mean age of 45 also reported greater mean VAS score, 84.32, than the PanCan sample, 76.8.[151] In comparison to a United Kingdom sample, the working, 87.5, and unemployed, 82.0, both reported greater

mean VAS scores than PanCan participants (working=78.0, unemployed=66.9).[152] Current smokers in the United Kingdom sample also reported greater VAS scores than PanCan current smokers (mean United Kingdom sample=80.4 vs. mean PanCan=75.6).

5.3 Study limitations

5.3.1 Lack of longitudinal data

Baseline cross sectional data from the PanCan was analyzed in this investigation. Therefore, conclusions are associations between variables and may not represent causal relationships. Any potential causal relationships between socio-demographics, smoking exposure, morbidities and symptoms, and HRQoL require analysis of longitudinal data.

5.3.2 Normality assumption

The assumption of linear regression modelling that the dependent variable be normally distributed was modestly violated. Even though validating this assumption may not be important to samples as large as the PanCan, a normal distribution is preferred. Bootstrap validation was completed in order to provide added confidence to the conclusions based upon linear regression modelling. Transformations of the dependent VAS variable were also assessed. A cubed transformation was deemed to most closely resemble a normal distribution, however, regression results indicated the linear regression model with the original VAS function (R^2=0.284) explained greater variation in VAS score than the cubed transformation (R^2=0.251).

5.3.3 Predominantly white sample

The PanCan sample is approximately 97% white. Furthermore, the remaining 3% of participants (N=66) were divided between 6 different race/ethnicities. In the interests of study conclusions it would be best to avoid aggregating distinct race/ethnicities into a

single category. This would avoid potential misclassification bias and also makes findings relevant to wider audiences.

5.3.4 Few participants experiencing severe problems among domain level data

Ordinal logistic regression was planned for analysis of EQ-5D domain level data. This would allow for individuals experiencing some and severe problems to be compared with those experiencing no problems. However, few participants report severe problems. As noted in the EQ-5D-3L user guide, this is a common occurrence among population studies where participants are less likely to report severe problems compared to clinical investigations. Participants reporting some or severe problems were combined into a single category, as per EuroQol's recommendation.[124]

5.3.5 Unaccounted independent predictors

Some variables which have been shown to be important predictors of HRQoL in previous investigations were unavailable for inclusion. Particularly, marriage status, number of children, household income, mental health status and social relationships, including discrimination, are measures of interest.

5.4 Study strengths

5.4.1 Study sample and variation of independent predictors

The PanCan is a nationally representative sample of 2,537 former and current smokers who have provided a wide array of data describing their socio-economic status, current and previous smoking exposure, symptoms, morbidities and pulmonary function. Data was carefully collected through in-person interview. Many population-based studies investigating factors associated with HRQoL do not extend analysis to include measures of exposures, symptoms, morbidities and pulmonary function. Furthermore, smoking exposure is commonly limited to a single variable, often measured categorically, such as smoking duration, pack-years or average number of cigarettes smoked per day

throughout smoking duration. Studies which incorporate these detailed independent variables often focus on disease subgroups recruited from clinical settings.

5.4.2 Use of continuous independent variables as opposed to using cut points

Variables such as BMI, average number of cigarettes consumed per day (last 30 days) and FEV_1 percent predicted are often examined categorically rather than as a continuous function.[12] Analyzing continuous variables using cut points may allow for easy interpretation or comparison but the amount of information in analysis decreases, producing results which may not accurately describe true effects. Non-linear relationships may exist between independent and dependent variables. Observing trends in categorical data may suggest a non-linear relationship but plotting and reviewing continuous data allows researchers to clearly observe linear or non-linear relationships between variables. The MFP selection procedure used to identify non-linear relationships is relatively new and should continue to be implemented in further investigations.

5.4.3 Internal validation of independent predictors of VAS score with bootstrapping

Regardless of the recommendation to internally validate regression models, bootstrap analyses are often not completed. Analysis of bootstrap inclusion frequencies enables the assessment of a variable's importance to the regression model. Some variables which were insignificant or borderline significant in the full regression model proved to be important predictors of VAS score when analyzing bootstrap results. In addition to whether a variable is important to a regression model, the linearity, or lack thereof, can be assessed. For non-linear predictors, exponential frequencies can be analyzed to determine the most accurate and parsimonious function to describe the relationship in question. This was particularly useful when analyzing the relationship between FEV_1 and VAS score.

5.5 Future Research

Along with incorporating study strengths, future research can continue to build upon study results in several ways. Analysis of longitudinal data will allow researchers to observe temporal associations which are necessary for further and overall understanding. To strengthen causal conclusions, researchers should also investigate the direct relationships between socio-demographics, smoking and alcohol exposures, morbidities and symptoms. In addition to the variety of independent variables, measures of income, marital status, number of children, social health, mental health and a more racially/ethnically diverse sample should also be analyzed. Stratification was beyond the scope of this project. However, comparing differences within significant independent variable subgroups is a logical proceeding step.

5.6 Clinical and public health implications of the study findings

Policy makers should make known the potential for increasing HRQoL through smoking cessation. As preventative measures, continued effort must be made by policy makers to ensure that students stay in school, are properly educated from a young age on the effects of smoking as well as smoking cessation, and that school policies encourage student involvement in various activities as well as smoke free environments for students and teachers. Healthcare practitioners need to be sensitive to their patient's demographic, recognizing that minorities and the unemployed are more likely to report poor HRQoL than their counterparts. Policy makers in all levels of government should continue to focus on reducing unemployment, not only to improve the economy but also individual well-being. Furthermore, more employers should promote reducing harmful behaviours among employees. This is also beneficial to employers as research indicates current smokers add to employer healthcare costs and are also less productive than non-smokers.[191]

In addition to improving general health, smoking reduction campaigns need to make smokers aware that their general day-to-day HRQoL is also likely to improve if they are able to quit or reduce smoking. Immediate benefits in HRQoL may arise from reducing the number of daily cigarettes consumed and from an eventual reduction in morbidities and symptoms caused by smoking. Caregivers should tailor treatments of current and former smokers to focus on the symptoms (dyspnea, phlegm, chest pain, poor appetite and fatigue) and morbidities (hypercholesterolemia, myocardial infarction, arthritis and pulmonary function) most strongly associated with reduced HRQoL. Even when no morbidities have been diagnosed, symptom experiencing individuals and their medical caregivers should seek to treat symptoms in order to improve HRQoL.

5.7 Conclusion

Multiple socio-demographic, exposure, morbidity and symptom variables are shown to have significant associations with HRQoL. More specifically, being non-white, unemployed, disabled or on sick leave, and having non-normal BMI have significant upstream associations with HRQoL. Older age and lower educational attainment also exhibit negative upstream associations with HRQoL.

