Starkie, Helen Jane (2010) *Health economic aspects in the management of Chronic Obstructive Pulmonary Disease.* PhD thesis.

http://theses.gla.ac.uk/2154/

Health Economic aspects in the management of Chronic Obstructive Pulmonary Disease

Helen J Starkie BA (Hons) MSc

Submitted in fulfillment of the requirements for the Degree of Doctor of Philosophy
Department of Public Health and Health Policy
Faculty of Medicine
University of Glasgow

2010

"From a physician standpoint the Holy Grail of COPD disease modification is to halt, or at least slow down, the rate of decline of FEV_1. However, from a patient perspective the Holy Grail is simply to be able to breathe easier."(1)

Abstract

The broad aim of this thesis on 'Health Economic aspects in the management of Chronic Obstructive Pulmonary Disease' (COPD) was to study the natural history of the disease in order to inform the conceptualisation and development of a new economic model.

Existing economic evaluations for COPD were critiqued and information on the natural history of the disease gathered though literature searches and analyses of two large datasets, a COPD randomised controlled trial called TORCH and a general population observational dataset called the Renfrew/Paisley (MIDSPAN) study. Particular attention was paid to identifying the COPD population using different diagnostic criteria. The elicitation of utility estimates under a number of circumstances was considered. A regression based prediction model was conceptualised and developed.

Significant contributions of this thesis include, but are not limited to: a NICE COPD cohort were identified who were found to be at higher risk of all-cause and COPD mortality than a GOLD defined cohort; a mapping equation was successfully developed that predicts the EQ-5D from the SGRQ; and an entirely new concept for modelling COPD was developed that uses a series of regression equations to predict cost and effect based on lung function, symptoms and exacerbations and weighted by survival probability in order to generate a model with one arm representing current treatment and a second arm representing a comparator treatment.

The thesis successfully combined information gathered throughout the period of research on the natural history of COPD with treatment effects in a novel way in order to conceptualise and develop a new economic model for COPD.

Publications and Presented Work

The following publications and presentations are as a result of the PhD research:

Starkie HJ, Briggs AH, Chambers MG. Pharmacoeconomics in COPD: lessons for the future. The International Journal of COPD. 2008;3(1):71-88

Briggs AH, Starkie HJ, Wu O. Health Economics in Asthma and COPD. In: Barnes PJ, Drazen JM, Rennard SI, Thomson NC, editors. Asthma and COPD: Basic Mechanisms and Clinical Management. Elsevier Ltd, Academic Press, 2009: 751-760.

OHE Commission on NHS Productivity: outcome measures for the assessment of treatment results in Chronic Obstructive Pulmonary Disease. Published at www.ohe.org. Presented at the OHE in October 2007 and the European Conference on Health Economics in July 2008.

Starkie HJ, Briggs AH, Hart CL et al. GOLD versus NICE diagnostic criteria for Chronic Obstructive Pulmonary Disease: impact on disease prevalence and mortality risk within the Renfrew/Paisley study. Thorax 2008;63(Suppl VII):A65. Presented at the British Thoracic Society conference, December 2008.

Starkie HJ, Briggs AH, Hart CL et al. COPD mortality risk: applying GOLD and NICE criteria within a UK general population study. Responding to reviewers comments.

Starkie HJ, Briggs AH, Chambers,MG. Predicting EQ-5D values using the SGRQ. In draft form and presented at the Health Economists Study Group (HESG) in January 2008 and the Society for Medical Decision making conference in October 2008.

Starkie HJ, Briggs AH, Hart CL et al. Natural history data in economic modelling for COPD: a focus on hospitalisation. Presented at HESG, January 2009.

Acknowledgements

Abbreviations

AUC	Area under the curve
BMI	Body mass index
COPD	Chronic Obstructive Pulmonary Disease
EQ-5D	EuroQol 5-D
FEV_1	Forced expiratory volume in 1 second
FVC	Forced vital capacity
GLM	Generalised linear model
GOLD	Global Initiative for chronic obstructive Lung Disease
GRO	General Register Office
HR	Hazard ratio
HRQoL	Health related quality of life
ICD	International classification of disease
ICER	Incremental cost effectiveness ratio
MRC	Medical Research Council
NHS	National Health Service
NICE	National Institute for Health and Clinical Excellence
OLS	Ordinary least squares
PSA	Probabilistic sensitivity analysis
QALY	Quality adjusted life year
RCT	Randomised controlled trial
RMSE	Root mean squared error
SGRQ	St George's respiratory questionnaire
SMR	Scottish morbidity record
TTO	Time trade off
UD	Utility decrement

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

The broad aim of this thesis on Health Economic aspects in the management of chronic obstructive pulmonary disease (COPD) is to gain an in-depth understanding of the respiratory disease, COPD though literature searches and analyses of a number of datasets containing COPD patients, with the ultimate goal of using this information to inform the design and build of an economic model that represents the disease over time, so that the costs and effects of treatment on a defined COPD population can be assessed.

COPD is a chronic condition with a high prevalence in the older population. As the disease worsens over time, there is an increasing, detrimental impact upon health related quality of life and an increase in the cost burden to the health care provider. Advice on the treatment and management of COPD has been available since the 1980s and has evolved over time. There are currently a number of published guidelines advising on best practice.

The UK National Health Service (NHS) provides health care free at the point of delivery, based on clinical need, with a general aim of maximising population health, subject to the budget constraint for health care. With the emergence of high cost pharmaceuticals for the treatment of COPD over recent years, it is important to know whether these treatments demonstrate value for money. That is, do the benefits (or effectiveness) of the new treatment warrant the additional cost, compared to current treatment? The National Institute for Health and Clinical Excellence (NICE) makes reimbursement decisions about treatments on behalf of the UK NHS based on all available and relevant evidence. To date, no treatment for COPD has been analysed by NICE in terms of its cost effectiveness but new treatments for the disease are emerging and a formal appraisal is expected in the near future.

There are a number of published studies that demonstrate the cost effectiveness of current treatments for COPD, but there are inconsistencies between the methodologies employed and comparisons between studies are therefore difficult. There is a need for a common approach, which can be used to assess different treatments on the same platform.

Economic modelling of disease is a response to the need for making informed decisions about the cost effectiveness of treatment based on all the available and relevant evidence. Economic models are frequently developed based on existing models and structures, and tend to focus on the mechanism of action of the drug of interest, and in doing so important aspects of the natural history of the disease may be ignored. In developing an understanding of the epidemiology of disease, in addition to the synthesis of results from literature searches, data from clinical trials and observational datasets are useful to explore. Clinical trials generally contain information on the effect of treatment over a short period of time and often contain detailed data relating to the health related quality of life of the diseased population and on costs arising to the health care provider. Observational studies on the other hand can provide information on the natural history of disease over a longer period of time but in less detail than clinical trials, and may provide an opportunity for raising and answering questions on the natural history of the disease, such as how the disease progresses over time.

Economic models represent a simplification of reality so that only key events or aspects, which impact upon health related quality of life and/or costs are captured. For example, people with COPD tend to have the disease for a long duration of time, often running into several decades. During that time, some events may occur which have a large detrimental effect on health related quality of life and/or may increase health care costs, such as exacerbations requiring hospitalisation. It is important to represent such events within an economic model. Some other events that occur during the person's life may have little or

no bearing on health related quality of life and/or health care costs and are therefore irrelevant for the purpose of economic modelling and should not be included.

It is of interest to identify 'who' the COPD population are and within that, those at differential risk of disease related events, using data containing information on the natural history of COPD. Published guidelines on COPD recommend different ways of diagnosing the disease with potentially large differences in prevalence and ultimately, the identification of 'who' the population in the model are. Different populations are likely to lead to different cost effectiveness estimates. Therefore in developing an economic model, a sound understanding of the disease is required. Frequently economic models use specific groups of people that move through the model, such as those with mild, moderate and severe disease and have limited flexibility to incorporate further heterogeneity at the patient level. However some people are at higher risk of certain events than others. For example, people with a history of heavy smoking are more likely to be at risk of mortality than non-smokers or those with a light smoking history and people with respiratory symptoms may have different outcomes than those without respiratory symptoms. Age is also a consideration, for as people age they are more likely to die from any cause than younger counterparts and people who are older may be at increased risk of respiratory events such as exacerbations. Because of differences in risk between potential study populations, the cost effectiveness of treatment may differ by subgroup, and the ability of a model to identify subgroups in which treatment might be particularly cost effective would be a useful feature of a new economic model.

The thesis is split into six main chapters: Chapters 2 to 7. Chapter 2 provides background information on COPD, Chapter 3 introduces the subject of health economics and describes in detail key concepts of economic evaluation. Within Chapter 4, a literature review on the economic evaluations published for COPD is described, with reference to key concepts

within economic evaluation. Chapter 5 presents results of several analyses conducted on an observational longitudinal dataset, and Chapter 6 contains results from analyses using data from a COPD clinical trial. Based on all the information gathered on the disease within the preceding chapters, in Chapter 7, an economic model for COPD is conceptualised and developed, incorporating further analyses on both the longitudinal and clinical trial data. The rest of this introductory chapter presents information on the specific chapters.

The purpose of Chapter 2 is to provide background details on COPD and begins by introducing the disease and describing the associated health and financial burdens that arise, for: the patient; to the health care system; and to society at large. The clinical management of the disease is discussed and the changing pattern of management and treatment for COPD and the associated costs to the NHS are detailed. A section that describes the current and historical COPD diagnostic criteria and a review of how these criteria affect prevalence estimates is presented.

Within Chapter 3, the need for economic analysis in health care is discussed before exploring the branch of health economics, economic evaluation, which is focussed specifically on the evaluation of different treatments and can be used to inform health care providers about the relative value of competing courses of treatment action. In developing an economic evaluation, it is important to fully understand the disease under study, including: how the disease presents itself, how the disease changes over time and how treatment affects the disease. The chapter on health economics presents information on the key elements required for the economic evaluation of any treatment or intervention and these key concepts are drawn upon throughout the thesis. Topics explored include: a general overview of economic evaluation, the types of evaluation available, different structural models that can be used to develop an evaluation, costs and cost data, outcome

measures including utility and issues around uncertainty. Specific interest is paid to outcome measures and in the measurement of utility.

Chapter 4 contains a literature review of the published economic evaluations for COPD. The focus is on treatments where competing interventions exist for COPD patients and includes pharmacological and surgical treatment. The review is split into two sections: the first section focuses on economic evaluations that have been conducted either alongside a clinical trial or an observational study and is structured around the key concepts of economic evaluation identified in Chapter 3: perspective, patient group, comparators, outcome measures, extrapolation, the results of the evaluations and the handling of uncertainty. The second section examines in further detail those economic evaluations that develop an economic model and include, in addition to those key concepts mentioned above, a description of the model structure, design of the study, sources of input data and assumptions used within the model.

In developing economic models for COPD, often researchers have searched the literature or consulted clinical experts for data on the natural history of disease, rather than directly exploring primary data sources. The analyses within Chapter 5 are aimed at examining the natural history of COPD in a British population and makes use of the Renfrew/Paisley (MIDSPAN) dataset, a Scottish prospective cohort followed since the early 1970s, with ongoing linked hospitalisation and mortality records, with a view to using the information gathered to inform an economic model for the disease. The analyses conducted on the dataset are split into four sections, each containing section specific methods, results and a summary or discussion. Within the first section, summary statistics of the MIDSPAN dataset are presented, by COPD disease severity, including: mortality rates, survival curves and the major causes of mortality. The second analysis replicates and updates a 1996 study conducted by Hole et al, who published a paper on the link between reduced lung function

and subsequent mortality. The third analysis examines the diagnostic criteria for COPD and challenges the assumption that COPD is best diagnosed using lung function alone. The effect of including a risk factor such as smoking history and respiratory symptoms in the diagnostic criteria for COPD is investigated. The final analysis investigates the number of hospital admissions and length of stay in hospital, before determining hospitalisation rate by disease severity for the Renfrew/Paisley population.

Within economic evaluations of health technologies, decision makers use quality adjusted life years (QALYs) in order to compare treatments in different disease areas on the same scale. Because of the importance of utility and ultimately the QALY as an outcome measure, analyses within Chapter 6 are focused on utilities and on deriving values for the QALY from a large, multinational, multi-site randomised controlled trial called the Towards a Revolution in COPD Health (TORCH) trial. Summary utility values are derived from the TORCH trial by disease severity and QALYs by treatment group are calculated. Multivariate regression models are developed to predict utility scores and to predict QALYs. The final section of the chapter develops a mapping equation, which generates EQ-5D utility scores, for use when a utility-based measure is not collected within a clinical trial, and from which QALYs can be estimated.

Economic models combine information on the natural history of a disease, specifically relating to components that quantify the major drivers of cost and effect within the disease, with the effect of treatment, so that decision makers can make reimbursement decisions based on all the available and relevant information. Chapter 7 presents the development and results of a new concept in economic modelling for COPD. In the first section of the chapter, a conceptual framework for an economic model of COPD is developed in which key components affecting cost and effect for COPD patients are considered. The rationale for selecting the chosen economic structure, a regression based model, is explained. The

second section presents details relating to the development of regression equations that are used to construct the economic model for COPD; equations predicting individually: lung function, exacerbations, symptoms, EQ-5D utility, cost and survival. The development of each regression equation was considered with attention given to the dataset within which the equations were developed, choice of explanatory variables and the type of regression analysis used. When combined, these regression equations form an economic model that represents current treatment. The final section of Chapter 7 considers different treatment effects and gives a worked example of a treatment effect applied to the model in order to develop the comparator arm and to generate cost effectiveness values.

The overriding purpose of this thesis is to aid better decision-making, by developing a generic disease model that can be used to appraise COPD therapies on the same platform. The model is built from the bottom up, based by an understanding of the epidemiology of the disease as a result of analyses conducted on a wealth of data. The methods and principles used in this thesis are generalisable to modelling other disease areas.

Chapter 2. COPD

The purpose of this chapter is to provide background details on chronic obstructive pulmonary disease (COPD) and begins by introducing COPD and the associated health and financial burdens that arise as a result of the disease, for the patient, to the health care system and to society at large. Treatment for the disease is discussed and the changing pattern of treatment and associated costs to the National Health Service (NHS) are detailed and is followed by an overview of the clinical management of the disease and how this has changed over time. A section that describes the current and historical COPD diagnostic criteria and a review of how these criteria affect prevalence estimates is presented.

2.1 The disease

COPD is a lung disease that principally affects older people with a history of smoking. People with COPD initially complain of breathlessness and may also have cough and increased sputum production, which tend to worsen over time.

COPD is defined by the Global initiative for chronic Obstructive Lung Disease (GOLD) as:

> "…a preventable and treatable disease…characterised by airflow limitation that is not fully reversible. The airflow limitation is usually progressive and associated with an abnormal inflammatory response of the lung to noxious particles or gases."(2)

COPD is an umbrella term that covers both parenchymal destruction (emphysema) and small airways disease (obstructive bronchiolitis). Small airways disease is caused by structural changes following inflammation which cause the thickening of the bronchial walls and the narrowing of the airways and results in irreversible airflow obstruction.

Small airway diseases include bronchiectasis and bronchiolitis obliterans. As the damage to the airways increases, the patient is likely to suffer from breathlessness and because the body is more susceptible to infection, phlegm and cough may be produced in response. Parenchymal destruction or emphysema refers to the permanent enlargement of the alveoli to inflammation. The walls of the alveoli are stretched beyond repair, leading the alveoli to merge into larger sacs, reducing the surface area for gas exchange and to airflow limitation.

Lung function is a key indictor of disease and disease severity. Lung function naturally decreases over time with age and the decline has been found to occur faster in smokers than non smokers.(3) Lung function is measured using spirometry and is frequently used to derive values of forced vital capacity (FVC) and forced expiratory volume in one second (FEV_1). From this, the ratio of FEV_1 to FVC (FEV_1/FVC) is calculated. Most guidelines for COPD agree that COPD is present in patients where FEV_1/FVC < 0.7,(2;4-7) (discussed in detail in section 2.4.3).

FEV_1 prediction equations based on age and sex and developed within a healthy population are used to predict FEV_1 values for each patient (details of developing a prediction equation are described in Chapter 5, section 5.2.2). The observed FEV_1 value is divided by the predicted FEV_1 value to give a lung function value called FEV_1 % predicted which gives a relative lung function score compared to a healthy person of the same age and sex. FEV_1 % predicted is employed in order to determine COPD disease severity.

Within guidelines for diagnosing COPD, disease severity is categorised based on lung function. Different guidelines categorise disease severity in different ways. For instance, as illustrated in table 2.1, disease severity in the GOLD guidelines is categorised in terms of mild, moderate, severe and very severe according to the degree of respiratory impairment.(2) COPD is diagnosed based solely on a lung function FEV_1/FVC score <0.7 plus $FEV_1 \geq 80\%$ predicted for mild, $50\% \leq FEV_1 < 80\%$ predicted for moderate,

$30\% \leq FEV_1 < 50\%$ predicted for severe and $FEV_1 < 30\%$ predicted or $FEV_1 < 50\%$ predicted plus chronic respiratory failure for very severe disease. This is in contrast to the National Institute for Health and Clinical Excellence (NICE) diagnostic criteria where COPD is based on $FEV_1/FVC < 0.7$ plus evidence of a risk factor (principally smoking), and symptoms such as breathlessness, cough, regular sputum,[8] and where disease severity is based on FEV_1 % predicted score which can be either: mild ($50\% \leq FEV_1 < 80\%$ predicted), moderate ($30\% \leq FEV_1 < 50\%$ predicted) or severe ($FEV_1 < 30\%$ predicted). The comparison is made between the GOLD and the NICE criteria as the GOLD criteria are frequently used in many types of study, in observational datasets and in clinical trials, and as this thesis is primarily focussed in the United Kingdom (UK), the NICE guidelines for COPD represent current thinking on the disease within the UK. Details of these and other diagnostic criteria for COPD are presented towards the end of this chapter in section 2.4.3.

Table 2.1 Diagnostic criteria for COPD according to the GOLD and NICE guidelines.(2;7)

GOLD	Mild	Moderate	Severe	Very severe
	$FEV_1/FVC < 0.7$ $FEV_1 \geq 80\%$	$FEV_1/FVC < 0.7$ $50\% \leq FEV_1 < 80\%$	$FEV_1/FVC < 0.7$ $30\% \leq FEV_1 < 50\%$	$FEV_1/FVC < 0.7$ $FEV_1 < 30\%$ or $FEV_1 < 50\%$ + chronic respiratory failure
NICE		Mild	Moderate	Severe
		Risk factor Resp symptoms $FEV_1/FVC < 0.7$ $50\% \leq FEV_1 < 80\%$	Risk factor Resp symptoms $FEV_1/FVC < 0.7$ $30\% \leq FEV_1 < 50\%$	Risk factor Resp symptoms $FEV_1/FVC < 0.7$ $FEV_1 < 30\%$

FVC= Forced Vital Capacity. FEV_1= Forced Expiratory Volume in one second

Exposure to cigarette smoke is the most common risk factor for COPD, however other risk factors do exist, including occupational dusts and chemicals, and indoor air pollution as a result of cooking/heating using biomass fuels in homes with poor ventilation (particularly problematic for women in developing countries).(2)

Exacerbations are a characteristic of the disease, particularly in more severe COPD patients.(9) Exacerbations have been said to occur when there is a worsening of respiratory

symptoms requiring treatment with oral corticosteroids and/or antibiotics.(10) Severe exacerbations occur where worsening of symptoms requires that the subject is hospitalised. Reductions in quality of life occur as a direct result of exacerbations. Recovery periods can be protracted and during this period, Health Related Quality of Life (HRQoL) will be less than normal. The duration of recovery periods can vary substantially between individuals.

COPD and asthma often present with similar symptoms making differentiation potentially difficult, however there are differences and these are described in table 2.2. For example symptoms are rare in the under 35's for COPD but are common in asthma patients. The use of spirometry can aid differentiation between the two conditions. Administration of a short-acting bronchodilator prior to spirometry should cause reversibility for asthma sufferers whereas for COPD sufferers, the airflow obstruction is largely irreversible.

Table 2.2 Clinical features differentiating COPD and asthma, reproduced from the NICE guidelines.(7)

	COPD	Asthma
Smoker or ex-smoker	Nearly all	Possibly
Symptoms under age 35	Rare	Common
Chronic productive cough	Common	Uncommon
Breathlessness	Persistent and progressive	Variable
Night time waking with breathlessness and/or wheeze	Uncommon	Common
Significant diurnal or day to day variability of symptoms	Uncommon	Common

2.2 Burden of COPD

COPD is a major cause of morbidity and mortality worldwide and is the only major cause of morbidity that is increasing. A substantial increase in the global burden of COPD is expected for the future. The health burden of COPD including: mortality, morbidity, the effects of age, gender and smoking, together with the financial burden of the disease, with

particular emphasis on the key drivers of cost: disease severity, exacerbation frequency and severity, are outlined in this section.

2.2.1 Health Burden

Current estimates of COPD prevalence in the general population of Europe are large and variable at between 4 and 10%.(11) The difference in prevalence estimates is attributable to the population under study and the different diagnostic criteria used in epidemiological studies; this issue is covered in section 2.4.3 and in detail in 5.6 within Chapter 5. The population of diagnosed COPD cases within primary and secondary care is much lower and is approximately 1% in the UK,(12) and evidence suggest that COPD is heavily under diagnosed worldwide.(12-14) It is likely that a substantial number of people are unaware of having COPD (especially so in mild to moderate stages) and who, rather than seek help, may attribute problems such as breathlessness and fatigue to old age rather than on any underlying cause. Figure 2.1 shows the reported breakdown, by disease severity, of a known COPD population (Mediplus UK, see section 5.1 for details of this dataset).(15) 17% of the population had mild COPD compared to 48% with moderate and 27% with severe disease. The true breakdown of COPD within the general population is likely to be different from this pattern because it is assumed that the majority of COPD patients are in the mild group. These data support the viewpoint that COPD is under diagnosed worldwide,(12-14;16) particularly among mild cases of COPD.
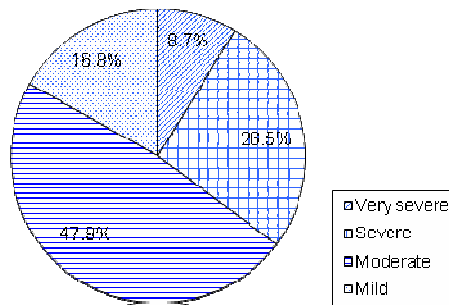
Figure 2.1 Proportion of COPD patients, by disease severity

COPD causes an accelerated depreciation of lung function over time compared to the average or predicted level for a healthy person, and is further accentuated by smoking.(3) Calverley et al suggest that around 20% of smokers are susceptible to some form of progressive lung disease,(12) but it may be larger than this: a different study reports that the absolute risk of developing COPD in smokers is at least 25%.(17) In China where COPD represents a major public health problem, smoking rates are very high: it was estimated that 67 % of the men in China smoke (approximately 300 million).(18) Approximately 15% of people in China who have ever smoked and 5% of people who have never smokers, are thought to have COPD.(14)

Smoking cessation has been shown to effectively slow the deterioration in FEV_1 and return the trajectory of lung function to one consistent with that of a non smoker.(3) The Lung Health study, at five years, found significantly lower all-cause mortality rates in a 'special intervention' group where smoking cessation was actively encouraged, compared to the non intervention group.(19)

Worldwide Burden of COPD

The World Health Organisation found global COPD deaths to be the fifth largest cause of death, accounting for 4.5% of deaths worldwide.(20) The proportion of deaths attributable to COPD varies between regions of the world. Of particular concern is the Western Pacific region (including China, Malaysia, the Philippines and Vietnam) where COPD accounts for 13.8% of all deaths and where COPD is ranked as the second leading cause of death, as shown in table 2.3 in the final row.

Table 2.3 Worldwide mortality burden of COPD

WHO region	% of deaths from COPD	Relative ranking of COPD mortality
African	1.1	15
Americas	3.5	6
Eastern Mediterranean	1.4	15
European	2.8	5
South East Asia	2.2	9
Western Pacific	13.8	2

% Percentage of total deaths attributable to COPD in each region. Rank for COPD deaths within region, compared to all other diseases/illnesses

Data from Europe suggest a lesser impact of the disease than in other regions; nevertheless, COPD was found to account for 2.8% (table 2.3, row 4) of all deaths and was ranked as the fifth cause of death within this region.

COPD prevalence increases with age and is generally a disease that occurs in an older population. Data from both the US and the UK show that the prevalence of COPD is comparatively small among the under 45's but increases markedly throughout later years.(21) In the UK, about 1% of the general population is diagnosed with COPD increasing with age to around 5% of men between 65 and 74 and rising to 10% in men aged 75 years and over.(12)

Gender specific mortality for COPD seems to be country specific. In Canada and in Northern Europe, there is little difference between death rates by gender.(11) In other countries there are notable differences. Many more men than women die of COPD in Eastern and Southern European countries.(14) Mortality rates for Europe as a whole suggest that two to three times as many men die from COPD as women.(11) In Singapore, female hospitalisation and mortality from COPD is significantly less than for male counterparts.(14) These differences are most probably attributable to historical reasons where smoking rates amongst men have been higher than for women. In recent years and in some countries, women now smoke as much as their male counterparts. Where smoking prevalence rates are equal, and have been for some time, similar mortality rates for COPD are expected. In countries where smoking rates have increased, the burden of COPD is expected to increase in the future.

2.2.2 Financial Burden

Within the UK, chronic conditions such as COPD place a major burden upon the NHS. Approximately 80% of all General Practitioner (GP) consultations are related to chronic disease, 60% of all hospital bed days are used by patients with chronic disease and two thirds of medical emergencies are either a result of an exacerbation of a chronic disease, or by the disease itself.(22) Total costs to the NHS for COPD have been estimated somewhere between around £486 million (23;24) and £848 million (25;26) per year. The major drivers of COPD burden are disease severity and exacerbations.

Disease Severity

Costs increase substantially as disease severity (measured by FEV_1% predicted) moves from moderate to severe (with a smaller increase between mild and moderate

groups).(24;27-32) Britton et al estimated that the direct cost for the three groups was €232 (£209) for mild disease, €477 (£430) for moderate and €2026 (£1826) for severe disease.(24) Total cost doubled with the inclusion of productivity costs, which was defined in the study as time lost from work for those under retirement age (see section on productivity costs below for more on this type of cost). Productivity cost was found to be €399 (£331) for mild, €202 (£168) for moderate and €2331 (£1934) for severe disease.(24) As FEV_1 deteriorates, a general shift from outpatient care to hospitalisation, an increase in the use of oxygen therapy and a subsequent increase in total costs, especially in the most advanced stages of the disease has been shown to occur.(33)

Exacerbations

Exacerbations are the leading driver of cost in COPD. A serious exacerbation leads to hospitalisation; indeed an exacerbation is the main reason why a COPD patient would attend hospital. COPD is responsible for approximately 1 million hospital bed days per year.(34) Mean unit cost was £1109 for a length of stay of four days for an admission with no complications and £1516 for a six day stay with complications,(35) or a total cost of approximately £142 million for non-elective inpatient COPD hospital admissions alone.

The cost of exacerbations has been found to increase in line with the severity of the exacerbation; a Swedish study reports: SEK 120 (£11) for mild, SEK 354 (£31) for mild/moderate, SEK 2111 (£185) moderate and SEK 21 852 (£1919) for severe exacerbations. Exacerbations account for between 35-40% of the total per capita health care costs for COPD.(36)

In the UK, the cost arising from acute exacerbations has been estimated to be £45 million at 1994 prices for a COPD population of 233 000.(37) Treatment which acts to reduce or prevent disease progression and or an exacerbation (particularly severe exacerbations) will

have a direct effect on the total cost for COPD.(38) In England and Wales, McGuire calculated that for every exacerbation related hospital admission avoided a total saving of approximately £1200 would be made.(37)

Productivity Costs

Productivity costs represent lost output/productivity within the economy due to ill health and mortality.(39) Productivity costs for COPD represent a burden on society as COPD is a cause of absenteeism from work.(24;40) People with COPD have a 'substantially shortened' work life compared to the population average.(41) Between 1994 and 1995 it was estimated that 24 million lost work days were attributable to COPD within the UK alone.(42) Within the 15 'original' EU member states, COPD was estimated to account for 41 300 lost work days per 100 000 people and productivity losses of around €28.5 (£25.7) billion per year. In Central and Eastern Europe, the number of lost work days was found to be 10% of this value; with a rate of 4300 per 100 000 people.(11)

In one UK study, 44% of COPD patients were below retirement age and because of the disease, 24% were completely prevented from working and 5% of patients' carers missed work. (24) Around 12 days were missed from work, per patient per year. Productivity costs were found to be almost equivalent in size to direct costs, imposing an additional £820 per patient per year upon the economy.(24) Whilst productivity costs can represent a considerable burden on both the individual and on society, their use in economic evaluation is an area of debate and will be discussed in detail in Chapter 3.

Treatment and Management Costs

In 2002, Britton et al estimated mean drug costs for COPD to be £152.84 per person per year.(24) Costs, especially drug costs, have increased significantly since then because of the change in the mix of prescribing and is shown in detail in section 2.3.1.

The cost impact arising from the change in the management of COPD (see section 2.4 for details on this change) is much harder to quantify, though it is anticipated that costs have increased: greater awareness of COPD, routine spirometry, the Hospital at Home guidelines (which are explained in section 2.4.2) and the other more general trends seen within COPD all demand additional resources. The British Lung Foundation (BLF) has met some costs pertaining to the "Hospital at Home" guidelines (Malcolm Shepherd, personal communication), however, extra nurses and respiratory specialists, numerous equipment costs, including nebulisers and improved prescribing of domiciliary oxygen therapy are required nationwide and this implies additional costs. No references were found that report the change in management costs for the UK over recent years.

2.3 Treatment of COPD

The aim of treatment for COPD, in the absence of a disease cure, is to prevent and control symptoms, reduce the frequency and severity of exacerbations, improve health status and improve exercise tolerance.(2) COPD, by definition, is a chronic condition, lasting over the course of a patient's life. There is no cure, patients symptoms may be managed in order to maximise quality of life, subject to resource limitations.

Treatment is dependent upon disease severity as previously described and disease management is additive. As the disease progresses through the stages, more treatments are

added, for example, in the European market in 2006 on average, mild patients (GOLD classified) were prescribed 1.4 products, moderate 1.8, severe 2.6 and very severe patients, 3 products.(43) Because COPD is a progressive disease, treatment must be continued throughout the lifetime of the patient.(38)

Table 2.4 Pharmacological classes and drugs used for treating COPD

Drug class	Generic name (commercial)
Inhaled Corticosteroids (ICS):	Beclomethasone
	Budesonide
	Triamcinolene
	Fluticasone
Short acting Bronchodilators:	
Short acting β2 agonist (SABA)	Salbutamol (Albuterol)
	Terbutaline
Short acting anticholinergic (AC)	Ipratropium
Long acting bronchodilators:	
Long acting β2 agonist (LABA)	Formoterol
	Salmeterol
Long acting anticholinergic	Tiotropium (Spiriva)
Combination products:	
SABA + AC	Salbutamol/Ipratropium
LABA +ICS	Formoterol/Budesonide (Symbicort)
	Salmeterol/Fluticasone (Seretide)
Methylxanthines:	
	Theophylline

As seen in table 2.4, there are five broad classes of drug treatments for COPD: inhaled corticosteroids (ICS), short acting bronchodilators, long acting bronchodilators, combination products (where two different substances have been combined in the same device) and methylxanthines. Table 2.4 above contains the major pharmacological treatment options available for COPD and provides the generic names of commonly used therapies for COPD within each class, together with some of the commercial names (in brackets) of new entrants.

Current treatment guidelines for the treatment of COPD as suggested by GOLD are outlined in table 2.5 and are similar to those proposed by NICE.(7) Both guidelines

approve of the use of smoking cessation programmes, rehabilitation and the administration of the influenza and pneumococcal vaccinations from early stages of the disease.

Table 2.5 Treatment guidelines for COPD, adapted from GOLD.(2)

	Mild	Moderate	Severe	Very severe
Lung function criteria	$FEV_1/FVC<0.7$ $FEV_1\geq80\%$ predicted	$FEV_1/FVC<0.7$ $50\%\leq FEV_1<80\%$ predicted	$FEV_1/FVC<0.7$ $30\%\leq FEV_1<50\%$ predicted	$FEV_1/FVC<0.7$ $FEV_1<30\%$ predicted or $FEV_1<50\%$ + chronic respiratory failure
Treatment	Influenza vaccination Short acting bronchodilator added when necessary: 1) Short acting β2 agonists: fenoterol, salbutamol, terbutaline. 2) Short acting anticholinergic: ipratropium			
		Add rehabilitation Add regular treatment with one or more long acting bronchodilators if needed: 1) Long acting β2 agonists: formoterol, salmeterol. 2) Long acting anticholinergic: tiotropium		
			Add inhaled corticosteroids if exacerbations are repeated: fluticasone, beclomethasone, triamcinolone, Budesonide	
				Add long term oxygen if chronic respiratory failure. Consider surgical treatments: lung volume reduction surgery, lung transplantation, bullectomy.

FVC= Forced Vital Capacity. FEV_1= Forced Expiratory Volume in one second

Smoking cessation is important in stemming the progression of COPD because quitting smoking has been shown to return subsequent decline in lung function to a normal rate, consistent with that of a non-smoker.(3)

> "Smoking cessation is the single most cost effective way to reduce exposure to COPD risk factors."(2)

Interventions to assist individuals in quitting smoking include: counselling, nicotine replacement products (gum, patches, spray, lozenges) and drug therapies: bupropion, nortriptyline and varenicline. Smoking cessation programmes are widely available in both

primary and secondary care settings and in some cases, through the pharmacy as well as via telephone and online support.

Rehabilitation is a multifaceted approach which incorporates a wide range of programs to improve quality of life and functional independence and to reduce symptoms and disability for the patient. Areas within the rehabilitation program can include exercise and physical training, psychological, social interactions, education around the disease, and about nutrition.(44) Rehabilitation usually occurs within secondary care.

NICE is less explicit about when treatments should be added in terms of stage of disease severity: nonetheless both guidelines suggest the initial use of short-acting bronchodilators for patients and then inhaled treatment should be intensified by adding a long-acting bronchodilator or combined therapy with a short-acting beta2 (β2)-agonist and a short-acting anticholinergic, in patients who remain symptomatic, as seen in table 2.5.

With regards to the use of inhaled corticosteroids, the GOLD guidelines recommend usage in patients with an $FEV_1 \leq 50\%$ predicted, or severe and very severe patients as illustrated in table 2.5, where exacerbations are repeated. NICE recommends combination therapies if the patient remains symptomatic on monotherapy.

Both guidelines suggest the patient be assessed for oxygen therapy when FEV_1 is less than 30 % predicted and it is used to increase the partial pressure of oxygen in the arterial blood (PaO2) with a view to: relieve dyspnea, aid oxygen intake during exercise and for long term, continuous treatment.(2) Portable oxygen therapy is available to enable greater mobility of the advanced COPD patient. Oxygen therapy forms a high proportion of outpatient costs for COPD patients.(45)

Lung surgery can be considered in some patients with advanced stages of the disease. There are three types of surgery: lung volume reduction surgery (LVRS), lung transplantation and bullectomy. LVRS involves cutting away around 30% of the diseased lung tissue in order to increase the effectiveness and efficiency of the remaining lung and surrounding muscle. With lung transplantation either a single or both lungs are replaced. Bullectomy may be an option when a large air filled bulla exists that fills half of the thoracic volume and compresses the relatively normal adjacent parenchyma.(46) Surgery for COPD patients is rare and is very expensive with variable outcomes. Suitable candidates must satisfy strict criteria,(2) to ensure that only patients likely to benefit and least likely to have associated complications, are selected. For example, to be eligible for a bullectomy under the NICE guidelines, patients must have a lung function of FEV_1 30–49% predicted, have breathlessness as a symptom and have a single large bulla on a CT scan.

A substantial part of the treatment and management for COPD occurs in the primary care setting; including diagnosis, prescription of appropriate pharmacotherapy, vaccination, smoking cessation programmes and repeat prescription oxygen therapy.

As the severity of the patient's COPD intensifies, there is an increased probability that the patient will be admitted to secondary care, primarily due to exacerbations. Other reasons why the patient would be admitted into hospital include surgery for COPD, rehabilitation, the prescription of oxygen therapy and any other co-morbid conditions that require hospitalisation. Patients may also be sent to hospital to confirm a diagnosis or to be tested for COPD, though increasingly first diagnosis is shifting away from secondary care into the primary care setting.

2.3.1 The Changing Face of Treatment for COPD

Since 2000 there has been an emergence of new treatments for COPD including long-acting bronchodilator drugs, respiratory rehabilitation services, and non-invasive ventilation in respiratory failure. Two combination products (combining a long acting β2 agonist and an inhaled corticosteroid) entered the market; the first in 2000 was Seretide, and the second, in 2001 was Symbicort. A third product, tiotropium (Spiriva), a long acting anticholinergic, entered the market in 2002. Clinical trials providing evidence on these new products are briefly described below, in order to provide information on the trials supporting these new treatments, the size of the patient population and the outcome measures used within the trial.

The TORCH (Towards a Revolution in COPD Health) (GlaxoSmithKline (GSK)) trial was published in early 2007, and followed a cohort of over 6000 patients for 3 years. The study compared the product Seretide with its component products, fluticasone and salmeterol and with placebo. The primary outcome measure was mortality. Other outcome measures included the St George's Respiratory Questionnaire (SGRQ),(47) and the EuroQoL-5D,(48) rate of exacerbations and post bronchodilator FEV_1.(49)

TRISTAN was a multi-centre, multi-national randomised controlled trial (RCT) of 1974 COPD patients followed for 12 months (GSK). The study was designed to compare the efficacy of salmeterol/fluticasone with salmeterol alone, fluticasone alone and placebo with primary endpoints including pre bronchodilator FEV_1, rate of exacerbations, and the SGRQ.(50)

UPLIFT (Understanding Potential Long-term Impacts on Function with Tiotropium) was a four year study of 6000 COPD patients, published in 2008, comparing tiotropium to

placebo (Boehringer Ingelheim). Outcome measures included lung function: FEV_1, FVC, and slow vital capacity (SVC), health status as measured by the SGRQ and rate of exacerbations.(51)

The tiotropium trials were two one-year trials that were run simultaneously in different countries to evaluate the efficacy and safety of tiotropium compared to ipratropium (Boehringer Ingelheim) in COPD patients. Outcome measures included: FEV_1, dyspnea using the transition dyspnea index (TDI) and the baseline dyspnea index (BDI). HRQoL was assessed using both the SGRQ and the short form 36 (SF-36). COPD exacerbations were also monitored.(52)

INSPIRE (53) followed 1270 patients from 20 countries who were randomised to either salmeterol/ fluticasone or to tiotropium over 104 weeks (GSK). The primary outcome was rate of exacerbations and secondary outcomes included post dose FEV_1, SGRQ, and all-cause mortality.

The OPTIMAL trial was a 52 week Canadian trial developed and implemented by the Canadian Thoracic Society Clinical Trials group (an academic group) in a sample population of 432 COPD patients, randomised to one of three treatments: tiotropium plus placebo, tiotropium plus salmeterol or tiotropium plus Seretide. The trial was designed to answer a question about which combination of products is the best for treating COPD patients. The primary outcome measure was the proportion of patients in each arm experiencing an exacerbation, secondary endpoints included the SGRQ, changes in dyspnea as measured by the BDI, the TDI and the dyspnea domain of the chronic respiratory disease questionnaire (CRQ), number of exacerbations and hospitalisations, time to first exacerbation, FEV_1 and FVC.(54)

The recent trials in COPD formed a considerable evidence base which supported the introduction of Seretide, Symbicort and tiotropium onto the market. Of all the trials, TORCH contained the most participants at over 6000 people, over the longest duration of follow up. All of the trials included measures of lung function, exacerbations and HRQoL. Economic evaluations from the literature which use the results from these trials are presented in Chapter 4.

Since 2000, on the basis of the trial results, Seretide, Symbicort and Spiriva have entered the market. The entrance of each product is clearly illustrate in figure 2.2 where the % of COPD patients taking the medication increased from 0 over time as the new products caused a change in prescribing, away from the more established treatments and towards the newer treatments.
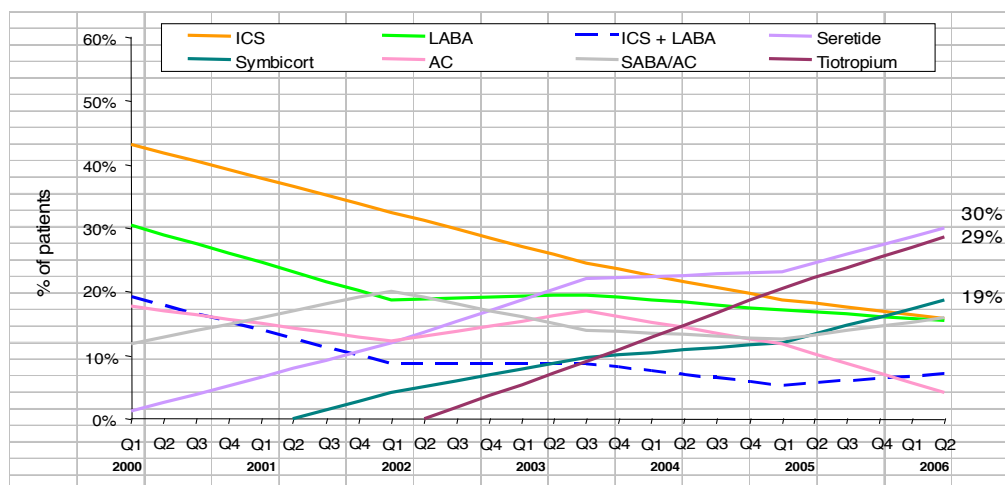
Figure 2.2 Therapy class shares (%) of the total market over time
Europe COPD market 2000-2006.(43)

Seretide, Symbicort and Spiriva were successfully adopted within the European market and by the second quarter of 2006, therapy class shares of the total patient market were 30% for Seretide, 29% for Tiotropium and 19% for Symbicort as seen on the right hand side of the

graph in figure 2.2. In response, as illustrated in figure 2.2, there was a gradual decline in the separate prescribing of ICS, LABAs and ACs (43) from approximately 43%, 30% and 18% down to 16%, 16% and 6% respectively and a fall in the use of ICS and LABA used together (but not in a combination product) from 20% to 8%.(43) Clinical trials assessing the efficacy and effectiveness of triple combination products are emerging,(55-57) and if found to be successful, further changes to the COPD market for treatments are expected in the future.

A clinical expert estimated the mix of treatments that would generally be prescribed to a moderate/severe patient in the year 2002 and then again for (March) 2009 (Malcolm Shepherd, personal communication). Using the British national formulary (BNF53),(58) the cost of each drug was determined (in 2009 prices) and is presented in table 2.6. Costs for these drugs are not assumed to have changed significantly over the period. In the 2009 prescription, it was assumed that a moderate/severe patient would be prescribed an inhaled corticosteroid, a β2 agonist, an anticholinergic and additionally, tiotropium.

Table 2.6 The changing treatment mix and corresponding cost increase of COPD drug prescriptions: 2002-2009

	Prescription	N doses pe day	Unit cost	30 day cost
2002				
Beclomethasone 100mcg	100 dose	4	£5.58	£6.70
Ipratropium 20mcg	200 dose	8	£4.21	£5.05
Salbutamol 100mcg	200 dose	8	£2.88	£3.46
	Total 30 day cost			£15.21
2009				
Seretide 250	120 dose	4	62.29	62.29
Tiotropium inhaler 18mcg	30 dose	1	37.62	37.62
Salbutamol 100mcg	200 dose	8	£2.88	£3.46
	Total 30 day cost			£103.37

Between 2002 and 2009, table 2.6 shows that there were significant increases associated with the cost of prescriptions for COPD due to the availability of newer and more

expensive treatments. The yearly prescription cost for moderate to severe patients, based on the assumptions used was £182.52 in 2002 compared to £1240.44 in 2009. That is, in 2009 prices, treatment costs are estimated to be almost seven times that of in 2002. Future treatment costs are expected to increase further still, if and when, triple therapy combination products are launched.

2.4 Management of COPD

Since the late 1980s, there has been a significant shift in the awareness and management of COPD. Prior to this period, COPD was largely ignored because of the widely held opinion that little could be done to treat COPD patients.(7) This section describes recent changes to the management of COPD.

> "Good chronic disease management offers real opportunities for improvements in patient care and service quality, and reductions in costs…and enabling people living with chronic conditions to attain the best possible quality of life."(59)

2.4.1 Routine Spirometry

Since 2003 there has been a move towards routine collection of spirometry data by GP's from those patients deemed 'at risk' of developing COPD. Within the general medical services contract (GMSC), clinical practices are rewarded payments for having records of COPD patients, an initial diagnosis of COPD (where diagnosis is confirmed by spirometry) and ongoing management of COPD.(60)

Little research has been published on the uptake of spirometry in routine practice. The earliest paper (published in 1999), a study around the uptake of spirometry in North Staffordshire, reported that while 21% of the general practices surveyed (88% response

rate) owned a spirometer, 12% were using it and two users out of 84 practices had received specific training in the use of spirometry.(61) Another study surveyed general practices in Leicestershire between January to March 2002 and found that very few practices within the area had sufficient resources to provide high quality care for COPD (65% response rate). 54% of those surveyed owned one or more spirometer(s) and 15% had one or more people with current training in spirometry.(62) A study using a similar survey, sent three years later (February 2005), after the introduction of the GMSC, to general practices within Grampian, Scotland revealed widespread use of spirometry (89% response rate): 93% of the practices had at least one spirometer and at least one person was trained in the use of spirometry in 92% of the responding practices.(63) As yet evidence is lacking as to whether the extensive use of spirometry as found in Grampian is a widespread trend across the country, emerging from the introduction of incentives offered by the GMSC, or whether it is location specific. However, expert opinion (Malcolm Shepherd, personal communication) supports the initial evidence that the pattern of first diagnosis of COPD is shifting away from secondary care and into primary care.

Concerns have been raised about the validity of spirometry results in general practice due to insufficient training and knowledge around the use and administration of spirometry.(61-65) The British Thoracic Society has issued a practical guide to encourage the use of spirometry amongst doctors and nurses in order to try and address this problem.(66)

2.4.2 Early Discharge

Guidelines from the British Thoracic Society (2007) recommend support of patients once they return home by way of the 'Hospital at Home' strategy. This is a process whereby treatment is delivered by specialist respiratory professionals (including nurses, occupational

therapists and physiotherapists) who visit the patient at home and treat the patient over a finite period, known as early supported discharge.(67)

Evidence from one study suggests that patients with uncomplicated acute exacerbations of COPD may be discharged earlier than is current practice provided that they are visited by a respiratory specialist.(68) The study found that hospital stay was reduced by almost a half, from 6.1 to 3.2 days, and found no associated increases in either re-admission or mortality rates.(68) Potentially large cost savings arising from a reduction in consumed resources could be achieved, for example by bed days avoided. A reduction in the number of bed days used would directly reduce costs to the NHS associated with COPD and in time, the resulting available beds would be redeployed for use in other disease areas.   In order to determine if this is a cost saving strategy, the additional costs associated with the "Hospital at Home" strategy such as the employment of a respiratory nurse to conduct the "Hospital at Home" visits, would need to be weighed against the cost saving of less time spent in hospital for the patient.

2.4.3 Guidelines: Diagnostic Criteria

Many guidelines advising on the diagnosis and treatment of COPD have been published and the effects of applying these criteria to the same population have often been examined, using prevalence statistics. This section describes some of the most frequently used diagnostic criteria and reports on studies that have compared criteria using prevalence statistics.

The first COPD guidelines were published in 1986 by the American Thoracic Society (ATS), and the British Thoracic Society (BTS) published the first UK guidelines for COPD in 1997. These guidelines resulted in significant improvements in the recognition and care

of COPD patients,(7) as they increased awareness of the disease, and for the first time presented a way of managing and treating people with COPD, as described above in section 2.4. Since then and on the back of these guidelines, numerous other guidelines have been developed, most notably, that of the Global initiative for chronic Obstructive Lung Disease (GOLD). From 2001, GOLD has produced, in collaboration with the US National Heart, Lung and Blood Institute (NHLBI) and the World Health Organisation, guidelines on the diagnosis and management of COPD. A total revision of the GOLD guidelines were published in 2006, and an update was published in 2007.(2) NICE produced their own guidelines for COPD in 2004.(7) Details of these criteria are described below.

It is important to consider the diagnostic criteria because differences in the definition of COPD can have large effects in terms of estimating the prevalence of COPD within a given population.(69) Different definitions of obstruction for diagnosing COPD could lead to estimates of prevalence that vary from one another by greater than 200%.(70)

Global Initiative for Obstructive Lung Disease (GOLD)

The GOLD guidelines state that "COPD is characterized by airflow limitation",(2) defined as $FEV_1/FVC<0.7$. Symptoms and risk factors are mentioned but are not explicitly included within the diagnostic criteria. In recent years there has been some consensus over the use of the criteria published in the GOLD guidelines and the GOLD criteria are frequently employed to identify COPD. This definition is based solely upon lung function; however it is widely known that the clinical diagnosis of COPD takes into account more information than lung function alone.

European Respiratory Society (ERS)

The European Respiratory Society state that the main risk factors for COPD are cigarette smoking and occupational exposure. The two main symptoms are breathlessness and cough, sometimes accompanied by wheezing or sputum production. Chronic cough is present in most patients. Patients with COPD are usually aged over 40 years and have functional evidence of moderate or severe airflow limitation. Airflow limitation is identified by a reduction in the ratio of FEV_1 to vital capacity (VC) or FVC: $FEV_1/VC<88\%$ in men, $FEV_1/VC <89\%$ in women and $FEV_1/FVC<70\%$.(6)

American Thoracic Society (ATS)

The American Thoracic Society produced standards for the diagnosis and care of patients with COPD in 1986.(71) The standards state that COPD is characterised by airflow obstruction with exposure to tobacco smoke being the primary cause of the disease. COPD patients have at least 20 pack years (20 cigarettes per day for 20 years) before symptoms develop and the symptoms commonly present from ages 50 with productive cough or an acute chest illness. Dyspnea on effort usually does not occur until the 60's or 70's. Sputum production is gradual and increases over time. Chest illness is characterised by increased cough, purulent sputum, wheezing, dyspnea and occasionally fever.(71) Spirometry values for diagnosis of COPD were published in a later paper, and were reported as $FEV_1/FVC<75\%$.(72)

More recently (2004), the ATS and the ERS issued a joint statement on diagnosis and treatment of COPD

ATS/ERS

The ATS/ERS guidelines state that a diagnosis of COPD should be considered in any patient who has the following: symptoms of cough, sputum production, dyspnoea; or history of exposure to risk factors for the disease. The diagnosis requires spirometry (post-bronchodilator) $FEV_1/FVC < 0.7$ and that the airflow limitation is not fully reversible. Spirometry should be obtained in all persons with the following history: exposure to cigarettes; and/or environmental or occupational pollutants; and/or presence of cough, sputum production or dyspnoea.(5)

British Thoracic Society (BTS)

The British Thoracic Society produced guidelines for COPD in 1997 and defined COPD as a chronic, slowly progressive disorder characterised by airways obstruction (FEV_1 <80% predicted and FEV_1/VC ratio <70%), with most COPD cases being caused by smoking.(4) The guidelines state that a diagnosis of COPD is normally suggested by symptoms but can only be established by an objective measure, preferably using spirometry. Treatment can improve symptoms and measured airflow limitation.

Symptoms and signs were thought to vary with the severity of the disease. In mild disease (FEV_1 60-80 % predicted), no abnormal signs would be present, a smoker's cough may occur and there would be little or no breathlessness. In moderate disease (FEV_1 40–59 % predicted) patients would generally be breathless (with or without wheeze) on moderate exertion and have a cough (plus or minus sputum). Some abnormal signs would present (general reduction in breath sounds and presence of wheeze). In severe disease (FEV_1 <40 % predicted) breathlessness on any exertion and/or at rest could occur and wheeze and cough are often prominent. Signs include lung over-inflation, cyanosis, peripheral oedema

and polycythemia in advanced disease, especially during exacerbations.(4) The BTS guidelines were superseded by the NICE guidelines in 2004.

National Institute for Health and Clinical Excellence (NICE)

The NICE guidelines recommend that a diagnosis of COPD should be considered in patients: aged over 35, with a risk factor (principally smoking), who present with one or more of the symptoms: exertional breathlessness, chronic cough, regular sputum production, frequent winter 'bronchitis' or wheeze, and who have airflow obstruction, considered present when both $FEV_1/FVC<0.7$ and $FEV_1<80\%$ predicted.(7)

Comparison of Prevalence Using Different Diagnostic Criteria

A study compared the ATS criteria (spirometry values) to those from the ERS (spirometry values) and used data on symptoms to validate the criteria in terms of overall accuracy, predictive value, sensitivity and specificity.(73) Prevalence rates (aged ≥ 46 years) were 14.5% (12.0%) for the men (women) with ERS, 33.1% (22.2%) using a clinical definition (symptoms) and 60.7% (53.4%) using the ATS criteria. The authors found that age, height and number of pack years were statistically significant in the prediction of prevalence of COPD. Prevalence of COPD was found to be dependent upon the definition used.(73)

The GOLD criteria (spirometry values) was compared to the BTS criteria (spirometry values) in a random population sample (aged ≥ 46 years). The prevalence of COPD was 8% using BTS and 14% using the GOLD criteria. Using the BTS criteria and stratifying by age, gave prevalence rates of: 1% (46-47 years), 2% (62-63 years) and 16% (76-77 years) compared to rate of: 5%, 24% and 45% respectively when employing the GOLD criteria. 94% of those diagnosed with COPD using BTS guidelines were symptomatic compared to

88% with the GOLD criteria. Age and smoking history were found to be important risk factors for the disease.(74)

In a study that employed five different diagnostic criteria in a random population sample (aged $\geq$ 40 years) COPD prevalence ranged from 8.2% to 26% for the males and 6.6% to 28% for the women.(75) Criteria assessed were: 1) based upon the symptoms of chronic bronchitis[1] and no spirometry, 2) the symptoms of chronic bronchitis plus $FEV_1/FVC \geq 70\%$ (old GOLD guidelines) 3) $FEV_1/FVC < 70\%$ and $50 < FEV_1 < 80\%$ (old GOLD guidelines) 4) $FEV_1/FVC \geq 70\%$ 5) $FEV_1/FVC < 88\%$ predicted (males) and $FEV_1/FVC < 89\%$ (females) (ERS).(75)

A more recent study (2007) compared the criteria using the Lower Limit of Normal (LLN) as suggested by the ATS/ ERS Task Force, to the GOLD criteria in a general population study (aged $\geq$ 40 years). (76) The LLN uses a cut-off value for the FEV_1/FVC ratio set at the fifth percentile of the normal distribution. The prevalence of COPD using the GOLD defined criteria was 14.2% compared to prevalence of 9.0% using the LLN criteria.(76)

In a general population study that recruited participants with a smoking history of greater than 100 cigarettes (aged $\geq$ 36years), COPD was diagnosed based upon lung function measured by spirometry.(77) By taking into account smoking history, the authors found that the prevalence rates were lower than expected: 11.6% for men and 4.8% for women.(77)

In a meta-analysis of prevalence statistics, COPD was stated as being between 4 and 10%.(78). However the values reported within the study are conservative with higher estimates of prevalence of over 20%,(73) being reported in at least one of the studies that

---

[1] Cough with phlegm, on most days for three months of the year, over not less than two years.

was reported on within the meta-analysis. In all the studies that make use of the GOLD definition, the COPD prevalence estimate was up to 28% of the over 40 population: a value that is much larger than reported in the meta-analysis.(78)

BOLD

The Burden of Obstructive Lung Disease (BOLD) study was set up to measure the prevalence of COPD and its risk factors, in order to determine the burden of COPD, in various countries around the world using standardised techniques.(79)

BOLD employed the GOLD diagnostic criteria to identify COPD cases, however in the reporting that followed, GOLD stages II and above (FEV_1/FVC<0.7 and FEV_1<80% predicted) were frequently used to identify disease. Doing so reduced disease prevalence estimates by about a half.(80;81) The first group to publish, reporting on COPD prevalence in Salzburg, Austria, found that overall prevalence in a random sample of the general population (aged ≥ 40 years) was 26.1%,(82) but prevalence was 10.3% when criteria of GOLD stages II and above were applied.(80) This compares to a doctors diagnosis of COPD for 5.6% of the participants.(82)

2.5 Summary

COPD is a chronic respiratory disease that principally affects lung function. Lung function impairment is used to identify the disease and to define disease severity. The disease is manifested with respiratory symptoms such as breathlessness and wheeze and also through exacerbations where major exacerbations end in hospitalisation. Smoking is an important risk factor for COPD.

COPD prevalence estimates for Europe are between 4 and 10% of the population but the prevalence of physician diagnosed COPD is less and represents approximately 1% of the population, leading to the opinion that there is under diagnosis of the disease. Compared to all other illnesses/disease, within Europe, COPD mortality accounts for the fifth largest cause of death and in the Western Pacific region COPD is responsible for a larger proportion still, with COPD as the second highest cause of death and 14% of deaths being attributable to the disease. The high proportion of deaths attributable to the disease within the Western Pacific region are largely explained by smoking patterns, and with smoking rates in this region continuing to be high, the considerable burden of the disease looks set to continue.

COPD represents a large and increasing health and financial burden on both the individual and on the health care provider as the disease worsens over time. Exacerbations in particular cause cost implications for the provider, particularly major exacerbations where the patient is hospitalised.

Current guidelines are largely in agreement over treatment for the disease and recommend an additive approach as the disease worsens and to treat specific elements of the disease including respiratory symptoms and exacerbations.

Until recently there has been little hope for COPD patients because of the widespread belief that nothing could be done for these patients. Clinical trials that have been published over the past ten years have brought evidence on new treatments which have been shown to bring benefits to this population. The licensing of these new products, particularly combination products, has caused a changing mix of treatments.

These new pharmacological products successfully entered the market and currently enjoy large market shares, to the detriment of market share for older drugs. This changing mix of

pharmacotherapy caused an increase in costs for treatment for COPD, with a potential seven times increase in drug costs over a seven year timeframe. Costs of other forms of care are more difficult to quantify. Hospitalisation costs will increase with increasing prevalence of COPD. Nevertheless, costs of care at the patient level, particularly in secondary care, might be expected to decrease with improved pharmacotherapy; because of a potential reduction in the incidence of exacerbations. The benefit of a new drug treatment relative to its cost is of particular relevance within this thesis. Details of methods to evaluate treatments are discussed in further detail within Chapters 3 and 7.

The development of a number of COPD guidelines, each with their own diagnostic criteria for the disease, have contributed to the changing pattern of care for COPD as the clinical community seeks to define appropriate care for this challenging patient group. However, prevalence estimates fluctuate heavily depending upon the diagnostic criteria employed to identify COPD cases. For example, spirometry alone is often used to identify COPD cases, resulting in large prevalence estimates, yet many descriptions of the disease describe it as resulting from smoking and presenting via respiratory symptoms. However, no studies were identified that compared prevalence from a spirometry based diagnostic criteria of COPD, with that of a diagnostic criteria based on spirometry and symptoms and a risk factor (such as smoking history), such as that proposed by NICE. A disparity in the identification and diagnosis of COPD therefore exists because criteria based upon lung function, symptoms and smoking history is likely to be the approach used by a physician, yet a diagnosis based upon spirometry alone dominates the literature. In the one study that compared doctors' diagnoses to the GOLD diagnostic criteria, prevalence estimates were five times higher applying GOLD criteria compared to the doctors' diagnoses (26.1% vs 5.6%). In reality the actual prevalence is probably somewhere in between these two values. However, taken with the finding that the GOLD definition always produced the highest estimates of COPD prevalence in the papers reviewed, it may be the case that applying

GOLD criteria overestimates prevalence of COPD in the general population from the perspective of the health provider. This question is raised and examined in detail within Chapter 5.

Chapter 3. Health Economics

In this chapter, the use and need for economic analysis within health care is discussed before the branch of health economics, economic evaluation, which is focussed specifically on the evaluation of different treatments, in order to inform health care providers about the relative value of competing courses of action, is explained.

In developing an economic evaluation for a treatment, it is important to fully understand the disease under study including how the disease presents itself, how the disease changes over time and how treatment affects the disease. The purpose of the previous chapter was to ascertain information from the published literature around the natural history of COPD and on treatments for the disease in order to begin the conceptualisation around an economic evaluation for COPD; this chapter on health economics presents information on the key elements required for the economic evaluation of any treatment or intervention and these key concepts are drawn upon throughout the thesis.

Specific topics are explored including a general overview of economic evaluation, the types of evaluation available, different structural models that can be used in which to develop an evaluation, costs and cost data, outcome measures and issues around uncertainty. Specific interest is paid to outcome measures and in particular, the measurement of utility.

3.1 Introduction

When the UK NHS was founded in 1948 it was built upon three core principles: the NHS would aim to meet the needs of everyone, be free at the point of delivery and be based on clinical need, not ability to pay.(83) These founding principles still apply today. Every year the NHS is allocated a sum of money and it is up to budget holders to

allocate this money in an appropriate way in order to fulfill the founding principles. Because resouces are finite, difficult decisions are constantly being made including which treatments or procedures to finance (and to whom) and which treatments should be replaced.

The budget holders seek to maximise population health, represented by the notion of utility (a measure of the relative satisfaction or quality of life gained), subject to the budget constraint for health care. In a world of perfect information, the combination of goods and services that results in the highest utility gain for society is selected subject to the fixed budget. However perfect information is a notion of the text books and in reality decisions must be made based on available, imperfect information. These choices have to be made, and have in the past been made based on 'gut feeling', 'what we did last time' and 'educated guesses'.(84)

Economists can assist decision makers with these difficult decisions via the use of economic evaluation where the value of one treatment in terms of its costs and effects is compared to the costs and effects of another treatment, usually current practice. economic evaluation has been described as:

> 'the comparative analysis of alternative courses of action in terms of both their costs and consequences'.(84 p9)

Hence the concept of opportunity cost, which is the cost of foregoing the next best choice, is of fundamental importance within economic evaluation. The true cost of something is what is given up to get it. This includes not only the money spent in buying the something (treatment/procedure), but also the economic benefits (utility) foregone because of buying that particular something and one can therefore no longer buy something else. Everything has an opportunity cost. Where one drug is accepted for use within the NHS, other drugs or treatments within the system are displaced. When a treatment is evaluated in an economic evaluation, the opportunity cost is often considered to be the value of the current treatment.

Economists seek to make explicit one set of criteria that can be used to decide between different uses for scarce resources.(84) In this way, decisions about which treatments to accept and/or reject are made based on criteria of value rather than on a value judgement.

Many countries around the world include a role for the incorporation of economic evidence into the decision making process for health, including many European countries, Australia and Canada.(85) In England & Wales, the relevant decision making agency is NICE who is an organisation independent from Government, responsible for providing guidance on the use of health technologies and the implementation of public health programmes. In the process of developing guidance, NICE brings together all the available clinical and economic evidence in order to decide whether the adoption of the technology (drug or treatment) represents good value for the NHS.(86)

The broad content of an economic evaluation should be similar regardless of the disease of interest and the treatment(s) under study. A number of different checklists have been proposed for assessing economic evaluations.(84;87;88) Most well known within the field of health economics is the ten point checklist proposed by Drummond et al which is reproduced below in table 3.1.(84) The checklist ensures the researcher ascertains whether various desirable elements have been included within the economic evaluation including, that the aim of the evaluation is clearly stated, the inclusion of a statement about which treatments are to be compared, that the costs and effects of the different treatments under study are established from appropriate data sources and a comparative analysis has been made, and that consideration has been given to uncertainty in the parameter estimates.

Table 3.1 Ten point checklist.(84)

Questions
1) Was a well defined question posed in answerable form?
2) Was a comprehensive description of the competing alternatives given? (that is, can you tell who did what to whom, where and how often?)
3) Was the effectiveness of the programmes or services established?
4) Were all the important and relevant costs and consequences for each alternative identified?
5) Were costs and consequences measured accurately in appropriate physical units (for example, hours of nursing time, number of physician visits, lost work days, gained life years)?
6) Were costs and consequences valued credibly?
7) Were costs and consequences adjusted for differential timing?
8) Was an incremental analysis of costs and consequences of alternatives performed?
9) Was allowance made for uncertainty in the estimates of costs and consequences?
10) Did the presentation and discussion of study results include all issues of concern to users?

NICE has defined a 'reference case' in which submissions to the Institute should follow to ensure consistency of health technology assessments (HTA)/economic evaluations and for these to be in keeping with the NHS objective of health maximisation under a limited budget,(87) and is reproduced in table 3.2.

Table 3.2 NICE reference case.(87)

Element of HTA	Reference Case
Defining the decision problem	The scope developed by the Institute
Comparator	Therapies routinely used in the NHS, inc technologies regarded as best practice
Perspective on costs	NHS and Personal Social Services
Perspective on outcomes	All health effects on individuals
Type of economic evaluation	Cost effectiveness analysis
Evidence on outcomes	Based on a systematic review
Measure of health effects	QALYs
Source for HRQoL	Reported directly by patients and/or carers
Source for valuing HRQoL	Representative sample of the public
Discount rate	An annual rate of 3.5% on costs and QALYs

NICE is concerned with comparisons across studies and as described in table 3.2, normally requires that outcome measurements are in terms of Quality Adjusted Life Year (QALYs) (discussed in detail in section 3.2.5). The source for valuing QALYs should be that of the general population (rather than the patient population), since the NHS is acting on behalf of the general population in allocating resources for health.

Whilst the ten point checklist and the NICE reference case were designed with different uses in mind, they are both used to assess economic evaluations. Of note is how the focus has shifted between the older ten point check list and the NICE reference case. A basic methodological focus is apparent within the ten point check list that concentrates on where costs and outcomes come from and how the outcome and costs were elicited and valued. These issues are less of a focus in the NICE reference case as methodologies used to assess costs and effects have become increasingly established. The focus within the NICE reference case ensures comparisons across disease areas can be made.

