



THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

Title: Modelling the prevalence, healthcare costs and number of deaths in chronic obstructive pulmonary disease in England and Scotland

Author: Susannah McLean

**PhD
The University of Edinburgh
2015**

Declaration

In accordance with regulation 25 I declare that the thesis has been composed by myself and that the work is my own, and has not been submitted for any other degree or professional qualification.

Signed:.....

Susannah McLean

Abstract

Introduction

Chronic obstructive pulmonary disease (COPD) has emerged as a major policy focus for health systems throughout Western Europe. This reflects the increased prevalence, associated healthcare utilisation and costs of COPD, and the potential to substantially improve outcomes through achieving reductions in smoking. The aim of this PhD was to develop projections for the prevalence, healthcare costs and number of deaths in people with COPD in England and Scotland over a 20-year horizon (i.e. from 2011 to 2030).

Methods

I undertook a phased programme of work, which began with a systematic review of the published and unpublished literature to identify models that were suitable for estimating and/or projecting the prevalence and disease and economic burden from COPD. This involved searching Medline, Embase, CAB Abstracts, World Health Organization (WHO) Library and Information Services and WHO Regional Indexes, and Google over the time period 1980-2013. The models were then critically appraised for their quality of reporting. From these, I selected the Dutch Model developed by Erasmus University for generating projections. Suitable data sources from both England and Scotland were identified, sourced and carefully processed in order to run the modelling exercises. Rates of incidence and prevalence were calculated using English and Scottish healthcare datasets and population data were obtained from the Office for National Statistics (ONS) and the General Register Office for Scotland (GROS). Relative risks for all-cause mortality among people with COPD were calculated from the Clinical Practice Research Datalink and mortality data were obtained from the ONS and GROS. The Model was thus adjusted to apply to England and Scotland. I then travelled to the Netherlands to work with the developers of the Dutch Model and ran a baseline model and an array of sensitivity analyses with modified inputs to the Model. Finally, my Rotterdam colleagues calculated uncertainty intervals for some of the estimates using probabilistic analysis.

Results

Using the probabilistic means and uncertainty intervals, in England, the modelled prevalence of diagnosed COPD among males of all ages in 2011 was 1.8% (95% uncertainty interval 1.8-1.9) increasing to 2.0% (1.7-2.1) by 2030. In females, in England, the baseline estimate was 1.8% (1.7-1.8) in 2011 increasing to 2.4% (2.0-2.6) in 2030. In Scotland, the modelled prevalence among males was 1.9% (1.8-1.9) in 2011 and this was projected to stay the same

at 1.9% (1.7-2.2) by 2030. In females in Scotland, the estimated prevalence was 2.2% (2.1-2.3) in 2011 and was projected to increase to 2.5% (2.1-2.7) in 2030. Using the Model I estimated that overall in 2011 there were a total of 952,000 (941,000-966,000) people with diagnosed COPD in England and 106,000 (103,000-110,000) in Scotland and that these numbers would increase to 1,325,000 (1,117,000-1,408,000) in England in 2030 and 125,000 (113,000-136,000) in Scotland in 2030, respectively. The greatest increase in COPD was projected to be in females over 65 years of age in both countries.

The total annual direct healthcare costs of COPD in England were projected to increase from £1.60 (95% uncertainty interval 1.18-2.5) billion in 2011 to £2.35 (1.85-3.08) billion in 2030. In Scotland, costs were projected to increase from £170 (128-268) million in 2011 to £210 (165-274) million in 2030. These costs were calculated in terms of 2011 costs without the application of any economic trends (i.e. no annual increase applied for inflation).

The number of deaths among people with COPD in England was estimated to be 99,000 (93,000-129,000) in 2011, increasing to 129,000 (126,000-133,000) in 2030. In Scotland there were estimated to be 10,000 (9,000-12,000) deaths in 2011, increasing to 14,000 (13,000-15,000) in 2030.

The Dutch Model demonstrated a 39% increase in the number of people with COPD in England and a 17% increase in Scotland between 2011 and 2030. It provided an estimate of a 30% increase in deaths among people with COPD in England and of a 43% increase in Scotland. Overall, there was a projected 46% increase in the direct healthcare costs required to care for people with COPD in England and a 23% increase in Scotland between 2011 and 2030. The reasons for these differences are largely due to higher COPD-related excess mortality in Scotland and to differences in the data used for populating the model in both countries.

Conclusions

There are likely to be substantial increases in the number of people with COPD, associated morbidity, direct healthcare costs and mortality in both England and Scotland over the next two decades. These increases in numbers will predominantly occur in females over 65 years of age and are likely to have substantial societal impact in terms of organising the health and social care for this frail population.

Table of Contents

Abstract.....	3
Introduction	3
Methods.....	3
Results.....	3
Conclusions	4
List of Tables and Figures.....	13
Abbreviations.....	16
Acknowledgements.....	20
Thesis overview.....	21
Chapter 1 Background, rationale and aims.....	23
1.1 Background	23
1.2 Rationale for a PhD in modelling	23
1.3 Rationale for a PhD in COPD	24
1.4 Rationale for modelling in England and Scotland, 2011 to 2030.....	24
1.5 Aim	25
1.6 Phase one – systematic review	25
1.7 Phase two – modelling study	25
1.8 Conclusion to Chapter 1.....	26
Chapter 2: Basic concepts in COPD.....	27
2.1 Introduction	27
2.2 Definitions of COPD.....	27
2.2.1 Historical context of COPD definitions.....	27
2.2.2 Definition from the 2010 National Institute for Health and Care Excellence (NICE) guideline for COPD.....	32
2.2.3 Definition from the 2013 updated Global Initiative for Chronic Obstructive Lung Disease (GOLD) guidelines	33

2.2.4	How the definitions of COPD relate to diagnosis.....	33
2.2.5	Diagnosis of COPD in elderly patients.....	35
2.2.6	Co-existing asthma in COPD.....	35
2.2.7	Defining COPD for the purpose of this thesis.....	36
2.3	Epidemiology of COPD.....	37
2.3.1	Global epidemiology of COPD.....	37
2.3.2	Epidemiology of COPD in England.....	41
2.3.3	Epidemiology of COPD in Scotland.....	42
2.3.4	Conclusions to epidemiology section.....	43
2.4	Morbidity, mortality and costs of COPD.....	44
2.4.1	Morbidity from COPD.....	44
2.4.2	Mortality from COPD.....	45
2.4.3	Costs of COPD.....	47
2.5	Aetiology of COPD.....	47
2.5.1	Prenatal and postnatal lung growth.....	47
2.5.2	Bronchial hyper-responsiveness.....	48
2.5.3	Effects of childhood respiratory tract illness.....	48
2.5.4	Smoking and COPD.....	48
2.5.5	Phenotypic variation.....	49
2.5.6	Smoking and genetics.....	49
2.5.7	Smoking and the immune response.....	50
2.6	Pathophysiology of COPD.....	51
2.7	Clinical features.....	52
2.7.1	Frequent exacerbator phenotype.....	53
2.8	Prognosis.....	53
2.8.1	FEV1 decline.....	54
2.8.2	Computer Tomography (CT) defined lung density over time in COPD.....	56

2.8.3 Bronchodilator reversibility	56
2.8.4 Systemic inflammation.....	56
2.8.5 Exercise capacity and the BODE index.....	57
2.8.6 Co-morbidities.....	57
2.8.7 The DOSE Index.....	57
2.8.8 Assessment of COPD severity as per GOLD guidelines	58
2.8.9 Life expectancy in COPD	58
2.9 Management of stable COPD.....	58
2.9.1 Smoking cessation.....	59
2.9.2 Inhaled therapies	59
2.9.3 Pulmonary rehabilitation	60
2.9.4 Oxygen therapy.....	60
2.10 Management of COPD exacerbations.....	60
2.10.1 Structure of care for people with COPD exacerbations.....	61
2.11 Patient perspectives in COPD	61
2.12 Conclusion to Chapter 2.....	62
Chapter 3: Reusing healthcare data for research	64
3.1 Introduction	64
3.2 Healthcare data capture	64
3.2.1 Healthcare data capture in primary care.....	64
3.2.2 Healthcare data capture in hospital settings.....	68
3.2.3 Coded and narrative data	69
3.3 Validity of healthcare data for research	73
3.3.1 Validity of routine data used in this thesis	74
3.4 Data reuse for research	75
3.4.1 Data reuse in linked data studies.....	76
3.4.2 Data mining.....	77

3.4.3 Data reuse for trials.....	77
3.4.4 Data reuse for pharmacovigilance	78
3.5 Professional and public concerns regarding data privacy.....	79
3.5.1 Governance for eHealth data linkages and reuse	80
3.5.2 Governance relevant to this PhD	81
3.6 Conclusion to Chapter 3	81
Chapter 4: Concepts in modelling	83
4.1 Introduction.....	83
4.2 Introduction to Markov modelling.....	83
4.2.1 Cohort simulation.....	85
4.2.2 Individual simulation	86
4.2.3 Comparing cohort and individual simulation	86
4.3 Alternative model formats	87
4.3.1 Demographic models	87
4.3.2 Risk factor models	87
4.3.3 Time trend models	87
4.3.4 Other models.....	87
4.4 Outputs of health economics modelling	87
4.4.1 Quality adjusted life years (QALYs)	88
4.4.2 Disability adjusted life years (DALYs)	89
4.4.3 Advantages and disadvantages of QALYs and DALYs.....	89
4.5 Conclusion to Chapter 4	89
Chapter 5: Systematic review of the literature for models that calculate the prevalence and burden of chronic obstructive pulmonary disease (COPD).....	91
5.1 Introduction.....	91
5.2 Methods	91
5.2.1 Protocol	91

5.2.2 Eligibility criteria.....	91
5.2.3 Information sources and study selection.....	92
5.2.4 Data extraction.....	92
5.2.5 Quality appraisal framework.....	93
5.2.6 Scoring on the quality of reporting framework	96
5.2.7 Synthesis of results	96
5.3 Results.....	97
5.3.1 Introduction to tables of results	98
5.3.2 Narrative results.....	127
5.4 Selection of model for further study	132
5.5 Conclusion to Chapter 5.....	132
Chapter 6: Model description and data requirements	134
6.1 Introduction	134
6.2 Dutch Model description	134
6.3 Modelling data input overview.....	136
6.4 Model data inputs and sources.....	138
6.4.1 General population data	138
6.4.2 Incidence of COPD from general practice data.....	138
6.4.3 Prevalence of COPD	143
6.4.4 Proportions of the population who smoke.....	149
6.4.5 Start smoking proportions	152
6.4.6 Stop smoking proportions.....	152
6.4.7 Restart smoking proportions	152
6.4.8 Relative risks of smokers and non-smokers.....	153
6.4.9 Lung function decline rate	153
6.4.10 COPD-related maintenance costs	153
6.4.11 Mean utility scores by COPD severity stage	158

6.4.12 Excess mortality.....	158
6.5 Strengths and limitations of datasets	161
6.6 Conclusion to Chapter 6	162
Chapter 7: Modelling process and results.....	164
7.1 Introduction.....	164
7.2 Modelling process	164
7.3 Results of modelling base case scenario	165
7.3.1 England: COPD population results	165
7.3.2 England: overall prevalence results.....	166
7.3.3 England: age-related prevalence.....	166
7.3.4 England: direct healthcare costs from COPD	168
7.3.5 England: number of deaths	168
7.3.6 Scotland: COPD population	169
7.3.7 Scotland: overall prevalence results	170
7.3.8 Scotland: age-specific prevalence	170
7.3.9 Scotland: direct healthcare costs from COPD	171
7.3.10 Scotland: number of deaths results	172
7.4 Results of sensitivity analyses	172
7.4.1 Sensitivity analyses for England	173
7.4.2 Sensitivity analyses for Scotland	174
7.5 Uncertainty limits for results.....	175
7.5.1 Point estimates and overall prevalence	176
7.6 Conclusion to Chapter 7	178
Chapter 8: Discussion	179
8.1 Introduction.....	179
8.2 Statement of principal findings	179
8.2.1 Principal findings from the systematic review	179

8.2.2 Principal findings of modelling study.....	180
8.3 Comparison of modelling results from England with results from Scotland.....	180
8.3.1 Female COPD predominance, differences between England and Scotland, evidence from smoking habits.....	181
8.3.2 Female COPD predominance, differences between England and Scotland, differences in baseline input prevalence.....	182
8.3.3 Differences between England and Scotland, differences in relative increases over 20 years.....	183
8.4 Strengths and limitations.....	185
8.4.1 Strengths and limitations of the systematic review.....	185
8.4.2 Strengths and limitations of modelling study.....	187
8.5 Interpretation of the results of this study in the context of previous research.....	188
8.5.1 The shape of the smoking epidemic.....	188
8.5.2 Comparison with UK data from the Global Burden of Disease Study 2010.....	191
8.5.3 The lag effect of smoking: ex-smokers developing COPD.....	192
8.5.4 Comparison with results from the Netherlands.....	193
8.6 Interpreting this research: implications for clinicians and policymakers.....	195
8.6.1 Five, 10 and 20 year projections.....	195
8.6.2 Implications of the age of the COPD population.....	195
8.6.3 Implications in terms of NHS spending for England.....	196
8.6.4 Implications in terms of NHS spending for Scotland.....	197
8.6.5 Implications in terms of COPD inpatient care.....	198
8.6.6 Policy developments in COPD.....	199
8.7 Unanswered questions and future research.....	201
8.7.1 The COPD population and socio-economic status.....	201
8.7.2 Alternative modelling strategies.....	201
8.7.3 Further work that may require new models.....	201

8.7.4 Validation of projections with real data.....	202
8.7.5 Re-working the Dutch Model	202
8.7.6 The definition of COPD.....	203
8.8 Conclusion to thesis.....	203
Appendix 1.1 CPRD Original protocol.....	205
Appendix 1.2 CPRD Amendment protocol	216
Appendix 1.3 CPRD agreement for provision of incidence and prevalence data	226
Appendix 1.4 Covering letter to CPRD ISAC for approval.....	230
Appendix 1.5 ISAC Application form	232
Appendix 1.6 ISAC Protocol approval.....	241
Appendix 2.1 Level 1 Ethics approval from UoE.....	243
Appendix 3.1 Ethics waiver for use of Lothian COPD cohort database.....	245
Appendix 3.2 Email approval for use of the Lothian COPD Cohort database from Privacy Advisory Committee (PAC) representative Janet Murray	246
Appendix 3.3 Privacy Advisory Committee (PAC) Application for Lothian COPD Cohort Database.....	258
Appendix 4.1 Request for Caldicott Guardian approval for file transfer to the University of Edinburgh	312
Appendix 4.2 Request for Caldicott Guardian approval for PhD data use.....	315
Appendix 4.3 Initial reply from Caldicott Guardian May 2013.....	319
Appendix 4.4 Clarification of data use for PhD to Caldicott.....	321
Appendix 4.5 Caldicott Guardian Approval August 2014.....	324
Appendix 5 Description of method of generating random incidence and prevalence curves for COPD.....	325
Appendix 6 Contribution to science	331
Glossary	332

List of Tables and Figures

Table 1.1 Global ranking of causes of death (Mathers and Loncar 2006)	24
Figure 2.1 The natural history of FEV1 with age and smoking and the thresholds of disability and death. Reproduced with permission (.....)	28
Figure 2.2 Mean FEV1 values (expressed as percent of its value at the age of 25) by age in smokers who quit smoking before the age of 30 (Q<30), between 30 and 40 years of age (Q30-40) and after the age of 40 (Q40+). Curves from health never smokers (NS) and continuous smokers (CS) are included for comparison. A males, B females.....	31
Figure 2.3 Time-volume graph showing obstructive pattern. Reproduced under Creative Commons licence (Cheek 2014).....	34
Table 2.1 European prevalence of COPD from Atsou. Reproduced from BMC Medicine under open access licence (Atsou, Chouaid et al. 2011).....	38
Table 2.2 Results from the BOLD study, adults over 40 years (Buist, McBurnie et al. 2007)	40
Table 2.3 PLATINO results, adults over 40 years (Menezes, Lopez et al. 2009).....	41
Table 2.4 Prevalence of COPD in England from East of England Public Health Observatory	42
Table 2.5 Incidence and prevalence of COPD in England from Clinical Practice Research Datalink (CPRD).....	42
Table 2.6 Physician diagnosed prevalence of COPD in Scotland (SPHO 2012).....	43
Figure 2.4 Number of deaths identified from death certificates with codes of J40 to J47 identified as underlying cause of death in England and Wales 2006-2012 by sex	46
Figure 2.5 Number of deaths from GROS with codes of J40 to J47 identified as underlying cause of death in Scotland 2000-2012 by sex	46
Table 2.7 Thresholds of airways obstruction used to define severity categories in COPD ...	55
Figure 2.6 Typical end of life trajectory in COPD. Reproduced with permission from BJGP (Lehman 2004).....	62
Table 3.1 Primary care databases covering England and Scotland.....	66
Table 3.2 Clinical coding systems (de Lusignan and van Weel 2006; ConnectingForHealth 2013)	71
Figure 4.1 Example of a Markov Model. Reproduced with permission (Briggs and Sculpher 1998)	84
Table 4.1 Cohort simulation of Markov Model, Figure 4.1, with no drug therapy	86
Figure 5.1 PRISMA flow diagram.....	97

Table 5.1 Characteristics of modelling studies. Studies are in order of quality of reporting (Table 5.3)	99
Table 5.2 Data inputs and model outputs from all 22 modelling studies	107
Table 5.3 Quality of reporting framework scores for modelling studies.....	116
Table 5.4 Reported outcome of each modelling study	119
Figure 6.1 The Dutch Model. Reproduced with permission (Hoogendoorn, Rutten-van Molken et al. 2011)	136
Table 6.1 Model inputs and respective data sources for England and Scotland	137
Table 6.2 COPD Read Codes used to identify the Lothian COPD Cohort	140
Table 6.3 Prevalence of COPD in England from Clinical Practice Research Datalink 2011.	143
Table 6.4 COPD Read Code Grouping for Practice Team Information Database	145
Table 6.5 COPD prevalence for Scotland	148
Table 6.6 England smoking proportions from General Lifestyle Survey Data 2011	149
Table 6.7 Health Survey for England smoking prevalence among 8-15 year olds	150
Table 6.8 Scottish Health Survey 2011, proportion of smokers by age and sex	151
Table 6.9 Original COPD direct cost input data for modelling (Euros) from Dutch Model report (Hoogendoorn, Rutten-van Molken et al. 2010).....	154
Table 6.10 Unit costing from indacaterol study data converted to total cost per severity level (Price, Asukai et al. 2013).....	156
Table 6.11 Costs of managing COPD in the UK 2011 from tiotropium study.(Hettle, Wouters et al. 2012)	157
Table 6.12 Utility scores for Dutch Model (Hoogendoorn, Rutten-van Molken et al. 2010)	158
Table 6.13 Parameters for Cox Regression Model of COPD patient survival	160
Table 6.14 Age specific relative risk of dying per year if patient has COPD compared to an age and sex matched control as obtained from CPRD data extract.....	161
Table 6.15 Strengths and limitations of datasets	161
Table 7.1 One way sensitivity analyses for Dutch Model projections from 2011 to 2030, all run for both England and Scotland.....	165
Figure 7.1 Modelled population of males and females with COPD in England	166
Figure 7.2 Age-specific prevalence of males with COPD in England	167
Figure 7.3 Age-specific prevalence of females with COPD in England	167
Figure 7.4 Modelled costs for COPD in England.....	168
Figure 7.5 Number of deaths among people with COPD in England from all causes	169

Figure 7.6 Modelled population of males and females with COPD in Scotland	169
Figure 7.7 Age-specific prevalence of males with COPD in Scotland	170
Figure 7.8 Age-specific prevalence of females with COPD in Scotland	171
Figure 7.9 Modelled costs for COPD in Scotland	172
Figure 7.10 Number of deaths among people with COPD in Scotland from all causes	172
Figure 7.11 Sensitivity analyses for England, vertical line represents base case.....	173
Figure 7.12 Sensitivity analyses for Scotland, vertical line represents base case.....	174
Table 7.2 Uncertainty intervals for outputs from the Dutch Model for England.....	176
Table 7.3 Uncertainty intervals for outputs from the Dutch Model for Scotland	176
Table 7.4 Overall numbers of people with COPD, the population size and prevalence of COPD in 2011 and 2030 in England and Scotland (uncertainty intervals from Tables 7.2 and 7.3, with probabilistic mean).....	177
Figure 8.1 Age related prevalence of smoking in England and Scotland in males.....	181
Figure 8.2 Age related prevalence of smoking in England and Scotland in females.....	182
Figure 8.3 Age related prevalence at baseline in England and Scotland	183
Table 8.1 Absolute and relative differences in increases in the COPD population	184
Figure 8.4 Age related smoking prevalence in England and Scotland (a combination of figures 8.1 and 8.2)	184
Figure 8.5 Age related COPD excess mortality per 1000 population.....	185
Figure 8.6 Smoking epidemic diagram for high-income countries. Reproduced with permission (Thun, Peto et al. 2012)	189
Figure 8.7 Smoking-attributable deaths at ages 35-69 estimated indirectly from lung cancer rates and expressed as a percentage of all deaths. Reproduced with permission (Thun, Peto et al. 2012)	190
Table 8.2 Relative risks for ex-smokers of developing COPD compared to non-smokers (Van Oers 2002; SurgeonGeneral 2004).....	193
Table 8.3 Overall projected prevalence of COPD in 2025 in three countries.....	194
Table 8.4 Predictions from the Dutch Model.....	195
Table 8.5 Comparison of English cost data from the Dutch Model with PCT returns	197
Table 8.6 Comparison of Scottish cost data from the Dutch Model for severe exacerbations with ISD cost data for hospital admissions	198
Table 8.7 Estimated numbers of exacerbations as predicted by modelling for England and Scotland for selected years.....	198

Abbreviations

ACOS Asthma and COPD Overlap Syndrome

ATS American Thoracic Society

BMI Body Mass Index

BOLD Burden of Obstructive Lung Disease

BTS British Thoracic Society

CAG Confidentiality Advisory Group

CAT COPD Assessment Test

CCQ Clinical COPD Questionnaire

CDSS Computer Decision Support System

CHI Community Health Index

CHRNA3 Cholinergic Receptor Nicotinic Alpha 3

COAD Chronic Obstructive Airways Disease

COLD Chronic Obstructive Lung Disease

COPD Chronic Obstructive Pulmonary Disease

CPRD Clinical Practice Research Datalink

CRS Care Records Service

CT Computer tomography

DALY Disability adjusted life year

DISMOD The disease model from the Global Burden of Disease Study

DYNAMO-HIA the name of a European model by Lhachami et al

ECCS European Community Coal and Steel Cohort

EHI E-Health Insider

EHR Electronic Health Record

ERJ European Respiratory Journal

ERS European Respiratory Society

FAM13A Family with sequence similarity 13 member A

FDA United States Food and Drug Administration

FEV1 Forced Expiratory Volume in 1 Second

FHIR Fast Healthcare Interoperability Resource

FVC Forced Vital Capacity

GBD Global Burden of Disease

GDP Gross Domestic Product

GOLD Global Initiative for Chronic Obstructive Lung Disease

GP General Practice or General Practitioner

GPRD General Practice Research Database (now called CPRD)

GROS General Register Office for Scotland

GWAS Genome Wide Association Studies

HES Hospital Episode Statistics

HIRU Health Information Research Unit

HHIP Hedgehog Interacting Protein

HL7 Health Level 7

HSCIC Health and Social Care Information Centre

HSW Health Solutions Wales

HTA Health Technology Assessment

iBALT inducible Bronchus Associated Lymphoid Tissue

ICD -10 International Classification of Diseases version 10

ICER Incremental Cost Effectiveness Ratio

ICPC International Classification of Primary Care

ISD Information Services Division, Scotland

LLN Lower limit of normal of FEV1/FVC ratio

MHRA Medicines and Healthcare products Regulatory Agency

MRC Medical Research Council

nGMS new General Medical Services Contract

NHS National Health Service

NICE National Institute for Health and Care Excellence

NIHR National Institute for Health Research

NPfIT National Programme for Information Technology

ONS Office of National Statistics

PAS Patient Administration System

PAC Privacy Advisory Committee

PACS Picture Archiving and Communication System

PaO₂ Partial Pressure of Oxygen

PCR Polymerase Chain Reaction

PCCIU Primary Care Clinical Informatics Unit

PCT Primary Care Trust

PLATINO Latin American Project for the Investigation of Obstructive Lung
Disease

PTI Practice Team Information

QALY Quality adjusted life year

QOF Quality and Outcomes Framework

QOL Quality of Life

SAIL Secure Anonymised Information Linkage

SALSUS Scottish Schools Adolescent Lifestyle and Substance Use Survey

SCI-DC Scottish Care Information – Diabetes Collaboration

SGRQ St George’s Respiratory Questionnaire

SHIP Scottish Health Informatics Programme

SIMD Scottish Indices of Multiple Deprivation

SMR01 Scottish Morbidity Record 01

SNPs Single Nucleotide Polymorphisms

sRAGE soluble Receptor for Advanced Glycation Endproducts

THIN The Health Improvement Network

TORCH A COPD drug trial titled “Towards a revolution in COPD health”

UK United Kingdom

US United States of America

WHOLIS World Health Organization Library Information Service

YLL Years of life lost

Acknowledgements

Firstly, I would like to acknowledge the support and guidance of my supervisors: Professor Aziz Sheikh, Professor Sarah Wild and Dr Colin Simpson. These three have been extremely helpful in propelling me along the PhD track. The team who received me in Rotterdam and with whom I collaborated must also be acknowledged and thanked; they are Dr Martine Hoogendoorn, Professor Maureen Rutten-van Mölken, Dr Talitha Feenstra and Dr Rudolf Hoogenveen. Martine guided me closely in Rotterdam and we had frequent contacts on my return and Rudolf's work was key in preparing the uncertainty estimates. Other acknowledgements should go to Dr Victoria Barbour who acted as second author for the purpose of the systematic review selection and scoring system. Professor Simon Capewell who helped design the systematic review quality of reporting framework. Professor Robert West and Dr Jamie Brown made available the results of the smoking toolkit study, which were invaluable. Dr Leonie Hunter and Dr Rachel Hardie gave me access to incidence rates from the Lothian COPD study. Miss Rosalia Munoz-Arroyo prepared further smoking data for me. Dr Helen Strongman at Clinical Practice Research Datalink (CPRD) played a key role in preparing data on incidence and prevalence and provided advice regarding the validity of the data. Miss Leslie Harkins provided similar support for Information Services Division (ISD) data. Miss Donna Mikolajczak also provided information from ISD regarding NHS Scotland's spending on COPD. Dr Daniel Kotz allowed me access to his CPRD data from a COPD case-control study which was central to enabling the PhD.