Improving educational and employment opportunities, promoting normal BMI and preventing or reducing smoking and excess alcohol consumption are expected to lessen consequent morbidities and symptoms, thus broadly improving HRQoL.

References

1. Walters SJ. *Quality of life outcomes in clinical trials and health-care evaluation : a practical guide to analysis and interpretation.* Chichester: Wiley; 2009.

2. Fayers PM, Machin D. *Quality of life : the assessment, analysis and interpretation of patient-reported outcomes.* 2nd ed. Chichester: Wiley; 2007.

3. McDowell I. *Measuring health : a guide to rating scales and questionnaires.* 3rd ed. New York ; Oxford: Oxford University Press; 2006.

4. Rapley M. *Quality of life research : a critical introduction.* London: SAGE Pubications; 2003.

5. Measuring healthy days. In: Control CfD, ed. Atlanta, Georgia: CDC; 2000.

6. Ismailov RM, Leatherdale ST. Smoking cessation aids and strategies among former smokers in Canada. *Addict Behav.* Mar 2010;35(3):282-285.

7. Health Canada. Canadian Tobacco Use Monitoring Survey (CTUMS). 2011; http://www.hc-sc.gc.ca/hc-ps/tobac-tabac/research-recherche/stat/_ctums-esutc_2010/ann_summary-sommaire-eng.php. Accessed April 7, 2012.

8. Sindhu BS, Shechtman O, Tuckey L. Validity, reliability, and responsiveness of a digital version of the visual analog scale. *Journal of hand therapy : official journal of the American Society of Hand Therapists.* Oct-Dec 2011;24(4):356-363; quiz 364.

9. Curbing the epidemic: governments and the economics of tobacco control. The World Bank. *Tob Control.* Summer 1999;8(2):196-201.

10. United States. Public Health Service. Office of the Surgeon General. *How tobacco smoke causes disease : the biology and behavioral basis for smoking-attributable disease : a report of the Surgeon General.* Rockville, MD. Washington, DC: U.S. Dept. of Health and Human Services, Public Health Service. For sale by the Supt. of Docs., U.S. G.P.O.; 2010.

11. Mirolla M. The cost of chronic disease in Canada. In: Canada CDPAo, ed2004.

12. Vogl M, Wenig CM, Leidl R, Pokhrel S. Smoking and health-related quality of life in English general population: implications for economic evaluations. *BMC Public Health.* Mar 19 2012;12(1):203.

13. Strandberg AY, Strandberg TE, Pitkala K, Salomaa VV, Tilvis RS, Miettinen TA. The effect of smoking in midlife on health-related quality of life in old age: a 26-year prospective study. *Arch Intern Med.* Oct 13 2008;168(18):1968-1974.

14. Mody RR, Smith MJ. Smoking status and health-related quality of life: as findings from the 2001 Behavioral Risk Factor Surveillance System data. *Am J Health Promot.* Mar-Apr 2006;20(4):251-258.

15. McClave AK, Dube SR, Strine TW, Mokdad AH. Associations between health-related quality of life and smoking status among a large sample of U.S. adults. *Prev Med.* Feb 2009;48(2):173-179.

16. Guiterrez-Bedmar M, Segui-Gomez M, Gomez-Gracia E, Bes-Rastrollo M, Martinez-Gonzalez MA. Smoking status, changes in smoking status and health-related quality of life: findings from the SUN ("Seguimiento Universidad de Navarra") cohort. *Int J Environ Res Public Health.* Jan 2009;6(1):310-320.

17. Mulder I, Tijhuis M, Smit HA, Kromhout D. Smoking cessation and quality of life: the effect of amount of smoking and time since quitting. *Prev Med.* Dec 2001;33(6):653-660.

18. Bass PF, 3rd, Wilson JF, Griffith CH. The association of health-related quality of life and age of initiation of smoking. *J Ky Med Assoc.* Mar 2004;102(3):96-101.

19. Piper ME, Kenford S, Fiore MC, Baker TB. Smoking cessation and quality of life: changes in life satisfaction over 3 years following a quit attempt. *Ann Behav Med.* Apr 2012;43(2):262-270.

20. Hays JT, Croghan IT, Baker CL, Cappelleri JC, Bushmakin AG. Changes in health-related quality of life with smoking cessation treatment. *Eur J Public Health.* Apr 2012;22(2):224-229.

21. Heikkinen H, Jallinoja P, Saarni SI, Patja K. The impact of smoking on health-related and overall quality of life: a general population survey in Finland. *Nicotine Tob Res.* Jul 2008;10(7):1199-1207.

22. EuroQol G. EuroQol--a new facility for the measurement of health-related quality of life. *Health Policy.* Dec 1990;16(3):199-208.

23. Mandy van Reenen, Mark Oppe. EQ-5D-3L user guide: Basic information on how to use the EQ-5D-3L instrument. 2015; 5.1:http://www.euroqol.org/fileadmin/user_upload/Documenten/PDF/Folders_Flyers/EQ-5D-3L_UserGuide_2015.pdf.

24. Lam S TM, Tammemagi M, et al. The pan-Canadian early detection of lung cancer study. [Abstract]. *J Thorac Oncol.* 2009;4:S377-378.

25. McWilliams A, Tammemagi MC, Mayo JR, et al. Probability of cancer in pulmonary nodules detected on first screening CT. *N Engl J Med.* Sep 5 2013;369(10):910-919.

26. Garces YI, Yang P, Parkinson J, et al. The relationship between cigarette smoking and quality of life after lung cancer diagnosis. *Chest.* Dec 2004;126(6):1733-1741.

27. Das S, Becker AM, Perakis H, Prosser JD, Kountakis SE. The effects of smoking on short-term quality of life outcomes in sinus surgery. *Laryngoscope.* Dec 2007;117(12):2229-2232.

28. Jia H, Zack MM, Thompson WW, Dube SR. Quality-adjusted life expectancy (QALE) loss due to smoking in the United States. *Qual Life Res.* Feb 18 2012.

29. Robert SA, Cherepanov D, Palta M, Dunham NC, Feeny D, Fryback DG. Socioeconomic status and age variations in health-related quality of life: results from the national health measurement study. *J Gerontol B Psychol Sci Soc Sci.* May 2009;64(3):378-389.

30. Li C, Ford ES, Mokdad AH, Balluz LS, Brown DW, Giles WH. Clustering of cardiovascular disease risk factors and health-related quality of life among US adults. *Value Health.* Jul-Aug 2008;11(4):689-699.

31. Adachi JD, Adami S, Gehlbach S, et al. Impact of prevalent fractures on quality of life: baseline results from the global longitudinal study of osteoporosis in women. *Mayo Clin Proc.* Sep 2010;85(9):806-813.

32. Asakawa K, Senthilselvan A, Feeny D, Johnson J, Rolfson D. Trajectories of health-related quality of life differ by age among adults: results from an eight-year longitudinal study. *J Health Econ.* Jan 2012;31(1):207-218.