3.2 Key Concepts within Economic Evaluation

Within the rest of this chapter, the subject of economic evaluation is explained and specific topics explored, including: the decision problem, different types of evaluation, economic evaluation structures, cost, outcome measures including utility and issues around uncertainty, in order to introduce topics and concepts that will be drawn upon throughout the thesis. There is a specific interest in outcome measurements and in the measurement of utility. The subheadings of the NICE reference case are loosely followed so as to provide a framework for describing the important concepts within economic evaluation.

3.2.1 The Decision Problem and Comparators

An economic evaluation should begin with a clear statement of the decision problem.(87) Including details on the technologies to be compared and the relevant patient population. The statement could read along the lines of:

> "The aim of the study was to compare the cost effectiveness of drug X to drug Y in a COPD population from the perspective of the UK NHS"

An economic evaluation may look at different populations or subpopulations depending on the aim of the study. It is important to consider any subgroups in order to permit, where evidence allows, the identification of any specific group of patients to whom the technology is particularly cost-effective.

Most published guidelines for economic evaluations assert that the comparator of interest is current treatment.(89) It is important to choose an appropriate comparator because the comparative nature of economic analysis means that an inappropriate comparator can bias an analysis and render it of little value for decision making. A significant issue in economic evaluation is that because the primary endpoint of RCTs (to which economic evaluations are frequently piggybacked onto) is drug registration, trials often provide relative treatment effects compared to placebo only. This is potentially a fundamental problem of conducting economic analyses alongside clinical trials.(90)

3.2.2 Types of Economic Evaluation

An economic evaluation can either be cost effectiveness, cost utility analysis, cost minimisation or cost benefit analysis. A cost benefit analysis measures the value of an intervention by a monetary value such as the dollar, or the Euro, and compares it to the costs of providing the treatment. The decision criterion as to whether or not to adopt the technology is based upon whether the benefits are greater than the costs subject to budget limitations. Where the alternatives under consideration have equal benefit or effect, then it is possible to choose the optimal treatment based only upon their cost; this method is known as cost minimisation. However, this scenario occurs infrequently because more often than not, different treatments will cause different effects and unless this is the case, cost minimisation is inappropriate.(91) More useful are cost effectiveness and cost utility analyses, where cost utility is a special case of cost effectiveness analysis. This type of analysis allows a full economic evaluation to take place where both the costs and outcomes are analysed. The main difference between

them lies in the outcome measure. Cost effectiveness studies present results in terms of natural units such as the cost per exacerbation avoided,(92) or improvement in health status.(93) A cost utility evaluation identifies the change in health status measured by cost per unit of utility (usually the QALY). Utilities and the QALY concept are described in detail in section 3.2.5. The terms cost effectiveness and cost utility are used interchangeably within the thesis.

Presenting cost-effectiveness analysis in natural units, such as cost per exacerbation avoided, will potentially limit the scope for decision-making on the efficient allocation of resources between COPD and other diseases. Therefore, in order to inform the choice of whether to allocate more health care resources to the treatment of one disease compared to another disease area, a generic outcome using cost utility analysis is required, as this enables a comparison both within and across disease areas to be made. Both NICE and the US Public Health Service via the US panel, recommend cost-utility analyses as the appropriate way to make comparative assessments of value for money within a health system.(93-95) Nevertheless, the US tends to make decisions on the use of drugs based on 'medical appropriateness' without taking into account cost. For example, Medicare does not consider costs when making decisions about coverage.

3.2.3 Economic Evaluation Structures

Economic evaluation can take a number of different structures including: alongside a clinical trial and within a modelling framework: a decision tree, a Markov model, a simulation model and regression modelling. Economic evaluation alongside a clinical trial is used to evaluate the cost effectiveness of different treatments within the clinical trial itself whereas a modelling structure aims to represent the disease under study and the effects of treatment on the disease in an appropriate way. These structures can also be combined, for example regression equations can be used to predict values for an economic evaluation alongside a clinical trial or could be used within a Markov model to inform a specific parameter of the model.

There are at least two ways of developing an economic evaluation alongside a clinical trial. The first method is, for each treatment group, to calculate the area under the curve (AUC) in order to derive the mean QALY score (described in detail in Chapter 6), and then calculate mean costs for each treatment group. The difference in costs divided by the difference in effects can be calculated in order to derive a cost effectiveness statistic. This method does not allow for patient heterogeneity such as differences in age and sex, which may influence subgroup cost effectiveness. The second method uses regression equations to fit models in order to predict the HRQoL score (usually QALYs) with explanatory variables including patient characteristics and treatment group. The same method is used to derive a prediction equation for costs. Cost effectiveness can be predicted by treatment group but also for different patient subgroups. Whilst included here for completeness, economic evaluation alongside a clinical trial is not the same as economic modelling for several reasons. Whilst there are benefits to conducting economic evaluation alongside clinical trials due to the substantial amount of patient level data on costs, effects and key outcomes, there is a lack of external generalisability because the trial participants and treatment may not be representative of those found in a 'real life' setting. Importantly the timeframe of the economic evaluation is restricted by the length of the clinical trial, which usually runs for one year or less. Health outcomes typically take longer to manifest than the duration of the trial allows.

More useful for decision makers is the use of economic models. Health Economists use economic models to represent the real system in a simplified way. Disease processes can be highly complex and economic modelling provides a way in which key information on the natural history of the disease and on the effect of treatment can be combined in a potentially useful and meaningful way so that decision makers can make reimbursement decisions based on all the available and relevant information.

> "The purpose of a model structure is to characterise the consequences of
> alternative options in a way that is appropriate for the stated decision
> problem and boundaries of the model"(96)

The different model structures that can be used for conducting economic evaluations are described below together with some pros and cons of selecting that type of modelling structure.

Decision Tree

Decision trees offer one approach to modelling a disease. A decision tree represents individuals' possible prognoses following an intervention, by way of different pathways,(84) as shown in figure 3.1. The tree is characterised by decision nodes (the square boxes in figure 3.1) representing the decision being addressed in the model such as treat/no treat and chance nodes (the circles in figure 3.1) representing uncertainty. Connecting the various nodes to outcomes are branches. The branches have probabilities attached that represent how likely it is that the event of interest occurs. Each way through the tree represents a pathway. In figure 3.1 there are 7 pathways. Each pathway has a probability attached to it, a cost and an effect value and from these, expected values for each decision can be determined.
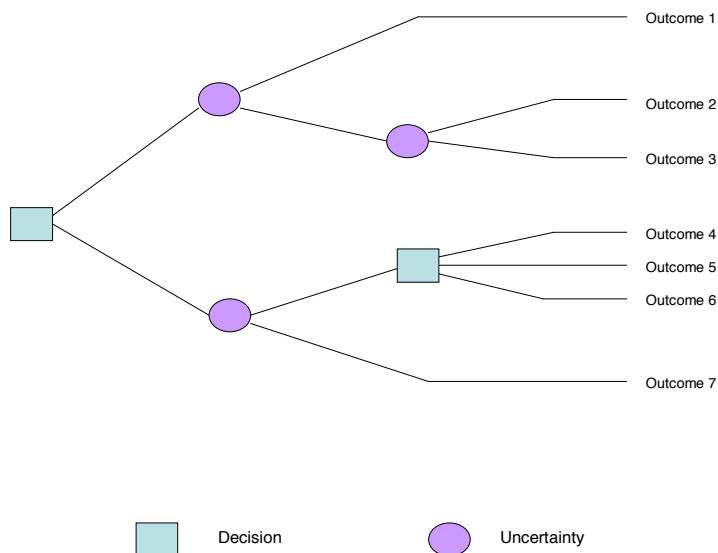
Figure 3.1 An example of a decision tree

Popular in the past, decision trees are rarely used nowadays in favour of other methods. The trees can become very bushy particularly in modelling chronic diseases, making them unwieldy and complex to programme and analyse.(84) In addition, and perhaps more importantly, the decision tree cannot incorporate time dependency. Time dependency is often a feature with modelling disease, particularly chronic conditions such as COPD where the disease tends to worsen over time. Another related problem is that the change in age related mortality cannot be modelled in this framework, nor can discounting occur (which is described in section 3.2.8). Sometimes decision trees are used within a Markov framework.(97)

Markov Model

Markov models have frequently been used in the economic evaluation of COPD. Markov modelling forces the characterisation of a disease into a number of discrete health states through which the disease will progress. The classification into a particular state is dependent on the natural history of a disease. In COPD, patients are frequently classified into a particular state depending upon FEV_1 % predicted. Movement between the states is dependent on decline in lung function and can be estimated in a number of ways such as with data from clinical trials or using a regression equation to predict the transition.

Markov models have cycles with durations that might last one day, one week, one month or one year, depending on the nature of the disease. At the beginning, middle or end of each cycle (subject to the modellers approach) the patient cohort moves through the model. The subjects being modelled move through the model in a number of ways, typically they either stay in the same state, move to a worse state (or a state in which a specific event occurs), or they die. The models extrapolate these cycles for a certain timeframe ie 30 yrs after which time the cohort is expected to have all but died and the marginal costs and effects are negligible. Costs and utilities are attached to each disease

state within the model. Events occurring within disease states such as exacerbations can be modelled by attaching (dis)utilities and costs, weighted by event rates.

A particular feature of the Markov model is that it is memory-less. That is, once a person has entered a state, the model cannot distinguish between them and others who were and remain in that state. A problem with this is that time, especially with chronic diseases, is often strongly correlated to worsening disease and that through the Markov approach, this way this correlation is ignored.

Discrete Event Simulation

Discrete Event Simulation (DES) contains three components: entities, events and time. Entities are usually the patients and are assigned attributes such as age, sex, type of disease. An event is something that occurs during the simulation, such as an exacerbation, and events can occur sequentially and/or simultaneously. The rate at which the event occurs can change and be dependent on any of the attributes of the entity. Time is an explicit element of a DES simulation and the model can be run for as long as is necessary.(98) DES has been described as:

> "…discrete event simulation models the pathway of an individual by sampling probabilities from an a priori distribution."(98)

DES is most successfully applied to processes such as modelling the introduction of new equipment within the setting of a GP clinic or for queuing. Another situation where DES is typically used is in infectious disease modelling where people with the disease infect others through interactions with them. DES explicitly allows for this interaction between patients.

A problem with DES is that the processes can get very complex, whilst the fundamental idea behind economic modelling is to present the disease in a simplified format and at

the same time, capturing the important elements that drive cost and effect: caution needs to be made to ensure that the model does not over-complicate the processes involved.

Regression Equation

Regression equation modelling can take a number of forms but principally involves the use of one or more prediction equations that predict the value of a dependent variable from a range of independent variables using patient level data such as a RCT or an observational dataset. Some authors have used a prediction equation for one parameter within a Markov model, for instance, cost, the probability of survival over time and transition probabilities (see Chapter 4 for details of these models). Glick et al have proposed the use of prediction equations for cost and effect based on patient level data within RCT data that when combined give a cost effectiveness statistic.(99) The use of regression equations to depict cost and effect, albeit outside of a modelling framework, dates back to at least 1977 where Weinstein and Stason represented net health care costs by the following expression:

$$\Delta C = \Delta C_{RX} + \Delta C_{SE} - \Delta C_{Morb} + \Delta C_{RX\Delta LE}$$

Where ΔC_{RX} includes all direct medical and health care costs, such as the cost of hospitalisation, medications and physician time, ΔC_{SE} comprises all health care costs associated with side effects occurring as a result of treatment, $-\Delta C_{Morb}$ is comprised of the savings from preventing or alleviating disease by improvements in morbidity and $\Delta C_{RX\Delta LE}$ is the cost of treating disease or illness that occur with the patient living for longer than otherwise. Weinstein and Stason represented the effectiveness parameter by:

$$\Delta E = \Delta Y + \Delta Y_{Morb} - \Delta Y_{SE}$$

which essentially represents a QALY, where ΔY represents the unadjusted number of life years gained following treatment, ΔY_{Morb} representing the adjustment to take into account quality of life and $-\Delta Y_{SE}$ factors in side effects of treatment.(100) Regression based modelling can be used to inform part of the economic evaluation or it can be used to inform the entire economic evaluation, if the model is built as a series of related regression equations that when combined give cost and effectiveness values for different treatments.

The main benefits of using a regression based model are that patient level heterogeneity (described in section 3.2.9) can be captured in much more detail compared to a Markov model and a regression based model allows for greater flexibility in the structure of the model. The main limitation of an economic model based on regression equations is the necessity for good quality data, with a large sample size that can be used to study the natural history of the disease. However, if good quality data are obtained, the use of a regression based structure makes better use of the available data than other modelling structures as the regression equations used to develop the model are developed from the data. The ultimate choice of model structure should be based on the natural history of the disease and the aim of the study.

3.2.4 Cost

The choice of perspective on cost depends upon the target audience for the study.(88) There are several perspectives: a specific provider/provider institution (ie the NHS), the patient/patient group, a third party payer (ie an insurer) or the perspective of society.(84) The perspective used will determine the costs employed and these must be appropriate and as comprehensive as required.

A societal perspective would typically include productivity costs due to absence from/inability to work, cost of and to carers and additional costs of the illness to

patients, but could exclude productivity costs. Whilst productivity costs can represent a considerable burden on both the individual and on society, their use in cost effectiveness analyses is an area of controversy. This is partly due to methodological uncertainty concerning the appropriate method of measuring productivity losses, especially where significant unemployment is evident.(101)

The US Panel on cost-effectiveness analysis recommends a societal perspective which can include productivity losses.(94) In contrast, NICE limits its perspective to the health service and personal social services perspective and explicitly excludes the use of productivity costs in its evaluations. It is often argued however, that economic studies should include all relevant costs associated with the illness,(102) and several countries including Sweden and the Netherlands suggest the use of a societal perspective.

3.2.5 Outcomes

The outcome of interest in an economic evaluation for health care is the effectiveness parameter which is usually described in terms of a utility measurement. NICE advise that all health effects should be accounted for within an economic evaluation. Where there is a belief that other non-utility based measures capture health effects that are not captured within the utility measure, the effectiveness of treatment on these measures can also be presented and there are a number of non-utility based HRQoL measures which are routinely measured in COPD patients.

Evidence on outcomes should ideally come from a RCT. If a clinical trial is unavailable or where or alternative options have not been looked at, then evidence should come from a meta-analysis or an indirect treatment comparison analysis. To be of use, a RCT should be externally valid, so that the results of the study are applicable to the intended treatment population in a real life setting.

In valuing outcomes, the aim is to capture the benefit of treatment to the patient population and this is achieved through the administration of a tool directly to the patients that can capture these HRQoL benefits. For utility based measures, for example with the EQ-5D questionnaire (which is discussed in detail in Chapter 6, section 6.2) the utility values attributable to each patient are based on previously collected scores elicited from the values that a representative sample of the population place on different states of health. The use of patient level data weighted by the views of the general population is recommended by NICE and represents the idea that the NHS is acting on behalf of the UK general population to allocate resources in its collective best interest.

There are a number of non-utility measures of COPD health effects including: hard endpoint data, clinical measures, and health outcome measures from health outcome questionnaires. Hard endpoint data consist of: mortality (survival), hospitalisations, and number of exacerbations. Clinical measures such as FEV_1 and FEV_1 % predicted, and exercise tolerance are limited in their scope to address questions of health benefit to patients and may only be weakly correlated to actual health benefit as experienced by the patient. Health outcome questionnaires on the other hand, specifically lend themselves to monitor health benefit as perceived by the patient.

Health outcome questionnaires fall into two groups, generic questionnaires and condition specific measures. Over the last twenty or so years, both generic and condition specific measures have increasingly been used within clinical and observational studies,(103) in order to capture health benefits to patients, usually as a result of treatment or an intervention.

Generic Questionnaires

Generic questionnaires were first developed during the 70s in an attempt to capture the HRQoL of different people in different health states. Many generic questionnaires have been applied to COPD patients, of which the following list captures the most common.

- Euroqol instrument (EQ-5D).(48)

- Health utilities index (HUI).(104)

- Short form 36 questionnaire (SF-36).(105)

- SF-12.(106)

- SF-6D.(107)

- General health questionnaire (GHQ).(108)

- Dartmouth COOP charts.(109)

- Nottingham health profile.(110)

- Quality of well being (QWB).(111)

- Sickness impact profile (SIP).(112)

Of these, the EQ-5D, SF-6D and HUI can be used to derive a utility score. Generic measures can be used on a variety of disease areas or health problems and can be used to compare HRQoL across disease areas. Generic measures can pick up important changes in HRQoL in all spheres of the patient's health: representing quality of life ideals that are common to all, regardless of age, disease or treatment, and measure the functional and mental well being.(113) However generic measures may be insensitive to small changes in specific components of COPD such as breathlessness or dyspnea (which may cause significant impact on HRQoL for COPD patients) because of the breadth of health areas they cover: particularly over shorter time periods.(114) In

addition, they may not pick up on the specific limitations experienced by COPD patients.(114)

HRQoL Measures Specific to COPD

Condition specific measures for COPD were developed in order to address the issue of the HRQoL specific to the COPD patient. There are many condition specific measures. In addition, there are questionnaires that target specific aspects of COPD such as dyspnea and breathlessness. The most frequently used questionnaires are listed below

- St George's respiratory questionnaire (SGRQ). (47)

- Chronic respiratory disease questionnaire (CRQ).(115)

- Breathing problems questionnaire (BPQ).(116)

- Seattle obstructive lung disease questionnaire (SOLG).(117)

- Sino-Nasal outcomes test.(118)

- Changes in dyspnea:

  - MRC dyspnea scale.(119)

  - Baseline Dyspnea Index (BDI).(120)

  - Transitional dyspnea index (TDI).(120)

Disease specific measures may be more sensitive than generic measures because the content is directly relevant to COPD patients.(121) However, the results are not easily generalisable or comparable across different diseases. In addition, disease specific instruments cannot be used on populations who do not have the disease and they may not pick up on co-morbidities or health problems that are associated with unexpected effects from treatment.(122) Most Health Economists prefer the use of a generic tool for outcome measurement from which utility scores can be derived, such as the EQ-5D, the SF-6D or the HUI.

Utility

As previously stated, utility is a measure of the relative satisfaction or quality of life gained. In economics, the notion of utility is frequently used to compare goods or services against one another in terms of preference. This could be the consumption of an orange over an apple. If the consumer prefers oranges to apples, they will place a higher value on the consumption of an orange than an apple; utilities are used to quantify this preference. However not all consumption gives enjoyment, goods give the user a degree of enjoyment but they may also fill a human need or may even be a negative experience. Within the economic evaluation of different treatments, analyses are conducted that determine the cost effectiveness of one drug treatment compared to one or several comparators. In the same way that utilities can be used to value the consumption of oranges and apples, so too can they be used to value outcomes from drug treatment.

In order to measure preferences in health economics, cardinal preferences are measured, that is a number needs to be attached to the effectiveness or outcome from the treatment so that the strength of preference for the treatment can be measured. These numbers should be measured onto an interval scale which has equal intervals, so that a move from a value of 0.3 to 0.4 has the same worth to the individual as a move from 0.8 to 0.9.

Measuring Utilities

Two methods are used to derive cardinal utilities from the patient population: standard gamble (SG) and time trade off (TTO). Generic utility based questionnaires were developed using either the SG or the TTO method and offer an indirect meausurement of HRQoL in a convenient and time saving format.

Standard Gamble

The SG technique has its foundations in von Neumann utility theory.(123) Within the SG, the subject is presented with a health state describing a particular condition and what it is like to be in that state. They are offered two choices. Either they can stay in the state with certainty, or they can take a treatment which involves an element of risk. The treatment is said to work immediately and either returns the subject to full health with a probability of p, or the subject dies with a probability of 1-p. The probability is varied until the subject is indifferent between the two alternatives.

Time Trade Off

Within the TTO technique, the participant is presented with a health state and they are again given two choices. This time the choice is between remaining in the state with certainty for a period of time t, such as ten years, followed by immediate death, or they are offered the choice of fewer years of life y in which they live in a state of full health, such that y<t, follwed by immediate death. The time y is varied until the subject is indifferent bewtween the two alternatives.

Generic Utility-Based Questionnaires

Generic health-related questionnaires are used to derive utilities and include the EQ-5D, the SF-6D and the HUI. These questionnaires are most freqently used to derive utilities because they are comparatively easy to administer and the results are easily generalisable across disease areas.

Quality Adjusted Life Years (QALYs)

In order to make resource allocation decisions across disease areas, a generic form of outcome measure that is measured over time is useful. The QALY is grounded in utility theory. In essence, a new treatment that leads to improved quality of life over a time period, compared to existing treatment, will generate greater utility than if the patient continued to take the existing treatment. The QALY quantifies changes in utility over

the life of the patient. The QALY has two components; quality and quantity of life. NICE recommends the use of QALYs within cost effectiveness analysis:

Measures that capture both quantity (duration) of life and quality of life… are best suited for use in a reference case analysis."(94)
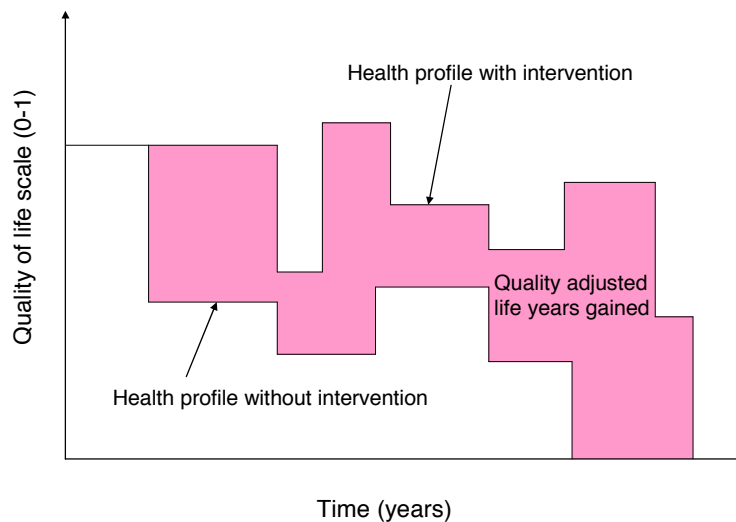
Figure 3.2 QALY profiles for a hypothetical patient with and without treatment
Profiles weight length of life by quality of life on a zero-one scale where one represents perfect health and zero death. In a chronic disease such as COPD, exacerbations might result in reduced quality of life for a period of time with incomplete recovery.

In figure 3.2 intervention 1 is associated with a higher quality of life (the curve is above the other) and, greater life expectancy (quantity) than the treatment with intervention 2. The shaded area between the two curves represents the difference in the number of QALYs between the treatments.

For example, a treatment may improve quality of life over the remainder of the patient's life by 0.03 QALYs, from 0.6 to 0.63 and extend the life of the patient from 10 years to 11. This will give:

New treatment	Existing treatment

10 years at 0.63 = 6.3   10 years at 0.6 = 6

Plus one extra year of life = 0.63

QALYs = 6.93   QALYs = 6

The new treatment will add a total of 0.93 QALYs compared to the existing treatment.
The QALY is by far the most accepted health-related utility measure and is the
preferred outcome measure in many countries including Canada, New Zealand, Sweden,
England and Wales (NICE HTA guidelines), the Netherlands and the US(85).

3.2.6 Incremental Cost Effectiveness Ratio (ICER)

The cost effectiveness statistic derived within economic evaluations is the incremental
cost effectiveness ratio (ICER). The ICER represents the difference in costs and
benefits between a new therapy (x) compared to current treatment (y). The ICER is
calculated using the following formula:

$$ICER = \frac{Cx - Cy}{Bx - By}$$

Using the previous example of a gain of 0.93 QALYs following treatment with
intervention 1, with a treatment cost of £15 000 over the remainder of the patient's life,
compared to the cost of the existing treatment at £1000, gives an ICER of:

$$\frac{15\ 000\text{-}1\ 000}{6.93\text{-}6} = £15\ 054$$

The ICER quantifies the additional cost of an extra QALY by taking the new treatment
rather than the existing one: the new treatment can provide one extra QALY for a cost
of £15 054.

Because the reimbursement of one treatment will displace monies spent from other
treatments within the health system, it is important to ensure that the treatment

reimbursed provides sufficient value for money. A 'guide price' or a threshold from which to decide whether or not any one treatment should be reimbursed by the health care system is required. In the UK, NICE have a threshold of between £20,000 and £30 000 per QALY, but this can vary. Below £20,000 per QALY there is a high probability of the technology being accepted and above £30 000 per QALY there is less chance of the technology being accepted.(87)

3.2.7 Timeframe

Economic evaluations attached to clinical trials are naturally constrained by the timeframe of the trial to which they are connected. Economic models have an advantage in that through applying adequate and transparent assumptions, the results can be extrapolated into the future, up to a lifetime timeframe. The timeframe of a study is crucial because the full benefit of treatment may not occur within the period of the trial: the timeframe of the model should extend far enough into the future so that the key differences between the comparators in the analysis can be established.(88) Restricting the timeframe may yield un-meaningful results:

> "There is no natural interpretation for life-years gained during a finite period of time, and the CE ratios that result from using different time horizons, such as one year and five years, cannot be compared in any meaningful way…researchers who truncate their analyses have made, perhaps unwittingly, the implausible alternative assumption that study subjects experience neither the costs nor the benefits of living beyond the period of study".(124 p191-192)

3.2.8 Discount Rate

A discount rate is often applied to cost effectiveness analyses in order to represent the fact that immediate health (or financial) gains are more highly valued in the present than in the future. The discount rate is currently set at 3.5% by NICE for both costs and effects though this value has not always been applied and there is still uncertainty

around the absolute value of the discount rate. It is common to see scenario analysis around the discount rate in order to assess the effect of different rates on the cost effectiveness result.

3.2.9 Uncertainty

Five different types of uncertainty exist in economic modelling: patient to patient variability, parameter uncertainty, structural uncertainty, heterogeneity, and decision uncertainty.

Of these, heterogeneity refers to differences between individuals that can be explained and as such, is a source of uncertainty that it is possible to reduce to some degree by accounting for patient level characteristics such as age and gender. Variability on the other hand, refers to differences between patients occurring for reasons that are unknown or that can not be captured, and within analyses, variability is covered within the error term.

Parameter uncertainty occurs around inputs into the model such as utilities, resource use, costs, event rates, and so on. Models have in the past used data deterministically, ie one value (usually the mean) is employed within the analysis and the uncertainty around the mean values was not accounted for. The use of Probabilistic Sensitivity Analysis (PSA) addresses the issue of parameter uncertainty by placing distributions around the parameter. Usually Beta distributions are applied to transition probabilities and to utilities, and gamma distributions are applied to cost parameters. The Cholesky decomposition can be used to keep the covariance structure between regression results in place whilst accounting for uncertainty in the model parameters.

The decision around the type of modelling structure to employ (or which mixture) is largely informed by the natural history of the disease in question. For instance infectious diseases are more successfully modelled using DES, whereas for chronic

conditions, Markov models are used to the same effect. Regression based modelling for cost effectiveness can be applied so that key drivers of the disease can be explicitly modelled. Often the chosen modelling structure is informed by historical models, carried out in the same disease area and this is an important starting point to building a new economic model, however this may lead to inflexibility in structure. It is often therefore useful to look to models carried out in other (similar) disease areas to investigate alternative approaches to model structure.

Where analyses are conducted using patient level data, a statistical analysis of the data should incorporate measures of uncertainty around cost-effectiveness results. If the study uses a modelling framework, sensitivity analysis, ideally PSA should be employed,(125) this allows the combined uncertainty of all the parameters in the model to be included,(126) by using the full probability distributions of each input into the model, rather than just the point estimates. A number of best practice guidelines (eg the BMJ, NICE and US Panel) state that the uncertainty surrounding estimates of cost effectiveness needs to be explored when presenting economic evaluation results.(95) PSA is now a formal requirement for cost effectiveness models submitted to NICE.(95)

Sponsorship

Industry sponsored studies represent a major source of funding for economic evaluations, however, concerns have been raised regarding potential biases arising as a result of this alliance because studies have found an association between pharmaceutical sponsorship and the probability of a favourable result.(127-129)

3.3 Summary

This chapter introduced the subject of health economics together with specific details about economic evaluation. Economic evaluation consists of a number of key concepts and within this chapter a focus has been given to: the types of economic evaluations, the

structure of model employed, cost, outcomes and utility, the ICER and the issue of uncertainty.

The aim of the economic evaluation should be clearly stated and the comparators considered should include current practice. There are a number of model structures that can be employed including Markov models and regression based models and it is important to select a structure that is appropriate for the disease under study and the question or aim of the economic evaluation. In estimating costs, all relevant costs should be considered and from the UK perspective this should represent all costs to the NHS and social services. The measuring of utility is a key component of economic evaluation and the QALY has been proposed as the appropriate measure of HRQoL for valuing treatments. Costs and QALYs from two different treatments can be combined in order to generate a cost utility value, the ICER. Decision rules exist where treatments are accepted if the ICER value falls below, between or above a documented threshold. Uncertainty is present in a number of forms including within parameter estimates, in the choice of model structure and between patient heterogeneity. Uncertainty should be considered in all economic evaluations, preferably using PSA. The timeframe of the study should be extended to lifetime in order to capture the full benefits of treatment.

In the following chapter these key concepts of economic evaluation are considered with reference to published economic evaluations in COPD, in order to assess the extent to which existing models meet current standards.

Chapter 4. Literature Review of Economic Evaluations in COPD

This chapter presents a review of the published literature on economic evaluations for COPD treatments. The focus is on treatments where competing courses of interventions exist for COPD patients and includes pharmacological and surgical treatment.

The review is split into two sections with the first section focusing on economic evaluations that have been conducted either alongside a clinical trial or an observational study, and is structured around the key concepts of economic evaluation identified in chapter 3: perspective, patient group, comparators, outcome measures, extrapolation, the results of the evaluations and the handling of uncertainty. The economic evaluations that develop an economic model are examined further in the second section and include, in addition to those key concepts mentioned above, a description of the model structure and design, sources of input data and assumptions used within each model. First the search strategy is presented.

4.1 Search Strategy

The focus of the search was on pharmacological therapies that are routinely used in the treatment of COPD, as described in table 2.5. Surgical interventions were included because surgery represents an alternative course of action to continued treatment. Other treatments such as smoking cessation, antibiotics, vaccinations and oxygen therapy are prescribed in addition to routine pharmacological treatment and as such would not be expected to affect the place in the treatment pathway of a routine pharmacological treatment. As such, these treatments are excluded from the scope of this chapter.

Four databases were used to search the literature for economic evaluations on interventions for COPD: MEDLINE, EMBASE, the Centre for Reviews and

Dissemination (CRD) and the Health Economics Evaluations Database (HEED) and restricted from 1990 until 2007. The search strategy was designed for MEDLINE and also applied to EMBASE. Included terms were: "Lung diseases, obstructive/ or bronchitis/ or pulmonary disease, chronic obstructive" or "COPD", and "costs and cost analysis"/ or "cost benefit analysis"/ or "cost of illness"/ or "health care costs"/ or "health expenditures" and "quality adjusted life years/ or cost utility" or "health adj4 utili#ation" or "economic$ or economics."

For CRD and HEED – where this type of search strategy was not possible to apply – "cost" or "effectiveness" and "chronic" or "COPD" and the names of individual drugs were used: "fluticasone", "salbutamol", "ipratropium", "formoterol", "tiotropium" etc. MEDLINE gave 732 hits; EMBASE = 159; CRD = 235 and HEED = 53. Removing duplicates, a total of 918 papers were found (March 2007). Reference lists from key papers were also searched for studies and papers: relevant papers were selected by first reviewing the abstracts and then if deemed appropriate, the papers were obtained. Papers satisfying the inclusion and exclusion criteria were selected. Inclusion criteria were: economic evaluations of pharmacotherapy or surgery for COPD and in the English language. Exclusion criteria included papers where the focus was upon: cost alone, antibiotics, smoking cessation, vaccinations and oxygen therapy. Fifteen pharmacoeconomic evaluations of COPD were identified and five papers evaluating surgical interventions. The search was updated in October 2009 and identified a further six pharmacological studies and three surgical studies, totalling 29 economic evaluations.

Table 4.1 Summary of COPD economic evaluations

First author	Ref	Yr	Type		Structure			Outcome measure				
			UA	EA	RCT	Obs	Model	QALY	Surv	SGRQ	Exac	FEV$_1$
Jubran	(97)	93		■			■					
Ramsey	(130)	95	■			■		■				
Rutten-v M	(131)	95		■	■				■			■
Al	(132)	98	■				■	■				
Friedman	(129)	99		■	■							■
Van den Boom	(133)	01	■		■			■				■
Anyanwu	(134)	02	■			■		■	■			
NETT	(135)	03	■		■			■				
Ayres	(136)	03		■	■						■	■
Hogan	(137)	03		■	■					■		■
Jones	(93)	03		■	■					■		■
Borg	(138)	04	■				■	■				
Groen	(139)	04	■				■	■				
Oostenbrink	(140)	04		■	■					■	■	■
Sin	(141)	04	■				■	■	■			
Gagnon	(142)	05		■		■		■				
Lofdahl	(143)	05		■	■						■	
Oostenbrink	(92)	05	■				■	■		■		
Spencer	(144)	05	■				■	■		■		
Briggs	(145)	06	■		■			■	■			
Maniadakis	(146)	06	■		■							
Ramsey	(147)	07	■		■			■				
Rutten-v M	(148)	07	■				■	■			■	
Chuck	(149)	08	■				■	■				
Earnshaw	(150)	08	■		■			■	■			
Najafzadeh	(151)	08	■		■			■			■	
Blough	(152)	09	■		■			■				
Briggs	(153)	09	■				■	■				
Oba	(154)	09	■				■	■				

UA=cost utility analysis. EA=cost effectiveness analysis. RCT=randomised controlled trial. Obs= observational. QALY=quality adjusted life year. Surv=survival. SGRQ=St George's respiratory questionnaire. Exac=exacerbation. FEV$_1$= Forced Expiratory Volume in one second

Table 4.1 provides a summary of the features of each of the 29 economic evaluations, including the type of study (cost utility analysis or cost effectiveness), the structure of the study (RCT/observational or model) and the main outcomes measures used (QALY, survival, change in SGRQ, rate of exacerbations and of change in FEV$_1$ score. Approximately half of the evaluations were conducted alongside clinical trials (13/29); three used data from observational studies and thirteen used an economic model. Evaluations were either cost utility (19/29) or cost effectiveness analyses (10/29).

The QALY was the most frequently used outcome measure but there were others, including: survival, reduction in rate of exacerbations, improvement in SGRQ and improvement in FEV$_1$. There has been a rapid increase in the number of economic evaluations published within the disease area over recent years, with 22/29 studies being published within the past six years, and in which an increasing focus on modelling and the use of the QALY as an outcome measure was seen.

4.2 Review of the Non-Modelling Papers

In total, sixteen economic evaluations were non-modelling papers, of which thirteen were conducted alongside a RCT and three used data from an observational study. Of these studies, five assessed the cost effectiveness of lung surgery, three of which were based on the national emphysema treatment trial (NETT) RCT. A summary of each of these studies is described within this section and each subheading refers to information contained in table 4.2.

As described previously in section 3.1, there are a number of different checklists for assessing economic evaluations.(84;87;88) A review that strictly followed one of these checklists was not used for a number of reasons including that in the health economics community, there is no consensus over which checklist should be used and the use of a checklist often turns into a box ticking exercise without getting to the heart of the issue of whether the study is of sufficient quality. As in Chapter 3, within this chapter, the NICE reference case was loosely followed so as to provide a framework for assessing whether the important concepts within economic evaluation have been captured, namely: perspective, patient population, size of the evidence base and extrapolation, comparators, outcomes, uncertainty and discounting.

4.2.1 Perspective

Three different perspectives were used: the provider (ie the NHS),(136;145;151) societal, (134-136;140;147;152) and the payer.(93;137;142;143) In three cases no reference to perspective was made.(129;130;133)

4.2.2 Patient Group

In order to compare patient severity for COPD, the patient population used within each study was classified into disease severity groups as defined by GOLD (mild, moderate, severe and very severe as previously described in Chapter 2). However, often scarce reporting within some papers around inclusion and exclusion criteria as well as heterogeneous classifications made it difficult to accurately convert the disease severities into GOLD severity groups. All the pharmacological studies included severe COPD patients with most focussing upon moderate to severe.(129;131;136;140;145;151) The cost effectiveness analysis alongside the observational study,(142) included COPD patients in each severity category, as did the study by Van den Boom et al.(133) No reference to patient severity was made in Hogan et al paper,(137) and so the original trial publication was searched for this information.(155)

Within the cost effectiveness analyses on surgical interventions, the patient population had either severe emphysema or end stage lung disease. Whilst FEV_1 % predicted values were not presented, this patient population represents very severe COPD as defined by GOLD.

4.2.3 Patient Numbers and Duration of the Study

Patient numbers in each of the studies varied substantially from 52,(130) to 1067 participants.(129) Most studies tended to have upwards of 500 participants, usually

equally distributed between the treatment arms. The shortest study durations were 3 months,(129;137) four studies were for 12 months and two were for 36 months.(142;145) The surgical analyses used data collected within a 3yr, 4yr or 5yr study.

4.2.4 Comparators

Perhaps due to the very nature of the RCT and the complexities that doing so would demand, none of the pharmacological studies used the preferred comparator of 'current treatment'. Conversely, all of the surgical cost effectiveness studies used current treatment as comparator defined as either being on the waiting list for surgery,(130;134) or a group receiving medical treatment.(135;147;152)

Placebo was a comparator in eight of the pharmacological studies. In five out of the sixteen studies the main point of interest was the cost effectiveness of a combined drug compared to its component parts and sometimes placebo.(129;131;142;143) However the more recent studies compared a number of different drugs such as one comparing tiotropium with combinations of salmeterol and/or fluticasone,(151) the study by Lofdahl compared the cost effectiveness of placebo, Budesonide, Formoterol and the combination Budesonide and Formoterol and finally Oostenbrink et al measured the effects of replacing the short acting anticholinergic, Ipratropium with the long acting equivalent, Tiotropium. Where two different dosages were examined, the dominated treatment was dropped from further study.(93;137)

4.2.5 Outcome Measures

A wide range of outcome measures were used within the studies. All of the surgical and four pharmacological cost effectiveness studies used QALYs. Other outcomes assessed were: survival, change in SGRQ, reduction in number of exacerbations and improvement in FEV_1. In addition to this there were others, including: proportion of

patients remaining free of exacerbations after six months,(136) number of symptom free nights,(93) a daytime symptom card of less than 2,(93) avoided exacerbation,(140;143;151) and improvement in dyspnea.(140)

4.2.6 Extrapolation

Four of the five surgical studies extrapolated to a longer time frame: either 10 years, (135;147) 15years,(134) or to a lifetime,(130) and used regression based models to predict survival.(130;134;135;147) These models were then used to calculate survival weighted costs and effects via the Kaplan-Meier Sample Estimator,(156) so that a cost effectiveness statistic could be calculated. One study also developed a regression equation to predict costs past the duration of follow up.(147)

4.2.7 Results

All the pharmacological studies reported a favourable outcome compared to the comparator(s) except for the Najafzadeh et al study. Six found the study drug cost effective compared to the comparator(s).(133;136;140;142;143;145) The other four reported improvements in outcome associated with the study drug compared to the comparator(s).(93;129;131;137) The results from the surgical economic evaluations were more conservative with all studies reporting that surgery was expensive but over time ICERs were shown to reduce.

4.2.8 Handling of Uncertainty

Five of the studies used probabilistic sensitivity analysis.(140;142;143;145;151) Five studies used univariate sensitivity analysis around the underlying assumptions such as adjusting the value for the cost per day,(136) and inflating/deflating the cost of treatment drugs and rescue medication by 50%.(137) Two studies used bootstrapping

of the results.(135;147) Four made no mention of the uncertainty surrounding the economic evaluation.(93;129;131;133)

4.2.9 Sponsorship

Most of the papers were sponsored by industry: five fully sponsored.(129;136;140;143;145) The paper by Jones et al,(93) did not state any information on financial support, though one of the authors was working for GlaxoSmithKline.(GSK) Three papers were partially supported by the industry.(131;137;142) GSK supported most of the studies (n=6) Boehringer Ingelheim (BI) (2) and Astra Zeneca (AZ) and Novartis sponsored one each. None of the economic evaluations for surgery were sponsored by industry.

4.2.10   Summary

The non-modelling papers were of mixed quality with the surgical evaluations tending to be of better quality in terms of duration of follow up, comparator, use of the QALY and extrapolation. For decision making the economic evaluations for pharmacological treatments would have limited use as the duration of these studies was too short (maximum three years), none of the studies used 'current treatment' as the comparator and a range of disparate outcome measures were used, only occasionally the QALY. The QALY and the effect of uncertainty on the ICER were increasingly seen in the more recent studies.

Table 4.2 Key features of the economic evaluations for COPD

	Perspective	Country a	Cost year	Duration (mths)	Patient group b	Interventions c (n by study arm)	Outcome measure	Outcome	+ve (d) outcome	Uncertainty	Sponsor
Ramsey. (130) (1995)	/	USA	/($)	3yrs life-time	emphysema patients	**Post transplant n=28** Waiting list n=24	Cost per QALY	ICER $176 817 per QALY	mixed	One way	Academic
Rutten-van Molken. (131) (1995)	Societal	NL	1989 ($)	30	II, III	**Combined β2 agonist & corticosteroid n=91** Combined β2 agonist & anticholinergic n=92 β2 agonist & placebo n=91	Improvement in FEV1(≥10%). Hyper-responsiveness. Restricted activity days. Symptom free days.	Cost per relevant improvement in FEV1: Co vs Pl = $200 and $5.35 per symptom free day.	✓	None	Partly by GSK
Friedman.(129) (1999)	/	USA	1998 ($)	≈3	II, III	Ipratropium n=362 Albuterol n=347 **Ipratropium & albuterol n=358**	Peak change in FEV1. Area under the FEV1 response time curve from time 0 to 4 hrs	ICER not calculated: Combination product striclty dominated albuterol	✓	None	BI
Van den Boon. (133) (2001)	/	NL	1999 ($)	12	I, II, III, IV	Placebo n=41 **Fluticasone 250μg (x2 daily) n=33**	Improvement in FEV1. Quality adjusted life years	ICER $13 016. Early detection and treatment: ICER $33 921 (direct costs) $14 031 (direct and productivity costs).	✓	None	Various Partly by GSK
Anyanwu. (134) (2002)	Societal	UK	1999 (£)	4yrs 15yrs extrap	End stage lung disease	Single transplant n=260 Double transplant n=199 Waiting list n=1030	Cost per QALY Life years	£29 415 single lung £20 002 double lung	mixed	One way	Government
Ayres. (136) (2003)	NHS Societal	UK	1998 (£)	6	I,II,III	Placebo n=139 **Fluticasone n= 142**	Improvement in FEV1 Proportion of patients remaining exacerbation free (moderate/ severe and mild)	Per patient per day: Cost per relevant improvement in FEV1= £0.25 (NHS perspective) = -£3.39 (societal). Cost per relevant proportion of patients remaining free of moderate & sev exacerbations = £0.25 (NHS) and £-3.28 (Societal).	✓	One way	GSK
Hogan. (137) (2003)	Payer	USA	2002 ($)	≈3	II, III*	Placebo n=200 Ipratropium n=194 **Formoterol 12μg n=194** Formoterol 24μg n=192*	Change in FEV1. Change in Quality of life as assessed via the SGRQ.	Cost per relevant improvement in FEV1: Pl vs Ip = $273.03. Ip vs Fo 12μg = $1611.32. Cost per relevant improvement in QoL: pl vs Fo 12μg = $25.20.	✓	One way	Partly by Novartis
Jones. (93) (2003)	Payer	UK	/(£)	≈4	II	Placebo n=227 **Salmeterol 50mcg n=229** Salmeterol 100mcg n=218	Improvement in FEV1 % of symptom free nights. Improvement in SGRQ (≥(-)4 points).	Per patient per day: Cost per relevant improvement in FEV1= £4.62. Cost per symptom free night = £5.67. Cost per daytime symptom card of<2 = £12.33. Cost per relevant increase in health status = £4.44.	✓	None	One author working for GSK

NETT. (135) (2003)	Societal	USA	2002 ($)	3yrs 5yrs 10yrs	Severe emphysema	**LVRS n=531** Medical treatment n=535	Cost per QALY	$190 000 per QALY at 3yrs $53 000 per QALY at 10 yrs	mixed	Bootstrap	Academic
Oostenbrink. (140) (2004)	Societal	NL BE	2001(€)	12	II, III	**Tiotropium n=344** Ipratropium n=175	Reduction in exacerbations. Improvement in SGRQ (≥(-)4 points). Improvement in trough FEV(≥ 12%). Improvement in the transitional dyspnoea index (≥1 unit).	Per patient per year: Cost per: exacerbation avoided = €667; improvement in health status = €1084; improvement in dyspnea = €1259; relevant improvement in FEV1 = €796. At a threshold of €2000, probability of being cost effective to avoid one exacerbation is 80%.	✓	One way	BI
Gagnon. (142) (2005)	Payer	USA	2001 ($)	36	I, II, III, IV	No ICS, no LABA (placebo) n=274 Inhaled corticosteroid (ICS) n=538 B2 agonists (LABA) n=130 **LABA & ICS n=212**	Life expectancy	3 yr: ICER: LABA vs ICS/LABA = $91 430. Lifetime: ICER: placebo vs LABA= $6110. ICER: LABA vs comb= $27 570. Discounted gain in life expectancy in days (within study/lifetime respectively) No ICS and no LABA = 2.41/3.88, LABA and ICS = 2.70/6.14.	✓	PSA	Partly by GSK
Lofdahl. (143) (2005)	Payer	SE	2001 (€)	12	III, IV	Placebo n=228 Budesonide n=243 Formoterol n=235 **Budesonide & Formoterol n=245**	Avoidance of an exacerbation requiring medical intervention.	Comb is cost effective vs placebo if a decision maker is willing to pay about €2 per day per avoided exacerbation.	✓	PSA	AZ
Briggs. (145) (2006)	NHS	UK	1998 (£)	36	II, III	Placebo n=370 **Fluticasone n=372**	Quality adjusted life expectancy over three years. Life expectancy.	Cost per additional life year gained = £17 700. Cost per QALY gained £9500. Gain in Life expectancy = 23 days	✓	PSA	GSK
Ramsey. (147) (2007)	Societal	USA	/($)	5yrs 10yrs	Severe emphysema	**LVRS n=538** Medical treatment n=540	Cost per QALY	ICER $140 000, 5yrs & $54 000, 10 yrs More CE for patients with predominantly upper lobe emphysema & low exercise capacity	mixed	Bootstrapping	Academic
Najafzadeh. (151) (2008)	Healthcare system	CAN	2006 ($)	12	II, III	**Tiotropium n=449/3** Tiotropium + salmeterol n=449/3 Tiotropium + salmeterol/fluticasone n=449/3	Cost per exacerbation avoided Cost per QALY	Treble combination vs monotherapy was $243 180 per QALY and $6510 per exacerbation avoided.	X	PSA	Research
Blough. (152) (2009)	Societal	USA	($)	3yrs	Severe emphysema	**LVRS n=531** Medical treatment n=535	Cost per QALY	ICER: $178 000 to $331 000 depending on imputation method for missing data	mixed	imputation of missingness	Academic

a NL stands for the Netherlands and BE for Belguim. b Patient grouping standardised in line with the GOLD guidelines. c Bold type corresponds to the primary drug of interest d Positive outcome for the primary drug of interest. QALY=quality adjusted life year. ICER=incremental cost effectiveness ratio. FEV$_1$= Forced Expiratory Volume in one second. SGRQ=St George's respiratory questionnaire. BI=Boehringer Ingelheim. GSK=GlaxoSmithKline. AZ=Astra Zeneca

4.3 Review of the Economic Models

In this section, the thirteen economic evaluations that were developed around a modelling framework were examined. The models were taken individually and are described below by way of the methodology/characteristics employed and some of the inputs used. Table 4.3 presents key elements of each study.

4.3.1 Jubran et al (1993, (97))

The Jubran paper is the earliest example of an economic model applied to COPD. The study compares the costs and cost effectiveness of theophylline vs ipratropium over one year from a societal perspective, based on three observational data sets, totalling 600 people with a diagnosis of moderate to severe COPD in three US practice based sites. Resource use was extracted from the data sets to include: the number and type of visits made, drug treatments, lab tests, consultations and toxic events. From this data, estimates for labour, non-labour and overhead costs were made. The datasets were uneven in duration (theophylline: 7.1 months and ipratropium: 5.9 months) but the results were extrapolated to one year.

Jubran et al incorporated a decision tree into the Markov framework. There were seven states within the model: stable; clinic visit, consult, ER visit, hospital, major toxicity and minor toxicity and during these cycles, any one of five outcomes could occur: exacerbation, toxicity, routine clinic visit, routine consultation and no event, from which a number of other possible outcomes could arise. The model assumed that the patient was in one of the states at any one time and transitions between these states took place at the end of a one month cycle. There were 12 cycles in total and the study was limited to a one-year framework. The Jubran et al model was focused on toxic outcomes for COPD treatments. Within the analysis, toxic events accounted for 8% of hospitalisation for those taking Theophylline compared to 0% of patients taking Ipratropium.

No ICER was calculated since ipratropium was found to be both less costly and more cost effective than theophylline. Univariate sensitivity analysis was used to represent the uncertainty around costs and probabilities into the model. The study was sponsored by Boehringer Ingelheim.

4.3.2 Al et al (1998, (132))

The model by Al et al compared the cost effectiveness of a situation with and without a lung transplantation programme from the perspective of society for up to 40 years using a simulation model that mirrored the movement of patients through the Dutch lung transplant program. The simulation modelled 100 patients and 17 donor lungs entering the program each year and consisted of 7 phases: outpatient screening, inpatient screening, pre-transplantation, waiting list, transplantation, inpatient follow up and outpatient follow up. The model predicted duration of stay for each patient within each of these phases using survival analysis. During each phase: death, referral to the next phase, rejection or contact lost with the team could occur. For the comparator analysis, only movement to the waiting list was permitted. For the treatment arm, each cycle (which was one year in duration) when the 17 lungs become available, the model checks the waiting list to match the donor lung to the person with the longest wait and with the appropriate body and blood type and the corresponding person moves through to the next phase of the model, to the transplant. Once the patient history of each patient has been simulated, costs and effects were added and were weighted by the probability of survival, which was predicted based on a Weibull parametric model. Sensitivity analyses were conducted using scenario analyses. The ICER for surgery compared to no surgery was calculated as G167 000 (£68 477).

4.3.3 Groen et al (2004 (139))

Groen et al employed the same model as Al et al, described above, in order to answer a question about the relationship between diagnosis and the cost effectiveness of lung

transplantation. In this model, the patients were split into seven categories, of which one was COPD and individual survival curves were fit based on diagnosis. Cost and utility data were the same as in the earlier model. The ICER for COPD patients was $118 200 compared to no transplant when a 3% discount rate was applied.

4.3.4 Borg et al (2004 (138))

This model aimed to evaluate new medicines to satisfy various payer requirements. The model was a two dimensional Markov model as shown in figures 4.1 and 4.2, and was comprised of two factors: irreversible decline in lung function and periods of exacerbations. Disease states were based on lung function, categorised according to GOLD guidelines by FEV_1 % predicted, split into four groups: 1) $FEV_1 \geq 80\%$, IIA) $50\% \leq FEV_1 < 80\%$, IIB) $30\% \leq FEV_1 < 50\%$ and III) $FEV_1 < 30\%$ predicted. Transition between disease severities was based on a decline in FEV_1.
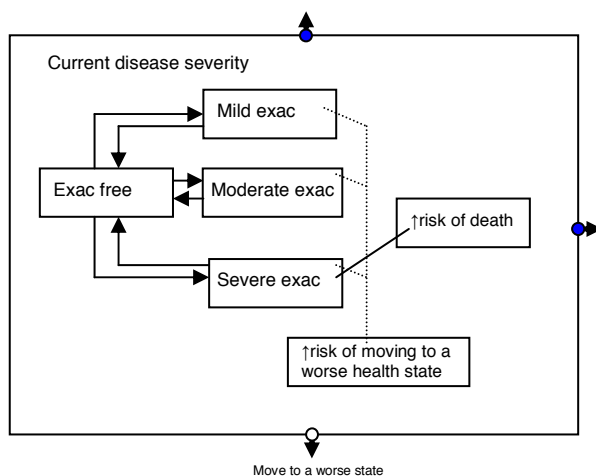
Figure 4.1 The first part of the disease model, Borg et al

The first part of the model was updated weekly as seen in figure 4.1, according to exacerbation activity and was based on current disease severity. Within each disease state Borg et al allowed any of four different events (exacerbation free, mild exacerbation, moderate exacerbation, severe exacerbation) to occur.

I

| 1 |

IIA

| 2 | | 3 |

IIB

| 4 | | 5 |

III

| 6 |

Death

| 7 |
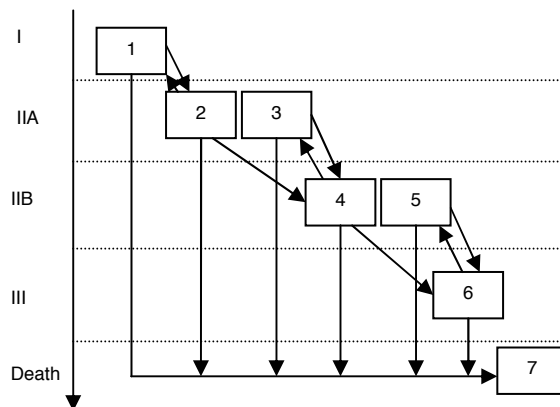
Figure 4.2 The second part of the Borg et al model
The normal progression was from 1, 2, 4, 6 then 7. It was also possible to regress one step (to 1, 3, 5) but no further

As shown in figure 4.2, the second part of the model was updated mid-yearly and within this part of the model, the patient could move: to the next state, remain in the same state, die or return to a previous health state.

4.3.5 Sin et al (2004, (141))

A societal perspective was adopted by Sin et al,(141) for their Markov Model which examined the effects of adding ICS to treatment for three groups; all COPD patients; patients with stage 2 or 3 disease and stage 3 disease. The model was three years in duration with twelve cycles; each cycle was for three months. Transitions between the states could occur after each cycle. Shown in figure 4.3, the model was split into three stages based on lung function: stage 1: $FEV_1 \geq 50\%$, stage 2: $35\% \leq FEV_1 < 50\%$ and stage 3: $FEV_1 < 35\%$. A final state of death was not explicitly included, though was inferred.
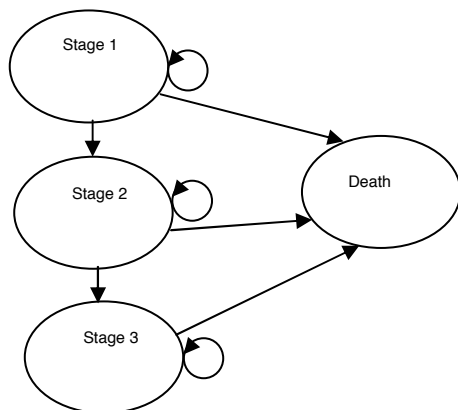
Figure 4.3 The Sin et al Markov model

FEV_1 was assumed to decline over time (47ml per year in each severity group) and from this the probability of progressing to the next stage was calculated and applied to all groups. Treatment was not assumed to affect the transition probabilities between states. Data from the third National Health and Nutritional Examination survey were used to estimate the proportion of patients in each state.(157) All-cause mortality was estimated from published data; risk increased subject to disease severity and varied between those treated with ICS compared to those who were not treated.

The rate and severity of exacerbations increased according to COPD stage, with the proportion of severe exacerbations, (compared to mild and moderate exacerbations) increasing as the disease severity worsened. Treatment with ICS was assumed to reduce the rate of exacerbations by 30%. Assumptions used in the model:

- Lung function decline was 11.75 ml per cycle regardless of disease severity group.

- QALY values applied were: 1.00 for stage 1, 0.92 for stage 2 and 0.84 for stage 3.

- From baseline, a reduction of 0.32 QALYs per exacerbation was applied regardless of exacerbation severity.

- The estimated duration of effect was 1 week (mild), 2 weeks (moderate) and 4 weeks (severe).

- The rate and severity of exacerbations increased according to COPD stage.

Whilst the viewpoint was that of society, direct marginal costs were included within primary analysis and an estimation of the productivity costs associated with work loss during exacerbations for those 65 years or younger was only conducted during a secondary analysis. Sin et al found that treatment was cost effective when given to patients with stage 2 or 3 disease. PSA was performed to account for the uncertainty around the inputs of the model.

4.3.6 Oostenbrink et al (2005, (92))

Oostenbrink et al developed an economic model for COPD in 2005. Since then this model has been adapted to different county settings in two subsequent studies, the first by Maniadakis et al in 2006,(146) and the second by Rutten-van Molken.(148) As these models are slightly different to the original they will be described later in this section.

The 2005 Oostenbrink et al *(92)* model was a one year Markov model around disease severity states (moderate, severe and very severe) and exacerbations, to compare the one year cost effectiveness of tiotropium compared to ipratropium and to salmeterol in the Netherlands and in Canada. The perspective of the model was the local health care reimbursement authorities in the Netherlands and in Canada. The Markov model had three health states: moderate ($50\% \leq FEV_1 < 80\%$ predicted), severe ($30\% \leq FEV_1 < 50\%$ predicted) and very severe ($FEV_1 < 30\%$ predicted) and did not include a mild state or a death state. The model had cycle lengths of one month. Within each state, non severe and severe exacerbations could occur and movement between states could be either: forwards, backwards or to remain in the same state. Effectiveness data for the tiotropium arm came from six RCTs and for the salmeterol and ipratropium arms, from the relative difference to tiotropium seen during the individual trials.

Tiotropium was assumed to improve symptoms and HRQoL and lead to fewer exacerbations in comparison to ipratropium and to placebo, and improvements in lung function in comparison to salmeterol. Exacerbations rates (severe and non severe) were dependent upon the treatment group and on disease severity.

- The length of the first cycle was 8 days.

- All subsequent cycles were one month.

- Only one exacerbation was allowed during any one cycle.

- Transitions between states were assumed to take place halfway through the cycle.

- Treatment was assumed to affect the transition probabilities; the average ipratropium patient was found to have a probability of 2.7 times more than the average tiotropium for movement from a moderate to severe state.

- Mean EQ-5D index scores used were: moderate=0.755; severe=0.748 and very severe= 0.549.

During a cycle in which an exacerbation occurred, utility was assumed to decrease (for the whole cycle) by 15% for a non severe exacerbation and 50% for a severe exacerbation.

In the Netherlands and in Canada, tiotropium was found to be associated with maximum expected net benefit for plausible values of the ceiling ratio.

To account for the uncertainties in the evaluation, Oostenbrink et al used PSA to test for the robustness of the result to changes in the baseline values of the model. PSA was employed by applying appropriate distributions around mean input values (for transition probabilities the Dirichlet (as proposed by Briggs et al, (158)), for rate of exacerbations and utilities the Beta, and for resource use, the Gamma distribution). Scenario analysis was employed to assess the impact on the model to different transition probabilities, utility values and around the cost of adding oxygen therapy.

4.3.7 Spencer et al (2005, (144))

The model by Spencer and et al (144) employed a Markov model in order to compare the cost effectiveness of the combination drug salmeterol/fluticasone, to usual care in patients with COPD. The model consisted of four disease states: mild COPD, $FEV_1 \geq 50\%$, moderate: $35\% \leq FEV_1 < 50\%$, severe: $FEV_1 < 35\%$ and death as illustrated in figure 4.4. The cycles were three months in duration and a maximum time horizon of 25 years was applied. Treatment for COPD was assumed to affect the risk of exacerbations, the risk of disease progression, risk of mortality and patient health status.
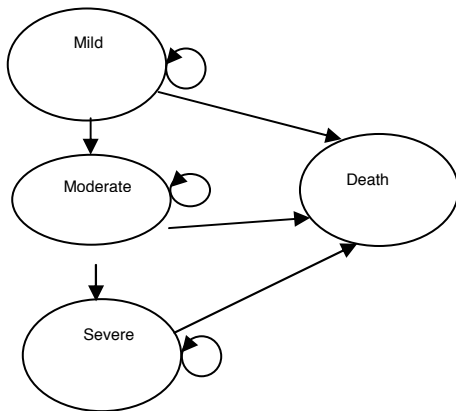
Figure 4.4 The Spencer et al Markov model

Baseline values for the model were sourced from GSK clinical trial data (TRISTAN),(159) published medical literature and from expert opinion. Estimates of health status by disease stage were mild: 0.81, moderate: 0.72 and severe: 0.67.

Table 4.3 Summary of the decision analytic models

	Perspective	Country	Cost year	Duration (mths)	Patient group a	Interventions	Outcome measure	Outcome	Extrapolation	Uncertainty	Sponsor
Jubran (1993)	Payer	USA	/($)	7.1 5.9	II,III	Theophylline n=311 Ipratropium n =289	Complication free therapy months.	Ipratropium was found to result in 11.30 complication free therapy months compared to Theophylline's 10.68. No ICER was generated: Ipratropium was found to strictly dominate Theophylline.	1 yr	One way	BI
Al (1998)	Societal	NL	1992 (G)	60	IV	Transplantation No transplantations	Life year gained QALY	ICER: G167 000 (£68 477) vs no transplant	40 yrs	One way	Academic
Borg (2004)	Various payers	UK & SE	1999 (SEK)	Up to 10yrs	I,II, III,IV	Tmt that reduces lung function decline. Tmt that reduces the number of exacerbations.	QALYs Life-years, Time without exac. N exacs	Reduction in lung function decline must be a long term treatment strategy compared to reducing the number of exacerbations.	Up to 30yrs	One way	Astra Zeneca
Groen (2004)	Societal	NL		60	IV	Transplantation No transplantations	Life year gained QALY	The ICER for COPD patients was $118 200 vs no transplant	40 yrs	One way	Academic
Sin (2004)	Societal	CAN	1999 ($)	36	I,II, III,IV	No patients treated ICS All patients treated with ICS. ICS - patients stage 2/3. ICS patients stage 3	Patient HRQoL. All-cause mortality.	Lifetime: Cost per QALY with a mortality effect: $4600 (All patients). $2900 (stage 2/3), $2000 (stage 3). Cost per QALY with no mortality effect: $26 200 (All patients), $21 200 (stage 2/3), $15 000 (stage 3).	Lifetime	MVA	GSK/ Institute of HE
Oostenbrink (2005)	/	NL CA	2001(€)	12 6 12	II, III	Tiotropium n=1296 Salmeterol n=405 Ipratropium n=175	Number of exacs. QAL months	NL: The prob tio is cost effective (CE) for: QALY's is almost indpendent of threshold, and for exacerbation avoided is 60% at €500. CA: The prob of tio being CE for: QALY's is highest from a threshold of ≥ €120, and for exacerbation avoided is highest from a threshold of ≥ €160.	1 yr	PSA	BI
Spencer (2005)	Payer	CAN	2002($)	12	II,III,IV	Usual care. Salmeterol/fluticasone.	Exacs Mortality. Patient health status.	25 years: Cost per QALY is $74 997 (basecase), $11 125 (survival effect) and $49 928 (delayed progression of disease)	25 yr	PSA	GSK
Maniadakis (2006)	NHS	Greece	2005(€)	12	II, III,IV	Tiotropium 18mcg Salmeterol	Number of exacs. QAL months.	The probability Tiotropium is cost effective is 77% at €1000 and 95% at €20 000.	No	PSA and one way	BI
Rutten-van Molken (2007)	NHS	Spain	2005(€)	12	II,III,IV	Tiotropium Salmeterol Ipratropium placebo	Exac free month. QALY	Tiotropium vs salmeterol €4118 salmeterol vs ipratropium €348 971	5 yrs	PSA	BI & Pfizer

Earnshaw (2008)	Pay er	USA	2006 ($)	36	II,III	Salmeterol and fluticasone Salmeterol Fluticasone Placebo	Life year saved QALY	ICERs vs placebo: salmeterol - $20 797, fluticasone - dominated, salmeterol & fluticasone - $33 865	Lifetime	PSA	GSK
Chuck (2008)		CAN	2006 CAN $	36	I, II, III	All patients treated LABA LABA + ICS to stage 3 LABA + ICS to stages 2& 3 LABA + ICS to all	QALY	LABA + ICS for stages 2 and 3 (lifetime) ICER: $50 571 LABA + ICS stage 3 ICER: $25 333 per QALY. LABA + ICS all stages was unlikely to be cost effective.	Lifetime	One way	GSK
Briggs et al (2009)	Pay er	US East EU Wes t EU	2007 ($)	36	II, III	Salmeterol and fluticasone Salmeterol Fluticasone Placebo	QALY	sal+ flut vs: placebo: ICER of $43 600 vs salmeterol: ICER $26 500 and vs fluticasone ICER $27 000.	No	Bootst rappin g	GSK
Oba (2009)	Pay er	USA	2006 ($)	36	II,III	Salmeterol and fluticasone Salmeterol Fluticasone Placebo	QALY	ICERs vs placebo: salmeterol - $56 519, fluticasone - $62 833, salmeterol & fluticasone - $52046	3 yrs	One way	None

a Patient grouping standardised in line with the GOLD guidelines. QALY=quality adjusted life year. ICER=incremental cost effectiveness ratio. FEV$_1$= Forced Expiratory Volume in one second. SGRQ=St George's respiratory questionnaire. BI=Boehringer Ingelheim. GSK=GlaxoSmithKline.

Transitions between states were based on decline in FEV_1. A regression equation was developed to predict disease progression based on FEV_1 at baseline, height, age and sex. Movement between states was unidirectional. Smoking status was assumed to affect the rate of decline of lung function and so affect the transition probability, which was calculated separately for smokers than ex-smokers. Smokers and ex-smokers were found to have increased FEV_1 rates of decline: 62ml and 31ml respectively.