I'd also like to acknowledge the contribution of Dr Niall Anderson for helping with some programming in R. Also Douglas Thompson and Margaret Horne were two other PhD students in statistics who helped to keep me right when I was undertaking analysis.

This research was supported financially by a Principal's Career Development PhD scholarship and by a grant from the Edinburgh and Lothian Health Foundation.

Thesis overview

I describe the rationale for this study in Chapter 1. Firstly, I describe the background, then the rationale for the choice of chronic obstructive pulmonary disease (COPD) as the condition under study and the selection of England and Scotland as the setting for the study. Then, in Chapter 1, I consider the selection of the Dutch Model and the timeframe for projections. The overall aim of the PhD was to select an epidemiological model to make projections for the prevalence and burden of COPD in England and Scotland from 2011 to 2030 using routinely collected healthcare data. I conclude Chapter 1 with a breakdown of the two phases of the PhD: Phase 1 – systematic review and Phase 2 – modelling study.

In Chapter 2, I introduce COPD as the condition under study. In it, I consider the history and definition of COPD and how this relates to its diagnosis. Then there is discussion of different definitions of COPD for research. This is followed by a section on the risk factors and epidemiology of COPD. The chapter continues with sections on the clinical features, prognosis and management of COPD.

In Chapter 3, I discuss the collection of healthcare data in England and Scotland with an overview of how these data may be reused for research purposes. I concentrate on particular issues: the current state of data gathering in England and Scotland; the validity of the data that are recorded, how these can be used; and the regulatory framework that is in place to secure patient privacy. I also consider the validity of the routine data that were selected for use with models to provide projections in this PhD.

In Chapter 4, I introduce different methods of modelling including Markov modelling, demographic modelling, risk factor modelling and time trend modelling. This is followed by an introduction to quality adjusted life years (QALYs) and disability adjusted life years (DALYs). This chapter introduces these models before the systematic review in Chapter 5 which identifies examples of these different types of models. In Chapter 4, I concentrate on Markov modelling because this is the type of model most often used in the studies found in the systematic review.

In Chapter 5, I present a systematic literature review of models for calculating the prevalence and burden of COPD. This was firstly undertaken in 2011 when the PhD was begun and updated in 2013. The aim was to identify all models for calculating the prevalence and burden of COPD, then to assess the quality of these models with a view to selecting one to use to provide projections of prevalence and burden of COPD in England and Scotland. As a first step to assessing the quality of the models a quality of reporting framework was

designed and subsequently published. The model that scored highest in relation to the quality of reporting framework was the Dutch Model at that time (2011), and so I decided to approach the Rotterdam team with a request for collaboration in order to use their model with English and Scottish Data to produce projections for England and Scotland. Chapter 5 presents 22 models for calculating the prevalence and burden of COPD in different contexts and settings. I include an appraisal of the overall quality of the models for the seven models which scored most highly on the quality of reporting framework assessment. The reason for this is that the overall quality of the model can only be appraised where the reporting is sufficiently detailed. From the 22 models, I selected the Dutch Model to proceed with estimating projections for the COPD population in England and Scotland.

In Chapter 6, I begin by describing the working of the Dutch Model then I give an overview of the data required by the Model and appropriate data sources. Then, I give detailed consideration to the sourcing of each of the appropriate data inputs and the permissions that were required to obtain each dataset. Finally, I discuss the strengths and limitations of these datasets.

In Chapter 7, I include a description of the computer programme used to run the Dutch Model followed by a list of the sensitivity analyses run on the programme following the base case. I present the results of the base case modelling followed by tornado diagrams detailing the sensitivity analyses. Then I present the uncertainty intervals. I then discuss the meaning of these findings.

In Chapter 8, I summarise the principal findings, followed by the strengths and weaknesses of the work presented in this thesis. I then discuss the meaning of the study in the context of other work, including work on the smoking epidemic and the Global Burden of Disease Study. This is followed by my discussion of the implications of the work for policy and practice and finally consideration as to what questions remain to be answered by further research. I end this thesis by summarising my main conclusions from this body of work.

Chapter 1 Background, rationale and aims

1.1 Background

Policymakers and governments must decide their healthcare priorities on the best information available to them. This PhD was begun in 2011 and, at that time, elements influencing healthcare were coming together to create an especially challenging environment for decision makers. These elements included demographic trends and an economic downturn.

The demographic trend is for populations globally to live longer lives with the result that they require substantial health and social care when they are frail and elderly. Globally, there has been a shift in the distribution of deaths from younger to older ages and from communicable, maternal, perinatal and nutritional causes to non-communicable disease.(Mathers and Loncar 2006) This is combined with a drop in the birthrate (see Glossary) so that there are fewer younger, economically active individuals in the population to provide and fund this care. In addition, these elderly populations often suffer from multiple long-term conditions.(Barnett, Mercer et al. 2012) This has implications for the complexity of care required.

2011 was characterised by a substantial economic downturn. Health expenditure across Organisation for Economic Co-operation and Development countries (OECD) did not increase in 2010 and 2011 as a result of the global economic slowdown.(OECD 2013)

1.2 Rationale for a PhD in modelling

In 2011, I wondered what the best and cheapest methods were to help governments focus their healthcare resources according to the greatest source of need. To me, it seemed important that policymakers and governments should have accurate information as to how to prioritise healthcare needs among their population. I had heard of the Global Burden of Disease (GBD) Study and at that time (2011) there were projections of the GBD until 2030.(Mathers and Loncar 2006) This study introduced me to the concept of modelling non-communicable diseases to predict future burden of disease.

I turned to Merriam-Webster's dictionary to find the definition of a "model" as "*a system of postulates, data and inferences presented as a mathematical description of an entity or state of affairs.*"(Merriam-Webster 2013)

As a non-mathematician, I realised that there would be a limit as to the level at which I could engage with any particular model. However, modelling clearly has an important role in predicting future population disease burden, and so I decided to pursue this track.

1.3 Rationale for a PhD in COPD

I have always been interested in respiratory disease, especially smoking-related disease as it is so potentially preventable. As a General Practitioner, I am always encouraging patients to stop smoking and I wanted to look at a condition which was relevant to my everyday practice in Lothian, Scotland. I was interested to see that in the Mathers and Loncar projections COPD was ranked as the fifth leading cause of death in 2002 and was modelled to rise to the fourth leading cause of death globally by 2030, see Table 1.1.(Mathers and Loncar 2006) This ranking is based upon modelling using World Health Organization (WHO) estimates for global tobacco-caused deaths in 2000.(Ezzati and Lopez 2003) COPD is, therefore, important in the context of the increasing prevalence of non-communicable diseases; and so I decided to focus upon COPD.

Table 1.1 Global ranking of causes of death (Mathers and Loncar 2006)

Condition	Rank in 2002	Projected rank in 2030
Ischaemic heart disease	1	1
Cerebrovascular disease	2	2
Lower respiratory infections	3	5
Human Immunodeficiency Virus/Acquired Immunodeficiency Disease	4	3
COPD	5	4

1.4 Rationale for modelling in England and Scotland, 2011 to 2030

The Mathers and Loncar projections were global; however, I was unable to find projections specifically for Scotland for COPD prevalence and burden which would be most relevant to my own clinical practice. I therefore decided to aim to model projections for Scotland. I realised that this work would have much more research impact if I could compare Scotland with its neighbour, England, therefore I decided to additionally model projections for England. The period of projection from 2011 to 2030 was taken to be comparable with the

Mathers and Loncar time period.(Mathers and Loncar 2006). I also hoped that I would be able to use a more up to date source for my projections than the WHO 2000 estimates.

1.5 Aim

The general aim of the PhD was to model the prevalence and burden of COPD. In order to fulfil this aim I undertook two phases of work as described in Sections 1.6 and 1.7. This then led to the specific aim of producing projections for the prevalence and burden of COPD in England and Scotland from 2011 to 2030 using an appropriate model.

1.6 Phase one – systematic review

The above rationale laid the foundations for phase one of my PhD. It was necessary to firstly investigate models for calculating the prevalence and burden of COPD. I decided to do this using a systematic review and defined the disease burden of COPD as being seen broadly from the perspective of the healthcare system. Therefore, the disease burden constitutes the numbers of people with COPD, numbers of COPD exacerbations in the population, the direct healthcare costs of these and also the numbers of COPD deaths.

I carried out the searches for the systematic review in 2011 and updated them in November 2013. There was no published scoring system for chronic disease models and so I constructed a quality of reporting framework for COPD models in discussion with Professor Simon Capewell of Liverpool University. At the time of the original searches the model which scored most highly on the quality of reporting framework was one (Hoogendoorn, Rutten-van Molken et al. 2005) of the three included models developed by a research team at Erasmus University, Rotterdam.(Feenstra, van Genugten et al. 2001; Hoogendoorn, Rutten-van Molken et al. 2005; Hoogendoorn, Rutten-van Molken et al. 2011). I then updated the systematic review searches in 2013 as I was writing up my thesis and found a study (Pichon-Riviere, Augustovski et al. 2011) which scored slightly higher on the quality of reporting framework; however, the Dutch Model remains the focus for phase two as a result of the work done in 2012 and 2013 towards using it for modelling.

1.7 Phase two – modelling study

The aim was to produce projections for the prevalence and burden of COPD in England and Scotland from 2011 to 2030 using an appropriate model. Following on from the systematic review, I contacted Erasmus University as the literature I had found suggested that their team had the most expertise in modelling in this area and suggested collaboration. They agreed that I could use the most up to date version of their Dutch COPD model with data from

England and Scotland to make projections of the prevalence and burden of COPD in these countries from 2011 to 2030. This would be the first time such projections had been made as previous models had either had global scope or, for English models, only modelled current prevalence instead of including future projections. In addition, none of the English models identified included detailed projections of COPD exacerbations or costs, which was something that the Dutch Model would be able to provide.

The inputs to the model would consist of a selection of routinely collected data. I had to consider what were the best available options for these inputs and which were the most valid. The inputs would then be modified and the model re-run to provide sensitivity analyses as is often done in modelling studies. This demonstrates the sensitivity of the output of the model to fluctuations in each of the different inputs separately and gives an indication as to whether the intrinsic assumptions that went into the model were accurate and stable.

1.8 Conclusion to Chapter 1

In this chapter I have outlined the context in which I began my PhD in 2011 in terms of an economic downturn and increasing demographic challenges. I then explained my rationale for choosing a PhD in modelling the prevalence and burden COPD. I also explained the setting of my studies in my local populations of England and Scotland. I then summarised my aim and the two phases of work which went into producing this PhD.

Chapter 2: Basic concepts in COPD

2.1 Introduction

COPD is an important chronic disease in the context of an ageing population. This chapter sets out to present an overview of the medical aspects of COPD. I begin by reviewing definitions of COPD followed by a discussion of its diagnosis. Then there is a section on the challenges of measuring the epidemiology of COPD. Then a summary of the aetiology of COPD leads into a discussion of its pathophysiology. I discuss the chief clinical features of COPD and its prognosis and management. I finish with a summary of patient perspectives in COPD and conclude this chapter.

2.2 Definitions of COPD

In any epidemiological study, it is important to review the definition of the index condition as incidence and prevalence will fluctuate widely depending on how many patients are determined as “having the condition” or “not having the condition”.

2.2.1 Historical context of COPD definitions

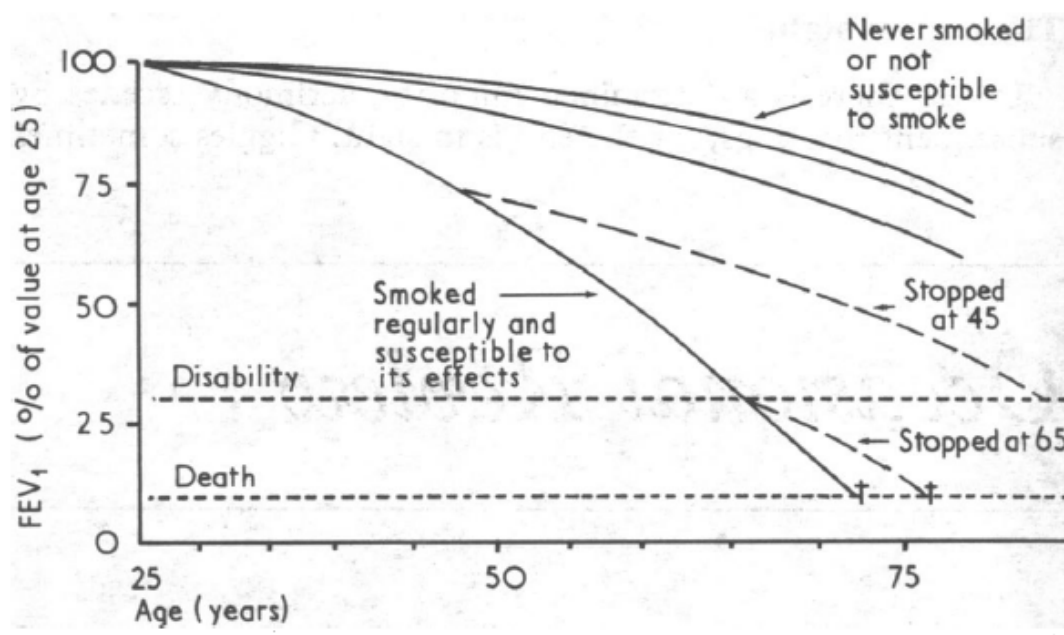
Since the beginning of the 20th Century, a relationship has been recognised between three overlapping conditions: emphysema, chronic bronchitis and asthma. In the 1950s, the London smog resulted in an increased number of deaths from respiratory causes clearly related to air pollution.(Fletcher 1978) Interest in defining this condition continued to increase and it emerged that Americans were labelling it “emphysema” while the British favoured the term “chronic bronchitis”.(Calverley and Wedzicha 2007). This group of conditions principally referred to three types of disorder: 1) chronic or recurrent excessive secretion of bronchial mucus; 2) intermittent obstruction to bronchial flow (asthma); and 3) persistent obstruction to bronchial airflow due either to emphysema – defined pathologically as the enlargement of the size of air spaces in the lung peripheral to terminal bronchioles – or to an obstructive form of bronchitis for which the underlying anatomical basis was not yet appreciated.(Fletcher 1978)

In 1960 the Medical Research council’s committee on the aetiology of chronic bronchitis concerned itself with the question of how smoking and other factors cause airflow obstruction. This led to Fletcher and Peto’s seminal work in which they followed prospectively, approximately 1000 working men in London, studying respiratory symptoms and changes in lung function over a period of 8 years.(Fletcher and Peto 1977) The men were aged 30-59 and few had any clinical disease. They were seen every 6 months and

questions and measurements were made to assess mucus hypersecretion and bronchial infections. Six monthly FEV1 readings were made, allowing estimates to be made of the average rate of decline of FEV1 for each man during the study. Only 103 of the men were non-smokers, none of these non-smokers had any evidence of even moderate airway obstruction.

Fletcher and Peto drew the following conclusions from their prospective study. Firstly, that FEV1 declines continuously and smoothly over an individual's life with a slight acceleration of rate of loss with ageing. Secondly, they identified that non-smokers lost FEV1 slowly and almost never developed clinically significant airflow obstruction. Thirdly, they found that some smokers were more or less susceptible to the effects of smoking on their FEV1 decline. Fourthly, they found that susceptible smokers who stopped smoking would not recover lost FEV1, but that the rate of decline in FEV1 would revert to being parallel with the non-smoking curve. These findings are illustrated diagrammatically in figure 2.1.

Figure 2.1 The natural history of FEV1 with age and smoking and the thresholds of disability and death. Reproduced with permission (Peto and Fletcher 1977)



Additionally, they found that after adjusting for FEV1, smoking, age and height there was no independent correlation between FEV1 slope and indices of either mucus hypersecretion or bronchial infections. They were surprised at this finding and so confirmed it by relating

changes in FEV1 level to changes in expectoration and to episodes of bronchial infection in individual men and found no consistent or significant effects: the loss of FEV1 that an individual man suffered from one six monthly survey to the next was on average the same if a chest cold chest illness or attack of sputum purulence intervened as if it did not. (Fletcher and Peto 1977)

This somewhat surprising finding laid the ground for further work that characterised airflow obstruction and mucus hypersecretion as two quite different diseases. Another study obtained follow up data over 20 years for 2718 British men with details of airflow obstruction and mucus hypersecretion and causes of death. (Peto, Speizer et al. 1983) The aim was to relate initial mucus hypersecretion and airflow obstruction to subsequent cause-specific mortality. They found that when the 108 men whose airflow obstruction values were more than two standard deviations below average for their age and area of residence were compared with the larger group of 1464 men whose airflow obstruction values were better than this average, 27% of the former were certified as dying of COPD and only 1% of the latter were so certified. The COPD death rate among the 108 was more than 50 times that among the 1464, since the former were at risk for fewer years because of an excess among them of premature deaths both from COPD and other causes. However, when comparing men who reported producing phlegm both in the morning and during the day with men who did not produce phlegm a relative risk of only 4.4 was seen for COPD mortality, which reduced to 1.4 with adjustment for airflow obstruction. Therefore the presence or absence of mucus hypersecretion is not predictive of subsequent mortality from COPD. The authors concluded that the type of mucus hypersecretion that leads to noticeable phlegm production is not a link in any important causal chain that accelerates the development of airflow obstruction. Mucus hypersecretion and airflow obstruction are therefore two distinct morbid processes. The authors then recommended that the term “chronic bronchitis” should not be used to indicate airflow obstruction. (Peto, Speizer et al. 1983)

Further studies have added to the evidence that it is the airflow obstruction and not mucous hypersecretion that has the relationship with mortality. (Calverley and Wedzicha 2007)

Airflow obstruction data is also available in evidence in the Framingham study which collected information on cigarette consumption, self-reported chronic cough, FEV1 and FVC. (Freund, Belanger et al. 1993) Heavy smokers, both men and women across all age groups, sustained diminished FEV1 in a dose-related fashion, data for FVC showed an equivalent trend. Cigarette smoking was an independent risk factor for chronic cough, reduction in FEV1 and FVC. The Framingham study did not consistently record the

incidence of COPD because of difficulties in definition and determination of onset, however a highly significant relationship between the rate of cigarette smoking and mortality was thought to represent the comorbid effect of the development of obstructive lung disease in smokers.(Freund, Belanger et al. 1993)

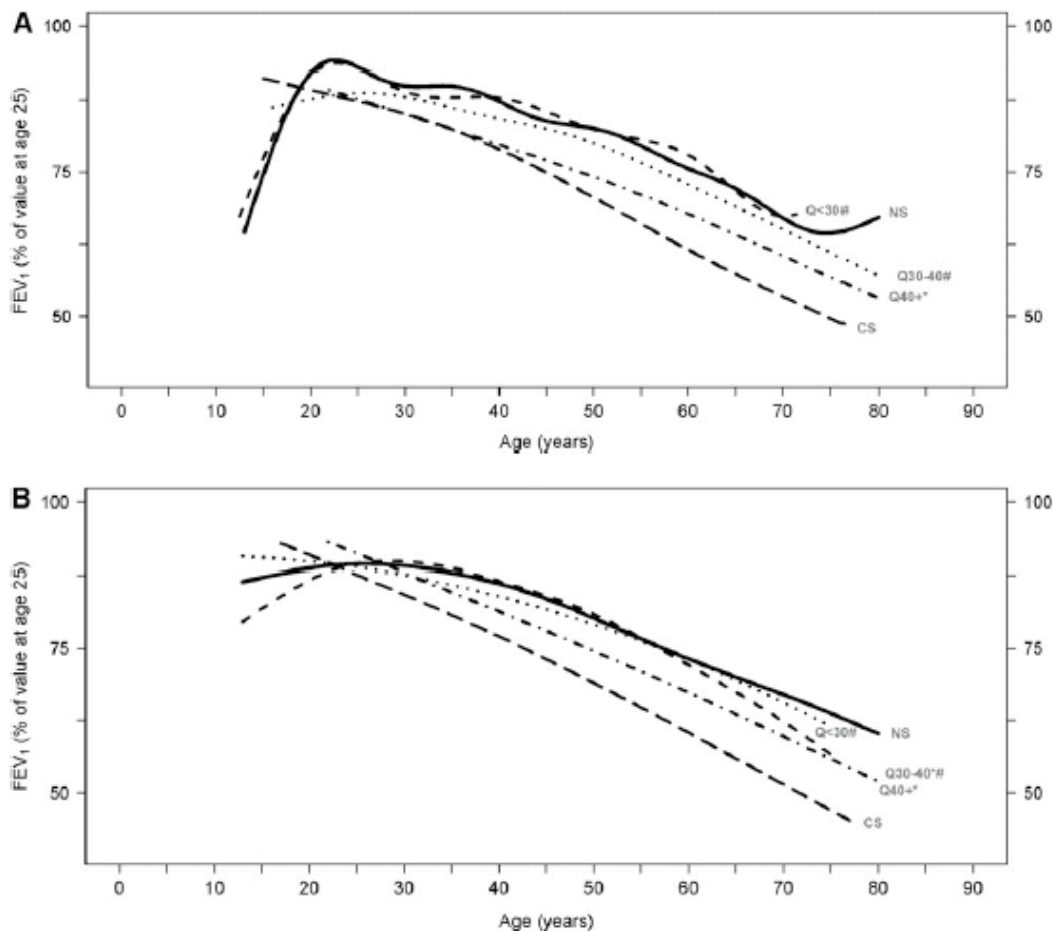
Further evidence from the Framingham study demonstrated that poorer lung function after 14 years of follow up was related to baseline lung function and cigarette smoking. (Sorlie, Lakatos et al. 1987) Follow up FEV1 was lower in smokers than non-smokers. Poorer lung function in smokers at follow up was also demonstrated following adjustment for initial lung function impairment, respiratory symptoms and stature. This study also showed that those who quit smoking have better lung function on follow up than those who continue to smoke. In this way the Framingham study helps to re-inforce the Fletcher-Peto evidence as to the significance of airflow obstruction from cigarette smoking and its association with mortality and morbidity.

An analysis of the Framingham Offspring Cohort, published in 2009, looked at spirometry measurements of 4391 males and females with an age range of 13-71 at baseline.(Kohansal, Martinez-Camblor et al. 2009) In order to revisit the natural history of COPD, this work recreated the Fletcher and Peto lung function curves but with the advantages that they included females, it was a larger cohort with a larger age range, participants were followed up regularly for up to 26 years and the spirometric measurements were standardized. Participants were considered to be “healthy” if they reported “no” to all questions concerning dyspnoea on exertion, increased dyspnoea, nocturnal dyspnoea, nocturnal cough or wheezing at exam 1 and were never reported by an examiner to have a clinical diagnostic impression of asthma, COPD, chronic bronchitis or emphysema during follow up examinations. By contrast if any of these criteria were fulfilled they were considered as “unhealthy”.

Figure 2.2 demonstrates the chief findings from this study. Firstly they found that health never smoker females achieved lung growth earlier than males and their rate of decline with age is slightly, but not significantly lower. Secondly, they clearly showed that smoking increased the rate of decline of FEV1 in both males and females. Thirdly they demonstrated a range of susceptibility to the negative effects of smoking on lung function and, in contrast to the Fletcher-Peto cohort, found that the presence of respiratory symptoms at baseline or follow up identified individuals who were more susceptible to the effects of smoking. This is an observation that may prove clinically useful. And fourthly, they confirmed that smoking cessation has a beneficial effect on lung function decline with age, especially in those who quit at younger ages.

Figure 2.2 Mean FEV₁ values (expressed as percent of its value at the age of 25) by age in smokers who quit smoking before the age of 30 (Q<30), between 30 and 40 years of age (Q30-40) and after the age of 40 (Q40+). Curves from health never smokers (NS) and continuous smokers (CS) are included for comparison. A males, B females.

Reprinted with permission of the American Thoracic Society. Copyright © 2015 American Thoracic Society. (Kohansal, Martinez-Cambor et al. 2009) The American Journal of Respiratory and Critical Care Medicine is an official journal of the American Thoracic Society.