33. Voll-Aanerud M, Eagan TM, Wentzel-Larsen T, Gulsvik A, Bakke PS. Respiratory symptoms, COPD severity, and health-related quality of life in a general population sample. *Respir Med.* Mar 2008;102(3):399-406.

34. Dumville JC, Lee AJ, Smith FB, Fowkes FG. The health-related quality of life of people with peripheral arterial disease in the community: the Edinburgh Artery Study. *Br J Gen Pract.* Nov 2004;54(508):826-831.

35. Voll-Aanerud M, Eagan TM, Plana E, et al. Respiratory symptoms in adults are related to impaired quality of life, regardless of asthma and COPD: results from the European community respiratory health survey. *Health Qual Life Outcomes.* 2010;8:107.

36. Soltoft F, Hammer M, Kragh N. The association of body mass index and health-related quality of life in the general population: data from the 2003 Health Survey of England. *Qual Life Res.* Dec 2009;18(10):1293-1299.
37. Stafford M, Soljak M, Pledge V, Mindell J. Socio-economic differences in the health-related quality of life impact of cardiovascular conditions. *Eur J Public Health.* Jun 2012;22(3):301-305.
38. Quek KF. Factors affecting health-related quality of life among patients with lower urinary tract symptoms. *International Journal of Urology.* Dec 2005;12(12):1032-1036.
39. Wang QQ, Wang JH. [Study on factors affecting the quality of life for nurses working in night shifts]. *Zhonghua Lao Dong Wei Sheng Zhi Ye Bing Za Zhi.* May 2011;29(5):353-357.
40. Science USCHCo, Technology Subcommittee on Energy D, Applications. Solar energy research institute and regional solar energy centers. Hearings. *Solar energy research institute and regional solar energy centers. Hearings.*
41. Constitution of the World-Health-Organization. *Public Health Rep.* 1946;61(35):1268-1277.
42. Coons SJ, Rao S, Keininger DL, Hays RD. A comparative review of generic quality-of-life instruments. *Pharmacoeconomics.* Jan 2000;17(1):13-35.
43. Shearer D, Morshed S. Common generic measures of health-related quality of life in injured patients. *Injury.* Mar 2011;42(3):241-247.
44. Jelsness-Jorgensen LP, Ribu L, Bernklev T, Moum BA. Measuring health-related quality of life in non-complicated diabetes patients may be an effective parameter to assess patients at risk of a more serious disease course: a cross-sectional study of two diabetes outpatient groups. *J Clin Nurs.* May 2011;20(9-10):1255-1263.
45. Kramer JA, Curran D, Piccart M, et al. Identification and interpretation of clinical and quality of life prognostic factors for survival and response to treatment in first-line chemotherapy in advanced breast cancer. *Eur J Cancer.* Aug 2000;36(12):1498-1506.
46. Mauer ME, Taphoorn MJ, Bottomley A, et al. Prognostic value of health-related quality-of-life data in predicting survival in patients with anaplastic oligodendrogliomas, from a phase III EORTC brain cancer group study. *J Clin Oncol.* Dec 20 2007;25(36):5731-5737.
47. Zafar SY, Alexander SC, Weinfurt KP, Schulman KA, Abernethy AP. Decision making and quality of life in the treatment of cancer: a review. *Support Care Cancer.* Feb 2009;17(2):117-127.
48. Pizzo E, Pezzoli A, Stockbrugger R, Bracci E, Vagnoni E, Gullini S. Screenee perception and health-related quality of life in colorectal cancer screening: a review. *Value Health.* Jan 2011;14(1):152-159.
49. Kahende JW, Adhikari B, Maurice E, Rock V, Malarcher A. Disparities in health care utilization by smoking status--NHANES 1999-2004. *Int J Environ Res Public Health.* Mar 2009;6(3):1095-1106.
50. Hvidtfeldt UA, Rasmussen S, Gronbaek M, Becker U, Tolstrup JS. Influence of smoking and alcohol consumption on admissions and duration of hospitalization. *Eur J Public Health.* Aug 2010;20(4):376-382.
51. Statistics Canada. Smokers, by sex, provinces and territories 2015; http://www.statcan.gc.ca/tables-tableaux/sum-som/l01/cst01/health74b-eng.htm. Accessed July 1, 2015.
52. Reid JL, Hammond D, Rynard VL, Burkhalter R. *Tobacco use in Canada: Patterns and trends, 2015 edition.* Waterloo, ON: Propel Centre for Population Health Impact, University of Waterloo.

53. Statistics Canada. Smoking, 2013. 2014; http://www.statcan.gc.ca/pub/82-625-x/2014001/article/14025-eng.htm. Accessed July 1, 2015.
54. Giovino GA, Mirza SA, Samet JM, et al. Tobacco use in 3 billion individuals from 16 countries: an analysis of nationally representative cross-sectional household surveys. *Lancet.* Aug 18 2012;380(9842):668-679.
55. Ng M, Freeman MK, Fleming TD, et al. Smoking prevalence and cigarette consumption in 187 countries, 1980-2012. *JAMA.* Jan 8 2014;311(2):183-192.
56. World Health Organization. *WHO report on the global tobacco epidemic, 2011: warning about the dangers of tobacco.* Geneva: World Health Organization; 2011.
57. United States. Public Health Service. Office of the Surgeon General. *Preventing tobacco use among youth and young adults: A report of the surgeon general.* Rockville, MD: U.S. Dept. of Health and Human Services, Public Health Service

For sale by the Supt. of Docs., U.S. G.P.O.; 2012.
58. O'Loughlin J, Karp I, Koulis T, Paradis G, Difranza J. Determinants of first puff and daily cigarette smoking in adolescents. *Am J Epidemiol.* Sep 1 2009;170(5):585-597.
59. Gough B, Fry G, Grogan S, Conner M. Why do young adult smokers continue to smoke despite the health risks? A focus group study. *Psychol Health.* Feb 2009;24(2):203-220.
60. United States. Public Health Service. Office of the Surgeon General. *The health consequences of smoking - nicotine addiction: a report of the Surgeon General.* Rockville, MD