Exacerbations were considered as a function of the disease state and were split into minor (contacts with primary care) and major (hospitalisation). Using the EQ-5D score, estimates for health status during an exacerbation were obtained from 27 respiratory physicians who completed the questionnaire from the perspective of their patients. For a minor exacerbation, health status was assumed to drop to 0.61 (mild), 0.61 (moderate) and 0.05 (severe). For a major exacerbation, health status fell to -0.26 regardless of disease severity. The study assumed a non linear recovery from the 'low point' to a position of 0.03 utility points below those of others in the study who did not have an exacerbation.

Routine/maintenance costs for each disease state were estimated and applied to the model. Costs associated with exacerbations for each disease state were estimated and included in the model by weighting by the rate of minor/major exacerbations in each disease state.

The study found that the combination therapy may represent a cost effective treatment in those patients who have a history of frequent exacerbations and poorly reversible COPD. A PSA was performed around the discount rate, exacerbation rate and the mortality benefit. The study was sponsored by GSK.

4.3.8 Maniadakis et al (2006 (146))

Maniadakis et al applied the model that was developed by Oostenbrink et al, to a Greek setting with the aim of comparing tiotropium to salmeterol over a one year time frame from the perspective of the Greek National Health Service. The probabilities and utilities applied to the model were the same as in the Oostenbrink et al model. New cost data were derived using resource utilisation obtained from the medical records at a Greek hospital and to this, unit costs were applied. Tiotropium was concluded to be cost effective, however there was no statistically significant difference found between the treatments. PSA was performed around the baseline values. As with the Oostenbrink et al paper, the study was supported financially by Boehringer Ingelheim.

The same issues around limitations of the Oostenbrink et al paper apply to this study, namely that there was no dead state applied to the model and that the model had only a one year time frame.

4.3.9 Rutten-van Molken et al (2007 (148))

The model by Rutten-van Molken et al expanded upon the earlier Oostenbrink et al model. The aim of the study was the same as in the Oostenbrink et al study, which was to examine the cost effectiveness of bronchodilator therapy with tiotropium, salmeterol or ipratropium for COPD, but from the perspective of the Spanish NHS.

The time frame was extended from one year in the earlier model to five years. The Markov model added the disease state of 'dead' compared to its predecessor, giving four disease states based on FEV_1 % predicted: moderate (50%$\leq FEV_1$<80% predicted), severe (30%$\leq FEV_1$<50%), very severe (FEV_1<30%) and dead. In each state, subjects were at risk of experiencing an exacerbation (moderate or severe) as illustrated in figure 4.5. Movements between states were based on the annual decline in FEV_1 derived from the trial data. Cycles were one month in duration. Backwards and forwards transitions

were allowed during the first year, and for subsequent years only forward transitions were permitted.
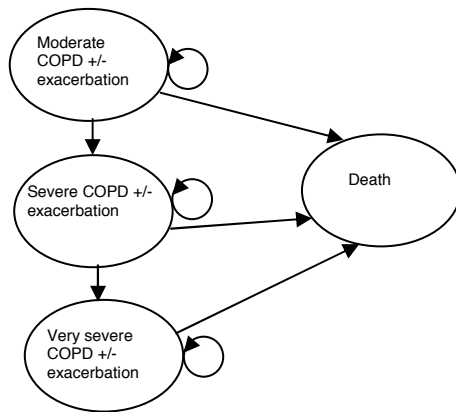
Figure 4.5 The Rutten-van Molken et al Markov model

Within the clinical trial, treatment with tiotropium was found to delay progression to the next disease state and to reduce the number of exacerbations and was applied to the model using scenario analysis. The base case scenario was that after the first year, the mean FEV_1 decline was the same for each treatment group (52ml per year). Exacerbation probabilities were based on first year rates. The second scenario applied transition and exacerbation probabilities from the first year throughout the five year model. The third scenario assumed that neither disease progression or exacerbation frequency/intensity was affected by treatment after the first year. Assumptions included:

- Movement between states was based on an annual decline in FEV_1 informed by trial data.

- The relative mortality rate (where severe mortality rate =1) was 3.754 for very severe COPD patients and 0.248 for moderate COPD patients.

- No differences in mortality risk were assumed between treatments.

- Mean (SE) utilities were 0.809 (0.008) for moderate disease, 0.762 (0.009) for severe disease and 0.655 (0.024) for very severe disease.

- Utility decreased by 15% for a moderate exacerbation and 50% for a severe exacerbation.

- The cost of a non severe exacerbation was EUR83 (£74) for a non severe exacerbation and EUR 2176 (£1941) for a severe exacerbation.

The ICER for tiotropium compared to salmeterol was EUR 4118 (£3787) and the ICER for salmeterol vs ipratropium EUR 348 971 (£320 896) in base case analyses. The research was financially supported by BI and by Pfizer.

4.3.10   Chuck et al (2008 (149))

Chuck et al developed a Markov model to determine the cost effectiveness of using combination therapy (long acting beta 2 agonists and inhaled corticosteroids) in different groups of COPD patients and was based on data from the literature. The model had three disease stages and a death state. Stage 1 was defined as $FEV_1 \geq 50\%$ predicted, stage 2 $35 \leq FEV_1 < 50\%$ predicted and stage 3, <35 % predicted. Cycle lengths were three months and were extrapolated to a lifetime horizon.

FEV_1 was assumed to decrease at a constant rate of 47ml per patient year and from this a transition probability for moving between states was calculated. Exacerbations were assumed to occur at each stage and were split into three types: mild, moderate and severe. The rate and severity of exacerbations increased in line with disease severity so that the total number of exacerbations per person per year was 0.17 for stage 1, 0.59 for stage 2 and 0.83 for stage 3. All-cause mortality rates increased with disease severity and were 3.92% for stage 1 disease, 6.16% for stage 2 disease and 9.24% for stage 3 disease.

Utility scores and costs were attached to exacerbations and utility was modelled as disutility where a QALY reduction would occur following an exacerbation: -0.17 for a mild exacerbation, -0.47 for a moderate and severe exacerbation.

Combination therapy for stages 2 and 3 over a lifetime perspective gave an ICER of CAN \$50 571 (£29 455) and restricting therapy to stage 3, an ICER of CAN \$25 333 (£14 755) per QALY. Including stage 1 patients was unlikely to be cost effective. The study was sponsored by GSK.

4.3.11   Earnshaw, S (2008 (150))

This Markov model was developed in order to examine the cost effectiveness of treating COPD patients with a combination of fluticasone and salmeterol compared to fluticasone alone, salmeterol alone and placebo, from the perspective of a third party US payer. The model had four states: moderate, severe, very severe and dead. Within each state the patient could either have no exacerbation, a mild exacerbation or a severe exacerbation. Annual cycle lengths were applied and the timeframe of the model was lifetime.

Transitions probabilities between states were based on data from the Lung Health Study (and were previously used in the model by Sin et al). Input data for exacerbation and for mortality risk were derived from the TORCH dataset. Utility values were taken from a study by Borg et al. The cost data were obtained from a range of sources.

The results from the model show that the ICER for the combination product compared to placebo was \$33 865 (£20 766). PSA was conducted to explore uncertainty around the parameters. The study was sponsored by GSK.

4.3.12   Oba. Y (2009 (160))

Oba developed a Markov model, designed to assess the cost effectiveness of inhaled medication (salmeterol, fluticasone and Seretide (salmeterol and fluticasone) compared to placebo) use in COPD from the third party payer's perspective in the US healthcare system. The model was based on four health states: stable, exacerbation requiring a

physician visit, severe exacerbation requiring hospitalisation, and death as illustrated in figure 4.6. Cycle lengths were three months and the total model duration was three years.
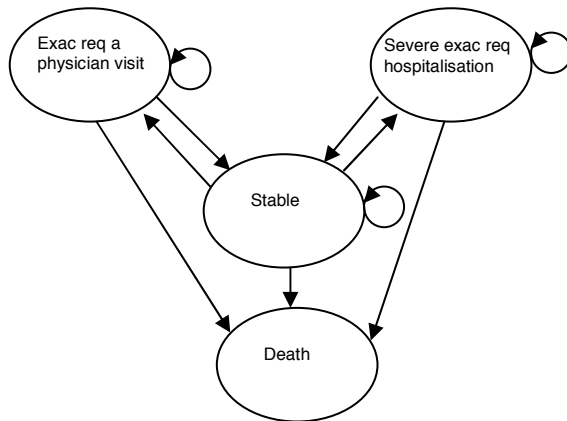
Figure 4.6 The Oba Markov model

Input data were derived from the TORCH clinical trial. Inputs into the model included: frequency of exacerbations requiring a physician visit, frequency of exacerbations requiring hospitalisation, all-cause mortality rates and utility scores (derived from the SGRQ and mapped to the EQ-5D). The three monthly all-cause mortality rate was modelled using an exponential approximation. A 3% discount rate was applied to costs and effects.

Compared to placebo, the ICER was $56 519 (£34 657) for salmeterol, $62 833 (£38 529) for fluticasone and $52 046 (£31 914) for the combination product. None of the arms were found to clearly dominate the other. One way sensitivity analysis was conducted. The study was developed within an academic setting and was not funded by the pharmaceutical industry.

4.3.13 Briggs et al (2009 (153))

Briggs et al developed an economic model using data from the TORCH trial to inform regression models to predict study medication cost, other medical cost, EQ-5D and

survival. The model compared the cost effectiveness of four treatments: salmeterol, fluticasone and the combination product salmeterol plus fluticasone to placebo.

Costs were inflated to a 2007 base year where necessary. Explanatory variables within the models included: age, body mass index, race, gender, count of items in medical history at baseline, count of pre randomised exacerbations requiring hospitalisation, baseline FEV_1 % predicted, Medical Research Council (MRC) dyspnea score, the SGRQ, smoking and region. The regression equation on survival was fit using a parametric Weibull survival model. A Generalised Linear Models (GLM) was used to estimate cost and an OLS regression equation was used to model EQ-5D scores.

Cost and QALY scores were weighted by the probability of surviving at the time of the visit. The method of recycled predictions was used to estimate costs and QALYs and ICERs. A discount rate of 3% was applied for both costs and effects. Bootstrapping was carried out to generate 95% confidence intervals (CIs) around the ICER results. Compared to placebo, the combination product had an ICER of $43 600 (£26 735), compared to salmeterol an ICER of $26 500 (£16 250) and to fluticasone, an ICER of $27 000 (£16 556).

4.4 Discrepancies in utility values

It is evident that there are many similarities in the approach adopted in the development of an economic model for COPD. COPD has been modelled using stages of the disease and at each stage; an exacerbation causes a drop in HRQoL for the modelled patient. The studies all used a utility measure in their analysis but there were sometimes large discrepancies around the inputs into the model, some of which have been described elsewhere.(161) Table 4.4 illustrates the differences between the utility values used in the papers by Sin et al, Oostenbrink et al, Rutten-van Molken et al and Spencer et al. The disease severity groupings were reported as per GOLD: $FEV_1>50\%$ (GOLD

moderate); 35%< $FEV_1 \leq$ 50% (approximate GOLD severe); and FEV_1<35% (approximate GOLD very severe).

Table 4.4 Health status by disease severity and the impact of an exacerbation

		Moderate	Severe	Very Severe
Rutten van Molken (2007)	Baseline	0.81	0.76	0.66
	↓ to (minor exac)	0.69	0.65	0.56
	↓to (major exac)	0.41	0.38	0.33
	Duration of effect	1 month	1 month	1 month
Oostenbrink (2005)	Baseline	0.76	0.75	0.55
	↓ to (minor exac)	0.64	0.64	0.47
	↓to (major exac)	0.38	0.37	0.27
	Duration of effect	1 month	1 month	1 month
Sin (2004)	Baseline	1.00	0.92	0.84
	↓ to (exac)	0.68	0.60	0.52
	Duration of effect	1 week	2 weeks	4 weeks
Spencer (2005)	Baseline	0.81	0.72	0.67
	↓ to (minor exac)	0.61	0.61	0.05
	↓ to (major exac)	-0.26	-0.26	-0.26
	Duration of effect	Recovery over six months		

Differences between the utility values by disease state differed between studies. For example, within table 4.4, utility values for very severe COPD were: 0.76 (Rutten-van Molken),0.55 (Oostenbrink), 0.84 (Sin) 0.67 (Spencer) as seen in the last column. In the same way, health state and duration of effect assigned to an exacerbation also differed. These differences exist due to the general lack of RCTs in COPD that have been designed to collect a generic HRQoL measure such as the EQ-5D and the subsequent need to derive utility estimates from elsewhere: from experts, from other studies or from mapping from a disease specific measure to the generic measure (none of these studies used mapping). Using expert opinion is unlikely to accurately capture utilities values and if other studies are used then there is a danger that the populations are different to one another. As described earlier in section 3.2.5, utility estimates for people with COPD are best obtained from RCT data using a generic HRQoL measure such as the EQ-5D.

The cost effectiveness analysis by Chuck et al applied dis-utilities to the model following an exacerbation: -0.17 for a mild exacerbation and -0.47 for a moderate or

severe exacerbation. Within the models, the reduction in HRQoL following a major exacerbation for a very severe patient was presumed to be: 0.33 (Rutten-van Molken), 0.27 (one month) Oostenbrink, 0.52 (one month) Sin and Spencer -0.26 (six months recovery time) as shown in table 4.4. The Spencer model assumed a non-linear recovery following an exacerbation to a position of 0.03 utility points below those of others in the study that did not have an exacerbation. Modelling recovery from an exacerbation in this way was novel and more likely to match the natural history of exacerbations compared to other approaches. Nevertheless, the approach taken by the authors of deriving utility estimates for health status during an exacerbation from the opinions of 27 respiratory physicians, is a major limitation of the Spencer study.

4.5 Discussion

The output of an economic evaluation is to inform and assist the decision maker in allocating scarce health care resources, but how far does the existing literature go in fulfilling this role?

Fundamental is the design of the study; RCT, observational study based or employing a model: this decision is all-important. Economic evaluations based entirely upon RCTs, such as those within the non-modelling section of the chapter, with tight inclusion and exclusion criteria may have limited generalisabilty to a wider population which may:

"seriously restrict their relevance for policy making"(162 p450)

The validity of alternative data sources, such as observational studies, depends upon the extent to which the study populations are representative and the conclusions of such studies may need supporting evidence from RCTs. It is suggested that modelling and the addition of observational data can enhance the external validity of the cost effectiveness study based on RCTs,(162) so that the study is generalisable to a wider

population than are contained in the RCT, and may be a useful avenue for future economic evaluations of COPD.

Within an RCT based study, efficacy data are confined to the length of the trial. To be of most value to clinicians and health care funding agencies, the costs and benefits should be considered over a period that reflects the longevity of the effects of the intervention.(163) All of the economic evaluations (with the exception of the surgical CE analyses) and almost half of the economic models had a duration of five years or fewer. A cost effectiveness ratio based solely upon the duration of the trial may fail to capture the longer-term effects of treatment, such as the extended life of study patients. This is important as there is currently an ongoing debate as to the impact of treatment on survival: recent evidence from trials such as TRISTAN,(50) UPLIFT,(51) and TORCH,(49) have suggested a survival effect of therapy in COPD patients following treatment with a combination product of salmeterol and fluticasone. Incorporation of a mortality effect into economic evaluations using models to extrapolate from trial evidence, may result in dramatic reductions of the resulting ICER.(141;144) Extrapolation is essential within surgical interventions because large costs in the short run lead to large HRQoL gains for those receiving successful treatment and so a long term follow up is necessary to capture all the benefits of treatment.

Several of the economic evaluations incorporated regression modelling, either for a particular variable, most predominantly survival, but also used for cost and for transition probabilities. Briggs et al used regression modelling within a clinical trial for their cost utility analysis. Several of the studies used a mapping equation to predict utility from a non-utility based HRQoL questionnaire, where utility data from an RCT was unavailable.

There is a clear and continuing role for the use of modelling in economic evaluation of COPD therapies. A modelling framework can produce externally valid studies (based on internally valid evidence of treatment effects), capturing the long-term effects of

treatment, thereby being useful in assisting the decision maker in allocating resources. Economic models in all their forms need to be methodologically sound, have relevant and valid inputs and to be well described and explained. In addition, the sensitivity analysis needs to be executed with care, distributions around the inputs explained and reasoned, and extrapolation needs to be adopted and presented with caution. It has been seen that within pharmacoeconomics for COPD, there has been a recent surge in the number of economic models for COPD. Economic evaluations in the future should ideally be based on sufficiently long RCT study durations and on modelling, so as to capture the relevant costs and effects so that results are useful and relevant for decision makers.

Results of economic evaluations are likely to vary according to the perspective employed. The perspective should be clearly stated within the paper and the results presented should be based upon the adopted perspective. Four evaluations did not mention perspective.(92;129;133)

RCTs for drug treatments were principally developed for the purposes of drug efficacy and more often than not an economic evaluation is piggybacked onto a RCT. As a result there are often problems when conducting cost effectiveness analyses where incremental effects and benefits arising from treatment are taken from the RCT, when the comparator(s) does not include 'current treatment'. If the comparator is not a real life existing/usual treatment or mix of treatment, the results have little value for the decision maker on which to base a decision. Only two of the pharmacological studies,(141;144) and all of the analyses for surgical interventions, included a range of relevant alternatives, including existing treatment. Decision makers need to know the full impact of the introduction of a new therapy to a disease or treatment area and this can best be achieved by using usual care as a comparator.

> 'Decisions on cost effectiveness should be based on the comparison of a
> new intervention with current practice, rather than with a placebo.'(164
> p711)

Earlier versions of the Oostenbrink 2005 model omitted the death state. Modelling a chronic condition longitudinally necessarily involves mortalities and death should always be modelled explicitly for COPD economic evaluations and in the case of the Oostenbrink et al model, was corrected in later versions.

A wide range of outcome measures have been used in economic evaluations of COPD. Although it may be the case that:

> "…it is neither known nor generally agreed which outcomes are most relevant"(165 p41)

A range of outcome measures causes a problem for the decision maker if they are faced with the problem of making a judgement based on disparate results which are not directly comparable. For example, to what extent is the avoidance of an exacerbation equivalent to an annual improvement in FEV_1 of 20ml? The economists' solution is the QALY. Within England and Wales, NICE has stated

> 'the QALY is considered to be the most appropriate generic measure of health benefit that reflects both mortality and HRQoL.'(95 p22).

As was described in Chapter 3, the QALY is a particularly useful outcome measure for economic evaluations. The strength of cost utility analyses based on QALYs depends upon the robustness of the derivation of the utility values. In each of the three modelling studies that used QALYs, the utility weights applied to different COPD states and to the impact of exacerbations varied quite considerably. The most likely reason behind these differences in utility values is the different methods of elicitation (and the population surveyed). The decrement in utility (from baseline) associated with each exacerbation and length of time to which this decrease is applied (duration of exacerbation), differs considerably between studies. It would be valuable to undertake further research into the derivation of utility values for COPD patients, and in particular, further research needs to be conducted in order to determine the effect of an exacerbation on utility.

An assessment of uncertainty should be included within an economic evaluation to reflect the uncertainty in the cost and health outcome results obtained from the study. In the case of modelling studies, there are further uncertainties, such as in the design of the model itself and the extrapolation of study data to a time horizon that extends beyond the life of the trial. PSA is preferred for assessing uncertainty in Markov models because it allows the combined uncertainty surrounding all of the parameters within the model to be assessed.(126)

Whilst all of the surgical interventions were developed and funded within academic settings, the majority of the pharmacological studies were sponsored to some extent by the pharmaceutical industry and the study drugs in each of these papers were reported to have a favourable cost-effectiveness result. Concerns about the outcomes of these studies, because of issues around: selection of study design, patient population and the potential for bias in the outcome and in the publication, are often raised. Nevertheless, the industry is an important provider of cost effectiveness data, especially to support submission for reimbursement in particular countries. In addition, the industry, because of tight regulating standards, may pay closer attention to quality control than academic institutes. Whatever the pros and cons are, industry-financed studies will continue to be a valuable source of data. However there is a gap for non-industry sponsored evaluations of pharmacotherapy in COPD and efforts should be made to provide the resources necessary in order to support non-industry bodies in producing such studies.

4.6 Conclusion

Within this chapter the published economic evaluations for COPD treatments were reviewed and critically appraised. The use of published economic evaluations in informing and assisting the decision maker to allocate scarce health care resources was then discussed.

The development of the reference case as discussed in Chapter 3 and recent methodological advances within the subject have gone some way to standardise the economic evaluations, as authors have increasingly: adopted the QALY as an outcome measure and used an economic model to carry out the evaluation.

Consistency between evaluations is necessary in order for comparisons to be made between different treatments over time,(95) however observed differences within the reviewed studies in terms of: study design, comparators, interventions, outcome measures and the analysis of uncertainty, make meaningful comparison between the studies very difficult. For the decision maker and for the clinician, it is of utmost importance that interventions are directly comparable. Decisions must be made as to the most suitable treatment; informed decisions, based upon and supported by all available knowledge and evidence of substitute or alternative treatments are most likely to be appropriate.

Efforts should be made for future economic evaluations to harmonise study design and methods, particularly towards adopting a universal modelling framework, using current treatment as comparator and adopting an effectiveness measure such as the QALY in order to produce results that are comparable across interventions and disease areas, and that are useful to a decision maker.

A generic model where an ICER could be derived in order to compare different treatments for COPD would be desirable. To maximise internal and external validity, combining different data including RCT and observational data is ideal within a modelling structure. The following two chapters investigate different data sources with the ultimate objective of developing an economic model. An observational dataset is studied to learn more about the natural history of COPD and from this assists the identification of an appropriate structure for an economic evaluation of the disease, and a RCT is analysed with a particular focus on utilities.

Chapter 5. Natural History of COPD in a British Population

The analyses within this chapter are aimed at examining the natural history of COPD in a British population, using the Renfrew/Paisley (MIDSPAN) dataset, a Scottish prospective cohort followed since the early 1970s, with ongoing linked hospitalisation and mortality records. It is guided by the importance of appropriately modelling the disease within the context of economic evaluation.

The process of evaluating COPD treatments for disease through economic evaluation, particularly decision analytic modelling, requires an understanding of the disease itself. Initially, what is the disease and how is it defined? How does the disease progress for the individual and more generally within the COPD population? Are number of hospitalisations influenced by disease severity and how many hospitalisations do COPD patients have? How frequent are hospitalisations and what is the length of hospital stay? Which factors are important for mortality risk? In other words, what is the natural history of the disease? In Chapter 4, it was seen that when developing economic models for COPD, researchers have often searched the literature or consulted clinical experts for answers to these questions rather than fully explore primary data sources. The purpose of this chapter is to analyse the natural history of COPD using a large longitudinal dataset from which the answers to these questions can be ascertained using primary evidence and four different analyses were conducted for this purpose.

Within the first analysis, summary statistics of the MIDSPAN dataset are presented. Prevalence of COPD in the dataset is reported. Particular attention is given to disease severity, and mortality rates, survival curves and the major causes of mortality are determined.

The second analysis replicates a previous study conducted by Hole et al,(166) who published a paper on the link between reduced lung function and subsequent mortality using the Renfrew/Paisley dataset with linked mortality records. One of the features of the dataset is that the follow-up of hospitalisations and mortality records is ongoing. The analysis in this chapter validates and updates the original analysis using all available data (up until December 2005).

The third analysis examines the diagnostic criteria for COPD. COPD is frequently identified based on impaired lung function and within this section the assumption that COPD is best diagnosed using lung function alone is questioned. The effect of including a risk factor such as smoking history and symptoms in the diagnostic criteria for COPD is investigated.

The final analysis investigates the number of hospital admissions and length of stay in hospital, before determining hospitalisation rates (or severe exacerbations) by disease severity for the Renfrew/Paisley population.

First, the benefits of using the Renfrew/Paisley (MIDSPAN) dataset compared to other datasets are explored before the MIDSPAN dataset is described. A section on epidemiology and statistics follows, explaining general concepts that are used throughout this chapter.

5.1 Datasets for Investigating the Natural History of COPD

Within this section, an overview of available datasets within the UK that have potential for study of the natural history of COPD is presented.

The MIDSPAN studies are comprised of three distinct large-scale epidemiological studies within the Scottish population and were designed to be used for researching issues of public health. At the time of conception in the 1960s, the development and use of such large population based studies was a novel concept. The importance of the MIDSPAN studies is not to be underestimated, with over 160 scientific papers written on findings from the studies. Of these three studies, of particular use for examining the natural history of COPD is the Renfrew/Paisley (MIDSPAN) study. Renfrew/Paisley was the last of the three studies to be operationalised and was the largest in terms of numbers recruited with 15 402 participants, of whom, 7048 were men and 8354 were women. The studies were originally developed in order to provide

> "…an evidence base for the detection and control of cardio-respiratory risks and diseases in whole populations in addition to improving the detection and control of tuberculosis."(167)

Nevertheless, the nature of the questions asked and data collected at the time of recruitment lend the dataset well to the investigation of COPD, with data collected on lung function, smoking history and respiratory symptoms. In addition, the study followed a general population with an appropriate age range for the study of COPD (45-64 years) in a UK population.

Other studies exist within the UK that could be used to investigate the natural history of COPD, including datasets based in primary care and secondary care:

The General Practice Research Database (GPRD) is based upon 3.4 million 'active' patients from approximately 450 primary care practices around the UK. GP Practices are not tied in to participating and may move into and out of providing data. The GPRD aims to provide continuous information on morbidity in primary care. In operation since 1988, it is said to be the largest computerised database of primary care medical records over time of its kind.(168) The GPRD records include: prescriptions (acute and repeat), adverse drug reactions, all consultations with medical personnel, family history,

diagnoses, symptoms, referrals (outpatient and emergency), hospital admissions, hospital diagnoses, operations, surgery, tests and investigations (such as FEV_1), contraception, pregnancies, births, deaths (cause and date) and patient lifestyle including smoking history, height and weight.(168)

The QRESEARCH database holds records for 3.3 million current and 4 million past, UK patients throughout 525 general practices, dating back to 1988. It is an anonymous database of GP records.(169) The database is similar to the GPRD database.

The Health Improvement Network (THIN) database contains anonymous patient information from over 400 general practices around the UK. The database is used for studies in: drug safety, epidemiology and health outcomes. THIN has been in operation since 1988.(170)

The Mediplus database is held by IMS. More than 500 GP's take part and it is regarded as being broadly representative of the GP population within the UK. Information available for analysis includes: patient age and sex, diagnoses (International Classification of Disease 10 (ICD-10)) and diagnoses linked to treatment, treatment linked with cost, test results linked with diagnosis and referrals.(171)

Secondary care based databases include the Scottish Morbidity Record (SMR1) and the Hospital Episode Statistics (HES). The SMR1 holds data on all patients admitted to Scottish hospitals since 1961,(172) a total of 6 million patients and 25 million individual episodes of care (to the end of 2006).(173) Each SMR1 episode (a single consultant in-patient episode) has an ICD-9, now ICD-10 code, recorded and for surgical operations, at least one operative code. The SMR1 data have been used to inform a wide range of decisions: to plan trust budgets, monitor year on year trends, to plan hospital bed numbers, for annual reports of individual consultants/ trusts and also for research purposes.(172) The SMR1 can be linked with corresponding mortality data

from the General Register Office (GRO) and to other Scottish datasets. The Renfrew/Paisley (MIDSPAN) dataset is linked to the SMR1 dataset.

The HES database collects information on all admitted patients treated within the English NHS. The range of data that the HES gathers is similar to that collected within the SMR1 with information on the care and treatment a patient receives whilst in hospital and coded using the ICD-10. Longitudinal data are available on patients' admissions. HES data are available since 1989, and currently approximately 12 million records are taken per year (approximately 150 million episodes of care had been reported by the end of 2003).(174) The HES data are said to be useful for a range of settings from developing, monitoring and evaluating departmental policies, identifying public health issues, monitoring improvements in public health, and for research purposes.(175) HES could be used to examine COPD hospitalisations (rates and frequency), length of hospital stay and survival for a COPD population (mortality data could be obtained from the Office of National Statistics). HES can be linked to other datasets.

Other databases available for examining the natural history of COPD include: the Health Survey for England and the Scottish Health Survey. The Health Survey for England has been running since 1991 and is an annual survey of a random sample of individuals from private households in England. The sample is thought to be representative of the general population. The Health Survey for England collects data on a range of socio-demographic, lifestyle, behavioural and biological variables. The data are collected via interview and physical examination. Key variables such as height, weight, smoking, drinking, blood pressure and general health are collected annually. In addition, each year sees a new focus for the survey, of particular interest is: asthma, accidents and disability in 1995, asthma, accidents and special measures of general health (including measurement of EQ-5D and SF-36) in 1996, and respiratory disease and atopic conditions, disability and non-fatal accidents in 2001. In the years, 1995, 1996, 1997, 2001 and 2002, several respiratory related questions were included as well

as lung function measurement (FEV$_1$). Analysis of this dataset would allow an estimation of the prevalence and incidence of COPD in relation to socio-demographic, lifestyle and behavioural factors. However there are limitations which exist because of the cross sectional nature of this survey, in particular, that the same people are not followed up at each survey.

The Scottish Health Survey aims to assess a representative sample of health and health related behaviours within private households in Scotland. It has been carried out three times: in 1995, 1998 and 2003. To date, more than 25 000 individuals have participated in the SHS. The survey involves two stages for each participant: an interview and a nurse visit. The interview covers a wide range of questions, including self assesed health and disability, health service utilisation, respiratory disease, smoking, drinking and socio-economic status. The nurse collects additional information including the use of prescribed medicine and takes clinical measures: blood pressure, lung function and collects blood and saliva samples.(176) It is possible to link the Scottish Health Survey to data from the SMR1 and the SMR4 (acute psychiatric hospital admissions), cancer registers and mortality data from the GRO.

Whilst it can be seen that there are a number of other databases available within the UK that could be used to study the natural history of COPD, the Renfrew/Paisley study is unique and is ideal for analyses around the natural history of COPD. The Renfrew/Paisley study provides a wealth of information on a sizeable population who have been continuously followed since the start of the study in the early 1970's until the end of 2005 in terms of both hospitailsation records and mortality. One particualar merit of the study is that follow up is either ongoing or complete for the majority of participants. Study participants continually residing within the UK had or will have complete follow up on mortality within the dataset. For participants continuously residing in Scotland, there is complete/ongoing follow up on both mortality records and hospitalisations. The time period and geographical location in which this study operated meant that few left the study location of Renfrew/Paisley, let alone the UK so

follow up is particularly strong. Added to that the high participation rate at the beginning of the study (78% of the general population in the two towns) and the result is an impressive and potentially highly informative UK general population study. The following section provides more detail on the Renfrew/Paisley (MIDSPAN) study.

5.2 Renfrew/Paisley (MIDSPAN) Study

The Renfrew/Paisley (MIDSPAN) study recruited from the general population of two towns, Renfrew and Paisley, located near Glasgow in Scotland between 1972 and 1976. All residents within the towns aged between 45 and 64 were invited to participate and asked to complete a self-reported questionnaire. They were then called for a screening examination. The total study population comprised 15 402 participants, representing 78% of the eligible population.

The self reported questionnaire asked a broad range of demographic and health questions, including: date of birth, sex, marital status, occupation, smoking history, the presence of phlegm and/or cough, breathlessness, wheeze, MRC bronchitis, angina, stroke symptoms and asthma. An example of the record card is reproduced in the appendix. The clinical examination measured, amongst other things: height, weight, blood pressure, plasma cholesterol and respiratory function (FEV_1 and FVC). Each participant had a chest x-ray. Details on occupation were used to determine occupational social class. Carstairs deprivation scores were derived based on the postcode of home address. BMI and FEV_1 % predicted were derived from the clinical examination data.

Since the start of the study in 1972, the MIDSPAN cohort have been linked to information held by the GRO for Scotland on all UK deaths,[177] and more recently, to all acute hospital discharges in Scotland through the SMR1 recording scheme.[178] Each hospitalisation and mortality is coded using ICD codes. Record linkage is on going and is set to continue until the study population has died out. The most recent

dataset available for analysis contains all mortality and hospitalisation records until the 31st December 2005 and is the dataset that is used within this chapter to conduct all analyses in order to investigate the natural history of COPD in a British population.

5.2.1 Missing Data

Two percent (n=368) of the participants were excluded from the analyses within this chapter, including participants who had data missing on any of: lung function (FEV$_1$ and FVC), respiratory symptoms (identified using the questions described in figure 5.3), height, age at starting smoking for current and ex-smokers and age at stopping smoking for ex-smokers. Other missing values by sex, were replaced by the mean value for diastolic blood pressure (n=8) and cholesterol (125), and by the modal value for social class (406). Participants who moved outside the UK were censored at the date of embarkation.(115) Twenty three participants were lost to follow-up and were also excluded.

Because there are only small amounts of missingness within the dataset, considering the patterns of missingness and appropriate responses is less important than if there were significant amounts of missing data. More space is given to discussions around missingness in Chapter 6.

5.2.2 Reference Values for Lung Function

As described earlier in Chapter 2, FEV$_1$ % predicted is used in diagnosing COPD and in stratifying by disease severity. Calculating a subject's FEV1 % predicted is done by dividing that person's observed FEV$_1$ by a predicted FEV$_1$. The predicted FEV$_1$ represents the expected lung function value for a 'healthy' person, with the same age and gender as the person under study.

Many equations have been developed that calculate predicted FEV_1 within various populations, usually with age and height as the independent variables,(179) and split by gender. The extent to which any of these equations relate to the Renfrew/Paisley (MIDSPAN) population is questionable as reference values are ideally calculated from measurements observed in a representative sample of healthy subjects within a general population.(180) A 'healthy' person can be identified as one with no respiratory disease or systematic disease and a lifelong non smoker (or who has no more than incidental smoking experience).(181) In order to produce representative % predicted lung function values for participants within MIDSPAN, equations for predicting FEV_1 were developed within the study population using only those subjects who met the 'healthy' criteria. Predicted FEV_1 for each subject was determined by a linear regression on age and height by sex using the healthy participants in the cohort. Those who answered 'yes' to questions on presence of wheeze, breathlessness, asthma, phlegm and the weather affecting their health as well as a smoking history were excluded. The remaining healthy participants comprised 870 men and 2792 women. The resulting regression equations with standard errors (SE) ($R2 = 0.26$) for men and women were:

FEV_1 (l) in men = -1.859 (0.532)- 0.029 (0.003) x age(yrs) + 0.037 (0.003) x height(cm)

FEV_1 (l) in women = -0.225 (0.230)- 0.029 (0.001) x age(yrs) + 0.024 (0.001) x height(cm)

FEV_1 % predicted was calculated for each subject using actual FEV_1 divided by predicted FEV_1. These equations slightly differ from previously published equations from the same dataset,(166) due to improved data recording in the dataset (details of which are described in section 5.5.1). STATA v10,(182) was used in all analyses.

5.3 Epidemiology and Statistics

Disease prevalence is defined as how much of a disease is present in a population at a point in time. Prevalence is calculated by dividing the number of people with a particular disease by the total number of people in the study at that time point. For example, if 500 people are diagnosed with COPD from a study population of 10 000 then disease prevalence would be equal to 500/10 000. Prevalence is usually expressed as a percentage, so in this case COPD prevalence would be 5%. Prevalence data can be used to compare prevalence between groups, for example it is probable that smokers are more likely than non smokers to have COPD, also that older people are more likely to have the disease than those who are younger.

Another measure that is frequently used in epidemiology studies is incidence. Incidence measures the rate of occurrence of new cases of a disease. It is calculated by dividing the number of new cases of a disease by the size of the disease free population. Because the Renfrew/Paisley dataset only contained information on patient characteristics from one time point, it is not possible to calculate incidence within the dataset.

Mortality rates are a measure of the number of deaths in a specific population, scaled to the size of the population, per unit of time. Mortality could be from either all-cause or disease specific causes. Mortality rates are given here per 10 000 individuals per year.

5.3.1 Survival Analysis

Survival analysis is used throughout this chapter for analysing time to event data. Some basic terminology is presented here, together with a discussion around the appropriate time frame to use within longitudinal studies.

Unique to the analysis of longitudinal data where time to event is of interest, is the notion of censored data. This concept was brought to light widely in the paper by Kaplan and Meier in 1958.(183) Within survival data, participants will either: 1) participate in the study until the event of interest 2) leave the study before the event of interest has occurred or 3) remain in the study until the study completion date and not have the event of interest. Survival analysis allows for censoring to occur.

An example of a longitudinal dataset with 10 participants, lasting for 25 years is presented below in table 5.1 and facilitates the explanation of some of the key concepts used within survival analysis. The event of interest is mortality and the dataset focuses around age at entry and age at exit, the rationale for which is described in the following section.

Within table 5.1, the age when each participant entered the study (column 2) and the age when they exited the study is recorded (column 3), together with information on if they were: alive, dead or whether they left the study before the end of the study (column 4). For example, participant A entered the study at age 45 and left the study 25 years later, aged 70 and alive, participant B entered the study at age 49 and died 30 years into the study, aged 79 and participant D entered the study at age 54 and exited just 3 years later. All participants who died have a value of one in the event column, or else zero (column 5), and all those who were either alive at the end of the study, or exited before study completion, have a value of one in the censored column, or else zero (column 6).

Table 5.1 Survival data for a hypothetical study

	Age enter study	Age exit study	Outcome	Event	Censored
A	45	70	Alive	0	1
B	49	79	Dead	1	0
C	51	76	Alive	0	1
D	54	57	Left study	0	1
E	56	69	Dead	1	0
F	58	83	Alive	0	1
G	60	85	Dead	1	0
H	60	72	Dead	1	0
I	62	87	Alive	0	1
J	63	81	Dead	1	0

Another way of presenting these data is shown in figure 5.1 where age is placed along the x axis and for each participant (represented by a bar), age of entry into the study, age of exit and outcome and events of interest are plotted. This way, the number of participants left in the study following an outcome can easily be established.

At each time point at which an event (mortality) occurs, a survival probability, S(t) is calculated and for the above cohort, given in table 5.2. S(t) is defined as:

$$S(t) = 1 - r(t) = \frac{n_t - d_t}{n_t}$$

where r(t) represents the estimated risk of mortality at the exact time of the event and is calculated by dividing the number of events, d_t by the total number at risk n_t. For example the risk of mortality for patient E, when that patient died at time t was calculated in table 5.2 by:

$$r(t) = d_t/n_t = 1/9 = 0.11$$
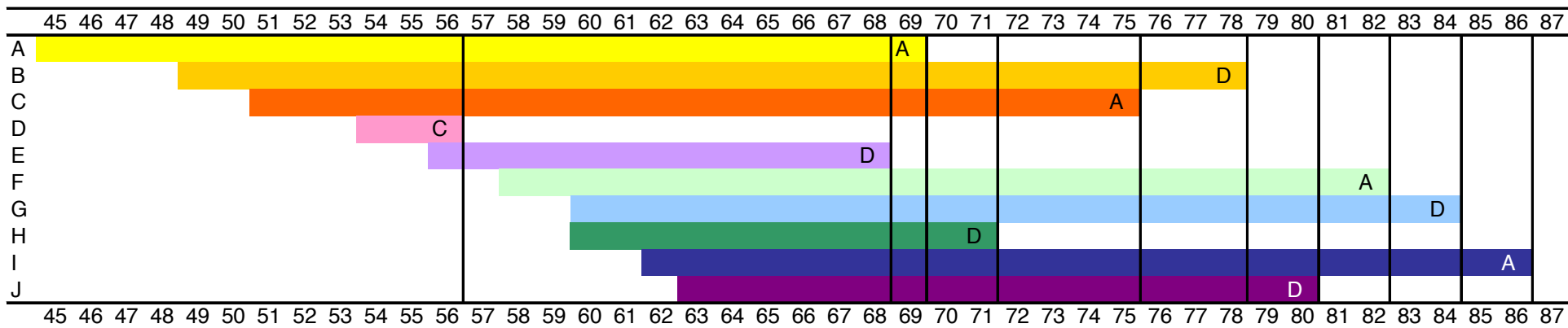
Figure 5.1 Visual representation of time in study and events, participants in a hypothetical study
Vertical lines represent outcomes of interest (A=alive, D=dead, C=censored)

If an event does not occur (for instance where censoring occurs) s(t) is always equal to 1. The survivor function is calculated by multiplying S(t) in the previous time point by s(t) at the current time point, and is formally described below:

$$S(t_j) = S(t_{(j-1)}) \times s(t_j) = s(t_1) \times s(t_2) \times ...s(t_j)$$

The survival probability S(t) is the probability that an individual survives until at least time t.(184) Because the survival probability only changes when there is an event, the value of the survival function, S(t), is constant between events and the estimated probability is a step function.(185) For example, for the time point at which the first person (who is censored) leaves the study, the S(t) remains at 1.00, this compares to the S(t) for the second observation which was calculated as:

$$S(t) = 1.00 \text{ x } 0.89 = 0.89$$

The values for r(t), s(t), S(t) based on n_t, d_t and c_t (censored at time t) for the ten participants are shown below in table 5.2

Table 5.2 Calculation of the survival function

Age (yrs)	n_t	d_t	c_t	r(t)	s(t)	S(t)
57	5	0	1	0.00	1.00	1.00
69	9	1	0	0.11	0.88	0.88
70	8	0	1	0.00	1.00	0.88
72	7	1	0	0.14	0.86	0.76
76	6	0	1	0.00	1.00	0.76
79	5	1	0	0.20	0.80	0.61
81	4	1	0	0.25	0.75	0.46
83	3	0	1	0.00	1.00	0.46
85	2	1	0	0.33	0.66	0.30
87	1	0	1	0.00	1.00	0.30

Survival data are usually described and modelled in terms of two rates: survival as previously described, and hazard. The hazard function, h(t) describes the event rate at time t conditional on survival up until time t, or beyond. The cumulative hazard, H(t) is the total hazard experienced up to time t and is estimated by the sum of the risks at each

time point at which an event occurs. S(t) is related to H(t) as described by the formula below:

$$S(t) = e^{-H(t)}$$

5.3.2 Kaplan-Meier Curves

The Kaplan-Meier survival curve is essentially a plot of the survival probability S(t) against time. It is easy to compare the survival probabilities in different groups using this visual method and also to determine the median survival time if applicable. Within this chapter, Kaplan-Meier curves are used to compare survival in different COPD severity groups and between groups identified using different diagnostic criteria.

The time frame to use within survival analysis is debated. Time in the study has often been used, however age, rather than time in the study has been recommended as the appropriate time scale within a longitudinal study.(186;187) This is because for some outcomes, such as COPD, it is expected that the hazard would change more as a function of age than as a function of time in the study.(186;187) To illustrate this point within the MIDSPAN study, it is argued that it is reasonable to assume that a participant who entered the study at 45 would have a lower hazard of mortality than a participant aged 65. Using the traditional method of time in the study as the time scale, there is an assumption that any two participants have the same hazard after a certain amount of time in the study (five years, ten years etc), however, it seems logical that in general, a fifty year old would, *ceteris paribus*, have a lower hazard compared to a seventy year old. For other studies, particularly clinical trials, time in study is likely to be a more appropriate timeframe because RCTs are looking for differences in treatment groups. As each treatment group is usually randomised, there should be no differences in the age distribution between groups and it is the time on treatment and subsequent effects of treatment that are of interest and which are studied through the use of time in study as the time frame. Using age as the time scale allows the median age of survival (at death)

to be read from the survival curve. The median age is found by reading off the x axis when the survival curve/function is equal to 0.5. The equivalent when using the traditional method is median time to event from the start of the study, the interpretation of which is less informative.

Note that at the start and at the end of the study, the numbers at risk are small. In standard survival analysis, precisely because the right hand tail of the Kaplan-Meier curve is based on comparatively fewer participants data than at other points on the curve, it is advised that when the number of observations is low (approximately five) either these values are omitted entirely from the analysis or that any interpretation is dismissed for these points. For the same reason, when using age along the x axis: because values at the left hand side of the curve are also based upon fewer observations, it is advised that interpretations of these areas are not made (discussed and illustrated in section 5.4.2).

A limitation of using age as the time variable, is that a person with moderate COPD entering the study aged 65 is treated the same a person with moderate COPD aged 45 on entry to the study, who survives for 20 years. This is problematic because after 20 years the person with moderate COPD at baseline may have worse disease severity such that they would be in the severe COPD group.

The corresponding survival curve for the data described in the previous section is shown in figure 5.2. With age on the x axis, the first event at age 69 causes a corresponding drop in the survival curve. Events occurring in the dataset at different ages are clearly shown using this method. Median survival is shown where the dotted line meets the survival curve, at approximately 80 years of age.
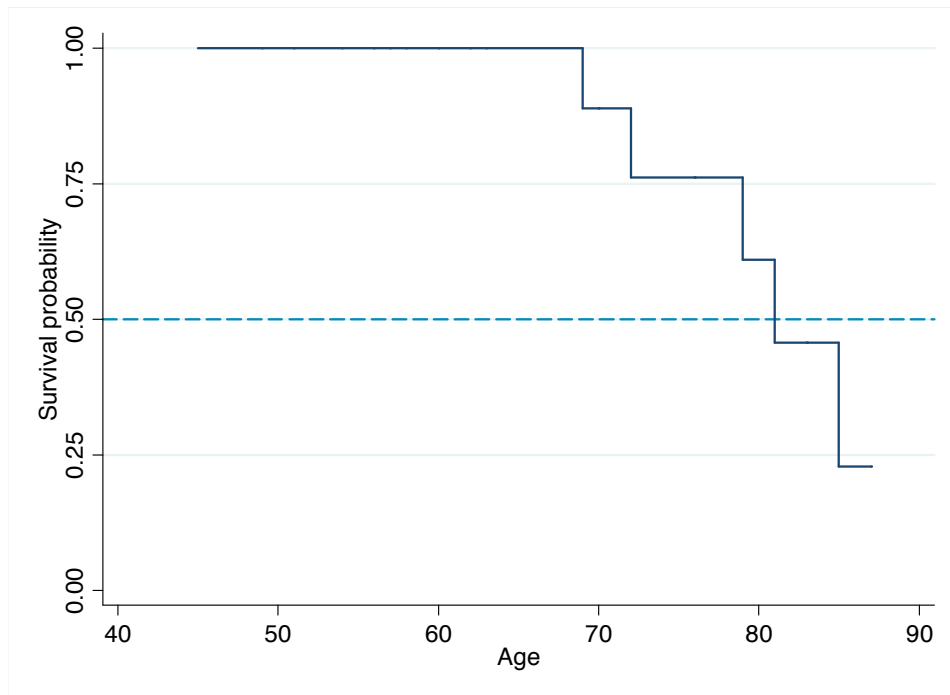
Figure 5.2 An example of a survival curve with age on the x axis

5.3.2.1 Cox Proportional Hazards Model

The Cox proportional hazards model is the most commonly used model for conducting survival analysis. The model describes the relationship between the event of interest (usually mortality) and the covariates, by way of Hazard Ratios (HR).

The Cox model is expressed as:

$$h(t) = h_0(t) \times \exp\{b_1 x_1 + b_2 x_2 + \ldots + b_q x_q\}$$

where the hazard function, h(t) is dependent upon a set of q covariates $(x_1, x_2, \ldots x_q)$ and whose impact is measured by the size of the coefficients $(b_1, b_2, \ldots, b_q)$ on those covariates. (188) The equation gives the hazard for the exposed group ($h_1(t)$). The hazard of the unexposed group, $h_0(t)$ is equal to the value of the hazard if all of the x's are equal to 0 (which when exponentiated gives a value of 1).

The HR from comparing the exposed group to the unexposed group at time t is given by the equation below and provides a measure of the relative survival experience between the two groups.(185)

$$HR(t) = \frac{h_1(t)}{h_0(t)} = \frac{h_0(t) \times \exp(\beta_1)}{h_0(t)} = \exp(\beta_1)$$

Because $h_0(t)$ appears in the top and bottom of the equation, they cancel one another out, and the HR is equal to the exponential of the linear equation. A HR with a value of one corresponds to the groups having equal hazards. A HR that is greater than one means that an explanatory variable is positively associated with the event of interest and conversely, if the HR is below one, that the explanatory variable is negatively associated with the event of interest.

One of the benefits of the Cox regression model is that because it is estimated non-parametrically, the model does not assume or impose any particular distribution on the dataset, however an important assumption of the model is that of proportional hazards: that the ratio of the hazards in the exposed group to the hazards in the unexposed group remains constant over time, as illustrated in the equation below.(189)

$$\frac{h_1(t)}{h_0(t)} = \text{constant}$$

Fulfilling the proportional hazards assumption is important when it is the absolute magnitude of effect of the independent variables that is of interest. If instead it is the relationship between variables that is of interest, then the assumption may be considered as having a secondary role.(190) If the assumption of proportional hazards is violated then the corresponding risk estimates may be inaccurate,(190) and the associated p values and CIs may be misleading.(191)

The assumption of proportional hazards can be tested either graphically or by using Schoenfeld residuals.(192) The graphical method involves plotting the log of the cumulative hazards function for each group against the log of time, which should give parallel lines.(193;194) With the graphical method, for the assumption to hold, the hazard curves for the groups should not cross and they should be proportional. The Schoenfeld residual is defined as the explanatory value for the individual that failed minus its expected value, and calculates separate residuals for each explanatory variable.(192) The Schoenfeld residuals are independent of time and as a result, the proportional hazards assumption can be assessed by testing the association between residuals and time.(193) A non-significant relationship between the residuals and time supports the proportional hazards assumption whereas a significant relationship suggests the proportional hazards assumption has not been met. Options to address the problem of violation of the Proportional Hazards assumption if it arises include: conducting the analysis stratified by the variable in question,(190) and including previously omitted variables or interaction terms into the model,(194) (if known).

5.4 Exploratory analysis

Within this section, exploratory analyses are conducted within the Renfrew/Paisley dataset in order to obtain information on the disease, such as prevalence of COPD in the general population, survival duration and mortality rates by disease severity, and causes of mortality for people with COPD.

5.4.1 Methods

COPD cases were identified using the NICE diagnostic criteria for COPD (which were outlined previously in Chapter 2 and the rationale for adopting these criteria is discussed later in detail in section 5.6), NICE suggest that a diagnosis of COPD should be considered in patients aged over 35 with airflow obstruction (FEV_1/FVC<0.7 and

FEV$_1$<80% predicted), a risk factor (principally smoking) and who present with one or more of: exertional breathlessness, chronic cough, regular sputum production, frequent winter 'bronchitis' or wheeze).(7) Assumptions were made to identify NICE COPD cases within the dataset:

- Airflow obstruction: FEV$_1$/FVC<0.7 and FEV$_1$<80% predicted.

- Respiratory symptom(s): one or more of breathlessness, wheeze, phlegm, identified using the questions listed in figure 5.3.

- Risk factor: either a smoking history of ten pack years or more, or being a pipe/or cigar smoker.

Disease severity was applied to the COPD cases according to FEV$_1$ % predicted. Mild COPD was identified in subjects with 50≤ FEV$_1$< 80% predicted, moderate COPD where 30≤ FEV$_1$< 50% predicted and severe COPD where FEV$_1$<30% predicted.

Presence of a self-reported respiratory symptom was considered if the participant answered positively to one or more of:

"Do you get short of breath walking with people of your own age on level ground?"
OR
"Does your chest sound wheezy or whistling on most days (or nights)?"
OR
"Do you usually bring up any phlegm from your chest first thing in the morning in the winter?" plus, yes to either: "do you bring up phlegm like this on most days for as much as three months in the winter each year?" or "in the past three years have you had a period of increased cough and phlegm lasting for three weeks or more?"

Figure 5.3 Inclusion criteria for respiratory symptoms within the NICE COPD diagnostic criteria

All analyses were conducted by disease severity groups and by sex. Prevalence of COPD was determined. Kaplan-Meier survival curves (with age on x axis) were produced to illustrate the impact of COPD and COPD disease severity on survival.

Mortality rates were calculated for all-cause mortality and for COPD mortality using person-years at risk. COPD mortality was identified where mortality records contained any of the following codes: ICD-9: 490-492 and 496 and ICD-10 J40-J44, in any position.

In a secondary analysis, first cause of death was identified in order to establish the degree of co-morbidity for COPD patients. The focus was on cardiovascular disease, respiratory disease and cancers as these three causes represent a high proportion of cause of mortality in the general population. ICD codes were: ICD-9 460-519 and ICD-10 J00-J99 for respiratory disease, ICD-9 140-208 and ICD-10 C00-C99 for cancer and ICD9 390-459 and ICD-I0 I00-I99 for cardiovascular disease. The proportion of deaths from these three causes of mortality was presented by disease severity.

5.4.2 Results

The Renfrew/Paisley (MIDSPAN) general population study comprised 15 034 participants of whom 46% were men and 54% were women. The majority of the participants have died: 78% of the men and 66% of the women. Most of the participants were smokers or had smoked, 77% of men and 46% of women had ten or more pack years (where one pack year is defined as 20 cigarettes smoked per day for one year). The participants tended to be from a working class background with the modal social glass group of III manual (IIIM) for men and IV for women. The mean BMI values show that on average both men and women were slightly overweight.

Table 5.3 Baseline characteristics of the study population

	Men	Women
Participants, n	6861	8173
Deceased, n (%)	5381 (78)	5378 (66)
Years of follow up, mean (SD)	19.8(9.6)	23.2(8.9)
Years of follow-up (range)	0-34	0-34
Age, mean (SD)	54.6(5.6)	54.9(5.6)
Social Class (mode)	IIIM	IV
Body Mass Index, mean (SD)	25.9(3.4)	25.8(4.5)
Smoking ≥ 10yr pack yrs, n (%)	5264 (77)	3757 (46)

Prevalence

COPD prevalence was lower for women than for men with overall COPD prevalence in the study of 12.4% (n=854) for men and 4.7% (n=388) for women. Of those diagnosed with COPD, the majority had mild rather than moderate or severe COPD. This is clearly illustrated in table 5.4, which shows that 8.5% of the men in the study and 3.3% of the women were diagnosed with mild COPD compared to 4% of men and 1.5% of women who were diagnosed with moderate or severe COPD at baseline.

Within the COPD cases, the proportion of subjects in each disease severity group was similar for men and women. As seen in table 5.4, 67.9% of the men were in the mild group compared to 68.8% of women, 23.9% of the men and 24.7% of the women were in the moderate group and 8.2% of the men and 6.4% of women were in the severe COPD group.

Table 5.4 Prevalence of COPD disease severity (%)
within: 1) the general population study and 2) COPD cases, men (left) and women (right)

	N in group	General pop (n=6861)	COPD cases (n=854)	N in group	General pop (n=8173)	COPD cases (n=388)
Mild	580	8.5	67.9	267	3.3	68.8
Moderate	204	3.0	23.9	96	1.2	24.7
Severe	70	1.0	8.2	25	0.3	6.4

Survival

The survival curves shown in figure 5.4 for all-cause mortality showed clear separation by disease severity. The severe COPD group had the lowest survival probability in men and women and the no COPD group, the highest survival probability. The median age of survival was associated with disease severity. On average, participants with severe disease had a life expectancy of 20 years fewer than the no COPD group for all-cause mortality.
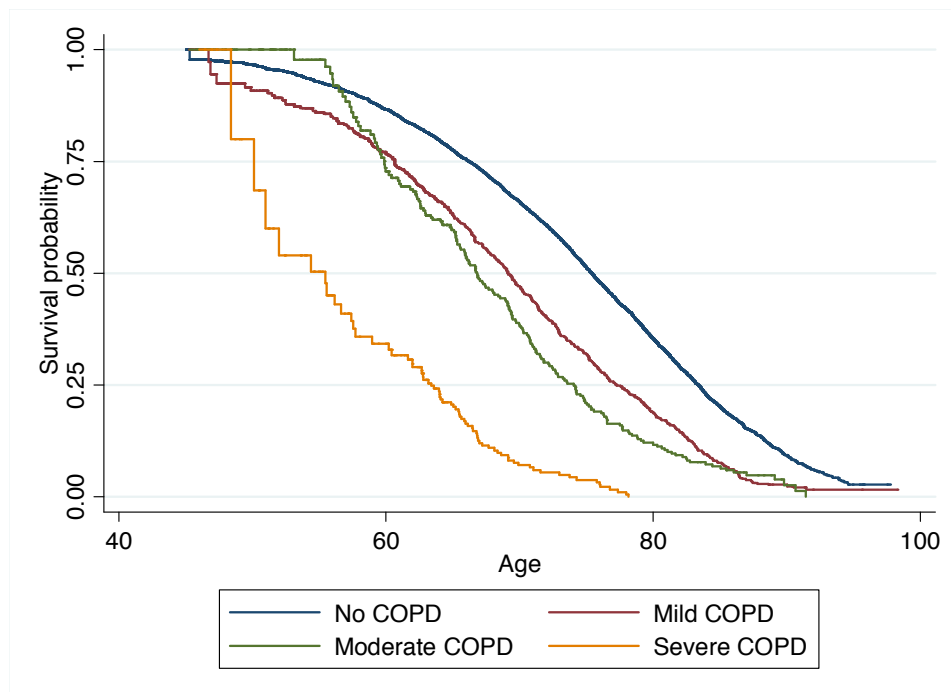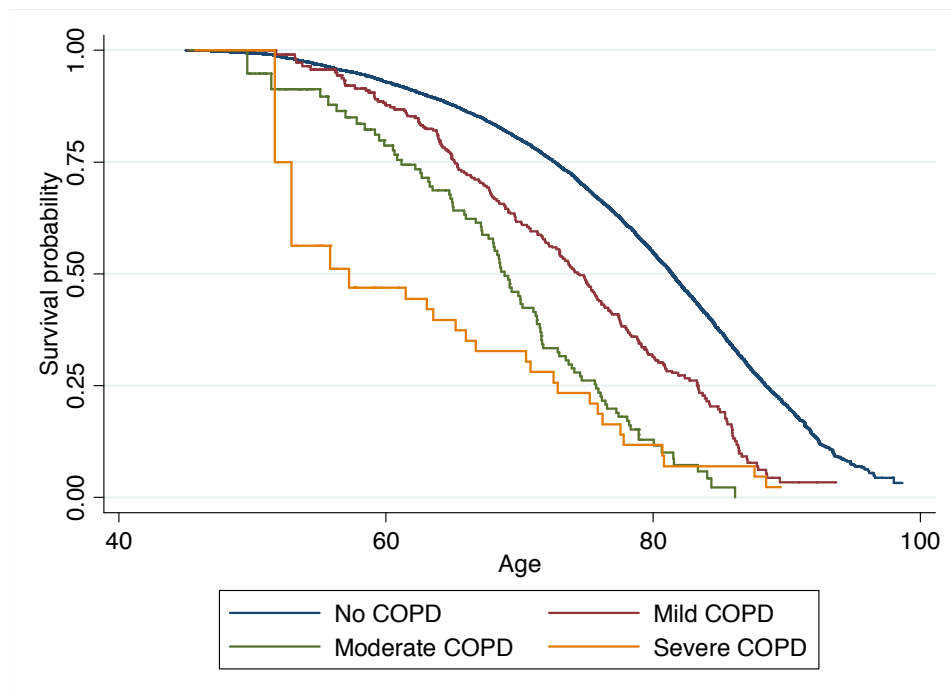
Figure 5.4 Survival curves for all-cause mortality for men (top) and women (bottom) by COPD disease severity

The large jumps at the start of these curves illustrate the point that was previously made in section 5.3.1. Because few subjects entered the study aged 46 and 47 (the youngest members of the study), when a 46 year old died within the first year of the study, this caused a large jump in the Kaplan-Meier curve because the number at risk, $n(t)$ was low and the corresponding event risk, $r(t)$ was high so the effect on $S(t)$ was large. This contrasts to other age groups such as age 65 where thousands of subjects were in the number at risk group and where one person dying would have a small impact on the $S(t)$. Therefore, for the youngest ages in the study, large drops in the $S(t)$ curve impact on the shape of the survival curves at the left hand side. Therefore when using this approach, for ages where there are only small numbers at risk, curves representing survival at these ages should be ignored. Adjustment methods can be used to limit the impact at the left hand side of these curves, and an approach for this is developed in the analysis comparing NICE to GOLD diagnostic criteria.

Mortality

Within this section, the effects of COPD disease severity on mortality rates for all cause and COPD mortality are described.

All-Cause Mortality Rates

All-cause mortality rates were seen to increase with disease severity, with a higher mortality rate by disease severity in men than in women. Table 5.5 shows that all-cause mortality rates in the no COPD group for men (women) were 372 (276) per 10 000 person years for men, this compares to 559 (417) deaths in every 10 000 person years at risk in the mild COPD group, 733 (582) in the moderate group and 1318 (634) deaths in every 10 000 person years at risk in the severe group.

Table 5.5 Mortality rates by disease severity, men and women

		N in group	N deceased	Person yrs at risk	Rate p/10 000 person yrs
Men & women	No COPD	13792	9635	305742	315
	mild	847	750	14733	509
	moderate	300	281	4148	677
	severe	95	93	902	1031
Men	No COPD	6007	4588	123209	372
	mild	580	532	9511	559
	moderate	204	192	2619	733
	severe	70	69	523	1318
Women	No COPD	7785	5047	182533	276
	mild	267	218	5222	417
	moderate	96	89	1529	582
	severe	25	24	378	634

COPD Mortality

The impact of disease severity on COPD related mortality is illustrated in table 5.6. The proportion of people with COPD related mortality rose as disease severity increased. In table 5.6 it can be seen that compared to 4% of the no COPD population, 21% of the mild group, 44% of the moderate group and 51% of the severe COPD population had

COPD recorded on their death certificate. This pattern was observed in both men and women.

Table 5.6 Percentage of the study with COPD coded mortality, by disease severity

		N	COPD mortality N (%)
Men & women	No COPD	13 792	557 (4)
	Mild	847	180 (21)
	Moderate	300	132 (44)
	Severe	95	48 (51)
Men	No COPD	6007	285 (5)
	Mild	580	123 (21)
	Moderate	204	93 (46)
	Severe	70	40 (57)
Women	No COPD	7785	272 (3)
	Mild	267	57 (21)
	Moderate	96	39 (41)
	Severe	25	8 (32)

Major Causes of Mortality

Primary causes of death for the study participants were compared by disease severity with a focus on cardiovascular disease, cancer and respiratory disease. Of the men and women, 70% and 56% respectively died from one of these three causes of mortality: 39% of the men died of cardiovascular disease compared to 32% of the women, 23% of men died from cancer compared to 18% of the women and 8% of the men died from respiratory disease and 6% of the women.

The three primary causes of mortality (cardiovascular disease, cancer and respiratory disease) accounted for 34% of deaths in the no COPD group, 45% of mild, 48% of moderate and 67% of severe COPD deaths and which is illustrated in figure 5.5. Cause of death changed with disease severity: over 50% of those in the severe COPD group died of respiratory disease which is represented in the final column in figure 5.5, compared to 6% in the no COPD group which is shown in the first column in figure 5.5. Mortality attributable to cancer was proportionally less for severe COPD patients compared to the no COPD group with 13% of deaths as a result of cancer within a severe COPD group compared to 20% in the no COPD group, though the absolute

number of people with cancer was small in the severe group, n= 12 compared with n=2735 in the no COPD group. The proportion of deaths due to cardiovascular disease was also less for the severe COPD group than the no COPD group with 2% of cardiovascular mortalities in severe COPD compared to 8% in no COPD participants.
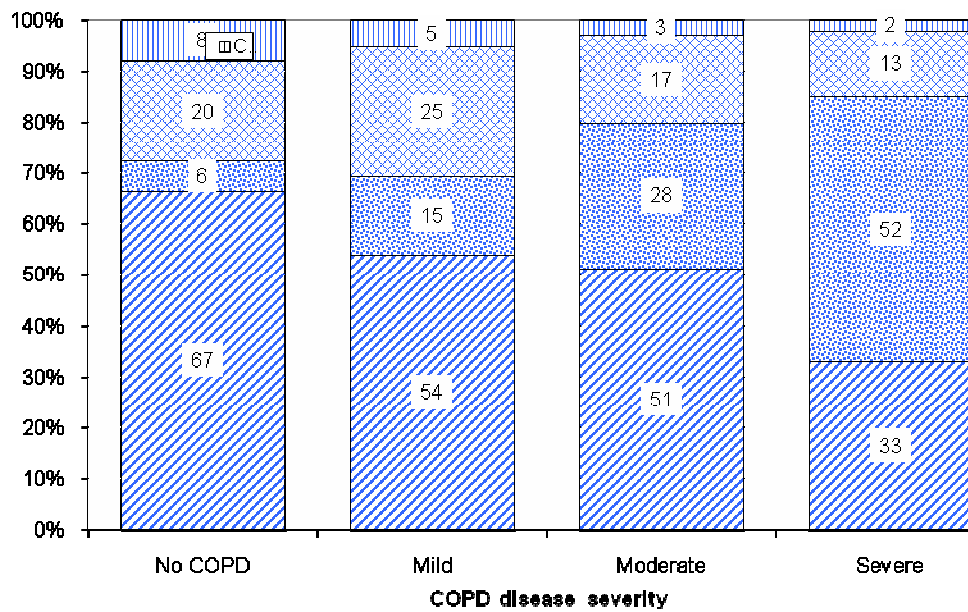
Figure 5.5 Cause of death by disease severity

5.4.3 Summary

People with COPD are at higher risk of all-cause mortality than the general population without COPD. This risk increases with disease severity and severe COPD patients have a mortality risk approximately three times that of the no COPD group. Increased risk affects survival duration and the data show that, adjusted for age, people with increasing disease severities have on average, a shorter life expectancy, such that for the severe group, life expectancy was as much as 20 years fewer than in the no COPD population. Survival curves for COPD by disease severity, presented as shown in this section with lengthy follow up of mortality data, have not previously been seen before

within the published literature. An economic model would need to factor in this increased risk of mortality by disease severity (which is directly related to FEV_1 % predicted) and take into account that it would be inappropriate to model a COPD population as having the same mortality risk as a non COPD population.

The proportion of people with COPD recorded on their death certificate was seen to increase in line with disease severity and was recorded in over 50% of death certificates for those people who had severe disease. Conducting analyses where COPD mortality is identified based on COPD coding in any diagnostic position on the death certificate is likely to overestimate the significance of COPD as cause of death. This is because the disease may have been recorded on the death certificate without actually playing a direct part in final cause of death. Thus it is more conservative to use primary cause of death on which to conduct analyses for diseases and because of this, subsequent analyses on mortality data focus on the principal cause of death.