In the Framingham Offspring Cohort, males started smoking at a mean age of 17.5 years, before the lung was fully mature, in contrast to females who started at a mean age of 18.8 years. In subsequent years teens have begun to start smoking at much younger ages which could cause more deleterious effects on lung growth. The small FEV₁ increase in males in the oldest categories could be due to survivor bias or smaller sample size as this trend is not seen in women. The percentage of male and female continuous smokers developing airflow

obstruction was similar, thus suggesting that males and females are equally susceptible to tobacco smoke. It was also observed that smoking appeared to reduce the FEV1 peak in males. No such peak was identified in females. However, for males, at least the results are consistent with the hypothesis that smoking during adolescence reduces lung growth.

In contrast to the model proposed by Fletcher-Peto, in which it was proposed that those who quit smoking would find their lung function decline revert to parallel with non-smokers, the Framingham Offspring Cohort produced data which demonstrated different, steeper, slopes for quitters at older ages in both males and females. These data highlight the importance of early quitting. (Kohansal, Martinez-Camblor et al. 2009)

These studies comprise the key evidence from which our understanding of the condition known as COPD is drawn. Other terms used included “chronic obstructive airways disease” (COAD) and “chronic obstructive lung disease” (COLD). “Chronic non-specific lung disease” and “chronic airflow limitation” persist in the literature but are less often used since the advent of formal agreed UK and international definitions such as the National Institute for Health and Care Excellence (NICE) definition and the Global organisation for Obstructive Lung Disease (GOLD) definition. These are reviewed below.

2.2.2 Definition from the 2010 National Institute for Health and Care Excellence (NICE) guideline for COPD

The following definition is from the 2010 update of the 2004 NICE / British Thoracic Society (BTS) guidelines on COPD: (NICE 2010)

“Chronic obstructive pulmonary disease (COPD) is characterised by airflow obstruction. The airflow obstruction is usually progressive, not fully reversible and does not change markedly over several months. The disease is predominantly caused by smoking.”

The guidelines then go on to clarify that airflow obstruction is defined as a reduced post-bronchodilator FEV1/FVC ratio < 0.7 (FVC = Forced Vital Capacity) when FEV1 is less than or equal to 80% of the predicted normal. A diagnosis of COPD should only be made in the presence of breathlessness and/or cough.

The guidelines state that COPD is now the preferred term for the conditions in patients with airflow obstruction who were previously diagnosed as having chronic bronchitis or emphysema. This preference for the term COPD is as a result of the importance of airflow obstruction in relation to mortality.

2.2.3 Definition from the 2013 updated Global Initiative for Chronic Obstructive Lung Disease (GOLD) guidelines

The updated 2013 GOLD guideline (GOLD 2013) gives the following definition for COPD:

“COPD, a common preventable and treatable disease, is characterised by persistent airflow limitation that is usually progressive and associated with an enhanced chronic inflammatory response in the airways and the lung to noxious particles or gases. Exacerbations and comorbidities contribute to the overall severity in individual patients.”(GOLD 2013)

The authors go on to stress that the definition does not use the terms chronic bronchitis and emphysema and excludes asthma. The symptoms of COPD include dyspnoea, chronic cough and chronic sputum production, and episodes of acute worsening of these symptoms (exacerbations) often occur. The same post-bronchodilator spirometry threshold of $FEV1/FVC < 0.7$ is given for persistent airflow limitation and thus COPD as in the NICE guideline above.

“Chronic bronchitis” overlaps with COPD, and its definition is clinical. It is defined as “cough with excessive production of sputum present for at least three months in each of two consecutive years.”(GOLD 2010)

2.2.4 How the definitions of COPD relate to diagnosis

The GOLD criteria describe “usually progressive airflow limitation which is not fully reversible” as the hallmark of COPD. GOLD recommends that this airflow obstruction is confirmed by a $FEV1/FVC$ ratio < 0.7 . GOLD criteria use this as a fixed threshold for the measurement of obstruction. The severity of the COPD may then be assessed by the $FEV1$ in the context of its percent of predicted normal values according to age, sex and height.(GOLD 2010)

GOLD guidelines recommend consideration of the diagnosis of COPD in any patient over 40 years of age who has persistent dyspnoea, a chronic cough with or without sputum production and a history of exposure to risk factors such as tobacco smoke, occupational dusts and chemicals or smoke from home cooking or heating fuel.(GOLD 2010) The diagnosis should be confirmed by measuring and calculating the following spirometry parameters:

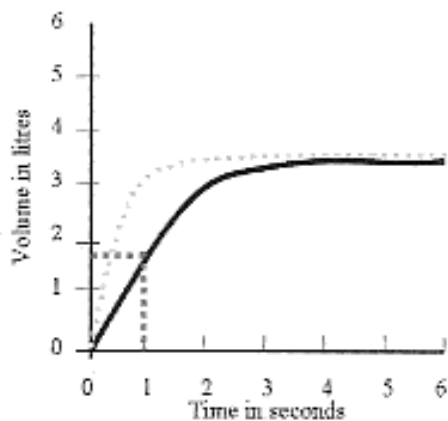
- $FEV1$
- FVC
- $FEV1/FVC$ ratio.

In order to exclude reversible obstruction, as may be present in asthma, the spirometry should be performed 15 minutes after administration of a bronchodilator.(GOLD 2010)

Both FEV1 and the ratio decrease in COPD. A reduction in the FVC alerts the physician to the presence of a possible restrictive defect.

The obstructive pattern of spirometry is most easily understood with reference to the volume-time graph in Figure 2.2. The dashed line shows an exhalation from 0 to a FVC of 3.5 litres with an FEV1 of 3.0 litres. FEV1/FVC ratio is $3.0/3.5 = 86\%$, within the normal range. The obstructive (solid) line shows an exhalation from 0 to an FVC of 3.5 litres with an FEV1 of 1.8 litres. FEV1/FVC is $1.8/3.5 = 51\%$, this is less than 70%, therefore it is an obstructive picture.

Figure 2.3 Time-volume graph showing obstructive pattern. Reproduced under Creative Commons licence (Cheek 2014)



The reason the clinical definition of COPD is based on spirometry is because of the relationship between airflow obstruction and mortality. Clinical patterns alone are too inconsistent to diagnose COPD. This was demonstrated by a study (Mannino, Etzel et al. 1993) of 4461 middle-aged Vietnam war veterans, which investigated whether the presence of clinical indications for spirometry would identify those with lung obstruction on spirometry.(Mannino, Etzel et al. 1993) All participants had both clinical and spirometric parameters assessed. Among those without a clinical indicator of COPD, 6% of the never smokers, 6% of the former smokers and 14% of the current smokers had lower than normal FEV1. This study demonstrates that using clinical indicators as a basis for pulmonary function in middle aged men misses many with low lung function and airways obstruction, especially in current smokers.(Mannino, Etzel et al. 1993) Clinical features alone are therefore a poor predictor of airways obstruction and thus mortality.

2.2.5 Diagnosis of COPD in elderly patients

It is important to be aware that both FEV1 and FVC decline proportionately with age and so although, graded as “moderate COPD” many older patients may often have had a “normal” FEV1 *for their age*. Therefore, reporting the FEV1 alone should be regarded as an unreliable marker of obstruction unless considered in the context of FEV1:FVC.(Editorial 2011)

It is now a well-established disadvantage of using the fixed 0.7 threshold that this can lead to a diagnosis of COPD in even the non-smoking elderly because of the age-related decline in FEV1.(Mohamed Hoesein, Zanen et al. ; Mannino, Sonia Buist et al. 2007) An alternative which has been suggested is to use the population’s lower limit of normal – the bottom 5% of the normal population distribution for FEV1/FVC ratio, although a consensus on the “representative population” has not yet been reached.(Price, Yawn et al. 2010) The most commonly used representative population for decisions regarding the percent of predicted FEV1 is the European Community for Coal and Steel Cohort (ECCS) and is over 20 years old, based only on a working white European population.(Quanjer PH, Tammeling GJ et al. 1993)

2.2.6 Co-existing asthma in COPD

An additional complication to the diagnosis of COPD is whether concomitant asthma may be present. Using two large cross-sectional population studies, Soriano et al., (Soriano, Davis et al. 2003) quantified the overlap between various obstructive lung diseases as diagnosed by either the patient’s GP or self-reported. They found that 17% of the United States (US) and 19% of the United Kingdom (UK) populations had been diagnosed with more than one obstructive airway disease. They also found that the prevalence of asthma-COPD overlap increased with age for both men and women.(Soriano, Davis et al. 2003)

How best to manage these older patients with concomitant asthma and COPD is uncertain as they are a group of patients who are often neglected by guidelines because they are often excluded due to co-morbidities and age from the trials that provide the evidence base.(McDonald, Higgins et al. 2013) In addition, there is confusion among clinicians about whether and how they should differentiate asthma from COPD. There is pervading uncertainty among clinicians as to the use of diagnostic labels for obstructive diseases and non-systematic use of terms.(Gibson and Simpson 2009)

The value of diagnosing asthma and COPD overlap syndrome (ACOS) is increasingly being recognised. The two conditions share important risk factors including increasing age, smoking, bronchial hyper-responsiveness and exacerbations. By understanding these risk

factors there may be increased scope for preventing the accelerated loss of lung function that leads to COPD.(Gibson and Simpson 2009)

One study found that subjects with the overlap syndrome had a statistically higher frequency of respiratory symptoms, functional limitation and hospitalisation with respect to subjects with the diagnosis of asthma or COPD alone.(de Marco, Pesce et al. 2013) These findings were echoed in another study that found that ACOS was associated with a lower health related quality of life.(Kauppi, Kupiainen et al. 2011)

2.2.7 Defining COPD for the purpose of this thesis

The definitions given above are primarily based on spirometry. These definitions have been used in a large number of population surveys in order to establish the epidemiology of COPD. In the Obstructive Lung Disease in Northern Sweden Studies, the prevalence of COPD among the over 45s was 14% according to the GOLD spirometry criteria. However, only 18% of 1500 patients invited to the survey for spirometry had previously been diagnosed as having either chronic bronchitis, emphysema or COPD (Lundback, Lindberg et al. 2003). There is, therefore, a major difference between the prevalence of spirometrically defined COPD and physician diagnosed and recorded COPD.

This under-ascertainment of COPD is because many people who have breathlessness or excess sputum production simply dismiss these symptoms as part of older age or “a smoker’s cough” and consider them to be normal: such people never seek medical attention for these symptoms. Following on from this, there have been studies of case-finding of COPD by performing spirometry in smokers (Tinkelman, Price et al. 2007). The yield for case-finding of undiagnosed COPD by spirometry has been shown to be 10-20% of smokers over 40 years of age in one study based in England and Colorado (Tinkelman, Price et al. 2007), and 20.7% (95% CI 18.3-23.4) in another, based in Toronto.(Hill, Goldstein et al. 2010)

For the purposes of this thesis, it is important to keep in mind how COPD has been defined, especially when a prevalence estimate has been provided. The prevalence based on spirometric survey, is the total prevalence of COPD including undiagnosed COPD, and is much higher than the prevalences based on healthcare records or prevalence of “physician diagnosed COPD”. The prevalences are also markedly different depending on the age group in which the prevalence was measured.

For the modelling phase of this thesis, the prevalence of COPD was based on a health services definition of COPD. As described in Chapter 6, a primary care dataset was used to establish the prevalence of physician diagnosed COPD based on a Read Code entry into the medical notes.

The rationale for concentrating on physician diagnosed COPD as opposed to spirometrically defined COPD is that there have been many spirometry-based surveys estimating the prevalence of COPD and several models based on them as can be seen in Section 2.3. However, only physician diagnosed COPD will contribute costs which can be identified as due to COPD. Therefore this is the most useful measure for determining the active burden of COPD on healthcare services. Unless medical practices change to result in greater rates of ascertainment, the burden of COPD on healthcare services in terms of costs and hospital admissions will be dependent on the rate of diagnosed COPD, therefore I have chosen to focus on physician diagnosed COPD. This also enables the use of the Dutch Model with routinely gathered healthcare data as I will expand upon in Chapter 6.

2.3 Epidemiology of COPD

2.3.1 Global epidemiology of COPD

The global epidemiology of COPD has been described in a series of spirometry survey studies under the banner of the Burden of Obstructive Lung Disease or BOLD study.(Chapmann, Mannino et al. ; Buist, McBurnie et al. 2007) These studies covered North America, Europe and Asia and were then extended into South America by the PLATINO studies.(Menezes, Perez-Padilla et al. 2008) These studies were dominated by cross-sectional surveys of prevalence. The aim of the BOLD studies was to standardise these surveys in terms of the use of methods including spirometry.

A recent systematic review summarised what is known about the prevalence of COPD in Europe.(Atsou, Chouaid et al. 2011) It grouped the prevalence estimates according to the criteria used to classify the disease; these estimates are shown in Table 2.1 below. There were three studies that had used questionnaires about symptoms of chronic bronchitis to judge prevalence (Cerveri, Accordini et al. 2001; Huchon, Vergnenegre et al. 2002; Cerveri, Accordini et al. 2003). There were six studies that judged prevalence based on physician reports (Lundback, Nystrom et al. 1991; Montnemery, Adelroth et al. 1998; Hedman, Kaprio et al. 1999; Montnemery, Nihlen et al. 2006; Schirnhofner, Lamprecht et al. 2007; Cazzola, Puxeddu et al. 2011) and 20 studies that used spirometry (Bakke, Baste et al. 1991; Brotons, Perez et al. 1994; Renwick and Connolly 1996; Marco Jordan, Martin Berra et al. 1998;

Dickinson, Meaker et al. 1999; Jaen, Ferrer et al. 1999; Pena, Miravittles et al. 2000; Viegi, Pedreschi et al. 2000; Hasselgren, Arne et al. 2001; Tzanakis, Anagnostopoulou et al. 2004; Murtagh, Heaney et al. 2005; Sichletidis, Tsiotsios et al. 2005; Shahab, Jarvis et al. 2006; Stavem, Sandvik et al. 2006; Buist, McBurnie et al. 2007; Bednarek, Maciejewski et al. 2008; Hansen, Pedersen et al. 2008; Roche, Dalmy et al. 2008; Miravittles, Soriano et al. 2009; van Durme, Verhamme et al. 2009). There were also three studies that had estimated the prevalence using modelling (Stang, Lydick et al. 2000; Feenstra, van Genugten et al. 2001; Peabody, Schau et al. 2005). In addition to the different definitions, the studies were of different sizes and covered different age groups in different countries. This resulted in very variable estimates of the prevalence of COPD. In Europe, the prevalence of COPD varied from 2.1% to 26.0% depending on age group and methods used.(Atsou, Chouaid et al. 2011)

Table 2.1 European prevalence of COPD from Atsou. Reproduced from BMC Medicine under open access licence (Atsou, Chouaid et al. 2011)

Criteria used to define COPD	Study	Country	Sample size	Age (yr)	Prevalence (%)*
Symptoms	Cerveri 2003	Italy	18645	20-44	9.5
	Huchon 2002	France	14076	≥25	4.1
	Cerveri 2001	16 Countries	14819	20-44	2.6
Physician reports	Cazzola 2009	Italy	15229	NS	2.5
	Schirnhofner 2007	Austria	1250	≥40	5.6
	Montnemery 2006	Sweden	3692	20-59	3.6
	Hedman 1999	Finland	3102	18-65	3.7
	Montnemery 1998	Sweden	8469	20-59	4.6
	Lundback 1991	Sweden	6610	35-66	4.1
Functional respiratory tests	Miravittles 2009	Spain	4274	56.6(10.7)	10.2
	van Durme 2009	Netherlands	7983	≥55	11.6
	Hansen 2008	Denmark	4757	45-84	12.0
	Bednarek 2008	Poland	1960	56.7(11.6)	9.3
	Roche 2008	France	4764	59.9(10.1)	2.6
	Buist 2007	Austria	1258	≥40	26.1

		Germany	683	≥40	13.3
		Poland	526	≥40	22.1
		Norway	638	≥40	18.8
	Shahab 2006	United Kingdom	8215	55.5(13.5)	13.3
	Sichletidis 2005	Greece	6112	21-80	16.4
	Murtagh 2005	Ireland	2484	53.3(8.6)	6.3
	Tzanakis 2004	Greece	888	≥35	8.4
	Hasselgren 2001	Sweden	4814	43(14.8)	2.1
	Peña 2000	Spain	3978	40-69	9.1
	Viegi 2000	Italy	1727	≥25	11.0 or 18.3
	Jaen 1999	Spain	497	20-70	7.2
	Dickinson 1999	United Kingdom	353	68.25	9.9
	Marco Jordán 1998	Spain	460	40-60	6.8
	Renwick & Connolly 1996	United Kingdom	783	66.1	9.0
	Brotons 1994	Spain	642	35-65	6.4
	Bakke 1991	Norway	1275	42(16.1)	4.5
Models **	Peabody 2005	Spain	NA	≥30	6.2
		Norway	NA	≥30	6.3
		Poland	NA	≥30	6.7
	Feenstra 2001	Netherlands	NA	≥20	1.5
	Stang 2000	Spain	NA	≥45	10.3
		Italy	NA	≥45	11.1
		France	NA	≥45	10.4
		United Kingdom	NA	≥45	15.0

Notes: Symptoms reported were of chronic bronchitis. Age was reported in various ways, either as a range or mean (SD) is given. Functional respiratory test definitions: FEV1/FVC ratio <70%, FEV1<80% of predicted, FEV1/FVC ratio <65%, FEV1/FVC ratio 70% or FEV1/FVC ratio <88%(male)/89%(female). NS =not specified(Atsou, Chouaid et al. 2011)

*95% Confidence intervals were not available for these values.

**These models used a variety of different techniques and definitions to generate their prevalence estimates. They are all reviewed in Chapter 5.

From the BOLD study, there are estimates available for the prevalence of COPD in other sites around the world, these estimates are shown in Table 2.2. This study used survey techniques to obtain a broadly representative sample of adults over 40 in each region. The study also had strict criteria, using only post-bronchodilator spirometry with a 0.7 FEV1/FVC ratio as definition.(Buist, McBurnie et al. 2007)

Table 2.2 Results from the BOLD study, adults over 40 years (Buist, McBurnie et al. 2007)

Region	Sample size	COPD stages I-IV Prevalence (%) Post- bronchodilator spirometry-based estimates*
Guangzhou, China	473	11.3%
Adana, Turkey	806	19.1%
Cape Town, South Africa	847	23.8%
Reykjavik, Iceland	755	17.9%
Vancouver, Canada	827	19.3%
Lexington, USA	508	19.6%
Manilla, Philippines	893	13.8%
Sydney, Australia	541	19.2%

*Confidence intervals were not published for these values

The authors of the BOLD study commented that their estimates were generally higher than had previously been found. They reflected that this may be a result of prevalence based on symptoms or physician reports leading to under-ascertainment of the prevalence of COPD, in comparison to that based on spirometry.

The PLATINO study added five Latin American cities using the BOLD methodology in adults over 40 years, see Table 2.3. (Blanc, Menezes et al. 2009)

Table 2.3 PLATINO results, adults over 40 years (Menezes, Lopez et al. 2009)

Region	Sample size	COPD prevalence post- bronchodilator spirometry-based estimates (%)*
São Paulo, Brazil	1000	15.8%
Santiago, Chile	1208	16.9%
Mexico City, Mexico	1063	7.8%
Montevideo, Uruguay	943	19.7%
Caracas, Venezuela	1357	12.1%

*95% Confidence intervals were not published for these values

The Global Burden of Disease (GBD) initiative by the World Health Organization (WHO) has included COPD in its list of conditions since its first estimates in 1994.(Murray 1994; Murray and Lopez 1994; Shibuya, Mathers et al. 2001; Lopez, Shibuya et al. 2006; Mathers and Loncar 2006) It uses data available from countries and modelling to calculate disability adjusted life years (DALYs) and life lost to disability or to premature mortality from COPD. They also collate statistics on COPD mortality. In 2001, they produced their estimates that COPD was the fourth most common cause of death in the world. This estimate was updated on the basis of figures from 2010 to the third most common cause of death.(GlobalBurdenofDisease 2012) The methods of the GBD Study differ from the methods used for most of the models explored in this dissertation and are discussed further in the systematic review, Chapter 5.

COPD is clearly a significant problem in the global context. Results from the BOLD study put the prevalence among the over 40 years age group most commonly at around 20%. And the GBD study puts COPD as the third most common cause of death in 2010. Now to explore the importance of COPD in the context of England and Scotland.

2.3.2 Epidemiology of COPD in England

The East of England Public Health Observatory has produced a modelled prevalence estimate of COPD from the 2001 Health Survey of England spirometry data, updated with 2009 demographic data and 2011 smoking rates, see Table 2.4.(ERPHO, Walford et al. 2011) This, therefore, uses a definition of COPD based on spirometry thresholds. This is an

estimate of undiagnosed prevalence and was calculated using Nacul's method.(Nacul, Soljak et al. 2007) This study is reviewed as part of the systematic review in Chapter 5.

Table 2.4 Prevalence of COPD in England from East of England Public Health Observatory

East of England Public Health Observatory(ERPHO, Walford et al. 2011)	Prevalence (%)
Males 16+	4.49
Females 16+	2.28
All persons	3.64

The second part of this thesis which uses the Dutch Model to project the prevalence of COPD in England involved obtaining data from the Clinical Practice Research Datalink (CPRD), limited to its English population. This large database includes over 6% of English general practices. In this case, COPD was determined according to how patients had been coded in their electronic GP records. Therefore this definition only included physician-diagnosed cases, see Table 2.6.

Table 2.5 Incidence and prevalence of COPD in England from Clinical Practice Research Datalink (CPRD)

Clinical Practice Research Datalink 2011	Incidence per person year
All English patients over age 35*	0.00385
Clinical Practice Research Datalink	Prevalence (%)
All English patients over age 35	3.35

*Age over 35 was used because consensus is that COPD cannot reliably be diagnosed in younger age groups.

For the purposes of this thesis, I used the CPRD prevalence of COPD based on physician diagnosed cases. A prevalence of 3.35% among those aged over 35 in England is therefore the baseline prevalence.

2.3.3 Epidemiology of COPD in Scotland

There are two main Scottish sources of primary care data for the diagnosed prevalence of COPD in Scotland both shown in Table 2.6. The first is the Practice Team Information Programme (PTI), which uploads data from 6% of the General Practice (GP) practices in Scotland from the whole practice team including nursing entries. The PTI defines COPD

prevalence as patients who have consulted a GP or practice nurse at least once during the year and coded with a COPD Read Code.(SPHO-PTI 2012) Detailed data of how the PTI prevalence changes with age was sought as part of the modelling phase of this thesis.

A second Scottish source of primary care diagnosed COPD prevalence is the Quality and Outcomes Framework (QOF), which records data from all Scottish practices which are on the New General Medical Services Contract (nGMS) introduced in 2004 who have chosen to submit data for scrutiny and for the award of financial incentives if the data reflect high enough quality practice.(GMC 2003) The COPD prevalence rate from QOF is the proportion of patients who are included on the COPD register across all practices. It is a crude unadjusted rate: practices with more elderly people or greater socioeconomic deprivation are likely to have a higher prevalence of COPD. In 2010/11, there were 857 Scottish practices with patients on the COPD register out of a total of 876 nGMS Scottish practices and 1002 Scottish practices in total.(SPHO 2012)

Table 2.6 Physician diagnosed prevalence of COPD in Scotland (SPHO 2012)

Practice Team Information 2010/11	Prevalence (%)
Males all ages	1.49
Females all ages	1.76
Quality and Outcomes Framework 2010/11	
Males and females, all ages, crude rate	2.04

Incidence data on COPD in Scotland were harder to source. The main source used by government statistics is the Scottish Morbidity Record 01 (SMR01). However, this uses admission to hospital with a COPD episode as a proxy for incidence and so is not a true incidence figure. Better estimates of incidence as number of new diagnoses of COPD were obtained from the Lothian COPD Cohort database. The Lothian COPD Cohort database was constructed from extracts of primary care data from practices in Lothian and is fully described in Chapter 6. The age related COPD incidence rates in the Lothian COPD cohort database was calculated as the number of new diagnoses per time period in the database. The formula for this is presented in Chapter 6.

2.3.4 Conclusions to epidemiology section

Depending on the ages included in the prevalence estimate and whether the estimate is based on a spirometry survey or physician-diagnosed COPD, there is variation in the published prevalences of COPD from 1.5% to 26.1% across the world. There is, thus, a pressing need for standardisation. In those aged over 40, most prevalence estimates are around 20% in

spirometry surveys. The BOLD study in London, UK found a prevalence of 18.9% among men aged over 40 and 16.0% among women aged over 40.(Hooper, Burney et al. 2012)

Among the physician-diagnosed population in Scotland estimates are around 2% (all ages) and in England around 3% (age over 35). This demonstrates that there is likely to be substantial under-diagnosis of COPD. However, the physician-diagnosed population are likely to be those whose symptoms affect them enough to engage in healthcare seeking behaviour. This diagnosed section of the COPD population is thus the section that will consume the most healthcare resources. Therefore, I focused on the physician-diagnosed population for the modelling phase of the PhD to allow projections of healthcare resources to be made according to where they are most likely to be used. Unless COPD screening is introduced or there are much greater resources spent on healthcare it is likely that the rate of COPD ascertainment will remain fairly constant. Alternatively, it can be taken that the projections for COPD prevalence in England and Scotland made by this thesis are very conservative estimates.