Washington, DC: U.S. Dept. of Health and Human Services, Public Health Service

For sale by the Supt. of Docs., U.S. G.P.O.; 1988.
61. Rabinoff M, Caskey N, Rissling A, Park C. Pharmacological and chemical effects of cigarette additives. *Am J Public Health.* Nov 2007;97(11):1981-1991.
62. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans : Tobacco Smoke and Involuntary Smoking, International Agency for Research on Cancer. *Tobacco smoke and involuntary smoking.* Lyon: International Agency for Research on Cancer; 2004.
63. Rustemeier K, Stabbert R, Haussmann HJ, Roemer E, Carmines EL. Evaluation of the potential effects of ingredients added to cigarettes. Part 2: chemical composition of mainstream smoke. *Food Chem Toxicol.* Jan 2002;40(1):93-104.
64. Glantz SA, Parmley WW. Even a little secondhand smoke is dangerous. *JAMA.* Jul 25 2001;286(4):462-463.
65. National Cancer Institute. Secondhand Smoke and Cancer. 2011; http://www.cancer.gov/cancertopics/factsheet/Tobacco/ETS. Accessed April 7, 2012.
66. Congress of the U.S W. Social behavioral effects of violence on television. Hearing before the Subcommittee on Telecommunications, Consumer protection, and Finance of the Committee on Energy and Commerce. House of Representatives, Ninety-Seventh Congress, First Session (October 21, 1981). *Social behavioral effects of violence on television. Hearing before the Subcommittee on Telecommunications, Consumer protection, and Finance of the Committee on Energy and Commerce. House of Representatives, Ninety-Seventh Congress, First Session (October 21, 1981).*
67. World Health Organization. *WHO Report on the Global Tobacco Epidemic, 2008: the MPOWER package.* Geneva: World Health Organization; 2008.
68. Kenfield SA, Stampfer MJ, Rosner BA, Colditz GA. Smoking and smoking cessation in relation to mortality in women. *JAMA.* May 7 2008;299(17):2037-2047.

69. Doll R, Peto R, Boreham J, Sutherland I. Mortality in relation to smoking: 50 years' observations on male British doctors. *BMJ.* Jun 26 2004;328(7455):1519.

70. Statistics Canada. Ten leading causes of death by selected age groups, by sex, Canada - All ages. 2011.

71. Smoking-attributable mortality, years of potential life lost, and productivity losses--United States, 2000-2004. *MMWR Morb Mortal Wkly Rep.* Nov 14 2008;57(45):1226-1228.

72. National Cancer Institute. Smoking. http://www.cancer.gov/cancertopics/tobacco/smoking. Accessed April 19, 2012, 2012.

73. Health Canada. Smoking and Strokes. 2011; http://www.hc-sc.gc.ca/hc-ps/tobac-tabac/legislation/label-etiquette/stroke-avc-eng.php.

74. Price JF, Mowbray PI, Lee AJ, Rumley A, Lowe GD, Fowkes FG. Relationship between smoking and cardiovascular risk factors in the development of peripheral arterial disease and coronary artery disease: Edinburgh Artery Study. *Eur Heart J.* Mar 1999;20(5):344-353.

75. Lee AJ, Fowkes FG, Carson MN, Leng GC, Allan PL. Smoking, atherosclerosis and risk of abdominal aortic aneurysm. *Eur Heart J.* Apr 1997;18(4):671-676.

76. Chouraki V, Wagner A, Ferrieres J, et al. Smoking habits, waist circumference and coronary artery disease risk relationship: the PRIME study. *Eur J Cardiovasc Prev Rehabil.* Dec 2008;15(6):625-630.

77. United States. Office on Smoking and Health., United States. Public Health Service. Office of the Surgeon General. *Preventing tobacco use among young people : a report of the Surgeon General.* Repr., with corrections, July 1994. ed. Washington, D.C.: U.S. Dept. of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion For sale by the Supt. of Docs., U.S. G.P.O.; 1994.

78. Bolliger CT, Zellweger JP, Danielsson T, et al. Influence of long-term smoking reduction on health risk markers and quality of life. *Nicotine Tob Res.* Nov 2002;4(4):433-439.

79. Forey BA, Thornton AJ, Lee PN. Systematic review with meta-analysis of the epidemiological evidence relating smoking to COPD, chronic bronchitis and emphysema. *BMC Pulm Med.* 2011;11:36.

80. American Lung Association. Health Effects. 2012; http://www.lung.org/stop-smoking/about-smoking/health-effects/. Accessed April 7, 2012.

81. Tian X, Tang Z, Jiang J, et al. Effects of smoking and smoking cessation on life expectancy in an elderly population in Beijing, China, 1992-2000: an 8-year follow-up study. *J Epidemiol.* Sep 5 2011;21(5):376-384.

82. American Cancer Society. Questions about smoking, tobacco and health. 2011; www.cancer.org.

83. Wilson DH, Chittleborough CR, Kirke K, Grant JF, Ruffin RE. The health-related quality of life of male and female heavy smokers. *Soz Praventivmed.* 2004;49(6):406-412.

84. Sohn YJ, Sohn HS, Kwon JW. Gender differences among middle-aged Koreans for health-related quality of life related to metabolic syndrome. *Qual Life Res.* May 2011;20(4):583-592.

85. Li YP, Lin SI, Chen CH. Gender differences in the relationship of social activity and quality of life in community-dwelling Taiwanese elders. *J Women Aging.* Oct 2011;23(4):305-320.

86. Gallicchio L, Hoffman SC, Helzlsouer KJ. The relationship between gender, social support, and health-related quality of life in a community-based study in Washington County, Maryland. *Qual Life Res.* Jun 2007;16(5):777-786.
87. Cherepanov D, Palta M, Fryback DG, Robert SA. Gender differences in health-related quality-of-life are partly explained by sociodemographic and socioeconomic variation between adult men and women in the US: evidence from four US nationally representative data sets. *Qual Life Res.* Oct 2010;19(8):1115-1124.
88. Ross NA, Garner R, Bernier J, et al. Trajectories of health-related quality of life by socio-economic status in a nationally representative Canadian cohort. *J Epidemiol Community Health.* Mar 24 2011.
89. Cherepanov D, Palta M, Fryback DG, Robert SA, Hays RD, Kaplan RM. Gender differences in multiple underlying dimensions of health-related quality of life are associated with sociodemographic and socioeconomic status. *Medical care.* Nov 2011;49(11):1021-1030.
90. Mrus JM, Williams PL, Tsevat J, Cohn SE, Wu AW. Gender differences in health-related quality of life in patients with HIV/AIDS. *Qual Life Res.* Mar 2005;14(2):479-491.
91. Aparicio VA, Ortega FB, Carbonell-Baeza A, et al. Are there gender differences in quality of life and symptomatology between fibromyalgia patients? *Am J Mens Health.* Jul 2012;6(4):314-319.
92. Kendel F, Dunkel A, Muller-Tasch T, et al. Gender differences in health-related quality of life after coronary bypass surgery: results from a 1-year follow-up in propensity-matched men and women. *Psychosom Med.* Apr 2011;73(3):280-285.
93. Belloch A, Perpina M, Martinez-Moragon E, de Diego A, Martinez-Frances M. Gender differences in health-related quality of life among patients with asthma. *J Asthma.* Dec 2003;40(8):945-953.
94. Nicolucci A, Cucinotta D, Squatrito S, et al. Clinical and socio-economic correlates of quality of life and treatment satisfaction in patients with type 2 diabetes. *Nutr Metab Cardiovasc Dis.* Jan 2009;19(1):45-53.
95. Zaninotto P, Falaschetti E, Sacker A. Age trajectories of quality of life among older adults: results from the English Longitudinal Study of Ageing. *Qual Life Res.* Dec 2009;18(10):1301-1309.
96. Sun S, Chen J, Johannesson M, et al. Regional differences in health status in China: population health-related quality of life results from the National Health Services Survey 2008. *Health Place.* Mar 2011;17(2):671-680.
97. Walker AE. Multiple chronic diseases and quality of life: patterns emerging from a large national sample, Australia. *Chronic Illn.* Sep 2007;3(3):202-218.
98. Burstrom K, Johannesson M, Diderichsen F. Health-related quality of life by disease and socio-economic group in the general population in Sweden. *Health Policy.* Jan 2001;55(1):51-69.
99. Klocek M, Kawecka-Jaszcz K. [Quality of life in patients with essential arterial hypertension. Part I: The effect o socio-demographic factors]. *Przegl Lek.* 2003;60(2):92-100.
100. Pappa E, Kontodimopoulos N, Papadopoulos AA, Niakas D. Assessing the socio-economic and demographic impact on health-related quality of life: evidence from Greece. *Int J Public Health.* 2009;54(4):241-249.
101. Hultman B, Hemlin S, Hornquist JO. Quality of life among unemployed and employed people in northern Sweden. Are there any differences? *Work.* 2006;26(1):47-56.
102. Lev-Ran S, Imtiaz S, Taylor BJ, Shield KD, Rehm J, Le Foll B. Gender differences in health-related quality of life among cannabis users: Results from the national epidemiologic