The three primary causes of death investigated (cardiovascular disease, cancers and respiratory disease) represent a significant proportion of cause of death for people with COPD, indicating the degree of co morbidities present in COPD cases and the range of causes of death that the COPD population may suffer. When designing an economic model for COPD it is important to consider the causes of death within the analyses. The analyses conducted within this section have shown that it would be unwise to focus exclusively on COPD related mortality for a COPD population as this only represents a proportion of cause of death. More appropriate would be to consider all-cause mortality.

5.5 Impaired Lung Function and Mortality Risk

One of the most influential analyses of the Renfrew/Paisley (MIDSPAN) dataset was by Hole et al and was published in 1996. The study assessed the relationship between impaired FEV_1 and mortality.(166) As described in Chapter 2, impaired lung function

is a key factor within COPD and is an indicator of disease severity. As shown in the previous section, as disease severity (which is defined by FEV_1 % predicted alone) worsens, mortality risk increases. This section analyses the extent to which impaired lung function affects mortality risk, away from disease severity, by replicating and extending the Hole et al study. The duration of available follow up between the original study and 2005 has increased from a mean of 15 years and now contains mean follow up duration of approximately 20 years.

Over recent years, some of the original coding of sex within the dataset was found to be incorrect and the total number of participants in the current version is slightly fewer than previously reported (15 402 vs 15 411). Therefore study numbers differ in the most recent version as previously described compared to those reported in the Hole et al paper. Most importantly there are now 7049 men compared to the 7058 as was reported in the Hole et al paper. The number of women has stayed the same.

The degree to which impaired lung function affects mortality risk (represented by HRs) within a general population for all-cause, respiratory and COPD mortality was calculated. The original study is described below before the updated analysis is presented in section 5.5.2. Secondary analyses investigated the extent to which having age on the x axis within the survival analysis as opposed to time, affected HR estimates.

5.5.1 Review of the 1996 Hole et al Paper

The aim of the Hole et al study was to assess the relationship between baseline FEV_1 % predicted and subsequent mortality using the MIDSPAN population with 15 years of linked mortality data. The main outcome measure was all-cause mortality. Secondary analyses were carried out for other causes of death however the focus in this section is on all-cause and respiratory mortality in line with the themes of the thesis.

Methods

The dataset contained an average of 15 years of follow up of mortality. Variables of interest within the study included mortality, FEV_1 % predicted, age at screening, history of cigarette smoking, diastolic blood pressure, cholesterol concentration, BMI and social class. Analyses were conducted based on FEV_1 % predicted.

Predicted FEV_1 for each subject was obtained from linear regressions on age and height on 878 men and 2796 women, classified as 'healthy' using the same method as described earlier in section 5.2.2. The resulting predicted FEV_1 equations for men and women are reproduced below and were used to determine FEV_1 % predicted (observed FEV_1 divided by % FEV_1). The prediction equations slightly differ in the constant term from those reported in section 5.2.2 because of differences in the dataset as previously described.

FEV_1 (l) in men = -1.9302 – (0.0290 x age(yrs)) + (0.0373 x height(cm))

FEV_1 (l) in women = -0.2662 – (0.0289 x age(yrs)) + (0.0238 x height(cm))

Participants were split into five groups according to FEV_1 % predicted. Those with the lowest 20% FEV_1 % predicted values went in group 1 and those with the highest (least impaired) FEV_1 % predicted values went into group 5. All other participants went into groups 2, 3 and 4 depending upon degree of lung function impairment.

The association between FEV_1 % predicted and mortality (adjusted for age, cigarette smoking, diastolic blood pressure, cholesterol concentration, BMI and social class) was examined using Cox's proportional hazards models, by group. Group 5, the least impaired group, was the base case.

Results

At the start of the study there were 15 411 participants of which 7058 were men and 8353 were women. After the follow up period, 4439 of the participants had died, 2454 of the men and 1844 of the women. Five equally sized groups (quintiles) were derived based on FEV_1 % predicted values. For men, FEV_1 % predicted values <73 meant that they were in group 1, values between 73 and 86 placed participants in group 2 , 87-98 scores went into group 3, 97-107 values into group 4 and a FEV_1 % predicted score of ≥108 placed participants in group 5, which was the base case and represented those with the least impairment. For women, the corresponding FEV_1 % predicted values were <75 for group 1, 75≥group 2<90, 90≥group 3<101, 101≥group 4<113 and participants with FEV_1 % predicted values ≥113 went into group 5.

Table 5.7 HRs by mortality from all-cause, respiratory and lung cancer and by lung function group, in men and women. Adapted from Hole et al.(166)

Cause of death	Sex	N dead	HRs (95%CI) vs group 5(HR=1)			
			Group 1	Group 2	Group 3	Group 4
All-cause	M	2545	1.92(1.68-2.20)	1.51(1.31-1.74)	1.45(1.26-1.68)	1.28(1.1-1.48)
	W	1894	1.89(1.63-2.20)	1.52(1.30-1.77)	1.21(1.03-1.42)	1.17(0.99-1.38)
Respiratory	M	198	9.35(4.87-17.97)	2.02(0.96-4.25)	1.45(0.65-3.32)	1.12(0.48-2.60)
	W	115	6.47(3.17-13.19)	2.95(1.38-6.29)	1.00(0.40-2.54)	1.04(0.41-2.61)

Groups were split by FEV_1 % predicted. For men: <73=group 1, 73≥group 2<87, 87≥group 3<97, 97≥group 4<108 and ≥108 = group 5. For women: <75=group 1, 75≥group 2<90, 90≥group 3<101, 101≥group 4<113 and ≥113 = group 5.

As seen within table 5.7, those participants with FEV_1 % predicted values <108 in men (groups 1-4) and <101 in women (groups 1-3) were found to have a significantly increased mortality risk for all-cause mortality compared to those in group 5. The HR for all-cause mortality in group 1 was 1.9 for men and women. For deaths from respiratory disease, group 1 for men and women and group 2 for men were found to have increased mortality risk, table 5.7 shows that the HRs for group 1 were 9.35 for men and 6.47 for women.

Increased mortality risk was identified in subjects whose FEV_1 was moderately lower than the FEV_1 % predicted, regardless of group. The authors found that participants smoking fifteen or more cigarettes daily, with low cholesterol, placed in low social class

groups and who had reduced FEV_1 % predicted, had the highest mortality risk. Fifteen year all-cause mortality was: 48% (29%) in men (women) smoking 20 cigarettes a day with poor FEV_1 % predicted; 28% (17%) for those smoking 20 a day and with good FEV_1 % predicted and 15% (10%) for non smokers and in summary, the authors found that:

> "Impaired lung function is a major clinical indication of mortality risk in men and women for a wide range of diseases(166)."

5.5.2 Update of the Hole et al Paper, 1996

This section describes the methods and results for the updated analysis that built on and extended analyses conducted in the Hole et al paper, with follow up of mortality until 31st December 2005.

Methods

The methods described in the review of the Hole et al paper were largely replicated. To identify quintiles of impaired lung function, the same cut off limits as in the Hole et al study were used. All analyses were run with the least impaired group ($FEV_1 \geq 108$ % (113) predicted for the men (women)) as the base case.

Cox proportional hazards models were used to investigate the impact of lung function on all-cause, respiratory and COPD mortality (adjusted for age at start of study, cigarette smoking, diastolic blood pressure, cholesterol concentration, body mass index and social class).

The primary analysis was conducted using time in study for the Cox model (as was used within the Hole et al study). A secondary analysis examined the effect of changing the time frame of the survival analysis to age in study for reasons explained in section 5.3.1.

The linked follow-up mortality data employed in the Hole et al study was coded using the ICD-9, which was the only classification system used for mortality during that period. The current ICD codes, known as ICD-10 were introduced in 1994 and enter the dataset from then on. The ICD codes used to identify different causes of mortality were: ICD-9 460-519, ICD-10 J00-J99 for respiratory mortality and ICD-9 490-492 and 496, ICD-10 J40-J44) for COPD mortality.

Results

Of the 15 402 participants in the study, 368 were excluded (as previously described) because of missing data. The study cohort consisted of 6861 men and 8173 women, of whom 10 759 had deceased: 5381 men and 5378 women.

As in the Hole et al paper, group 1 contained those with the most impaired lung function, group two the second most impaired and so. Group 5 was the base case group and comprised those people with the greatest FEV_1 % predicted.

Table 5.8 Percentage of group deceased, men and women by FEV_1 group (n)

Group	Men	Women
1	89 (1303)	78 (1299)
2	83 (1152)	69 (1204)
3	78 (1029)	64 (1030)
4	74 (1023)	59 (950)
5	67 (874)	58 (895)

Table 5.8 illustrates how the proportion of deceased participants increased in relation to the degree of lung impairment, with the highest proportion of deceased participants in the group with the lowest FEV_1 % predicted values. These results suggest that there is an increased risk of mortality for those in group 1 compared to those in group 5.

The results from the primary analysis with time in study as the time horizon are presented in table 5.9, and table 5.10 contains the results of the survival analysis in which age as the time horizon was employed.

Table 5.9 HRs by cause of death and group, in men and women (time in study)

Cause	Sex	Dead (n)	HRs (95%CI) vs group 5(HR=1)			
			Group 1	Group 2	Group 3	Group 4
All-cause	Men	5381	1.77(1.62-1.93)	1.45(1.33-1.59)	1.34(1.23-1.47)	1.24(1.13-1.36)
	Women	5378	1.77(1.62-1.93)	1.40(1.28-1.53)	1.24(1.14-1.36)	1.12(1.02-1.23)
Respiratory	Men	542	5.48(4.02-7.45)	2.25(1.61-3.15)	1.45(1.01-2.10)	1.38(0.96-1.99)
	Women	535	3.64(2.76-4.81)	1.78(1.32-2.40)	1.22(0.88-1.69)	1.03(0.74-1.44)
COPD	Men	425	27.97(12.29-63.66)	7.41(3.15-17.41)	2.68(1.04-6.92)	2.14(0.81-5.64)
	Women	180	19.34(8.45-44.24)	5.00(2.10-11.93)	2.95(1.18-7.41)	1.28(0.44-3.69)

Groups were split by FEV_1 % predicted. For men: <73=group 1, 73≥group 2<87, 87≥group 3<97, 97≥group 4<108 and ≥108 = group 5. For women: <75=group 1, 75≥group 2<90, 90≥group 3<101, 101≥group 4<113 and ≥113 = group 5.

Table 5.10 HRs by cause of death and group, in men and women (age in study)

Cause	Sex	Dead (n)	HRs (95%CI) vs group 5(HR=1)			
			Group 1	Group 2	Group 3	Group 4
All-cause	Men	5381	1.81(1.66-1.98)	1.46(1.33-1.60)	1.34(1.22-1.47)	1.24(1.13-1.36)
	Women	5378	1.78(1.63-1.94)	1.40(1.28-1.53)	1.24(1.13-1.35)	1.11(1.01-1.22)
Respiratory	Men	542	5.64(4.15-7.66)	2.26(1.61-3.15)	1.44(1.00-2.08)	1.37(0.96-1.98)
	Women	535	3.63(2.75-4.79)	1.77(1.32-2.39)	1.22(0.88-1.69)	1.04(0.74-1.45)
COPD	Men	425	28.68(12.62-65.20)	7.38(3.14-17.33)	2.65(1.03-6.84)	2.12(0.81-5.59)
	Women	180	19.11(8.37-43.66)	5.05(2.12-12.05)	2.99(1.19-7.50)	1.32(0.46-3.80)

Groups were split by FEV_1 % predicted. For men: <73=group 1, 73≥group 2<87, 87≥group 3<97, 97≥group 4<108 and ≥108 = group 5. For women: <75=group 1, 75≥group 2<90, 90≥group 3<101, 101≥group 4<113 and ≥113 = group 5

Tables 5.9 and 5.10 show that regardless of method adopted, participants with FEV_1 % predicted values of <108 for men and <113 for women (or groups 1 to 4) were found to have a statistically significant, increased mortality risk for all-cause mortality compared to those in group 5, the base case. The HRs for all-cause mortality were 1.8 in group 1 for both men and women. This compares to a slightly larger HR of around 1.9 for men and women in the Hole et al study.

Participants with FEV_1 % predicted of <97 in men and <101 in women (groups 1 to 3) had significantly larger HRs for COPD mortality than those in group 5. Those participants whose lung functions placed them in group 1 (FEV_1 % predicted <73 for men and <75 for women) had very large HRs for COPD mortality with HRs of 28.68 for men and 19.11 for women, compared to those in group 5. The size of the HRs for COPD mortality illustrates the importance of impaired lung function on COPD mortality.

The HRs for each quintile were found to be similar when analysing the data using time in study for the survival analysis compared to using age in study. The CIs were almost identical. This finding supports those presented elsewhere, that using time in study rather than age, yields approximately unbiased proportional hazards regression coefficients.(186)

There is a clear difference in hazards between men and women. The men tended to have larger HRs compared to the women in each of the groups and this manifests itself as a higher percentage of deceased men in the study compared to deceased women for COPD, respiratory and all-cause mortality.

The conclusions of the Hole et al study, that impaired lung function was found to be a major clinical indication of mortality risk for a wide range of diseases in men and women,(166) is as true today as it was in 1996 and is backed up by an even greater amount of evidence.

5.5.3 Summary

An updated version of the study by Hole et al was presented that replicated the previous analysis, but with a longer follow up period. With 2/3 (n=10 759) of the original population deceased in the current version of the dataset as opposed to 1/3 (4439) at the time when Hole et al analysed the data, the HRs had changed. For all-cause mortality, the HR in group 1 was lower than the same group using the shorter follow up period. Within the other groups, the HRs all differed slightly to those previously reported: either larger or smaller than those published in the original paper. Nevertheless, in all cases, the HRs calculated using the current version of the dataset lay within the 95% CI of those reported in the original paper.

The use of time in study rather than age as the time frame for the survival analysis produced almost identical results, indicating that if time in study were used, it would produce approximately unbiased estimators of the hazard ratios.

Analysing the data by sex found that men and women have different mortality risks, especially for respiratory and for COPD mortality. Therefore it would be useful to conduct analyses on men and women separately when building a cost effectiveness model. Whilst policy makers are unlikely to differentiate provision of treatments according to gender if a treatment was found to be cost effective in females and not in males, or vice versa, because of ethical and political implications, it is nonetheless of interest to examine the impact of cost effectiveness by sex because of the differences in the natural history of the disease between men and women. The aim when building an economic model is to simplify reality whilst still maintaining the fundamental elements of the disease. Gender seems to play an important role and so it would be appropriate to model men and women separately where possible. In addition, if certain groups of the population can be identified at higher risk than others, then preventative treatment and awareness campaigns can be targeted to these populations.

The findings from this study entirely support those published a decade earlier, namely that there is an important link between impaired lung function and mortality. The extension to the original work confirmed the link between impaired lung function and COPD mortality with large and highly significant HRs particularly for group 1, which were much larger than any for other diseases replicated here or published by Hole et al in their paper. Group 1 are likely to closely match a COPD population, diagnosed based on impaired lung function. However it is known that a COPD case may be identified based on more than impaired lung function, and some guidelines for diagnosing COPD (as presented in Chapter 2), explicitly include other elements for inclusion within the diagnosis of COPD. The mortality risk for subjects with impaired lung function along with other risk factors is investigated in detail in the following section.

5.6 NICE vs GOLD

It has been previously stated (in Chapter 2) that between 4 and 10% of the adult population worldwide have COPD,(78) with population estimates varying considerably according to the age group under study and the diagnostic criteria used,(69;73) and there is no clear agreement over which criteria are most appropriate. As seen in Chapter 2 most international guidelines agree on the central importance of the ratio of FEV_1 to FVC, however sole reliance on this measure is believed to under-diagnose in the young and over-diagnose the disease in the elderly.(195) GOLD diagnoses COPD on the basis of airflow limitation alone ($FEV_1/FVC<0.7$).(2) Reporting of disease prevalence from the BOLD study (see Chapter 2), a large study on the burden of COPD worldwide, used GOLD stage II and above, defined as airflow obstruction ($FEV_1/FVC<0.7$ and $FEV_1<80\%$ predicted).(79;81) NICE in the UK suggests that identification of disease be based on airflow obstruction, a risk factor and respiratory symptoms.(7)

NICE and GOLD have different remits which may explain the differences in criteria, GOLD is "to raise awareness of COPD and to improve prevention and treatment",(196)

and NICE is to provide recommendations on appropriate treatment and care for people with specific diseases (such as COPD) within the UK NHS, based upon 'best' available evidence.(86)

Previous studies have compared COPD prevalence under different COPD diagnostic criteria,(70;73-76;78;82;197;198) but none have compared the GOLD, GOLD II+ and NICE criteria. Some studies have used respiratory symptoms to identify COPD either alone or in combination with airflow limitation/obstruction,(78;199) but a COPD diagnostic criteria incorporating a risk factor, in particular smoking history, has not previously been investigated. In addition, while the impact of airflow limitation on mortality has been demonstrated by Hole et al,(166) and again in the previous section, the effect of $FEV_1/FVC<0.7$ and $FEV_1<80\%$ predicted, in combination with smoking history and respiratory symptoms is unknown.

Diagnostic criteria for COPD are difficult to validate, partly due to the chronic nature of the disease. The Renfrew/Paisley study offers a unique opportunity to compare GOLD, GOLD II+ and NICE diagnostic criteria because of the data gathered at baseline and the lengthy follow-up of mortality.

When conducting an economic evaluation of a treatment it is important to correctly identify the patient population before any modelling of the disease gets underway. Within this section an analysis to compare NICE criteria to the most frequently used diagnostic criteria, the GOLD diagnostic criteria, is presented. The principal aim of the analysis was to determine how diagnosing COPD based on a smoking history and respiratory symptoms in addition to airflow limitation and/or airflow obstruction, impacts upon prevalence estimates and mortality risk for all-cause and COPD mortality, compared to a diagnosis based only on lung function. A secondary aim was to determine if smoking history, respiratory symptoms, $FEV_1/FVC<0.7$ and $FEV_1<80\%$ predicted are independently predictive of all-cause and COPD mortality risk in the Renfrew/Paisley (MIDSPAN) dataset.

5.6.1 Methods

Three different diagnostic criteria for COPD were applied to subjects in the dataset. These have previously been stated, but are reproduced here to enable comparisons between the criteria to be made:

GOLD:"COPD is characterized by airflow limitation"(2) defined as $FEV_1/FVC<0.7$.

GOLD II+: COPD is characterised by airflow obstruction ($FEV_1/FVC<0.7$ and $FEV_1<80\%$ predicted).

NICE: a diagnosis of COPD should be considered in patients aged over 35 with airflow obstruction ($FEV_1/FVC<0.7$ and $FEV_1<80\%$ predicted), a risk factor (principally smoking) and who present with one or more of: exertional breathlessness, chronic cough, regular sputum production, frequent winter 'bronchitis' or wheeze.(7)

Assumptions were made to identify NICE COPD cases within the dataset:

- Airflow obstruction: $FEV_1/FVC<0.7$ and $FEV_1<80\%$ predicted.

- Respiratory symptom(s): one or more of breathlessness, wheeze, phlegm (see figure 5.3).

- Risk factor: either a smoking history of ten pack years or more, or being a pipe/or cigar smoker.

All subjects were aged 45 and over, therefore the age criterion was satisfied. As all three diagnostic criteria require airflow limitation ($FEV_1/FVC<0.7$), the more restrictive NICE criteria form a subgroup within the GOLD groups. It was of interest to follow those subjects who had airflow limitation/obstruction but who did not fulfil all requirements to enter the NICE COPD group. These subjects are referred to as the 'low risk' group.

Statistical Analysis

Comparisons were made between the NICE and GOLD and the NICE and GOLD II+ COPD diagnostic criteria. All analyses were carried out on men and women separately. Prevalence estimates were calculated by age group and diagnostic criteria. Kaplan-Meier curves were produced for the full follow-up period for the 'no COPD' vs 'low risk' vs NICE groups only (as GOLD='low risk'+NICE), with age along the time scale. As previously described with reference to figure 5.4, where the denominator is small (in the dataset at the extremes of age), small numbers of events can cause large effects on the Kaplan-Meier curve. To avoid this, the graphs were started from age 46 years onwards. Kaplan-Meier curves were used to compare mortality risk using the different diagnostic criteria, by disease severity.

Cox proportional hazards models were used, adjusted for risk factors (age, diastolic blood pressure, cholesterol, BMI, social class and ex-smoker). The assumption of proportional hazards was violated when using the full follow-up period, therefore the data were analysed in two time periods (≤ 20 years and >20 years follow-up). The split represents the point at which half of all the COPD deaths occurred. Entrance to the 2nd period was conditional on surviving/remaining in the study for the first 20 years. All Cox regressions were run in both time periods. The proportionality assumption (tested using Schoenfeld residuals) was found to perform better following the split.

Analyses were run using both methods of time scale within the survival analysis. Employing time in study was found to meet the proportionality assumption more times over all the analyses in the study than employing age in study as the time scale. Analyses with age in study were therefore conducted to represent the survival profile visually. All other analyses were conducted with time in study as the time scale and adjusted for age.

Two sets of Cox proportional hazards models were run, the first set used GOLD and NICE to identify a 'low risk' group and compared 1) 'no COPD' vs GOLD and 2) 'no COPD' vs 'low risk' vs NICE. The second set used GOLD II+ and NICE to identify a different 'low risk' group and compared 1) 'no COPD' vs GOLD II+ and 2) 'no COPD' vs 'low risk' vs NICE.

An analysis was carried out on the components of the NICE diagnostic criteria using a Cox proportional hazards model to predict all-cause and COPD mortality with $FEV_1/FVC<0.7$, $FEV_1<80\%$ predicted, smoking history and respiratory symptoms as independent variables and adjusted for risk factors. Statistical significance was defined at the p=0.05 level.

5.6.2 Results

The general population study comprised of 6861 men and 8173 women and is summarised in table 5.11. As can be seen in the table, of the men and women 77% and 46% respectively had 10 or more pack years and a substantial minority, 39% of men and 29% of women, had one or more respiratory symptom. Compared to the women, the men were more likely to suffer from phlegm and wheeze and less likely to report breathlessness. Of the study population, 3542 of the men and 5518 of the women remained in the study after 20 years of follow-up and these formed the dataset for the second period of follow-up.

Table 5.11 Baseline characteristics of the study population, Renfrew/Paisley (MIDSPAN)

	Men	Women
Participants, n (%)	6861 (46)	8173 (54)
Participants >20yrs follow up, n (%)	3542 (39)	5518 (61)
Years of follow up, mean (SD)	19.8(9.6)	23.2(8.9)
Years of follow-up (range)	0-34	0-34
Age, mean (SD)	54.6(5.6)	54.9(5.6)
Social Class (mode)	IIIM	IV
Body Mass Index, mean (SD)	25.9(3.4)	25.8(4.5)
Smoking ≥ 10yr pack yrs, n (%)	5264 (77)	3757 (46)
Presence of symptoms, n (%)	2644 (39)	2389 (29)
Presence of phlegm, n (%)	2107 (31)	1382 (17)
Presence of breathlessness, n (%)	920 (13)	1332 (16)
Presence of wheeze, n (%)	1206 (18)	961 (12)

Prevalence

The prevalence of COPD in the study population was high. Overall the prevalence of COPD was 25% using the GOLD diagnostic criteria, 15% with the GOLD II+ and 8% with the NICE criteria. Prevalence of COPD was higher in men than in women and this is clearly shown in table 5.12. COPD prevalence generally increased with age and was highest in men aged 60-64.

Table 5.12 COPD prevalence by diagnostic criteria and age, men and women

	Age category	Total number	Prevalence, % (n)		
			GOLD	GOLD II+	NICE
Men	45-49	1774	25(441)	16(282)	9(156)
	50-54	1934	28(546)	18(346)	10(197)
	55-59	1641	34(554)	23(385)	13(217)
	60-64	1512	41(622)	29(437)	19(284)
	Overall	*6861*	*32(2163)*	*21(1450)*	*12(854)*
Women	45-49	1997	17(338)	12(238)	5(94)
	50-54	2249	19(431)	13(300)	5(107)
	55-59	1997	23(463)	16(320)	6(127)
	60-64	1930	21(401)	13(243)	3(60)
	Overall	*8173*	*20(1633)*	*13(1101)*	*5(388)*

Figure 5.6 illustrates diagrammatically how the COPD population is affected by the diagnostic criteria. The 'no COPD' population is dependent upon the criteria used and is larger when applying the NICE criteria than either the GOLD II+ or GOLD criteria.
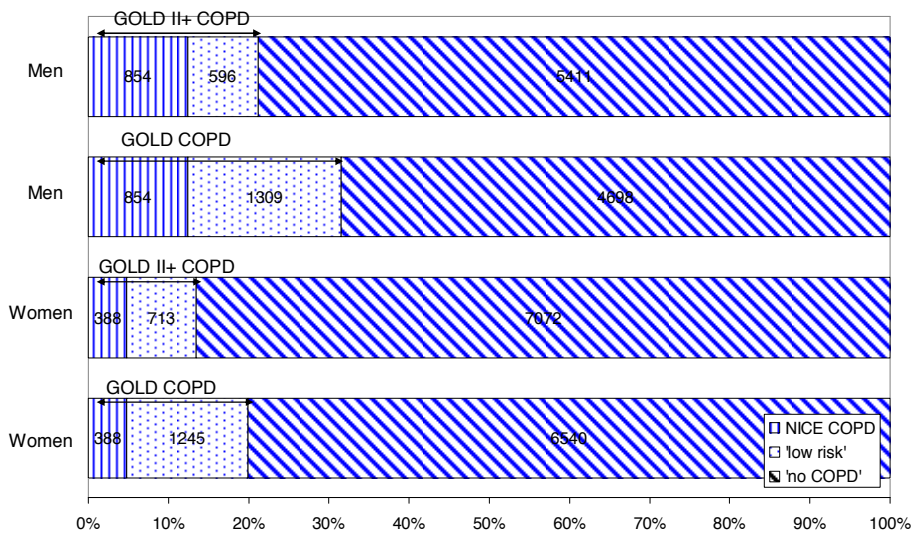
Figure 5.6 COPD prevalence applying the GOLD, GOLD II+ and the NICE diagnostic criteria to the Renfrew/Paisley dataset in men and women, n

The 'low risk' group is smaller when using the GOLD II+ criteria in conjunction with the NICE criteria in both men and women, as opposed to applying the GOLD criteria in conjunction with the NICE criteria as illustrated in figure 5.6. The 'low risk' population shown in the top column for the men and the top column for the women of figure 5.6 is made up of study participants with airflow obstruction but without both respiratory symptoms and a smoking history. Because of the high proportion of the dataset with a smoking history: 77% of men and 46% of women had 10 or more pack years as described in table 5.12, it is likely that the majority of those in this 'low risk' group would have airflow obstruction and a smoking history, but no respiratory symptoms.

The 'low risk' population shown in the bottom column in figure 5.6 for men and the bottom column for women, potentially contains any subject who has an $FEV_1/FVC<0.7$ but not all of: $FEV_1<80\%$ predicted plus a smoking history and respiratory symptoms. The difference between the top and bottom columns is that lung function must be $FEV_1<80\%$ predicted in the top columns. Therefore it can be seen that the addition of this lung function criteria is highly influential in the diagnosis of COPD.

Survival

Survival curves for all-cause, and COPD mortality, which are shown below in figures 5.7 to 5.10 showed clear separation between the NICE, 'low risk' and 'no COPD' groups (log rank p<0.001) in men and women. This is important because it clearly illustrates that the groups are different from one another in terms of survival probability with the NICE COPD group having the lowest survival probabilities. The 'low risk' group is separated from the no COPD group, highlighting the increased mortality risk relative to the 'no COPD' group of those within this 'low risk' classification.
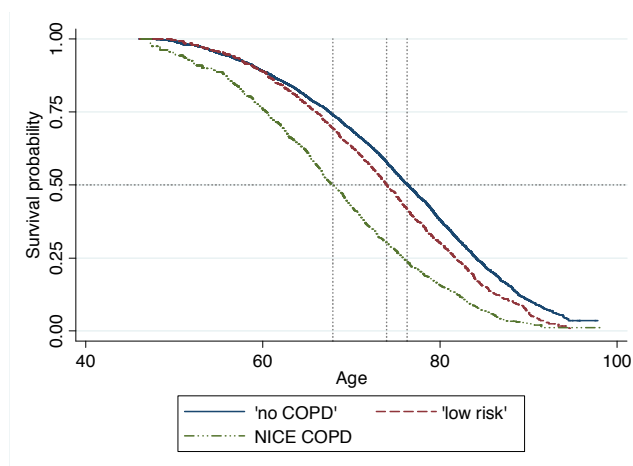
Figure 5.7 Survival curves for all-cause mortality (men)
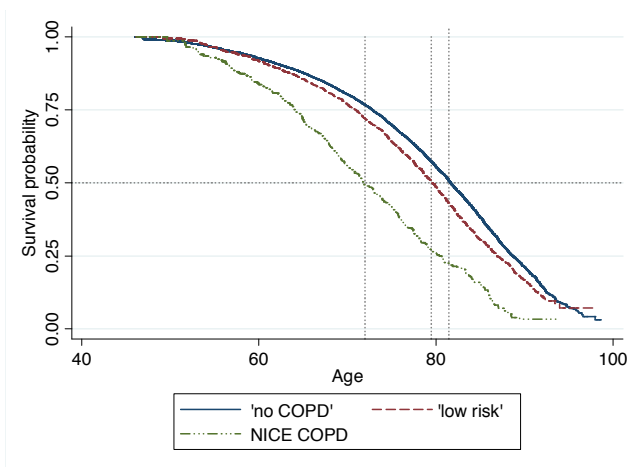
Figure 5.8 Survival curves for all-cause mortality (women)
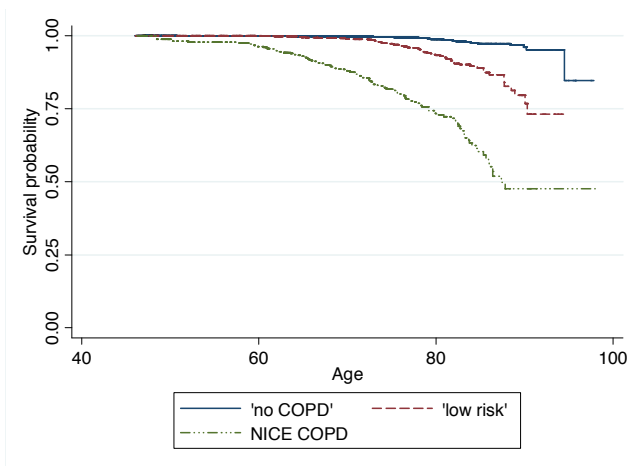
Figure 5.9 Survival curves for COPD mortality (men)
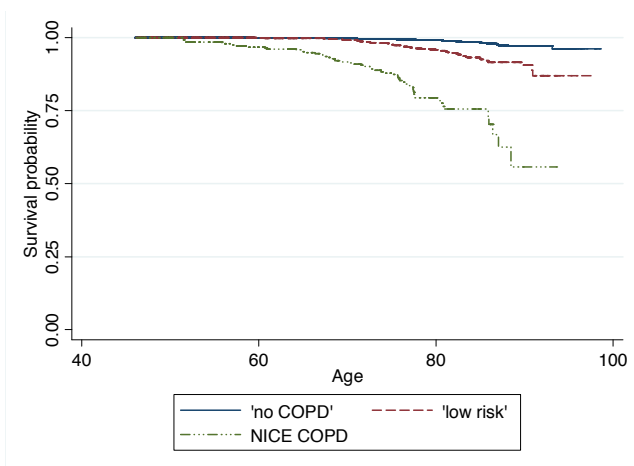
Figure 5.10 Survival curves for COPD mortality (women)

The median age of survival can be read off the x axis when the survival curve is equal to 0.05 (where the dotted horizontal line meets the dotted vertical line). The horizontal line running through the 0.50 survival probability for all-cause mortality gives a median age of survival for men of 68 years if they are categorised in the NICE COPD group, 74 if they are in the 'low risk' group and 76 for those with 'no COPD'. For women, the median age was higher at 72 years, 80 years and 82 years respectively.

Because the survival curves for COPD mortality incorporates censoring as a result of death from causes other than COPD, these curves illustrate the degree of co-morbidity within COPD patients who often die from causes other than COPD (as previously shown in figure 5.5). Survival probability for the 'no COPD' is close to 1 throughout the observed period.

All-Cause Mortality

Table 5.13 below presents the results from four separate Cox proportional hazards models, comparing HRs between the different diagnostic criteria in terms of all-cause and COPD mortality. Compared to 'no COPD' (where the HR=1) HRs for men (women) were 1.41 (1.47) for GOLD and 1.48(1.64) for GOLD II+ in ≤20 years of follow-up, suggesting a slightly higher mortality risk in the GOLD II+ group than GOLD.

As shown in table 5.13, the 'low risk' group (GOLD) and the 'low risk' (GOLD II+) had statistically significant but lower HRs for mortality than the HRs in the NICE groups for follow up ≤20yrs. In men (women) HRs were 1.19 and 1.81 (1.24 and 2.42) for 'low risk' (GOLD) and NICE respectively compared to 'no COPD', and 1.15 and 1.76 (1.32 and 2.40) for 'low risk' (GOLD II+) and NICE. In both time periods, the HRs for the NICE group were considerably larger than the HRs in the 'low risk' group, suggesting that participants with airflow obstruction, respiratory symptoms and risk had a higher all-cause mortality risk than those with airflow limitation and/or obstruction

alone. The low risk group from the GOLD II+ versus NICE analysis with follow up >20 years in men had a mean value of 1.11 (95%CI: 0.95-1.30) and confidence interval that included 0, implying that this group cannot be considered significantly different to the no COPD group.

Table 5.13 HRs from four separate Cox PH regression models:
1) 'no COPD' vs GOLD 2)'no COPD' vs 'low risk' vs NICE (GOLD='low risk'+NICE) 3) 'no COPD' vs GOLD II+ 4)'no COPD' vs 'low risk' vs NICE (GOLD II+='low risk'+NICE), follow-up ≤ 20 years (n=6861 men and 8173 women) and > 20 years (n=3542 men and 5518 women) for all-cause and COPD mortality, men (top) and women (bottom)

Cause of death/ follow-up	Dead %(n)	GOLD Hazard Ratios (95%CI)			GOLD II+ Hazard Ratios (95%CI)		
		GOLD	'low risk'	NICE	GOLD II+	'low risk'	NICE
All-cause							
≤20 yrs	48 (3275)	1.41(1.31-1.51)	1.19(1.09-1.30)	1.81(1.65-1.99)	1.48(1.37-1.60)	1.15(1.03-1.30)	1.76(1.61-1.93)
>20 yrs	59 (2106)	1.34(1.22-1.48)	1.26(1.12-1.40)	1.58(1.36-1.84)	1.29(1.15-1.45)	1.11(0.95-1.30)	1.52(1.31-1.77)
COPD							
≤20 yrs	2 (139)	20.57(11.02-38.41)	6.69(3.27-13.67)	48.81(25.87-92.10)	22.03(13.30-36.50)	7.67(3.97-14.80)	35.08(20.98-58.66)
>20 yrs	3 (106)	6.14(4.03-9.34)	4.58(2.87-7.31)	10.94(6.63-18.05)	5.41(3.64-8.04)	3.72(2.22-6.23)	7.83(4.96-12.37)
All-cause							
≤20 yrs	32(2593)	1.47(1.34-1.60)	1.24(1.12-1.38)	2.42(2.10-2.80)	1.64(1.48-1.81)	1.32(1.16-1.50)	2.40(2.08-2.77)
>20 yrs	50(2785)	1.27(1.15-1.39)	1.16(1.04-1.28)	1.95(1.62-2.36)	1.44(1.29-1.61)	1.29(1.13-1.47)	1.95(1.62-2.35)
COPD							
≤20 yrs	1 (73)	8.87(5.25-15.00)	4.38(2.34-8.20)	27.73(15.54-49.51)	9.86(6.05-16.07)	4.28(2.19-8.34)	22.45(13.17-38.26)
>20 yrs	2 (107)	5.22(3.55-7.67)	3.83(2.47-5.93)	12.86(7.67-21.56)	5.91(3.99-8.76)	4.10(2.51-6.68)	11.22(6.79-18.55)

Regressions were adjusted for age, blood pressure, cholesterol, body mass index, social class and ex-smoker .

COPD Mortality

Of those deceased, COPD mortality was the cause of death for 17% (16%) of men (women) in the NICE group, 6% (6%) of the 'low risk' (GOLD) and 11% (9%) of those diagnosed with COPD by GOLD. This compared to 8% (7%) of the 'low risk' (GOLD II+) group and 13% (10%) of the GOLD II+ group.

HRs in both GOLD groups for COPD mortality were large and comparable and were smaller for women than men as seen in table 5.13. The HR for GOLD men (women) was 20.57 (8.87) and for GOLD II+ was 22.03(9.86) compared to 'no COPD'.

Splitting participants diagnosed with COPD according to the GOLD/GOLD II+ criteria into a 'low risk' group and a NICE group resulted in a clear trend of increasing risk with increasing disease severity. The 'low risk' groups had statistically significant but smaller HRs for COPD mortality than the HRs for NICE in follow up ≤20yrs for men and in both time periods for the women. In men (women) HRs were 6.69 and 48.81 (4.38 and 27.73) for 'low risk' (GOLD) and NICE respectively, and 7.67 and 35.08 (4.28 and 22.45) for 'low risk' (GOLD II+) and NICE. These results illustrate that participants meeting NICE criteria had a higher COPD mortality risk than those with airflow limitation and/or obstruction alone.

Individual Components of the NICE Diagnostic Criteria

The regression model on mortality (all-cause and COPD) identified that in general, smoking history (≥10 pack years), respiratory symptoms, FEV_1/FVC <0.7 and predicted FEV_1<80% were found to independently predict all-cause and COPD mortality in men and women with large and significant HRs. Results from the analysis are presented in table 5.14.

Table 5.14 HRs from a Cox PH regression model on the individual components of the NICE criteria:
FEV_1/FVC, $FEV_1 <80\%$ predicted, smoking history and respiratory symptoms and with follow-up split into ≤ 20 years and > 20 years for all-cause and COPD mortality, men (top) and women (bottom)

Cause of death/ follow-up	Hazard Ratios (95%CI)			
	$FEV_1/FVC<0.7$	$FEV_1<80\%$ predicted	≥ 10 pack years	Respiratory symptoms
All-cause				
≤20 years	1.10(1.01-1.20)	1.26(1.16-1.37)	1.71(1.55-1.89)	1.30(1.20-1.40)
>20 years	1.16(1.03-1.30)	1.08(0.96-1.22)	1.49(1.34-1.65)	1.29(1.17-1.42)
COPD				
≤20 years	5.65(2.92-10.95)	6.13(3.26-11.55)	1.78(0.92-3.44)	6.86(3.83-12.28)
>20 years	3.22(1.96-5.29)	2.02(1.25-3.24)	7.10(2.85-17.72)	1.75(1.17-2.63)
All-cause				
≤20 years	1.13(1.02-1.25)	1.35(1.23-1.48)	1.64(1.51-1.78)	1.37(1.26-1.50)
>20 years	1.02(0.92-1.14)	1.33(1.21-1.46)	1.63(1.50-1.76)	1.21(1.11-1.32)
COPD				
≤20 years	3.01(1.70-5.32)	5.00(2.53-9.88)	3.62(1.85-7.09)	5.24(2.87-9.56)
>20 years	2.52(1.62-3.92)	2.59(1.65-4.07)	6.02(3.54-10.23)	1.93(1.29-2.88)

The regressions were adjusted for age, blood pressure, cholesterol, BMI, social class and ex-smoker
FEV1 = forced expiratory volume in one second, FVC= forced vital capacity

Larger HRs for the individual components of the NICE criteria were seen in the first period of follow-up (≤20 years) than in the second (>20 years), except for the HRs relating to smoking history for COPD mortality where a smoking history was seen to have a stronger relationship with mortality in the second period of follow up compared to the first. Results presented in table 5.14 show that in the first 20 years the HR for smoking history in men was not statistically significant: HR:1.78 (95%CI:0.92-3.44) yet for follow-up of over 20 years, the HR was large and highly significant at 7.10 (95%CI:2.85-17.72). The table also provides evidence showing an increase in hazards for women with a smoking history, where the HR was seen to increase over time from 3.62 to 6.02.

5.6.3 Discussion

This study assessed how the presence of respiratory symptoms and a risk factor in addition to airflow limitation and/or airflow obstruction impacts upon disease

prevalence and mortality risk for all-cause and COPD mortality within a general population cohort whose survival was followed from the 1970s.

Prevalence of COPD was heavily dependent on the diagnostic criteria used and was found to be anywhere between 8% (NICE) and 25% (GOLD). In terms of estimates of burden of disease and subsequent resource allocation, the implications of potentially treating an additional 17% of the population are substantial. Misclassified cases with people free from the disease are likely and treating false positives may result in quality of life decrements to the individuals involved and unnecessary resource use for the payer. Nevertheless, in treating any disease it is important to identify those most at risk of key events such as hospitalisations and disease specific mortality, which would not only lead to a drop in quality of life or ultimately a premature end to life for the individual, but also to a financial burden for the NHS.

The implications of splitting a GOLD COPD diagnosed population into a NICE group and a 'low risk' group were examined and it was found that doing so identified groups that are statistically significantly different. The NICE group contained participants meeting the NICE diagnostic criteria who had respiratory symptoms, smoking histories of >10 pack years and airflow obstruction ($FEV_1/FVC<0.7$ and $FEV_1<80\%$ predicted), and were found to be at high risk from all-cause and COPD mortality compared to a 'low risk' group. The 'low risk' group contained subjects with either airflow limitation ($FEV_1/FVC<0.7$) (GOLD) or airflow obstruction (GOLD II+), without both of respiratory symptoms and a smoking history. Thus it is the inclusion of respiratory symptoms and smoking history that identifies those at highest risk. The 'low risk' group is of interest and raises the question, to what extent, if any, should clinical management for the 'low risk' group differ from the approach pursued for the NICE COPD group? The answer to this question lies with the clinical experts and should be debated given the evidence presented here.

The components of the NICE criteria: respiratory symptoms, smoking history, $FEV_1/FVC<0.7$ and $FEV_1<80\%$ predicted, were identified as independent predictors of all-cause and COPD mortality within a Cox proportional hazards model.

A substantial number of the study population were diagnosed with COPD using GOLD criteria but 'no COPD' with GOLD II+, 10% (n=713) of men and 7% (n=532) of women (representing those with $FEV_1/FVC<0.7$ and $FEV_1≥80\%$ predicted). Applying GOLD II+ instead of GOLD reduced prevalence estimates from 25% to 15%. HRs for those meeting GOLD criteria to 'no COPD' in terms of mortality remained similar to HRs for GOLD II+ compared to 'no COPD'.

Strengths and Limitations

GOLD criteria stipulate the use of post bronchodilator spirometry to measure lung function but this method was not standard procedure at the time of the study. This may have some impact on the absolute numbers diagnosed with COPD but is unlikely to affect the key messages from the study.

The diagnostic criteria used here are based on an interpretation of the NICE guidelines with specialist input from a respiratory clinician (Malcolm Shepherd), firstly around the assumption that smoking history is the only risk factor and additionally that it should be set at 10 pack years. Secondly that evidence of respiratory symptom(s), set by the inclusion criteria applied, was limited by the questions asked within the study.

This comparative analysis is the first to use NICE diagnostic criteria in a general population cohort and has shown that these are practical to employ in a large prospective study. The information necessary to diagnose based upon the NICE criteria are readily available to the clinician and there is evidence that the NICE diagnostic criteria reflect physician diagnosed COPD more closely than GOLD criteria.(200) In particular, fewer false positive diagnoses will be made using NICE criteria than with

GOLD, especially in the elderly who are most at risk of misclassification.(195) The problem of false negatives may arise but the recognition of a 'low risk' group would resolve this issue.

The organisations behind these guidelines arguably succeed in their respective remits. GOLD adopts a blanket approach and flags up a large population of COPD cases, thus raising awareness of the disease as a substantial issue. NICE on the other hand is focused on providing recommendations on appropriate treatment and care, so identifying those most in need of treatment is important. However by restricting COPD cases to those identified using NICE criteria ignores a group at 'low risk' but nonetheless higher mortality risk than those with 'no COPD'. Recognising higher mortality risk in those identified by NICE whilst recognising increased risk in those at 'low risk', bridges the gap between the GOLD and the NICE diagnostic criteria.

5.6.4 Conclusion

More restrictive COPD criteria identify subjects at higher risk of all-cause and COPD mortality than using the GOLD criteria alone. Diagnosing COPD based upon airflow obstruction, respiratory symptoms and smoking history as opposed to airflow limitation/airflow obstruction splits the COPD population as diagnosed by GOLD/GOLD II+, into two, one at 'high risk' and another at 'low risk'. These two groups have statistically significant mortality risks and may benefit from different clinical management strategies.

5.7 Hospital Admissions

The major drivers of costs to the NHS arising from COPD are disease severity and severe exacerbations, for example as described in Chapter 2, COPD patients have been found to occupy approximately one million bed days annually.(201) Severe exacerbations have been described as when the patient (or caregiver) recognises an

obvious and/or a rapid deterioration in their condition and requires hospitalisation.(202) They are typified by one or more of: increased shortness of breath; increased volume and purulence of sputum; increased cough and shallow/rapid breathing.(202) An exacerbation is the main reason why a COPD patient would attend hospital.

Reducing or preventing disease progression and/or hospital admission would have a direct effect on the total cost burden for COPD,(38) and on quality of life for patients. Therefore, reducing or preventing disease progression and/or hospital admission are often principal outcome measures for clinical trials in COPD.

As seen in Chapter 4, due to the physical detriment and economic cost associated with exacerbations, exacerbations are often modelled within cost effectiveness analyses of COPD. In brief, exacerbations are predominantly built into the disease states (usually mild, moderate and severe).(92;141;144;146;203) Each health state has an exacerbation (usually mild and severe) probability attached which varies by disease state and by treatment group.(92;141;144;146;203) To date all information on COPD exacerbations used within decision analytic models have come from clinical trials.(92;141;144;146;203)

As previously explained in section 5.2, the Renfrew/Paisley dataset with its sizeable COPD population has been linked to hospitalisation data as well as to death records. Within this section the feasibility of using observational data to study the course of severe hospitalisation for COPD patients was investigated, for the purpose of informing an economic model. The rate of hospitalisation, length of stay in hospital and number of hospitalisations per participant were studied.

5.7.1 Methods

All linked respiratory coded hospital discharges for the MIDSPAN dataset (SMR1 recording scheme) occurring in Scotland between 1972 and December 1995 were

obtained. All COPD coded hospital admissions within the dataset were regarded as severe exacerbations on the grounds that an exacerbation is the main reason why a COPD patient would attend hospital (clinical opinion - Malcolm Shepherd, personal communication).

COPD hospital admissions (ICD-8 491-492, ICD-9 490-492 & 496, ICD-10 J40-J44) were identified, in any diagnostic position (in any of the principal or secondary (up to five) diagnostic coding positions). Hospital admission stays of more than 200 days (n=11) were removed from the dataset as these were unlikely to represent (acute) COPD hospital exacerbations. COPD cases and corresponding disease severity were identified within the dataset according to NICE guidelines, as previously described in section 5.4.

Statistical Analysis

Missing data was dealt with as outlined in section 5.2.1. The percentage of participants in each COPD severity group at baseline, with a COPD hospitalisation at any point within the follow up period was identified from the data. The percentage of participants within each COPD severity group with a COPD coded death was also determined.

Hospitalisation rates were calculated by first summing the frequency of hospitalisations in each severity group and then dividing this value by the number of person years at risk for the severity group. The number of times each participant was hospitalised over the duration of follow up was determined and depicted in a histogram. Length of stay in hospital was derived from the data and also illustrated using a histogram. Length of stay was calculated by subtracting the date of discharge from the admission date (once any transfers had been accounted for). A value of one was added to all length of stay values to adjust for day cases, because the admission date is equal to the discharge date and otherwise, the length of stay would have been 0 days.

5.7.2 Results

Of the study population, 10% of men and 8% of women had a COPD hospital admission. The proportion of study participants with a COPD hospital admission was shown to increase in line with baseline disease severity. COPD hospitalisation was proportionally higher in women than in men, regardless of disease severity. Table 5.15 shows that 33% of the women with mild COPD at baseline had a COPD coded hospital admission in comparison to 25% of the men in the same group, this compares to 42% of the women in the moderate group and 36% of men, and in the severe group, 56% and 49% respectively had a hospitalisation.

Table 5.15 Percentage of participants with a COPD coded hospital admission in any diagnostic position by COPD disease severity at baseline

		N	% (N) with COPD hosp admission(s)
Men & women	No COPD	13 792	7 (915)
	Mild	847	27 (232)
	Moderate	300	38 (114)
	Severe	95	51 (48)
Men	No COPD	6007	7 (441)
	Mild	580	25 (144)
	Moderate	204	36 (74)
	Severe	70	49 (34)
Women	No COPD	7785	6 (616)
	Mild	267	33 (88)
	Moderate	96	42 (40)
	Severe	25	56 (14)

COPD hospitalisation rates were seen to increase with disease severity as shown in table 5.16, in both men and women. In the male (female) group, for every 10 000 person years at risk in the severe COPD group, there were 1452 (1163) hospitalisations, compared to 373 (458) for every 10 000 person years at risk in the mild group and 89 (73) in the no COPD group.

Table 5.16 COPD hospitalisation rates by disease severity, men and women

		N in group	N hospitalisations	Person yrs at risk	Rate per 10 000 person yr
Men & women	No COPD	13792	2426	305742	79
	mild	847	594	14733	403
	moderate	300	392	4148	945
	severe	95	120	902	1331
Men	No COPD	6007	1093	123209	89
	mild	580	355	9511	373
	moderate	204	227	2619	867
	severe	70	76	523	1452
Women	No COPD	7785	1333	182533	73
	mild	267	239	5222	458
	moderate	96	165	1529	1079
	severe	25	44	378	1163

Of those hospitalised with COPD, the modal number of hospital admissions in both men and women was one, as seen in figures 5.11 and 5.12: 51% of men and 47% of women had one hospital admissions over the follow up period. The distribution around number of hospitalisations was positively skewed in men and women. The largest number of COPD hospital admissions was seen in the female group, with one woman having 39 admissions compared to one man having 34 admissions. Nevertheless, most of the participants (98%) had eleven or fewer hospital admissions over a maximum follow up period of 34 years. The mean number of COPD hospitalisations was three for both men and women, in those people hospitalised for COPD.
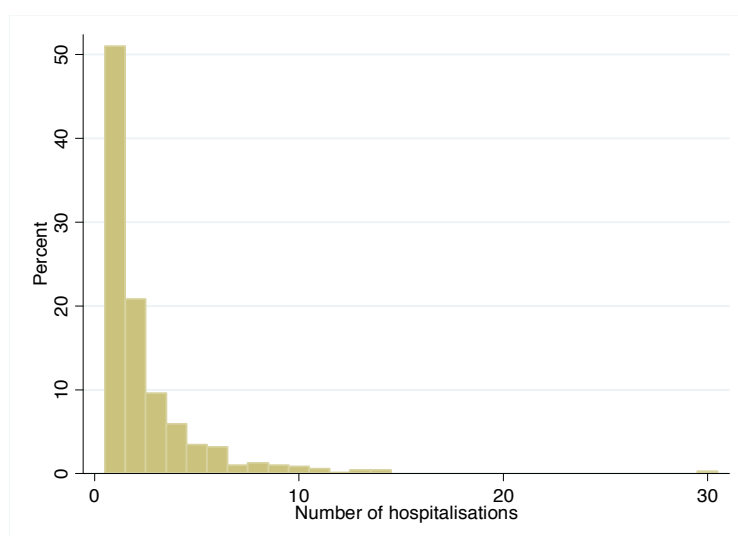
Figure 5.11 Number of times hospitalised, men
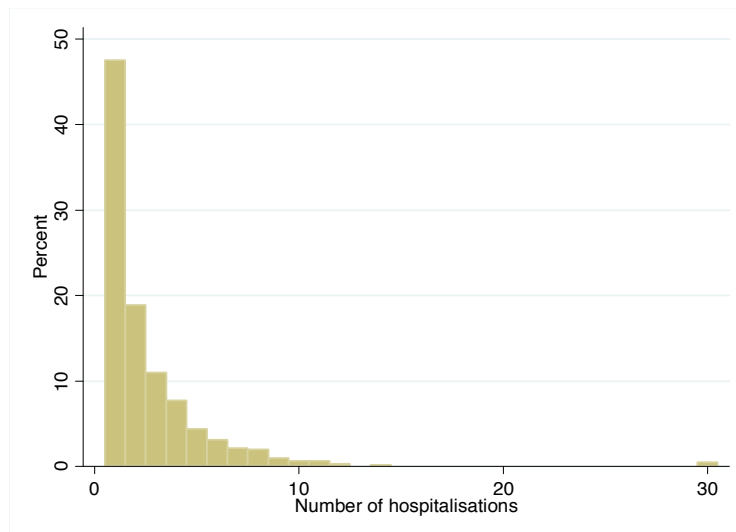
Figure 5.12 Number of times hospitalised, women

Figures 5.13 and 5.14 illustrate the length of hospital stay for men and women over the study period, for those hospitalised with COPD. Of the men (women), 50% had a length of stay of 8 (9) days or fewer and 90% of hospital admissions were less than or equal to 28 (27) days. The distribution around duration of stay was positively skewed. Less than 4% of hospital stays in both men and women were for more than 50 days.
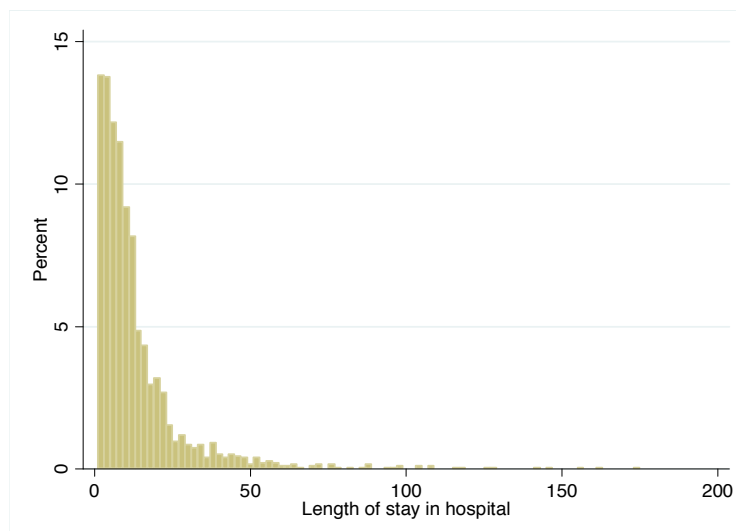
Figure 5.13 Length of stay in hospital, men
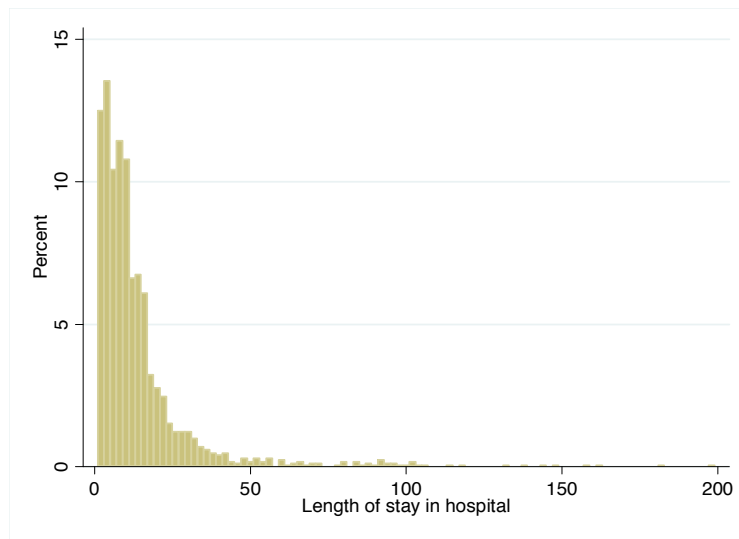
Figure 5.14 Length of stay in hospital, women

5.7.3 Summary

Hospitalisation rates increased with disease severity in both men and women. Most of those hospitalised during the follow up period had one hospitalisation, and length of stay in hospital was found to be nine days or fewer for 50% of those with a COPD hospitalisation. A significant proportion of participants diagnosed with COPD did not experience a hospital admission.

Whilst this dataset has value in studying the pattern of hospitalisation over a longer time frame, a limitation of using the data within economic evaluation is that only severe exacerbations could be studied. Often in economic models it is of interest to study the impact of treatment on both minor and major exacerbations. Whilst data on severe or major hospitalisations could be obtained using the MIDSPAN dataset, data on minor exacerbations would have to be sourced from elsewhere. Another limiting factor is that because there is no information on HRQoL or FEV_1 values, other than at baseline, the relationship between these variables and hospitalisation could not be modelled. The other option with regards to obtaining data on exacerbations is from a RCT. The hospitalisation rates reported in this chapter are likely to be lower than those reported in

a RCT because of the design of each type of study. The extent to which this is the case would be of interest for further study. The use of COPD coding in any position for a hospitalisation may have led to the over reporting of hospitalisation rates, however because the linked hospitalisation data obtained included information on respiratory coded hospitalisation only, the degree to which this is the case is less than otherwise.

5.8 Discussion

The MIDSPAN dataset with over 30 years of follow-up is among the longest and largest yet available for studying the natural history of COPD in a general population. COPD is by definition a chronic condition and most mortality studies are limited by the duration of follow-up. This study reports on 425 COPD deaths from a deceased population of 10 759 (72%) compared to a previously published longitudinal cohort study reporting on 242 of 6709.(204)

Since follow-up began in 1972, all mortality records and (Scottish) hospitalisations for the MIDSPAN cohort who remained in the UK have been linked to the dataset, however coding errors and under-reporting of the disease may exist, particularly in older mortality records as recent shifts in the awareness of COPD may have led to increased recording of the disease on death certificates.

The use of age as the time scale produced mixed results. The proportionality assumption proved to be more difficult to fulfil when using age compared to time in study. The reason behind this is unclear and additional research into this would be beneficial in order to understand this further. Nonetheless, in presenting survival analysis visually via graphs such as Kaplan-Meier curves, the use of age along the x axis is more intuitive than time in study for displaying the survival history of participants within an observational based study design.

One of the biggest issues with the dataset is that information on the participants was only collected at one time point. Therefore as the study population grew older, lung function was bound to deteriorate and a proportion was likely to develop respiratory symptoms. The extent to which any participant with no COPD went on to develop COPD is unknown, but is likely to, in part, explain the no COPD population with COPD coded mortality and the 7% of the no COPD group with a COPD hospitalisation. Of those with mild or moderate COPD at baseline, the proportion developing more severe disease is unknown and is a limitation of the study and the extent to which this impacted on subsequent mortality and hospitalisations is unknown.

5.9 Conclusion

Analyses conducted within this chapter have uncovered useful information on the natural history of disease that can be used to inform an economic model. The data and analyses on mortality are of particular value because of the length of follow up and the size of the study population. The breakdown of cause of death by disease severity revealed a range of causes of death that the COPD patient may face. It is not the case that people with COPD all tend to die of COPD (although a significant proportion will die from COPD). Because of this fact, the use of all-cause mortality, as opposed to disease specific mortality would be appropriate within a modelling framework: the numbers of people with COPD dying from COPD would mean that a focus on COPD mortality may miss important benefits of treatment.

The extension of the Hole et al paper to include COPD mortality reinforced the importance of impaired lung function as a major indicator of COPD mortality. This concept was further examined in the diagnostic criteria section. It is important not to underestimate the importance of diagnostic criteria. As previously shown in Chapter 2 large variations in prevalence rates are seen when using different diagnostic criteria. NICE criteria were found to identify a group of COPD patients at higher risk of all-

cause and COPD mortality, than the COPD population identified using GOLD criteria. Applied to a modelling scenario and looking at treatment in moderate disease, using NICE criteria to diagnose moderate COPD would identify a cohort at higher risk of mortality and possibly events than a GOLD diagnosed moderate group. This is important to note because COPD treatments are often only licensed to specific disease severity groups so it is important at the outset to identify COPD cases accurately. Of note is that pharmaceutical companies employ tight disease criteria for entry into their RCTs which tend to be similar to the diagnostic criteria recommended by NICE and the closer the profile of study participants is to the diseased population, the more likely that values obtained within the study will be externally valid.

Where there is a range of data options for eliciting inputs to the model, it is important that the appropriate dataset is used for answering the question being asked. For example a RCT may be more useful for eliciting information on exacerbations because information on minor and major exacerbations, lung function and HRQoL are collected over time. Epidemiological data can be useful to inform the design and inputs into an economic model for COPD and as seen in this chapter, the Renfrew/Paisley (MIDSPAN) observational dataset can provide important information, particularly around survival and also on identifying the patient population. For an economic model, it is desirable to also use data that has been collected repeatedly over time so that data on aspects that cannot be derived from an observational study are obtained. The following chapter investigates a different type of dataset, a RCT and looks at deriving an important component of cost utility studies, utilities.

Chapter 6. Utilities

The effectiveness of a new drug in relation to current treatment is an important element in demonstrating its cost effectiveness. It has already been shown (Chapter 3) that there is a wide range of outcome measures that can be used within an economic evaluation and the most appropriate for ease of comparison across disease areas is a utility based HRQoL measure such as the EQ-5D, which can be employed to derive QALYs. Analyses within this chapter are focused on deriving values for utilities and QALYs.

Within the context of economic modelling, the utility parameter can be derived from RCTs because trials provide accurate patient level information on the effectiveness of treatment over time. Where a utility based measure such as the EQ-5D has been routinely collected it is relatively straightforward to generate QALYs from these scores. Another option for obtaining QALYs from within a trial is to develop a prediction equation that predicts QALY score based on patient characteristics. Ultimately QALYs are used by decision makers in order to compare treatments in different disease areas on the same scale (discussed in detail in Chapter 3).

Because the primary aim of the investigators who develop RCTs is drug registration rather than cost effectiveness, few of the COPD RCTs have included a utility based measure for HRQoL, instead choosing a disease specific measure, usually the SGRQ (see section 2.3.1). Where economic evaluation of these RCTs is required, a mapping algorithm can be used to link the disease specific measure to a utility based measure, and then this predicted utility score can be used to derive QALYs.

The dataset that was used within this chapter was a large, multinational, multi-site RCT called the Towards a Revolution in COPD Health (TORCH) trial (and was introduced in Chapter 2). The TORCH clinical trial data are described below and issues around missing data discussed. An example of how to obtain QALYs in the face of missing

data are presented followed by a statistical section containing details of the methods that are employed within this chapter.

Summary utility values were derived from the TORCH trial by disease severity and QALYs from the trial data were calculated by treatment group. A number of multivariate regression models were developed to predict: EQ-5D utility scores and QALYs using patient characteristics, so that the utility/QALY scores could be adjusted to take into account patient to patient heterogeneity. The final section of the chapter develops a mapping equation which generates EQ-5D utility scores, for use when a utility based measure is not collected within a clinical trial.

6.1 Data and Modelling Methods

The TORCH trial assessed the efficacy of the combination product salmeterol/fluticasone compared to salmeterol, fluticasone and placebo in 6112 COPD patients over the course of three years.(49) Inclusion criteria included that the subject had to have at least 10 pack years and FEV1 ≤60 % predicted. Of particular importance was that the EQ-5D was collected at regular time periods throughout the three years of the trial, the only major COPD trial to have done so. Because of this, data from this RCT are of use in informing the QALY parameter of economic evaluation for COPD treatments.

Other outcome measures included: the SGRQ, rate of exacerbations and post bronchodilator FEV_1.(49) Measures were repeatedly collected throughout the trial period with full follow up being achieved if the subject had full responses from seven different study days. Due to translations of the EQ-5D questionnaire being unavailable in some languages, the EQ-5D was administered to 4237 out of the 6112 respondents. Therefore the base case dataset for the analyses conducted within this chapter contained those 4237 participants. Attrition occurred over the three years of the trial with fewer observations taken at every successive session as individuals ceased to participate in the

study, or missed a session date. All analyses were carried out using the TORCH dataset.

The EQ-5D was developed in the late 1980s. The EQ-5D defines health in terms of five dimensions: mobility, self care, usual activities, pain/discomfort and anxiety/depression and three possible responses: no problem, a moderate problem or an extreme problem (see appendix for details).(48) This gives 243 different health states.(205) An additional two states of unconscious and dead are included in the final number of health states and bring the total to 245 states.(206) From the EQ-5D score, EQ-5D utility is calculated from a scoring mechanism that uses values that have been derived from the elicitation of population preferences. Originally this elicitation of population preferences was carried out for 42 EuroQoL health states using time trade off methods and then a tariff was developed for all the 245 health states.(207;208) For all EQ-5D scores, the tariff is employed to get the utility value; a utility of one indicates perfect health and a value of zero represents a health state equal to death. Sometimes the EQ-5D drops below 0 and these states indicate a health state worse than death. Within the TORCH study, the EQ-5D scores were transformed into utilities based on the UK tariff, which was developed using responses from a sample of 3000 members of the UK population.

The SGRQ is used to record health status on patients with chronic airflow limitation (including COPD and asthma). Published in 1992, the SGRQ provides a self-administered HRQoL questionnaire for patients, designed to improve upon the original interview administered, non-standardised Chronic Respiratory Disease Questionnaire, by being comparable across different studies and patient populations and contains questions which are used to inform three health domains of: symptoms, activity and impact upon daily life, and a total score (see appendix for details).(121) The total score ranges from 100 which is the worst possible health state, to zero, which is considered the best possible health state.