In order to perform the modelling undertaken in the later stages of this PhD, it was necessary to have sources for physician diagnosed incidence and prevalence data in England and Scotland. The following sources were therefore selected: English prevalence and incidence data both came from a large sample COPD case-control study extract from the CPRD (see Appendix 1.1). The denominators were from the whole population of CPRD patients. The CPRD population has been shown to be representative of the whole UK population; this is discussed further in Chapter 3. (Campbell, Dedman et al. 2013) Scottish incidence came from the Lothian COPD Cohort Database as using physician recorded primary care Read Codes was felt to be more accurate to provide true new diagnoses rather than the option of SMR01 hospital codes for admissions to provide the incidence rate. Not all cases of COPD require hospital admission and so the hospital coded rate may substantially underestimate the prevalence of COPD. Scottish prevalence came from the PTI Database as discussed with Information Services Division and they recommended this source as more representative than the QOF crude estimates. Further details of all these sources are given in Chapter 6.

2.4 Morbidity, mortality and costs of COPD

2.4.1 Morbidity from COPD

It is estimated that diagnosed COPD is responsible for 1.4 million GP consultations per year across the UK. This is four times more than the number of consultations for angina. (HealthcareCommission 2006)

COPD is the second largest cause of emergency admission in the UK, one in eight (130,000) emergency admissions to hospital is for COPD.(BLF 2007) Overall, respiratory disease is estimated to account for 5.2 million bed days, COPD itself accounts for more than one million bed days each year in hospitals in the UK.(BTS 2006)

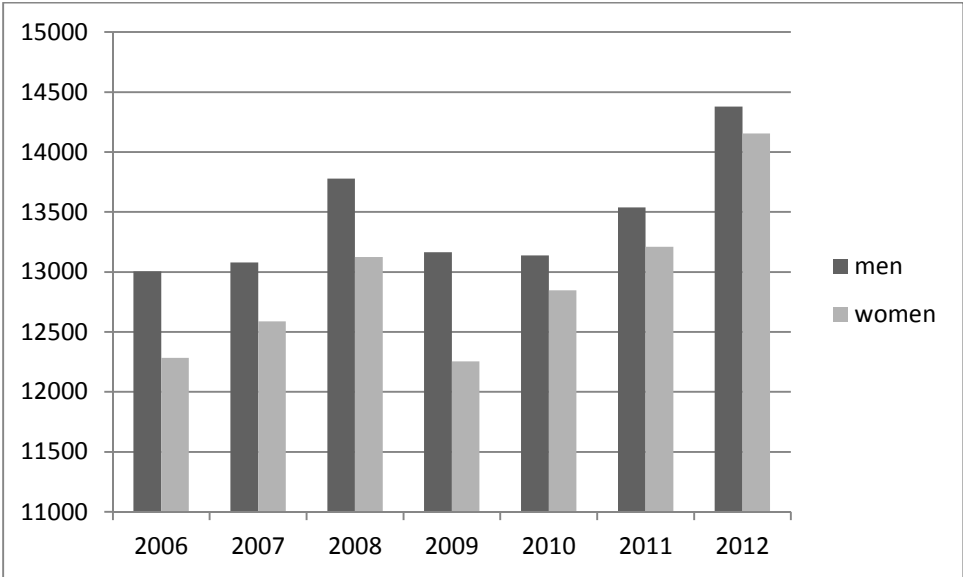
About 30% of patients admitted with COPD for the first time will be readmitted within three months.(BTS 2006) Admission patterns vary by up to five times in different parts of England; this reflecting differences in the underlying prevalence of COPD and wide variations in the quality of care that is provided in the community.(HealthcareCommission 2006)

2.4.2 Mortality from COPD

It is estimated that over 30,000 people die from COPD every year in the UK.(HealthcareCommission 2006) There is some controversy over the official figures for mortality from COPD as many patients die with the disease rather than of the disease, as discussed below.

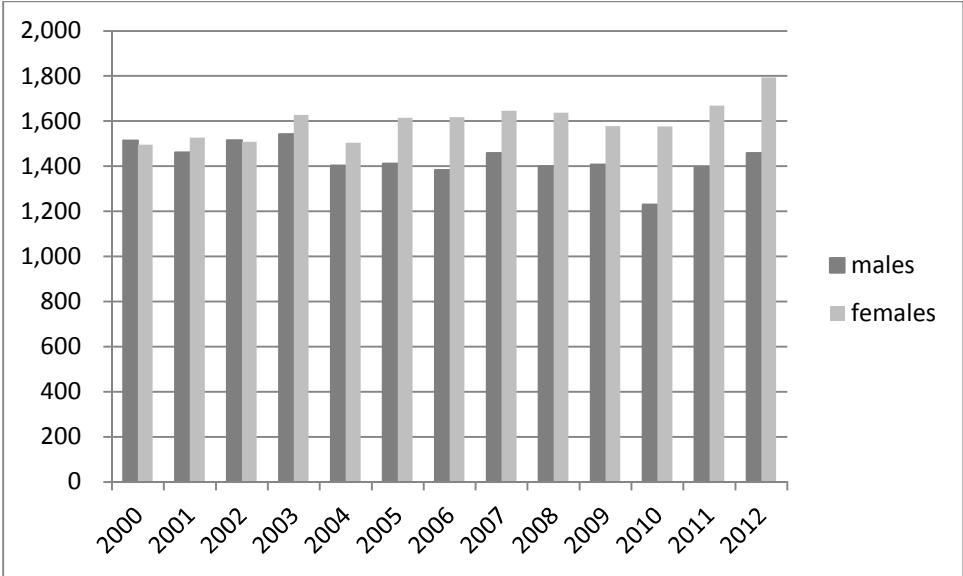
There follows data for England and Wales, these figures are coded in International Classification of Diseases 10 (ICD-10) codes from the death certificate.(WHO 2010) Codes J40 to J47 are the chronic lower respiratory diseases to which COPD is the major contributor. If COPD is recorded as the underlying cause of death on the death certificate then the death will be registered as a chronic lower respiratory disease death. In England and Wales it appears that there are more deaths among males, and that deaths are increasing overall among both sexes, see Figure 2.3.

Figure 2.4 Number of deaths identified from death certificates with codes of J40 to J47 identified as underlying cause of death in England and Wales 2006-2012 by sex



The General Register Office for Scotland (GROS) has released the following figures for mortality from COPD in Scotland, see Figure 2.4.(GROS 2012)

Figure 2.5 Number of deaths from GROS with codes of J40 to J47 identified as underlying cause of death in Scotland 2000-2012 by sex



Here it can be seen that deaths from COPD are slightly higher among females than males, in Scotland, and that deaths are perhaps continuing to increase for females while remaining steady for males.

However, overall these mortality figures present a misleading picture of the burden of mortality among COPD patients because so many patients with COPD have significant smoking-related co-morbidities and, therefore, die with COPD rather than of COPD (Hansell, Walk et al. 2003). It has been shown that 8.0% or 312,664 deaths in England and Wales over a six year period had obstructive lung disease (including asthma) mentioned on the death certificate. Of these, 59.8% had obstructive lung disease comprising the underlying cause of death. In the remainder of deaths the leading causes were acute myocardial infarction, other ischaemic heart disease, lung cancer and bronchopneumonia.(Hansell, Walk et al. 2003) It can be concluded from this study that information on COPD from death certificates significantly underestimates the burden of the disease.(BLF 2007)

2.4.3 Costs of COPD

The direct cost of COPD to the UK healthcare system was previously estimated to be between £810-£930 million a year.(HealthcareCommission 2006) Direct costs are the cost of medical care for routine and emergency management of COPD. More than half these costs relate to the provision of care in hospital. COPD is among the most expensive conditions treated by the National Health Service (NHS) in the UK.(BTS 2006; NICE 2010)

The indirect costs of COPD are substantial, however, they are difficult to quantify. Indirect costs include the “social and economic costs” of lost days from work through ill health and informal caring for relatives with the disease. One estimate of the impact of annual productivity is 24 million lost working days per annum.(HealthcareCommission 2006; NICE 2010)

One important outcome from this PhD is an estimate of the direct cost of caring for COPD patients in hospitals and primary care for Scotland and England in 2011 and projections without inflation to 2030.

2.5 Aetiology of COPD

2.5.1 Prenatal and postnatal lung growth

Adult COPD results from an accelerated decline in lung function as proposed by Peto.(Peto and Fletcher 1977) However, failure to attain maximal airway growth also has a role. Prenatal lung growth is relevant; lung function is reduced among children of low birthweight whether as a result of prematurity or intra-uterine growth retardation. (Rona, Gulliford et al. 1993) In addition, lung function is reduced among children whose mothers smoked during pregnancy, and there is also an effect of childhood exposure to second-hand smoke on lung

function. (Cook, Strachan et al. 1998). Thus, both prenatal and postnatal airway growth affect the maximal lung growth reached in early adulthood and, therefore, the baseline from which lung function decline begins.

2.5.2 Bronchial hyper-responsiveness

The role of reversible airways narrowing related to asthma in the generation of COPD remains controversial. Bronchial hyper-responsiveness describes an inflammatory process in the airways initiated in atopic persons by allergen exposure. If this inflammation continues it is thought to promote remodelling of the airways, which leads to irreversible airflow obstruction, such as is found in COPD. In this way, people who do not smoke may develop COPD (or overlap syndrome as previously discussed in Section 2.2.6) as a result of remodelled asthma. However, in people who do smoke bronchial hyper-responsiveness may be a feature of the immune response to smoking.(Strachan and Sheikh 2004)

2.5.3 Effects of childhood respiratory tract illness

Barker's study found that bronchitis, pneumonia and whooping cough before the age of five years were associated with a reduced FEV1 in adulthood.(Barker, Godfrey et al. 1991) Another study found that a medical history of bronchitis, pneumonia or asthma in early life was associated with a more than twofold increase in COPD mortality (adjusted hazard ratio 2.37, 95% CI 1.16-4.83)(Galobardes, McCarron et al. 2008) Such findings could reflect lung damage due to early episodes of chest infection, a longstanding susceptibility to all forms of lung disease, continued environmental exposure or socioeconomic disadvantage.(Strachan and Sheikh 2004) Of these suggested mechanisms, longstanding susceptibility or continuing environmental exposure are favoured. This is as a result of some research that scrutinised the impact of childhood infections with respect to season of birth. Babies born in the autumn are at substantially greater risk of viral bronchiolitis, bronchitis and pneumonia as a result of seasonal fluctuations in environmental viruses. If such early infections have a direct impact on adult lung function or increased respiratory symptoms then these outcomes should be more frequently observed among autumn births. However, such an association was not found in several large datasets. (Strachan, Seagroatt et al. 1994)

2.5.4 Smoking and COPD

The definition of COPD characterises it as arising from *enhanced chronic inflammatory response in the airways and the lung to noxious particles or gases*(GOLD 2013).

In the UK and most other economically-developed nations, cigarette smoke is overwhelmingly the most common risk factor. The epidemic of COPD in the Western world

has followed the epidemic of smoking. The classic Doll et al. study of smoking doctors in the 1950s demonstrated a dose-response relationship between severity of lung disease and number of cigarettes smoked.(Doll, Peto et al. 2004)

Other exposure factors may have a role in specific populations where rates of cigarette smoking are insignificant or absent, for example in China, South East Asia and parts of Africa exposure to biomass cooking fuel smoke in poorly ventilated indoor settings results in COPD.(Norman, Barnes et al. 2007) This is particularly manifest in non-smoking women. Outdoor air pollution from factories and traffic may also have a role.

In addition “noxious particles or gases” may come from industry and occupational dusts. The lines of definition and diagnosis are not clear between COPD due to smoking and fixed lung obstruction as a result of occupational exposure to dusts and particles; however, diseases such as silicosis, pneumoconiosis, asbestosis and psitticosis are generally considered to be distinct from COPD. For the purposes of this PhD, COPD is defined as overwhelmingly caused by smoking; exposure to other dusts are not taken into account.

2.5.5 Phenotypic variation

One of the great mysteries of COPD is to explain why there are some people who may smoke all their lives and yet be only mildly, if at all, affected by COPD. FEV1 decline is present in 20-50% of smokers.(Lundback, Lindberg et al. 2003) It needs to be explained why 50-80% of smokers do not suffer a demonstrable decline in FEV1. Recent genetic and immunology studies (see Sections 2.5.6 and 2.5.7) are only just beginning to shed light on the mechanisms by which some people succumb to COPD whereas others remain well, despite a lifetime of smoking.(Lundback, Lindberg et al. 2003)

2.5.6 Smoking and genetics

Much work has been going on over the past few years to try and discover the genetic associations of COPD. It has been known for some time that the genetically determined absence of alpha-1-anti trypsin is associated with accelerated lung function decline and premature death, especially in smokers. However, more recently, the sequencing of the human genome enables Genome Wide Association Studies (GWAS). This methodology enables candidate single nucleotide polymorphisms (SNPs) to be mapped from the genome to clinical phenotypes, such as FEV1 decline, in population cohort studies.(Pillai, Ge et al. 2009)

For example, the first GWAS study for COPD found two candidate polymorphisms on chromosome 15 at a locus which has been previously described as formally associated with smoking behaviour and lung cancer.(Pillai, Ge et al. 2009) The so-called *AGPHDI/ cholinergic receptor nicotinic alpha 3 (CHRNA3)* on chromosome 15q25 was thought to be a genetic determinant of smoking intensity. It has recently been confirmed using statistical mediation studies (see glossary) with a large cohort of genetically profiled COPD smokers that *CHRNA3* affects COPD via pathways affecting smoking intensity as measured by pack-years. In other words, this gene mediates nicotine addiction and has been shown to have a relationship with COPD. (Siedlinski, Tingley et al. 2013) Other identified gene variants *FAM13A* (family with sequence similarity 13 member A) and *HHIP* (Hedgehog interacting protein) SNPs have been shown not to act through these pathways although they are still postulated to be COPD susceptibility genes.(Siedlinski, Tingley et al. 2013)

Also under investigation is how different genes control different people's level of inflammatory response to an inflammatory stimulus such as cigarette smoke or an infection. Some cell studies have shown that cells from those with COPD are prone to an exaggerated and potentially more locally harmful immune response to a common infectious agent such as rhinovirus.(Baines, Hsu et al. 2013)

2.5.7 Smoking and the immune response

Inhalation of smoke deposits carbon, oxygen free radicals and other particles in the lung; to which there is an immune reaction. The body increases the production of mucous and the number of goblet cells in an attempt by the mucociliary ladder to clean the lung of noxious and foreign particles.(Ganesan, Comstock et al. 2013) As smoking continues there will be release of pro-inflammatory cytokines by airway epithelial cells which will recruit innate immune cells to the lungs(John, Kohse et al. 2014). These neutrophils and macrophages then release oxygen radicals and proteolytic enzymes such as neutrophil elastase and matrix metalloproteinases. The adaptive immune system also has a role: CD8+ T cells and natural killer cells also destroy lung parenchyma by releasing perforin and granzyme B. There also begins a process of chronic inflammation with lymphoid neogenesis: the iBALT – inducible Bronchus Associated Lymphoid Tissue which organises adaptive immune cells into lymphoid follicles around the small airways. In these follicles antigen retention, immunoglobulin class switching and affinity maturation occur.(Brusselle, Joos et al. 2011) The chronic inflammation spurs airways remodelling with increased deposition of fibrin which also reduces the lung's elastic recoil (known as compliance, see 2.6) and eventually there is squamous metaplasia and consequent loss of gas exchange surface (Herfs, Hubert et

al. 2012). There is increasing evidence that these changes represent a type of accelerated ageing in COPD.(MacNee 2009; MacNee and Tudor 2009)

Recently, scientific consensus has grown that the lungs in COPD are colonised by bacteria and viruses even when the disease is stable. Techniques such as quantitative polymerase chain reaction (PCR) have enabled the bacteria *H influenzae*, *S pneumonia*, and *Moraxella catarrhalis* to be detected in about 25% of patients with stable COPD and in more than 50% of patients during exacerbations.(Garcha, Thurston et al. 2012) Macrophages from patients with COPD have demonstrated impaired phagocytic responses to bacteria compared with non-smokers and smokers who do not have COPD. So it seems that those who develop COPD have an abnormal immune response.(MacNee 2009; Garcha, Thurston et al. 2012)

The role of the abnormal or exaggerated immune response is also being investigated in exacerbations. It has been discovered that during exacerbations there is an increase in the concentration of bacteria that colonise the lower airways and that colonisation with a new bacterial strain is also often a key element in the development of an exacerbation.(Brusselle, Joos et al. 2011) Viruses can be detected in 10-15% of sputum samples in patients with stable COPD and 30-60% of those with COPD exacerbations. Rhinovirus and influenza virus are most commonly associated with viral exacerbations. Sputum eosinophils are increased during viral exacerbations.(Brusselle, Joos et al. 2011) Exacerbations are an important feature of COPD and an important contributor to the overall disability from COPD.

2.6 Pathophysiology of COPD

In terms of the pathophysiology of COPD the key feature is airways resistance, also termed airways obstruction. The destruction of the lung parenchyma and, in particular, the loss of elastin reduces a property of the lung known as “compliance”. The compliance of the lung is the elastic recoil that returns the lung to its resting volume following normal inspiration. Firstly, during inspiration, the diaphragm descends and the rib cage expands resulting in a low pressure in the lung cavity. Air is drawn down the trachea, along the bronchi, the bronchioles and into the lung parenchyma. As the air pressure inside the chest cavity equalises to that outside the chest cavity and the rib cage and diaphragm are no longer moving dynamically, the expiration phase begins. In contrast to inspiration, expiration is passive. The diaphragm and intercostal muscles relax and the only force acting is the elastic recoil of the lungs, (i.e. their compliance). The compliance of the lung tissue holds open the lung air sacs and bronchioles by what is known as the “guy rope effect”, (as in the way of guy ropes on a tent) as the lung inflates and deflates thereby maximising the time for gas

exchange to take place. Since the resistance to airflow is inversely and exponentially related to the diameter of the airway, small decreases in airway diameter lead to large increases in resistance.(Holleman and Simel 1995) Thus, in COPD, the “guy rope effect” is reduced as a result of the reduction of elastin in the lung and so the distal airways narrow, causing airway resistance and obstruction. The airways may even collapse trapping deoxygenated air in the terminal bronchioles and air sacs, thereby inhibiting efficient gas exchange.

2.7 Clinical features

Traditionally patients with COPD have been described as “pink puffers” or “blue bloaters”.

The “pink puffers” are affected by air trapping in bullae which results in the high residual volumes which manifest clinically as a barrel shaped chest and palpable heart apex.(Lemyze and Bart 2011) Patients cope naturally with the air trapping by pursing their lips on expiration, thus maximising the pressure inside the chest cavity for as long as possible to prevent the collapse of the distal air spaces. Patients remain pink because their PaO₂ (partial pressure of oxygen) is normal enough for their haemoglobin to be saturated with oxygen and they are described as puffers because in later stages of the disease they have to work hard with pursed lip breathing and an increased respiratory rate to maintain this PaO₂. They also use their accessory muscles to help with this increased work of breathing. Hoover’s sign is common; this is the paradoxical inward movement of the lateral diameter of lower part of the rib cage with inspiration and is a sign of airflow obstruction.(Bruyneel, Jacob et al. 2011)

“Blue bloaters” have blue lips due to deoxygenated haemoglobin circulating as they are unable to maintain blood oxygen saturation and their PaO₂ falls centrally. They often suffer with co-existent right heart failure leading to fluid retention, peripheral oedema and ascites: hence the appearance of bloating. The presence of central cyanosis is an important clinical finding as it indicates that the brain is likely to have switched from a hypercapnic respiratory drive for respiration to a hypoxic drive. In this setting the provision of supplemental oxygen, if not carefully titrated, may improve the oxygen saturation to a level that the hypoxic drive in the brainstem switches off and the patient stops breathing. This may have the unintended consequence of potentially fatal acidosis.

Emphysema (pink puffer) and chronic bronchitis (blue bloater) is the traditional clinical dichotomy but it is unsatisfactory for many reasons. The pathologies of emphysema and mucous hypersecretion co-exist in the majority of smokers and whether the pink puffer or blue bloater clinical features are in evidence is often to do with the length of time the patient has been smoking. A pink puffer is at an earlier stage of the disease as their hypercapnic

drive is still functioning in their brainstem and they consequently raise their respiratory rate to compensate. The blue bloater has lost sensitivity to high CO₂ and so their blood pH is likely to be slightly acidotic as this gas accumulates.

2.7.1 Frequent exacerbator phenotype

COPD exacerbations are an important cause of COPD morbidity and hospital admissions. The frequent exacerbator phenotype is a more useful clinical identity than either pink puffer or blue bloater. COPD exacerbations affect FEV1 decline independently of cigarette smoking. Exacerbations are associated with increased airway inflammation, lung function decline and increased bacterial colonisation. Following an exacerbation there is increased susceptibility to further exacerbations. Exacerbations have been observed to cluster together in time in patients.(Wedzicha and Donaldson 2012)

It emerged from the ECLIPSE study that one relatively consistent group of patients were particularly prone to frequent exacerbations.(Vestbo, Agusti et al. 2014) The frequent exacerbators were identified by a history of two or more exacerbations in the past year. They were found to represent a stable phenotype and also exhibited decreasing FEV1, reflux heartburn and a high white cell count. Conversely, patients who had had no exacerbations in year one or two of the study had a 75% chance of no exacerbations in the third year.(Vestbo, Agusti et al. 2014) In the future it may prove possible to trial new therapies and intensive therapies specifically targeted to help this group (i.e. using a personalised medicines approach).

2.8 Prognosis

Recent research has provided a basis for the following statements regarding the prognosis of COPD and has updated the Fletcher-Peto trajectory.(Vestbo, Agusti et al. 2014) COPD is a very heterogeneous disease with a very variable course over time. COPD is now regarded as a syndrome-type disease in which symptoms, health status and exercise capacity fluctuate over time, but may be poorly related to the FEV1.(Editorial 2007) The problem is that the longitudinal FEV1 does not correlate very highly with patient-reported outcomes as assessed by a health status tool (for example the St George's Respiratory Questionnaire). Change in FEV1 is only weakly correlated with symptom scores over time.(Kesten, Celli et al. 2011) This suggests that the declining FEV1 does not necessarily capture how the patient feels "in themselves" as reflected in their level of coughing, breathlessness and exercise capacity therefore, spirometry alone provides a very incomplete picture of COPD. (Westwood, Bourbeau et al. 2011) There is increasingly recognition that COPD is a multi-system

disorder and that factors such as Body Mass Index (BMI) and cardio-respiratory exercise reserve indicate prognosis more accurately than FEV1 alone because they consider COPD in the context of its common and important co-morbidities.(Editorial 2007; Wood 2011) Similarly, improvements in clinical symptoms do not necessarily correlate with slowing of FEV1 decline as assessed by spirometry.(Qaseem, Wilt et al. 2011) Such considerations have led to the development of new ways of assessing prognosis. In this context, there is a reframing of the importance of FEV1 decline and other physiological measurements including CT-measured lung density as a quantification of emphysema, bronchodilator reversibility, systemic inflammatory markers and exercise capacity. These will be discussed below.

2.8.1 FEV1 decline

The original Fletcher-Peto trajectory assumed a smooth longitudinal progressive decline in lung function. In this context, the airflow obstruction as determined by the % of predicted FEV1 was used as a guide to the severity of the COPD, classifying a patient as “mild, moderate, severe or very severe” as below. Spirometry is performed in patients in their stable non-exacerbated state to guide the classification. The latest NICE guidelines harmonised the definitions of each stage of airflow obstruction with the GOLD/ATS/ERS guidelines(Qaseem, Wilt et al. 2011). The thresholds of airways obstruction used in the NICE and GOLD/ATS/ERS guidelines are shown in Table 2.7.(GOLD 2010; Qaseem, Wilt et al. 2011)

Table 2.7 Thresholds of airways obstruction used to define severity categories in COPD

		National Institute for Clinical Excellence (NICE) Paper 12 2004	American Thoracic Society/ European Respiratory Society (Qaseem, Wilt et al. 2011)	Global initiative on Obstructive Lung Disease (GOLD 2010)	NICE Paper 101 update (NICE 2010)
Post-bronchodilator FEV1/FVC	FEV1 % predicted	Severity of airflow obstruction			
< 0.7	≥ 80%		Mild	Stage 1 – Mild	Stage 1 – Mild*
	50-79%	Mild	Moderate	Stage 2 – Moderate	Stage 2 – Moderate
	30-49%	Moderate	Severe	Stage 3- Severe	Stage 3- Severe
	< 30%	Severe	Very Severe	Stage 4- Very Severe**	Stage 4- Very Severe**

* symptoms should be present to diagnose COPD in people with NICE 2011 mild airflow obstruction.

**or FEV1<50% with respiratory failure

However, there have been few longitudinal studies to confirm the Fletcher-Peto trajectory and the latest evidence shows that the FEV1 decline with time may be less and more variable than was stated in their original paper. The ECLIPSE study was a longitudinal prospective observational study of moderate, severe and very severe COPD patients which showed a mean rate of change (+/- SD) in FEV1 as a decline of 33(+/-2) ml with significant variation

across the study.(Vestbo, Edwards et al. 2011) Over the three year study period 38% of patients had a decline in FEV1 of more than 40ml per year, 31% had a decline of 21ml to 40ml, 23% had a change between a decline of 20ml per year and an increase of 20ml per year, and 8% had an increase of more than 20ml per year. The mean rate of decline was 21(+/-4) ml greater in current smokers than current non-smokers, 13ml (+/-4) greater among patients with emphysema on Computer Tomography (CT) Scan and 17(+/-4)ml per year greater in those with bronchodilator reversibility than in those without bronchodilator reversibility. An association with lung function decline of low levels of club cell secretory protein was also found, confirming findings from the Lung Health Study.(Park, Churg et al. 2013) In the ECLIPSE study, all patients were receiving treatment for their COPD from their doctors (Vestbo, Edwards et al. 2011). It would seem that some of the variable decrease in lung function demonstrated may be accounted for by a treatment effect in conjunction with different underlying susceptibilities. And it might also be expected that the FEV1 decline might be larger in untreated patients.