survey on alcohol and related conditions. *Drug Alcohol Depend.* Jun 1 2012;123(1-3):190-200.

103. Fiz J, Duran M, Capella D, Carbonell J, Farre M. Cannabis use in patients with fibromyalgia: effect on symptoms relief and health-related quality of life. *PLoS One.* 2011;6(4):e18440.

104. Kaplan MS, Huguet N, Feeny D, et al. Alcohol use patterns and trajectories of health-related quality of life in middle-aged and older adults: a 14-year population-based study. *J Stud Alcohol Drugs.* Jul 2012;73(4):581-590.

105. Byles J, Young A, Furuya H, Parkinson L. A drink to healthy aging: The association between older women's use of alcohol and their health-related quality of life. *J Am Geriatr Soc.* Sep 2006;54(9):1341-1347.

106. Volk RJ, Cantor SB, Steinbauer JR, Cass AR. Alcohol use disorders, consumption patterns, and health-related quality of life of primary care patients. *Alcohol Clin Exp Res.* Aug 1997;21(5):899-905.

107. World Health Organization. 10 facts on obesity. 2012; http://www.who.int/features/factfiles/obesity/facts/en/index.html. Accessed July 6, 2012.

108. Khanna D, Maranian P, Palta M, et al. Health-related quality of life in adults reporting arthritis: analysis from the National Health Measurement Study. *Qual Life Res.* Sep 2011;20(7):1131-1140.

109. Hill CL, Parsons J, Taylor A, Leach G. Health-related quality of life in a population sample with arthritis. *J Rheumatol.* Sep 1999;26(9):2029-2035.

110. Uhlig T, Loge JH, Kristiansen IS, Kvien TK. Quantification of reduced health-related quality of life in patients with rheumatoid arthritis compared to the general population. *J Rheumatol.* Jun 2007;34(6):1241-1247.

111. de Oliveira Ferreira N, Arthuso M, da Silva R, Pedro AO, Pinto Neto AM, Costa-Paiva L. Quality of life in women with postmenopausal osteoporosis: correlation between QUALEFFO 41 and SF-36. *Maturitas.* Jan 20 2009;62(1):85-90.

112. Picavet HS, Hoeymans N. Physical disability in The Netherlands: prevalence, risk groups and time trends. *Public Health.* Jul 2002;116(4):231-237.

113. Heijdra YF, Pinto-Plata VM, Kenney LA, Rassulo J, Celli BR. Cough and phlegm are important predictors of health status in smokers without COPD. *Chest.* May 2002;121(5):1427-1433.

114. Larsson U, Taft C, Karlsson J, Sullivan M. Gender and age differences in the relative burden of rhinitis and asthma on health-related quality of life--a Swedish population study. *Respir Med.* Jun 2007;101(6):1291-1298.

115. Kalpaklioglu AF, Baccioglu A. Evaluation of quality of life: impact of allergic rhinitis on asthma. *J Investig Allergol Clin Immunol.* 2008;18(3):168-173.

116. Chen H, Cisternas MG, Katz PP, et al. Evaluating quality of life in patients with asthma and rhinitis: English adaptation of the rhinasthma questionnaire. *Ann Allergy Asthma Immunol.* Feb 2011;106(2):110-118 e111.

117. Frost MH, Hoskin TL, Hartmann LC, Degnim AC, Johnson JL, Boughey JC. Contralateral prophylactic mastectomy: long-term consistency of satisfaction and adverse effects and the significance of informed decision-making, quality of life, and personality traits. *Ann Surg Oncol.* Oct 2011;18(11):3110-3116.

118. Jones PW, Brusselle G, Dal Negro RW, et al. Health-related quality of life in patients by COPD severity within primary care in Europe. *Respir Med.* Jan 2011;105(1):57-66.