There are 50 questions in the SGRQ and each questionnaire response has a unique empirically derived 'weight'. The lowest possible weight is zero and the highest is 100. Each domain of the SGRQ (symptoms, activity and impact) is worked out separately in three steps: first the weighted values for all the items are summed; the weights for missed items are then subtracted from the maximum possible weight for each component. The score for the domain is then calculated using the formula:

$$\text{Score} = \frac{\text{Sum of weights from positive items in the domain}}{\text{Sum of weights for all items in the domain}} \times 100$$

The total score is calculated using the same formula, but based on the entire questionnaire rather than any specific domain within it. Each domain carries a different weight when the total score is calculated: approximately one sixth of the total score comes from the symptoms domain (weight = 662.5), one third from activity (weight = 1209.1) and half from impacts (weight = 2117.8). The total score has a weight of 3989.4. The symptoms domain can handle two items of missing data, the activity domain, four and the impact domain six. For this reason the total score and domain scores may be available even when some items are missing.

6.1.1 Missing Data

In this section the issue of missing data is described before the TORCH dataset is examined for patterns of missingness around the EQ-5D and SGRQ variables.

Missing data are a feature of applied analyses and whilst there are various ways of dealing with the problem of missing data, including dropping observations with missing values, imputation, last observation carried forward and inserting the mean value, the issue of missing data nevertheless should be considered.

There are three different types of missingness: Missing Completely At Random (MCAR), Missing at Random (MAR) and Missing Not At Random (MNAR). MCAR

is said to occur when the probability of an observation being missing is unrelated to either observed or unobserved variables. Consistent results can be obtained using the same analyses that would have been used had there been no missing data. Analysing data in which the missingness is MCAR leads to valid inferences.(209) Missing at Random (MAR) occurs when given the observed data, the missingness mechanism does not depend on the unobserved data, and for example missing data may be related to gender, with men more likely to miss study day appointments than women. When working with a regression model with missing observations that are MAR, as long as the variables predictive of missing responses are included as independent variables within regression equations, the analysis will be valid.(209) Missing Not At Random (MNAR) occurs when neither MCAR or MAR hold. This means that even accounting for all the available observed information, the reason for the missing observations is dependent on unobserved observations.

Checking for Patterns of Missingness

Three sets of the TORCH data were compared for patterns of missing data: datasets 1, 2 and 3.

Dataset 1 used all the data on participants to whom an EQ-5D was administered (minus one subject who had missing FEV_1 and FVC values and who was dropped from all analyses). Dataset 1 contained 4236 participants and 22 532 observations.

Dataset 2, a subset of dataset 1, contained observations with complete data on EQ-5D and SGRQ total score, comprising 3854 participants and 18 503 observations.

Dataset 3, a subset of dataset 2, contained observations with complete data on EQ-5D index, SGRQ total score, SGRQ domain scores and SGRQ item scores and was comprised of 3640 subjects and 14 612 observations.
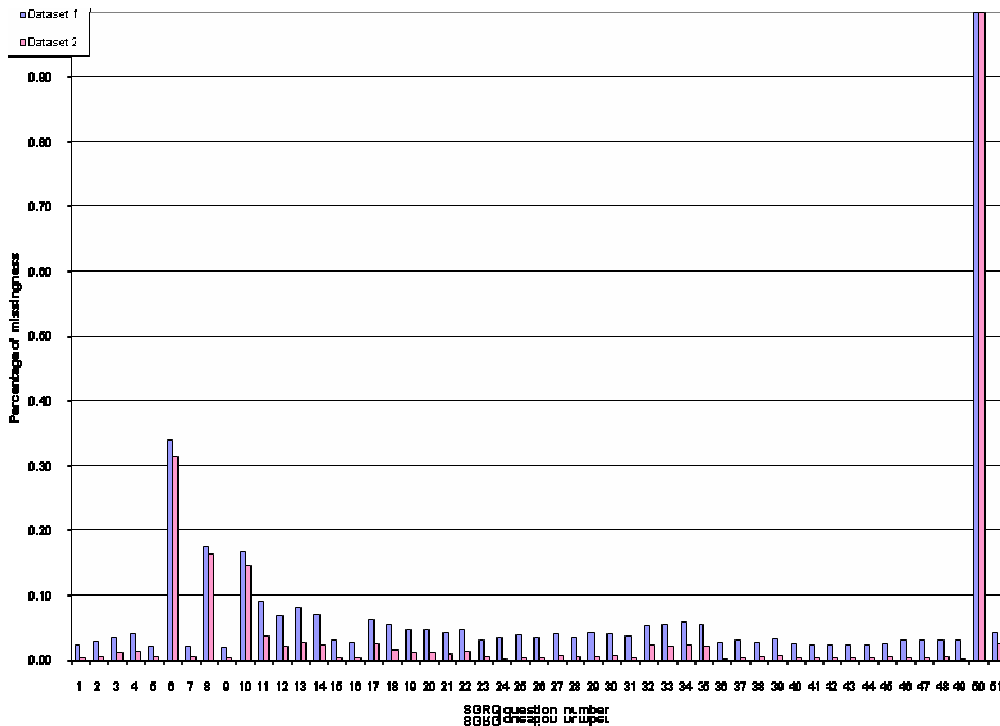
Figure 6.1 Percentage of missingness within each SGRQ question, datasets 1 and 2

Figure 6.1 above illustrates the degree of missingness for each item score in datasets 1 and 2. Before dropping any observation with missing SGRQ item scores to form dataset 3, it was observed that questions 6, 8, 10 and 50 each had a large proportion of missing responses (clearly illustrated in figure 6.1) and that responses were either conditional or optional (see figure 6.2). Because of this, these questions were excluded within dataset 3. This act was justified on the grounds that not only can the total score and the domain scores deal with missing responses for up to 12 items within the SGRQ, but dropping questions 6,8,10 and 50 prevented the subsequent loss of even more observations in dataset 3 and aimed to minimise any potential bias. If question 50 were not removed then the total study population for dataset 3, which contains only complete cases, would amount to one person. As questions 6, 8, and 10 were predominantly related to more severe disease, the result of dropping observations where answers to these questions were not recorded would bias the resulting dataset towards people with worse health.

177

Q6: How long did the worst attack of chest trouble last? (go to Q6 if you had no severe attacks)

Q8: If you have a wheeze, is it worse in the morning?

Q10: If you ever had paid employment…

Q50: Please write down any other important activities that your chest trouble may stop you doing

Figure 6.2 Questions omitted from the dataset because of a high degree of missingness and the conditionality of the question

Table 6.1 presents summary statistics from the three datasets in order to aid identification of any differences between the datasets. The proportion of men in each of the three datasets was the same at 71% and there was no difference in age with 27% of participants aged less than 60 years, 41% were aged 60-69 and 32/33% were 70 years or older. BMI slightly increased from dataset 1 to datasets 2 and 3, though was not statistically significant. Most of the participants in the TORCH study were Caucasian (93-94%) with few people from Black, Asian and 'other' races recruited, and this was the same for all three datasets. The number of deceased participants after the three years was the same in each dataset at 13% of the total study population.

Fewer of the subjects withdrew in dataset 2 (40%) and dataset 3 (38%) compared to dataset 1 (42%), suggesting that subsequent withdrawers were more likely to provide incomplete data during study visits than non-withdrawers during their study visits. The proportion of mild COPD participants was slightly less in dataset 3 than in dataset 1 and there were more participants with severe disease in dataset 3 than in dataset 1, suggesting some potential correlation between disease severity and missingness.

Table 6.1 Patterns of missingness, summary statistics by dataset % (n)

	Dataset 1	Dataset 2	Dataset 3
N	4236	3854	3640
N observations	22 532	18 503	14 612
Men, % (n)	71(2994)	71(2735)	71(2586)
<60yrs, % (n)	27(1134)	27(1026)	27(979)
60-69yrs, % (n)	41(1733)	41(1573)	41(1498)
70yrs +, % (n)	32(1369)	33(1255)	32(1163)
BMI (sd)	25.95(5.21)	26.01(5.28)	26.04(5.30)
Caucasian, % (n)	94(3973)	93(3603)	93(3402)
American Hispanic, % (n)	3(113)	3(105)	3(102)
Black , % (n)	2(91)	2(57)	1(53)
Other, % (n)	1(59)	1(89)	2(83)
Dead, % (n)	13(555)	13(505)	12(451)
Withdrew, % (n)	42(1760)	40(1546)	38(1366)
Mild COPD, % (n)	38(8541)	37(6896)	36(1298)
Moderate COPD, % (n)	50(11175)	50(9243)	50(1822)
Severe COPD, % (n)	12(2696)	12(2263)	14(497)

The three datasets were found to give very similar summary statistics, datasets 1 and 2 were almost identical and dataset 3 was only slightly different. The largest disparities between the datasets were for disease severity and for withdrawing. The reason for missingness may be explained by disease severity and/or by withdrawing, or it could be explained by some unobserved variable. Within this chapter, datasets 2 and 3 are used. It was assumed that withdrawal was related to disease severity and that missingness can be explained by disease severity, therefore an assumption that the data are MAR was made. Therefore analyses contain, where possible, a variable for adjusting by disease severity.

6.1.2 Calculating QALYs

The EQ-5D can be used within a RCT to measure HRQoL and where used, the instrument is usually administered to each participant at pre-scheduled intervals throughout the duration of the trial. Within the TORCH trial, the EQ-5D was administered every six months. The EQ-5D scores over time can be combined in order to derive QALYs. QALYs are easily calculated for each participant using the method of Area Under the Curve (AUC) using data on utility over time. For example, in a RCT

lasting for three years, subject X was administered the EQ-5D seven times at pre scheduled intervals. To calculate QALYs for subject X, for each observation, the date of the assessment day, (d) was subtracted from the date of the previous assessment date, (d-1) and this was divided by 365.25 to get a value between 0 and 1 representing the proportion of one year between assessments. The average utility value between the two observations was calculated by adding the EQ-5D score from the current observation, (e) to the EQ-5D score from the previous observation (e-1) and dividing the value by two. The two differences are multiplied together to give a time weighted QALY score. The time weighted QALY scores are summed to get a QALY score out of 3 (for 3 years).

To further illustrate this, a worked example is shown in table 6.2. The duration of follow up between the first and second observations is calculated as 169-1 to give 168 and is divided by 365.25 to get 0.46 (A in table 6.2), representing the duration of time in years between the first and second observation. At baseline, the participant had an EQ-5D utility of 0.796 and in the second observation, had a score of perfect health, or 1. To estimate the average utility between the two study days, 0.796 is added to 1 to give 1.796. This is divided by two to give 0.898 (B). To weight this value by time, the utility is multiplied by the value previously derived for the duration of time between the first and second observation to give 0.413 (A*B). All the weighted utilities are summed together for the duration of the trial to give a within trial QALY score. In this example, patient X had a QALY score of 2.526 from a maximum of 3.

Table 6.2 QALY calculation for a hypothetical patient over a three year trial

Follow up day	EQ-5D utility	Diff in date/ 365.25 (A)	Difference in EQ-5D/2 (B)	(A*B)
1	0.796	/	/	
169	1	0.460	0.898	0.413
337	0.725	0.460	0.863	0.397
505	0.883	0.460	0.804	0.370
673	0.848	0.460	0.866	0.398
841	1	0.460	0.924	0.425
1093	0.516	0.690	0.758	0.523
			QALYs	**2.526**

To calculate mean QALYs for the TORCH trial, the process described above is carried out for each participant. Mean QALYs by treatment group can be compared to determine if a treatment is effective in terms of improving quality and quantity of life.

Missing Data and Deriving QALYs

Where the participant is followed up at every pre-scheduled time point for the duration of the RCT, complete information is available for this person. In real life however some values are missing and it is necessary to use imputation methods in order to create a full three year follow up history for each participant. The TORCH dataset had missing observations on EQ-5D utilities because some of the participants were not observed on the first study day, others were not observed on the last study visit (they were lost to follow up) and a large number had missing study days in between. The problem of missing study visits between the first and last observation was solved by finding the AUC between each observation and summing together as previously described. Imputation for missing first observations was achieved by using first observation carried backwards. Missing observations where participants withdrew from the trial or were lost to follow up had EQ-5D index values inputted using last observation carried forwards. Participants who died during the trial were given an EQ-5D value of 0 from the date of death for the remainder of the trial. Whilst this method of imputation may not be regarded quite as highly as other methods such as multiple imputation, the aim of the analyses on QALYs was to develop a multivariate regression equation to predict QALY scores, rather than the precise estimation and comparison of QALYs by treatment group per se. Where the ultimate aim is to accurately predict relative treatment gains, a variety of imputation methods may be attempted in order to improve the reliability of results.

6.1.3 OLS and GLM modelling

Within this section, different structures for developing regression equations that predict utility values are described. Because utility data (like cost data) are likely to be heavily skewed, the appropriateness of Ordinary Least Squares (OLS) in particular which relies upon the normal distribution, should be considered. A more flexible approach such as the Generalised Linear Model (GLM) may be more appropriate.

OLS Models

OLS is frequently used to model relationships between variables. The classic normal multiple regression model can be written as:

$$Y_i = \alpha + \sum_j \beta X_{ji} + \varepsilon_i$$

$$\varepsilon_i : NID\left(0, \sigma^2\right) \quad i = 1,...n$$

Where the error term is normally and independently distributed (NID). The OLS estimators, α and β, are chosen to get the 'best fit' in terms of minimising the sum of the squared residuals:

$$S\left(\alpha, \beta 1,..., \beta k\right) = \sum_i \left[y_i - \left(\alpha + \sum_j \beta_j X_{ji} \right) \right]^2$$

For OLS modelling, several important assumptions about the property of estimators are made.

- The model is linear in parameters, that is, the model can be written as described above with α and β as unknown parameters, and the error term is an unobservable random error.

- There is a random sample of observations from the population.

- The error term has a zero conditional mean, so that given any values of the explanatory variables the error has an expected value of zero.

- There is no perfect collinearity between any of the explanatory variables, in other words, none of the explanatory variables is constant and there is no exact linear relationship among the explanatory variables.

If these assumptions hold then the OLS estimators are unbiased estimators of the population parameters.(210)

Homoscedasticity, which refers to the variance of each error term being constant, conditional on the explanatory variables, is necessary for OLS to be efficient. If all these assumptions hold in the analysis then the OLS estimators are regarded as being the Best Linear Unbiased Estimators (BLUE). For inference, normality is also assumed. The condition of BLUE is important because where all these assumptions hold, for determining unbiased estimators, OLS will do as well, or better than any other method.

Any breaches of the assumptions will lead to the condition of BLUE failing. There are circumstances in which OLS may no longer be 'best' which refers to the smallest variance, but may still give unbiased estimators, for example including irrelevant variables in a regression model will have no effect on the unbiasdness of the estimators, but can cause the variances of the OLS estimators to increase because of multi collinearity. Sample size is important to achieve the 'best' condition because the larger the total variation in X, the smaller the variance of β will be. It is desirable therefore to gain as much sample variation as possible ,which is achieved through a large sample size.

If the assumption of homoscedasticity fails so that there is heteroskedasticity; for example where the error term varies, this will result in unbiased coefficients but inefficient estimates. Excluding a relevant variable can cause the estimators to be biased as the expected value of β (E(β)) will not equal β. The assumption of independence of the error term may be violated if there are multiple observations from the same subject.

Utility data are negatively skewed and because OLS operates assuming a normal distribution, predicted values from OLS may not be suitable for working with utility data, in this case it may be useful to explore other methods.

GLM Models

An alternative approach is GLM modelling. A feature of the GLM is that the OLS model is a special example of the GLM in which the family is Gaussian and the link, identity. Thus the OLS model can be identified within the GLM framework if it is found to fit the data well. GLM modelling has advantages over OLS because GLM directly models the mean and variance function on the original scale of EQ-5D. The GLM is based on the following equation:

$$g\{E(y)\} = x\beta, \quad y \sim F$$

With covariates x, g() is the link function and F is the distributional family.(211) The link function specifies the relationship between the mean (E(y|x)) and the linear specification of the covariates (xβ). Link functions include, but are not limited to: identity, log and square root links. The family used within the GLM should be one that fits the distribution of the data. Families include: Gaussian, Poisson and gamma. The GLM model with a Gaussian family and an identity link takes the form:

$$E(y) = x\beta, \quad y \sim \text{Normal}$$

Which is the equation for the OLS model.(211)

The family and a link function are selected, based on outcomes from statistical tests. Testing for distributional family is done using the Parks test,(212) using a GLM model with any link and family. Once the test is run, the family is identified based on the lowest Chi^2 value and is inputted into the model, at which point the Parks test is re-run to check whether the chosen family still gives the lowest Chi^2 value. Once the family is determined, tests are carried out to identify link using the Pearson correlation test,(213) the Pregibon link test,(214) and the modified Hosmer–Lemeshow test.(215) Where all three tests yield non-significant p-values, the link function is said to fit well.(99) The appropriate trade-off between the three links is currently unknown. A decision rule was used so that if any power function dominates in two or more tests then it is the chosen link function to take forward.

6.2 Utilities and QALYs from Trial Data

Utility data are an important component in a cost utility analysis, for reasons previously explained within Chapter 3. Therefore it is important to derive utility values from an appropriate patient population such as TORCH. Within this section EQ-5D utility scores were summarised from the data. OLS and GLM multivariate modelling were used to derive equations that predict EQ-5D utility and the predicted results were compared to the observed results of EQ-5D utility by disease severity. Using the AUC method, QALYs over three years were obtained for each subject and were summarised by treatment group. A multivariate analysis was developed in order to predict within trial QALY scores.

6.2.1 Methods

Subjects within TORCH were classified into COPD disease severity groupings according to the following spirometry classifications, which are consistent with the

NICE diagnostic criteria, given the inclusion criteria of the trial. COPD cases were identified where airflow obstruction was $FEV_1/FVC<0.7$ and $FEV_1<80\%$ predicted. For each COPD case, mild COPD was defined as $50\leq FEV_1< 80\%$ predicted, moderate COPD as $30\leq FEV_1< 50\%$ predicted and severe COPD as $FEV_1<30\%$ predicted. Because the spirometry inclusion criteria was $FEV_1\leq60\%$ predicted for the trial, in practice the mild group contained, in the most part, subjects with a lung function capacity of $50\leq FEV_1\leq 60\%$ predicted. The TORCH dataset 2 was used for all analyses because it was important that there was complete data on EQ-5D, which entered the equation as the dependent variable. Mean EQ-5D utility by disease severity was derived from the data.

A prediction equation for EQ-5D utility using patient specific variables was developed using OLS and GLM models. Utility decrement (ud) scores were used to conduct the analyses in order to bind the distribution to a positive scale. This was achieved by subtracting the index scores from 1 (ud=1-index). The advantage of using ud rather than utility is that standard methods (particularly GLM models) can be used to deal with skewed data by constraining values onto a positive scale. Multivariate analysis was conducted rather than univariate analysis because it enables subgroup analyses to be carried out. Exploratory analysis identified a number of candidates for inclusion within the model to predict EQ-5D utility and these were included within the regression models; COPD severity group, sex, ex-smoker, BMI and country. Because the utility scores of interest are those for the UK population, a dummy variable for UK was included within the list of covariates. Statistical tests previously described were used to identify family and link functions.

QALYs were derived from the TORCH dataset using the method of AUC (as described in section 6.1.3). QALYs were derived for each participant and mean QALYs by treatment group were compared. A multivariate analysis was developed to predict QALYs using the QALY scores estimated from the AUC method.

Once the best fitting model had been identified, the method of recycled predictions was used to estimate average EQ-5D values, by treatment group. Glick et al argue that on the transformed scale the effect of the treatment group is estimated holding all else equal but retransformation in order to estimate EQ-5D can reintroduce covariate imbalances and so the method of recycled predictions gets around this problem by creating an identical covariate structure for each treatment group, by coding everyone as if they were in the first treatment group (1) and predicting QALYs for each subject, then coding everyone in the other treatment groups in turn (2, 3 or 4) and predicting QALYs for each person. The difference in the arithmetic mean between the treatments, predicts QALYs holding all else equal during model estimation and retransformation.(99)

The best fitting GLM equation and the method of recycled predictions were replicated 1000 times using bootstrapping to obtain 1000 average QALY scores in each of the four treatment groups. These average scores were ordered from lowest to highest where the 26th and the 975th values represent the 95% CIs. Non parametric tests such as bootstrapping of arithmetic means may be more suitable for tests of significance,(99) than standard parametric tests such as the t test which assume that the data are normally distributed. QALY data are likely to have a skewed distribution which potentially invalidates the assumptions underlying the use of parametric tests.

6.2.2 Results

EQ-5D utility values from dataset 2 are illustrated in figure 6.3 which shows that the EQ-5D utility values are not normally distributed and are negatively skewed. The distribution is tri-modal, with almost 3800 of the observations taking the value 1. Only 8% of the responses fall below 0.50. The distribution of utility scores, particularly the apparent gaps, represents an artefact of the scoring algorithm; it is well known that the scoring algorithm does not generate values above 0.88 but below 1.
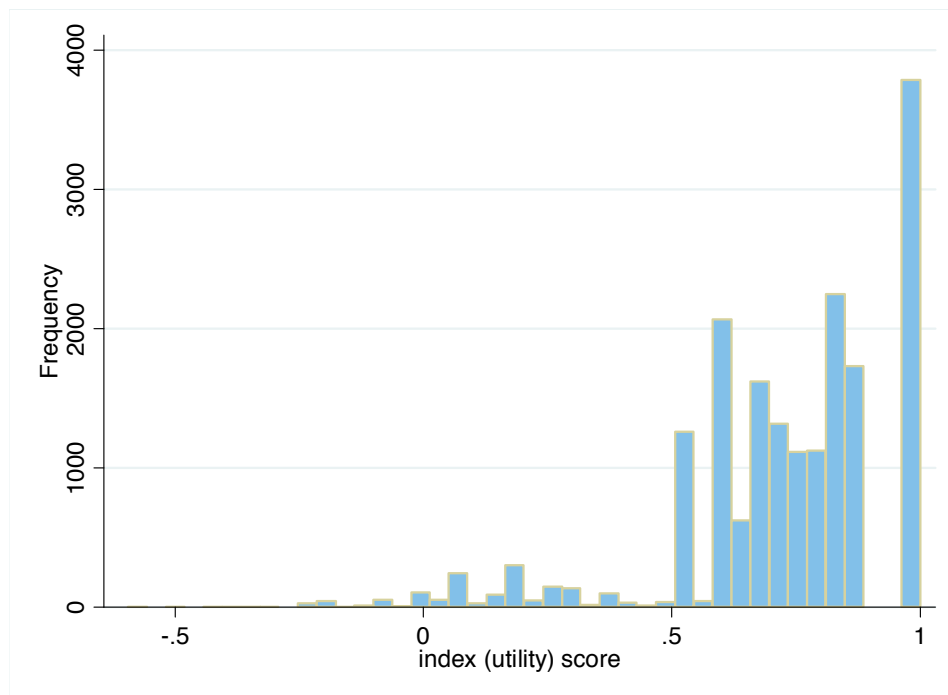
Figure 6.3 Distribution of EQ-5D responses

The study population comprised of 3854 subjects and 18 503 observations of whom, as shown in table 6.3, 2735 (71%) were male and 1119 (29%) were female. Mean age at the first observation was 64.8 (62.7) for men (women). Of the men and women, 39% and 42% respectively withdrew from the trial and 14% of the men died during the three years of the trial compared to 10% of the women.

Table 6.3 Summary statistics, dataset 2
n=3854, n observations= 18 503

	Men	Women
Number in study, n (%)	2735 (71)	1119 (29)
Age, mean (sd)	64.8 (8.3)	62.7 (8.7)
BMI, mean (sd)	26.2 (4.9)	25.5 (6.0)
Withdrew, n(%)	1072 (39)	474 (42)
Dead, n(%)	391 (14)	114 (10)

EQ-5D utility by disease severity was 0.773 (SD=0.22) for mild COPD, 0.713 (0.23) for moderate COPD patients and 0.650 (0.23) for severe COPD patients.

Utilities from the Prediction Equation

An OLS and GLM were developed during exploratory analyses. The GLM model identified that a Poisson family (Chi^2=0.96, p=0.33) with an identity link was appropriate for the data, rather than the Gaussian family (OLS) and so the GLM model was taken forward and is described here. The GLM model had ud as the dependent variable with: BMI, UK, ex-smoker and COPD disease severity group as independent variables. Based on three statistical tests, the identity link was found to perform relatively well (Pearson correlation test=0.87, Pregibon link test=0.44 and Modified Hosmer and Lemeshow=0.61). Table 6.4 contains the coefficients, standard errors and p-values for the prediction equation for EQ-5D utility with Poisson family and identity link.

Table 6.4 Prediction equation for EQ-5D ud
with independent variables, coefficients, SEs and p values from a GLM regression with Poisson family and identity link

Variable	Coefficient	SE	P
BMI (range12-57)	0.0016	0.0007	0.03
UK (0=no 1=yes)	0.0907	0.0264	0.00
Ex-smoker(0=no 1=yes)	-0.0292	0.0080	0.00
Moderate (0=no 1=yes)	0.0613	0.0083	0.00
Severe (0=no 1=yes)	0.1268	0.0139	0.00
Cons	0.1974	0.0204	0.00

Gender was originally included within the list of coefficients and was removed due to statistical insignificance. The variable on ex-smoker shows that ex-smokers had a small but statistically significant, larger utility compared to smokers. COPD moderate and severe disease severity was statistically significant compared to mild disease.

Because the regression was developed on the ud scale, transformation of the resulting ud values to EQ-5D utility was conducted (1-ud), in order to compare against observed utility, and for reporting purposes. Predicted EQ-5D utility by disease severity was calculated using the prediction equation described above and is shown in table 6.5 alongside mean EQ-5D utility by disease severity as recorded in the trial.

Table 6.5 Observed and predicted EQ-5D utility by COPD disease severity

Disease severity	Predicted	Observed
Mild	0.773	0.773
Moderate	0.713	0.713
Severe	0.649	0.650

As can be seen in table 6.5, utility was highest for people with mild disease and lowest for those with severe disease, suggesting HRQoL deteriorates as the disease progresses. The prediction equation was found to perform well in predicting EQ-5D utility by disease severity when compared to the observed results. The predictions matched the observed scores, accurate to 2 decimal places.

QALYs from Trial Data

QALY scores were derived for each subject using the AUC method. QALYs by treatment group were: 1.988, 1.980, 2.019 and 2.087 for treatment groups 1, 2, 3 and 4 respectively. The histogram of QALYs, reproduced in figure 6.4, shows that the QALY data were left skewed with over 80% of scores between 1.5 and 3. The distribution was characterised by a long left tail and for a minority of subjects, QALY scores were less than 0
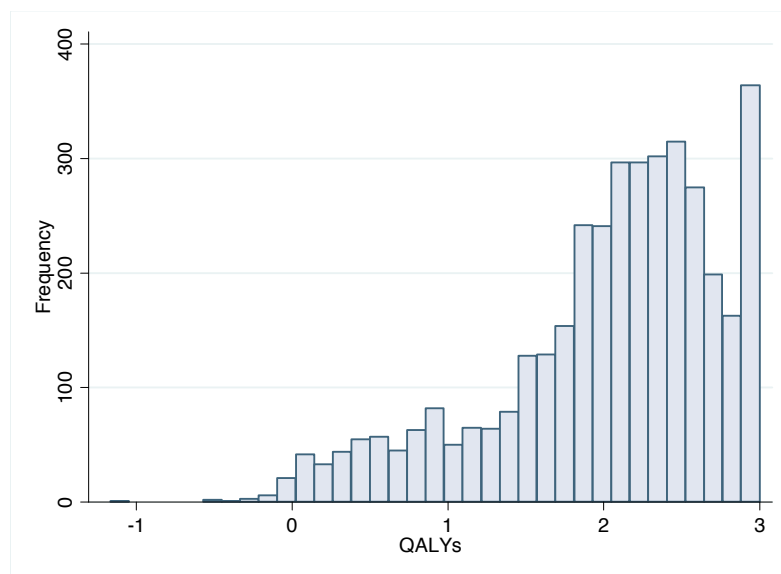
Figure 6.4 Distribution of QALYs

QALYs from Modelling

GLM regression models were developed to predict QALY decrement ((3-QALY) and were retransformed to QALYs for reporting). Explanatory variables of: treatment group; COPD disease severity; sex; BMI; UK and ex-smoker were used.

The Poisson family was identified using the Parks test with a Chi^2 of 3.37 (p=0.07). The GLM on ud with Poisson family was run with a number of different link functions and the outputs of the resulting statistical tests are reproduced in table 6.6. Table 6.6 shows that for the identity link (where power=1), the p-value from the Pregibon link test was 0.3214 and 0.8887 for the modified Hosmer and Lemeshow test. It is clearly seen in the table that these numbers are the largest within each test, and based on the *a priori* decision rule, the identity link is said to fit the data well.

Table 6.6 P-values from statistical tests

Results from the Pearson Correlation, Pregibon link and Modified Hosmer and Lemeshow tests applied to a GLM on QALY decrement with a Poisson family and different link functions

Power	Pearson Correlation test	Pregibon Link test	Modified Hosmer and Lemeshow test
0	0.8837	0.1716	0.1398
0.25	0.8653	0.1953	0.1555
0.50	0.8526	0.2278	0.2505
0.65	0.8478	0.2518	0.4825
0.75	0.8458	0.2698	0.4426
1	0.8447	0.3214	0.8887
-1	1.0000	0.1500	0.6853
-2	0.8707	0.2407	0.0724

The GLM model for QALY decrement with Poisson family and identity link is shown in table 6.7 below. BMI was not found to be statistically significant and so was removed from the analysis. Disease severity, sex, UK and ex-smoker were all found to be statistically significant in the prediction of QALY decrement. Treatments 2 and 3 were not found to be statistically significant, but the p value for treatment 4 suggested some QALY gain arising from this treatment.

Table 6.7 Prediction equation for QALY decrement
with independent variables, coefficients, SEs and p values from a GLM regression with
Poisson family and identity link

Variable	Coefficient	SE	P
Tmt 2 (0=no 1=yes)	-0.0195	0.0500	0.695
Tmt 3 (0=no 1=yes)	-0.0455	0.0495	0.358
Tmt 4 (0=no 1=yes)	-0.0924	0.0487	0.058
Sex (0=women 1=men)	-0.0889	0.0390	0.023
UK (0=no 1=yes)	0.2670	0.1242	0.032
Ex-smoker(0=no 1=yes)	-0.0759	0.0353	0.032
Moderate (0=no 1=yes)	0.1884	0.0370	0.000
Severe (0=no 1=yes)	0.4199	0.0598	0.000
Cons	0.9702	0.0519	0.000

Using the resulting GLM model as described, with the identity link and Poisson family together with the method of recycled predictions, gave predicted QALY scores: 1.981, 1.991, 2.021 and 2.081 for treatments 1,2,3 and 4 respectively as shown in table 6.8 below. The predicted means were similar to those from the dataset. Bootstrapping 1000 times with replacement gave 95% CIs around the predicted mean QALYs and the results from this analysis are presented in table 6.8. Analyses from the bootstrapped results suggested that treatment 4 has a higher mean QALY than the other 3 treatments (p<0.001 vs. treatment 1, p=0.008 vs. treatment 2, p=0.070 vs. treatment 3), and there is no evidence of differences between treatments 1 to 3 (all p>0.2).

Table 6.8 Predicted QALYs and bootstrapped 95% CIs

Treatment	Observed	Predicted	95% CI
1	1.988	1.981	1.935-2.030
2	1.980	1.991	1.946-2.036
3	2.019	2.021	1.978-2.063
4	2.087	2.081	2.036-2.123

6.2.3 Discussion

Previous studies have used a range of utility scores within their COPD models. As discussed in Chapter 3, all of the COPD models currently published are structured on a Markov framework and within the Markov model; many authors have presented the disease in terms of COPD disease severity based on FEV_1 % predicted. Table 6.9

illustrates the array of utility values used within these models, alongside the results from the current study, which are consistent with those previously estimated.

Table 6.9 Utilities used in economic models, by COPD disease severity

	>80% predicted	Mild 50-80	Moderate 30-50	Severe <30
Rutten van Molken (2007)	0.81	0.81	0.76	0.66
Borg et al (2004(138)) (sd)	0.8971 (0.1117)	0.7551 (0.2747)	0.7481 (0.2991)	0.5493 (0.3129)
Spencer (2005)(144) (se)	0.81(0.02)	0.81(0.02)	0.72(0.03)	0.67(0.05)
Sin (2004)(141)	1.00	1.00	0.92	0.84
Observed		0.773	0.713	0.649
Predicted		0.773	0.713	0.650

QALYs derived from the TORCH trial using means and predicted means, by treatment group were found to be similar. Deriving QALYs directly from the dataset may have limited usage because the resulting QALYs cannot take account of, amongst other things, age, sex or disease severity. The use of regression models to predict QALYs allows these variables to be explicitly accounted for.

A prediction equation for QALYs using a GLM model with Poisson family and the identity link was found to fit the data well. Non-parametric tests of statistical significance were preferred and a bootstrap was carried out to determine the 95% CIs around the predicted mean scores.

The methods described above to estimate and predict utilities can be applied in exactly the same way to cost data. Extensions to this study on utilities would be to use the same methods on costs, which would lead to costs and QALY estimates being available for each treatment group. The differences between the treatment groups would be calculated to derive ICERs. Uncertainty around the ICER would be established using the method of bootstrapping as previously described.

6.3 Utilities and QALYS Indirectly from Trial Data

For reasons previously described in Chapter 3, for cost utility analyses, data on a utility based measure such as the EQ-5D is necessary in order to assess the cost effectiveness of an intervention. However, clinical studies assessing the effectiveness of new drugs or treatments for COPD often only employ the SGRQ (section 2.3.1), a disease specific HRQoL instrument, in order to examine changes in HRQoL: of the major COPD RCTs, only TORCH has collected information on EQ-5D. Therefore, in order to use the HRQoL data gathered within trials where EQ-5D was not collected, it is useful to employ a mapping algorithm in order to predict utility from the SGRQ. NICE have recently approved of mapping and have stated:

> "Where EQ-5D data are not available, other methods should be used to estimate EQ-5D utility data. Consideration will be given to mapping EQ-5D utility data from other HRQL measures included in the relevant clinical trial(s) where an adequate mapping function can be demonstrated and validated." (87)

Responses to the SGRQ questionnaire and the EQ-5D questionnaire were routinely collected within the three years of the TORCH trial and these measures have previously been described at the start of this chapter.

6.3.1 Literature Review

This section reviews the literature on both existing SGRQ to EQ-5D mappings and on other mappings more generally.

Two mapping algorithms already exist that map the SGRQ to utilities, and are available in abstract form only.(216;217) The first by Meguro et al derives utilities directly from a COPD patient population using the TTO method (this method is described in Chapter 3). The states were created using a combination of nine of the questions/statements

within the SGRQ in order to simulate different levels of disease severity for a small number of sample states.(216) A multiple regression model was developed to impute values for the other possible states. A few issues exist about the study results, including: that of the patients recruited, fewer than 50% provided reliable data (69 out of 150); an adjustment factor is required to normalise the resulting utility estimate; and the algorithm is currently restrictive in that it only applies to 65 year old males.

The second mapping by Vardeva et al models the relationship between the SGRQ and EQ-5D utilities using a regression model.(217) OLS regression was used for the mapping. Various alternatives were also investigated, including Tobit regression and logarithm transformation of the dependent variable. A validation sample was used based upon health status at the end of the study. Predictive ability by COPD severity was estimated. R^2 and Root Mean Squared Error (RMSE) were used to assess the model with regards to fit and predictive ability. The preferred model (OLS no log transformation) combined domain scores and individual item scores, resulting in some element of double-counting. The model has been shown to be problematic (GSK statisticians, Stockley Park) when applied to a recent dataset, particularly at the upper and lower extremes of self-reported HRQoL. There is therefore room for a soundly developed and thoroughly validated mapping algorithm from the SGRQ to the EQ-5D.

Structured Review

There has been an increase in the literature on mapping within recent years. A literature search in OVID was carried out using the terms "utility" "mapping" "health" since 1996 until January 2008 which identified 50 papers. Of these, 24 abstracts were read and of these 10 were found to be relevant. Reference lists were also searched in order to identify potentially relevant papers. Five additional papers were identified via this search. The inclusion criterion was 'HRQoL questionnaires mapped to utility' and the exclusion criterion was 'mapping based upon VAS or SG studies'.

In total, 14 papers were identified and table 6.10 below provides a summary of each of these papers, including for each study: the HRQoL measures used to carry out the mapping; the statistical model(s) used; how missing data was dealt with; the various coefficients that were included within the model; details of the validation sample used; how goodness of fit and predictive ability compared; the final model; and the method chosen for each study.

Table 6.10 illustrates that whilst there is much variability in the HRQoL measures used to conduct the mapping, the methods used are similar. Without exception, all of those papers published before 2005 solely employed OLS in order to develop regression equations. A number of other methods have been employed more recently including Generalised least squares, Random Effects, Generalised Estimating Equations, and Tobit models.

Where stated within the studies, only complete case data were used on which to run the analyses, thus removing all missing data. Different combinations of explanatory variables were investigated, particularly around options with total, domain or individual item scores. Some analyses included squared coefficients and/or interaction terms. The majority included demographic terms but these were found to differ between the studies (most used as a minimum age and gender).

The validation samples used were usually internal following a random split within the dataset: either 1/2 fitting and 1/2 validation or 2/3 fitting and 1/3 validation. Other studies used an external clinical trial, or bootstrapping from the original dataset. R^2 was the usual method used to measure goodness of fit; ranging from 0.34 to 0.69. RMSE was most frequently used for testing predictive ability and ranged from 0.09 to 0.21.

Brazier et al Systematic Review

A systematic literature review by Brazier et al,(218;219) identified 28 mapping studies, representing 119 models, including studies from so called 'grey' literature'. The Brazier et al review had similar findings to those previously outlined.

Brazier at al found that most of the studies employed OLS, though some used more complex techniques including GLM with random effects, adjusted least squares, Tobit, censored least absolute deviation and non-linear models.

The most popular measure to map to was the EQ-5D Index (16 out of 28 studies). The sample size used within the studies varied from 98 to 23 547. Regression models were usually additive, with total/dimension/item scores as independent variables (some including the item scores as categorically as well as continuous). Of the 119 models, 34 included personal characteristics such as age, sex and race, 33 included interaction terms and 19 used squared transformations of the dependent variable. Increasing the complexity of the model was found to often achieve only negligible improvements in predictive ability.

Table 6.10 Summary of the mapping papers

First author, ref	Dataset, n	HRQoL measure	Model	missing data	Coefficients	validation	validation stat	Mapping algorithm
Wu (220)	Obs study, 270	FACT-P and the EORTC QLQ-C30 to EQ-5D Index	1) OLS 2) OLS for EQ-5D scores <1 3) median regression	Baseline measures	1) FACT-P score, EORTC QLQ-C30 component scores, demographics 2) FACT-P score, EORTC QLQ-C30 component scores, demographics and interaction terms 3) FACT-P component scores, EORTC QLQ-C30 component scores, demographics 4) FACT-P component scores, EORTC QLQ-C30 component scores, demographics and interaction terms.	Split into 10: 90% used to fit, 10% validation. Each group of 10% used successively to test	R2 and SSE Absolute deviation	OLS with FACT-P component scores, EORTC QLQ-C30 component scores and demographics (R2=0.58)
Yang, (221)	MOS study, 2825 and 2 RCTs 199, 124	MOS sleep scale to SF-6D	non parametric and OLS	/	1) non parametric regression 2) SLP9 3) SLP9, SLP9 squared 4) SLP9, SLP9 squared, gender, age 5) slp9, slp9 squared, disease status 6) SLP9, SLP9 squared, gender, age, disese status	random split	R2, observed vs predicted SF-6D	linear regression: SLP, SLP squared (R2=0.34)
Buxton (222)	Two RCTs, 905	IBDQ, CDAI, EQ-5D Index and SF-6D Index	GLS	not stated	1) IBDQ to SF-6D 2) IBDQ to EQ-5D 3) CDAI to SF-36 4) CDAI to EG-5D plus for the IBDQ mappings, squared IBDQ and demographics	external RCT	R2, MA% Error. Comparisons of means	IBDQ with the inclusion of IBDQ squared to SF-6D (R2=0.69). And IBDQ to EQ-5D (R2=0.45). Both with no demographics
Grootendorst (223)	RCT 255	WOMAC and HUI-3	OLS RE	Complete cases.	1) WOMAC individual items (cat) 2) WOMAC domain scores 3) WOMAC Domain scores plus interaction terms 4)WOMAC total + total^2 + demographics for each	Random split: 2/3 training, 1/3 validation	MAE, RMSE, ICC	WOMAC domain scores + interactionn + demographics. (R2= 0.39) (RMSE= 0.21) (MAE=0.16)
Bansback (224)	Survey, 319 CAN, 151 UK	HAQ, SF-36 and EQ-5D Index	GEE	/	Each to EQ-5D and SF-6D: 1) Total HAQ 2) HAQ domain scores (cont) 3) HAQ individual items (cont) 4) HAQ domain scores (ca) 5) HAQ individual items (cat) + patient demographics	Random split. 2/3 training, 1/3 validation.	RMSE, R2	SF-6D: Models 2 and 4 performed well, no demographics (RMSE=0.089) EQ-5D: All with marginal R2>0.5
Brennan (225)	Survey, 375	OHIP to EQ-5D Index	Tobit	/	individual items (categorical and continuous) plus age and sex, nonlinear (age2) and inclusion of interaction terms	Random split: 2/3 fitting 1/3 validation	forecast errors.	continuous model + age and sex (R2=0.62)
Sullivan (226)	MEPS 2000 14 286 & 2002, 23 647	SF-12 to EQ-5D Index	CLAD, Tobit, OLS	complete cases	PCS-12, MCS-12, PCS-12*2 MCS-12*2 + demographics (age, sex, ethnicity, income, education, no of chronic conditons)	MEPS 2002 for fitting, MEPS 2000 for validation	T PE, mean PE	CLAD PCS-12, MCS-12 + demographics
Longworth (227)	Survey, 510	CCS, Breathlessness grade to EQ-5D Index	OLS	complete cases	CCS score, breathlessness score, gender, age, treatment centre, type/number of revascularisation experiences, no current medications and interaction terms is statistically significant	/	R2	OLS with CCS(cat) Breathlessness score(cat), No of drugs
Franks (228)	Survey, 12 998	SF-12 to EQ-5D Index (UK pref)	OLS	complete cases	SF-12 domain scores, with and without inclusion of squared domain scores and interaction terms, up to 4th degree polynomials and using linear splines of each domain (dividing each into 20 equally saced intervals and demographics	Bootstrap and external validation on a dataset of 240	R2.	A second degree polynomial without demographics (R2=0.62).

Brazier (229)	survey, RCTs, surgery cand, 1972	IWQOL-Lite to SF-6D	OLS	complet e cases	1) IWQOL-Lite total score 2) IWQOL-Lite domain scores 3) IWQOL-Lite individual scores (cont) 4) IWQOL-Lite individual scores (cat) 5) the best from above + age, gender, bmi	Random split: 50:50	RMSE and R2	Model 4 plus age and bmi (R2= 0.521) (RMSE=0.0946)
Lawren ce (230)	MEPS (2000), 14 580	SF-12 to EQ-5D	OLS	complet e case	1) PCS-12, MCS-12, 2) PCS-12, MCS-12, PCS-12*MCS-12 3) PCS-12, MCS-12, PCS-12*MCS-12, PCS-12^2, MCS-12^2, PCS-12^3	Random split: 50:50	R2	PCS-12, MCS-12 (R2= 0.612)
Sengup ta (231)	Survey, 6923	SF-12 to the HUI3 and VAS	OLS	complet e case	1) SF-12 individual items (cat) 2 SF-12 individual items (cont) 3) MCS-12, PCS-12, MCS-12^2, PCS-12^2, PCS-12*MCS-12 + age and gender	Random split: 50:50	R2 MAD, mean deviation[2]	1) SF-12 individual items + demographics (cont) (HUI3 R2= 0.47) (VAS R2= 0.55)
Franks (232)	Survey, 240	SF-12, HUI3 to EQ-5D	OLS	complet e case	PCS-12 and MCS-12, squared and interaction terms, demographics and clinical variables	Bootstrapping with 1000 replications	R2	EQ-5D: PCS-12, MCS-12, PCS-12^2, MCS-12^2, PCS-12*MCS-12 (R2=0.59) HUI: PCS-12, MCS-12, MCS-12^2 (R2 = 0.51)
Nichol (233)	Survey, 6921	SF-36 to HUI2	OLS	complet e case	SF-36 domain scores, interaction terms, age and gender	/	R2	SF-36 domain scores + age (R2=0.51)

MOS = Medical outcomes study, FACT-P = Functional Assessment of Cancer Therapy - Prostate, EORTC-C30 = European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire, AQoL = Assessment of quality of life, IBDQ = Inflammatory Bowel Disease Questionnaire, CDAI = Crohn's Disease Activity Index, WOMAC=Western Ontario and McMaster University Osteoarthristis Index, FACT-G = Functional Assessment of Cancer Therapy - General, HAQ=Health Assessment Questionnaire, OHIP = Oral Health Impact Profile, MEPS= Medical Expenditure Panel Survey, CCS = Canadian Cardiovascular Society score

R^2 and/or adjusted R^2 were usually reported. For models mapping from a generic to a generic measure (such as the SF-12 to EQ-5D), the studies achieved an R^2/adjusted R^2 of greater than 0.5. However, for those studies mapping from a condition specific to a generic measure, R^2 was lower, with a range of 0.17-0.51. Brazier et al question the use of the R^2 statistic by stating that:

> "Measures of explanatory power, for example R^2 and adjusted R^2, are not a useful basis for assessing model performance of mapping functions as they focus upon how well the model explains the dataset it was estimated on... A better method for assessing mapping functions is to examine the difference between predicted and observed values by calculating Mean Error (ME), Mean Absolute Error (MAE) or Root Mean Squared Error (RMSE). These can provide an indication of the size of the prediction errors."(234)

Within the studies, ME ranged from 0.0007 to 0.042, MAE ranged from 0.0011 to 0.19 and RMSE ranged from 0.084 to 0.2. Overall, the level of error was found to be far greater when predicting more severe health states. Scatter plots in several of the studies found that there is a tendency for the predicted EQ-5D values to over-predict observed values at the lower end and under-predict at the upper end of the EQ-5D utility scale.

It has been seen that whilst previous studies have mapped disease specific measures to the EQ-5D, the extent to which these algorithms are developed and tested varies and is often limited. Most studies have employed and stop at using OLS to derive a prediction equation for the mapping. Where validation has been attempted, the majority of studies used a random split of the data on which to test their resulting algorithm. Here it is argued that such a split is inappropriate, with preference made for a non-random split of the data. Within this section, a rigorous approach to the development and validation of a prediction model using a dataset in which both the EQ-5D and the SGRQ were routinely collected in COPD patients was carried out.

6.3.2 Methods

The methods used to develop and validate a mapping algorithm from the SGRQ to the EQ-5D are described below.

The TORCH dataset 3 (as previously described) was used to conduct all the analyses within this dataset, complete data were available on EQ-5D utility and on all SGRQ responses. It was important to only use the complete cases because it was the relationship between the SGRQ variable(s) and the EQ 5D variable that was of interest and imputing missing values for these variables potentially risks blurring the relationship between them. Within the TORCH study, SGRQ and EQ-5D were collected at baseline and every 24 weeks over three years. Data were split non-randomly into a 'fitting' sample (all non-USA subjects, 67% of study sample) and a 'validation' sample (all USA subjects, 33% of study sample). Splitting the dataset in this way ensures a more robust validation than would be provided using a random split.

During earlier exploratory analyses, the dataset was randomly split into two to give a fitting dataset and a validation dataset. Two thirds of the data went to the fitting dataset (6987 observations on 2098 participants) and the remainder to the validation dataset (3727 observations on 1082 participants). However, following a conference discussion as to the usefulness of this approach, this method was reconsidered. If the data were randomly split so that there was a fitting and a validation dataset, it is expected that these groups would be almost identical in terms of population characteristics. Therefore a prediction equation developed in one of these groups should predict well in the other group, precisely because the two groups are similar. Thus when comparing the actual to predicted values in the validation dataset, the model should predict well. Therefore it was decided to split the data non-randomly. In deciding how to split the data, 'country' came out as a reasonable choice

as the country variable was found to be statistically insignificant within exploratory analyses of the dataset. The number of USA individuals within the trial equalled about 1/3 of the complete case dataset. Therefore the data were split into two, with the data from all of the countries, apart from the USA, going in the fitting dataset and just those USA participants' data making up the validation dataset. The prediction dataset was used to conduct all the analyses. The validation dataset was used to test the predictive ability of the mapping algorithm.

Basic summary statistics were derived from the dataset. COPD disease severity states were used for comparing the prediction equations, and were classified in terms of degree of airway restriction as previously described in section 6.2.1. As before, EQ-5D utility values were transformed onto a ud scale so that commonly used distributions could be fit to the (right) skewed data.

Model Development

Models were developed with increasing complexity, in terms of explanatory variables. As previously described, the SGRQ contains three levels of questions: item scores corresponding to individual questions, domain scores (symptoms, activity and impact) and a total score. Together with demographic variables of gender, age and BMI, each of these levels in turn were offered to the model, to give a general algorithm for the prediction of EQ-5D ud of:

$$\text{EQ-5D ud} = \alpha + \sum \beta(\text{SGRQ(total/domain/questions)}) + \text{age} + \text{BMI} + \text{gender}$$

All models were developed in the fitting sample. Fractional polynomial analysis was used to identify non-linear terms for SGRQ within the above model, for equations where either total SGRQ score or the SGRQ domain scores were the independent variable(s). As the

starting point, OLS was used to develop the prediction algorithms using the algorithm described above. Significant terms that arose when employing fractional polynomials were used to develop additional prediction equations. For example fractional polynomials can identify if total$^{0.5}$, total^2 and total^3 are statistically significant in the model that predicts EQ-5D index, with SGRQ total as the primary independent variable. Backwards stepwise selection was used on all models to identify statistically significant variables.

The following algorithms were developed:

EQ-5D ud =SGRQ total score + age + gender + BMI

EQ-5D ud = impact + activity + symptoms + age + gender + BMI

EQ-5D ud = SGRQ individual item scores (continuous variables) + age + gender + BMI

EQ-5D ud = SGRQ individual item scores (categorical variables) + age + gender + BMI

Models were developed using in turn: OLS, GLM and two part models. The GLM model used the same methods as previously set out in 6.1.4 in order to identify family and link. The two part model,(235) provided an alternative approach to modelling skewed data. Of the EQ-5D ud scores, a large proportion of values were zero, corresponding to a full health status. The first part of the two part model uses a logistic regression to determine the probability of a non-zero value. The second part uses a GLM regression to predict EQ-5D from each SGRQ value. A GLM model for the second part of the two part model was found in exactly the same way as previously described: using fractional polynomials, backwards stepwise selection and then the modified Parks test and link tests to identify family and link.(99) However, because it is the non zero values that are of interest, only data with positive values for EQ-5D ud were used to develop the model. The predicted EQ-5D ud score from the two part model was found by multiplying the predicted values from the first part of the model by the predicted values from the second part.

Validation

The OLS, GLM and two part prediction models developed within the previous section were used to predict EQ-5D index scores within the validation sample. The maximum and minimum predicted EQ-5D scores values are presented to illustrate the predictive ability of the resulting models. A summary measure of model fit; RMSE was calculated,(236) using each mapping equation in turn to predict EQ-5D values within the validation sample. Fit was decided based on lowest RMSE. Predicted scores from the OLS, GLM and two part procedure were compared to observed EQ-5D score by disease severity.

Recommended Model

The best fitting model from the validation exercise was identified, based on the lowest RMSE. In cases where RMSEs were very similar, the decision criterion was to select the simplest model. The selected model structure with associated explanatory variables was refit onto as much of the data as possible in order to determine the best fitting model. For example, if the model containing SGRQ total score was selected, then only variables with missing total and missing EQ-5D utility would be excluded from the dataset (ie dataset 2 would be used). The model was fit using all of the resulting dataset. The resulting prediction equation is referred to as the recommended model.

QALYs

In order to further validate the prediction equation within the context in which it could ultimately be used, ie the generation of QALYs within a cost utility analysis, QALY scores by treatment group were calculated. The recommended model was used to predict EQ-5D utility within the dataset. QALYs were estimated for each subject in the whole study population, by treatment arm: 1, 2, 3 or 4. For both the observed and the predicted values,

QALYs were calculated using the AUC method using EQ-5D scores over a three year history. Where missing values occurred, the last observation was carried forward and/or the first observation carried backwards in order to create a full follow up history of three years for each subject as described earlier in 6.1.3.

6.3.3 Results

As previously described, the TORCH study consisted of 6112 participants (4236 with EQ-5D administered) of whom 3855 (Obs=18 505) had data on both the EQ-5D index score and the SGRQ total score. For this analysis, dataset 3 was used (described in section 6.1.2) which contained observations with complete responses on any of the SGRQ individual responses, the SGRQ domain scores, the SGRQ total score and/or in the EQ-5D index score to give a study sample containing 3640 subjects (2586 men, 1054 women), and 14 612 observations, and is summarised in table 6.11. To minimise loss due to missing data, as previously explained, four of the SGRQ questions (6, 8, 10 and 50) were removed because of high proportions of missing data. The fitting dataset consisted of 9724 observations and the validation dataset, 4888.

Table 6.11 Summary statistics, TORCH study

Number in study	3640
Observations, n	14 612
Validation sample, n (%)	1278 (33)
Men	2586 (71)
Age, mean (sd)	64.7 (8.4)
3yr survival, n (%)	3189 (88)
Withdrawal, n (%)	1366 (38)
Mild COPD, n (%)	1315 (36)
Moderate COPD, n (%)	1828 (50)
Severe COPD, n (%)	497 (14)
EQ-5D utility, mean (SD)	0.73 (0.23)
SGRQ Total score, mean (SD)	46.5 (17.9)
SGRQ Impact score, mean (SD)	34.3 (19.5)
SGRQ Activity score, mean (SD)	61.5 (20.6)
SGRQ Symptom score, mean (SD)	58.3 (21.3)

SGRQ= St George's respiratory questionnaire

The mean age of subjects was 65 years as can be seen in table 6.11. After three years of follow-up, 88% of the study population were still alive and 38% had withdrawn from the study. Most of the participants had either mild (36%) or moderate (50%) disease with a smaller proportion of the study in the severe group (14%). Mean (SD) EQ-5D utility score was 0.73 (0.23). SGRQ Impact scores were lower than the activity and symptom scores for both men and women, suggesting that the impact domain was the least affected compared to the domains of activities and symptoms.

The scatter graph of EQ-5D responses against SGRQ total score in figure 6.5 shows that there is some correlation between the two measures and the overlaid histogram illustrates the tri-modal nature of the responses to the EQ-5D: 3068 observations were at 1 (full health), 10 372 were between 0.5-0.88 and 1172 observations were lower than 0.5.

Figure 6.5. Scatter of SGRQ total and EQ-5D utility together with the distribution of EQ-5D utility responses
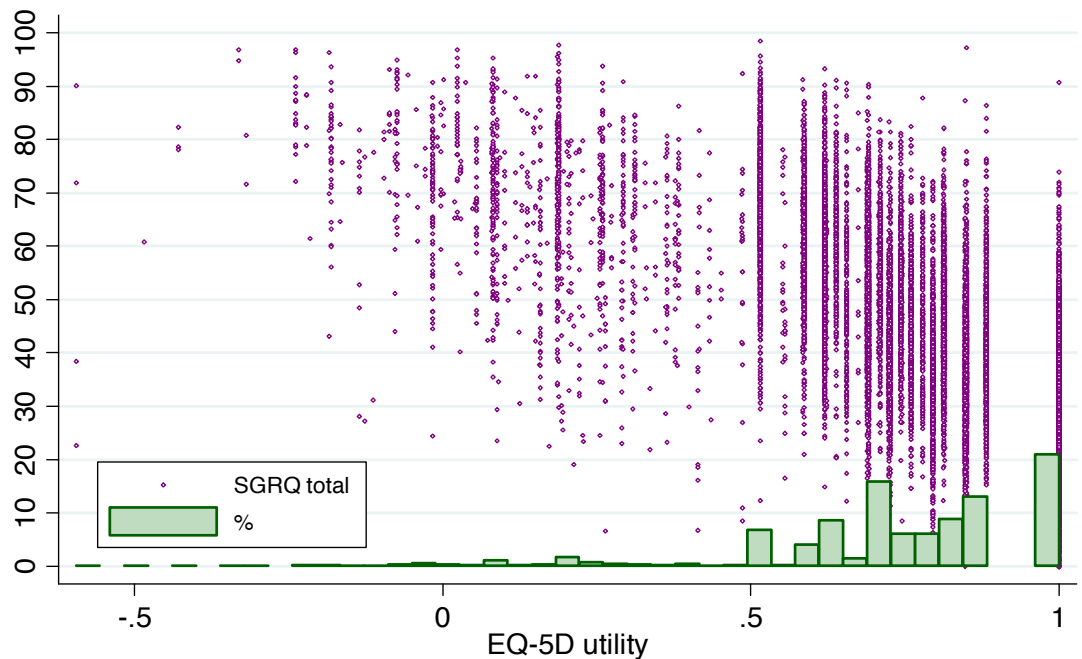
Model Development

Models that mapped to the EQ-5D ud scale from the SGRQ were developed using increasingly more explanatory variables leading to a more complex structure. Fractional polynomial analysis found SGRQ total^2 was statistically significant in the model with SGRQ total as the primary independent variable and within the regression with SGRQ domains as the independent variables, symptoms^2 was found to be statistically significant. Using backwards stepwise selection, gender was the only non SGRQ variable that was found to be statistically significant ($p<0.05$).

The following six models were developed in the fitting dataset:

1. EQ-5D ud =SGRQ total score + gender

2. EQ-5D ud =SGRQ total score + SGRQ total score^2 + gender

3. EQ-5D ud = impact + activity + symptoms + gender

4. EQ-5D ud = impact + activity + symptoms + symptoms^2 + gender

5. EQ-5D ud = SGRQ individual item scores (continuous variables) + gender

6. EQ-5D ud = SGRQ individual item scores (categorical variables) + gender

OLS was used for each of the six models in order to predict EQ-5D ud using SGRQ. GLM modelling was used for the six models and included additional steps of selecting the chosen family based on the results from the modified Parks test and the link was identified using the link function tests: the Pearson correlation test, the Pregibon link test and the modified Hosmer–Lemeshow test. Based on these results, the Poisson family with the square root link were identified as appropriate for modelling each of the six models.

In order to build up a two part model, for each of the six models described above, a corresponding logistic regression and GLM were identified. The logistic regression used statistically significant explanatory variables (determined by backward stepwise selection) to predict the probability of non-zero EQ-5D ud values. The GLM was fit onto positive EQ-5D ud values. Results from the modified Parks test and the three link tests found that for three of the six models (1,3 and 4), the Poisson family with a log link fit well, for models 5 and 6, a gamma family with a log link was chosen and for model 2, a gamma family with an identity link. The predicted values from the logistic regression and the GLM predicted values for each algorithm were multiplied together to give predicted values for the two part model.

Validation

The 18 models (six using each of OLS, GLM and two part) were used to predict EQ-5D scores within the validation sample, from the SGRQ. Summary statistics from the different models are presented in table 6.12. Predicted mean EQ-5D scores for all 18 models performed well, predicting either the same value as was observed of 0.72, else 0.71. The range of predicted values was found to be generally wider when using the GLM models compared to OLS and wider still when using the two part model. OLS model 2 was the only OLS based model to predict scores within the possible range of EQ-5D utility (≤ 1), with all other OLS models predicting values of greater than 1. All the GLM and two part models predicted scores of less than 1.

Table 6.12 Observed utility compared to predicted utility scores using the 18 models in terms of mean scores, minimum and maximum values and RMSE for each of the model, validation sample (number observations=4888).

	Mean	Min	Max	RMSE
Observed	0.72	-0.43	1.00	
OLS1	0.71	0.30	1.14	0.1746
OLS 2	0.71	0.13	0.98	0.1723
OLS 3	0.72	0.28	1.11	0.1740
OLS 4	0.72	0.27	1.09	0.1739
OLS 5	0.71	0.24	1.01	0.1726
OLS 6	0.72	0.22	1.02	0.1746
GLM 1	0.71	0.13	0.99	0.1724
GLM 2	0.71	0.11	0.99	0.1724
GLM 3	0.72	0.13	0.99	0.1722
GLM 4	0.72	0.13	0.99	0.1722
GLM 5	0.71	0.14	0.97	0.1724
GLM 6	0.72	0.14	0.98	0.1728
2 part 1	0.71	0.15	0.99	0.1725
2 part 2	0.71	0.11	0.98	0.1723
2 part 3	0.72	0.11	0.99	0.1722
2 part 4	0.72	0.12	0.99	0.1720
2 part 5	0.71	0.10	0.97	0.1737
2 part 6	0.71	0.07	0.97	0.1741

RMSE= root mean squared error

Based on RMSE it was seen in table 6.12 that algorithm 2 using OLS, algorithm 3 and 4 with the GLM and algorithm 4 from the two part model performed well, with RMSE values of: 0.1723, 0.1722 and 0.1720 respectively. RMSE within the 18 models ranged from 0.1720-0.1746. Algorithms 2 (OLS) and 4 (GLM and two part) were compared in terms of predictive ability for utility scores by disease state. The predicted scores were very close to the observed scores, irrespective of model employed.

Table 6.13 Mean (SD) observed EQ-5D utility compared to mean predicted EQ-5D utility scores by disease severity using the best fitting OLS, GLM and two part models, validation sample

Model	Disease severity		
	Mild	Moderate	Severe
Obs, n	1681	2380	827
Observed	0.752 (0.22)	0.708 (0.23)	0.672 (0.22)
Pred -OLS	0.752 (0.14)	0.704 (0.15)	0.667 (0.15)
Pred-GLM	0.754 (0.15)	0.705 (0.15)	0.667 (0.14)
Pred-2 part	0.755 (0.15)	0.706 (0.15)	0.666 (0.14)

As seen in tables 6.12 and 6.13, algorithm 2 using OLS, predicted as well as the algorithms developed using more complex model structures. The 0.0003 difference between the RMSE of the OLS model and the RMSE of the best fitting two part model (calculated from table 6.12), was small and the mean score and range of minimum maximum values were comparable with the best fitting models using either GLM or two part, as seen in table 6.12. Table 6.13 shows that all the algorithms predicted utility scores close to observed scores by disease severity. On the grounds of parsimony, OLS algorithm 2 was seen to have a good fit for mapping the EQ-5D from the SGRQ and was taken through for further validation.

Recommended Model

OLS equation 2 was refit onto dataset 2, where observations had complete EQ-5D and SGRQ total scores (n=3855, Obs=18 505), to give a best fit mapping algorithm for predicting EQ-5D index scores from SGRQ which is described below:

EQ-5D utility = 0.9617 – 0.0013 SGRQ Total – 0.0001 SGRQ Total^2 + 0.0231M

M= male

QALYs

Within the full dataset, QALYs derived using observed utility scores were compared to QALYs derived from utility scores predicted using the algorithm described above, by disease severity and by treatment arm. Results by disease severity are presented in table 6.14 and show that whilst predicted QALY scores were slightly larger than the observed values, the absolute difference between QALYs from the mild to moderate groups was the same at 0.17. However, the difference between the moderate and severe groups was larger in the observed than the predicted values (0.24 compared to 0.21). Mean QALY scores by treatment group were slightly larger for the predicted QALY scores than for observed scores.

Table 6.14 Mean (SD) QALY scores by COPD disease severity group, using the observed utility scores compared to the predicted utility scores to generate QALY scores in the full dataset

	Mild	Moderate	Severe
Observed	2.16 (0.68)	1.99 (0.74)	1.75 (0.75)
Predicted	2.18 (0.52)	2.01 (0.57)	1.80 (0.58)

The absolute difference between treatment 1 and treatment 2 was the same (0.07) using observed and predicted scores and the difference between the most and least effective treatments in terms of QALY gain was equivalent at 0.11. However, using the predicted utility scores to derive QALYs gave a different relative ordering of treatments in terms of QALY gain, than using the observed utility scores. This is demonstrated in table 6.15 where subjects taking treatment 3 had a marginally larger QALY gain than those on treatment 4 within the observed data, but when using predicted scores, treatment 4 was associated with more QALY gain than treatment 3. These differences are slight.

Table 6.15 Mean (SD) QALY scores by treatment group, full dataset using the observed utility scores compared to the predicted utility scores to generate QALY scores

Treatment	Study result	Predicted
1	2.09 (0.69)	2.11 (0.54)
2	2.02 (0.73)	2.04 (0.57)
3	1.99 (0.74)	2.00 (0.59)
4	1.98 (0.75)	2.02 (0.56)

6.3.4 Discussion

A wide range of potential model structures were considered to ensure that a systematic and robust approach was taken in identifying an appropriate model. Most previous studies on mapping stop after developing an algorithm based on OLS. This paper went beyond OLS and carried out a rigorous approach to model fitting using not only OLS but GLM and two part models. Use of fractional polynomials in the development stage of the study identified significant squared terms to include within the models. Fitting the GLM models employed a number of statistical tests to determine the appropriate family and link function for each algorithm. A two part model was developed based on a logistic regression and GLM in order to better fit the distribution of the data, with the intention of improving predictive accuracy. The predictive ability of each of the models was validated using RMSE in the validation sample. A prediction equation developed using OLS was found to perform as well as more complex model structures in terms of RMSE within the validation sample, with RMSE values comparable to RMSE from other mapping studies.[218] The finding that increasing the complexity of the model only achieved slight improvements in predictive ability was consistent with that reported elsewhere.[218] The chosen model predicted a range of values that remained within the possible range for EQ-5D utility scores (≤ 1). The equations were not made to take account of multiple observations per participant on the grounds that it was the relationship between the SGRQ and the EQ-5D that was of

interest. There was no reason to assume that this relationship would change if there was more than one observation per participant.

Validation occurred in a sample that was not used to develop the model. Importantly, the validation sample was identified from a non random split of the data. If the data were randomly split into a fitting and a validation dataset, these groups would be almost identical in terms of population characteristics, therefore a prediction equation developed in one of the groups should predict well in the other group. A non random split was chosen based on country, because country was found to be statistically insignificant within exploratory analyses of the dataset and the number of USA subjects within the trial totalled about 1/3 of the complete case dataset so non-USA/USA country was an appropriate candidate on which to split the data. Using a non-random split as opposed to a random split did not affect the final choice of model: during exploratory analyses in which a random dataset was used, the final equation included both a total and a total^2 term.