2.8.2 Computer Tomography (CT) defined lung density over time in COPD

It is now known that lung density changes over time in COPD are dependent on smoking status and gender. Smokers and women lose lung density more rapidly; however, variability between patients is significant. Change can be assessed using CT and sRAGE (soluble Receptor for Advanced Glycation Endproducts) may potentially be a biomarker for assessment of emphysema development in the future.(Vestbo, Agusti et al. 2014)

2.8.3 Bronchodilator reversibility

The latest evidence is that this is very variable among COPD patients. It is also unstable within individual patients and therefore difficult to use clinically. At present it is not thought that bronchodilator reversibility will be useful in defining subtypes of COPD.(Vestbo, Agusti et al. 2014)

2.8.4 Systemic inflammation

The ECLIPSE study confirmed that not all COPD patients have elevated biomarkers of systemic inflammation, one-third of participants had no evidence of elevated biomarkers of systemic inflammation for the three year follow-up. It appears the IL-8 and TNF Alpha are markers of smoking rather than markers of systemic inflammation. However, there was a small group (16%) of COPD patients in the ECLIPSE study who had persistent systemic inflammation and this was associated with a six-fold increase in all-cause mortality at three

years of follow-up. This finding makes way for using systemic inflammatory markers to improve prediction of mortality.(Vestbo, Agusti et al. 2014)

2.8.5 Exercise capacity and the BODE index

Limb muscle dysfunction is a major component of reduced exercise tolerance and also affects quality of life and mortality.(Maltais, Decramer et al. 2014) COPD results in metabolic and structural changes in the muscle. Physical inactivity contributes to muscle dysfunction but inflammation, oxidative stress, nutritional imbalance and hypoxaemia also have a role. Muscles are also compromised during COPD exacerbations.

Exercise capacity is of prognostic value in COPD. B -BMI, O-airflow Obstruction as measured by FEV1, D-Dyspnoea as gauged by the Medical Research Council's (MRC) dyspnoea scale, and E-Exercise capacity by way of the six-minute walk test, form the BODE index prognostic tool. (Puhan, Garcia-Aymerich et al. 2009) The BODE index has been evaluated recently for prognostic prediction and found to compare favourably with the GOLD 2011 rubric for combining predictive elements (see section 2.8.8)(de Torres, Casanova et al. 2014)

2.8.6 Co-morbidities

There is increasing recognition of the value of co-morbidities in predicting prognosis in COPD. The 2013 GOLD guidelines have been adapted to include consideration of comorbidities when assigning a severity class to a COPD patient. COPD patients have a higher prevalence of osteoporosis, anxiety and panic attacks, ischaemic heart disease and heart failure.(GOLD 2013)

2.8.7 The DOSE Index

Another important assessment tool is the DOSE index.(Jones, Donaldson et al. 2009) This is D-Dyspnoea (MRC dyspnoea scale), O-airflow Obstruction, S-Smoking status and E-Exacerbations. This is a pragmatic test designed to be used in clinical settings as a multicomponent prognostic tool. According to multiple regression analysis, the four DOSE variables explain 48% of the variance in health status. There is no need for the six minute walk test which limits the applicability of the BODE index in routine clinical settings where space, staffing and time are limited. The DOSE index is not only a measure of COPD severity but a guide to management. If the patient is suffering from dyspnoea they should be referred for pulmonary rehabilitation. If they have a certain level of obstruction they should be offered steroid inhalers, if they suffer recurrent exacerbations then long-acting bronchodilators would be more appropriate and if they are still smoking they should be given

smoking-cessation help. The DOSE index has been validated against other markers, for example, against healthcare utilisation in large databases, and has been shown to be useful for predicting future events such as hospital admissions and exacerbations.(Jones, Donaldson et al. 2009)

2.8.8 Assessment of COPD severity as per GOLD guidelines

The above considerations resulted in the latest update of the GOLD guidelines recommending that the following four components be assessed separately in order to categorise a patient's severity of COPD: degree of airflow limitation using spirometry (as above), symptoms, risk of exacerbations and comorbidities according to a rubric in the latest GOLD guidance.(GOLD 2013)

Symptoms are assessed with a validated questionnaire such as the COPD Assessment Test (CAT), the Modified British Medical Research Council (MRC) breathlessness scale or the clinical COPD Questionnaire (CCQ). A history of frequent exacerbations and comorbidities are also taken into account.

Although there is evidence that all the components identified by the GOLD algorithm influence prognosis, the level of effect is usually not exactly known. How the factors interact and their combined prognosis is only crudely predicted by the rubric and further research is needed in this area to refine predictions.

2.8.9 Life expectancy in COPD

A recent Swedish study(Stallberg, Janson et al. 2013) of retrospective linked electronic medical record data from COPD patients has confirmed that people with COPD have a markedly reduced life expectancy. Overall life expectancy (standard deviation) for the COPD population was 8.3 (6.8) years lower than in the average Swedish population; 9.4 years lower for women and 7.4 years lower for men. All-cause mortality was 3.5 times higher in the study population (age-standardised) compared with the death rate in the Swedish population.(Stallberg, Janson et al. 2013)

Life expectancy for COPD has not so far been described for Scotland and England.

2.9 Management of stable COPD

The 2011 Clinical Practice update from the European Respiratory Society, American College of Physicians, American College of Chest Physicians and American Thoracic Society emphasises the wide variation between individuals with COPD.(Qaseem, Wilt et al. 2011)

They state that spirometric decline cannot be used to measure individual long term response to treatment. Instead, the goals of COPD treatment are described as follows:

1. Reduce lung function decline
2. Prevent and treat exacerbations
3. Reduce hospitalisations and mortality
4. Relieve disabling dyspnoea
5. Improve exercise tolerance and health related quality of life.

2.9.1 Smoking cessation

Smoking cessation is the most important strategy for preventing and improving COPD. (NICE 2010) A Cochrane review (van der Meer, Wagena et al. 2003) found that psychosocial therapies in combination with pharmacological therapies were more effective than psychosocial therapies alone among COPD patients at five years (risk difference = 0.17, 95% CI 0.14 to 0.19, relative risk 4.19, 95% CI 3.41 to 5.15). However, the review identified a need for further research to test the effectiveness of psychosocial therapies alone and to determine the most effective psychosocial intervention.

2.9.2 Inhaled therapies

The mainstay of COPD treatment (apart from smoking cessation) is inhaled therapies. There is a consensus that long-acting bronchodilators, inhaled corticosteroids or combination therapies reduce the annual rate of decline of FEV1 more than placebo. (Qaseem, Wilt et al. 2011) A Cochrane review of 23 studies found evidence for a significant change in FEV1 with salmeterol 50 mcg twice daily of 51 mls (95% CI 32 to 70).(Appleton, Poole et al. 2006) A Cochrane network meta-analysis put inhaled treatments in the following order of efficacy according to their effects on the St George's Respiratory Questionnaire and trough FEV1: 1) inhaled corticosteroid/long acting beta agonist combination, 2) long acting muscarinic antagonist, 3) long acting beta-agonist and 4) inhaled corticosteroid. However, the seeming advantage of combination inhalers is tempered by their expense and their capacity for adverse events including pneumonia.(Kew, Dias et al. 2014; Kew and Seniukovich 2014)

In terms of exacerbations, treatment with salmeterol combined with fluticasone, tiotropium, other long acting beta-agonists and other inhaled corticosteroids reduced annual rates of exacerbations.(Wilt, Niewoehner et al. 2007). Tiotropium also reduced the number of hospitalisations compared with placebo.(Niewoehner, Rice et al. 2005)

However, in terms of mortality the TORCH trial (TOwards a Revolution in COPD Health) showed no difference in pulmonary cause mortality with salmeterol or fluticasone or combined therapy versus placebo.(Calverley, Anderson et al. 2007) One meta-analysis has found an increase in pulmonary cause mortality in association with long-acting beta-agonists compared with placebo.(Salpeter, Buckley et al. 2006) Overall, there remains insufficient evidence to choose between a combination of inhaled steroid and long-acting beta agonist and tiotropium in terms of their relative efficacy and safety.(Welsh, Cates et al. 2013)

Inhaled therapies are most likely to benefit patients who have FEV1<60% predicted and symptoms, above this level there is little to no improvement in exacerbations, health related quality of life, COPD hospitalisations or mortality. In all patients, combination therapies do not consistently demonstrate benefits over monotherapy although they may be useful in some patient subgroups. However, adverse events may be more common.(Qaseem, Wilt et al. 2011)

2.9.3 Pulmonary rehabilitation

In people who have symptoms and FEV1<50% predicted pulmonary rehabilitation has been shown to improve symptoms, quality of life and the distance on the six minute walk test. There is also some moderate quality evidence that pulmonary rehabilitation is an effective intervention to reduce hospital readmission after exacerbations.(Puhan, Scharplatz et al. 2009; Puhan, Gimeno-Santos et al. 2011) Pulmonary rehabilitation also reduces mortality after exacerbations (Odds Ratio 0.28; 95% CI 0.10 to 0.84).(Puhan, Gimeno-Santos et al. 2011)

2.9.4 Oxygen therapy

There is widespread consensus that supplemental oxygen used for 15 or more hours daily in order to maintain partial pressure of oxygen above 60mmHg in patients with severe resting hypoxaemia <55mmHg reduces mortality. Physiological indications for oxygen therapy include cor pulmonale or polycythaemia with PaO₂ 55-59mmHg.(Qaseem, Wilt et al. 2011)

2.10 Management of COPD exacerbations

COPD exacerbations are significant events in the natural history of COPD. They are upsetting for patients and account for much of the costs associated with caring for COPD patients. Commonly, symptoms include worsening breathlessness, cough, increased sputum production and change in sputum colour. Bacteria and viruses are common causes of exacerbations. The value of antibiotics for less severe exacerbations remains controversial, particularly in the context of increasing antibiotic resistance, and further research is needed

to identify those who will best respond to antibiotics.(Vollenweider, Jarrett et al. 2012)
Therapy for exacerbations include inhaled and nebulised therapies, oral steroids, supplemental oxygen and, when required, non-invasive ventilation techniques as well as the full armoury of modern intensive care units when considered appropriate.(NICE 2010)

2.10.1 Structure of care for people with COPD exacerbations

In the UK, there has been a move to setting up anticipatory care services for COPD.(NICE 2010) This is with a view to preventing hospital admissions for exacerbations and managing a greater proportion of patients in primary care. However, the ethos of such services is very different from the rest of the NHS where a more traditionally “reactive” approach is the norm. There are consequently many challenges to making such services holistic, patient-centred and able to ensure good communication between the relevant multidisciplinary professionals.(Cleland, Moffat et al. 2012)

Other approaches to managing COPD in the community include increased use of telehealthcare, where the patient uses information technologies to communicate with the healthcare professional at a distance. A Cochrane systematic review has found that telehealthcare can decrease hospital admissions for COPD.(McLean, Nurmatov et al. 2011; McLean, Nurmatov et al. 2012)

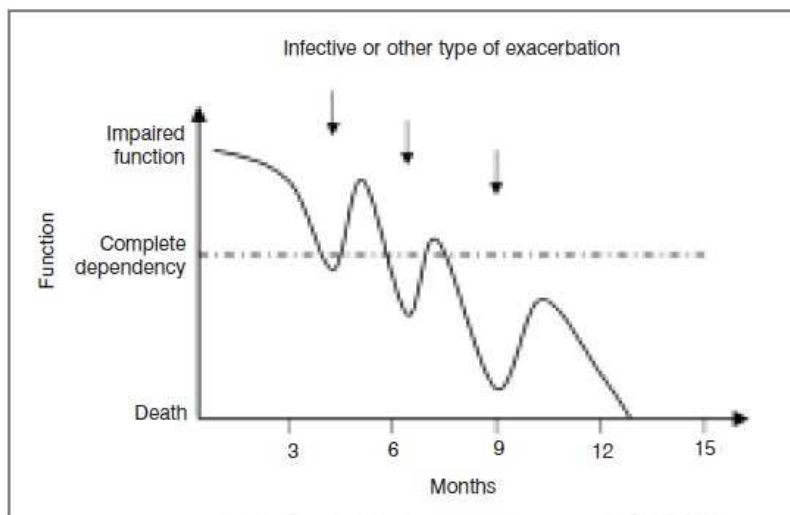
2.11 Patient perspectives in COPD

Of course patients do not measure their COPD in terms of the FEV1 and FVC, but in terms of how they are feeling “in themselves” and what they are able to do. The characteristic respiratory symptoms of COPD include dyspnoea (at rest and during exercise), chronic cough, sputum production, wheeze and chest tightness. Patients report dyspnoea to be the most bothersome symptom and it is often the primary reason for their seeking help.(van der Molen, Miravittles et al. 2013) The onset of dyspnoea is gradual and patients often attribute it solely to aging or a lack of fitness. However, it is progressive and soon becomes an intrusive symptom, causing disability, especially as patients limit their activity to avoid dyspnoea which results in deconditioning and worsening dyspnoea.(van der Molen, Miravittles et al. 2013) Cough is another troublesome symptom in COPD and has notable effects on quality of life.(Calik-Kutukcu, Savci et al. 2014) The extra-pulmonary impact of COPD includes fatigue, muscle weakness, weight loss and sleep disturbance. Symptoms fluctuate for patients within the day with mornings often being particularly bad, but this may vary from day to day.(Giacomini, DeJean et al. 2012)

In contrast to other conditions, qualitative research has shown that patients with COPD view their condition as a “way of life” or “health problem” rather than an illness which disrupts life and from which they may expect to recover. Patients have great difficulty identifying the beginning of their illness and describe a “chaos narrative” of their illness: disjointed events associated with COPD in which they can see no clear purpose or direction.(Pinnock, Kendall et al. 2011)

Disease progression in COPD is not smooth from the patient’s point of view. Gradual decline in the early years becomes more notable with exacerbations. There are steep deteriorations in function with exacerbations followed by recovery to impaired function (see Figure 2.5). Death is often unpredictable and relatives and doctors often express surprise when someone dies because they have rallied so many times before. (Lehman 2004; Giacomini, DeJean et al. 2012) This trajectory and the “chaos narrative” have challenged the traditional palliative care approach which was developed for cancer patients and anticipates a clear transition to palliative care for a terminal decline. Instead it has been suggested that discussions about palliative care be entered into at milestones in the patient’s COPD narrative such as on discharge from hospital for an exacerbation or at installation of home oxygen therapy.(Murray, Kendall et al. 2005; Pinnock, Kendall et al. 2011)

Figure 2.6 Typical end of life trajectory in COPD. Reproduced with permission from BJGP (Lehman 2004)



2.12 Conclusion to Chapter 2

In this chapter, I have discussed the major aspects of COPD in order to give an overview appropriate to this PhD. I began by discussing the definition of COPD and outlined how

COPD is diagnosed with the use of spirometry for FEV1 and FVC and the controversies that arise when using fixed spirometry thresholds for diagnosis. I then defined COPD for this thesis using a health services definition. I went on to explain the difference between epidemiology estimates from spirometry surveys and those from physician diagnosis. Then I presented various estimates for the prevalence of COPD globally, followed by estimates for England and Scotland and overall summaries. Then, I reviewed the burden of COPD in the UK in terms of morbidity, mortality and cost of COPD. Next, I reviewed the aetiology of COPD and the focus made upon smoking as the main cause of COPD in high income countries. I then discussed the pathophysiology of COPD and I summarised the clinical features of COPD, followed by the state of the art in terms of predicting COPD prognosis. I then considered the management of stable COPD and COPD exacerbations. Finally, I took a brief look at the experience of COPD from the patient's perspective. It is hoped that this chapter will help to put the remainder of this dissertation in context.

Chapter 3: Reusing healthcare data for research

3.1 Introduction

Chapter 2 introduced the condition of COPD on which this PhD concentrates. As the overall aim of the PhD was to model population projections of COPD, I review such models in Chapter 5. Such models need to be populated with healthcare data. Therefore, I want to introduce the ways in which healthcare data are captured by health systems in England and Scotland. In this chapter I will describe how healthcare data are captured in primary care and in hospital settings and then how they are captured in coded or narrative form. I describe how the validity of the data that are gathered may be assessed and improved. Then I discuss the validity of the data used in this PhD. I go on to describe the ways in which healthcare data may be re-used for research. Healthcare data may be linked with other data for research, I describe some of the opportunities and challenges in carrying out such linked data studies. Healthcare data may be reused in the context of trials to speed recruitment and linked data may be used as a method of recording trial endpoints. In addition, linked healthcare data can be used as a basis for pharmacovigilance. Next I describe the concerns with regards to privacy that such data reuse may pose and the process and regulatory responses in place to mitigate such concerns. I discuss the specific governance required for this PhD and conclude the chapter.

3.2 Healthcare data capture

3.2.1 Healthcare data capture in primary care

Although UK general practices vary somewhat in their degree of paperless processing, there have been incentives since the 2003 General Medical Services (GMS) Contract for them to computerise their healthcare records.(Thiru, De Lusignan et al. 2002; GMC 2003) UK NHS general practice now relies on these computer-based records. Extracts from the primary care computer records are used to inform eligibility for part of the remuneration package for general practices according to the Quality Outcomes Framework (QOF), where practices are rewarded for meeting certain standards of good medical care and organisation.(GMC 2003) There are four main computer systems for general practices in the UK: EMIS (EMIS 2015), Systmone (TPP-Systmone 2015) Torex (which was acquired by the company MICROS in 2012) (MICROS 2015) and Vision (INPS4 2015). GPASS was a system used in Scotland which has now been superseded by other systems.

In order that these data can be used for research, these primary care systems are harvested for information by several large national primary care databases as summarised in Table 3.1. These databases may be accessed by researchers for reuse of the data according to protocol, this will be discussed further in the section on data reuse.

Table 3.1 Primary care databases covering England and Scotland.

Database name	History	Size	Content	Coverage and representativeness
Clinical Practice Research Datalink, CPRD GOLD and CPRD Silver	Formerly General Practice Research Database, GPRD	5.69 million active patients (i.e. alive and registered patients.	Extracts from Vision GP electronic health records. Contains patient registration information and all care events that GPs record including diagnoses, referrals to specialists and secondary care settings, prescriptions, immunisations, diagnostic testing and lifestyle information (smoking, alcohol etc.)	680 General practices across the UK (England Scotland Wales and Northern Ireland), 8.5% of the UK population, Age and gender structure of the CPRD GOLD population is generally similar to the UK.(Campbell, Dedman et al. 2013)
The Health Improvement Network, THIN	A collaboration between two companies: In Practice Systems Ltd (INPS) and CSD Medical Research UK (formerly EPIC). THIN data collection began in 2003	11.1 million patients (3.7 million active patients) representing 75.6 million patient years of data.	Data recorded in Vision software, the GP IT system, is recorded in the form of Read Codes concerning symptoms, diagnoses, referrals etc and also prescription information is available. In addition 158,037 free text comments have been anonymised and added to the database this is 35% of all comments in the records.	562 general practices in the UK. Covers 6.2% of UK population. Broadly representative of the UK population.

QResearch	A not-for-profit initiative between the University of Nottingham and EMIS (supplier of IT systems to primary care)	13 million patients from 754 practices in the UK	Data recorded in the EMIS GP IT system. Patients' data is pseudo-anonymised to prevent identification. Also available are linked data to cause of death, cancer registries and hospital data – however, such data must be analysed at the University of Nottingham.	As well as medical data the system contains socio-economic details of each patient's postcode.
Primary Care Clinical Informatics Unit, PCCIU	Based at the University of Aberdeen	From 318 GP practices in spring 2007 to 133 practices in 2010	Data recorded in GPASS GP IT system including details of patient encounters, Read Codes, prescriptions and measurements. Month and year of birth and gender.	As PCCIU has postcode to sector level this can be linked to Scottish Index of Multiple Deprivation (SIMD) to assess representativeness of the dataset.
Practice Team Information, PTI	Curated by Information Services Division (ISD) Scotland	60 GP practices in Scotland	Practice information collected by GPs, doctors in training and practice nurses. In March 2012, 19 practices were using INPS Vision and 40 were using EMIS as their IT system.	6% of Scotland's population. The PTI population represents Scotland in terms of age, gender and deprivation. Imbalances may be addressed by standardisation in analysis.

Secure Anonymised Information Linkage, SAIL, System	Curated by Health Information Research Unit (HIRU) School of Medicine, Swansea University in partnership with Health Solutions Wales (HSW)	Over 500 million anonymised person-based records from numerous sources, (Ford, Jones et al. 2009)	A wide variety of data sources connected with health and well-being including National Datasets, the NHS including primary care general practices that have signed up to SAIL, social care datasets and others.	Welsh National Screening programmes, the cancer registry, registers of births and deaths, national community child health emergency services, primary care general practices, secondary care hospitals, and social care.
---	--	---	---	--

3.2.2 Healthcare data capture in hospital settings

Hospitals in the UK are approximately two decades behind general practices in adopting Electronic Health Records (EHRs). (Sheikh, Jha et al. 2014) This is despite, in 1998, the Department of Health making the commitment to adopting life-long EHRs for all NHS patients. The NHS National Programme for IT (NPFIT) Care Records Service (CRS) has faced many setbacks and much controversy in trying to implement electronic records in hospitals across the NHS. (Cresswell, Ali et al. 2011) (Robertson, Bates et al. 2011) Hospital staff still want EHRs for hospitals, however opinions differ as to the type of electronic record and scale of data sharing. Centrally negotiated contracts have proven inflexible and have contributed to the delay in the delivery of nationwide EHRs. (Robertson, Cresswell et al. 2010) The main hospital EHR systems commissioned by the NPFIT CRS were Lorenzo Regional Care, RiO and Cerner Millennium. (Sheikh, Cornford et al. 2011) As well as the EHR systems, other computerisation in hospitals takes the form of the Patient Administration System (PAS) which manages appointments, admissions and discharges and the Picture Archiving and Communications Service (PACS) which is used to manage the radiology output of a hospital. Adoption of all of these systems by hospitals has been significantly scaled back since the 2010 spending review by the coalition government in 2010 which identified £0.7 billion savings from the NHS NPFIT (HMTreasury 2010) However, the think tank E-Health Insider (EHI) Intelligence estimates the total IT market for acute and mental

health trusts in England to still be around £1.3 billion per year, of which £560 million is locked up in NPfIT contracts until 2015. (EHIIntelligence 2012).

The bulk of English hospital statistics used in research come from Hospital Episode Statistics (HES) which is derived from the hospital records of admissions and discharges, outpatient appointments and accident and emergency attendances. Data from all of these services are downloaded monthly by the Health and Social Care Information Centre (HSCIC) from the hospitals.(HSCIC 2014) Scottish hospital statistics are collected in the Scottish Morbidity Record (SMR01) by a special health board: The Information Services Division (ISD). The SMR01 has data on the admissions and discharges for all inpatient and day case episodes in Scotland. Both HES and SMR01 have identifier data for the patients, such as name, date of birth, postcode, NHS number (and Community Health Index (CHI) number in Scotland), the databases also contain International Classification of Disease version 10 (ICD-10) codes (see Table 3.2) for the illness, condition or procedure for which the person was admitted and some co-morbidities and specific geographical information, including measures of deprivation and responsible NHS board.(ISD 2014) The ICD-10 information is coded by a team of professional hospital coders who extract data from paper records or directly by clinicians where the hospital has an EHR system.

3.2.3 Coded and narrative data

The data that are collected in EHRs are either in the form of “narrative” or “coded” data. Narrative is the original way of recording information. (Morrison, Fernando et al. 2012; Morrison, Fernando et al. 2013) It is the most faithful because every aspect of the patient’s episode can be recorded. In paper records, the patient’s history and problems were recorded in the form of structured narrative text. The writer of the text had some discretion as to how best to indicate clinical importance of symptoms and signs and clinical context. (Williams and Morgan 1995) Narrative text relies on the semantics of the language to enable the reader to extract useful information.

One review of the advantages of structuring a narrative text such as a patient history in the EHR found that more complete clinical histories are obtained. It also found that structuring improves the accuracy of patients’ self-documented histories and enables better associated decision-making by professionals. No studies were found examining the benefits of coding patient histories.(Fernando, Kalra et al. 2012) Another review examined the evidence of value for direct patient care of structuring and coding of clinical information within the EHR and found eight studies demonstrating improved clinical outcomes where a structured or coded EHR was combined with a prompting or alerting system in a targeted clinical domain.

Domains included the management of a long term condition, a preventive intervention or appropriate choice of therapy. There were three studies demonstrating improvement in safety outcomes in the prescribing domain.(Kalra, Fernando et al. 2012)

Computerisation of the sophisticated task of understanding narrative text is only relatively recently being realised. Computational linguistics is the field of computer science that seeks to understand and represent such narrative language as an interoperable set of semantics. (Elkin, Trusko et al. 2010) Natural language processing is the part of this field pertaining to task-based analysis of narrative text.(Jurafsky and Martin 2009) Natural language processing depends on a “language model”. Language models are drawn from the areas of computer science, mathematics and linguistics and are familiar to those who have trained in these areas. Models include state machines, rule systems, logic models, probabilistic models and vector space models. Models underlie many of the tasks of language processing. For example: where the meaning of a phrase or sentence is ambiguous, language processing models choose the most probable meaning. Search engines such as “Google” and voice recognition software such as “Siri” use language models to decipher the meaning from the input words.(Jurafsky and Martin 2009)

However, computers still have many short-comings when it comes to “reading” health records. In terms of being able to extract useful information from health records, the use of “coding” is therefore widespread. Coding is related to standardisation and structuring.(Morrison, Fernando et al. 2012) Words are chosen from a terminology system or controlled vocabulary. This is a system of labels for different diseases and conditions. There is a hierarchy in the labels in order to give more detailed information when required. Also more than one code can be selected to improve the specific meaning of the coding. Some coding systems go beyond diagnosis alone to include aspects of patient management and administration. Examples of coding systems are given in the table3.2.