119. Polinder S, Haagsma JA, Belt E, et al. A systematic review of studies measuring health-related quality of life of general injury populations. *BMC Public Health.* 2010;10:783.
120. Chung HY, Machado P, van der Heijde D, D'Agostino MA, Dougados M. Smokers in early axial spondyloarthritis have earlier disease onset, more disease activity, inflammation and damage, and poorer function and health-related quality of life: results from the DESIR cohort. *Ann Rheum Dis.* Jun 2012;71(6):809-816.
121. Geijer RM, Sachs AP, Verheij TJ, Kerstjens HA, Kuyvenhoven MM, Hoes AW. Quality of life in smokers: focus on functional limitations rather than on lung function? *Br J Gen Pract.* Jun 2007;57(539):477-482.
122. Tammemagi MC, Freedman MT, Pinsky PF, et al. Prediction of true positive lung cancers in individuals with abnormal suspicious chest radiographs: a prostate, lung, colorectal, and ovarian cancer screening trial study. *J Thorac Oncol.* Jun 2009;4(6):710-721.
123. Tammemagi CM, Pinsky PF, Caporaso NE, et al. Lung cancer risk prediction: Prostate, Lung, Colorectal And Ovarian Cancer Screening Trial models and validation. *J Natl Cancer Inst.* Jul 6 2011;103(13):1058-1068.
124. Mandy van Reenen, Mark Oppe. EQ-5D-3L User guide: Basic information on how to use the EQ-5D-3L instrument. 2015; 5.1:http://www.euroqol.org/fileadmin/user_upload/Documenten/PDF/Folders_Flyers/EQ-5D-3L_UserGuide_2015.pdf, 2015.
125. Johnson JA, Pickard AS. Comparison of the EQ-5D and SF-12 health surveys in a general population survey in Alberta, Canada. *Medical care.* Jan 2000;38(1):115-121.
126. Johnson JA, Coons SJ. Comparison of the EQ-5D and SF-12 in an adult US sample. *Qual Life Res.* Feb 1998;7(2):155-166.
127. Luo N, Johnson JA, Shaw JW, Feeny D, Coons SJ. Self-reported health status of the general adult U.S. population as assessed by the EQ-5D and Health Utilities Index. *Medical care.* Nov 2005;43(11):1078-1086.
128. Kontodimopoulos N, Pappa E, Niakas D, Yfantopoulos J, Dimitrakaki C, Tountas Y. Validity of the EuroQoL (EQ-5D) instrument in a Greek general population. *Value Health.* Dec 2008;11(7):1162-1169.
129. Horowitz E, Abadi-Korek I, Shani M, Shemer J. EQ-5D as a generic measure of health-related quality of life in Israel: reliability, validity and responsiveness. *Isr Med Assoc J.* Dec 2010;12(12):715-720.
130. van Agt HM, Essink-Bot ML, Krabbe PF, Bonsel GJ. Test-retest reliability of health state valuations collected with the EuroQol questionnaire. *Social science & medicine.* Dec 1994;39(11):1537-1544.
131. Pickard AS, Wilke C, Jung E, Patel S, Stavem K, Lee TA. Use of a preference-based measure of health (EQ-5D) in COPD and asthma. *Respir Med.* Apr 2008;102(4):519-536.
132. Janssen MF, Lubetkin EI, Sekhobo JP, Pickard AS. The use of the EQ-5D preference-based health status measure in adults with Type 2 diabetes mellitus. *Diabet Med.* Apr 2011;28(4):395-413.
133. Dyer MT, Goldsmith KA, Sharples LS, Buxton MJ. A review of health utilities using the EQ-5D in studies of cardiovascular disease. *Health Qual Life Outcomes.* 2010;8:13.
134. Brazier JE, Walters SJ, Nicholl JP, Kohler B. Using the SF-36 and Euroqol on an elderly population. *Qual Life Res.* Apr 1996;5(2):195-204.
135. Tammemagi MC, Lam SC, McWilliams AM, Sin DD. Incremental value of pulmonary function and sputum DNA image cytometry in lung cancer risk prediction. *Cancer Prev Res (Phila).* Apr 2011;4(4):552-561.

136. Standardization of spirometry--1987 update. Statement of the American Thoracic Society. *Am Rev Respir Dis.* Nov 1987;136(5):1285-1298.
137. *Stata Statistical Software: Release 13* [computer program]. College Station, TX: StataCorp LP.; 2013.
138. Cuzick J. A Wilcoxon-type test for trend. *Statistics in medicine.* Jan-Mar 1985;4(1):87-90.
139. Carsten F. Dormann, Jane Elith, Sven Bacher, et al. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography.* 2013;36(1):27 - 46.
140. Desai M, Begg MD. A comparison of regression approaches for analyzing clustered data. *Am J Public Health.* Aug 2008;98(8):1425-1429.
141. Mickey RM, Greenland S. The impact of confounder selection criteria on effect estimation. *Am J Epidemiol.* Jan 1989;129(1):125-137.
142. Sauerbrei W, Meier-Hirmer C, Benner A, Royston P. Multivariable regression model building by using fractional polynomials: Description of SAS, STATA and R programs. *Comput Stat Data An.* Aug 1 2006;50(12):3464-3485.
143. Vittinghoff E. *Regression methods in biostatistics : linear, logistic, survival, and repeated measures models.* 2nd ed. New York: Springer; 2012.
144. Casson RJ, Farmer LD. Understanding and checking the assumptions of linear regression: a primer for medical researchers. *Clinical & experimental ophthalmology.* Aug 2014;42(6):590-596.
145. Royston P, Sauerbrei W. *Multivariable model-building : a pragmatic approach to regression analysis based on fractional polynomials for modelling continuous variables.* Chichester, England ; Hoboken, NJ: John Wiley; 2008.
146. Allison PD. *Multiple regression : a primer.* Thousand Oaks, Calif. ; London: Pine Forge Press; 1999.
147. Fox J, Fox JAralm, related m. *Applied regression analysis and generalized linear models.* 2nd ed. ed. London: SAGE; 2008.
148. Abreu MN, Siqueira AL, Cardoso CS, Caiaffa WT. Ordinal logistic regression models: application in quality of life studies. *Cadernos de saude publica / Ministerio da Saude, Fundacao Oswaldo Cruz, Escola Nacional de Saude Publica.* 2008;24 Suppl 4:s581-591.
149. Burns R, Burns R. *Business research methods and statistics using SPSS.* Los Angeles ; London: SAGE; 2008.
150. Hensher DA, Stopher PR. *Behavioural travel modelling.* London: Croom Helm; 1979.
151. Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. *Medical care.* Mar 2005;43(3):203-220.
152. Kind P, Dolan P, Gudex C, Williams A. Variations in population health status: results from a United Kingdom national questionnaire survey. *BMJ.* Mar 7 1998;316(7133):736-741.
153. StataCorp. Stata user's guide: Release 13. 2013; http://www.stata.com/manuals13/u20.pdf#u20.21Obtainingrobustvarianceestimates.
154. Lumley T, Diehr P, Emerson S, Chen L. The importance of the normality assumption in large public health data sets. *Annual review of public health.* 2002;23:151-169.
155. Peto J. That the effects of smoking should be measured in pack-years: misconceptions 4. *British journal of cancer.* Jul 24 2012;107(3):406-407.
156. Jason W. Osborne & Amy Overbay. The power of outliers (any why researchers should always check for them). *Practical Assessment, Research & Evaluation.* 2004;9(6).
157. Royston P, Sauerbrei W. Improving the robustness of fractional polynomial models by preliminary covariate transformation: A pragmatic approach. *Computational Statistics and Data Analysis.* 2007;51(9):4240 - 4253.