The prediction equation was refit onto the whole dataset and was used to predict QALYs using the AUC method. Predicted QALY scores were compared against QALY scores obtained from observed utility data, derived using the same AUC method. Applying the prediction equation was found to produce slightly higher QALY scores than using observed utility data by disease severity and by treatment group, but the absolute differences between disease severity groups were similar using either observed or predicted values to derive QALYs. These inflated QALY scores may have been due to rounding error in the coefficient for the SGRQ total^2 term. A way to improve the accuracy of the prediction equation might be to rescale the SGRQ onto a scale of -1 to +1, and is of interest for future research.

The relative treatment ordering of the four treatments was different using observed utility scores to generate QALYs compared to using predicted utility scores. Whilst the resulting

prediction equation appears to perform well, given the size of QALY gains to be had between treatments, small differences in QALY scores can lead to large impacts on cost effectiveness outcomes. Therefore it is important to measure utility scores as accurately as possible and the best way of doing this is to routinely value the impact of treatment on HRQoL using a utility based measure, such as the EQ-5D within the clinical trial. A mapping algorithm such as the one developed here can be used to predict EQ-5D utility scores from the SGRQ and may be useful in some situations, however for use within a HTA submission where precision of estimation is important, it is recommended that utility scores are directly derived from the clinical trial population and that mapping is used as a second best solution only. This conclusion differs from an earlier study which reported on a mapping between a different HRQoL measure and the EQ-5D in estimating QALYs.(237)

Missing data was dealt with by using complete cases for the development of the algorithm because it was the relationship between the SGRQ variable and the EQ-5D variable that was of interest, and this was not thought to be affected by reasons for missingness (in contrast to if the focus of the study was to accurately estimate EQ-5D utility scores), rather is was thought that if imputation methods were used, this might blur the relationship between the variables. Imputation of missing values was used for the estimation of QALYs and this was achieved using last observation carried forward and first observation carried backwards in order to create full follow up of SGRQ total and EQ-5D utility scores.

A recent study has shown the impact that 10 different imputation methods for inputting missing quality of life scores had on the resulting ICER. ICER values for the same procedure were between \$178 000 and \$433 000 depending on the imputation method used. The frequently used method of last value carried forward resulted in an ICER in the middle of these estimates at \$292 000 and restriction to a complete case analysis resulted in the largest ICER of \$433 000. Because of the large difference in results, the authors

recommend using two imputation methods, one of which to always include multiple imputation.(152) An extension to this current study would be to assess different imputation methods for utility, for the purpose of deriving QALYs from clinical trial data in order to assess the impact on the QALY estimates.

6.4 Conclusion

Utility data are integral to economic evaluation in health care. Within this chapter, two strategies for deriving utilities and QALYs were employed. The first strategy was to derive utility data directly from a relevant clinical trial in which a utility based measure, the EQ-5D, was routinely collected. Deriving utility based values directly from trial data can either be simple, as was initially carried out such that mean EQ-5D utility scores are summarised by disease severity, or alternatively, deriving utility values can be more complex. More complex techniques are appropriate when, for example, it is important to account for subject heterogeneity in utility scores. Within this chapter, a GLM model with a Poisson family and identity link with independent variables of treatment group, disease severity, sex, BMI and race was developed to generate EQ-5D utility scores. The derivation of QALYs by treatment group is frequently the next step after deriving utility scores. Mean QALY scores by treatment group were estimated using the method of AUC. A GLM model was used to derive predicted QALY scores, by treatment group and the results were compared.

Eliciting utilities from trial data becomes a problem when utility based measures such as the EQ-5D have not been collected within a trial and instead only disease specific questionnaires such as the SGRQ have been employed. For reimbursement decisions, it is necessary to show value for money in terms of effectiveness, preferably using a generic effectiveness measure based on utility. The second strategy employed within this chapter

was to elicit utilities and QALYs for application where a utility based HRQoL measure such as the EQ-5D has not been routinely collected, by developing a mapping equation so that where a utility based measure has not been collected, a mapping equation can be used to predict EQ-5D utility from a disease specific HRQoL measure. However, the development and validation of a prediction equation is only an option where both measures have been collected, such as in the TORCH study.

In applying the resulting mapping equation it was found that in some situations, such as predicting mean EQ-5D utility scores, the mapping predicts well. Therefore if the use of the algorithm was to derive simple EQ-5D scores by disease severity, then the use of such an equation is likely to give a fairly accurate mean score. In other situations the prediction equation should be used with caution. The algorithm was used to derive QALY scores by treatment group and found that by employing the prediction equation to predict utility scores and then to work out AUC led to a slight underestimation of the relative treatment effect and a difference in the relative ordering of treatments. Mapping is only advisable when no utility data have been collected and where there is no alterative, as it is regarded as a second best solution. As EQ-5D was routinely collected within the TORCH dataset, utility values can be derived directly from trial data and this is the approach used in the following chapter on modelling.

Chapter 7. A New Economic Model for COPD

Resource allocation decisions for healthcare are made with respect to opportunity cost: decision makers need to know what the impact of a new treatment will be in terms of impact on health and impact on cost in comparison to the costs and effects of current treatment or current best practice. The real system is highly complex and Health Economists use economic modelling in order to combine information on the natural history of the disease and the effect of treatment in a potentially useful and meaningful way.

There are many challenges in developing an economic model, some of which can be overcome early on in the process by clear conceptualisation. During conceptualisation, it is useful to consider how the disease could be modelled and to identify key aspects of the disease that affect disease progression and impact upon HRQoL and/or cost. Following conceptualisation, an appropriate modelling structure can be selected which allows for these components to be included within the model. With structural decisions made, the model can be developed using available and relevant data sources.

This chapter presents the development and results of a new concept in economic modelling for COPD. The model employs a series of regression equations to parameterise key drivers of cost effectiveness: lung function, symptoms and exacerbations and from these, informs regression equations for cost and effectiveness that predict cost and effect values per cycle. Values for a current treatment arm and a comparator arm are calculated separately and for each arm, costs and effects are summed over time to give cost effectiveness statistics. This chapter is focussed on conceptualising and developing an economic model and is split into three sections.

In the first section, a conceptual framework for an economic model of COPD is developed which considers incorporating the natural history of the disease (as studied within the thesis) into the model. The section principally considers the major components that are likely to bring about an impact on either cost or HRQoL, or both. Based on this, an economic structure is selected and the rationale for preferring the chosen economic structure, a regression based model, for modelling the disease area is explained. The first section also contains information pertaining to all economic models, including use and perspective, timeframe, discounting and uncertainty, and draws on the experience gained from the published literature on COPD economic models as described in Chapter 4.

The second section describes the methods used in developing the model, starting with the generation of individual regression equations for each of: lung function; exacerbations; symptoms; EQ-5D utility; cost; and survival probability. For each regression equation, the rationale for the selection of the dataset within which the equations were developed, the type of regression model used and choice of explanatory variables, as well as the resulting model and interpretation of the results, is given. These regression equations are combined to form an economic model for COPD, representing current treatment.

In the final section, potential effects of treatment on each component are considered in relation to the model so that a treatment arm can be developed. Exploratory analyses around a hypothetical treatment for COPD are presented and resulting ICERs produced.

7.1 Conceptual framework

The conceptual framework presented within this section is important within the model development process as it enables various known components of the disease to be considered and pieced together, in a structured way, before the model is built.

7.1.1 Use and Perspective

In developing a new economic model it is important to decide at the outset what the end use of the model will be: what it will assess and who it is for. The aim is to develop a model that represents the natural history of the disease for a range of COPD patients, and on top of that, layer in a treatment effect. This new COPD model will be flexible with regards to the type of intervention it will be employed for and the treatment effect will be dependent upon the intervention's mechanism of action. It will also be able to incorporate heterogeneity in the disease population. Comparing costs and utilities in the population receiving current treatment, to those receiving the new treatment, will give cost effectiveness statistics.

Primary interest lies in the UK market and because the UK is one of a small number of countries that requires HTA for treatments, the NICE guidelines for HTA are followed, which recommend that the most appropriate provider perspective is the UK NHS (see Chapter 3). NICE also recommends that all economic evaluations should be based on UK population preferences and this was adopted within the chapter.

7.1.2 Key Components of the Model

Within cost utility analysis the aim is to determine the cost effectiveness of one treatment in comparison to another. A case for promoting one treatment over another, especially in COPD, has historically been based on improvements in effectiveness rather than a reduction in cost.

It is important to bear in mind the role of treatment for COPD as this has a bearing as to how manufacturers target their products and subsequently provides an insight into the potential effects of treatment. The role of treatment for COPD:

"…in the absence of a disease cure, is to prevent and control symptoms, reduce the frequency and severity of exacerbations, improve health status and improve exercise tolerance."(2)

Different treatments have different mechanisms of action, targeting any one or more of the factors identified in the quote above. Through the treatment action there is an element of causality whereby, for example, treatment eases symptoms and as a result HRQoL improves and impacts upon QALYs. In the same way, cost is impacted on through (other than cost of the specific treatment), the benefits (or side effects) of taking a particular treatment. For example, if treatment causes an absolute reduction in the rate of exacerbations requiring hospitalisation, the financial burden should be less than otherwise.

As described in the quote above, and as has been seen throughout this thesis, there are three important factors that impact quality and quantity of life, affect costs and influence mortality risk for people with COPD. These are: exacerbations, symptoms and lung function. The statements reproduced below, extracted from the NICE COPD guidelines, further illustrate that each of these elements are tied up in the effectiveness of treatment:

"The effectiveness of bronchodilator therapy should not be assessed by lung function alone but should include a variety of other measures such as improvement in symptoms, activities of daily living, exercise capacity, and rapidity of symptom relief."(7)

"[The aim of treatment with] inhaled corticosteroids…is to reduce exacerbation rates and slow the decline in health status and not to improve lung function per se."(7)

"The clinical effectiveness of combined treatments can be assessed by improvements in symptoms, activities of daily living, exercise capacity and lung function."(7)

Most of the published models, as described in Chapter 4, have used lung function as the driver of progression through the model, with episodes of exacerbations occurring at a pre-defined rate within each health state, with increased frequency as disease severity worsens. Symptoms have not previously been modelled explicitly. Nevertheless, because these three elements are each important components for the treatment of COPD, it is desirable that they are all included within the model because of their individual impact on costs and effects.

In the following sections, further details are given that support the use of each component for modelling COPD.

Lung Function

FEV_1 naturally declines over time and is frequently used to define disease severity groupings for COPD based on FEV_1 % predicted (as seen in Chapter 2). These groupings are also used for treatment allocation; with more treatments being prescribed the more severe the disease (see section 2.3). The more severe the disease, generally the more exacerbations occur (Chapters 2 and 5) and the worse symptoms are. The majority of existing models have chosen lung function as the key driver of movement through the model (see literature review on published COPD models in Chapter 4) with the use of Markov states representing disease severity groups based on FEV_1 % predicted.

Costs increase substantially as disease severity moves from (GOLD) moderate to severe.(24;27-32) One study estimated the average direct cost for a mild COPD patient to be €232 (£207) for moderate disease, €477 (£427) and for severe disease €2026 (£1812).(33) As FEV_1 deteriorates, a general shift from outpatient care to hospitalisation, an increase in the use of oxygen therapy and a subsequent increase in total costs, especially in the most advanced stages of the disease has been shown to occur.(33)

Symptoms

The majority of people with COPD suffer from respiratory symptoms including breathlessness, phlegm (and cough) and wheeze, and some treatments work explicitly on improving symptoms. Symptoms have not previously been incorporated into an economic model for COPD despite the obvious detrimental HRQoL effects of respiratory symptoms. Ignoring symptoms, an important element of HRQoL, within an economic model, risks excluding a major target area for some COPD treatments.

Symptoms are a more subjective outcome measure of treatment for COPD than either COPD exacerbations or disease progression and the impact of symptoms on HRQoL is more difficult to capture within an economic model in a meaningful way and is perhaps why to date it has largely been ignored within economic models. Nonetheless it is known with certainty that a treatment that reduces symptoms will have a beneficial impact upon HRQoL for that person. Therefore it is important to capture this effect in the disease model. The following quote illustrates the importance of symptom control from the perspective of the patient and how this can conflict with the more commonly used approach of disease measurement of lung function.

> "From a physician standpoint the Holy Grail of COPD disease modification is to halt, or at least slow down, the rate of decline of FEV_1. However, from a patient perspective the Holy Grail is simply to be able to breathe easier."[1]

One of the aims of this new model is to bring together these components of the disease.

Exacerbations

As discussed in Chapter 2, exacerbations are the leading driver of cost in COPD. Exacerbations account for between 35-40% of the total per capita health care costs for

COPD.(36) Treatment which acts to reduce or prevent disease progression and/or an exacerbation (particularly severe exacerbations) will have a direct effect on the total cost for COPD.(38) The cost of exacerbations increases with the severity of exacerbations. An exacerbation is the main reason why a COPD patient would attend hospital. In England and Wales, it has been calculated that for every exacerbation-related hospital admission avoided, a total saving of approximately £1200 would be made.(37) The prevention of a minor exacerbation would reduce costs by avoiding a visit to the GP.

In addition to COPD exacerbations being a major driver of cost for the disease, they also negatively impact HRQoL. Treatments for COPD that reduce the frequency of exacerbations and/or the intensity of the exacerbation will improve the HRQoL for the person involved.

Survival

In economic models for COPD it is important to incorporate mortality, and within the literature review this was identified as a weakness in several published economic models (see Chapter 4). It is known that disease severity (as defined by FEV_1 % predicted) is related to mortality and that as the disease worsens, mortality risk increases (see Chapter 5). It is also known that whilst some people with COPD die of COPD, many die of other causes (Chapter 5), therefore it is important to incorporate mortality, but to incorporate mortality in a way that it is a function of disease severity and also to model all-cause mortality as opposed to disease-specific mortality. Within economic models, survival probability can be used to represent the risk of mortality, which was seen for some of the models reviewed in Chapter 4.

Cost and Utility

Costs and utilities are a function of disease so that, *a priori*, as disease worsens, costs would be expected to increase and utility worsen. Throughout this section it has been argued that the natural history of COPD can be represented in terms of the three components of: lung function, symptoms and exacerbations, and within the model it is these components that should be incorporated and combined in a way in which cost and utility are elicited. The risk of mortality and the resulting impact on costs and utilities needs to be accounted for within the model. The way in which these strands are interwoven is dependent upon the chosen model structure.

7.1.3 Model Structure

Economic evaluation in COPD has traditionally been carried out either alongside a clinical trial or through decision analytic modelling (as described in detail within Chapter 4) with all published economic models adopting a Markov structure, with one exception.(153) The overriding difference between the aim of this model and previously published models is that a main aim is to describe the natural history of disease within a modelling framework, before layering on top (potentially any) COPD treatment effect. This contrasts with usual approaches of modelling one pre-specified treatment against another.

The Markov approach to modelling is limited in a number of ways. First that in order to incorporate each of the components (lung function, exacerbations and symptoms) into the model, there would be an increase in the size and complexity of the Markov model compared to previously described models. For any different set of combinations of components, a health state representing that combination would be required. For example, one could imagine a health state in which the subject had moderate COPD, with severe

respiratory symptoms and who had a severe exacerbation in the space of one year. Another person could have moderate COPD, moderate respiratory symptoms and have had no exacerbations over one year and yet another could have moderate COPD, mild respiratory symptoms and an exacerbation. These states would have different utilities and different costs associate with them and should be represented by different health states, or by incorporating an additional exacerbating state. Ultimately there could be a large number of different states in the model. Even with a large number of states, the resulting model may be insufficiently flexible to account for population heterogeneity in the way that would meet the aim of the new economic model, as was set out earlier in this chapter.

A second limitation is that lung function, which is usually employed as the key driver through the model, is forced into discrete states, and once in those states the model is memoryless, there is no count of how long somebody has spent in the state. As COPD is a progressive disease, people who have just entered a disease state are less likely to enter a worse state than somebody who has remained in that state for a period of time. The Markov model cannot treat these two scenarios differently where they are grouped in the same Markov state.

A third reason why a Markov structure may be limited is that as one of the aims of the model is to incorporate a treatment effect that is dependent upon the intervention's mechanism of action, it is important that the model can incorporate outcomes of COPD clinical trials. In a Markov structure it is difficult to model the impact of, for example, absolute improvements in lung function into the model. Often this effect has been modelled by assuming a blanket reduction in the transition probabilities to the next state. However of particular interest is how to incorporate the reported gains in FEV, for example a 100ml improvement over one year, into a model framework.

An alternative to Markov modelling that has been used in asthma and once in COPD (summarised in Chapter 4) is regression based modelling,(238;239). As previously described within Chapter 3, the regression based model operates through a series of regression equations that predict values for key components. The published regression based model for COPD directly predicts costs and effects using similar methods to those described in the utilities chapter.(153) However, a regression based model does not need to be limited to just cost and effectiveness equations. Taking this a step further, a regression based model structure could be more flexible because the key components identified previously: lung function, exacerbations and symptoms, can be individually modelled and then separate equations for cost and effect can be developed that employ the predicted values from the various component regression equations. This approach can account for patient level heterogeneity and relationships between the components. Like a Markov model, the regression equation based model can be extrapolated to take up to a lifetime perspective so every year, as the cohort ages, the outcome values of each regression equation change.

A limitation of the regression based approach is the necessity to have access to large and relevant datasets through which regression equations can be developed in order to model the disease. The Renfrew/Paisley (MIDSPAN) dataset and the TORCH RCT are suitable datasets on which regression equations can be developed, therefore a regression based structure was employed for the COPD model and was developed around the key components of COPD as previously identified, namely lung function, exacerbations and symptoms.

7.1.4 Generic Modelling Considerations

Some additional considerations for the model, as set out in the NICE reference case (Chapter 3) are described below, including the rationale for the exclusion of productivity costs and side effects, the timeframe of the model and discounting.

Productivity Costs

Days lost from work due to illness, representing productivity costs to the economy, are a financial cost of illness to society. Productivity costs for COPD represent a significant burden on society as COPD is a major cause of absenteeism from work.(24;240) As previously discussed in Chapter 2, people with COPD have a "substantially shortened" work life compared to the population average,(41) and within the UK, 44% of COPD patients were below retirement age. Productivity costs were found to be almost equivalent in size to direct costs.(24)

Whilst days lost from work represent a substantial burden on the individual and on society, their use in cost effectiveness analyses is dependent upon the perspective adopted for the study, as discussed in Chapter 3. NICE limits its perspective to the health service and personal social services perspective and explicitly excludes the use of productivity costs in its evaluations. As the economic model that will be developed is focussed in the UK, for potential use for submissions to regulatory bodies such as NICE, productivity costs are mentioned here, but are not of primary concern for inclusion within the model developed in this chapter.

Side Effects

Major side effects occurring from COPD treatment are relatively few in number compared to other disease areas. For example, the economic model by Jubran et al,(97) considered the toxicity of theophylline compared to ipratropium and found toxicity rates of 8% and 0% respectively.(97) Theophylline has been on the market for a long time and is unlikely to be considered as a comparator in an economic analysis; whereas ipratropium is frequently used within economic evaluations (see Chapter 4). It is assumed that side effects for COPD treatments are relatively insignificant and that there is little between treatment variation in side effects. Therefore side effects were not included within the model.

Timeframe and Discounting

The timeframe of the model was lifetime (to a maximum of 90 years or at the point where $S(t) \leq 0.05$, whichever occurs sooner) because of the reasons previously outlined in section 3.2.7. Cycle length was set at one year, and is consistent with cycle lengths used in published COPD models (refer to Chapter 4 for details). A discount rate of 3.5% for costs and effects was applied, as recommended by NICE (see section 3.2.8).

7.1.5 The Conceptual Model

A conceptual model was developed in order to piece together the relationship between the individual components, and the relationship between the components and costs and QALYs. The paths of effect between the components and HRQoL and cost are unlikely to be completely independent from one another: a degree of interdependence is likely to exist. For example, is has already been seen in Chapter 5 that there is a relationship between lung function and symptoms, so that if lung function worsens, symptoms will usually deteriorate, and as a result, the expectation is that HRQoL score would decrease.

Another relationship is between lung function and exacerbation severity; that as lung function becomes more limited, the severity and frequency of exacerbations tends to increase,(36;241;242) which in turn affects HRQoL. The relationship between the components and HRQoL is difficult to untangle, however one of the benefits of a regression based model is that this interdependence can be modelled explicitly because predicted scores from one regression equation such as lung function, can be used as explanatory variables in regression equations that predict symptoms and exacerbation rates.

Within figure 7.1, FEV_1 has been conceptualised as the key component of the disease that drives symptoms and exacerbations, and this is an assumption used throughout this chapter. The rationale for this assumption is that previously lung function impairment (as defined by FEV_1 % predicted) was shown to be related to the number of exacerbations experienced and to the severity of respiratory symptoms (Chapter 5). The links between FEV_1 and exacerbations and FEV_1 and symptoms are assumed to be unidirectional, so exacerbations are a function of lung function but lung function is not a function of exacerbations, and symptoms are a function of FEV_1 but FEV_1 is not a function of symptoms.

So far the conceptualisation of a mechanism for modelling the natural history of the disease has been presented, which will represent the baseline non-treated/current treatment population. For the treated population, treatment effects are layered on top of the natural history model in a comparator arm. Figure 7.1 illustrates ways in which treatment (Tx) effects may occur that influence HRQoL. Costs are assumed to be affected by similar relationships, but are not presented here. Treatment that instantly and directly affects HRQoL (first line of figure 7.1) has already been ruled out, and the second to the fourth treatment effects shown in figure 7.1 are simple relationships, so that treatment which: improves symptoms; or reduces exacerbations; or increases lung function, should improve HRQoL. In previous models and in RCT reports, some of the relationships between the

components may have been overlooked. For example, trials report absolute change in
FEV_1 and change in exacerbation rates separately, yet it may be that the effect of treatment
on FEV_1 is responsible, at least in part, for the change in the rate of exacerbations, as
illustrated by the second to last relationship in figure 7.1. Another possibility is that the
treatment reduces exacerbations with a knock on effect on symptoms as shown in the last
line of figure 7.1, which jointly improves HRQoL. Figure 7.1 is purely illustrative and
other relationships between treatment and the three components are possible.

Tx ⇸ HRQoL

Tx ⟶ Symp ⟶ HRQoL

Tx ⟶ Exacs ⟶ HRQoL

Tx ⟶ FEV1 ⟶ HRQoL

Tx ⟶ FEV1 ⟶ Symp ⟶ HRQoL

Tx ⟶ FEV1 ⟶ Exacs ⟶ HRQoL

Tx ⟶ Exacs ⟶ Symp ⟶ HRQoL
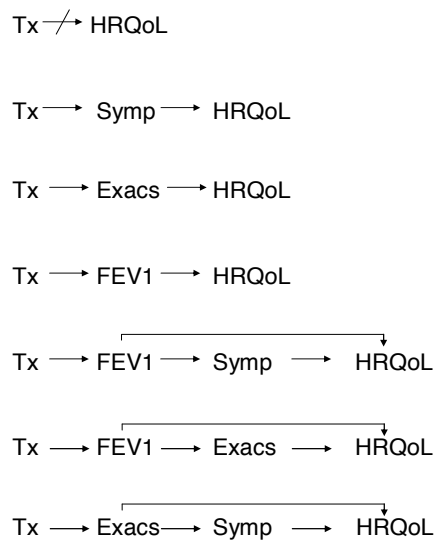
Figure 7.1 Possible treatment effects

Of those treatments assessed with regards to mortality gain in RCTs; the existing evidence
around treatment extending life is not statistically significant, therefore the effect of
treatment on the regression equation for survival will not be considered within this model,
though extensions to the model could be made to incorporate this.
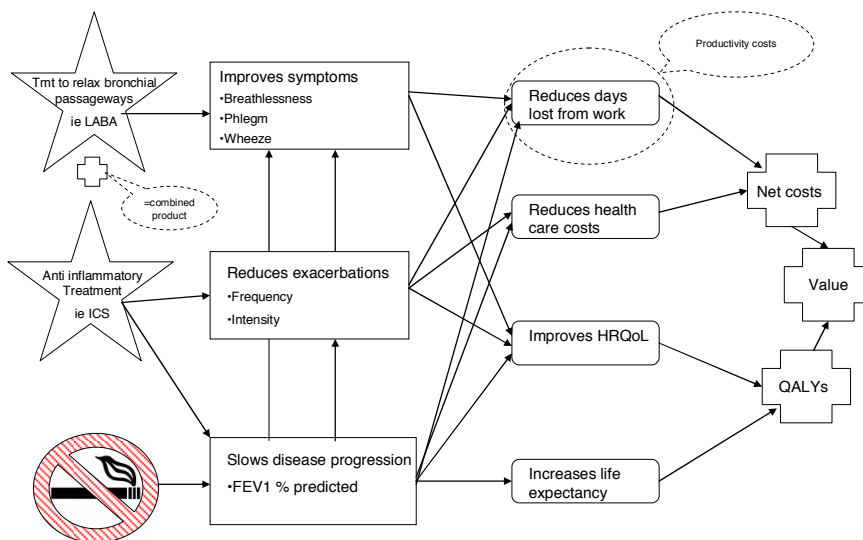
Figure 7.2 Conceptual model for COPD

Figure 7.2 above illustrates how four different interventions may work on the different components of COPD and how in turn these may impact upon costs and QALYs. As previously noted, the effect of treatment on each of the components is dependent upon the treatment's mechanism of action. For example, a LABA such as salmeterol has a mechanism of action that relaxes the bronchial passageways, which in turn has an impact upon symptoms; the airways open and breathing becomes easier. A different product like an ICS such as fluticasone, has an anti-inflammatory effect that reduces the number of exacerbations and may also slow disease progression. A dual product such as Seretide (salmeterol plus fluticasone) combines the effects of the LABA and the ICS, essentially working on all three of the components.

Smoking cessation impacts directly on disease progression by slowing the decline in lung function,(3) as depicted in figure 7.2. Smoking cessation is a non-pharmacotherapy

intervention that has been shown to effectively slow the deterioration in FEV_1 and return the trajectory of lung function to one consistent with that of a non smoker.(3) This model could be used for the economic evaluation of smoking cessation, in order to inform decisions about public health interventions for a COPD population. This is outside the scope of this thesis but is of interest for further application of the model.

7.1.6 The Prediction Model

Turning the conceptual model into an economic model within a regression based structure is relatively straightforward and is described here in terms of a two phase process. The first phase is the development phase in which the regression equations are determined and the second phase refers to how the model is populated.

The development of the prediction equations was carried out in either the TORCH or the Renfrew/Paisley (MIDSPAN) dataset depending on suitability (see section 7.2 for details), and the selected study is referred to as the 'dataset' within figure 7.3. Phase one is split into seven steps. In step 1, potential explanatory variables are identified within the 'dataset', including: age, sex, disease severity at baseline, presence of symptoms and smoking. These are used to develop the equation for lung function (step 2), and to inform a Weibull survival model that is used to represent survival probability, as shown in step 3. Step 4 uses explanatory variables from step 1 and a continuous variable for FEV_1 in a regression equation for exacerbation rate. Step 5 uses explanatory variables from step 1, a continuous variable for FEV_1 and a variable for exacerbations (from the equation in step 4) to generate a prediction equation for symptoms   Explanatory variables of: symptoms, exacerbations and lung function from the dataset (step 6) together with some of those from step 1 are used in order to determine the coefficients on an equation for cost and on an equation for EQ-5D utility as seen in step 7.

1

Explanatory variables from the 'dataset':

- Age
- Gender
- Disease severity at baseline
- Respiratory symptoms
- Smoking history
- FEV_1

2

Are used to determine the coefficients in an equation for FEV_1:

$$FEV_1 = \alpha + \Sigma\beta X$$

3

Are used to derive a Weibull function for survival:

$$S(t) = EXP(-\lambda t^{\gamma})$$

4

Are used to determine the equation for exacerbations

$$Exacs = \alpha + \Sigma\beta X$$

5

And then for symptoms:

$$Symp = \alpha + \Sigma\beta X$$

6

Explanatory variables from the 'dataset' plus:

- Symptoms
- Exacerbations
- Lung function

7

Are used to determine the coefficients in an equation for cost and for utility:

$$Cost = \alpha + \Sigma\beta X$$

$$Utility = \alpha + \Sigma\beta X$$

Figure 7.3 Phase 1, development of the model

Phase 2 refers to how the model is populated and is split into eight steps and is illustrated in figure 7.4. The model is populated in cycles. Each cycle represents one year (see sections 3.2.7 and 7.1.4 for details on timeframe) and for the first cycle, the model runs as shown in figure 7.4. User defined patient characteristics (described in detail in section 7.2), are entered into the model as explanatory variables for FEV_1 (step 1). Predicted values for FEV_1 and values for patient characteristics are inputted as explanatory variables into the equation for exacerbations as shown in step 3. Step 4 uses the exacerbation rate from the model in step 3 with predicted values for FEV_1 and values for patient characteristics as explanatory variables into the equation for symptoms. Predicted values from these equations (step 5) are used as explanatory variables for the cost and EQ-5D utility equations (step 6) to allow for the interdependence of the various components of the disease on one another. The cost equation, and the EQ-5D utility equation are each multiplied by the survival probability (step 7) to give survival-adjusted cost and survival adjusted utility for the cycle (step 8). In subsequent cycles, the model is populated in a similar way to the first cycle with the prerequisite that that the user defined patient characteristics remain the same as for the first cycle, with the exception of age which increases by one each cycle. As time increases, the survival probability, which is a function of time, decreases. Section 7.2 presents detailed information on how the economic model was developed.

To get costs and QALY estimates, the survival-adjusted cost and utility from each cycle are summed and the mean value derived (once discounting has occurred). Once all these steps are completed, a fully operational model exists that represents the natural history of COPD for a defined patient population on current treatment. A treated population is developed in exactly the same way as described above except that a treatment effect is incorporated into the model. The treatment effect is dependent upon the treatment's mechanism of action and the modeller specifies which component(s) is affected by the treatment being modelled. Details of treatment effects are contained in section 7.3. The ICER is found by calculating the difference between the costs and QALYs in the new treatment group to those in the current treatment group.

1

User defined Patient Characteristics (PC):

•Age

•Gender

•Disease severity at baseline

•Respiratory symptoms

•Smoking history

2

Are used as explanatory variables in equation on FEV$_1$:

$FEV_1 = α + ΣβX$

3

Are used with predicted FEV$_1$ values in equation for exacerbations:

$Exacs = α + ΣβX$

4

Are used with predicted FEV$_1$ and exac rates in equation for symptoms:

$Symp = α + ΣβX$

5

Predicted values from:

$Symp = α + ΣβX$

$Exacs = α + ΣβX$

$FEV1 = α + ΣβX$

6

Become explanatory variables in:

$Cost = α + ΣβX$

$Utility = α + ΣβX$

7

Multiplied by survival probability:

$S(t) = EXP(-λt^γ)$

8

Gives survival adjusted cost and effect estimates for each cycle:

▬ Survival adjusted cost

▬ Survival adjusted utility

Figure 7.4 Phase 2, population of the model

7.1.7 Uncertainty

Each of the prediction equations predicts a value based on selected patient characteristics. Each coefficient within the equations has a confidence interval attached representing the uncertainty attached to its value. These measures of uncertainty were included within the model and can be used to examine uncertainty by selecting values within the distribution (assumed to be normal) on that variable.

The benefit of the regression based modelling approach used here, in terms of uncertainty, is that a wide range of user defined patient characteristics can be inputted and the impact of this on costs and QALYs can be assessed. In this way, uncertainty around patient level heterogeneity is diminished. The other side of this flexibility is that an average ICER is not immediately available. For example potential populations in which a new treatment could be assessed could be the different disease severity groups, as defined at baseline. To calculate cost effectiveness statistics by disease severity group, one method would be to enter mean values into the model as user defined patient characteristics for each disease severity at baseline. Alternatively, Monte Carlo simulation could be employed in order to select (with replacement) user defined patient characteristics from distributions.

7.2 Model Development: the Natural History of COPD

Within this section, the methods and results from the development of each regression equation are presented. A range of different statistical models were employed to develop the regression equations for use within the economic model, in order that when combined create a model that represents the natural history of COPD. All analyses were carried out

on either the observational dataset, Renfrew/Paisley (MIDSPAN) or on the TORCH RCT, which have previously been described in full in chapters 5 and 6 respectively.

In order to piece together the economic model using each of the prediction equations, the software package, Microsoft Excel™ was used. Following the development of each regression equation, examples of how each regression equation predicts within the economic model for men aged 55 years at baseline, with NICE diagnosed mild and moderate COPD, are presented. An example of the full economic model in these groups, which represents current treatment, is presented at the end of the section. In brief, each new row of the model represents an additional year of age/time and each column contains predicted values from each regression equation. Column headings include: age, time, lung function, exacerbations, symptoms, cost, utility, survival, survival adjusted cost and survival adjusted utility. The model was run with a lifetime perspective and costs and utility were summed over the duration of the study. A discount rate of 3.5% was applied to both costs and utility.

7.2.1 User Defined Patient Characteristics

The model was designed so that different user defined patient characteristics could be offered to the model to enable the cost effectiveness of treatment within different subgroups to be assessed and therefore to minimise uncertainty due to heterogeneity.

Depending upon the specific regression equation, the user defined explanatory variables (at baseline) could include any of the following: age, sex, height, FEV_1 % predicted, respiratory symptoms, UK, disease severity, smoker/ex-smoker and greater than or less than a 10 pack year history, as described in table 7.1 below. Age is entered into the model in years and is restricted to people aged over 45 years, which represents the typical

minimum age of a COPD patient and also the patient population within both the Renfrew/Paisley (MIDSPAN) study and the TORCH RCT. Height can be entered into the model as any feasible value but the mean height in England, which is 175cm for men and 162 for women was employed within this chapter.(243) Disease severity at baseline was split into four groups: 0, 1, 2 and 3, defined by FEV_1 % predicted, where group 0 is the baseline group which represents people with $FEV_1 \geq 80$% predicted, group 1 contains people with mild COPD ($50 \leq FEV_1 < 80$ % predicted), group 2 with moderate COPD ($30 \leq FEV_1 < 50$ % predicted) and group 3 people with severe COPD ($FEV_1 < 30$ % predicted). The rest of the user defined characteristics are dummy variables, taking the value of 0 or 1. Sex is 0 for men and 1 for women, the variable ex-smoker is 0 for not an ex-smoker and 1 for ex-smoker, over 10 pack years is 0 for less than 10 pack years and 1 for more than 10 pack years. When using the TORCH dataset, because the RCT was multinational and because country-to-country variances are known to exist, a UK variable was also offered to the model where 0 was non-UK and 1 was UK.

Table 7.1 Baseline user defined patient characteristics applied within the modelling

Patient characteristics	Description	Example
Age	years (>45 yrs)	60
Height	cm	162
Disease severity	FEV_1 % predicted (0-1 scale) possible (0): >=0.80, mild (1): 0.50-0.79, moderate (2): 0.30-0.49, severe (3) <0.30	3
Sex	0: men, 1: women	1
Ex-smoker	0: not ex-smoker 1: ex-smoker	1
Over 10 pack years	0: <10 1: ≥10	1
Respiratory symptoms	0: no symptoms 1: symptoms	1
UK	0: no 1: yes	1

FEV_1= Forced Expiratory Volume in one second.

For respiratory symptoms, where the MIDSPAN dataset was employed, a dummy variable for symptoms as previously described in Chapter 5 (where 0 represents no symptoms and 1 presence of respiratory symptoms), was used. When using the TORCH dataset, the same

respiratory symptoms classification was not available, so responses to the MRC dyspnea scale questionnaire were used to identify symptomatic breathless patients. Subjects answering yes to statements 3, 4 or 5 (see figure 7.7) were considered as having breathlessness. Coding for the breathless variable was 0 for no breathless and 1 for breathless.

An example of a specific patient group that could be modelled is presented in table 7.1 which represents a cohort containing 60 year old women, with mean height of 162 cm, ex-smokers, with more than 10 pack years, respiratory symptoms, (therefore NICE diagnosed) severe COPD at baseline and who live in the UK. Characteristics offered to the model can take on any value from the range presented in the description column of table 7.1.

Each regression equation described in this section on model development, contains slightly different explanatory variables depending on the statistical significance for predicting the dependent variable, and on a valid rationale for including or excluding explanatory variables. Specific details of explanatory variables contained within each regression equation are included in the relevant sections.

As seen previously, the populations within the TORCH and the MIDSPAN datasets differ. TORCH contains a COPD population only and MIDSPAN is a general population study that includes but is not limited to COPD patients. Because it is the natural history of the COPD population that is of interest, it was important to only use data from people with COPD in developing the regression equations within the MIDPSAN population. As discussed in Chapters 2 and 5, there are a number of different COPD diagnostic criteria which when applied, identify differing COPD populations. The results from the study on diagnostic criteria (see Chapter 5) revealed that employing the NICE criteria identified a population at higher risk of all cause and COPD mortality than a population identified

using the GOLD criteria. Nevertheless because there is widespread use of the GOLD criteria and to keep the model as flexible as possible, the regression equations were developed within datasets containing GOLD defined COPD, with options for specifying risk factors of over 10 pack years and respiratory symptoms so that either a GOLD COPD group, a GOLD II COPD group or a NICE COPD group (or different combinations of risk factors and lung function) can be modelled, depending on the patient group of interest. In the following sections on the development of each regression equation, summaries of the patient populations within each dataset are not presented on the grounds that they have previously been described in detail within Chapters 5 and 6.

Lung function was assumed to be the primary driver of the model affecting: the rate of exacerbations, symptoms, mortality, QALYs (through EQ-5D utility) and cost. In practical terms this means that the prediction equations for each of these components could potentially contain lung function as an explanatory variable.

7.2.2 Lung function

This section describes how the equation that predicts lung function was developed. It has been seen in previous chapters that both the TORCH and the Renfrew/Paisley (MIDSPAN) study have collected information on lung function, particularly FEV_1 and FVC. This is because these measures give a clear indication as to the respiratory health of an individual and are relatively easy to collect. For developing the lung function equation, the Renfrew/Paisley dataset was chosen in preference to the TORCH dataset because of the wider range of lung function values recorded in the Renfrew/Paisley study: the TORCH trial excluded subjects with a FEV_1 % predicted > 60%, whereas the Renfrew/Paisley study did not. Because one of the aims of this model was flexibility in terms of the population who could be assessed, it is desirable that the model could be used to represent patients

with less severe disease, where FEV_1 % predicted is greater than 60. An equation that predicts lung function values outside of the range in which it was originally developed, could give unreliable estimates, therefore using the Renfrew/Paisley dataset in the development phase is appropriate as it allows for a full range of lung functions to be predicted when the model is operationalised. In addition, because the Renfrew/Paisley (MIDSPAN) study was comprised of a general population group, the COPD cases are more likely to be representative of a general COPD population than those identified in the TORCH trial, which had tight inclusion and exclusion criteria.

RCTs for COPD, including TORCH, usually report improvement in lung function in terms of FEV_1, rather than FEV_1 % predicted. However, as seen in Chapter 2, COPD cases are frequently classified into disease severity groups according to FEV % predicted in order to inform treatment decisions. Therefore when modelling the disease, a question arises: should FEV_1 or FEV_1 % predicted be used, or both? On the one hand, employing FEV_1 in its own right within an economic model may be useful: it is known that FEV_1 values are reported as primary outcomes from clinical trials and that pharmaceutical companies are looking for a way of incorporating FEV_1 improvement explicitly into an economic evaluation for COPD (Helen Rudge, GSK, personal communication). However COPD disease severity is almost always defined in terms of FEV_1 % predicted and so the inclusion of FEV_1 % predicted into the model is useful for consistency and is important for face validity. One solution, and the one which is adopted here, is to develop and employ equations for both FEV_1 and for FEV_1 % predicted within the model.

The Renfrew/Paisley dataset are used for developing an equation that predicts FEV_1 given defined explanatory variables. As previously described in Chapter 5, to calculate FEV_1% predicted from FEV_1, a reference equation is used that predicts expected FEV_1 for a 'healthy person' based on age and height. There are a number of reference equations for

determining predicted FEV_1, including the reference equations previously derived in Chapter 5 for use within the MIDSPAN dataset and a different set for use within the TORCH trial and there are always different reference equations for men and women. The explanatory variables within the MIDSPAN and TORCH reference equations are the same, namely age and height. The coefficients on the explanatory variables are similar in the male equations, however for the female equations, the coefficients differ, particularly in respect to the constant term. In exploratory analysis, the impact of this difference in the reference equations was that the equations used in TORCH were found to give subjects comparatively lower FEV_1 % predicted scores than when applying the MIDSPAN equations on the same population. Because the Renfrew/Paisley dataset was selected for developing the FEV_1 equation, the reference equations from the Renfrew/Paisley (MIDSPAN) population were used to generate predicted FEV_1, on the grounds that they were more likely to be valid for the MIDSPAN population than an externally generated set of equations.

Methods

This section on lung function sets out the methods used to develop an equation for FEV_1 and how the predicted values resulting from this equation were converted into FEV_1 % predicted scores. Equations from this section were taken forward to inform the economic model for COPD.

The analyses to derive a FEV_1 equation were carried out within the Renfrew/Paisley (MIDSPAN) study using subjects satisfying the GOLD defined COPD criteria (as previously described in Chapter 5) for reasons outlined in section 7.2.1. To develop the regression equation for FEV_1, the method of OLS was used with FEV_1 (litres) as the dependent variable, adjusted for user defined patient characteristics of: respiratory symptoms, over 10 pack years, being an ex-smoker, height, age and sex, with a categorical

variable for disease severity at baseline in terms of: 0) $FEV_1 \geq 80\%$ predicted (used as the reference case); 1) $80 < FEV \geq 50\%$ predicted; 2) $50 < FEV \geq 30\%$ predicted; and 3) $FEV_1\%$ predicted $<30\%$. Given the user defined patient characteristics, the resulting equation predicts a corresponding lung function value as FEV_1 in litres.

The reference equations that were used for converting FEV_1 to FEV_1 % predicted are reproduced below from Chapter 5:

FEV_1 (l) in men = -1.859 (0.532)- 0.029 (0.003) x age(yrs) + 0.037 (0.003) x height(cm)

FEV_1 (l) in women = -0.225 (0.230)- 0.029 (0.001) x age(yrs) + 0.024 (0.001) x height(cm)

The estimated FEV_1 as determined by the OLS regression equation was divided by the predicted FEV_1 (for the same baseline characteristics) derived from the reference equations above and multiplied by 100, to get a FEV_1 % predicted value for each cycle within the model.

Results

The coefficients and standard errors for the OLS regression equation on FEV_1 are reproduced in table 7.2. As seen in the last column of the table, all of the explanatory variables were found to be statistically significant ($p<0.001$) with the exception of over 10 pack years and ex-smoker, which had large p-values of 0.610 and 0.383 respectively ($R^2=0.89$). Whilst statistically insignificant, these explanatory variables were retained so that a NICE COPD cohort could be defined.

Table 7.2 Prediction equation for FEV$_1$

Variable name	Coefficient	SE	P
Constant	-0.016	0.109	0.885
Height	0.022	0.001	0.000
Age	-0.022	0.001	0.000
Sex	-0.381	0.010	0.000
Over 10	-0.005	0.009	0.610
Ex-smoker	0.009	0.011	0.383
Symptoms	-0.050	0.008	0.000
Mild COPD	-0.594	0.008	0.000
Moderate COPD	-1.173	0.012	0.000
Severe COPD	-1.610	0.020	0.000

The table shows that for each year that passes, and as age increases by one, FEV$_1$ declines by 20ml. Women have a lower lung function than men by, on average 380ml, *ceteris paribus*. People with respiratory symptoms have, on average, 50ml less lung capacity than those without respiratory symptoms.

By incorporating disease severity at baseline as a categorical value (rather than producing separate equations for each disease severity group), the assumption was that the rate of FEV$_1$ decline (in other words the slope of the curve) was the same, regardless of disease severity. Starting lung function capacity was seen to differ by disease severity so that FEV$_1$ score was higher for subjects with less severe disease (indicating larger lung capacity) and smaller for people with more severe disease (more impaired lung function). As disease severity worsens, a downward shift in FEV$_1$ score is seen. The results presented in table 7.2 show that compared to the group with possible COPD which contains patients with FEV$_1$ % predicted $\geq$80, those with mild COPD had a lung function with 0.6 litres less, those with moderate COPD 1.2 litres less and those with severe COPD, 1.6 litres less capacity.

Tables 7.3 illustrates how FEV$_1$, predicted FEV$_1$, and FEV$_1$ % predicted enter into the model for a hypothetical cohort containing men with mild and men with moderate NICE diagnosed COPD with a starting age of 55 years. Age increases by one year at every cycle

within the model as shown in the first column of the table. FEV_1 scores predicted from the FEV_1 regression equation were inputted into the FEV_1 columns in the table. Predicted FEV_1 values were calculated from the reference equations based on information on height, age and sex and are shown in the predicted (pred) FEV_1 columns. FEV_1 % predicted values are presented in the FEV_1 % predicted columns (where FEV_1 % predicted is equal to FEV_1 divided by predicted FEV_1 multiplied by 100).

Table 7.3 An example of the decline in FEV_1 in men aged 55 with mild and moderate NICE diagnosed COPD at baseline

	Mild COPD			Moderate COPD		
Age	FEV_1	Pred FEV_1	FEV_1% pred	FEV_1	Pred FEV_1	FEV_1% pred
55	2.00	3.02	0.66	1.42	3.02	0.47
56	1.98	2.99	0.66	1.40	2.99	0.47
57	1.96	2.96	0.66	1.38	2.96	0.46
58	1.93	2.93	0.66	1.36	2.93	0.46
59	1.91	2.91	0.66	1.33	2.91	0.46
60	1.89	2.88	0.66	1.31	2.88	0.46

FEV_1= Forced Expiratory Volume in one second.

Table 7.3 shows that the predicted FEV_1 % predicted for men in this mild cohort is 66% and in the moderate group is 47% which are consistent with expected FEV1 % predicted values in the corresponding COPD disease severity groups. Table 7.3 shows that FEV_1 has a steady decline of 20ml per year regardless of disease severity group.

Fewer data were available for developing the regression equations for people with severe COPD and this caused inconsistencies when predicting lung function in this population. In particular, FEV_1 % predicted scores were slightly higher in men than expected. For men aged 55 with severe disease, FEV_1 % predicted was 33%, when it was expected to be less than 30% predicted. In women aged 55 years with severe disease, FEV_1 % predicted scores were as expected (below 30% predicted) at 25% predicted.

7.2.3 Exacerbations

Within this section, prediction equations for the rate of exacerbations are developed and the resulting equations are incorporated into the economic model. There are a number of different definitions of an exacerbation. Consistent with other definitions used in economic modelling for COPD (see Chapter 4), exacerbations were said to occur when there is a worsening of respiratory symptoms requiring treatment with oral corticosteroids and/or antibiotics.(10) There were assumed to be two types of exacerbation: minor and major exacerbations. A minor exacerbation is considered present when a subject is managed exclusively in primary care. A major exacerbation occurs when the worsening of symptoms require that the subject is hospitalised.(144)

The TORCH dataset reports on both minor and major exacerbations whereas the MIDSPAN dataset contains details of major COPD exacerbations only (through the linked hospitalisation data which were described and analysed in Chapter 5). Therefore whilst the MIDSPAN dataset could be used to derive the rate of major exacerbations from a number of explanatory variables, because data on minor and major exacerbations is available in TORCH, TORCH was considered to be the most suitable dataset in which to develop the predication equations for the rate of minor and the rate of major exacerbations. Furthermore in TORCH, data on lung function and exacerbations were repeatedly collected over the three year study and hospitalisation was carefully followed. This contrasts with the MIDSPAN dataset in which measurement of lung function was made at baseline and information on exacerbations came in the form of COPD recorded hospitalisation records that in Chapter 5 were assumed to represent major exacerbations. Therefore as lung function is, *a priori*, expected to influence the rate of exacerbations, it is important to determine this relationship as accurately as possible, which is more likely within the TORCH dataset.

With reference to the conceptual model, it was assumed that the rate of exacerbations was related to lung function, which in turn was expected to affect symptoms, QALYs (via EQ-5D utility) and cost. Exacerbations are incorporated into the model as an annual rate.

Methods

The TORCH dataset was used to derive the prediction equations for rates of mild and rates of major exacerbations. Patients were dropped from the TORCH dataset where follow up time was less than 3 months in order to avoid biases introduced due to situations in which an event was observed over a short follow up period and then the patient left the study. These subjects would contribute a high event rate even though, if they had full follow up, it would be unlikely that the actual number of events would be close to the event rate estimated.

Because the exacerbation data in the TORCH study are count data, the Poisson distribution was employed. A goodness of fit test was used to check for over-dispersion (where the sample variance is greater than the sample mean) in order to decide whether a Poisson regression model or a negative binomial model was appropriate for modelling the data.

In developing the regression model for rate of major exacerbations, an assumption was made that previous exacerbations have no impact on future exacerbations. Explanatory variables included: age, FEV_1 % predicted, ex-smoker, sex and a UK variable. FEV_1 % predicted was calculated for each observation within the TORCH dataset, using the prediction equation developed within the MIDSPAN population (section 5.2.2) the rationale for use within the MIDSPAN study has been previously described, but use within the TORCH study was done on the grounds of consistency.

The resulting equations for major and for minor exacerbations were incorporated into the economic model as previously described in figure 7.4, so that as age and FEV_1 % predicted change over time, the rate of exacerbations also change over time, dependent upon the user defined patient characteristics.

Results

Major Exacerbations

Testing was carried out to assess the appropriateness of the Poisson model to model major exacerbations. It was found that the probability these data would be observed conditional on alpha being 0 (assumption for Poisson model) is close to zero with a $\chi^2 = 193.87$, and an alpha value within the equation of 3.4, which suggests that the negative binomial model is more appropriate for modelling this type of data, than the Poisson model.

The coefficients and standard errors from the negative binomial model are reproduced below in table 7.4. The variables ex-smoker and sex were not statistically significant in predicting the rate of major exacerbations and were therefore excluded from the analyses. As seen in the last column of table 7.4, the explanatory variables: age, breathless, FEV_1 % predicted and UK, were statistically significant and formed the prediction equation for rate of major exacerbations.

Table 7.4 The prediction equation for rate of major exacerbations

Variable name	Coefficients	SE	P
Age	0.023	0.006	0.000
Breathless	0.599	0.109	0.000
FEV % predicted	-1.524	0.391	0.000
UK	0.621	0.237	0.009
Constant	-5.286	0.448	0.000
Time (offset)			

Offset=log(time). Cycle lengths = 1yr. Therefore the offset term is one. FEV_1= Forced Expiratory Volume in one second.

This prediction equation was entered into the economic model using the algorithm below which gives the mean rate of exacerbation from the negative binomial model based on the explanatory variables and coefficients described in table 7.4. The rate of exacerbations from the negative binomial model is equal to the exponential of the linear equation plus the offset term. As the cycle lengths are one year, the value for the offset term is one.

$$Rma = e^{(cons + age*\beta age + breath*\beta breath + predFEV*\beta predfev + UK*\beta UK + offset)}$$

Minor Exacerbations

The prediction equation for minor exacerbations was developed in the same way and using the same dataset as the prediction equation for major exacerbations. In this equation, age was found to be statistically insignificant in the prediction of minor exacerbations and so was excluded. The remaining explanatory variables: breathless, FEV_1% predicted and UK were seen to be statistically significant (p<0.001) as shown within table 7.5 below.

Table 7.5 The prediction equation for rate of minor exacerbations

Variable name	Coefficient	SE	P
Breathless	0.451	0.062	0.000
FEV_1 % predicted	-1.552	0.220	0.000
UK	0.880	0.143	0.000
Constant	-1.983	0.121	0.000
Time (offset)			

FEV_1= Forced Expiratory Volume in one second.

The prediction equation from table 7.5 was incorporated into the economic model using the formula described below for the negative binomial model:

$$Rmi = e^{(cons + breath*\beta breath + predFEV*\beta predfev + UK*\beta UK + offset)}$$

Mild and moderate exacerbation rates were incorporated into the model. As an example, the predicted rates for men aged 55 years with mild or moderate NICE diagnosed COPD at baseline are shown in table 7.6, for the first six years of the model.

Table 7.6 An example of rate of major and minor exacerbations in men aged 55 with mild and moderate NICE diagnosed COPD at baseline

	Mild COPD		Moderate COPD	
Age	Minor	Major	Minor	Major
55	0.51	0.06	0.68	0.08
56	0.51	0.06	0.69	0.08
57	0.51	0.06	0.69	0.09
58	0.51	0.06	0.69	0.09
59	0.51	0.07	0.69	0.09
60	0.51	0.07	0.70	0.09

Tables 7.6 show that exacerbation rates generally increase over time and that minor exacerbations are more frequent than major exacerbations, irrespective of disease severity. For mild COPD, the rate of major exacerbations increased from 0.06 at baseline to 0.07 at 60 years. In the moderate COPD group, the rate of major exacerbations was higher than for the mild group at 0.08 per year at age 55 and increasing to 0.09 at age 60. Rate of minor exacerbations stayed constant over the first six years in the mild cohort, at 0.51 per year whereas the minor exacerbation rate for those with moderate COPD at baseline increased from 0.68 at 55 years to 0.70 at 60 years.

7.2.4 Symptoms

A prediction model for symptoms was developed within this section and the methods and results are presented.

Measuring symptoms and incorporating this measure in a meaningful way into an economic model is not without difficulty. One commonly used measure for breathlessness is the MRC dyspnea questionnaire which is shown in figure 7.5. Changes in breathlessness occur over time as lung function deteriorates and the disease progresses. The differences between the five levels of the MRC questionnaire are large and realistically anything other than a successful surgical intervention (or perhaps significantly increased exercise capacity) is unlikely to impact the patient so considerably that there would be an upward movement

to a less impaired level. The MRC questionnaire is limited to breathlessness and is to all intents and purposes, a way of classifying the degree of (im)mobility in a patient. Respiratory symptoms can however include sputum production, cough and wheeze, in addition to breathlessness.

1	I only get breathless with strenuous exercise
2	I get short of breath when hurrying on the level or walking up a slight hill
3	I walk slower than people of the same age on the level because of breathlessness, or I have to stop for breath when walking at my own pace on the level
4	I stop for breath after walking about 100 yards or after a few minutes on the level
5	I am too breathless to leave the house or I am breathless when dressing or undressing

Figure 7.5 MRC dyspnea scale

A frequently used method for measuring symptoms is to ask the respondent to answer a questionnaire. An example is the self-reported questionnaire within the Renfrew/Paisley study which contained specific questions on symptoms and which are stated in figure 7.6 below, where responses are either yes or no (the full questionnaire is reproduced in full within the appendix). As with the MRC scale, the improvement in symptoms needed to, for example, move from answering 'yes' to the question, 'do you get short of breath with people your own age on level ground?', to then answer 'no' following treatment, is unlikely to occur because it is a very large jump. Instead a more sensitive tool is required for the purposes of modelling.

Presence of a self-reported respiratory symptom was considered if the participant answered positively to one or more of:

"Do you get short of breath walking with people of your own age on level ground?"
OR
"Does your chest sound wheezy or whistling on most days (or nights)?"
OR
"Do you usually bring up any phlegm from your chest first thing in the morning in the winter?" plus, yes to either: "do you bring up phlegm like this on most days for as much as three months in the winter each year?" or "in the past three years have you had a period of increased cough and phlegm lasting for three weeks or more?"

Figure 7.6 Symptoms questions within the Renfrew/Paisley study

As seen in Chapter 2, the SGRQ is routinely collected within COPD RCTs and has three domains of symptoms, activity and impact which are used to give a total score (see Chapter 6 for further details). Questions 1 to 8 of the SGRQ (as shown in the appendix), ask about symptoms experienced over the past year and form the symptoms domain of the SGRQ. Therefore the topics covered within the symptoms part of the questionnaire are likely to be sensitive to changes in symptoms experienced and fits in with the time frame of the model. As such, the symptoms domain of the SGRQ is a feasible option for measuring changes in the symptoms of COPD.

As described within section 7.1, there is likely to be a link between exacerbations and symptoms because of the definition of an exacerbation, which is a worsening of respiratory symptoms requiring treatment. The respiratory symptoms that occur as a result of an exacerbation are acute, rather than chronic, which is what this section is primarily referring to. Nevertheless, because of the nature of the questions asked within the SGRQ, the measure is likely to be sensitive to both types. In order to account for this possible relationship, a variable for exacerbation is included within the regression equation on symptoms.

Methods

The TORCH dataset was used to develop the prediction equation for symptoms, using the same population used in the previous section on exacerbations. The SGRQ symptoms domain was used as a surrogate measure for measuring symptoms experienced by COPD patients, and was the dependent variable in the regression equation. The prediction equation for SGRQ symptoms was adjusted for a range of explanatory variables including: sex, age, ex-smoker and breathless at baseline. Because symptoms were *a priori*, expected to be influenced over time by exacerbations and by lung function, variables for major exacerbation rate and FEV_1 % predicted were also included within the model. GLM regression models were used as previously described in Chapter 6, in order to identify a model that fit the data well.

The resulting prediction equation on symptoms was employed within the economic model. For each cycle, the values from the FEV_1 % predicted and major exacerbation prediction equations were used as explanatory variables, in addition to values for baseline user defined patient characteristics.

Results

The SGRQ symptoms scores were seen to approximately fit a normal distribution as shown in figure 7.7, with a mean (SD) of 57 (21), a minimum value of 0 and a maximum value of 100, where 0 represents no symptoms.
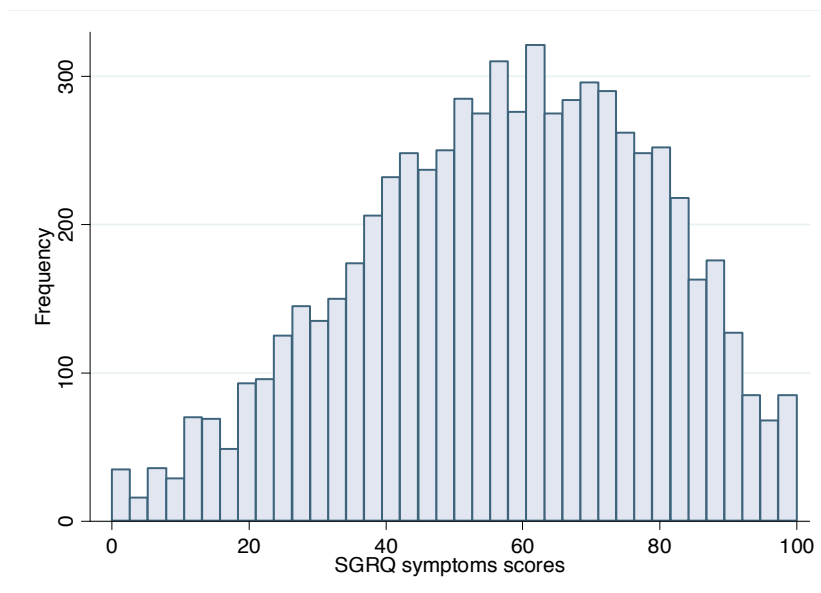
Figure 7.7 Frequency of SGRQ symptoms scores in the TORCH study

Given the distribution of the data, a GLM model with Gaussian family and identity link (OLS) was found to fit the data well. The prediction equation from the regression is shown in table 7.7 and shows that major exacerbations, sex, FEV_1 % predicted, being an ex-smoker, based in the UK and having breathlessness at baseline were all statistically significant in the prediction of SGRQ symptoms score.

Table 7.7 shows that: a major exacerbation increases the symptoms score by 6.1 points; a 1% increase in FEV_1% predicted lowers the SGRQ total score by -0.21; females have a higher symptoms score by on average 2.4; being an ex-smoker compared to a smoker lowers the symptoms score by 7.9 and a person with breathlessness has a higher SGRQ score by 7.4, compared to somebody without breathlessness. Age was assessed for inclusion within the prediction equation; however, whilst statistically significant, it was removed from the analyses on the grounds that increasing age was found to be associated

with better symptoms, rather than worse, as was originally expected. This relationship with age is explained by changing expectations of health as age increases, in that people expect their health to be worse as they get older and adjust their expectations accordingly, with a greater tolerance of their ailments. There is no evidence to suggest that symptoms per se improve with age.

Table 7.7 Prediction equation for symptoms

Variable name	Coefficients	SE	p
Major exac	6.100	0.788	0.000
Sex	2.382	0.686	0.001
FEV_1 % predicted	-0.213	0.023	0.000
Ex-smoker	-7.041	0.613	0.000
UK	7.870	1.554	0.000
Breathless	7.365	0.625	0.000
Constant	68.435	1.246	0.000

FEV_1= Forced Expiratory Volume in one second.

The prediction equation for symptoms was included within the economic model as described earlier. Predicted symptoms scores from the model for a male cohort aged 55 years with mild and moderate NICE diagnosed COPD at baseline are shown in table 7.8 below.

Table 7.8 Symptoms scores for men aged 55
with mild and moderate NICE diagnosed COPD at baseline

Age	Mild	Moderate
55	62.9	67.1
56	62.9	67.2
57	62.9	67.2
58	63.0	67.3
59	63.0	67.4
60	63.0	67.5

Table 7.8 shows that symptoms marginally worsen over time as the disease progresses. The cohort with moderate COPD has higher symptoms scores than those with mild disease, indicating worse respiratory symptoms. With reference to table 7.8, symptoms scores were 62.9 in the mild group at baseline, rising to 63.0 by 60 years. The corresponding symptoms

scores in the moderate group were 67.1 rising to 67.5, indicating that symptoms worsen at a faster rate in people with more severe disease than those with less severe disease.

7.2.5 EQ-5D Utility

Within the section, an equation that predicts expected EQ-5D utility, given: lung function, exacerbations and symptoms, is developed and the results presented.

Two prediction equations for EQ-5D utility were developed in Chapter 6. The first used explanatory variables of: height, age, sex, over 10 pack years, ex-smoker and disease severity, and the second used explanatory variables of sex and the SGRQ (total, domain or item) scores. In these equations it was seen that the SGRQ symptoms domain and disease severity were statistically significant in the prediction of utility. The prediction models presented in Chapter 6 form the starting point for this equation, with two key differences, which came to light following the conceptualisation of the model: exacerbations were included as an explanatory variable within the prediction equation, and rather than the SGRQ total score being an explanatory variable, the SGRQ symptoms domain alone was used. The TORCH dataset was used to derive the prediction equation for EQ-5D utility as no HRQoL data were collected in the Renfrew/Paisley (MIDSPAN) study.

Methods

The TORCH dataset was used to develop the prediction equation for EQ-5D utility. EQ-5D utility at one year time periods were used together with information on events and experiences occurring within the previous year to predict EQ-5D utility. Explanatory variables assessed for incorporating into the EQ-5D utility prediction model included: SGRQ symptoms; FEV_1 % predicted; breathless at baseline, sex, UK, ex-smoker, age and major exacerbations within the last year (where 0=no, 1=yes). The prediction equation for

EQ-5D utility was developed using GLM models, using the same methods as previously applied within Chapter 6 (the equations did not adjust for multiple observations per participant for reasons previously described in Chapter 6). The equation was estimated on the utility decrement scale (1-EQ-5D utility), in order to allow greater flexibility in identifying the appropriate family for the GLM model, and converted back for reporting.

The resulting EQ-5D utility equation was applied within the economic model using the predicted values from the regressions on lung function, exacerbations and on symptoms (where statistically significant) plus user defined patient characteristics in order to predict EQ-5D scores for every cycle.

Results

The GLM prediction equation for EQ-5D was fit with a Gaussian family and identity link. Table 7.9 shows the resulting model. Major exacerbations within the last year, FEV_1 % predicted, SGRQ symptoms scores, breathlessness at baseline, UK and sex were all found to be statistically significant in the prediction of EQ-5D utility. Age and being an ex-smoker at baseline were not found to be statistically significant and were excluded from the regression model.

Table 7.9 Prediction equation for EQ-5D decrement score

Variable name	Coefficients	SE	P
Major	0.053	0.010	0.000
FEV_1 % predicted	-0.056	0.020	0.006
SGRQ Symptoms	0.004	0.000	0.000
UK	0.049	0.017	0.004
Breathless	0.087	0.006	0.000
Sex	0.031	0.006	0.000
Constant	-0.003	0.013	0.829

FEV_1= Forced Expiratory Volume in one second. SGRQ= St George's respiratory questionnaire.

Because the analysis was conducted using EQ-5D decrement, in order to generate EQ-5D scores, the predicted score was subtracted from 1. The table shows that for each major exacerbation, EQ-5D utility decreases by 0.053, people with respiratory symptoms at baseline have worse EQ-5D utility and that women tend to have worse utility scores than men.

Using this equation, EQ-5D utilities were predicted within the economic model for men aged 55 years with mild and moderate COPD at baseline and the results are shown in table 7.20. The model shows that over this short time frame, EQ-5D utility is relatively stable. As seen in the table, moderate COPD was associated with lower EQ-5D utility than mild COPD, with a utility value of 0.61 in the moderate group compared to 0.64 in the mild group.

Table 7.10 EQ-5D utility for men aged 55
with mild and moderate NICE diagnosed COPD at baseline

Age	Mild	Moderate
55	0.64	0.61
56	0.64	0.61
57	0.64	0.61
58	0.64	0.61
59	0.64	0.61
60	0.64	0.61

7.2.6 Cost

The economic model developed in this section represents current treatment practice and it is assumed that there are costs accruing to the NHS as a result of the disease. Costs can be split into two types: treatment costs and all 'other costs', which includes the cost of hospitalisation, GP contacts and costs associated with adverse events. In this section, both types are considered but are dealt with in different ways.