Table 3.2 Clinical coding systems (de Lusignan and van Weel 2006; ConnectingForHealth 2013)

Coding System	History	Use
Read Codes	Over 100000 codes which have gone through several rounds of updates.	Version 2 and version 3 are used in UK General Practices. They allow more detail to be recorded than ICD-10 coding. Include codes for more precise diagnoses and also management options.
International Classification of Primary Care (ICPC)	The World Organisation of National Colleges and Academies of Family Practice developed these codes. They have been translated into over 20 languages.	These codes have also undergone updating and evolution. It is organised into body systems and components covering patient-orientated aspects of primary care. They are used in mainland Europe and Australia.
International Classification of Diseases-10 (ICD-10)	Distributed by the World Health Organization at very low cost thereby avoiding financial barriers to their use.	Good for recording diagnoses. Used in the UK for Hospital Episode Statistics and various data outputs from the Office of National Statistics including Mortality Statistics
Systematized Nomenclature for Medicine-Clinical Terms (SNOMED-CT)	This is a joint venture between the UK Department of Health and the American College of Pathologists. Created out of a merger of Read Code V3 and an American Coding Classification SNOMED-RT.	A very extensive terminology, there are over 300000 medical concepts, multiple axes and over 7 million relationships.

The main advantage of coding the health record is that it then becomes amenable to computational analysis. Simple computer programmes can determine how many patients have each condition for audit purposes, patterns and trends within patient populations can be

identified during clinical practice, thus identifying problems e.g. epidemics and accelerating research.(Morrison, Fernando et al. 2013) In addition, computer programmes can look for combinations of codes, especially where the coding has been done from a rich terminology, in order to aid clinicians in diagnosis and other tasks. This is the basis of clinical decision support. Clinical decision support systems (CDSS) are defined as “systems that use two or more items of patient data to generate case specific advice”. CDSS becomes possible when the record is coded.(Wyatt and Spiegelhalter 1991)

Disadvantages of coding include that the coding system may be limited and therefore unable to capture a specific situation or complex unstructured problem. In addition, coding is more time consuming than writing narrative text as it involves an interpretive process in firmly establishing the diagnosis and its subsequent implications. For example, in a qualitative study looking at the coding of depression, GPs were worried about the social stigma, inclusion on the mental health register and ramifications in terms of insurance reports and occupational health screening.(Cresswell, Morrison et al. 2012) There may be a preference for deliberately vague coding in order to reflect diagnostic uncertainty. (Morrison, Fernando et al. 2013)

Problems also arise when accurate coding is a low priority for an organisation. Many clinicians do not see coding as important. In hospitals, clinical coding is a low status administrative job and there is little interaction with clinical staff. In general practice, there may be time constraints which limit the quality of the coded part of the clinical record and clinical users sometimes feel that coding distracts from good direct patient care. (Cresswell, Morrison et al. 2012; Morrison, Fernando et al. 2012) There is evidence that more complete coding can be achieved with targets and financial incentives – however, these can distort what is coded, especially when small numbers of additional coded patients will enable the achievement of targets. Addressing these problems requires education of professionals, preferably with audit-type feedback from their own coded data.(de Lusignan 2005)

In addition there can be problems with having rich coding vocabularies in terms of the reliability of the coding. Rates of agreement between different coders for the precise choice of code may be low when there are lots of options for codes. Reliability is different from validity. Coders may pick equally valid codes but these are just not the same, hence lower reliability. This is important when selected codes are reported in audit and public quality reports from hospitals including performance statistics.(Coory, Youlden et al. 2002; Stausberg, Lehmann et al.)

3.3 Validity of healthcare data for research

The quality of routine data is encapsulated in the concept of validity. Validity refers to the agreement of the routine data with a Gold Standard.(Stausberg, Lehmann et al.) Historically, there have been many problems with the validity of routine data (Williams, Cheung et al. 2003). A Health Technology Assessment (HTA) review of routine data in 2003 concluded that “the validity of routinely collected data is suspect, particularly in systems that are not under clinical and professional control.”(Williams, Cheung et al. 2003) The report also found difficulties with identifying accessing and extracting data and problems as a result of a lack of uniformity in data structuring and coding and definitions. A need for further research into several key areas was identified, including the need to define standards to ensure the uniformity and validity of data collected by different local and national systems. A subsequent HTA report in 2005 found over 1000 routinely collected databases and datasets potentially available for research in the UK.(Raftery, Roderick et al. 2005) They also found that, although internal consistency checks were common within databases, there was little regular external audit.

How may the validity of routine data be improved? In order to be useful to research it may be necessary to “clean” data. This involves making checks on the data’s internal and external validity. In order to check the internal validity – measures of consistency may be developed and tested – for example, the consistent use of the same codes in specific situations. In order to check external validity it may be necessary to triangulate the routine data source with other sources. For example, in one study by Sinha et al they checked the validity of stroke diagnoses in a cohort database by reviewing the case notes of patients (Sinha, Myint et al. 2008). There were 520 incident strokes identified by death certification and hospital record linkage. This increased to 726 when they included self-reported strokes. However, the authors were only able to validate 250 cases of true stroke because only 250 sets of paper case notes were retrieved. This was due to logistical reasons, the case notes that could not be retrieved were not in the health records department at the time they were requested – they were in other hospital departments. Out of the 250 clinically validated cases only 185 had had a CT scan and were therefore considered to be radiologically validated. This study illustrates that validation of routine data is time consuming and logistically challenging.

Many of the problems that occur when trying to adapt routine data to research stem from the fact that the data are collected for one purpose – either administrative or clinical – but used in research as a proxy indicator for another. For example, tiotropium inhaler prescriptions are recorded as part of the clinical record, but may be used as a proxy measure for COPD

diagnosis. Another example would be where a survey of mental health patients may include questions about their alcohol consumption; however, it would be stretching the validity of the routine data to draw conclusions about their physical health from this alone. The validity of routine data therefore varies depending on the original purpose of why it was collected. A study by Soo et al. used routine hospital administrative data, SMR01, and linked it to a cohort database for chronic kidney disease.(Soo, Robertson et al. 2014) The SMR01 data were screened for selected comorbidities and risk factors and the validity of this approach checked by case note review. Kappa agreement between the SMR01 data and the case notes was substantial for ischaemic heart disease (K=0.80) and cerebrovascular disease (K=0.63) and moderate for six other comorbidities, but only fair for hypertension (K=0.28) and poor for smoking. This is because hypertension and smoking, as risk factors, are not often coded in acute admission records. Had the database been additionally linked to a primary care dataset then recording of risk factor variables would be expected to improve significantly, therefore the underlying purpose for the data collection substantially affects the completeness and validity of the data.

Another problem in the context of using routine data is what to do when data are missing. Statistical power is reduced and bias may be introduced if cases with missing data are dropped from the dataset. It may, therefore, be necessary to use data imputation – a family of statistical techniques which build a probable total dataset from the base of the data that are present.(Sterne, White et al. 2009)

3.3.1 Validity of routine data used in this thesis

In terms of this thesis, the validity of COPD data in routine datasets requires discussion. In 2003, Hansell et al. compared mortality data for England 1991-1995 with HES data 1990/91-1994/95 with General Practice Research Database (GPRD – now CPRD) data and Health Survey for England 1995 data and found broadly consistent patterns across all data sources.(Hansell, Hollowell et al. 2003) Incidence rates for hospital admissions and mortality, and diagnosis and inhaler prescriptions in GPRD all increased with age and were generally higher in males than females. Mortality and hospital admissions for COPD shared similar seasonal patterns with higher levels in autumn and winter and peaks at the end of the year. Hansell et al.'s study therefore established the external validity of these data sources and indicated that they are likely to be valid input sources for modelling. A similar exercise could be undertaken again today with cross-validation between prevalence data in one data source (i.e. CPRD) and another (e.g. HES) for a single condition such as COPD. However, I have not found further examples of this. I discussed some of the potential pitfalls of using

COPD mortality data in Chapter 2. Many patients die “with COPD” rather than “of COPD” and so do not have COPD on the death certificate. This means that the prevalence of COPD is underestimated where codes from the death certificate are used to guide estimates of prevalence, because many death certificates do not have COPD listed as one of the causes of death even though the patient had COPD. It is possible that this feature of mortality data has implications for the UK part of the Global Burden of Disease modelling study 2010, that I will discuss in Chapter 8.

There has also been some more recent validation work on the CPRD GOLD database (the GOLD database is the higher quality CPRD primary care database, see Table 3.1) from which the COPD cohort for this study was taken. Presented in 2013, Campbell et al. compared the number of patients actively registered on 27th March 2011 (census day) to the UK census data, overall and by age, sex and region. (Campbell, Dedman et al. 2013) (They compared overall numbers, including all CPRD patients, not just those with COPD.) Age-specific mortality rates were calculated in five year intervals for the CPRD data. Age-standardised mortality rates were then calculated using the European Standard Population, by sex, overall and by UK country. Their results showed that, on census day, 7.3% of the UK population were registered in GOLD with little variation by sex, but a lower representation of younger age patients compared to the census data. Crude death rates were nearly the same as national rates 8.7 per 1,000 versus 8.69 in GOLD. (Campbell, Dedman et al. 2013) Age-standardised mortality rates were lower in GOLD than national rates and age-specific mortality rates among the GOLD population were also below those reported for the UK population. (Campbell, Dedman et al. 2013) This may have implications for the results of the modelling in this thesis. If the CPRD GOLD population is generally older than the UK population then the CPRD prevalence of COPD may be higher with greater costs associated. This may lead to overestimates of the prevalence and costs when extrapolating from the GOLD database to the total English and Scottish populations by modelling. If mortality is lower in GOLD than nationally then there may be an underestimate following modelling.

3.4 Data reuse for research

As growing volumes of routine data are recorded in databases and in datasets there is anticipated to be a huge expansion in the data potentially available for research analysis. (Bobrow 2013) The usefulness of these data is tempered somewhat by their validity as previously discussed, however, several potential avenues for medical researchers have opened up for reuse of data in linked studies, for data mining, trials and for

pharmacovigilance in addition to generation of estimates of disease burden as described in this thesis.

3.4.1 Data reuse in linked data studies

Linking data from different sources is a very fertile ground for both medical research and collaborative interdisciplinary research. Primary care datasets may be linked with secondary care data including hospital admissions, and accident and emergency records. Other datasets may also be linked including the Office of National Statistics (ONS) registers of births and deaths, educational datasets, social care datasets and judicial datasets.

One example of this is the Scottish Care Information – Diabetes Collaboration (SCI-DC), a web-based diabetes record used across Scotland, which has been linked to national morbidity (hospital admission and cancer) and mortality records for use in a wide range of research studies. (Cunningham, McAlpine et al. 2011; Kennon, Leese et al. 2012)

The advantages of using unselected population-based healthcare databases for research include that the results reflect true clinical practice, additional variables can be collected for hypothesis-generation and they are an efficient and cost-effective means to undertake research. (Thuesen, Jensen et al. 2013)

There are challenges in linking databases. Linking the correct individual's records across the database may be done in healthcare records in Scotland by use of the Community Health Index number or CHI. As this is a unique identifier per person, then records may be linked by a deterministic matching process. In datasets which do not have a CHI or equivalent identifier the records may be matched using probabilistic algorithms and a combination of different data fields including name, date of birth and information from the address or other fields used together to construct an identifier. This type of probabilistic matching is more often used for English records which do not always have a unique identifier. This process is somewhat controversial, as some people feel strongly that they do not want their records to be identifiable, especially without their consent. I shall explore the issue of privacy in routine data studies in Section 3.5.

There are also technical challenges in linking databases. For example, EHRs may be stored on different IT systems or written with different standards of software code. Attempts have been made to introduce standardisation among NHS EHRs so that even where record software is produced by different IT providers, it will be able to function alongside each other in the clinical environment. The standard for this is Health Level 7 (HL7), a set of system-independent data interchange specifications and messaging standards supporting

clinical applications. These standards also aim to facilitate linkage in a pragmatic fashion.(de Lusignan, Pearce et al. 2011) This helps to open up the possibility of using the EHRs for research. These standards are evolving rapidly there is HL7 version 2 and version 3 and the new specification FHIR (Fast Healthcare Interoperability Resource) which enables App development to be linked to EHRs amongst other features.(HL7 2015)

Another technical challenge is that many EHRs have limited predefined sets of query-options for patient retrieval e.g. unable to specify time-oriented retrieval of clinical datasets. Until sophisticated ad-hoc querying tools are developed to provide research support to clinical records, the usefulness of clinical records for research will remain constrained.(Prokosch and Ganslandt 2009)

3.4.2 Data mining

Where a linked dataset exists, there is potential for data mining. Data mining is the practice of searching through large amounts of computerised data to find useful patterns or trends.(Merriam and Webster 2014) This type of searching uses sophisticated statistical analyses and may use artificial intelligence approaches such as machine learning. The aim is to find previously unrecognised clinical or other relationships. This type of database analysis is practised widely by corporations in order to target their advertising and sell more of their products, so-called “business intelligence”. However, an analogous approach has potential in clinical informatics, but to date has not been widely employed.(Prokosch and Ganslandt 2009)

Data mining could help identify and explain patterns in the variation of services. Clinicians and managers may then be able to adapt where services are underused or overused. For example, identifying where and why some children are not receiving their full complement of vaccinations or why there is a high rate of caesarean section in an institution and what may be done to address this.(Elkin, Trusko et al. 2010)

3.4.3 Data reuse for trials

The reuse of routine data has several potential employments during the course of clinical trials. Firstly, routine data can significantly improve recruitment rates for prospective trials. Many randomised controlled trials fail to recruit their target number of patients, such that the ability to undertake clinical research in the UK is threatened.(Treweek, Lockhart et al. 2013) A national research register has been suggested as a potential solution to this. The register would hold the registration and contact preference of volunteer patients.(Grant, Ure et al. 2013) The volunteers give permission for these data to be linked to primary and secondary

care EHRs and for their EHRs to be securely searched in the NHS to assess whether they meet the eligibility criteria for studies which have NHS ethical and governance approval. This enables automatic identification of potential subjects and streamlines patient recruitment. Thus also enhancing the speed of recruitment and enabling greater accuracy of consideration of eligibility.(Prokosch and Ganslandt 2009; Dugas, Lange et al. 2010)

Secondly, the reuse of routine data can mean that some endpoints may be recorded from routine data, rather than in a time-consuming and expensive separate trial data recording exercise. The validity of this approach is still being explored. For example, the West of Scotland Coronary Prevention Study compared the cardiovascular event rates recorded by the trial with those from linked routine data. They concluded that the researchers would have drawn very similar conclusions using linked routine data alone compared with trial recorded outcomes.(Barry, Dinnett et al. 2013)

3.4.4 Data reuse for pharmacovigilance

Drug adverse events were traditionally passively monitored via systems such as the Yellow Card Scheme or databases of reported adverse events. However, active surveillance is increasingly being used to identify potential safety problems. Active surveillance systems include drug-based or condition-based registries which actively collect safety information on a defined population.(Wise, Parkinson et al. 2009)

Alternatively, databases from routine clinical records or from community pharmacies can provide a basis for pharmacovigilance. Such databases contain information on the drug exposure and also on outcomes. In some cases databases may be linked to provide these two pieces of information from a variety of sources including primary and secondary care and pharmacies as well as drug registries. The next step is to use the data for follow-up studies of pharmacoepidemiological outcomes. For example, an association was discovered between the thiazolidinediones, rosiglitazone and pioglitazone, and hip fractures using the SCI-DC database linked to hospital admissions data.(Colhoun, Livingstone et al. 2012)

In the USA, the Food and Drug Administration (FDA) have developed a post-marketing surveillance system which includes a component called “Mini-Sentinel”, started in 2009. Mini-Sentinel is a data foundation which uses data from eighteen private health plans supplemented by other sources of data including clinical sources, insurer data, full-text medical records and electronic laboratory result data when needed. In 2014, it covered more than 380 million person-years of data. The network is used by the FDA to answer its new queries every week.(Curtis, Brown et al. 2014)

Also in the USA, there is a programme of research and capacity building named informatics for integrating biology and the bedside (i2b2) which began in 2002.(Kohane, Churchill et al. 2012) It sought to provide the instrumentation for using the information and biological materials accumulated during healthcare delivery. The aim was to create standard shared tools to allow the conduct of research over millions of patients including measurements of their biological material. The i2b2 tools include core and optional software components which are web compatible and build up databases. Shared health research informatics network SHRINE is a clinical querying protocol which can query multiple diverse i2b2 databases. SHRINE can select patient records with defined inclusion/exclusion criteria for laboratories, diagnoses, demographics and medications.(Kohane, Churchill et al. 2012) In terms of pharmacovigilance, i2b2 tools were able to retrospectively identify a large spike in cardiovascular mortality which was associated with the cyclooxygenase 2 inhibitor treatments rofecoxib and celecoxib. This then returned to baseline with the withdrawal of the drugs. The authors suggest that a population monitoring strategy might enhance traditional pharmacovigilance.(Brownstein, Sordo et al. 2007) In addition, a study examining rosiglitazone identified high relative risk for myocardial infarction with this drug in comparison to other drugs including pioglitazone.(Brownstein, Murphy et al. 2010) This latter study was cited by the US FDA in their decision to impose further prescribing controls on rosiglitazone. These studies (Brownstein, Sordo et al. 2007; Brownstein, Murphy et al. 2010) are proof of concept studies in modern pharmacovigilance.

3.5 Professional and public concerns regarding data privacy

There has been much recent media focus on professional and public concerns regarding the privacy of data that is to be reused for research. NHS England has launched a plan called Care.data to upgrade the potential for use of routinely collected primary care data and there are similar plans for Scotland.(NHS_England 2014) However, many GPs remain very concerned as to how the data would be used and the confidentiality arrangements for their patients. There was much controversy in the press as to whether the data might be sold to pharmaceutical companies for financial gain or sold to insurance companies who would then use it to hike premiums.(Borland 2014) There is controversy too over the opt-out model of consent and whether patients had sufficient awareness of how their data were being used and the lack of opportunity to exercise their choices. This has resulted in a delay in the implementation of the Care.data programme to collect additional data from GP practices while a public education campaign is ongoing.

In general, focus groups have found that the public appears supportive of the use of their routine data for medical research. However, they are much more suspicious of research by insurance companies or pharmaceutical companies where the outcomes of research might lead to financial gain for a company.(Haddow, Bruce et al. 2011; Hill, Turner et al. 2013) In the specific arena of a national genetic database, Generation Scotland, findings included that people who had an illness or cared for someone with an illness were more supportive of the involvement of pharmaceutical companies as a “necessary evil” than those with little or no personal experience of illness(Haddow, Laurie et al. 2007). The authors recommended that developing the genetic database should include obligations for private companies to share the profits from any output in order to help to restore public confidence in for-profit research and thereby encourage participation.(Haddow, Laurie et al. 2007)

3.5.1 Governance for eHealth data linkages and reuse

The governance for eHealth data linkages and reuse has been criticised for being unwieldy and overly complex. In 2008 the Department of Health identified 43 pieces of relevant legislation, 12 sets of relevant standards and eight professional codes of conduct, altogether fuelling “a culture of caution, confusion, uncertainty and inconsistency.”

(Science_and_Technology_Committee 2009) Often data custodians and researchers are afraid of misunderstanding the law and risking sanctions and so are over-cautious in their interpretations of what can be done with data. There has also been a tendency to view consent as a panacea that alone addresses the concerns around using data for research.(Sethi and Laurie 2013) This has resulted in governance solutions being shaped around the “consent moment” and, where this is not possible, anonymisation of data is the approach preferred. However, although anonymisation makes identification of research subjects from their data less likely it does not make it impossible especially as other data may be released in the future that make identification more possible for a malignant analyst who is determined enough. (Information_Commissioner's_Office 2012; Lewis 2014) In addition to the above solutions, the European Data Protection Directive is under review and may result in an even more complex regulatory framework.

A model of proportionate governance has been proposed by Sethi and Laurie following their governance work on the Scottish Health Informatics Programme (SHIP)(Sethi and Laurie 2013). They propose balancing the two principles of (1) promotion of the public interest and (2) protection of the privacy and other interests of citizens as the central goal of governance. Consent and anonymisation are the starting points to consider within a proportionate governance approach. Decision makers can then depart from these two mechanisms and use

other routes where it can be shown that consent and anonymisation would be disproportionate. For example where there is only very low risk of harm following identification. This is a risk-based approach and risk assessment is a key role of authorising bodies such as the Caldicott Guardians whose functions are outlined in the Data Protection Act 1998 (OPSI 1998) and Caldicott report and review. (Dept_of_Health 1997; Dept_of_Health 2013) Other authorising bodies include the Confidentiality Advisory Group (CAG) in England which considers data access applications submitted to the HRA, and the Privacy Advisory Committee (PAC) which performs a similar function in Scotland. Such bodies provide oversight to the linking and sharing of data, assess the risks to privacy and determine the role of consent and anonymisation in healthcare data research. These bodies also oversee how the rules regarding how long datasets may be kept, how they may be kept, who has access to them and when they are destroyed, are interpreted. In providing the mechanism for balancing privacy and public interest CAG and PAC are ideally placed to ensure that research is not stifled by burdensome governance yet proceeds with due respect for privacy.(Sethi and Laurie 2013)

3.5.2 Governance relevant to this PhD

I had to satisfy the CPRD data curators that I would take reasonable care of the COPD data and treat it responsibly, and that the risk of identification was minimised. I discuss this further in Chapter 6 when I discuss the process for obtaining these data, see also Appendices 1.1 to 1.6. In addition, I required ethical approval for the PhD (see Appendix 2.1). For the Lothian COPD Cohort data there was a need to satisfy the Privacy Advisory Committee (PAC) of NHS National Services Scotland (See Appendices 3.2 and 3.3).

3.6 Conclusion to Chapter 3

This chapter has outlined the main ways in which data are captured in UK healthcare in both primary care and secondary care. It introduced several primary care databases which are used as a basis for research. There then followed an introduction to the differences between narrative and coded data and emphasis that the majority of data captured in the health service which may be reused in research, is of the coded form. There was then a discussion of the ways in which both the internal and external validity of routine data could be tested and improved, including by triangulation with other data, followed by a discussion of sources of triangulation for the CPRD data used for modelling later in this thesis and the implications of this for the validity of these data. I then went on to outline the ways in which routine data can be reused for research, namely: in linked data studies, for data mining, in trials and for pharmacovigilance. Finally, I discussed the issues regarding privacy in routine data research,

including the controversy around the NHS Care.data programme, and the regulatory responses to these concerns and how these have been applied to the data in my thesis. I will expand on the precise approvals which were required for each dataset in Chapter 6.

In summary, I hope that this chapter has made clear the origins of healthcare routine data and outlined how such data may be employed in research and the issues that need to be considered when reusing routine data in healthcare research.

Chapter 4: Concepts in modelling

4.1 Introduction

The previous chapter served to introduce how routine healthcare data are being reused for research. In the next chapter, Chapter 5, I will systematically review models for projecting the population with COPD, therefore, the purpose of this chapter is to introduce some of the concepts used in chronic disease modelling. The field of quantitative chronic disease modelling is growing rapidly as a new scientific discipline. Many chronic disease models are based on health economic models. The most common underlying structure of these health economic models is the Markov model. I will introduce the Markov model below, from the health economics perspective, followed by the two main ways it can be used in model simulations, namely cohort simulation and individual simulation. I then discuss alternative model formats namely demographic models, risk factor models, time trend and other models.

Next I will discuss one of the main concepts of modelling termed quality adjusted life years (QALYs). Finally, I discuss an alternative modelling concept: disability adjusted life years (DALYs). (See also glossary).

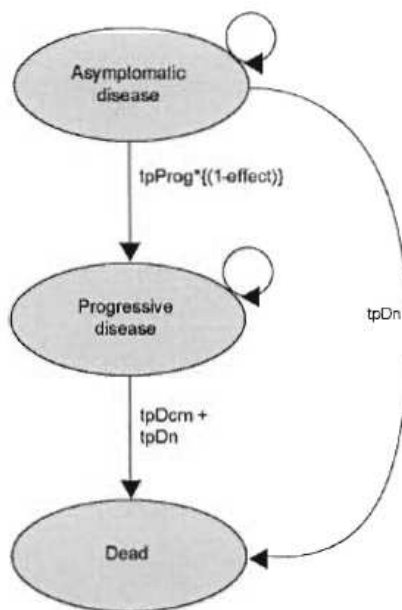
4.2 Introduction to Markov modelling

Markov models are suited to modelling the progression of chronic disease. They are used to model stochastic processes, or random processes that evolve with time. (Briggs and Sculpher 1998) The chronic disease is divided into a set of mutually exclusive disease states. Transition probabilities for movement between these states are assigned per time period that the Markov model is run, i.e. each “Markov cycle”. In health economics, resources and health outcomes are attached to each state. Running the model over a large number of cycles then allows estimation of the long-term costs and outcomes associated with that particular disease.