158. Barnett V, Lewis T. *Outliers in statistical data.* 3rd ed ed. Chichester: Wiley; 1994.
159. Wilson D, Parsons J, Wakefield M. The health-related quality-of-life of never smokers, ex-smokers, and light, moderate, and heavy smokers. *Prev Med.* Sep 1999;29(3):139-144.
160. Laaksonen M, Rahkonen O, Martikainen P, Karvonen S, Lahelma E. Smoking and SF-36 health functioning. *Prev Med.* Mar 2006;42(3):206-209.
161. Sales MP, Oliveira MI, Mattos IM, Viana CM, Pereira ED. The impact of smoking cessation on patient quality of life. *Jornal brasileiro de pneumologia : publicacao oficial da Sociedade Brasileira de Pneumologia e Tisilogia.* May 2009;35(5):436-441.
162. Tsiligianni I, Kocks J, Tzanakis N, Siafakas N, van der Molen T. Factors that influence disease-specific quality of life or health status in patients with COPD: a review and meta-analysis of Pearson correlations. *Primary care respiratory journal : journal of the General Practice Airways Group.* Sep 2011;20(3):257-268.
163. Ross NA, Garner R, Bernier J, et al. Trajectories of health-related quality of life by socio-economic status in a nationally representative Canadian cohort. *J Epidemiol Community Health.* Jul 2012;66(7):593-598.
164. Murtagh KN, Hubert HB. Gender differences in physical disability among an elderly cohort. *Am J Public Health.* Aug 2004;94(8):1406-1411.
165. Coste J, Quinquis L, D'Almeida S, Audureau E. Smoking and health-related quality of life in the general population. Independent relationships and large differences according to patterns and quantity of smoking and to gender. *PLoS One.* 2014;9(3):e91562.
166. Vogl M, Wenig CM, Leidl R, Pokhrel S. Smoking and health-related quality of life in English general population: implications for economic evaluations. *BMC Public Health.* 2012;12:203.
167. Zack MM, Centers for Disease C, Prevention. Health-related quality of life - United States, 2006 and 2010. *Morbidity and mortality weekly report. Surveillance summaries.* Nov 22 2013;62 Suppl 3:105-111.
168. Pereira CCdA, Palta M, Mullahy J. Health domains and race in generic preference-based health-related quality of life instruments in the United States literature. *Revista Brasileira de Estudos de População.* 2010;27:425-437.
169. Maralani V. Understanding the links between education and smoking. *Social science research.* Nov 2014;48:20-34.
170. Merchant JA, Kelly KM, Burmeister LF, et al. Employment status matters: a statewide survey of quality-of-life, prevention behaviors, and absenteeism and presenteeism. *Journal of occupational and environmental medicine / American College of Occupational and Environmental Medicine.* Jul 2014;56(7):686-698.
171. Extremera N, Rey L. Health-related quality of life and cognitive emotion regulation strategies in the unemployed: a cross-sectional survey. *Health Qual Life Outcomes.* 2014;12:172.
172. Faheem M, Qureshi S, Ali J, et al. Does BMI affect cholesterol, sugar, and blood pressure in general population? *Journal of Ayub Medical College, Abbottabad : JAMC.* Oct-Dec 2010;22(4):74-77.
173. Yatsuya H, Toyoshima H, Yamagishi K, et al. Body mass index and risk of stroke and myocardial infarction in a relatively lean population: meta-analysis of 16 Japanese cohorts using individual data. *Circulation. Cardiovascular quality and outcomes.* Sep 2010;3(5):498-505.

174. Derraik JG, de Bock M, Hofman PL, Cutfield WS. Increasing BMI is associated with a progressive reduction in physical quality of life among overweight middle-aged men. *Scientific reports.* 2014;4:3677.
175. Anandacoomarasamy A, Caterson ID, Leibman S, et al. Influence of BMI on health-related quality of life: comparison between an obese adult cohort and age-matched population norms. *Obesity.* Nov 2009;17(11):2114-2118.
176. Yan LL, Daviglus ML, Liu K, et al. BMI and health-related quality of life in adults 65 years and older. *Obesity research.* Jan 2004;12(1):69-76.
177. Ostbye T, Taylor DH, Jr., Krause KM, Van Scoyoc L. The role of smoking and other modifiable lifestyle risk factors in maintaining and restoring lower body mobility in middle-aged and older Americans: results from the HRS and AHEAD. Health and Retirement Study. Asset and Health Dynamics Among the Oldest Old. *J Am Geriatr Soc.* Apr 2002;50(4):691-699.
178. Nova E, Baccan GC, Veses A, Zapatera B, Marcos A. Potential health benefits of moderate alcohol consumption: current perspectives in research. *The Proceedings of the Nutrition Society.* May 2012;71(2):307-315.
179. Huang C, Zhan J, Liu YJ, Li DJ, Wang SQ, He QQ. Association between alcohol consumption and risk of cardiovascular disease and all-cause mortality in patients with hypertension: a meta-analysis of prospective cohort studies. *Mayo Clin Proc.* Sep 2014;89(9):1201-1210.
180. Blinderman CD, Homel P, Billings JA, Tennstedt S, Portenoy RK. Symptom distress and quality of life in patients with advanced chronic obstructive pulmonary disease. *Journal of pain and symptom management.* Jul 2009;38(1):115-123.
181. Leander M, Cronqvist A, Janson C, Uddenfeldt M, Rask-Andersen A. Non-respiratory symptoms and well-being in asthmatics from a general population sample. *J Asthma.* Aug 2009;46(6):552-559.
182. Willemse BW, Postma DS, Timens W, ten Hacken NH. The impact of smoking cessation on respiratory symptoms, lung function, airway hyperresponsiveness and inflammation. *The European respiratory journal.* Mar 2004;23(3):464-476.
183. Simmons MS, Connett JE, Nides MA, et al. Smoking reduction and the rate of decline in FEV(1): results from the Lung Health Study. *The European respiratory journal.* Jun 2005;25(6):1011-1017.
184. Nishitsuji M, Fujimura M, Oribe Y, et al. [Influence of smoking on longitudinal decline in one-second forced expiratory volume in clinically healthy Japanese men: a longitudinal study]. *Nihon Kokyuki Gakkai zasshi = the journal of the Japanese Respiratory Society.* Oct 2003;41(10):691-695.
185. Khan A, Shabbir K, Ansari JK, Zia N. Comparison of forced expiratory volume in one second (FEV1) among asymptomatic smokers and non-smokers. *JPMA. The Journal of the Pakistan Medical Association.* Mar 2010;60(3):209-213.
186. Stojkovic J, Stevcevska G. Quality of life, forced expiratory volume in one second and body mass index in patients with COPD, during therapy for controlling the disease. *Prilozi / Makedonska akademija na naukite i umetnostite, Oddelenie za biloski i medicinski nauki = Contributions / Macedonian Academy of Sciences and Arts, Section of Biological and Medical Sciences.* Jul 2009;30(1):129-142.
187. Wheaton AG, Ford ES, Thompson WW, Greenlund KJ, Presley-Cantrell LR, Croft JB. Pulmonary function, chronic respiratory symptoms, and health-related quality of life among adults in the United States--National Health and Nutrition Examination Survey 2007-2010. *BMC Public Health.* 2013;13:854.