As previously described in Chapter 2, current practice is to treat COPD based on disease severity with therapies added as the disease worsens. In order to use a relevant value for treatment costs, the treatment pathway as described in table 2.5 and treatment costs as estimated in table 2.6 (Chapter 2) were used to derive costs for a person with mild and moderate COPD before the 'new' COPD therapies entered the market (Seretide, Spiriva and Symbicort). It was assumed that a person with mild (NICE diagnosed) COPD would be prescribed ipratropium and salbutamol at an annual cost of £102 ($167) and for treating people with moderate COPD, that beclomethasone would been added to the treatment mix, so to include beclomethasone, ipratropium and salbutamol with a one year cost of treating this group of £183 ($299).

As pharmacological costs are fixed, given the quantity prescribed, they are entered into the model as a user defined value that occurs every cycle and does not have any uncertainty surrounding it. If instead incremental treatment costs were of interest as opposed to specific costs, the model allows for the treatment costs in the base case to be left empty (entered as zero into the model) and the incremental treatment cost to be added to the treatment costs within the new treatment arm of the economic model.

A different option for estimating treatment costs is to use the TORCH dataset to derive a prediction equation. However, this method was not pursued because the aim of this new economic model is to produce a general model for COPD, rather than looking at the cost effectiveness of any drug treatment from TORCH. Any new treatments that come onto the market for COPD would be priced differently to treatments in TORCH and so this approach would lack external generalisability.

For 'other costs', summary statistics were derived by type of exacerbation, before a regression equation was developed to predict cost based on patient events and experiences. The cost equation for 'other costs' was developed within the TORCH dataset because the

study collected detailed information on all treatment and disease related costs over six monthly time frames and the Renfrew/Paisley (MIDSPAN) dataset did not contain any cost data. The TORCH dataset measured costs in American dollars.

The cost equation was estimated using GLM regression methods as used within this chapter and previously described and applied in Chapter 6. 'Other costs' was the dependent variable and explanatory variables offered to the model included: major exacerbations, minor exacerbations, SGRQ symptoms, sex, UK, FEV % predicted, breathless at baseline, age and ex-smoker.

Results

'Other costs' were found to be higher in people who had had an exacerbation compared to those people who hadn't had an exacerbation, with costs highest in people who had experienced a major exacerbation. In people who had had a major and a minor exacerbation in the previous year, the mean cost was £1976 (SD=3498), where there was a major but no minor exacerbation, cost was £1929 (2670), where there was a minor and no major exacerbation cost was £339 (814). This compares to where there was neither a minor nor a major exacerbation, when 'other cost' was found to be £97 (388). The large standard deviations show the extent to which there was variability in other costs within the dataset.

The Gaussian family and identity link were found to fit the data well. Details of the resulting regression coefficients are presented in table 7.11 and the table shows that major exacerbations, minor exacerbations, age, SGRQ symptoms score, FEV % predicted were all found to be statistically significant in the prediction of 'other costs'. From the table it can be seen that a major exacerbation is estimated to cost $2498 and a minor exacerbation, $330. On average, increasing age was associated with an increasing cost burden of $5 year on year. For every one point increase in symptoms score (associated with deterioration in

health) an extra \$1.50 is added to the cost burden. A 1% increase in FEV_1% predicted is seen to be associated with a \$267 reduction in 'other costs'.

Table 7.11 Prediction equation for 'other costs'

Variable name	Coefficient	SE	P
Major exac	2497.836	178.266	0.000
Minor exac	330.223	36.7009	0.000
Age	5.376	1.2448	0.000
SGRQ symptoms	1.480	0.672	0.028
FEV_1 % predicted	-267.176	90.409	0.003
Constant	-150.835	100.960	0.135

The application of the prediction equation for 'other costs' is shown in table 7.12 for the male cohort aged 55 years at baseline with mild or moderate COPD. Mild and moderate treatment costs are \$167 and \$299 respectively and are assumed to be constant over time as illustrated in table 7.12.

Table 7.12 Treatment and 'other costs' for men aged 55 years with mild and moderate NICE diagnosed COPD at baseline (\$)

Age	Mild COPD		Moderate COPD	
	Tx	Other	Tx	Other
55	167	379	299	545
56	167	388	299	558
57	167	398	299	571
58	167	408	299	584
59	167	418	299	597
60	167	428	299	611

Table 7.12 shows that 'other costs' increase from the mild to the moderate groups and that 'other costs' increase over time as the disease progresses. 'Other costs' start at \$379 at age 55 years in the mild group and \$545 in the moderate group, and rise over time to \$428 and \$611 for the mild and moderate group respectively by 60 years of age.

7.2.7 Survival

In Chapter 5 it was shown that those meeting NICE diagnostic criteria, which includes people with respiratory symptoms, airflow obstruction and over 10 pack years, had a higher mortality risk than people with no COPD, or people with GOLD defined COPD. It was also seen that people with more severe disease had a higher mortality risk than those with less severe disease, for all cause and COPD mortality. In addition, age is a well known driver of mortality risk, so that as the cohort ages, the probability of survival in the next period, decreases. Within this section, a model that predicts survival probability, that incorporates all of these factors, is described and developed before being applied within the economic model for COPD.

A traditional way of incorporating mortality into an economic model is to use life tables. However because within Chapter 5 more factors than age alone have been shown to affect mortality risk, the use of life tables would be inappropriate within this context as only information on mortality rates for the general population, by age are available within lifetables. A regression model predicting survival given the presence of respiratory symptoms, over 10 pack years, COPD disease severity and age would be more suitable for representing mortality in a COPD population given the earlier findings of this thesis.

The Renfrew/Paisley (MIDSPAN) dataset was used in Chapter 5 in order to investigate mortality and for the same reason, the MIDSPAN dataset is used here in preference to the TORCH dataset. Within the dataset, only those meeting the GOLD COPD criteria were used, for reasons explained earlier in section 7.2.1. As seen in Chapter 5 there are a range of causes of mortality for COPD patients and cause of mortality is not limited to COPD mortality. Therefore, the survival model was developed using data on all-cause mortality from the MIDSPAN study, rather than on COPD specific mortality.

Methods

The GOLD COPD cases of the MIDSPAN dataset were used to develop the regression model on survival. Explanatory variables included sex, FEV % predicted at baseline, age at baseline and dummy variables for: fewer than/over 10 pack years, ex-smoker and respiratory symptoms. The survival function was modelled using a Weibull regression where the survival function is given by:

$$S(t) = EXP\left(-\lambda t^{\gamma}\right)$$

The resulting equation gives the gamma variable within the output (γ) and $\ln \lambda$ is the linear predictor of covariates, such that:

$$\ln \lambda = \alpha + \sum \beta_j X_{ij}$$

The survival function was used within the model to predict the probability of survival at each cycle using user defined patient characteristics at baseline, with the exception of FEV_1 % predicted score which was obtained from the prediction of FEV_1 % predicted from the model in the first time period.

Results

The explanatory variables from the Weibull model are shown in table 7.13 together with the coefficients, standard errors and p values for each. Supporting the previous findings on the importance of these factors on mortality risk within Chapter 5, respiratory symptoms, a smoking history and FEV_1 % predicted, along with age, sex and ex-smoker were all found to be statistically significant in the prediction of survival ($p<0.001$).

Table 7.13 The prediction equation for survival

Variable	Coefficient	SE	P
Constant	-9.700	0.241	0.000
Sex	-0.362	0.039	0.000
FEV$_1$ % pred	-0.008	0.001	0.000
Age	0.078	0.003	0.000
Over10	0.455	0.047	0.000
Ex-smoker	-0.276	0.053	0.000
Symptoms	0.263	0.038	0.000
Gamma	1.827	0.028	0.000

The survival curves plotted in figure 7.8 show how all-cause mortality within the male cohort of the MIDSPAN study is affected by COPD disease severity at baseline and shows real data. This figure is a replication of figure 5.4 in Chapter 5 with the exception that figure 7.8 presents survival probability by disease severity with time in study along the x axis as opposed to age in study. Figure 7.9 demonstrates how the survival function described above predicts for the mean male population for 55 year old men (mean age in MIDPSAN study) and different disease severities.
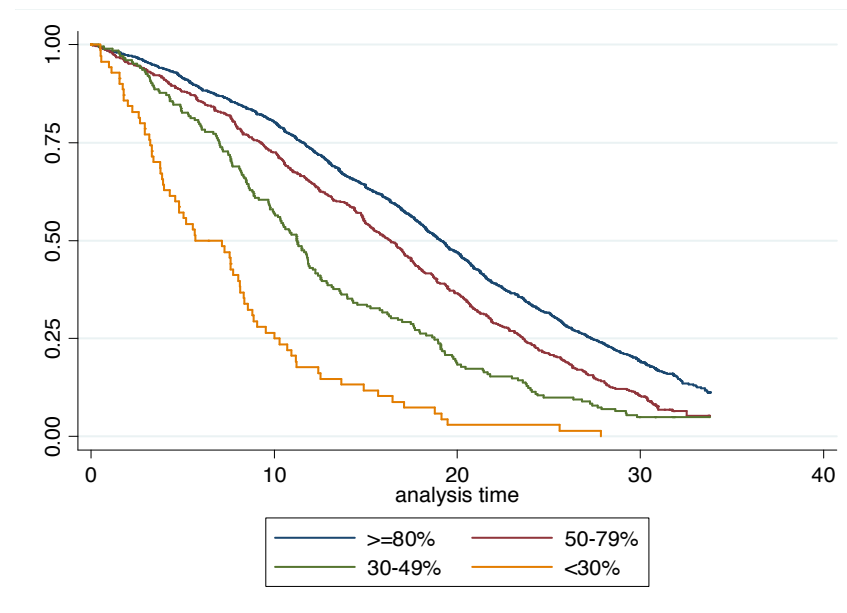
Figure 7.8 Survival curves for all cause mortality for men by COPD disease severity within the MIDSPAN dataset
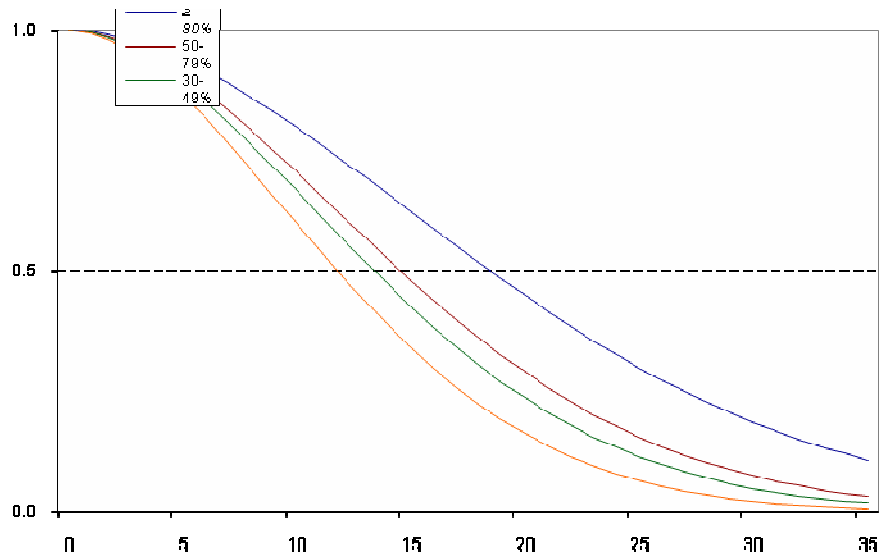
Figure 7.9 An example of the modelled survival function applied to 55 yr old males, FEV$_1$ ≥ 80% predicted assumed a smoking history and no symptoms and the other lung function groups were assumed to have a smoking history and respiratory symptoms

The top line in each figure illustrates the survival probability of the cohort with FEV$_1$ ≥80% predicted with a smoking history of greater than 10 pack years. The other lines restrict the population to those with respiratory symptoms. The red line corresponds to the NICE mild group, the green line to moderate COPD and the orange line to severe disease. The modelled curves in figure 7.9 are smooth over time in comparison to the jagged lines seen in the observed data in figure 7.8.

The figures capture different patient groups, the observed data represent all men with GOLD defined COPD, whereas the modelled curves represent survival of men aged 55 years at the start of follow, as such the curves are expected to be slightly different. Nevertheless, the two groups can be roughly compared and it can be seen that the survival probabilities are similar between the observed male patients and the average modelled male cohort, particularly in the least severe cohort where median survival is approximately 20

years. The biggest difference between the groups is seen in the severe COPD group, and the spread of the curves is less in the modelled than in the observed curves.

Table 7.14 Survival probability for men aged 55 years
with mild and moderate NICE diagnosed COPD at baseline

Age	S(t) mild	S(t) mod
55	1.00	1.00
56	0.99	0.98
57	0.97	0.97
58	0.95	0.94
59	0.93	0.92
60	0.90	0.89

The mild and moderate disease populations plotted in figure 7.9 match those in table 7.14. The survival probabilities in table 7.14 show that probability of survival decreases with disease severity, so that moderate COPD patients are at higher risk of all-cause mortality than mild COPD patients. At age 60, survival probabilities for men in the mild and moderate groups are 0.90 and 0.89 respectively.

7.2.8 Survival Adjusted EQ-5D Utility and Cost

Survival adjusted EQ-5D utility and total costs were estimated using the Kaplan-Meier sample average estimator.(156) For each interval (one year within the model) the estimator calculates the mean cost and the mean EQ-5D utility for patients alive at the beginning of the interval, weighted by the probability of surviving to the beginning of the interval and sums these values over all the time intervals.(130) Survival adjusted total cost was calculated for each time period by multiplying the cost by the survival function, and EQ-5D utility was calculated in the same way.

7.2.9 The Economic Model for Current Treatment

Snapshots of the deterministic economic model representing current treatment for 55 year old men with mild NICE diagnosed COPD and moderate COPD at baseline have been presented above for individual sections of the model. Within tables 7.15 and 7.16 below, the model for these two groups is presented in full over a lifetime perspective with specified patient characteristics including: ex-smoker, over 10 pack years, presence of respiratory symptoms and based in the UK. The model combines the individual regression equations developed earlier, to predict values for lung function, exacerbations and symptoms and from these to predict costs and utility, which are weighted by survival and are discounted.

From tables 7.15 and 7.16 it can be seen that as age increases, FEV_1 decreases year on year in both cohorts, as does predicted FEV_1 and FEV_1 % predicted and lung function is more impaired in the moderate than in the mild group. Whilst the incremental change in FEV_1 is the same for the two cohorts, FEV_1 % predicted deteriorates quicker in the moderate cohort than in the mild cohort. Symptoms are seen to worsen over time with a deterioration that is quicker in the moderate group compared to the mild COPD group. The rate of minor exacerbations was comparatively steady over time for the mild cohort, whereas in the moderate group, the rate of minor exacerbations increased over time. The rate of major exacerbations in the mild and moderate cohort more than doubled. EQ-5D utility was almost constant over time for the modelled mild cohort, whereas in the moderate group, EQ-5D utility at baseline was lower and worsened at a quicker rate than in the mild group. Treatment costs are assumed to be constant over time. 'Other costs' are seen to increase over time and are higher in the moderate cohort than the mild. Total costs are equal to the sum of treatment cost and 'other costs'. The survival probability is higher for the mild group than the moderate group.

Table 7.15 Economic model for current treatment, men aged 55 yrs with mild COPD

Time	Age	FEV	pred FEV1	FEV1% pred	Symp	Exac minor	Exac major	EQ-5D utilitys	Tmt cost	Other cost	Total cost	S(t)	S(t)* utility	S(t)*total cost
1	55	2.00	3.02	0.66	62.9	0.51	0.06	0.64	167	379	546	1.00	0.62	525
2	56	1.98	2.99	0.66	62.9	0.51	0.06	0.64	167	388	555	0.99	0.59	511
3	57	1.96	2.96	0.66	62.9	0.51	0.06	0.64	167	398	565	0.97	0.56	495
4	58	1.93	2.93	0.66	63.0	0.51	0.06	0.64	167	408	575	0.95	0.53	477
5	59	1.91	2.91	0.66	63.0	0.51	0.07	0.64	167	418	585	0.93	0.50	457
6	60	1.89	2.88	0.66	63.0	0.51	0.07	0.64	167	428	595	0.90	0.47	436
7	61	1.87	2.85	0.66	63.1	0.51	0.07	0.64	167	438	605	0.87	0.44	414
8	62	1.85	2.82	0.66	63.1	0.51	0.07	0.64	167	448	615	0.84	0.41	392
9	63	1.83	2.79	0.65	63.1	0.51	0.07	0.64	167	459	626	0.80	0.38	369
10	64	1.80	2.76	0.65	63.2	0.51	0.07	0.64	167	469	636	0.77	0.35	347
11	65	1.78	2.73	0.65	63.2	0.51	0.08	0.64	167	480	647	0.73	0.32	324
12	66	1.76	2.70	0.65	63.2	0.52	0.08	0.64	167	490	657	0.69	0.29	301
13	67	1.74	2.67	0.65	63.3	0.52	0.08	0.64	167	501	668	0.65	0.27	279
14	68	1.72	2.64	0.65	63.3	0.52	0.08	0.64	167	512	679	0.61	0.24	258
15	69	1.70	2.62	0.65	63.3	0.52	0.08	0.64	167	524	691	0.58	0.22	237
16	70	1.67	2.59	0.65	63.4	0.52	0.09	0.64	167	535	702	0.54	0.20	217
17	71	1.65	2.56	0.65	63.4	0.52	0.09	0.64	167	546	713	0.50	0.18	199
18	72	1.63	2.53	0.64	63.5	0.52	0.09	0.64	167	558	725	0.46	0.16	181
19	73	1.61	2.50	0.64	63.5	0.52	0.09	0.63	167	570	737	0.43	0.14	164
20	74	1.59	2.47	0.64	63.5	0.52	0.10	0.63	167	582	749	0.39	0.13	148
21	75	1.56	2.44	0.64	63.6	0.52	0.10	0.63	167	594	761	0.36	0.11	133
22	76	1.54	2.41	0.64	63.6	0.52	0.10	0.63	167	607	774	0.33	0.10	119
23	77	1.52	2.38	0.64	63.7	0.53	0.10	0.63	167	619	786	0.30	0.09	107
24	78	1.50	2.35	0.64	63.7	0.53	0.11	0.63	167	632	799	0.27	0.08	95
25	79	1.48	2.33	0.64	63.8	0.53	0.11	0.63	167	645	812	0.25	0.07	84
26	80	1.46	2.30	0.63	63.8	0.53	0.11	0.63	167	658	825	0.22	0.06	75
27	81	1.43	2.27	0.63	63.9	0.53	0.11	0.63	167	671	838	0.20	0.05	66
28	82	1.41	2.24	0.63	63.9	0.53	0.12	0.63	167	685	852	0.18	0.04	58
29	83	1.39	2.21	0.63	64.0	0.53	0.12	0.63	167	699	866	0.16	0.04	51
30	84	1.37	2.18	0.63	64.0	0.53	0.12	0.63	167	713	880	0.14	0.03	44
31	85	1.35	2.15	0.63	64.1	0.54	0.13	0.63	167	727	894	0.12	0.03	38
32	86	1.33	2.12	0.62	64.1	0.54	0.13	0.63	167	742	909	0.11	0.02	33
33	87	1.30	2.09	0.62	64.2	0.54	0.13	0.63	167	757	924	0.10	0.02	29
34	88	1.28	2.06	0.62	64.2	0.54	0.14	0.63	167	772	939	0.09	0.02	25
35	89	1.26	2.04	0.62	64.3	0.54	0.14	0.63	167	787	954	0.07	0.01	21
													7.72	**7710**

Table 7.16 Economic model for current treatment, men aged 55 yrs with moderate COPD

Time	Age	FEV	pred FEV1	FEV1% pred	Symp	Exac minor	Exac major	EQ-5D utility	Tmt cost	Other cost	Total cost	S(t)	S(t)* utility	S(t)*total cost
1	55	1.42	3.02	0.47	67.1	0.68	0.08	0.61	299	545	844	1.00	0.59	812
2	56	1.40	2.99	0.47	67.2	0.69	0.08	0.61	299	558	857	0.98	0.56	787
3	57	1.38	2.96	0.46	67.2	0.69	0.09	0.61	299	571	870	0.97	0.53	758
4	58	1.36	2.93	0.46	67.3	0.69	0.09	0.61	299	584	883	0.94	0.50	726
5	59	1.33	2.91	0.46	67.4	0.69	0.09	0.61	299	597	896	0.92	0.47	692
6	60	1.31	2.88	0.46	67.5	0.70	0.09	0.61	299	611	910	0.89	0.44	655
7	61	1.29	2.85	0.45	67.6	0.70	0.09	0.61	299	625	924	0.85	0.41	618
8	62	1.27	2.82	0.45	67.6	0.70	0.10	0.61	299	639	938	0.81	0.37	580
9	63	1.25	2.79	0.45	67.7	0.71	0.10	0.61	299	653	952	0.77	0.34	541
10	64	1.23	2.76	0.44	67.8	0.71	0.10	0.61	299	668	967	0.73	0.31	503
11	65	1.20	2.73	0.44	67.9	0.71	0.11	0.60	299	683	982	0.69	0.29	465
12	66	1.18	2.70	0.44	68.0	0.72	0.11	0.60	299	698	997	0.65	0.26	428
13	67	1.16	2.67	0.43	68.1	0.72	0.11	0.60	299	713	1012	0.61	0.23	392
14	68	1.14	2.64	0.43	68.2	0.73	0.12	0.60	299	729	1028	0.56	0.21	358
15	69	1.12	2.62	0.43	68.3	0.73	0.12	0.60	299	745	1044	0.52	0.19	325
16	70	1.09	2.59	0.42	68.4	0.73	0.12	0.60	299	761	1060	0.48	0.17	294
17	71	1.07	2.56	0.42	68.5	0.74	0.13	0.60	299	778	1077	0.44	0.15	265
18	72	1.05	2.53	0.42	68.6	0.74	0.13	0.60	299	795	1094	0.40	0.13	238
19	73	1.03	2.50	0.41	68.7	0.75	0.13	0.60	299	813	1112	0.37	0.11	213
20	74	1.01	2.47	0.41	68.8	0.75	0.14	0.60	299	831	1130	0.33	0.10	189
21	75	0.99	2.44	0.40	68.9	0.76	0.14	0.60	299	849	1148	0.30	0.09	168
22	76	0.96	2.41	0.40	69.0	0.76	0.15	0.60	299	868	1167	0.27	0.08	148
23	77	0.94	2.38	0.40	69.1	0.77	0.15	0.59	299	887	1186	0.24	0.07	130
24	78	0.92	2.35	0.39	69.2	0.77	0.15	0.59	299	907	1206	0.22	0.06	114
25	79	0.90	2.33	0.39	69.4	0.78	0.16	0.59	299	927	1226	0.19	0.05	99
26	80	0.88	2.30	0.38	69.5	0.78	0.16	0.59	299	948	1247	0.17	0.04	86
27	81	0.86	2.27	0.38	69.6	0.79	0.17	0.59	299	969	1268	0.15	0.03	75
28	82	0.83	2.24	0.37	69.8	0.79	0.17	0.59	299	991	1290	0.13	0.03	65
29	83	0.81	2.21	0.37	69.9	0.80	0.18	0.59	299	1014	1313	0.11	0.02	55
30	84	0.79	2.18	0.36	70.0	0.81	0.19	0.59	299	1037	1336	0.10	0.02	47
31	85	0.77	2.15	0.36	70.2	0.81	0.19	0.59	299	1060	1359	0.09	0.02	40
32	86	0.75	2.12	0.35	70.3	0.82	0.20	0.58	299	1085	1384	0.07	0.01	34
33	87	0.72	2.09	0.35	70.5	0.83	0.20	0.58	299	1110	1409	0.06	0.01	29
34	88	0.70	2.06	0.34	70.7	0.83	0.21	0.58	299	1136	1435	0.06	0.01	25
35	89	0.68	2.04	0.33	70.8	0.84	0.22	0.58	299	1162	1461	0.05	0.01	21
													6.91	**10 977**

Costs and QALYs after costs and EQ-5D utilities have been discounted and survival adjusted, are shown in bold in the bottom right hand corner of table 7.15 and 7.16, and are: 7.72 QALYs and $7710 for the mild cohort and 6.91 QALYs and $10 977 for those with moderate COPD at baseline. As expected, mild disease is associated with more favourable health outcomes and lower cost than moderate COPD.

7.3 Model Development: Treatment Effects

In the previous section, a model representing the natural history of COPD was created and was assumed to represent the current treatment arm of the model. In order to produce a comparator arm for the model, treatment effects need to be incorporated. In this section, methods for building in a treatment effect are considered and an example of a treatment effect on lung function is applied to the model.

7.3.1 Introduction

Treatments have different courses of action that may impact on one or more of the components of COPD (lung function, exacerbations, symptoms) and potential effects of treatment on the components are described in this introduction. It is important to bear in mind that an effect of a treatment whose mechanism of action is to treat one component may have spill over effects onto the other components, so that a treatment that improves, for example, lung function may also reduce exacerbations and/or improve symptoms as described earlier in figures 7.1 and 7.2. Therefore when identifying treatment effects from data, it is useful to understand whether a treatment effect is independent, ie once controlling for all other components, a statistically significant effect on one component is seen, or whether it is a dependent effect.

Exacerbations

In this section, ways in which a treatment effect on exacerbations can be incorporated into the model are considered, with reference to the published literature from RCTs. The regression equations developed within section 7.2 which predict rates of mild and major exacerbations, given user defined patient characteristics, represent the rate of mild and major exacerbations over time, given current treatment. In line with what is known about the natural history of the disease (Chapter 5) and as was seen within tables: 7.6, 7.15 and 7.16, the rate of exacerbations is expected to increase over time. As illustrated earlier in figures 7.1 and 7.2, treatment for COPD can affect the rate of exacerbations. The effect of treatment on rate of exacerbations can take a number of forms including: 1) no effect, such that the rate of exacerbations is the same in the new treatment arm as in the current treatment arm and is represented in figure 7.10 by the navy line (with diamond markers). Alternatively treatment could; 2) reduce the rate of exacerbations such that the curve shifts downwards as illustrated by the pink line (square markers) in figure 7.10; 3) cause a change in the slope of the curve which represents a slowing down in the increasing rate of exacerbations over time as shown in the green line (triangle markers) or; 4) treatment may cause the rate of exacerbations to shift downwards and to change slope as represented by the pale blue curve (cross markers). In order to apply treatment effects to the model, the starting point is to look to the exacerbation rates from RCTs in order to inform the size of the treatment effect.
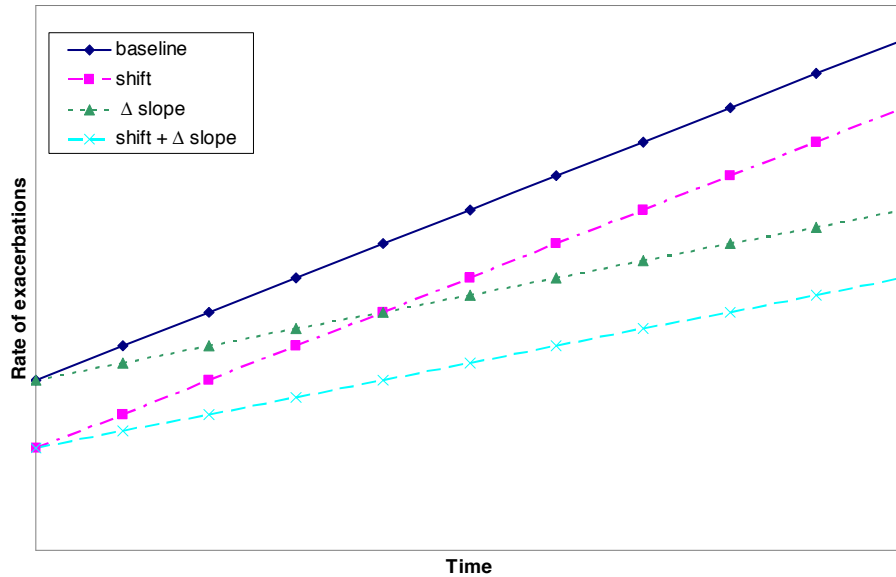
Figure 7.10 Effect of treatment on exacerbations over time

Table 7.17 reports on outcomes of the TORCH trial with respect to rates of exacerbations according to resource utilisation by treatment group. The placebo group were found to have a higher rate of unscheduled health care contacts than the other groups, particularly the combination therapy and rates for Accident and Emergency visits were almost double that of the combination therapy group (143 per 1000 person year vs 72 per 1000).

In a 2003 paper, describing outcomes of a RCT, Calverley et al reported that when treated with placebo, the mean rate of total exacerbations (minor plus major) per patient year was 1.30 compared to a rate of 0.97 for the treatment group who were prescribed the combination therapy of salmeterol and fluticasone.(50)

Table 7.17 Rates of healthcare contacts within the TORCH trial.

	Placebo (N=1524)	SAL 50 (N=1521)	FP 500 (N=1534)	SFC 50/500 (N=1533)
Unscheduled Healthcare contacts, n(%)	949 (62%)	918 (60%)	937 (61%)	896 (58%)
Rate/1000 years of exposure				
ER/A & E Visits	143	75	81	72
Out-patient clinic visits	245	196	197	209
GW admissions	200	172	195	180
ICU admissions	18	15	16	17
GW and/or ICU admissions	208	179	201	187
GP office/practice visits	593	450	603	426

ER = Emergency Room A & E = Accident and Emergency ICU = Intensive Care Unit GP = General Practitioner GW= General Ward
Table adapted from the TORCH study report.(244)

In a more recent paper (2009) Calverley et al describe exacerbation event rates within two one year clinical trials (M2-124 and M2-125) which compare roflumilast to placebo. Mean (95% CI) severe exacerbation rates in M2-124 were 0.11 (0.07-0.15) per patient year for treatment compared to 0.12 (0.09-0.16) for placebo. For the M2-125 trial these rates were 0.14 (0.10-0.20) and 0.18 (0.13-0.25) respectively. Moderate exacerbation rates per patient year were: 0.94 (0.83-1.06) and 1.11 (1.00-1.25) in the M2-124 trial and 1.04 (0.92-1.18) and 1.27 (1.13-1.42) in the M2-125 trial.(245)

To incorporate treatment effects on exacerbations, the relative difference in the rates between the treated population and a relevant comparator within the trial could be determined and then this effect applied onto exacerbations in the new treatment arm of the model.

Treatment effects may cause a permanent change (improvement) to the curve, which continues over time, or over time treatment effects may diminish such that the trajectory attenuates to the baseline curve, or even to a different trajectory. Scenario analysis could be conducted in order to test various possibilities for the effect of treatment over time on exacerbation rate, and should be informed by observed effects from studies. Scenario analysis can be used to examine the impact of different durations of effect on the cost

effectiveness statistic, for example by limiting the effect of treatment to the duration of the trial, such as one year.

Symptoms

Some treatments for COPD work explicitly on alleviating symptoms, for example LABAs relax the bronchial passageways and through this action, breathing becomes easier for the patient. As seen in tables 7.15 and 7.16, over time symptoms scores gradually worsen as the disease progresses (symptoms scores get closer to 100). In much the same way as has been described in figure 7.10, treatment for symptoms can have a number of effects: 1) there may be no effect on symptoms such that the symptoms scores remain unchanged between the current treatment population and the new treatment arm (as in the blue line in figure 7.10). Alternatively, 2) an immediate and constant improvement in symptoms following treatment could occur, which would be modelled by a downward shift in the curve for symptoms over time (such as is illustrated by the pink line), 3) the change in symptoms score over time such that the rate at which symptoms worsen slows down, as shown by the green line in figure 7.10, or 2) and 3) could both occur such that symptoms immediately improve and continue to improve over time relative to the non treated arm (pale blue line). Over time, the treatment curve may attenuate to the baseline curve for symptoms if the effectiveness of treatment is assumed to diminish with continued exposure. Or, the shifted curve could remain over time. Scenario analysis could be employed in order to investigate the impact of treatment duration on the outcomes of the model. SGRQ total score rather than SGRQ symptoms scores are generally reported in the clinical trial reports, as such these data are not as readily available as data on exacerbations and on lung function.

Mortality Risk

Little information is available on the effect of COPD treatment on mortality and it is an area of contention. The TORCH trial reported on mortality and found that the absolute risk reduction for death in the combination therapy group compared to the placebo group was 2.6% (12.6% vs 15.2%). The corresponding HR for all-cause mortality for the combination product compared to placebo (HR=1) was 0.825 (95% CI 0.681-1.002) and was fractionally non-significant. The corresponding reduction in the risk of mortality between the two treatments was found to be 17.5%. Secondary analysis using Cox proportional-hazards gave a significant HR of 0.811 (95% CI 0.670-0.982).(49)

Because of this uncertainty, treatments directly affecting survival per se are not incorporated into this model, nonetheless it would be relatively simple to incorporate a direct survival effect by applying a HR to the survival curve of the current treatment arm to elicit the survival curve for the new treatment arm, if this were of interest. The way this model is set up, an impact on survival could be experienced indirectly as a result of improved lung function, by multiplying the treatment adjusted FEV_1 % predicted, observed within the first year, with the coefficient on FEV_1 % predicted from the survival equation.

Lung Function

This section considers how treatment may affect lung function and how this could be incorporated into the model.

Effects of treatment on FEV_1 include: 1) no effect on FEV_1. As seen earlier in figures 7.1 and 7.2, different COPD treatments have different paths of effect and treatment may affect rate of exacerbations, symptoms or mortality rather than lung function, or could be ineffective in the patient. If there is no effect on FEV_1, then FEV_1 would necessarily be the

same in the treated group as in no treatment/current treatment arm of the model and is illustrated by the navy curve in figure 7.11. Alternatively, treatment could: 2) cause an outward shift in the curve of FEV_1, for example, to the pink line as shown in figure 7.11. which shows that whilst FEV_1 declines at the same rate regardless of treatment, the treatment group has a comparatively higher FEV_1, 3) the rate of change in FEV_1 could be altered by treatment, illustrated by the green line in figure 7.11 so that FEV_1 declines at a slower rate than the non-treatment group or; 4) a shift in the curve could occur changing the rate of FEV_1 decline as seen in the light blue curve.
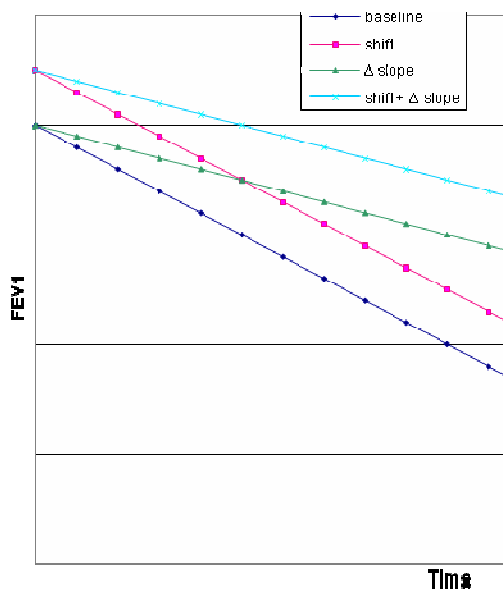
Figure 7.11 Effect of treatment on FEV$_1$ rate of decline

Some treatments for COPD have been shown to affect the decline in lung function.(246) In two recently reported roflumilast clinical trials (M2-127 and M2-128) the primary endpoint was change in prebronchodilator FEV_1. Study M2-127 found that combining salmeterol with roflumilast improved mean FEV_1 (SE) by 39 (9) ml compared to salmeterol with placebo in which FEV_1 worsened by 10 (9) ml. Study M2-128 examined the combination of tiotropium with roflumilast and reported a 65 (12) ml improvement in lung function

compared to the tiotropium plus placebo arm where lung function dropped by 16 (12) ml over the 24 weeks of the trial.(247) In terms of the effect of treatment on lung function, the trials showed that following treatment, lung function increased, such that a shift in the curve for FEV_1 was seen, peaking at around 12 weeks before a decline, with a similar rate of decline in the treatment in the placebo group. A similar trajectory is captured by the pink line in figure 7.11. Calverley et al reported similar shifts in the FEV curve as a result of treatment, including a 10% increase in pre-bronchodilator FEV_1 ($\approx$150ml) over a one year trial,(248) and Celli et al reported on an outward shift in FEV_1 within the TORCH study following treatment.(246) Therefore a curve for FEV_1 which shifts outwards following treatment, as illustrated by the pink curve in figure 7.11 appears to represent the mechanics of the situation well and is applied to the model within this section.

Duration of effect is an important consideration as this affects the duration for which the curves would be separated. An analysis of the TORCH dataset has shown that treatment effects occur as shown by the pink line in figure 7.11, for the duration of the three year trial.(246) As to date, TORCH is the longest trial within COPD, the duration of treatment effect beyond three years is unknown. A conservative analysis of the economic model would use the same duration of effect as observed within the RCT, after which time a return to the curve for the baseline FEV_1 may be modelled. Another scenario would be to keep the shift constant as long as treatment is continued. Scenario analysis can be used to investigate the implications on CE of different durations of effect.

If treatment improves FEV_1, a HRQoL gain and a reduction in costs would be expected. Mean QALYs are higher for people with larger lung functions (as defined by disease severity, see Chapter 6) and as the disease progression slows compared to the current treatment arm, exacerbation rates (and associated costs of hospitalisation) and symptoms should reduce.

7.3.2 Methods

The 'new treatment' arm of the model was developed based on the current treatment arm as described previously and is demonstrated applied to the same cohort: men aged 55 years with mild and moderate (NICE diagnosed) COPD at baseline. There are two differences between the 'new treatment' and the current treatment arms, the first is the incorporation of a treatment effect on lung function using the new treatment and the second is different treatment costs are applied to the different arms.

In order to assess the impact of an improvement in lung function on the cost effectiveness estimate, a one off increase in lung function of 150ml (as found by Calverley et al, see previous section) was incorporated into the 'new treatment' arm of the model as a dummy variable into the equation for FEV_1. This improvement was assumed to be sustained for the duration of the model or as long as treatment was continued.

With reference to section 7.2.6, in order to use a relevant value for new treatment costs, the treatment pathway as described in table 2.5 and treatment costs as estimated in table 2.6 (Chapter 2) were used to derive costs for a person with mild and moderate COPD after the 'new' COPD therapies entered the market (Seretide, Spiriva and Symbicort). Following the introduction of the 'new' therapies, it is anticipated that tiotropium has largely replaced the use of ipratropium, so that a person with mild COPD, is likely to be prescribed salbutamol plus tiotropium at a yearly cost of £493 ($808) and treatment for moderate COPD would be likely to include salbutamol, tiotropium and Seretide at an annual cost of £1240 ($2032). These treatment costs were used in the 'new treatment' arm of the model. Cost effectiveness statistics (ICERs) were calculated for the cohort containing men aged 55 years with mild COPD, and men with moderate COPD at baseline by taking the difference

between costs and QALYs in the current treatment arm as calculated in section 7.2 and those in the 'new treatment' arm.

Analyses were conducted around the user defined characteristics of: age, sex, symptoms and treatment duration, keeping all else the same and for each, ICERs were calculated. In further detail: age entering the model was changed from 55 years to 65 years; the effect of no respiratory symptoms at baseline, as opposed to presence of respiratory symptoms was assessed by changing the user defined input on symptoms from one to zero; the effect of changing the patient population to women instead of men was assessed and accordingly height was changed from 175 cm to 162 cm to reflect the difference in mean height between men and women; and the impact of a three year treatment duration was looked at by incorporating the treatment dummy variable for the first three years and then taking it out, different treatment costs were also assumed to occur for the first three years only, after which time they became the same as in the current treatment arm.

Results

A snap shot of the economic model representing the new treatment for 55 year old men with mild and moderate NICE diagnosed COPD is presented in tables 7.18 and 7.19 and combines all the individual regression equations developed in section 7.2 plus a treatment effect on lung function, to predict costs and effect, which are weighted by survival and are discounted.

The costs and QALYs over the modelled lifetime for the new treatment are shown in the bottom right hand corner of tables 7.18 and 7.19 and are £10 233 ($15 164) and 7.94 for the cohort with mild COPD at baseline, and £20 667 ($30 617) and 7.10 for those starting with moderate COPD. This compares to the cost and QALYs from current treatment, which

were £5204 ($7709) and 7.77 for mild disease and 7408 ($10 975) and 6.91 for moderate disease, as shown in tables 7.15 and 7.16.

Therefore, a new treatment which improves and sustains lung function by 150ml compared to current treatment leads to improved HRQoL in that more QALYs are gained compared to current treatment, but at higher cost. Corresponding ICERs are £22 888 ($34 300) per QALY in the mild group and £63 396 ($95 004) per QALY in the moderate group, compared to current treatment. Based on these ICERs, it is likely that the treatment would be accepted for use in the UK NHS for the management of mild (NICE diagnosed) COPD but not for the management of moderate COPD.

Table 7.18 Economic model for new treatment, men aged 55 yrs with mild COPD

Time	Age	FEV	pred FEV1	FEV1% pred	Symp	Exac minor	Exac major	EQ-5D utility	Tmt cost	Other cost	Total cost	S(t)	S(t)* utility	S(t)*total cost
1	55	2.15	3.02	0.71	61.8	0.47	0.06	0.65	808	341	1149	1.00	0.62	1106
2	56	2.13	2.99	0.71	61.8	0.47	0.06	0.65	808	350	1158	0.99	0.60	1066
3	57	2.11	2.96	0.71	61.8	0.47	0.06	0.65	808	359	1167	0.97	0.57	1023
4	58	2.08	2.93	0.71	61.9	0.47	0.06	0.65	808	368	1176	0.95	0.54	977
5	59	2.06	2.91	0.71	61.9	0.47	0.06	0.65	808	377	1185	0.93	0.51	929
6	60	2.04	2.88	0.71	61.9	0.47	0.06	0.65	808	386	1194	0.91	0.48	880
7	61	2.02	2.85	0.71	61.9	0.47	0.06	0.65	808	395	1203	0.88	0.45	829
8	62	2.00	2.82	0.71	61.9	0.47	0.07	0.65	808	405	1213	0.85	0.41	779
9	63	1.98	2.79	0.71	62.0	0.47	0.07	0.65	808	414	1222	0.81	0.38	728
10	64	1.95	2.76	0.71	62.0	0.47	0.07	0.65	808	424	1232	0.78	0.36	679
11	65	1.93	2.73	0.71	62.0	0.47	0.07	0.65	808	434	1242	0.74	0.33	630
12	66	1.91	2.70	0.71	62.0	0.47	0.07	0.65	808	444	1252	0.70	0.30	582
13	67	1.89	2.67	0.71	62.0	0.47	0.07	0.65	808	454	1262	0.67	0.27	537
14	68	1.87	2.64	0.71	62.1	0.47	0.08	0.65	808	464	1272	0.63	0.25	493
15	69	1.85	2.62	0.71	62.1	0.47	0.08	0.64	808	474	1282	0.59	0.23	451
16	70	1.82	2.59	0.71	62.1	0.47	0.08	0.64	808	484	1292	0.55	0.20	411
17	71	1.80	2.56	0.70	62.1	0.47	0.08	0.64	808	495	1303	0.51	0.18	373
18	72	1.78	2.53	0.70	62.1	0.48	0.08	0.64	808	505	1313	0.48	0.17	338
19	73	1.76	2.50	0.70	62.2	0.48	0.09	0.64	808	516	1324	0.44	0.15	305
20	74	1.74	2.47	0.70	62.2	0.48	0.09	0.64	808	527	1335	0.41	0.13	274
21	75	1.71	2.44	0.70	62.2	0.48	0.09	0.64	808	538	1346	0.38	0.12	245
22	76	1.69	2.41	0.70	62.2	0.48	0.09	0.64	808	549	1357	0.34	0.10	219
23	77	1.67	2.38	0.70	62.3	0.48	0.09	0.64	808	560	1368	0.31	0.09	195
24	78	1.65	2.35	0.70	62.3	0.48	0.10	0.64	808	572	1380	0.29	0.08	173
25	79	1.63	2.33	0.70	62.3	0.48	0.10	0.64	808	583	1391	0.26	0.07	153
26	80	1.61	2.30	0.70	62.3	0.48	0.10	0.64	808	595	1403	0.24	0.06	135
27	81	1.58	2.27	0.70	62.4	0.48	0.10	0.64	808	607	1415	0.21	0.05	119
28	82	1.56	2.24	0.70	62.4	0.48	0.11	0.64	808	619	1427	0.19	0.05	104
29	83	1.54	2.21	0.70	62.4	0.48	0.11	0.64	808	631	1439	0.17	0.04	91
30	84	1.52	2.18	0.70	62.5	0.48	0.11	0.64	808	644	1452	0.15	0.03	79
31	85	1.50	2.15	0.70	62.5	0.48	0.11	0.64	808	656	1464	0.14	0.03	69
32	86	1.48	2.12	0.70	62.5	0.48	0.12	0.64	808	669	1477	0.12	0.03	59
33	87	1.45	2.09	0.69	62.6	0.48	0.12	0.64	808	682	1490	0.11	0.02	51
34	88	1.43	2.06	0.69	62.6	0.48	0.12	0.64	808	695	1503	0.09	0.02	44
35	89	1.41	2.04	0.69	62.6	0.48	0.13	0.64	808	708	1516	0.08	0.02	38
													7.94	**15 162**

Table 7.19 Economic model for new treatment, men aged 55 yrs with moderate COPD

Time	Age	FEV	pred FEV1	FEV1% pred	Symp	Exac minor	Exac major	EQ-5D utility	Tmt cost	Other cost	Total cost	S(t)	S(t)* utility	S(t)*total cost
1	55	1.57	3.02	0.52	66.0	0.63	0.07	0.62	2032	499	2531	1.00	0.59	2435
2	56	1.55	2.99	0.52	66.1	0.63	0.08	0.62	2032	511	2543	0.98	0.57	2337
3	57	1.53	2.96	0.52	66.1	0.64	0.08	0.62	2032	523	2555	0.97	0.54	2230
4	58	1.51	2.93	0.51	66.2	0.64	0.08	0.62	2032	535	2567	0.95	0.51	2116
5	59	1.48	2.91	0.51	66.3	0.64	0.08	0.62	2032	547	2579	0.92	0.48	1997
6	60	1.46	2.88	0.51	66.3	0.64	0.09	0.62	2032	560	2592	0.89	0.45	1876
7	61	1.44	2.85	0.51	66.4	0.65	0.09	0.62	2032	572	2604	0.86	0.41	1753
8	62	1.42	2.82	0.50	66.5	0.65	0.09	0.61	2032	585	2617	0.82	0.38	1631
9	63	1.40	2.79	0.50	66.5	0.65	0.09	0.61	2032	598	2630	0.78	0.35	1511
10	64	1.38	2.76	0.50	66.6	0.65	0.10	0.61	2032	612	2644	0.74	0.32	1393
11	65	1.35	2.73	0.50	66.7	0.66	0.10	0.61	2032	625	2657	0.70	0.29	1278
12	66	1.33	2.70	0.49	66.7	0.66	0.10	0.61	2032	639	2671	0.66	0.27	1168
13	67	1.31	2.67	0.49	66.8	0.66	0.10	0.61	2032	653	2685	0.62	0.24	1063
14	68	1.29	2.64	0.49	66.9	0.67	0.11	0.61	2032	668	2700	0.58	0.22	963
15	69	1.27	2.62	0.48	67.0	0.67	0.11	0.61	2032	682	2714	0.54	0.20	869
16	70	1.24	2.59	0.48	67.1	0.67	0.11	0.61	2032	697	2729	0.50	0.17	781
17	71	1.22	2.56	0.48	67.1	0.67	0.12	0.61	2032	713	2745	0.46	0.16	699
18	72	1.20	2.53	0.48	67.2	0.68	0.12	0.61	2032	728	2760	0.42	0.14	623
19	73	1.18	2.50	0.47	67.3	0.68	0.12	0.61	2032	744	2776	0.38	0.12	553
20	74	1.16	2.47	0.47	67.4	0.68	0.13	0.61	2032	760	2792	0.35	0.11	489
21	75	1.14	2.44	0.47	67.5	0.69	0.13	0.61	2032	777	2809	0.32	0.09	431
22	76	1.11	2.41	0.46	67.6	0.69	0.13	0.61	2032	793	2825	0.29	0.08	378
23	77	1.09	2.38	0.46	67.7	0.70	0.14	0.60	2032	811	2843	0.26	0.07	331
24	78	1.07	2.35	0.45	67.8	0.70	0.14	0.60	2032	828	2860	0.23	0.06	288
25	79	1.05	2.33	0.45	67.9	0.70	0.14	0.60	2032	846	2878	0.21	0.05	250
26	80	1.03	2.30	0.45	68.0	0.71	0.15	0.60	2032	865	2897	0.18	0.04	216
27	81	1.01	2.27	0.44	68.1	0.71	0.15	0.60	2032	883	2915	0.16	0.04	186
28	82	0.98	2.24	0.44	68.2	0.72	0.16	0.60	2032	903	2935	0.14	0.03	160
29	83	0.96	2.21	0.44	68.3	0.72	0.16	0.60	2032	923	2955	0.13	0.03	136
30	84	0.94	2.18	0.43	68.5	0.73	0.17	0.60	2032	943	2975	0.11	0.02	116
31	85	0.92	2.15	0.43	68.6	0.73	0.17	0.60	2032	964	2996	0.10	0.02	99
32	86	0.90	2.12	0.42	68.7	0.74	0.18	0.60	2032	985	3017	0.08	0.02	83
33	87	0.87	2.09	0.42	68.8	0.74	0.18	0.60	2032	1007	3039	0.07	0.01	70
34	88	0.85	2.06	0.41	69.0	0.75	0.19	0.59	2032	1029	3061	0.06	0.01	59
35	89	0.83	2.04	0.41	69.1	0.75	0.19	0.59	2032	1053	3085	0.05	0.01	49
													7.12	**30 618**

Although the improvement was applied purely to FEV_1, knock on effects are seen to the majority of the components within the model, reflecting the interdependence between lung function, symptoms, exacerbations and mortality, and how these impact on costs and QALYs as previously described in section 7.1. Comparing the mild COPD cohort receiving the new treatment (table 7.18) to those receiving current treatment (table 7.15), at 89 years of age, symptoms scores are more favourable (62.6 instead of 64.3), there are fewer minor and fewer major exacerbations (0.48 minor exacerbations and 0.13 major exacerbations as opposed to 0.54 minor exacerbations and 0.14 major exacerbations), EQ-5D utility is marginally higher (0.65 compared to 0.64) and other costs are lower (£478 ($708) instead of £531 ($787)), but treatment costs are higher (£545 ($808) rather than £113 ($167)). There is a slight survival advantage in the new treatment arm at age 89 years (0.08 instead of 0.07). Similar improvements in symptoms, exacerbations rates, QALYs and other costs are seen in the modelled moderate COPD cohort receiving new treatment, in comparing table 7.16 to table 7.19, but there was no difference seen in survival between the groups.

Table 7.20 Results from analyses on heterogeneity, mild and moderate COPD
Costs and QALYs represent average costs over the model timeframe

			Current treatment		New treatment	
		ICER $ (£)	Cost ($)	QALYs	Cost ($)	QALYs
Mild	Base case	34 300 (22 888)	7710	7.72	15 162	7.94
	Aged 65 yrs	27 529 (18 371)	6334	5.61	11 784	5.81
	Women	24 429 (16 302)	9180	8.04	17 214	8.37
	No symptoms	34 125 (22 773)	6305	10.10	15 575	10.34
	Tx for 3yrs	13 268 (8854)	7710	7.72	9506	7.86
Moderate	Base case	95 004 (63 396)	10 977	6.91	30 618	7.12
	Aged 65 yrs	77 786 (51 909)	8756	4.89	22 829	5.07
	Women	68 377 (45 630)	14 466	6.81	35 736	7.12
	No symptoms	92 820 (61 943)	9404	9.18	31 194	9.42
	Tx for 3 yrs	37 384 (24 948)	10 977	6.91	15 820	7.04

Analyses were conducted around the user defined patient characteristics analysis for treatment on lung function. As seen from table 7.20, there were considerable differences in the cost effectiveness of the same treatment, dependent on user defined patient

characteristics. In general, the use of this treatment on a moderate COPD cohort is less likely to be cost effective than in a mild COPD cohort, with the ICERs for the mild cohort in the range of £9000 to £23 000 per QALY and those for the moderate population between £25 000 and £63 000 per QALY. The analyses on women with mild COPD gave lower ICERs than for men, with an ICER of £16 302 compared to £22 888 for the mild COPD cohort and £45 630 compared to £63 396 for the moderate COPD group. Using a cohort with no reported respiratory symptoms at baseline slightly reduced the ICERs compared to the base case in both the mild and the moderate cohort. Restricting treatment duration to three years reduced the ICER per QALY by more than half, from £22 888 to £8854 in the mild group and from £63 396 to £24 948 in the moderate group.

7.4 Discussion

In this chapter an economic model for COPD was developed, based around the key components of COPD that were identified within the thesis. The focus of the chapter was the conceptualisation of an economic model for COPD, which was successfully accomplished. From that, a working model developed and operationalised, which works and predicts sensible results.

This model is unique within COPD and breaks the mould for modelling the disease as Markov models are usually employed as opposed to regression based models. It has a number of advantages over earlier models, including: the incorporation of patient level heterogeneity; it can be used to assess a wide range of treatments; the explicit inclusion of symptoms into the model; and allowing for interdependence. These advantages are described in detail below.

A variety of people with COPD can be modelled with the model developed in this chapter, including those with: mild, moderate and severe disease; presence/absence of respiratory symptoms; and different smoking histories at baseline. A range of different ages can be modelled and men and women enter the model separately. Therefore the uncertainty pertaining to patient level heterogeneity is minimised. Because of this flexibility the resulting model can assist with the identification of subgroups in which a product is cost effective.

The model can be used to assess the cost effectiveness of different types of treatments, which is important because as was seen in Chapter 4, if different treatments can be assessed on the same platform then decisions around cost effectiveness can be standardised.

An important advantage of this model over other models for COPD is that this is the first to include a specific component for symptoms. Symptoms have a major impact on HRQoL for people with COPD and many treatments for COPD have mechanisms of action that work specifically on alleviating respiratory symptoms. Thus it is important to incorporate symptoms explicitly into the model.

The model developed in this chapter allows interdependence to exist between the different components of COPD and is novel in modelling COPD. Lung function, symptoms, exacerbations, survival, costs and utility are all intertwined so that a treatment which specifically works on one component, can have a knock on effect on the other components and this interdependence was clearly demonstrated in the final section of this chapter where treatment effects were incorporated into the model.

The country variable for UK was highly significant for both major and minor exacerbations. This is more likely to be explained by a higher propensity to hospitalise in

the UK and/or to seek treatment/support for the exacerbation, than the absolute risk of experiencing an exacerbation being higher in the UK than in other countries.

This type of model is demanding in terms of the need for good quality real data on which to derive the regression equations. The availability of the TORCH and the MIDSPAN datasets was of huge importance in enabling this type of model to be pursued and to be developed. The MIDSPAN observational study and the TORCH RCT are complementary types of dataset each with their own strengths and limitations. The benefit of using both types to develop the model is that where one dataset may not provide the information required, such as on long term survival, the other does, so that the resulting model is built on reliable data and is well supported. Because of this, the use of observational data together with RCT data can enhance the external validity of the cost effectiveness study.(162) For model building, the combined usage of such datasets is the gold standard for developing economic models.

The regression models within this chapter are developed using data from each of the treatment groups within the TORCH trial, not just the placebo group. This means that data from the treatment group (those randomised to Seretide) in TORCH were included to develop equations for the current treatment arm. Whilst it may be argued that only patients receiving the placebo should form the current treatment arm, this goes against the recommendation that it is current treatment that forms the comparator. It is argued here that whilst at the present time, Seretide may be one of the newest products on the market, it is anticipated that newer products, particularly triple combination therapies, will soon supersede this. In which case, benefits of new treatment will need to be compared to products currently on the market, which this model represents.

The Renfrew/Paisley (MIDSPAN) dataset, which recruited study participants between 1972 and 1976, was used for developing the equations on lung function and on survival. The use

of longitudinal data is important for accurately predicting long-term survival but by the very nature of the data, it will always be 'out of date'. Nonetheless, because no treatments for COPD have shown a clear survival benefit, it is likely that the equations presented here represent the natural history of COPD, even within the current treatment arm of the model.

The component regression equations would benefit from further research into model fit and validation; using methods similar to those demonstrated in Chapter 6, in order to ascertain whether the regression equations represent the natural history of the disease well. For example, the 20ml annual decline in FEV_1 that was predicted by the regression equation on lung function was consistent with some studies on lung function decline, however other studies have reported larger declines in lung function of around 50ml per year, as such it may be beneficial to conduct further research on the FEV_1 equation. One avenue for further research is around the type of dataset used. It might be that the use of cross-sectional data to predict longitudinal change in FEV_1 has underestimated the decline. One possible explanation is that because the population in MIDSPAN are those who survived long enough to participate in the study, a selection bias exists as those who died prior to the survey are likely to be in worse health with poorer lung function. As described in Chapter 5, there are a number of longitudinal datasets that exist such as the GPRD QRESEARCH, LINK and Mediplus databases. These datasets could be used to derive an equation for FEV1, though they also have limitations. For example, because they only contain data taken in GP surgeries, the resulting equations are unlikely to be generalisable to a wider COPD population.

Fine-tuning the equations is relatively straightforward to execute and was not pursued further in this chapter due to both time constraints and the fact that such methods were previously described and applied within Chapter 6. Potential study around each equation may include, but is not limited to: the use of fractional polynomials to identify any power

terms, the use of interaction terms to identify any relationships between the explanatory variables, the incorporation of treatment effects, the selection of best fitting explanatory variables and thorough testing of relationships between the components. Extensions to the survival model may include producing separate models by baseline disease severity and exploring the exponential model as an alternative to the Weibull model. Currently, because of the inconsistencies in predicting FEV_1 for men with severe COPD, as identified in the section on lung function in this chapter, and on survival, caution should be taken if the model is to be used to assess the cost effectiveness of treatment in people with severe (NICE diagnosed) COPD. Further work in the severe group would be of benefit.

This chapter focused on one kind of uncertainty; patient level heterogeneity, but as described in Chapter 3 there are others, including parameter uncertainty and structural uncertainty. It would be interesting to compare the outcomes of this modelling approach with that of another approach, such as a Markov model, so that the costs and outcomes could be compared between different methods. This would give some indication as to the degree that structural uncertainty has on the cost effectiveness estimate. Each of the coefficients in each of the equations within the model has a standard error attached to it. Fitting distributions to these coefficients and re-running the model whilst sampling from within those distributions in a probabilistic sensitivity analysis would address parameter uncertainty (maintaining the covariance structure between coefficients in the regression equations may be achieved using the Cholesky decomposition). Net benefit statistics can easily be calculated from the model by taking the value of the threshold, for example £20 000 and multiplying it by the change in QALYs, then subtracting the difference in costs between the current and new treatment arms of the model. A treatment is considered cost effective if the net benefit statistic is greater than zero.

This model was designed for the economic evaluation of different treatments for supporting reimbursement decisions to the UK NHS. However, there are a number of applications of this model in addition to establishing the cost effectiveness of treatment. One possible use of the model is in the 'early modelling' arena. At the early stages of a products life, there is uncertainty about the products efficacy and whether that product is worth the initial investment to try and take it to launch. Because of the flexibility to incorporate measurable outcomes within this model, such as absolute reduction in FEV_1 (something that could be predicted at the early stages of a products life), the cost effectiveness of such an improvement could be assessed. Taking it one step further, the model could be used to estimate a UK price for the treatment in order to bring it in under the UK threshold. Another possible use of this model would be for public health decisions relating to the COPD population, such as how much of an investment should be given over to smoking cessation programme, where one arm would consider current smokers and the treated arm would contain ex-smokers. Different smoking cessation costs could then be assessed to determine how much investment would bring the intervention under and over the UK threshold for cost effectiveness. These different uses of the model are all of interest and could be taken further in future research.

7.5 Conclusion

Within this chapter, data from an observational study and a RCT were used in a complementary manner in order that equations predicting lung function, exacerbations, symptoms, cost, EQ-5D utility and survival could be obtained. The resulting equations were combined so that an economic model representing the natural history of COPD under current treatment was built. A hypothetical treatment effect was combined into a second arm of the economic model to give a new treatment comparator arm. Costs and effects of

current treatment were compared to those from the new treatment to obtain cost

effectiveness statistics.

Chapter 8. Discussion

The broad aim of this thesis on 'Health Economic aspects in the management of COPD' was to combine information on the natural history of the disease with health economics in order to inform an economic model. The resulting model is a generic model that has the potential to facilitate better decision making, as different COPD therapies can be appraised on the same platform. The methods and principles used are generalisable to modelling other disease areas.

This thesis has shown that when observational and RCT datasets are used in partnership, the result is a unique and powerful body of evidence that can be and was successfully explored in order to develop a different kind of economic model than has previously been presented. The novelty has come through the harmonious use of two large and powerful data sets, TORCH and MIDSPAN. In Chapter 5, using the Renfrew/Paisley (MIDSPAN) study, the natural history of COPD was studied, including: an assessment of risk factors for mortality, the investigation of different diagnostic criteria for COPD, how disease severity affects hospitalisation rates and the rate and causes of mortality for people with COPD. Within Chapter 6, patient level data on EQ-5D utility from TORCH were used and particular attention was paid to predicting utility and the calculation of QALYs from the utility data. All of this information was brought together in order to inform the conceptualisation of a model for COPD and the data were used to develop a series of regression equations that when combined, formed a model representing the natural history of COPD as seen within Chapter 7.

A second aim of the thesis was to improve upon existing models for COPD. In this way, published models were reviewed and critiqued and within Chapter 4, desirable characteristics of cost effectiveness studies were identified, namely: the use of economic modelling, the need for a model to extend to a lifetime timeframe, the role of the comparator to be current practice rather than placebo, the importance of the QALY as the HRQoL outcome measure and the incorporation of capturing uncertainty. All these aspects

were used in order to guide the conceptualisation of a new economic model for COPD, as lessons learned, to avoid the same limitations being repeated in the new model.

This body of research is unique in a number of ways, which are described below:

The review of the published economic evaluations for COPD in Chapter 4, which highlights the weaknesses of current studies and suggesting ways in which a new model could improve upon existing ones has not previously been seen before.

The update of the Hole et al paper at the beginning of Chapter 5, not only replicated and validated existing analyses, but presented new results on the relationship between COPD and lung function. Analyses were conducted using both the traditional time along the x axis of the survival curve and, age on the x axis and found that the two ways of analysing survival data give similar results but that for presenting the data, age along the x axis is perhaps more useful for representing longitudinal observational data.

No other research is known of that compares the use of the NICE diagnostic criteria compared to the GOLD diagnostic criteria in order to identify people with COPD as was carried out in Chapter 5. This was made possible because of the relevance of the questions asked at the start of the Renfrew/Paisley (MIDSPAN) study and the lengthy follow-up of mortality.

The mapping algorithm developed in Chapter 6 goes beyond the scope of previous mapping equations which often start and stop with the use of OLS to develop the mapping equation. The use of the GLM and the two part models to develop the equation is a new idea in the area of mapping equations. The resulting model has been shown to predict well, in comparison to other mapping equations for COPD and this has been recognised in a study within the updated full NICE guidelines for COPD (anticipated publication date is June 2010),(249) which employs the algorithm developed in Chapter 6 in order to predict EQ-5D utility. Validation of the resulting mapping equation using QALYs in four different treatment groups has not been done before.

Within COPD and as far as is known, within health economics, the economic model that was developed is unique. Other models have used regression analyses to inform parts of their model and the Briggs et al model,(153) used a regression based model to do a within trial analysis on costs and effects. None have used regression models to predict values for each component of the disease and then combine these values into equations for cost and effect. The development of a baseline cohort representing current treatment, is rather different than a group taking a placebo treatment, or a specified comparator.

Access to such data is a privilege and by no means a right. The PhD Industrial Collaborative studentship that funded this research gave an advantageous position from which to access data held by both the pharmaceutical industry and by academia, the likes of which may be difficult in other situations. From the pharmaceutical perspective, clinical trials are expensive and potential income from successful drugs, very lucrative. As such, pharmaceutical companies must be very careful with their data. Observational studies such as the MIDSPAN dataset, developed within a university setting are incredibly valuable from a research perspective and it is essential that the integrity of the data is maintained. In both cases it is important that results are consistent with previous outcomes of the studies and that they are analysed and reported on with sensitivity and that the data are kept safe. Because of these issues, model building of this nature in which both types of data are employed, is perhaps best placed to occur within an academic setting.

This work confirms and/ or challenges the research previously been seen in that:

The methodologies used within some of the existing economic models for COPD were challenged in terms of the rationale for not including a death state within the model, the chosen timeframe being less than lifetime, no consideration given to uncertainty and the choice of outcomes, with support given to the use of the QALY as the appropriate measure for use in COPD.

The analyses replicating and building on the Hole et al study in Chapter 5 confirmed the earlier published results, including the reference equation as the updated results were found to be very similar to those reported. It has been confirmed that people with COPD die of

different causes, including but not exclusively from COPD mortality. This outcome has previously been documented, however within Chapter 5, causes of mortality by disease severity were also examined and it was found that whilst mortality from all other major causes of mortality decreases, mortality from respiratory disease increases. Whilst perhaps intuitive, from a modelling perspective these findings are important because they support and confirm the use of all-cause mortality as opposed to COPD mortality within economic models for the disease. The study on diagnostic criteria for COPD challenged the use of the GOLD diagnostic criteria, suggesting the NICE criteria in its place, but confirms that applying the GOLD criteria does identify people at higher risk of mortality than a non COPD defined population. The study supports the use of GOLD II diagnostic criteria to identify a COPD population, as has been used for the reporting in the BOLD study.

Whilst challenging the sole reliance of OLS in order to develop a mapping algorithm, the study confirmed the view that increasingly complex structures may only slightly improve the predictive ability of the model: a simple model developed using OLS was selected on the grounds that a complex structure such as the two-part model only offered a 0.0003 improvement in RMSE. In the past, where internal validation has been conducted on the same dataset, the data have been split randomly. This study dismissed this as an option in favour of a random split of the data for reasons previously described. The mapping study was cautious in its conclusions about overly promoting the use of mapping, and confined it as a second best solution, which is a more conservative approach than others have made.