An example of a Markov model for a chronic disease is shown in Figure 4.1. This disease has three states: asymptomatic, progressive and dead. A patient can only be in one disease state at any one time. The arrows between the ovals represent the transitions between the states. A patient in the asymptomatic state can move to the dead state (with transition probability equal to all-cause mortality excluding the disease in question) or to the progressive disease state. The progressive disease state represents more unpleasant aspects of the disease and the patient experiences symptoms and has an increased risk of death over-

and-above the all-cause mortality risk. The state of death is impossible to leave from and so is known as an “absorbing state”. There are also circular arrows returning to the state that they left, which represent patients remaining in the same state as they were in for the previous cycle. It is not illustrated here, but it is possible to represent clinical improvement by having a patient with progressive disease move back to an asymptomatic state.

Figure 4.1 Example of a Markov Model. Reproduced with permission (Briggs and Sculpher 1998)



Mathematically, in a model of k states, a $k * k$ transition matrix would give all the transition probabilities of each cycle. However, many of these probabilities are set to 0. In this model there are three states, thus nine probabilities to be estimated, but patients here cannot recover, so there are no transitions from dead to progressive, progressive to asymptomatic and dead to asymptomatic. The probability of moving to states in each cycle must sum to 1, as patients must be only in 1 state at a time, thus the probability of staying in the same state in a given cycle is one minus the probability of leaving that state. This model, therefore, requires the estimation of only three probabilities: $tpProg$ the probability of moving from asymptomatic to progressive; $tpDn$ the probability of dying when in the asymptomatic state from a condition other than the index disease (natural death); the probability of dying when in the progressive disease state from the disease itself, $tpDcm$, or from an unrelated condition, $tpDn$.

In Markov models the probability of moving out of a state is not dependent on the patient's previous states. This is the "memoryless" feature of Markov models or the "Markovian assumption" and is sometimes a limiting factor when constructing models.

In "Markov chain models" the transition probabilities are assumed to be constant over time. However, this may not be very realistic when trying to represent chronic diseases as it is likely that risk of death will increase with time in a disease state. This last may be represented with so-called "time-dependent Markov processes" which assume that transition probabilities are a function of the number of cycles the model has run and use a run of temporary states which must be visited in a fixed sequence as "tunnel states" to reflect the time dependency of the transition probabilities.(Briggs and Sculpher 1998)

In terms of health economics, Markov models are often used to compare a model of disease progression without drug therapy to one including drug therapy, which is assumed to reduce the transition probabilities to disease and death. The model is run in both situations and the results of the treatment arm are compared with the results of the control arm. In economic analysis the key parameter is the incremental cost-effectiveness ratio (ICER).(McCabe, Claxton et al. 2008) This is calculated by the following formula where C_t and E_t are the costs and effects associated with the treatment arm and C_c and E_c are associated with the control arm. The ICER is the difference in the costs divided by the difference in effects.

$$ICER = \frac{C_t - C_c}{E_t - E_c}$$

4.2.1 Cohort simulation

In Markov model cohort simulation, a cohort of for example, 1000 patients will illustrate the experience of patients in the model.(Briggs and Sculpher 1998) The whole cohort begins the model at time 0 in the asymptomatic state, then at each cycle the appropriate transition probabilities are applied and the numbers of patients in each Markov state is altered. Each patient must be in one and only one disease state at any one time. This cohort's progression is represented in Table 4.1.

Table 4.1 Cohort simulation of Markov Model, Figure 4.1, with no drug therapy

Cycle	Asymptomatic state	Progressive state	Dead	Total
0	1000	0	0	1000
1	976	10	14	1000
2	943	28	29	1000
3	902	52	46	1000
4	854	79	67	1000
5	799	109	92	1000

For health economic purposes, the progression of the cohort with no drug therapy is represented and compared with the progression of the cohort with drug therapy

4.2.2 Individual simulation

The alternative complementary method for Markov modelling is individual (or Monte Carlo) simulation.(Briggs and Sculpher 1998) This is where a large number of patients are followed through the model individually. In this case, the individual patients are subject to the same transition probabilities, but can only be in one state at a given time and may or may not move along states during any given cycle. Therefore the paths followed by each individual patient will have random differences. A profile of costs for each patient’s path through the model will be built up. Averaging these costs and effects over a large number of patients will give an estimate of the average costs and effects for both drug therapy and no drug therapy arms of a model.(Briggs and Sculpher 1998)

4.2.3 Comparing cohort and individual simulation

In contrast to the exact result obtained from the cohort method, the individual simulation method will not give the same results on any two occasions because the simulations vary randomly. However, as long as a large number of individual simulations are performed and averaged the differences between the two methods are likely to be small and insignificant. The individual simulation method gives an estimate of the likely variance associated with the parameters estimated by the model. This uncertainty is due to the intrinsic probabilistic uncertainty of the model and is often termed “first order” Monte Carlo simulation. (Second order Monte Carlo simulation also allows the parameters of the model to vary over a range with a specific distribution and are sometimes useful).(Briggs and Sculpher 1998)

4.3 Alternative model formats

4.3.1 Demographic models

Demographic models are one of the most straightforward models that can be constructed to represent chronic disease. The national population is taken and each age and sex stratum is multiplied by the age and sex specific prevalence data. It is the reverse of a prevalence calculation. It is useful when prevalence is only known for a sample of the national population and where the sample needs to be multiplied up to represent the full population. This is done using age and sex strata so that the resulting disease prevalence represents the age and sex distribution within the national population.

4.3.2 Risk factor models

One example of a model of this type features in the tables in the systematic review in Chapter 5.(Peabody, Schau et al. 2005) An algorithm is formed by the accumulation of risk factors for COPD such as smoking, outdoor air pollution, indoor air pollution, age and sex. Evidence is gathered to quantify these risks as contributing to COPD and the algorithm is multiplied out according to the age and sex profile of the population. In the end an estimate of the prevalence of COPD is obtained.

4.3.3 Time trend models

These models use historical data to graph the trend of COPD prevalence for a country then use computer software and mathematical techniques to project the trend into the future. There is an example of this type of model included in the systematic review in Chapter 5.(Erbas, Ullah et al. 2012)

4.3.4 Other models

The Markov model is a flexible type of model. Many models in Chapter 5 are similar in their structure are not necessarily identified as “Markov”. Some state that they are “Markov-type” or “Markovian”. In general, the studies identified in Chapter 5 state that they are running cohorts, or micro-simulation, or a combination of the two approaches; which are complementary rather than mutually exclusive.

4.4 Outputs of health economics modelling

The results of Markov and other models may be calculated and expressed in terms of either QALYs or DALYs.

4.4.1 Quality adjusted life years (QALYs)

Quality adjusted life years (QALYs) and disability adjusted life years (DALYs) are similar terms for concepts which health economists and epidemiologists use in slightly differing situations to give an idea of the impact of an illness on people's and a population's health. They are both used as outputs from modelling. Both begin with the time period of a year and reduce it to allow comparisons to be made between different functional states due to illness, as explained below.

Seminal work by Klarman et al. quantified the quality of experience of kidney patients who received a transplant as compared with those who continued on dialysis in the context of cost effectiveness analysis, and helped develop the QALY (Klarman, Francis et al. 1968). However, the term "QALY", for a life year which had been modified to reflect illness, was not coined until 1976. (Zeckhauser and Shepard 1976)

QALYs rely on a generic quantification of health status. The health status is a functional level or "utility". This health status is quantified by assuming that the person without illness is unimpaired or has 1.0 utility and that a person in state 0 is dead. The original presumption by Klarman was a time-trade off where the participant traded length of life for quality of life. Other techniques currently in use allow participant use of visual analogue scales or of the standard gamble. The standard gamble is an assessment involving probabilities e.g. what proportion of people choose 100% chance of half "full" health over a 50% chance of "full" health and vice versa.

Currently, a variety of questionnaires are used to quantify health status. The EQ-5D (Brooks, Rabin et al. 2003) or Short Form 36 (SF36) (Ware and Sherbourne 1992) are examples of these. Completion of these forms varies depending on who fills them in. Members of the public, caregivers and health professionals rate illness states differently to people actually suffering these illnesses. The consensus is that the conditions be calibrated by lay people who are "healthy". (Matza, Boye et al. 2014)

QALYs are calculated from the utility or reduced utility as a result of a health condition, multiplied by the time spent at this utility, the life years. Then, by incorporating costs, health economic analysis may be performed. Cost-effectiveness is the concept that a finite amount of money may be spent on a treatment that improves utility for a number of years (measured in QALYs in one way or another), and the "cost per QALY" is calculated to enable comparison of treatments, or other interventions. (Gold, Stevenson et al. 2002) These cost-effectiveness calculations are often used in developed nations to decide which drugs should

be paid for on behalf of the population in the face of budgetary constraints.(McCabe, Claxton et al. 2008)

4.4.2 Disability adjusted life years (DALYs)

QALYs are aspirational in that they measure health expectancy and the aim is to maximise the QALY score to 1.0, or as close to 1.0 or full health as possible. DALYs are the opposite: a measure of health gap. On the DALY scale, 1.0 is total disability and 0 is the target of zero disability.(Gold, Stevenson et al. 2002)

DALYs were described in the 1990s in a World Health Organization (WHO) publication that proposed how the global burden of disease and injury could be quantified and reduced in developing nations.(Murray and Lopez 1994) DALYs measure the gap between a population's health and a hypothetical ideal for health achievement. DALYs place different value weights on populations that are very young or very old as these are supposedly less economically productive than mid-age populations.(Murray 1994) DALYs rely on secondary data and expert opinion to identify and describe diseases and different conditions are placed according to their level of disability. The WHO is the most prolific source of DALY estimates, the Global Burden of Disease Study update in 2012 quantified disease burden across the world's countries in terms of DALYs.(GlobalBurdenofDisease 2012)

4.4.3 Advantages and disadvantages of QALYs and DALYs

Advantages of QALYs and DALYs include that they combine an estimate of the extra length of life gained from an intervention and of the quality of the extra life thus enabling quantification of health gains associated with interventions. This means that one intervention may be compared with another for effectiveness in relation to the same problem. Comparison of health burdens from different diseases or across different countries is also possible – this enables priority setting by policymakers.(Olsen 2009)

Disadvantages of QALYs and DALYs are that patients who are actively receiving the intervention may have different views on the values assigned to quality of life from the sample of the general population who were questioned in order to assign the values. This generates controversy as to whose values should be used. There are also some who argue that QALYs and DALYs oversimplify complex resource allocation decisions.(Malek 2001)

4.5 Conclusion to Chapter 4

This short chapter has introduced some of the concepts in modelling that come from the field of health economics. To begin with the Markov model was explored in some detail. This was

followed by a brief discussion of other types of model. Then, there followed an introduction to the key concepts of QALYs and DALYs. These concepts should be helpful when making sense of the results from the systematic review to look for models in Chapter 5.

Chapter 5: Systematic review of the literature for models that calculate the prevalence and burden of chronic obstructive pulmonary disease (COPD)

5.1 Introduction

The aim of this thesis was to project the prevalence and burden of COPD in England and Scotland. This will require a model to use with routine data to make projections. Thus, this chapter describes a systematic search for models for calculating the prevalence and burden of COPD.

I begin this chapter by summarising the methods that were used in order to conduct a systematic review of the literature. Then I outline the quality of reporting framework that I developed in order to assess the quality of reports of the included models. I decided to assess the quality of reporting as a first step in assessing model quality on the basis that if a model is poorly reported there may not be enough information available on which to critique the model itself. Summary tables of the characteristics, and specific data inputs and outputs of the models are presented. This is followed by a table detailing the scores of each model on the quality of reporting appraisal framework. Then a table containing the specific findings by each model in its own context is given. I then present a narrative summary of the seven models which achieved the highest relative scores on the quality of reporting framework. I describe the features of each model and its functioning. These descriptions include an assessment of the quality of each model's development in terms of sensitivity analyses and validation. Finally, I make a selection of one of the models for further study within the PhD and justify its selection.

5.2 Methods

5.2.1 Protocol

In keeping with best practice recommendations in conducting systematic reviews, I developed and published a detailed protocol for this work.(Appendix 6)(McLean, Wild et al. 2013) I also registered the protocol in the PROSPERO database with number CRD42012002623. (PROSPERO 2013)

5.2.2 Eligibility criteria

Any modelling study which used demographic and epidemiological data to estimate the prevalence and disease burden was eligible for inclusion. The outcomes of interest were

incidence, prevalence, disease burden and mortality. For the purposes of this review, “disease burden” was considered from the perspective of the health system. The “disease burden” can also potentially be considered from the perspective of the individual or from the perspective of society (e.g. in terms of lost productivity). Such perspectives were considered to be outwith the scope of this PhD.

The types of “models” of interest included demographic models, risk factor models and models which have evolved from health economic models including Markov-type models (see Chapter 4 for an overview). Models were excluded if they described animal cell lines, clinical series or estimates of individual risk (such as individual prognostic models). Decision analytical models or decision support models were excluded where they referred to clinical decision-making for individuals. Models comparing one intervention with another intervention were also excluded as the aim was to estimate the prevalence, disease burden and mortality rather than to investigate the effectiveness of interventions. Regression models which were quantifying the effect of risk factors on COPD rather than projecting prevalence and/or disease burden were also excluded.

5.2.3 Information sources and study selection

A search strategy was developed using search terms to include the three concepts of “modelling”, “disease burden” and “chronic obstructive pulmonary disease”. The full search strategy is published in the protocol which is included in Appendix 6.(McLean, Wild et al. 2013) Searches were conducted in Medline, Embase, CAB Abstracts, World Health Organization (WHO) Library and Information Services (WHOLIS – library catalogue of books and reports), WHO Regional Indexes (AIM (AFRO, LILACS, (AMR/PAHO), IMEMR (EMRO), IMSEAR(SEARO), (WPRIM (WPRO). A modified search strategy was used to identify reports from the WHO home website and Google. Searches were for both published and unpublished modelling studies from January 1980 (when modelling methods first began to be widely used) to November 2013. All studies were independently reviewed against the stated inclusion criteria by me and by an associate researcher, Dr Victoria Barbour (VB), and all disagreements were resolved by discussion between reviewers, with arbitration by a third reviewer if a decision could not be reached.

5.2.4 Data extraction

A piloted data extraction form for each study was used by me and checked by VB. The following data were extracted from each study: author and email address, year, institution and funding source, the purpose of the model, model title, model type, model setting, time

period and population. Model inputs and source of input data details of processing of the model, were also extracted, along with COPD outcomes (i.e. incidence, prevalence, mortality, primary care visits, emergency department visits, hospitalisations, treatment costs), model output/ results, details of the model's availability, any comparisons with other studies, social and economic policy implications of model's output and future research recommendations.

5.2.5 Quality appraisal framework

A standard procedure in systematic reviewing is to assess the quality of studies included in the review. However, there are not yet any quality appraisal frameworks for modelling studies. I therefore decided to first assess the quality of reporting of the modelling studies. I did this by designing a quality of reporting framework in discussion with Professor Simon Capewell from Liverpool University. The quality of reporting framework was designed following review of key guidelines as to good practice in modelling.(Weinstein 1989; Weinstein, Toy et al. 2001; Weinstein, O'Brien et al. 2003; Philips, Bojke et al. 2006) A scoring mechanism was devised to weight the different elements required in reporting the production of a relevant and high quality model.(Unal, Capewell et al. 2006) This scoring framework was described in the published protocol for this systematic review and is included in Appendix 6.(McLean, Wild et al. 2013) The quality appraisal framework's elements are described as follows:

5.2.5.1 Purpose and aim

- There should be a clear statement of the question which the model is trying to answer and its scope, its limits or boundaries.
- The structure of the model should be relevant to the decision making perspective. That is whether decisions are considered from the patient's perspective, the healthcare system's perspective or the perspective of broader society, and the appropriate implications should be included.
- The time horizon of the model should be long enough to engender the differences between strategies – even if there is a lack of long term follow up input data – as long as this absence is made explicit in the model assumptions.

5.2.5.2 Transparency

- This is a key detail – if the model is publicly available it will be possible to use it for modelling COPD. Public availability is also important for peer review of the quality of the model before it was published. However, sometimes there are intellectual

property concerns and models are not made available as they may be profitable for their authors.

5.2.5.3 Data input

- It is important to consider the quality of data used to develop the model and to consider if the source, variables and range are justified.
- Even where data are weak, models can be helpful in aiding decision-making as they represent a logical framework of assumptions.
- Inclusion or exclusion of data sources should always be justified.
- The conclusions of the model should be framed as conditional upon the range of inputs, and special attention should be paid to attaining accurate inputs for those data to which the results of the model are particularly sensitive.
- Where data are absent, expert opinion may be used to estimate parameters, for example, following a Delphi exercise.

5.2.5.4 Data modelling

- There should be a transparent method for how the raw data are transformed into data inputs.
- There should be discussion of the model's derivation and methods of update to reflect fresh evidence.
- The assumptions on which a model is based should be documented and justified.
- Any model should be consistent with accepted techniques of statistics and epidemiology.

5.2.5.5 Data incorporation

- Data may be incorporated as point estimates or as distributions. It should be clear which method has been used.
- It should be clear whether a deterministic or probabilistic modelling approach has been taken and the resulting impact on the uncertainty of the outputs from the modelling. Where the data are incorporated as distribution for probabilistic analysis the choice of distribution and how the distribution has been defined should be described. The advantage of a probabilistic analysis is that it allows statistical statements to be made regarding the importance of input uncertainties.

5.2.5.6 Sensitivity analysis

- Structural uncertainties should be evaluated using sensitivity analysis: multi-way sensitivity analysis is preferred over one-way sensitivity analysis as it is more rigorous.
- Parameter uncertainty of point estimate inputs should be investigated by sensitivity analysis with explicitly stated and justified inputs. Probabilistic sensitivity analysis can assess the joint effect of uncertainty over all parameters.

5.2.5.7 Internal validation and calibration

- Verification of internal consistency (debugging) should be undertaken by testing with extreme values of inputs and checking that outputs are logical. Sensitivity analysis with extreme or null values is a part of this.
- External consistency (calibration) should be tested by using retrospective data from an independent source (not used in model derivation, e.g. from national health statistics) to see that the model works to give logically consistent outputs. Where discrepancies cannot be explained the model may need to be recalibrated.

5.2.5.8 Predictive validity

- Predictive validity is testing the accuracy of the model's prediction and can only be done with time and long term follow up data. Often these do not exist and so predictive validity cannot be confirmed. This does not mean that the model is not useful it must simply be recorded as a caveat.
- Models should be flexible enough to evolve with the science of the subject they capture. As new research is undertaken they should incorporate new information. However, it should be recognised that the circumstances in which the model was derived may change significantly over time and the conditions in which the model was designed may no longer apply. In such circumstances predictive validity is impossible.
- Models which cannot evolve or be recalibrated and remain inconsistent with new empirical evidence should be abandoned or replaced.

5.2.5.9 Potential limitations

- The authors should show an awareness of the potential limitations of their model and discussion of the following in this context is desirable: assumptions, confounding, lag times and competing causes.

5.2.5.10 Involvement of policymakers, planners and decision makers

- Were policymakers, planners and decision makers involved at all?
- How were they involved: commissioning, influencing approach or methodology?
- Will policymakers, planners and decision makers have an opportunity to act on the findings of the model?

5.2.6 Scoring on the quality of reporting framework

Following discussion with Professor Simon Capewell, it was decided to allocate scores to give a relative importance weighting to the domains. If one domain of the model is flawed or absent it is possible that it may not invalidate the whole model if the other domains are fully present and accurately reported.

The most important consideration was the purpose of the model and whether it aligned with the aim of projecting burden of COPD. This is reflected by the four points available for this domain. Next follows model transparency, and so on. The domains of this framework and associated scores were as follows:

1. Model purpose and aim (4)
2. Transparency (3)
3. Data input (1)
4. Data modelling (3)
5. Sensitivity analysis (2)
6. Internal validation (1)
7. Calibration (1)
8. Predictive validity (3)
9. Potential limitations (1)
10. Involvement of policy makers, planners and decision makers in model. (1)

Thus these 10 domains contributed to a total score out of 20 points.

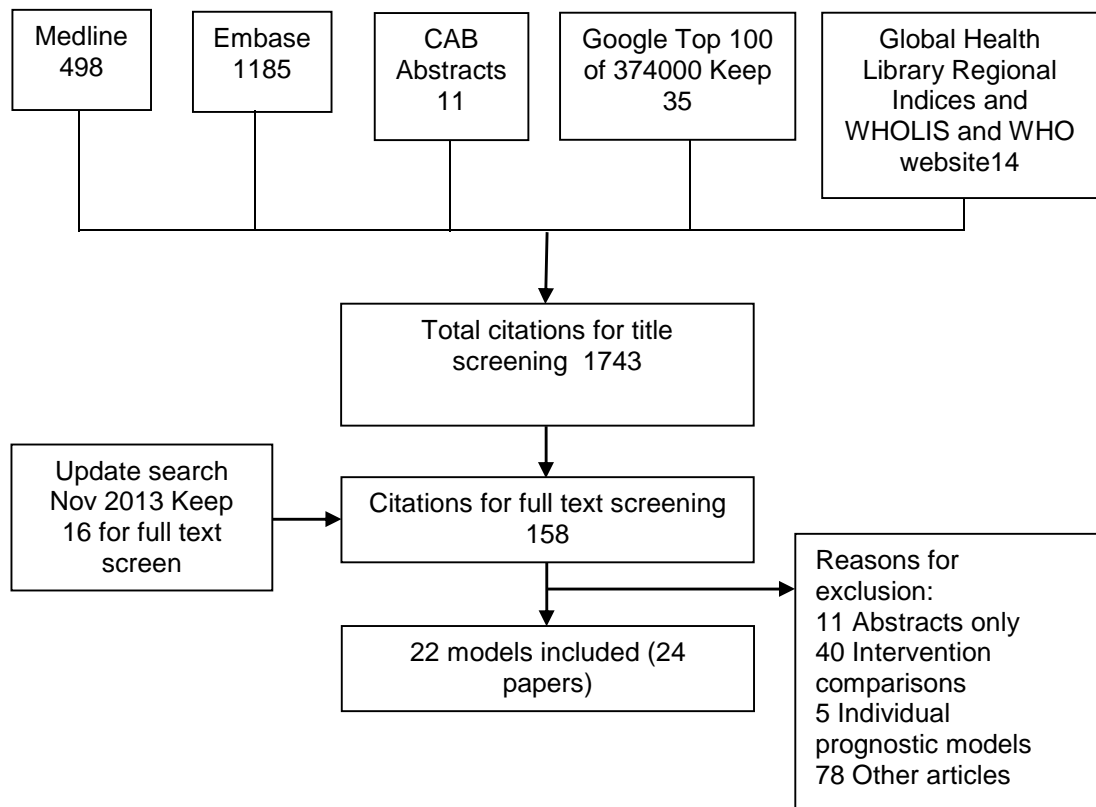
5.2.7 Synthesis of results

As the models had different purposes and were context specific and based in different settings with different input data, I present results tables to summarise the features, inputs and outputs, and reported outcomes of all of the models. This is followed by a narrative synthesis of the seven highest scoring models on the quality of reporting scoring framework as heterogeneity precluded any meaningful combination of results.

5.3 Results

1743 studies were title-screened. 158 titles and abstracts were selected for full text review. In the event that a paper used a model published previously to make a new projection, update a projection or make a projection in different setting or geography, I included only the original explanation of the model unless the model itself was substantially altered or updated in any of the newer papers. In the end, 22 models were selected for inclusion, as described in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram in Figure 5.1.

Figure 5.1 PRISMA flow diagram



5.3.1 Introduction to tables of results

This section begins with summary tables of the characteristics (Table 5.1), and specific data inputs and outputs of all 22 models (Table 5.2) (Bronnum-Hansen and Juel 2000; Stang, Lydick et al. 2000; Feenstra, van Genugten et al. 2001; Shibuya, Mathers et al. 2001; van Genugten, Hoogenveen et al. 2003; Hoogendoorn, Rutten-van Molken et al. 2005; Peabody, Schau et al. 2005; Chapmann, Mannino et al. 2006; Lopez, Shibuya et al. 2006; Hurley and Matthews 2007; Nacul, Soljak et al. 2007; IHME 2010; Atsou, Chouaid et al. 2011; Holt, Zhang et al. 2011; Hoogendoorn, Rutten-van Molken et al. 2011; Pichon-Riviere, Augustovski et al. 2011; Erbas, Ullah et al. 2012; Lhachimi, Nusselder et al. 2012; Najafzadeh, Marra et al. 2012; Perera, Armstrong et al. 2012; Pham, Ozasa et al. 2012; Stanciole, Ortegon et al. 2012). These are followed by Table 5.3 detailing the scores of each model on the quality of reporting appraisal framework. One model is missing from this table as it was not possible to quality appraise it because an appropriate report on which to base the quality appraisal was not found.(IHME 2010)Then there follows Table 5.4, which contains the specific findings of each model in its own context.