188. Xie G, Li Y, Shi P, Zhou B, Zhang P, Wu Y. Baseline pulmonary function and quality of life 9 years later in a middle-aged Chinese population. *Chest.* Oct 2005;128(4):2448-2457.
189. Tomioka H, Imanaka K, Hashimoto K, Iwasaki H. Health-related quality of life in patients with idiopathic pulmonary fibrosis--cross-sectional and longitudinal study. *Internal medicine.* 2007;46(18):1533-1542.
190. Ibn Yacoub Y, Amine B, Laatiris A, Hajjaj-Hassouni N. Health-related quality of life in Moroccan patients with rheumatoid arthritis. *Clinical rheumatology.* Oct 2012;31(10):1471-1477.
191. Sherman BW, Lynch WD. The relationship between smoking and health care, workers' compensation, and productivity costs for a large employer. *Journal of occupational and environmental medicine / American College of Occupational and Environmental Medicine.* Aug 2013;55(8):879-884.

Appendices

Appendix 1a. Bootstrap inclusion fractions for all variables included in regression modelling, 1000 replications, N=2397

	1st Polynomial			2nd polynomial		
Independent variable	**% Inclusion**	**Exponential Mean**	**Mean β**	**% Inclusion**	**Exponential Mean**	**Mean β**
Age	30.2	1.01	-114	6.3	2.13	247
Gender (male vs. female)	77.7	1	-1.96			
Race/ethnicity (non-white vs. white)	60.2	1	-4.69			
Education	8.8	-1.07	-5.17	2.8	-.589	12.2
BMI	99.9	-.365	-39.2	52.2	-1.92	225
Retired	100.0	1	-1.67			
Unemployed	100.0	1	-6.56			
Disabled	100.0	1	-11.2			
Sick leave	100.0	1	-14.4			
Other	100.0	1	.645			
Familial cancer history	44.2	1	1.56			
Cigarettes per day (smoking period)	30.5	.033	-.149	7.7	1.14	1.06
Cigarettes per day (last 30 days)	73.6	.995	-.778	11.0	1.3	.289
Smoking duration	40.4	2.06	-.356	22.9	3.0	-.345
Time since quit	30.1	.806	2.7	2.9	-.431	.392
Average alcohol per week	65.4	.696	-3.68	4.4	-.523	1.1
Average marijuana use	17.8	.183	-4.27	7.8	1.1	-5.24
Dyspnea	100	1	-3.1			
Cough	14.4	1	1.45			
Phlegm	55.6	1	-1.69			
Hemoptysis	22.9	1	4.79			
Hoarseness	49.4	1	-1.76			
Wheeze	32.2	1	-1.67			
Chest pain	91.6	1	-2.55			
Poor appetite	96.6	1	-4.4			
Fatigued	100.0	1	-5.77			
Stroke	9.2	1	-2.67			
CHD	31.1	1	-3.31			
Angina	10.4	1	-2.37			
Myocardial infarction	51.7	1	-3.69			
Hypercholesterolemia	46.8	1	-1.55			
Hypertension	44.9	1	-1.58			
Congestive heart failure	27.9	1	7.69			
PVD	46.3	1	-3.87			
Asthma	22.2	1	-2.27			
COPD	7.9	1	-.619			
Pulmonary fibrosis	58.2	1	-26.8			
Pneumonia	44.5	1	-1.6			

Appendix 1b. Bootstrap inclusion fractions for all variables included in regression modelling, 1000 replications, N=2397

	1st Polynomial			2nd polynomial		
Independent variable	**% Inclusion**	**Exponential Mean**	**Mean β**	**% Inclusion**	**Exponential Mean**	**Mean β**
Liver disease	7.2	1	-.645			
Kidney disease	7.3	1	-.222			
Osteoporosis	27.7	1	-2.02			
Arthritis	92.0	1	-1.97			
Weight loss	9.4	1	1.5			
Allergy	6.4	1	-1.36			
Diabetes	27.0	1	-2.39			
Anemia	10.0	1	-1.89			
Blood disease	25.8	1	-4.77			
$FEV_{1,}$ % predicted	99.5	-.492	.652	2.8	2.86	-3.51

Abbreviations: β, beta-coefficient; BMI, Body Mass Index; CHD, coronary heart disease; PVD, peripheral vascular disease; COPD, Chronic obstructive pulmonary disease; $FEV_{1,}$ Forced Expiratory Volume in one second

Appendix 2. Frequency of first and second polynomial exponent combinations for BMI and VAS score, 1000 bootstrap replications

	2nd polynomial exponent							
1st polynomial exponent	**-2**	**-1**	**-0.5**	**0**	**0.5**	**2**	**3**	**Insignificant**
-2	500	11	2	4	1	2	2	0
1	0	0	0	0	0	0	0	376
3	0	0	0	0	0	0	0	101
Insignificant	0	0	0	0	0	0	0	1

Appendix 3. Frequency of first and second polynomial exponent combinations for daily cigarettes smoked in last 30 days and VAS score, 1000 bootstrap replications

	2nd polynomial exponent							
1st polynomial exponent	**-1**	**-0.5**	**0**	**0.5**	**1**	**2**	**3**	**Insignificant**
-2	1	0	2	0	2	0	0	0
-1	4	9	3	1	0	0	0	0
-0.5	0	9	6	0	0	0	0	2
0	0	0	8	5	0	0	0	3
0.5	0	0	0	1	0	0	1	1
1	0	0	0	0	0	6	1	607
2	0	0	0	0	0	20	16	6
3	0	0	0	0	0	0	15	8
Insignificant	0	0	0	0	0	0	0	264

Appendix 4. Frequency of first and second polynomial exponent combinations for weekly alcohol exposure and VAS score, 1000 bootstrap replications

	2nd polynomial exponent							
1st polynomial exponent	**-2**	**-1**	**-0.5**	**0**	**1**	**2**	**3**	**Insignificant**
-2	17	9	4	3	0	0	0	3
-1	0	1	0	0	0	0	0	3
-0.5	0	0	0	0	0	1	0	6
0	0	0	0	0	0	0	0	54
0.5	0	0	0	0	0	0	0	55
1	0	0	0	0	2	1	0	489
2	0	0	0	0	0	1	2	0
3	0	0	0	0	0	0	3	0
Insignificant	0	0	0	0	0	0	0	347

Appendix 5. Frequency of first and second polynomial exponent combinations for FEV_1, percent predicted and VAS score, 1000 bootstrap replications

	2nd polynomial exponent			
1st polynomial exponent	**1**	**2**	**3**	**Insignificant**
-2	0	0	22	229
-1	0	0	0	342
-0.5	0	0	0	35
0	1	0	0	0
1	0	0	1	361
2	0	2	0	0
3	0	0	2	0
Insignificant	0	0	0	5

Appendix 6. Transformations of VAS score