The research within this thesis has confirmed the use of lung function as an important risk factor for COPD and the application of it in order to define disease severity groupings from which summary statistics can be compared. The model developed in Chapter 8 rejected the use of lung function to define Markov states, in favour of a more flexible approach where lung function continually changes over time. Previously symptoms have not been included within an economic model for COPD. However, it has been seen that symptoms are an important component of disease and that therefore they should be captured within an economic model. This study has shown that a regression based approach is feasible and can be used to model COPD.

As well as being the first COPD model to use a proxy for symptoms, this model improves upon other models in COPD by allowing for the incorporation of absolute changes to lung function, exacerbation rates and symptoms scores directly into the model. In addition, the model has a flexible yet standard framework that permits the cost effectiveness of alternative COPD treatment strategies to be established for different patient populations, such as by disease severity, which is of use for decision-making.

Another benefit of this new approach in terms of decision-making is that the model attempts to match the disease population more closely than traditional models have been able to do. As a result, uncertainty associated with heterogeneity is reduced, whilst none of the functionality of more traditional approaches in dealing with parameter uncertainty is lost.

From a theoretical standpoint, this model has opened up a new dimension in model building, with a strong focus on longitudinal data on the natural history of the disease to inform the model. The model was built from the bottom up, based by an understanding of the epidemiology of the disease as a result of analyses conducted on a wealth of data. The methods and principles used are generalisable to developing and modelling other disease areas, particularly to chronic conditions where there are large amounts of data, such as in arthritis and cancer.

There is scope for further development of the research conducted and presented within this PhD thesis. The list below represents those areas in which further research would be both of interest and of use, in the order in which they would be approached.

The modelling chapter represents the area in which the most research would be beneficial, in particular, further development and validation of the individual regression equations that make up the economic model. This would be the priority for future research. Chapter 6 illustrates ways in which equations can be developed and validated. Methods for further development of each regression equation would follow the methods used in Chapter 6 and have been described previously in Chapter 7. For validation, the dataset could be split into two to form a fitting and a validation dataset, where fit would be determined using RMSE

as used within Chapter 6. An alternative approach for external validation is to compare the results from an alternative dataset. For example by comparing the results from the equation predicting FEV_1, or predicted survival probability from the Weibull model, which were developed within the MIDSPAN dataset, and validating them using the TORCH dataset. Once the prediction equations have been further developed and validated, in order to support an application of this model for example, within a HTA submission, then analysis of parameter uncertainty, involving the employment of the standard errors around the coefficients of each equation should be assessed.

Further work arising from Chapter 6 would include an analysis of a complete case cohort to predict QALYs, based on: observed EQ-5D data, and predicted EQ-5D scores from employing the mapping equation. The results of this analysis would be compared to the results presented in Chapter 6, in which QALYs were predicted using a dataset in which there were missing data.

A variety of analyses could be conducted on the Renfrew/Paisley (MIDSPAN) dataset in order to further study COPD. For example, research into different diagnostic criteria could be undertaken, particularly with regards to alternative diagnostic methods such as the Lower Limit of Normal classification, instead of FEV_1/FVC.

This thesis has contributed to knowledge in a number of ways. The new mapping equation for COPD is the first to predict plausible results and can be used to generate EQ-5D utility where only data on the SRGQ were collected.

The NICE COPD criteria identified a group at higher risk of all-cause and COPD mortality compared to the more frequently used GOLD diagnostic criteria, and importantly the resulting COPD population is much smaller when the NICE criteria are applied compared to when using the GOLD criteria.

It has been shown that observational and RCT data can be combined in a meaningful and useful way in order to model disease. The published economic models for COPD were critically appraised and the results used to inform the new economic model for COPD as

presented in Chapter 7. It is thought that no other model for COPD exists where results from RCTs can be directly plugged into the model, such as an improvement in FEV_1 of x ml per year. The model developed in Chapter 7 is based on a series of regression equations that predict outcomes for key components of disease: lung function, respiratory symptoms and exacerbations. This is an entirely new concept and approach for modelling COPD.

Appendix

PAISLEY

midspan health plan

conducted by the
Renfrewshire King Edward Memorial Trust, the
Glasgow MMR Cardiorespiratory Screening Unit
and the Department of Community Medicine
of the University of Glasgow.

For official use only

1.01	1.02	1.03	1.04	1.05	1.06	1.07

MMR reference number

Questionnaire and Health Record

This form may look a little long and difficult, but in fact
it takes only a few minutes to answer.

Please answer all the questions in this health record and
bring it with you when you come to the unit.
Write a tick (√) in the appropriate answer box.
If you cannot give an exact answer, give the best guess you can.

---

1 Surname (Mr., Mrs., Miss) . . . .
(BLOCK LETTERS)

If married or divorced woman—
Birth Surname (Maiden Name) . . . .

First Name(s) . . . .

Address. . . . .

. . . .

Home Telephone Number . . . .

National Health Service Number. . . . .
(if known)

Place of Birth . . . .

1.08	1.09	1.10	1.11	1.12	1.13	1.14
1.15	1.16	1.17	1.18	1.19	1.20	

1.21

2 Sex. . . . .

3 Are you

1 ☐ Married? 3 ☐ Single?
2 ☐ Widowed? 4 ☐ Other?

1.22

4 What is your date of birth ?

Day of Month   Month   Year

. . . .

1.23	1.24	1.25	1.26	1.27	1.28

5 What is your usual occupation?

If retired, please give last full-time occupation.
If you are a whole-time housewife, what is your husband's occupation?
Please describe as fully as possible.

....
....
....
....

6 Family Doctor. Dr...

Address...

....

7 How much walking altogether do you do on the way to and from work on an average day? If you are a housewife, how much walking altogether do you do to and from the shops?

.... minutes

8 Do you usually bring up any phlegm from your chest first thing in the morning in the winter?

1 ☐ Yes   2 ☐ No

If 'No', go to question 11.

9 Do you bring up phlegm like this on most days for as much as three months in the winter each year?

1 ☐ Yes   2 ☐ No

10 In the past three years have you had a period of increased cough and phlegm lasting for three weeks or more?

1 ☐ Yes—1 period   3 ☐ No

2 ☐ Yes—2 or more periods

11 Do you get short of breath walking with people of your own age on level ground?

1 ☐ Yes   2 ☐ No

12 Does your chest sound wheezy or whistling on most days (or nights)?

1 ☐ Yes   2 ☐ No

13 Does the weather affect your breathing?

a 1 ☐ Yes   2 ☐ No

If 'No', go to question 14.

b If 'Yes':

1 Specify type of weather

....

2 Does it make your breathing

1 ☐ worse?   2 ☐ better?

14 Do you smoke ~~more than 5 cigarettes each day~~ now ?

   a 1 ☐ Yes   2 ☐ No

   b If 'Yes':

   1 How many cigarettes do you usually smoke on an average day ?

   . . . .

   2 At what age did you start smoking ?

   . . . .

   3 Do you inhale ?

     1 ☐ Yes   2 ☐ No

   c What brand of tobacco do you normally smoke ?

   . . . .

---

15 If you do not smoke cigarettes now, did you ever smoke ~~more than 5 a day~~ regularly ?

   a 1 ☐ Yes   2 ☐ No, never

   b If 'Yes':

   1 At what age did you start smoking ?

   . . . .

   2 At what age did you stop smoking ?

   . . . .

   3 What is the most you ever smoked *per day* for as long as a year ?

   . . . .

---

16 a Have you ever had any pain or discomfort in your chest ?

    1 ☐ Yes   2 ☐ No (If 'No', go to question 17)

   b Do you get this pain or discomfort when you walk uphill or hurry ?

    1 ☐ Yes   2 ☐ No

   c Do you get it when you walk at an ordinary pace on the level ?

    1 ☐ Yes   2 ☐ No

   d When you get any pain or discomfort in your chest what do you do ?

    1 ☐ Stop

    2 ☐ Slow down

    3 ☐ Continue at the same speed

   e Does it go away when you stand still ?

    1 ☐ Yes   2 ☐ No

   f If 'Yes', how soon ?

    1 ☐ 10 minutes or less

    2 ☐ More than 10 minutes

   g Where do you get this pain or discomfort ? Mark the place(s) with X on the diagram

Right side   Left side

1.55

1.56

1.57

1.58

1.59

1.60

1.61

---

17 Have you ever had a severe pain across the front of your chest lasting for half an hour or more ?

    1 ☐ Yes   2 ☐ No

18 Are you, or have you ever been diabetic?

1 ☐ Yes
2 ☐ No
3 ☐ Do not know

19 Have you ever been in hospital?

a 1 ☐ Yes   2 ☐ No
If 'Yes' please give the following information for each time you went into hospital.

b   Year   Nature of illness or what was wrong?

1 ....

2 ....

3 ....

4 ....

(Continue on a separate sheet of paper if necessary.)

20 1 Have you ever had a stroke (shock)?

1 ☐ Yes   2 ☐ No

2 Have you ever, without warning:

a suddenly lost the power of an arm?
1 ☐ Yes   2 ☐ No

b suddenly lost the power of a leg?
1 ☐ Yes   2 ☐ No

c suddenly been unable to speak properly?
1 ☐ Yes   2 ☐ No

d suddenly lost consciousness?
1 ☐ Yes   2 ☐ No

3 Do you ever complain of headaches on one side of your head?
1 ☐ Yes   2 ☐ No

21 a Have you ever had high blood pressure?
1 ☐ Yes   2 ☐ No

b Are you receiving treatment for your blood pressure?
1 ☐ Yes   2 ☐ No

22 Do you suffer from, or have you ever suffered from:

a Asthma?
1 ☐ Yes   2 ☐ No

b Hayfever?
1 ☐ Yes   2 ☐ No

23 How long have you stayed in your present home?

.........................years

I understand that everything which could identify me will be treated as strictly confidential.

I do not object to any of my hospital records being looked at by members of the team to help them in the survey.

I wish / I do not wish the results of my examination to be sent to my family doctor.

Signature ....

Thank you for your help

For official use only

Name .... Pla ....

2.01	2.02	2.03	2.04	2.05	2.06	2.07

M.M.R. reference number

2.08	2.09	2.10	2.11	2.12	2.13
4 3	4 4		4 5		

Date . . . .

BLOOD PRESSURE

1 Systolic . . . .

2.14	2.15	2.16	(57)
4 6			

2 Diastolic . . . .

2.17	2.18	2.19	(52)
4 7			

Blood Pressure Observer

2.20	2.21	(53)
4 8		

CHEST X-RAY

2.22	2.23	(52)
4 9		

HISTORY

2.24	(55)
5 0	

TINE TEST
Self reading

2.25	(56)
5 1	

Observer reading

2.26	(57)
5 2	

Tine tester

2.27	2.28	(58)
5 3		

Tine Test Observer

2.29	2.30	(59)
5 4		

SPUTUM
Direct

2.31	(60)
5 5	

Culture

2.32	(61)
5 6	

BLOOD
Plasma Serum cholesterol

2.33	2.34	2.35	(62)
	5 7		

Sugar

2.36	2.37	2.38	(63)
	5 8		

Blood sampler

2.39	2.40	(64)
5 9		

Heart width
Thoracic width

2.41	2.42	2.43	2.44
	6 0		
(65)		(66)	

ANGINA

2.45	(67)
6 1	

ANTHROPOMETRY
Height (without shoes)   mtrs.   cms.   (F3.2)

2.46	2.47	2.48	(68)
	6 2		

Weight (indoor clothing)   kilos.   (F4.0)

2.49	2.50	2.51	2.52	(69)
	6 3			

Observer

2.53	2.54	(70)
6 4		

E.C.G.   i . . . .   ii . . . .   iii . . . .

           iv . . . .   v . . . .   vi . . . .

           vii . . . .   viii . . . .   ix . . . .

E.C.G. Summary

Operator

Coder

Adjudicator

M.M.R. reference number

RESPIRATORY FUNCTION
F.E.V. 1.0

F.V.C.

Respiratory Function Tester

CATEGORY

Interviewer or questionnaire checker

Coder

Checker

	(71)		(72)		(73)
i 2.55	2.56	ii 2.57	2.58	iii 2.59	2.60
iv 2.61	2.62	v 2.63	2.64	vi 2.65	2.66
	(74)		(75)		(76)
vii 2.67	2.68	viii 2.69	2.70	ix 2.71	2.72
	(77)		(78)		(79)

2.73	(80)
65	

2.74	2.75	(81)
66		

2.76	2.77	(82)
67		

2.78	2.79	(83)
68		

2.80
2

3.01	3.02	3.03	3.04	3.05	3.06	3.07

3.08	3.09	3.10	(84)
	69		

3.11	3.12	3.13	(85)
	70		

3.14	3.15	(86)
7	1	

3.16	3.17	(87)
7	2	

3.18	3.19	(88)

3.20	3.21	(89)

3.22	3.23	(90)

SGRQ (250)

These questions are about how much chest trouble you have had over the past year. Please tick one box in response to each question.

1. Over the last year, I have coughed:
   - Most days a week
   - Several days a week
   - A few days a week
   - Only with chest infections

2. Over the last year, I have brought up phlegm (sputum):
   - Most days a week
   - Several days a week
   - A few days a week
   - Only with chest infections

3. Over the last year, I have had shortness of breath:
   - Most days a week
   - Several days a week
   - A few days a week
   - Only with chest infections

4. Over the last year, I have had attacks of wheezing:
   - Most days a week
   - Several days a week
   - A few days a week
   - Only with chest infections

5. During the last year, how many severe or very bad unpleasant attacks of chest trouble have you had?
   - More than three attacks
   - 3 attacks
   - 2 attacks
   - 1 attack
   - No attacks

6. How long did the worst attack of chest trouble last? (if you had no severe attacks go to question 7)
   - A week or more
   - 3 or more days
   - 1 or 2 days
   - Less than a day

7. Over the last year, in an average week, how many good days (with little chest trouble) have you had?
   - None
   - 1 or 2
   - 3 or 4
   - Nearly every day
   - Every day

8. Do you have a wheeze?
   - No
   - Yes

   If yes, is it worse in the morning?
   - No
   - Yes

9. How would you describe your chest condition?
   - The most important problem I have

- Causes me quite a lot of problems
- Causes me a few problems
- Causes no problem

10. If you have ever had paid employment:
- My chest trouble made me stop work
- My chest trouble interferes with my work
- Or made me change my work
- My chest trouble does not affect my work

11. Questions about what activities usually make you feel breathless. Please tick one box in response to each question as it applies to you recently:
a) Sitting or lying
- True
- False

b) Getting washed or dressed
- True
- False

c) Walking around the home
- True
- False

d) Walking outside on the level
- True
- False

e) Walking up a flight of stairs
- True
- False

f) Walking up hills
- True
- False

g) Playing sports or games
- True
- False

12. More questions about your cough and breathlessness. Please tick one box in response to each question as it applies to you recently:
a) My cough hurts
- True
- False

b) My cough makes me tired
- True
- False

c) I get breathless when I talk
- True
- False

d) I get breathless when I bend over
- True
- False

e) My cough or breathing disturbs my sleep
- True

- False

f) I get exhausted easily
- True
- False

13. Questions about other effects your chest trouble may have on you. Please tick one box in response to each question as it applies to you recently:
a) My cough or breathing is embarrassing in public
- True
- False

b) I get afraid or panic when I cannot get my breath
- True
- False

c) I feel that I am not in control of my chest problem
- True
- False

d) I do not expect my chest to get any better
- True
- False

e) I have become frail or an invalid because of my chest
- True
- False

f) Exercise is not safe for me
- True
- False

g) Everything seems too much of an effort
- True
- False

14. Questions about your medication (if you are not receiving medication for your chest trouble, go to question 15). Please tick one box in response to each question as it applies to you recently:
a) My medication does not help me very much
- True
- False

b) I get embarrassed using my medication in public
- True
- False

c) I have unpleasant side effects from my medication
- True
- False

d) My medication interferes with my life a lot
- True
- False

15. Questions about how activities may be affected by your breathing. Please tick one box in response to each question as it applies because of your breathing:

a) I take a long time to get washed or dressed
- True
- False

b) I cannot take a bath or shower, or I take a long time
- True
- False

c) I walk more slowly than other people, or I stop for rests
- True
- False

d) Jobs such as housework take a long time, or I have to stop for rests
- True
- False

e) If I walk up one flight of stairs, I have to go slowly or stop
- True
- False

f) If I hurry or walk fast, I have to stop or slow down
- True
- False

g) My breathing makes it difficult to do things such as walk up hills, carry things up stairs, light gardening such as weeding, dance, play bowls or play golf.
- True
- False

h) My breathing makes it difficult to do things such as carry heavy loads, dig the garden or shovel snow, jog or walk at 5 mile per hour, play tennis, or swim.
- True
- False

i) My breathing makes it difficult to do things such as very heavy manual work, run, cycle, swim fast or play competitive sports.
- True
- False

16. We would like to know how your chest trouble usually affects your daily life, Please tick one box for each question as it applies because of your chest trouble:

a) I cannot play sports or games
- True
- False

b) I cannot go out for entertainment or recreation
- True
- False

c) I cannot go out of the house to do the shopping
- True
- False

d) I cannot do housework
- True
- False

e) I cannot move far from my bed or chair
- True
- False

Here is a list of other activities that your chest trouble may prevent you from doing (You do not have to tick these, they are just examples of ways in which breathlessness may affect you)

- Going for walks or walking the dog
- Doing things in the home or in the garden
- Sexual intercourse
- Going out to church, pub or place of entertainment
- Going out in bad weather or into smokey rooms
- Visiting family or friends or playing with the children

Please write down any other important activities that your chest trouble may stop you doing: …………………………………………………………………

17. Tick the statement which you think best describes how your chest affects you. Please tick only one box:
- It does not stop me doing anything I would like to do
- It stops me doing one or two things I would like to do
- It stops me doing most of the things I would like to do
- It stops me doing everything I would like to do

EQ-5D (251)

1. Mobility:

   - No problems walking about (1)
   - Some problems walking about (2)
   - Confined to bed (3)

2. Self care
   - No problems with self-care (1)
   - Some problems washing or dressing myself (2)
   - Unable to wash or dress myself (3)

3. Usual activities
   - No problems with performing usual activities (eg work, study, housework, family or leisure activities) (1)
   - Some problems with performing usual activities (2)
   - Unable to perform usual activities (3)

4. Pain/discomfort
   - No pain or discomfort (1)
   - Moderate pain or discomfort (2)
   - Extreme pain or discomfort (3)

5. Emotions
   - Not anxious or depressed (1)
   - Moderately anxious or depressed (2)
   - Extremely anxious or depressed (3)

References

(1)   Brines R, Thorne M. Clinical consensus on COPD. Respiratory Medicine 2007; COPD update 3:42-48.

(2)   GOLD. Global Initiative for Chronic Obstructive Lung Disease: global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. 2007.

(3)   Fletcher C, Peto R. The natural history of chronic airflow obstruction. British Medical Journal 1977; 1:1645-1648.

(4)   British Thoracic Society. BTS guidelines for the management of Chronic Obstructive Pulmonary Disease. Thorax 1997; 52(S1):S28.

(5)   Celli BR, MacNee W, Augusti A, Anzueto A, Berg M, Buist AS et al. Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper. European Respiratory Journal 2004; 23(932):946.

(6)   Siafakas N, Vermeire P, Pride NB, Paoletti P, Gibson P, Howard JC et al. Optimal assessment and management of chronic obstructive pulmonary disease (COPD). European Respiratory Journal 1995; 8(1398):1420.

(7)   National Collaborating Centre for Chronic Conditions for the National Institute of Health and Clinical Excellence. Chronic obstructive pulmonary disease. National clinical guideline on management of chronic obstructive pulmonary disease in adults in primary and secondary care. Thorax 2004; 59 Suppl 1:1-232.

(8)   National Institute of Health and Clinical Excellence. Diagnosing COPD. Chronic obstructive pulmonary disease: national clinical guideline on management of chronic obstructive pulmonary disease in adults in primary and secondary care. Thorax 2004; 59:27-38.

(9)   Burge S, Wedzicha JA. COPD exacerbations: definitions and classifications. European Respiratory Journal Suppl 2003; 41:46s-53s.

(10)   Burge PS, Calverley PM, Jones PW, Spencer S, Anderson JA et al. Randomised, double blind, placebo controlled study of fluticasone propionate in patients with moderate to severe chronic obstructive pulmonary disease: the ISOLDE trial. British Medical Journal 2000; 320:1297-1303.

(11)   European Respiratory Society, European Lung Foundation. Chronic Obstructive Pulmonary Disease. European Lung White Book. Huddersfield: ERSJ, 2003.

(12)   Calverley PM, Bellamy D. The challenge of providing better care for patients with chronic obstructive pulmonary disease: the poor relation of airways obstruction? Thorax 2000; 55:78-82.

(13)   Fukuchi Y, Nishimura M, Ichinose M, Adachi M, Nagai A, Kuriyama T et al. COPD in Japan: The Nippon COPD Epidemiology study. Respirology 2004; 9(4).

(14)   Zielinski J, Bednarek M, Gorecka D, Viegi G, Hurd SS, Fukuchi Y et al. Increasing COPD awareness. European Respiratory Journal 2006; 27:833-852.

(15)   IMS disease analyzer-Mediplus UK. IMS, editor. 2006.

(16)   Starkie HJ, Briggs AH, Chambers MG. Pharmacoeconomics in COPD: lessons for the future. The International Journal of COPD. 2008; 3(1): 71-88.

(17)   Lokke A, Lange P, Scharling H, Fabricius P, Vestbo J. Developing COPD: a 25 year follow up study of the general population. Thorax 2006; 61:935-939.

(18)   Mackay J, Eriksen M. *The Tobacco Atlas*. World Health Organization. 2002.

(19)   Anthonisen N, Skeans MA, Wise RA, Manfeda J, Kanner RE, Connett JE. The effects of a smoking cessation intervention on 14.5 year mortality. Annals of Internal Medicine 2005; 142(4):233-239.

(20)   Murray Christopher JL, Lopez AD, Mathers CD, Stein C. *The Global Burden of Disease 2000 project: aims, methods and data sources*. World Health Organization. 2001.

(21)   US Department of Health and Human Services. Summary Health Statistics for US Adults: National Health Interview Survey, 2004. 2004.

(22)   Department of Health. Improving Chronic Disease Management.[Online] 2004. Available from http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicy AndGuidance/DH_4075214 [Accessed 18th July 2010].

(23)   Calverley PM. The burden of obstructive lung disease in the UK - COPD and asthma. Thorax 1998; 53 Suppl 4.

(24)   Britton M. The burden of COPD in the U.K.: results from the Confronting COPD survey. Respiratory Medicine 2003; 97 Suppl C:S71-S79.

(25)   Guest JF. The annual cost of chronic obstructive pulmonary disease to the UK's National Health Service. Disease Management & Health Outcomes 1999; 5(2):93-100.

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

(27)   Halpern MT, Stanford RH, Borker R. The burden of COPD in the U.S.A.: results from the Confronting COPD survey. Respiratory Medicine 2003; 97 Suppl C:S81-S89.

(28)   Izquierdo JL. The burden of COPD in Spain: results from the Confronting COPD survey. Respiratory Medicine 2003; 97 Suppl C:S61-S69.

(29)   Wouters EF. The burden of COPD in The Netherlands: results from the Confronting COPD survey. Respiratory Medicine 2003; 97 Suppl C:S51-S59.

(30)   Dal Negro R, Rossi A, Cerveri I. The burden of COPD in Italy: results from the Confronting COPD survey. Respiratory Medicine 2003; 97 Suppl C:S43-S50.

(31)   Chapman KR, Bourbeau J, Rance L. The burden of COPD in Canada: results from the Confronting COPD survey. Respiratory Medicine 2003; 97 Suppl C:S23-S31.

(32)   Piperno D, Huchon G, Pribil C, Boucot I, Similowski T. The burden of COPD in France: results from the Confronting COPD survey. Respiratory Medicine 2003; 97(Suppl C):S33-S42.

(33)   Jansson SA, Andersson F, Borg S, Ericsson A, Jonsson E, Lundback B et al. Costs of COPD in Sweden according to disease severity. Chest 2002; 122(6):1994-2002.

(34)   Lung and Asthma Information Agency. Trends in COPD. 2003.

(35)   Department of Health. NHS reference costs 2005-2006. 2006. London, Department of Health.

(36)   Andersson F, Borg S, Jansson SA, Jonsson AC, Ericsson A, Prutz C et al. The costs of exacerbations in chronic obstructive pulmonary disease (COPD). Respiratory Medicine 2002; 96(9):700-708.

(37)   McGuire A, Irwin DE, Fenn P, Gray A, Anderson P, Lovering A et al. The excess cost of acute exacerbations of chronic bronchitis in patients aged 45 and older in England and Wales. Value Health 2001; 4(5).

(38)   Starkie HJ, Briggs AH, Chambers MG. Pharmacoeconomics in COPD: lessons for the future. International Journal of COPD 2008; 3(1):71-88.

(39)   Pritchard C, Sculpher M. Productivity costs: principles and practice in economic evaluation. London: Office of Health Economics, 2000.

(40)   Halpin DM. Health economics of chronic obstructive pulmonary disease.
        Proceedings of the American Thoracic Society. 2006; 3(3):227-233.

(41)   Yelin E, Katz Patricia, Balmes J. Work life of persons with asthma, rhinits, and
        COPD: A study using a national, population based sample. Journal of Occupational
        Medicine and Toxicology 2006; 1(2).

(42)   Calverley PM, Sondhi S. The burden of obstructive lung disease in the UK - COPD
        and Asthma. Thorax 1998; 53 Suppl 4:A83.

(43)   Adelphi Group. Asthma/COPD DSP V USA & Europe COPD. 21-11-2006.

(44)   British Thoracic Society standards of care subcommittee on pulmonary
        rehabilitation. Pulmonary rehabilitation. Thorax 2001; 56(827):834.

(45)   Pelletier-Fleury N, Lanoe JL, Fleury B, Fardeau M. The cost of treating COPD
        patients with long-term oxygen therapy in a French population. Chest 1996;
        110:411-416.

(46)   Meyers BF, Patterson GA. Chronic obstructive pulmonary disease v 10:
        Bullectomy, lung volume reduction surgery, and transplantation for patients with
        chronic obstructive pulmonary disease. Thorax 2003; 58:634-638.

(47)   Jones PW, Quirk FH, Baveystock CM, Littlejohns P. A self-complete measure of
        health status for chronic airflow limitation. American Review of Respiratory
        Disease 1992; 145(1321):1327.

(48)   EuroQol group. EuroQol - a new facility for the measurement of health-related
        quality of life. Health Policy 1990; 16:199-208.

(49)   Calverley PM, Anderson JA, Celli B, Ferguson GT, Jenkins C, Jones PW et al.
        Salmeterol and Fluticasone proprotionate and survival in Chronic Obstructive
        Pulmonary Disease. New England Journal of Medicine 2007; 356(8):775-789.

(50)   Calverley PM, Pauwels R, Vestbo, Jones P, Pride N, Gulsvik A et al. Combining
        salmeterol and fluticasone in the treatment of chronic obstructive pulmonary
        disease. Lancet 2003; 361:449-456.

(51)   Decramer M, Celli B, Tashkin DP, Pauwels RA, Burkhart D, Cassino C et al.
        Clinical trial design considerations in assessing long term functional impacts of
        Tiotropium in COPD: the UPLIFT trial. COPD:Journal of Chronic Obstructive
        Pulmonary Disease 2004; 1(2):303-312.

(52)   Vincken W, Van Noord JA, Greefhorst APM, Bantje ThA, Kesten S, Korducki L et
        al. Improved health outcomes in patients with COPD during 1 yr's treatment with
        tiotropium. European Respiratory Journal 2002; 19(209):216.

(53)   Seemungal TA, Stockley R, Calverley PM, Hagan GW, Wedzicha JA. Investigating new standards for Prophylaxis in reduction of exacerbations - the INSPIRE study methodology. COPD 2007; 4(3):177-183.

(54)   Aaron sD, Vandemheen K, Fergusson D, FitzGerald M, Maltais F, Bourbeau J et al. The Canadian Optimal Therapy of COPD Trial: Design, organization and patient recruitment. Canadian Respiratory Journal 2004; 11(8):581-585.

(55)   Welte T, Hartman L, Polanowski T et al. Budesonide/formoterol added to tiotropium is well tolerated and reduces risk of severe exacerbations in COPD patients. Abstract presented at American Thoracic Society International Conference, San Diego, USA. May 2009

(56)   Welte T, Miravitlles M, Hernandez P, Ericsson G, Peterson S, Polanowski T et al. Efficacy and tolerability of Budesonide/Formoterol added to Tiotropium in patients with Chronic Obstructive Pulmonary Disease. American Journal of Respiratory and Critical Care Medicine 2009; 180:741-750.

(57)   Aaron SD, Vandemheen K, Fergusson D, Maltais F, Bourbeau J, Goldstein R et al. Tiotropium in combination with placebo, salmeterol or fluticasone-salmeterol for treatment of chronic obstructive pulmonary disease. Annals of Internal Medicine 2007; 146(545):555.

(58)   Joint Formulary Committee (2007). British National Formulary. 54th edition. London: British Medical Association and Royal Pharmaceutical Society of Great Britain. British National Formulary.

(59)   Department of Health. Improving Chronic Disease Management.[Online] 2004. Available from http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicy AndGuidance/DH_4075214 [Accessed 18th July 2010].

(60)   The Department of Health. Investing in general practice: the new general medical services contract. 30-12-0003. London, The NHS Confederation and the BMA.

(61)   Dowson LJ, Yeung A, Allen MB. General practice spirometry in North Staffordshire. Monaldi Archives for Chest Disease 1999; 54(2):186-188.

(62)   Ma Lwin A, McKinley RK. Management of COPD in Primary Care in Leicestershire. Primary Care Respiratory Journal 2005; 14(1):38-41.

(63)   Cleland J, Mackenzue M, Small I, Douglas G, Gentles I. Management of COPD in primary care in North-East Scotland. Scottish Medical Journal 2006; 51(4):10-14.

(64)   Dowson LJ, Yeung A, Allen MB. Most practices would use open access spirometry in hospitals. British Medical Journal 1998; 317(209).

(65)   Jones RCM, Freegard S, Reeves M, Hanney K, Dobbs F. The role of the practice nurse in the management of chronic obstructive pulmonary disease (COPD). Primary Care Respiratory Journal 2001; 10:106-108.

(66)   Bellamy D. Spirometry in Practice. 2nd Ed. 2005. London, British Thoracic Society (BTS) COPD Consortium.

(67)   British Thoracic Society. Intermediate Care - Hospital-at-Home in Chronic Obstructive Pulmonary Disease. Thorax 2007; 63(3):200-210.

(68)   Cotton MM, Bucknall CE, Dagg KD, Johnson MK, MacGregor G, Stewart C et al. Early discharge for patients with exacerbations of Chronic Obstructive Pulmonary Disease: a randomised controlled trial. Thorax 2000; 55:902-906.

(69)   Mannino D. COPD: epidemiology, prevalence, morbidity and mortality and disease heterogeneity. Chest 2002; 121:121S-126S.

(70)   Celli BR, Halbert RJ, Schau B. Population impact of different definitions of airway obstruction. European Respiratory Journal 2003; 22(268):273.

(71)   American Thoracic Society. Standards for the diagnosis and care of patients with chronic obstructive pulmonary disease. American Journal of Respiratory and Critical Care Medicine 1995; 152(S77):S120.

(72)   American Thoracic Society. Evaluation of impairment/disability secondary to respiratory disorders. American Review of Respiratory Disease 1986; 133:1205-1209.

(73)   Viegi G, Pedreschi M, Pistelli F, Di Pede F, Baldacci S, Carrozzi L et al. Prevalence of airways obstruction in a general population. European Respiratory Society vs American Thoracic Society Definition. Chest 2000; 117:339S-345S.

(74)   Lundback B, Lindberg A, Lindstrom M, Ronmark E, Jonsson AC, Jonsson E et al. Not 15 but 50% of smokers develop COPD? - report from the obstructive lung disease in northern Sweden studies. Respiratory Medicine 2003; 97:115-122.

(75)   Menezes AM, Macedo S, Gigante DP, Costa JD, Olinto MT, Fiss E et al. Prevalence and risk factors for chronic obstructive pulmonary disease according to symptoms and spirometry. COPD 2004; 1(2):173-179.

(76)   Shirtcliffe P, Weatherall M, Marsh S, Travers J, Hansell A, McNaughton A et al. COPD prevalence in a random population survey: a matter of definition. European Respiratory Journal 2007; 30(232):239.

(77)   Tzanakis N, Anagnostopoulou U, Filaditaki V, Christaki T, Siafakas N. Prevalence of COPD in Greece. Chest 2004; 125(892):900.

(78)   Halbert RJ, Isonaka S, George D, Iqbal A. Interpreting COPD prevalence estimates: what is the true burden of disease? Chest 2003; 123:1684-1692.

(79)   Buist AS, Vollmer WM, Sullivan SD, Weiss KB, Lee TA, Menezes AMB et al. The Burden of Obstructive Lung Disease Initiative (BOLD): Rationale and design. COPD 2005; 2(2).

(80)   Buist SA, McBurnie MA, Vollmer WM, Gillespie S. International variation in the prevalence of COPD (The BOLD Study): a population-based prevalence study. The Lancet 2007; 370(741):750.

(81)   Nizankowska-Mogilnicka E, Mejza F, Buist AS, Vollmer WM. Prevalence of COPD and tobacco smoking in Malopolska region-results from the BOLD study in Poland. Polskie Archiwum Medycyny Wewnetrznej 2007; 117(9):402-409.

(82)   Schirnhofer L, Lamprecht B, Vollmer WM, Allison MJ, Studnicka M, Jensen RL et al. COPD prevalence in Salzburg, Austria. Results from the Burden of Obstructive Lung Disease (BOLD) Study. Chest 2007; 131:29-36.

(83)   NHS. NHS core principles. [online] 2009. Available from http://www.nhs.uk/NHSEngland/thenhs [Accessed 18th July 2010].

(84)   Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. Methods for the Economic Evaluation of Health Care Programmes. 3rd ed. Oxford University Press, 2005.

(85)   ISPOR. Pharmacoeconomic guidelines around the world. [Online] 2007. Available from http://www.ispor.org/peguidelines/index.asp . [Accesssed 18th July 2010].

(86)   National Institute of Health and Clinical Excellence. A guide to NICE. [Online] 2005. Available from http://www.nice.org.uk/nicemedia/pdf/NewguidetoNICEApril2005.pdf [Accessed 18th July 2010].

(87)   National Institute for Health and Clinical Excellence. Guide to the methods of technology appraisal. 2008. London, National Institute for Health and Clinical Excellence.

(88)   Philips Z, Bojke L, Sculpher M, Claxton K, Golder S. Good practice guidelines for decision-analytic modelling in Health Technology Assessment: A review and consolidation of quality assessment. Pharmacoeconomics 2006; 24(4):355-371.

(89)   Drummond M, Sculpher M. Common Methodological Flaws in Economic
         Evaluations. Medical Care 2005; 43(7(S)):II-5-II-14.

(90)   Sculpher M, Claxton K, Drummond M, McCabe C. Whither trial-based economic
         evaluation for health care decision making? Health Economics 2006;(15):677-687.

(91)   Briggs AH, O'Brien BJ. The death of cost-minimisation analysis? Health
         Economics 2001; 10:179-184.

(92)   Oostenbrink JB, Rutten-van Molken MP, Monz BU, FitzGerald JM. Probabilistic
         Markov model to assess the cost-effectiveness of bronchodilator therapy in COPD
         patients in different countries. Value Health 2005; 8(1):32-46.

(93)   Jones PW Wilson KS. Cost-effectiveness of salmeterol in patients with chronic
         obstructive pulmonary disease: an economic evaluation. Respiratory Medicine
         2003; 97(1):20-26.

(94)   Gold M, Siegel JE, Russel LB, Weinstein MC. Cost-effectiveness in Health and
         Medicine. New York: Oxford University Press, 1996.

(95)   National Institute for Clinical Excellence. Guide to the Methods of Technology
         Appraisal. 2004. London, National Institute for Clinical Excellence.

(96)   Briggs AH, Sculpher M, Claxton K. Decision Modelling for Health Economic
         Evaluation. First ed. Oxford: Oxford University Press, 2006.

(97)   Jubran A, Gross N, Ramsdell J, Simonian R, Schuttenhelm K. Comparative cost-
         effectiveness analysis of theophylline and ipratropium bromide in chronic
         obstructive pulmonary disease. A three-centre study. Chest 1993; 103:678-684.

(98)   Le Lay A, Despiegel N, Francois C, Duru G. Can discrete event simulation be of
         use in modelling major depression? Cost Effectiveness & Resource Allocation
         2006; 4(19).

(99)   Glick HA, Doshi JA, Sonnad SS, Polsky D. Economic Evaluation in Clinical Trials.
         New York: Oxford University Press Inc, 2007.

(100)   Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health
         and medical practices. The New England Journal of Medicine 1977; 296(13):716-
         721.

(101)   Sculpher MJ. The role and estimation of productivity costs in economic evaluation.
         In: Drummond MF, McGuire A, editors. Theory and practice of economic
         evaluation in health. Oxford: Oxford University Press, 2001.

(102)   Sullivan SD. The burden of illness and economic evaluation for COPD. European Respiratory Journal 2003; 22(Suppl 43):7S-8S.

(103)   Siafakas NM, Vermeire P, Pride NB, Paoletti P, Gibson J, Howard P et al. Optimal assessment and management of Chronic Obstructive Pulmonary Disease (COPD). European Respiratory Journal 1995; 8:1398-1420.

(104)   Feeny D, Furlong W, Boyle M, Torrance G. Multi-attribute health status classifications systems: health utilities index. Pharmacoeconomics 1995; 7(490):502.

(105)   Brazier J. The SF-36 Health Survey Questionnaire - a tool for economists. Health Economics 1993; 2:213-215.

(106)   Ware JE, Kosinski M, Keller SD. A 12 item short-form health survey: construction of scales and preliminary tests of reliability and validity. Medical Care 1996; 34:220-233.

(107)   Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. Journal of Health Economics 2002; 21:271-292.

(108)   Goldberg DP. Manual of the general health questionnaire. Windsor, England: NFER Publishing, 1978.

(109)   Nelson EC, Landgraf JM, Hays RD, Wasson JH, Kirk JW. The functional status of patients: how can it be measured in physicians' offices? Medical Care 1990; 28(12):1111-1126.

(110)   Hunt SM, McEwan J. The development of a subjective health indicator. Sociology of Health and Illness 1983; 2:231-245.

(111)   Kaplan RM, Anderson J. A general health policy model: update and application. Health Services Research 1988; 23:203-235.

(112)   Bergner M, Bobbitt RA, Carter WB, Gilson BS. The Sickness Impact Profile: development and final revision of a health status measure. Medical Care 1981; 19(8):787-805.

(113)   Ware JE, Sherbourne CD. The MOS 36-Item Short-Form Health Survey (SF-36). Medical Care 1992; 30(6):473-483.

(114)   Larson JL, Kapella MC, Wirtz S, Covey MK, Berry J. Reliability and validity of the functional performance inventory in patients with moderate to severe chronic obstructive pulmonary disease. Journal of Nursing Measurement 1998; 6(1):55-73.

(115) Guyatt G, Berman LB, Townsend M. A measure of quality of life for clinical trials in chronic lung disease. Thorax 1987; 42:773-778.

(116) Hyland ME, Bott J, Singh S, Kenyon CAP. Domains, constructs and the development of the breathing problems questionnaire. Quality of Life Research 1994; 3(4):245-256.

(117) Tu SP, NcDonell MB, Spertus JA, Steele BG, Fihn SD. A new self-administered questionnaire to monitor health-related quality of life in patients with COPD. Chest 1997; 112:614-622.

(118) Piccirillo JF, Merritt MG, Richards ML. Psychometric and clinimetric validity of the 20 item sino-nasal outcome test (SNOT-20). Otolaryngology - head and neck surgery 2002; 126:41-47.

(119) Fletcher C. Standardised questionnaire on respiratory symptoms: a statement prepared and approved by the MRC Committee on the aetiology of chronic bronchitis (MRC breathlessness score). British Medical Journal 1960; 2:1665.

(120) Mahler DA, Weinberg DH, Wells CK, Feinstein AR. The measurement of dyspnea: contents, inter-observer agreement, and physiologic correlates of two new clinical indexes. Chest 1984; 85(751):758.

(121) Jones PW, Quirk FH, Baveystock CM, Littlejohns P. A self-complete measure of health status for Chronic Airflow Limitation. The American review of respiratory disease 1992; 145(6):1321-1327.

(122) Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. Vol 2 No 14. 1998. Basingstoke, Core Research on behalf of The National Coordinating Centre for Health Technology Assessment. Health Technology Assessment.

(123) von Neumann J, Morgenstern O. Theory of Games and Economic Behavior. Princeton,NJ: Princeton University Press, 1947.

(124) Garber AM. Advances in cost-effectiveness analysis of health interventions. In: Culyer A, Newhouse J, editors. Handbook of Health Economics. Amsterdam: Elsevier BV, 2000: 181-221.

(125) Claxton K, Sculpher M, McCabe C, Briggs AH, Akehurst R, Buxton M et al. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. Health Economics 2005; 14:339-347.

(126) Briggs AH. Probabilistic Analysis of Cost-Effectiveness odels: Choosing between Treatment Strategies for Gastroesophageal Reflux Disease. Medical Decision Making 2002; 22:290-308.

(127)   Baker CB, Johnsrud MT, Lynn Crismon JM, Rosenheck RA, Woods SW. Quantitative analysis of sponsorship bias in economic studies of antidepressants. British Journal of Psychiatry 2003; 183:498-506.

(128)   Bell CM, Urbach DR, Ray JG, Bayoumi A, Rosen AB, Greenberg D et al. Bias in published cost effectiveness studies: systematic review. British Medical Journal 2006; 332:699-703.

(129)   Friedman MSCM. Pharmacoeconomic evaluation of a combination of ipratropium plus albuterol compared with ipratropium alone and albuterol alone in COPD. Chest 1999; 115(3):635-641.

(130)   Ramsey SD, Patrick DL, Albert RK, Larson EB, Wood DE. The cost-effectiveness of lung transplantation. A pilot study. University of Washington medical centre lung transplantation study group. Chest 1995; 108:1594-1601.

(131)   Rutten-van Molken M, Van Doorslaer EK, Jansen MC, Kerstjens HA, Rutten FF. Costs and effects of inhaled corticosteroids and bronchodilators in asthma and chronic obstructive pulmonary disease. American Journal of Respiratory and Critical Care Medicine 1995; 151(4):975-982.

(132)   Al MJ, Koopmanschap NA, van Enckevort PJ, Geertsma A, Van der Bij W, de Boer WJ et al. Cost-effectiveness of lung transplantation in the Netherlands. Chest 1998; 113:124-130.

(133)   van den Boom GR. The cost effectiveness of early treatment with fluticasone propionate 250 microg twice a day in subjects with obstructive airway disease. Results of the DIMCA program. American Journal of Respiratory and Critical Care Medicine 2001; 164(11):2057-2066.

(134)   Anyanwu AC, McGuire A, Rogers CA, Murday AJ. An economic evaluation of lung transplantation. The Journal of Thoracic and Cardiovascular Surgery 2002; 123(3):411-420.

(135)   Ramsey SD, Berry K, Etzioni R, Kaplan RM, Sullivan SD, Wood DE et al. Cost effectiveness of lung-volume-reduction surgery for patients with severe emphysema. New England Journal of Medicine 2003; 348(21):2092-2102.

(136)   Ayres JG, Price MJ, Efthimiou J. Cost-effectiveness of fluticasone propionate in the treatment of chronic obstructive pulmonary disease: a double-blind randomized, placebo-controlled trial. Respiratory Medicine 2003; 97(3):212-220.

(137)   Hogan TJ, Geddes R, Gonzalez E. An economic assessment of inhaled formoterol dry powder versus ipratropium bromide pressurized metered dose inhaler in the

treatment of chronic obstructive pulmonary disease. Clinical Therapeutics 2003; 25(1):285-297.

(138) Borg S, Ericsson A, Wedzicha J, Gulsvik A, Lundback B, Donaldson GC et al. A Computer Simulation Model of the Natural History and Economic Impact of Chronic Obstructive Pulmonary Disease. Value in Health 2004; 7(2).

(139) Groen H, van der BW, Koeter GH, TenVergert EM. Cost-effectiveness of lung transplantation in relation to type of end-stage pulmonary disease. Am J Transplant 2004; 4(7):1155-1162.

(140) Oostenbrink JB, Rutten-van Molken MPMH, Al MJ, Van Noord JA, Vincken W. One-year cost-effectiveness of tiotropium versus ipratropium to treat chronic obstructive pulmonary disease. European Respiratory Journal 2004; 23(2).

(141) Sin DD, Golmohammadi K, Jacobs P. Cost-effectiveness of inhaled corticosteroids for chronic obstructive pulmonary disease according to disease severity. American Journal of Medicine 2004; 116(5):325-331.

(142) Gagnon YM, Levy AR, Spencer MD, Hurley JS, Frost FJ, Mapel DW et al. Economic evaluation of treating chronic obstructive pulmonary disease with inhaled corticosteroids and long-acting beta2-agonists in a health maintenance organization. Respiratory Medicine 2005; 99(12):1534-1545.

(143) Lofdahl CG, Ericsson A, Svensson K, Andreasson E. Cost effectiveness of budesonide/formoterol in a single inhaler for COPD compared with each monocomponent used alone. Pharmacoeconomics 2005; 23(4):365-375.

(144) Spencer M, Briggs AH, Grossman RF, Rance L. Development of an economic model to assess the cost-effectiveness of treatment interventions for chronic obstructive pulmonary disease. Pharmacoeconomics 2005; 23(6):619-637.

(145) Briggs AH, Lozano-Ortega G, Spencer S, Bale G, Spencer MD, Burge PS. Estimating the cost-effectiveness of fluticasone propionate for treating chronic obstructive pulmonary disease in the presence of missing data. Value in Health 2006; 9(4):227-235.

(146) Maniadakis N, Tzanakis N, Fragoulakis V, Hatzikou M, Siafakas N. Economic evaluation of tiotropium and salmeterol in the treatment of chronic obstructive pulmonary disease (COPD) in Greece. Current Medical Reseacrh Opinion 2006; 22(8).

(147) Ramsey SD, Shroyer AL, Sullivan SD, Wood DE. Updated evaluation of the cost-effectiveness of Lung Volume Reduction Surgery. Chest 2007; 131(3):823-832.

(148) Rutten-van Molken MPMH, Oostenbrink JB, Miravitlles M, Monz BU. Modelling the 5-year cost effectiveness of tiotropium, salmeterol and ipratropium for the treatment of chronic obstructive pulmonary disease in Spain. European Journal of Internal Medicine 2007; 8:123-135.

(149) Chuck A, Jacobs P, Mayers I, Marciniuk D. Cost-effectiveness of combination therapy for Chronic Obstructive Pulmonary Disease. Canadian Respiratory Journal; 2008; 15(8):437-443.

(150) Earnshaw SR, Wilson MR, Dalal AA, Chambers MG, Jhingran P, Stanford R et al. Cost effectiveness of fluticasone propionate/ salmeterol (500/50mg) in the treatment of COPD. Respiratory Medicine 2008; 103:12-21.

(151) Najafzadeh M, Marra CA, Sadatsafavi M, Aaron SD, Sullivan SD, Vandemheen KL et al. Cost effectiveness of therapy with combinations of long acting bronchodilators and inhaled steroids for treatment of COPD. Thorax 2008; 63:962-967.

(152) Blough DK, Ramsey S, Sullivan SD, Yusen R. The impact of using different imputational methods for missing quality of life scores on the estimation of the cost effectiveness of lung volume reduction surgery. Health Economics 2009; 18:91-101.

(153) Briggs A, Glick HA, Lozano-Ortega G, Spencer M, Calverley PM, Jones PW et al. Is treatment with ICS and LABA cost-effective for COPD? Multinational economic analysis of the TORCH study. European Respiratory Journal 2010; 35: 532-539

(154) Oba Y. Cost effectiveness of long term oxygen therapy for Chronic Obstructive Pulmonary Disease. The American Journal of Managed Care 2009; 15(2):97-104.

(155) Dahl R, Greefhorst LA, Nowak D. Inhaled formoterol dry powder versus ipratropium bromide in chronic obstructive pulmonary disease. American Journal of Respiratory and Critical Care Medicine 2001; 164(5):778-784.

(156) Lin DY, Feuer EJ, Etzioni R, Wax Y. Estimating medical costs from incomplete follow-up data. Biometrics 1997; 53(2):419-434.

(157) Sin DD, Stafinski T, Ng YC. The impact of chronic obstructive pulmonary disease on work loss in the United States. Critical Care Medicine 2002; 165:704-707.

(158) Briggs AH, Ades AE, Price MJ. Probabilistic sensitivity analysis for decision tress with multiple branches: the use of the Dirichlet distribution in a Bayesian framework. Medical Decision Making 2003; 23:341-350.

(159)   Calverley PM, Pauwels R, Vestbo, Jones P, Pride N, Gulsvik A et al. Combined salmeterol and fluticasone in the treatment of chronic obstructive pulmonary disease: a randomised controlled trial. Lancet 2003; 361:449-456.

(160)   Oba Y. Cost-Effectiveness of salmeterol, fluticasone and combination therapy for COPD. American Journal of Managed Care 2009; 15(4):226-232.

(161)   Rutten-van Molken M, Lee TA, Authors FN, Lee TA. Economic modelling in chronic obstructive pulmonary disease. Proceedings of the American Thoracic Society 2006; 3(7):630-634.

(162)   Baltussen R, Leidl R, Ament A. Real World Designs in Economic Evaluation: Bridging the gap between Clinical Research and Policy-Making. Pharmacoeconomics 1999; 16(5 (1)):449-458.

(163)   Halpin DM, Authors FN. Health economics of chronic obstructive pulmonary disease. Proceedings of the American Thoracic Society 2006; 3(3):227-233.

(164)   Claxton K, Sculpher M, Drummond M. A rational framework for decision making by the National Institute for Clinical Excellence (NICE). The Lancet 2002; 360:711-715.

(165)   Gross N. Outcome measures for COPD Treatments: A critical Evaluation. COPD 2004; 1(1):41-57.

(166)   Hole DJ, Watt GCM, Davey-Smith G, Hart CL, Gillis CR, Hawthorne VM. Impaired lung function and mortality risk in men and women: findings from the Renfrew and Paisley prospective population study. British Medical Journal 1996; 313:711-715.

(167)   Hart CL, MacKinnon PL, Watt GCM, Upton MN, McConnachie A. The Midspan studies. International Journal of Epidemiology 2005; 34:28-34.

(168)   GPRD Division of the Medicines and Healthcare products Regulatory Agency (MHRA). The General Practice Research Database. www.gprd.com . 2007.

(169)   The University of Nottingham, EMIS National User Group, EMIS. Qresearch. www.qresearch.org . 2007.

(170)   THIN. The Health Improvement Network website. [Online]. Available from: www.thin-uk.com [Accessed 15th June 2010]. 2010.

(171)   IMS. A brief description of IMS databases. http://research.imhhealth.com/databases . 2002.

(172) Raza Z, Holdworth RJ, McCollum PT. Accuracy of the recording of operative events by the Scottish morbidity record 1 (SMR1) for a teaching hospital vascular unit. Journal of the Royal College of Surgeons of Edinburgh 1999; 44(96):98.

(173) The Information Centre for health and social care by the National Centre for Health Outcomes Development. Clinical and Health Outcomes Knowledge base: Scottish Morbidity Record 1. [Online]. Available from: www.nchod.nhs.uk . [Accessed on 24th September 2007]. 2003

(174) The Information Centre for health and social care by the National Centre for Health Outcomes Development. Clinical and Health Outcomes Knowledge base: Hospital Episode Statistics. [Online]. Available from: www.nchod.nhs.uk . [Accessed on 24th September 2007].2003.

(175) North East Public Health Observatory. Hospital Episode Statistics. [Online]. Available from: www.nepho.org.uk . [Accessed on 24th September 2007]. 2004.

(176) ISD Scotland. The Scottish Health Surveys. [Online]. Available from: www.isdscotland.org. [Accessed on 24th September 2007]. 2006.

(177) General Register Office. Classification of Occupations. London: HMSO, 1966.

(178) Hanlon P, Walsh D, Whyte BW, Scott SN, Lightbody P, Gilhooly LM. Hospital use by an ageing cohort: an investigation into the association between biological, behavioural and social risk markers and subsequent hospital utilization. Journal of Public Health Medicine 1998; 20(4):467-476.

(179) Predicted Forced Expiratory Volume in 1 second (FEV1). [Online]. Available from: http://www.medal.org/visitor/www%5CActive%5Cch8%5Cch8.01%5Cch8.01.01.a spx . [Accessed on 10th October 2007]. 2008.

(180) Pellegrino R, Viegi G, Brusasco V, Crapo RO, Burgos F, Casaburi R et al. Interpretative strategies for lung function tests. European Respiratory Journal 26, 948-968. 2005.

(181) Stocks J, Quanjer PH. Reference values for residual volume, functional residual capacity and total lung capacity. ATS workshop on lung volume measurements. Official statement of the European Respiratory Society. European Respiratory Journal 8, 492-506. 1995.

(182) Stata Statistical Software: Release 10.0. College Station, TX: StataCorp LP, 2008.

(183) Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. Journal of the American Statistical Association 1958; 53(282):457-481.

(184) Ahmed FE, Vos PW, Holbert D. Modeling survival in colon cancer: a methodological review. Molecular Cancer 2007; 6(15).

(185) Clark TG, Bradburn MJ, Love SB, Altman DG. Survival analysis part 1: basic concepts and first analyses. British Journal of Cancer 2003; 89:232-238.

(186) Korn EL, Graubard BI, Midthune D. Time-to-event analysis of longitudinal follow-up of a survey: choice of the time scale. American Journal of Epidemiology 1997; 145:72-80.

(187) Kirkwood BR, Sterne JAC. Essential Medical Statistics. 2nd ed. Blackwell Publishing, 2006.

(188) Bradburn MJ, Clark TG, Love SB, Altman DG. Survival analysis part II: multivariate data analysis - an introduction to concepts and methods. British Journal of Cancer 2003; 89:431-436.

(189) Kay R. An explanation of the hazard ratio. Pharmaceutical Statistics 2004; 3:295-297.

(190) Concato J, Feinstein AR, Holford TR. The risk of determining risk with multivariable models. Annals of Internal Medicine 1993; 118(201):210.

(191) Shepherd BE. The cost of checking proportional hazards. Statistics in Medicine 208; 27:1248-1260.

(192) Schoenfeld D. Partial residuals for the proportional hazards regression model. Biometrika 1982; 69(1):239-241.

(193) Bradburn MJ, Clark TG, Love SB, Altman DG. Survival analysis part III: multivariate data analysis-choosing a model and assessing its adequacy and fit. British Journal of Cancer 2003; 89:605-611.

(194) Christensen E. Multivariate survival analysis using Cox's regression model. Hepatology 1987; 7(6):1346-1358.

(195) Roberts SD, Farber MO, Knox KS, Phillips GS, Bhatt NY, Mastronarde JG et al. FEV1/FVC ratio of 70% misclassifies patients with obstruction at the extremes of age. Chest 2006; 130:200-206.

(196) GOLD. Global Initiative for Chronic Obstructive Lung Disease home page. [Online]. Available from: http://www.goldcopd.com. [Accessed on 28th August 2008]. 2008

(197) Lindberg A, Jonsson A-C, Ronmark E, Lundgren R, Larsson LG, Lundback B. Prevalence of Chronic Obstructive Pulmonary Disease according to BTS, ERS,

GOLD and ATS criteria in relation to Doctor's diagnosis, age, gender, and smoking habits. Respiration 2005; 72:471-479.

(198) Lindberg A, Bjerk-Backlund A, Ronmark E, Larsson LG, Lundback B. Prevalence and under diagnosis of COPD by disease severity and the attributable fraction of smoking. Report from the Obstructive Lung Disease in Northern Sweden studies. Respiratory Medicine 2006; 100:264-272.

(199) Mannino DM, Buist AS, Petty TL, Enright PL, Redd SC. Lung function and mortality in the United States: Examination Survey follow up study data from the First National Health and Nutrition. Thorax 2003; 58:388-393.

(200) Currie GP, Legg JS. Diagnosis. British Medical Journal 2006; 332:1261-1263.

(201) Starkie HJ, Briggs AH. OHE Commission on NHS Productivity. Outcome measures for the assessment of treatment results in Chronic Obstructive Pulmonary Disease. 2007.

(202) Rodriguez-Roisin R, Authors FN. Toward a consensus definition for COPD exacerbations. Chest 2000; 117(5 Suppl 2):398S-401S.

(203) Rutten van-Molken MP, Feenstra TL, Authors FN, Feenstra TL. The burden of asthma and chronic obstructive pulmonary disease: data from The Netherlands. Pharmacoeconomics 2001; 19 Suppl 2:1-6.

(204) Jensen HH, Godtfredsen NS, Lange P, Vestbo J. Potential misclassification of causes of death from COPD. European Respiratory Journal. 2006; 28:781-785.

(205) Kind P, Dolan P, Gudex C, Williams A. Variations in population health status: results from a United Kingdom national questionnaire survey. British Medical Journal 1998; 316:736-741.

(206) Kind P, Hardman G, Leese B. Measuring health status: information for primary care decision making. Health Policy 2005; 71:303-313.

(207) Dolan P. Modelling valuations for EuroQol health states. Medical Care 1997; 35(11):1095-1108.

(208) Williams A. The measurement and valuation of health: a chronicle. 1995. York, Centre for Health Economics.

(209) Carpenter J, Kenward M. Missing Data. [Online]. Available from: www.missingdata.org.uk . [Accessed on 4th February 2009].2009.

(210) Wooldridge JM. Introductory Econometrics: a modern approach. 2E ed. Ohio: South western, 2003.

(211) StataCorp. STATA base reference manual A-J. Release 9 ed. 2005.

(212) Manning WG, Mullaly J. Estimating log models: to transform or not to transform? Journal of Health Economics 2001; 20:461-494.

(213) Pearson ES, Please NW. Relationship between the shape of population distribution and the robustness of four simple test statistics. Biometrika 1975; 62:223-241.

(214) Pregibon D. Goodness of link tests for generalized linear models. Applied Statistics 1980; 29(15):24.

(215) Hosmer Jn.D.W, Lemeshow S. Applied Logistic Regression. 2nd ed. New York: John Wiley & Sons, Inc, 2000.

(216) Meguro M, Jones PW. Elicitation of utility weights for a COPD specific preference measure. [unpublished] 2007.

(217) Vardeva K, Spencer MD, Punekar Y. Development of preference based EQ-5D utility values for the St George's Respiratory Questionnaire - chronic obstructive pulmonary disease (SGRQ-COPD). Value in Health 10[6], A311. 2007.

(218) Brazier J, Yang Y, Tsuchiya A. Review of methods for mapping between condition specific measures onto generic measures of health. 2007. www.ohe.org, OHE Commission.

(219) Brazier J, Yang Y, Tsuchiya A, Rowen D. A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. European Journal of Health Economics 2009; 11(2): 215-225.

(220) Wu E, Mulani P, Farrell MH, Sleep D. Mapping FACT-P and EORTC QLQ-C30 to patient health status measured by EQ-5D metastatic hormone-refractory prostate cancer patients. Value in Health 2007; 10(5):408-414.

(221) Yang M, Dubois D, Kosinski M, Sun X, Gajria K. Mapping MOS Sleep Scale scores to SF-6D utility index. Current Medical Research Opinion 2007; 23(9):2269-2282.

(222) Buxton MJ, Lacey LA, Feagan BG, Niecko T, Miller DW, Townsend RJ. Mapping from disease-specific measures to utility: an analysis of the relationships between the Inflammatory Bowel Disease Questionnaire and Crohn's Disease Activity Index in Crohn's disease and measures of utility. Value in Health 2007; 10(3):214-220.

(223) Grootendorst P, Marshall D, Pericak D, Bellamy N, Feeny D, Torrance G. A Model to Estimate Health Utilities Index Mark 3 Utility Scores from WOMAC Index

Scores in Patients with Osteoarthritis of the Knee. Journal of Rheumatology 2007; 34:534-542.

(224) Bansback N, Marra CTA, Anis A, Guh D, Hammond T, Brazier J. Using the Health Assessment Questionnaire to Estimate Preference-Based Single Indices in Patients With Rheumatoid Arthritis. Arthritis & Rheumatism (Arthritis Care & Research) 2007;(963):971.

(225) Brennan DS, Spencer AJ. Mapping oral health realted quality of life to generic health state values. BMC Health Services Research 2006; 6(Art. no. 96).

(226) Sullivan P, Ghushchyan V. Mapping the EQ-5D Index from the SF-12: US General Population Preferences in a Nationally Representative Sample. Medical Decision Making 2006; 26(4):401-409.

(227) Longworth L, Buxton MJ, Sculpher M, Smith DH. Estimating utility data from clinical indicators for patients with stable angina. European Journal of Health Economics 2005; 6:347-353.

(228) Franks P, Lubetkin EL, Gold MR, Tancredi DJ, Jia H. Mapping the SF-12 to the EuroQol EQ-5D Index in a National US Sample. Medical Decision Making 2004; 24(3):247-254.

(229) Brazier J, Kolotkin RD, Crosby RD, Williams R. Estimating a Preference-Based Single Index for the Impact of Weight on Quality of Life-Lite (IWQOL-Lite) Instrument from the SF-6D. Value in Health 2004; 7(4):490-498.

(230) Lawrence WF, Fleishman JA. Predicting EuroQoL EQ-5D Preference Scores from the SF-12 Health Survey in a Nationally Representative Sample. Medical Decision Making 2004; 24(2):160-169.

(231) Sengupta N, Nichol MB, Wu J, Globe D. Mapping the SF-12 to the HUI3 and VAS in a managed care population. Medical Care 2004; 42(9):927-937.

(232) Franks P, Lubetkin EL, Gold MR, Tancredi DJ. Mapping the SF-12 to preference-based instruments: convergent validity in a low-income, minority population. Medical Care 2003; 41(11):1277-1283.

(233) Nichol NB, Sengupta N, Globe D. Evaluating Quality-Adjusted Life Years. Medical Decision Making 2001; 21(2):105-112.

(234) Brazier J, Yang Y, Tsuchiya A, Rowen D. A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. The European Journal of Health Economics 2010; 11(2):215-225.

(235) Lachenbruch PA. Analysis of data with excess zeros. Statistical Methods in Medical Research 2002; 11:297-302.

(236) Lipscomb J, Ancukiewicz M, Parmigiani G, Hasselblad V, Samsa G, Matchar DB. A comparison of alternative models applied to stroke. Medical Decision Making 1998; 18(2):S39-S66.

(237) McKenzie L, van der Pol M. Mapping the EORTC QLQ C-30 onto the EQ-5D Instrument: The potential to estimate QALYs without Generic Preference Data. Value in Health 2009; 12(1):167-171.

(238) Briggs AH, Bousquet J, Wallace MV, Busse WW, Clark TJH, Pederson SE et al. Cost-effectiveness of asthma control: an economic appraisal of the GOAL study. Allergy: European Journal of Allergy and Clinical Immunology 2006; 61(5):531-536.

(239) Briggs A, Glick HA, Lozano-Ortega G, Spencer M, Calverley PM, Jones PW et al. Is treatment with ICS and LABA cost-effective for COPD? Multinational economic analysis of the TORCH study. European Respiratory Journal 2010; 35(532):539.

(240) Halpin DM. Health economics of chronic obstructive pulmonary disease. Proceedings of the American Thoracic Society 2006; 3(3):227-233.

(241) Donaldson GC, Seemungal TA, Bhowmik A, Wedzicha JA. Relationship between exacerbation frequency and lung function decline in Chronic Obstructive Pulmonary Disease. Thorax 2002; 57(847):852.

(242) Miravitlles M, Guerrero T, Mayordomo C, Sanchez-Agudo L, Nicolau F, Lois Segu J. Factors associated with increased risk of exacerbation and hospital admission in a cohort of ambulatory COPD patients: a multiple logistic regression anlaysis. Respiration 2000; 67:495-501.

(243) Joint Health Surveys Unit. Health Survey for England 2007 Latest Trends. 16-12-2008.

(244) TORCH steering Committee. TORCH Study report. GlaxoSmithKline 2006.

(245) Calverley PM, Rabe KF, Goehring U-M, Kristiansen S, Fabbri LM, Martinez F. Roflumilast in symptomatic chronic obstructive pulmonary disease: two randomised clinical trials. The Lancet 2009; 374:685-694.

(246) Celli B, Thomas NE, Anderson JA, Ferguson GT, Jenkins CR, Jones PW et al. Effect of pharmacotherapy on rate of decline of lung function in Chronic Obstructive Pulmonary Disease. American Journal of respiratory and Critical Care Medicine 2008; 178:332-338.

(247)   Fabbri LM, Calverley PM, Izquierdo-Alonso JL, Bundschuh DS, Brose M,
         Martinez FJ et al. Roflumilast on moderate to severe chronic obstructive pulmonary
         disease treated with long acting bronchodilators: two randomised clinical trials. The
         Lancet 2009; 374:695-703.

(248)   Calverley PM, Pauwels R, Vestbo, Jones P, Pride N, Gulsvik A et al. Combined
         salmeterol and fluticasone in the treatment of chronic obsructive pulmonary disease:
         a randomised controlled trial. Lancet 2003; 361:449-456.

(249) National Clinical Guideline Centre. (2010) Chronic obstructive pulmonary disease:
         management of chronic obstructive pulmonary disease in adults in primary and
         secondary care. London: National Clinical Guideline Centre. Available from:
         http://guidance.nice.org.uk/CG101/Guidance/pdf/English

(250)   Health Status Research Team. St George's respiratory Questionnaire population
         survey.[Online] 2007. Available from: www.healthstatus.sgul.ac.uk . [Accessed on
         20th July 2010].

(251)   Brazier J, Roberts J, Tsuchiya A, Busschbach J et al. A comparison of the EQ-5D
         and SF-6D across seven patient groups. Health Economics 2004; 13(9):873-884.