Table 5.1 Characteristics of modelling studies. Studies are in order of quality of reporting (Table 5.3)

Author	Model Title	Model Type	Country	Purpose of model	Structure
(Pichon-Riviere, Augustovski et al. 2011)		Health economic Monte Carlo microsimulation model	Seven Countries in Latin America	To describe the development and validation of a health economic model to address the tobacco disease burden and the cost-effectiveness of smoking cessation interventions in seven Latin American countries.	This model considers all tobacco-related diseases (i.e. heart, cerebrovascular, COPD, pneumonia/influenza, lung cancer and nine other neoplasms). COPD was a special case in terms of data because as national statistics are known to significantly underestimate COPD-related mortality, its incidence and prognosis were estimated based on the Stang model and a Mannino epidemiological study. A first order Monte Carlo probabilistic microsimulation of individual subjects was built, incorporating the prognosis, costs and quality of life of the above diseases. The model underwent extensive internal and external validation.
(Feenstra, van Genugten et al. 2001)	The chronic disease model.		Netherlands	To single out the impact of changes in demography and smoking behaviour on COPD morbidity, mortality and healthcare costs.	Shows the impact of changing patterns in aging and smoking on the incidence and prevalence of COPD. COPD is modelled in a dynamic multistate life table model with other smoking related diseases: lung cancer, stroke, CHD and asthma. Multiple birth cohorts are followed year on year through the model. Two scenarios are modelled for 2015, the first assumes that the incidence of COPD remains constant, the second models changes in smoking behaviour.

(Hoogendoorn, Rutten-van Molken et al. 2005)	Dynamic population model	Builds on Feenstra's single state model with a dynamic multi-state life table model	Netherlands	To project the future burden of COPD in the Netherlands by disease severity and to evaluate the impact of different smoking cessation interventions on the national burden of COPD	Part of the Dutch Chronic Disease Model. This model projects the incidence, prevalence, mortality, progression and healthcare costs of COPD per GOLD severity stage as well as change in the healthy population. A new birth cohort is followed each year and within each cohort patients smoke or stop smoking, transition along COPD severity and die all according to transition probabilities based on research conducted in the Dutch population including the Dutch Foundation for Smoking and Health and three Dutch cohort studies.
(Peabody, Schau et al. 2005)	Prevalence estimation model	Multiplicative	Spain, Norway, Nepal and Poland.	To model COPD prevalence from demographic data and smoking prevalence and risk factors.	Data were taken from a literature review to quantify the relationships between COPD and its risk factors and prevalence. There were five input risk factors: Age, gender, (by inclusion of gender specific smoking rates) percentage of smokers, COPD severity and percentage of population exposed to non-smoking risk factors (biomass fuel, occupational airborne dust and air pollution).
(Shibuya, Mathers et al. 2001; Barendregt, Van Oortmarssen et al. 2003)	DISMOD 2	Burden estimation model	Worldwide	To estimate internally consistent values for incidence, prevalence and mortality due to COPD in developing world countries.	Disease occurrence is inferred from mortality figures with the help of the mathematical consistency required by the following epidemiological relationships: prevalence, incidence, remission and mortality rates.

(Atsou, Chouaid et al. 2011)		Multistate Markov	England	To estimate the specific burden of continuous smoking as well as the effectiveness and cost effectiveness of smoking cessation.	Follows a cohort aged 40-89. Two parallel cohorts are modelled. All have COPD but one cohort stop smoking at cohort initialisation and remain sustained quitters. Patients are simulated individually. The severity stages of COPD are modelled and the patients transition along severity to death according to probabilities from the Framingham cohort.
(Kulik, Nusselder et al. 2012; Lhachimi, Nusselder et al. 2012)	Dynamo-HIA	Markov based multi-state	EU-wide	To be a standard tool which is publicly available allowing researchers and policy makers to quantify the impact of policies using epidemiological evidence.	Multiple smoking related diseases modelled. A closed real-life population is modelled according to the explicit risk factor states of each simulated individual. Incidence and prevalence of a disease are required as inputs at the population level and the module back-calculates the risk factor specific values using the relative risk from each risk factors state. Multiple risk factors states are modelled as a partial micro-simulation, this is combined with a deterministic macro approach for the disease life table.
(Najafzadeh, Marra et al. 2012)		Dynamic population projection model	Canada	To project the total burden of COPD (epidemiology, cost, morbidity and mortality) over the next 25 years using the population of Canada as a case study and test three hypothetical intervention strategies.	A dynamic population simulation model was built based on population growth projections from Statistics Canada and parameters from the literature. Demographic data for Canadians of 40 years or older with information on rates of birth, immigration, emigration and mortality plus information on smoking status, presence of respiratory symptoms and prevalence of mild, moderate and severe COPD. The prevalence data came from the BOLD spirometry survey of the population and so included both diagnosed and undiagnosed COPD patients. Progression rates across mild to moderate and severe were calculated based on changes in lung function related to age and smoking status. Minor and major exacerbation rates were modelled. The impact of smoking status, age and disease severity on COPD-related mortality was captured indirectly by their effects on exacerbations.

(Perera, Armstrong et al. 2012)		Generalised linear model	USA hospital care, data from 2006	To determine the direct inpatient burden of COPD in 2006, to characterize hospitalizations and determine the top ten diagnoses for COPD exacerbation cases and determine the association of comorbidities with exacerbations and their impact on costs.	Two generalised linear models were fitted independently. One had outcome hospitalisation costs and the other: hospital mortality. The variables fitted included patient demographics, co-morbidity index, hospital region, ventilation type, primary payer, admission month and length of stay. The optimal model for each outcome was selected based upon the evaluation of residual analyses and goodness of fit statistics.
(Hoogendoorn, Rutten-van Molken et al. 2011)	Stochastic Dynamic Population Model	Builds on Hoofendoorn and Feenstra's earlier models	Netherlands	To include exacerbations and probabilistic sensitivity analysis into a stochastic dynamic population model of COPD.	This model differs from the earlier model in that it is able to include the effects of COPD exacerbations and also allows for probabilistic sensitivity analysis. Changes in the COPD population are the result of incidence, smoking status changes disease progression along severity according to an algorithm for lung function decline and mortality. In each severity stage COPD patients have an annual probability of experiencing exacerbations. It is a high quality model as many of the input parameters have come from systematic reviews and meta-analyses of the literature and it was not industry sponsored and so should be independent.

(Hurley and Matthews 2007)	The Quit Benefits Model	Markov cycle tree model	Australia	To assess the consequences of quitting smoking in terms of cases avoided in four of the most common smoking-associated diseases.	Initially all subjects are considered to be in the health state “well”, including both smokers and quitters. Then every 1 year cycle a proportion of the simulated population will transition to illness states. The illnesses modelled for these purposes are stroke, lung cancer, acute MI and COPD. Smokers and quitters are analysed separately and there are 14 five-year age group categories up to 80-84 years. In this model no subjects are followed beyond the age of 85 years.
(Erbas, Ullah et al. 2012)	Australia forecasting model	Functional data analysis model	Australia	To project the age related forecasts for men and women separately for COPD mortality in Australia 2006-2025	The annual COPD mortality rates were calculated as a function of age. The age and mortality for each year were plotted and defined as mortality-age curves. The log of COPD mortality was taken and functional data analysis techniques were used to model the curves collectively as a functional time series. Each curve was expressed as a linear function of non-parametric basis functions, defined using a principal component decomposition. The age effects were separated from time effects and this enabled the age-specific mortality rates to be forecast into the future while taking the correlations between ages and across time into account.
(Stanciole, Ortegon et al. 2012)	POPMOD	Multistate population model	Sub-saharan Africa and SEAsia	To examine the population level costs, effects and cost effectiveness of individual interventions to combat COPD in low and middle income countries.	This is the precursor to DISMOD. Regional populations were categorised into two live health states: those with the disease and those susceptible but without disease; and dead. Cohort members are born into the population and move along by age. They develop disease and die according to incidence and mortality rates respectively. The population is modelled as males and females separately with one year age groups up to age 100. Values for incidence prevalence, case fatality and background mortality were obtained from the Global Burden of Disease study estimates.

(van Genugten, Hoogenveen et al. 2003)	Dynamic model of antismoking interventions	Dynamic multistate model	Netherlands	To describe the impact of increasing smoking prevalence over time in terms of future burden and costs of smoking related diseases.	This is a multistate model with four diseases: lung cancer, coronary heart disease, stroke and COPD. If you follow one subject through their health states on a one year cycle, in each year they may start smoking and may have one or more diseases. Transition probabilities govern these changes and the incidence of the smoking related diseases depends on age, gender and smoking behaviour. Life years lost and quality of life losses were calculated.
(Lopez, Shibuya et al. 2006)	Bayesian Age-period-cohort model	Bayesian Age-period-cohort model	England and Wales	To extrapolate from past trends. (Best suited to short term projections.)	Period and cohort effects were represented in a Bayesian software programme. Period effects represented factors affecting people of all ages at a particular point in time (e.g. treatment advances or influenza epidemics) while cohort effects represent factors more common in people born at particular time points (e.g. smoking habits). The model combines these effects with a constant rate plus some other covariates. Different prior beliefs about the age period and cohort parameters could be incorporated into the model.
(Bronnum-Hansen and Juel 2000)	PREVENT	Simulation model	Denmark	To estimate the mortality due to cigarette smoking with the PREVENT model and compare it with an estimate from a 1992 Peto model (Lancet 339:1268-1278)	The PREVENT model used data from the Danish National Bureau of Statistics, survey data on the prevalence of smoking and estimates on relative risks for different diseases from the CPS-II American study (with over a million patients) to estimate the mortality due to cigarette smoking in Denmark in 1993. Calculations were based on the “potential impact fraction” which indicates the proportion of new cases of disease in the population after changes resulting from the intervention. Alongside this was modelled the “trend impact fraction” to incorporate changes due to autonomous trends not due to interventions.

(Chapmann, Mannino et al. 2006)	The BOLD (Burden of Obstructive Lung Disease) Health Economic Model	Markov chain model	Norway and Iceland	To evaluate the costs of COPD disease and the value of new interventions	This model incorporated elements from Feenstra and Hoogendoorn's models to develop a health policy tool that can be used to calculate estimates of current and future impact of COPD. There were nine health states in this model and the transition probabilities amongst different stages of severity of COPD are informed by the Framingham cohort study. Exacerbations are also modelled.
(Pham, Ozasa et al. 2012)		Age-Period-Cohort model	Japan	To examine secular trends in COPD mortality, emphasising the contributions of age at death, time period and birth cohort to COPD mortality trends between 1950 and 2004 in Japan	Sex and age specific COPD mortality data were extracted from the Vital Statistics of Japan 1950-2004 for deaths at age 40 or older. The data were grouped into 10 five year age groups, 11 five year time periods (from 1950-1954 interval to 2000-2004 interval) and 20 five year birth cohorts (from 1865-1869 year interval to 1960-1964 year interval), The effects of age, time period and birth cohort on COPD mortality was analysed using and age-period-cohort model by means of Poisson Regression.
(Nacul, Soljak et al. 2007)	Health Survey for England model	Logistic regression analysis of risk factors	England	To estimate the population prevalence of COPD in England.	Explanatory variables from the Health Survey for England which were included in the final model were: sex, age, smoking status, ethnicity, area of residence and area based index of deprivation. A logistic regression model was used to derive the odds ratios for subjects with different combinations of risk factors in relation to baseline (baseline prevalence is the prevalence in non-smokers under 35 i.e. 0) and used to calculate the prevalence of COPD by gender for the relevant geographical area.

(Stang, Lydick et al. 2000)	COPD smoking prevalence model		Model derived using USA data then applied in Western Europe	To develop and validate a model based on smoking rates that will provide reliable estimates of the true prevalence of COPD including both clinically detected and undetected disease.	The model estimated the true COPD prevalence based on smoking rates. Studies included are those that report rates of COPD based on systematic spirometric examination of the population. The proportion of individuals in each smoking category (never, current or former) within specific age groups was multiplied by the proportion of individuals with COPD as estimated by regression lines. Among current smokers many studies report 10-15% have COPD, however, this varies with age from 17-43%.The two prevalences were compared.
(Holt, Zhang et al. 2011)		Spatial cluster analysis and Bayesian hierarchical spatial modelling	USA	To compare geographic variations in COPD hospitalizations across the USA in order to highlight explanatory environmental and social factors.	This model used Bayesian spatial mapping to analyse geographical variations in COPD hospitalisation risks for the medicare population in the USA. Regional characteristics and local characteristics were included in the analysis and indicators or spatial association were then computed to examine the local spatial dependence and discontinuities. The spatial model was implemented using Markov Chain Monte Carlo methods.
(IHME 2010)	DISMOD 3	Burden estimation model	International	To provide comparable estimates of the impact of diseases and injuries in all countries to inform priorities for policies, funding, and further research.	Models over 200 diseases in 22 world regions from regionally collected data in collaboration with the WHO. The model is fully Bayesian and internet based. COPD mortality was modelled using cause of death ensemble modelling (CODEm). DALYs were modelled separately in a multi-national collaboration.

Table 5.2 Data inputs and model outputs from all 22 modelling studies

Author	Model Title	Model Type	Data Inputs	Model Outputs
(Pichon-Riviere, Augustovski et al. 2011)		Health economic Monte Carlo microsimulation model	<ul style="list-style-type: none"> • Country-specific data for incidence and mortality. The model structure and parameter's calculation approach were validated and calibrated using a selected country dataset (Argentinean national Health Statistics year 2005). • COPD incidence was modelled as dependant of sex, age and smoking status, its progression as dependant of sex, years in current stage and smoking status. COPD deaths were modelled as dependant of sex and stage. • Where this was not available a common methodology was used to derive baseline incidence data in non-smokers from mortality data using assumptions from DISMOD II. 	<ul style="list-style-type: none"> • Results can be presented according to age, sex and previous cardiovascular history. • Benefit measures reported to compare the costs effectiveness of different smoking cessation strategies included: cost per quitter, cost per year of life gained, cost per event averted and cost per QALY. • A graphical depiction of decision uncertainty was made showing the cost effectiveness rate dispersion as 95% confidence intervals and also as cost-effectiveness acceptability curves.
(Feenstra, van Genugten et al. 2001)	The chronic disease model.		<ul style="list-style-type: none"> • Demographic data. • Age and sex specific COPD incidence, prevalence and excess mortality rates. (Including COPD all-cause mortality which was calculated including other smoking attributable cause). • Smoking data, start and stop rates, relative risks for smokers and former smokers, smoking prevalence rates. • Quality of life data. • Unit costs and trends per category of healthcare (GP, hospital, nursing home, medication). 	<ul style="list-style-type: none"> • Prevalence of COPD projected into the future for men women and total population. • DALYs lost owing to premature death among patients with COPD. • Future healthcare costs.

(Hoogendoorn, Rutten-van Molken et al. 2005)	Dynamic population model	Builds on Feenstra's single state model with a dynamic multi-state life table model	<ul style="list-style-type: none"> • As for Feenstra model plus: • Severity distribution of the prevalence. • Distribution of the incidence over the severity stages. • Decline in lung function by severity stage to model disease progression. • Mortality by COPD severity as per a model from meta-analysis. • COPD-related healthcare costs by severity. 	<ul style="list-style-type: none"> • Prevalence, mortality and costs specified by sex, age, smoking status, COPD severity and year.
(Peabody, Schau et al. 2005)	Prevalence estimation model	Multiplicative	<ul style="list-style-type: none"> • Six key risk factors as data inputs for the model: smoking percentage, age and gender, indoor air pollution, outdoor air pollution and occupational exposure to airborne particles. And a prevalence (age and sex specific) and relative risk for each of the key risk factors. E.g. percentage of population which are urban and rural for exposure to outdoor air pollution. 	<ul style="list-style-type: none"> • COPD prevalence was estimated for 12 countries.
(Shibuya, Mathers et al. 2001; Barendregt, Van Oortmarssen et al. 2003)	DISMOD 2	Burden estimation model	<ul style="list-style-type: none"> • Sex and age-specific mortality data on COPD and total mortality in a population. • Remission is zero for COPD. • Case-fatality/relative risk of death from comparing survival curves between groups of COPD patients and healthy population in the United States. • Risk factor variables by region: cigarette smoking and indoor air pollution from WHO's Comparative Risk Assessment database. • Years lived with a disability due to COPD by region calculated from prevalence and disability weights. 	<ul style="list-style-type: none"> • Estimation of incidence used to calculate years lived with disability and combined with years of life lost from the mortality estimates enabled disability adjusted life years due to COPD for each region to be estimated.

(Atsou, Chouaid et al. 2011)		Multistate Markov	<ul style="list-style-type: none"> • Prevalence of COPD data for five year age classes (Shahab). • Office of National Statistics English population details. • Transition probabilities from the Framingham cohort (Nielsen). • Transition probabilities according to smoking status (Hoogendoorn). • Probability of experiencing one or more exacerbations depending on severity stage (Soler-Cataluna). • Mortality for the UK general population in 2007, data from Mannino and data from Ekberg-Aronsson for smoking status of COPD patients. • Relapse rates from smoking status from Hoogendoorn. • Costs from Jansson, both direct and indirect costs. • Health utilities based of estimates reported by Borg and Cataluna. 	<ul style="list-style-type: none"> • Mean remaining lifespan of smokers versus quitters. • Additional disease-related cost of continuous smoking versus sustained abstinence. • Health gains associated with sustained abstinence. • Specific burden of continuous smoking during a 10 year period. • The above parameters for different scenarios of quitting during different stages of COPD severity.
(Kulik, Nusselder et al. 2012; Lhachimi, Nusselder et al. 2012)	Dynamo-HIA	Markov based multi-state	<ul style="list-style-type: none"> • Incidence and prevalence at the population level by age and sex of each disease. • Relative risks for each risk factor state for disease. • Exact configuration of disease life tables, e.g. the numbers and types of diseases, can be specified by the user. 	<ul style="list-style-type: none"> • Cohort disease life table for every simulated cohort or the period data for every simulated year. • Population pyramids or survival rates • Summary outcome measures e.g. cohort life expectancy, period life expectancy or disease free life expectancy.

(Najafzadeh, Marra et al. 2012)		Dynamic population projection model	<ul style="list-style-type: none"> • Prevalence of mild, moderate and severe COPD, according to age and smoking status. • Model of the likelihood of exacerbations according to disease severity to give annual rates of minor and major exacerbations. • Mortality rate for Canada and mortality rate according to disease severity and age and smoking status. • Prevalence of respiratory symptoms. • Progression rates to next COPD severity stage. • Quality of life proportional to disease severity. • Annual direct cost of COPD maintenance treatment and exacerbations according to disease severity. • Annual discount rate of 3%. 	<ul style="list-style-type: none"> • Number of COPD patients. • Annual societal cost of COPD. • Burden of COPD in terms of distribution of mild, moderate and severe COPD.
(Perera, Armstrong et al. 2012)		Generalised linear model	<ul style="list-style-type: none"> • Inpatient discharge records from the Agency for Healthcare Research and Quality Healthcare cost and Utilization Project Nationwide Inpatient Sample for 2006, United States. These records contain information on patient demographics, diagnoses, procedures, admission and discharge characteristics, payer, total charges, length of stay and hospital characteristics. 	<ul style="list-style-type: none"> • Hospitalisation costs. • In hospital mortality . • Factors associated with higher hospitalisation costs and higher rates of in-hospital mortality.

(Hoogendoorn, Rutten-van Molken et al. 2011)	Stochastic Dynamic Population Model	Builds on Hoogendoorn and Feenstra's earlier models	<ul style="list-style-type: none"> • As for Hoogendoorn 2005 plus: • Exacerbation frequency by COPD severity. • Case-fatality of exacerbations. • Relation between exacerbations and lung function decline. • QALYs lost for exacerbations. • Costs of exacerbations. 	<ul style="list-style-type: none"> • Total annual number of life years, quality adjusted life years, moderate and severe COPD exacerbations, total mortality and total COPD-related healthcare costs.
(Hurley and Matthews 2007)	The Quit Benefits Model	Markov cycle tree model	<ul style="list-style-type: none"> • Incidence and hospitalisation counts for four diseases: acute myocardial infarction, stroke, lung cancer and COPD, transformed into probabilities for disease specific incidence death and other events. These probabilities were initially obtained for smokers, ex-smokers and never smokers combined. Then probabilities for smokers were estimated by adjusting the population probabilities on the basis of smoking prevalence data and disease relative risks. Probabilities for quitters were estimated according to functions that describe the decreasing risk with time. • Probabilities for death for people with smoking related diseases and other causes. • Healthcare costs for smoking related diseases. • Utilities for smoking related diseases. 	<ul style="list-style-type: none"> • Incidence of four diseases: acute myocardial infarction, stroke, lung cancer and COPD. • Total deaths, including deaths attributable to all smoking related diseases and deaths due to the above four diseases. • Life expectancy. • Quality adjusted life expectancy. • Direct health costs from the above four diseases.
(Erbas, Ullah et al. 2012)	Australia forecasting model	Functional data analysis model	<ul style="list-style-type: none"> • Annual COPD (ICD10 J41-J44) death rates 2002 to 2005 for 5 year age groups from 50-54 to 85+ obtained from the General Record of Incidence of Mortality, Australia. 	<ul style="list-style-type: none"> • Forecasts of COPD mortality rates in Australia until 2025.
(Stanciole, Ortegon et al. 2012)	POPMOD	Multistate population model	<ul style="list-style-type: none"> • Mortality data. • Modelled demographic population data from Popmod. • Estimated disease rates prevalence and severity 	<ul style="list-style-type: none"> • DALYs averted, costs, cost effectiveness ratios for all interventions in both sub-Saharan Africa and South East Asia.

(van Genugten, Hoogenveen et al. 2003)	Dynamic model of antismoking interventions	Dynamic multistate model	<ul style="list-style-type: none"> • Demographic data on migration, birth and total mortality by gender and age from Statistics Netherlands. • Incidence rates for lung cancer, coronary heart disease, stroke and COPD, 1994 prevalence rates from several general practitioner registrations were combined. • Disease specific mortality rates, estimated as the difference between mortality in the general population and mortality among patients. • Gender and age-specific start and stop rates for smoking were estimated from observed trends 1987-1994 using age-period-cohort analysis. • Relative risks of smokers and former smokers for incidence of smoking-related diseases. • Dutch disease specific quality of life weights. • Cost of illness estimates per person from the Cost of Illness in the Netherlands study. 	<ul style="list-style-type: none"> • Prevalence of smoking among men and women for each of three tobacco control scenarios. • DALYs lost each year to smoking-related diseases. • Costs avoided by each of the tobacco control scenarios in comparison to the base-line.
(Lopez, Shibuya et al. 2006)	Bayesian Age-period-cohort model	Bayesian Age-period-cohort model	<ul style="list-style-type: none"> • COPD and population data for England and Wales 1945-1999 by five year age bands. • Population projections for 2000-2009 from the Government Actuarial Dept, London UK. • Age specific conversion factors for COPD ICD coding of mortality over time. 	<ul style="list-style-type: none"> • Graphs of trends in males and females of deaths from COPD in England and Wales by age.

(Bronnum-Hansen and Juel 2000)	PREVENT	Simulation model	<ul style="list-style-type: none"> • Based on data for the period up to 1993: the size of the Danish population and total mortality rates from the Danish National Bureau of Statistics. • Cause specific death rates from the Danish National Institute of Public Health for ischaemic heart disease, stroke, lung cancer and chronic bronchitis and emphysema. • Data on the prevalence of cigarette smoking by sex and age based on personal interviews carried out on 9,000-20,000 Danes annually by Gallup Market Analysis. • Relative risk estimates from Michael Thun, American Cancer Society. • Data from CPS-II: The American Cancer Society Prospective Studies. 	<ul style="list-style-type: none"> • Comparison between the numbers of death attributable to smoking as calculated using two different modelling methods from these data, for different diseases. • Life expectancy gains if there were no smoking attributable deaths.
(Chapmann, Mannino et al. 2006)	The BOLD (Burden of Obstructive Lung Disease) Health Economic Model	Markov chain model	<ul style="list-style-type: none"> • Aggregate estimates from the BOLD prevalence survey. • Local cost and population estimates. • Estimates of prevalence, smoking rates and healthcare utilisation rates. • Cost estimates based on local unit costs estimates for hospitalisations, physician (or other healthcare provider visits) and medications. • Incidence rates for COPD, mortality rates and smoking prevalence in younger populations. 	<ul style="list-style-type: none"> • Estimates on the costs related to the treatment of COPD. • The types of healthcare resources consumed. • Estimates of current and future costs of overall and per capita spending. • Costs stratified by severity. • Number of events in terms of hospitalisations, emergency dept visits and outpatient visits. • Estimates of mortality and quality of life.

(Pham, Ozasa et al. 2012)		Age-Period-Cohort model	<ul style="list-style-type: none"> • Sex and age-specific mortality data on COPD from the vital statistics of Japan 1950-2004 restricted to deaths at age 40 or older. 	<ul style="list-style-type: none"> • Graphs of the age-specific rate of COPD mortality per 100,000 persons by birth cohort. • Graph of effects of age, time periods and birth cohort on COPD population.
(Nacul, Soljak et al. 2007)	Health Survey for England model	logistic regression analysis of risk factors	<ul style="list-style-type: none"> • Health Survey England data 2001 findings for lung function parameters and their association with relevant risk factors. These data refer to 5269 men and 6133 women over age 15 with valid lung function measures. • Age group, smoking status, ethnicity and degree of urbanisation and deprivation score for the above sample. 	<ul style="list-style-type: none"> • Map of prevalence of COPD by age and gender in England.
(Stang, Lydick et al. 2000)	COPD smoking prevalence model		<ul style="list-style-type: none"> • Estimates of healthcare utilization rates for COPD from outpatient and inpatient data then divided by the mean number of visits per diagnosed COPD subject reported in the Third National Health and Nutrition Examination Survey to derive an estimate of diagnosed COPD prevalence. • Estimates from a literature review of population based surveys presenting information on smoking status and spirometry to estimate the true rate of COPD by age, sex and smoking status (smoothed by regression.) • Available information from government surveys on smoking rates in Germany, the UK, Spain and Italy and France. 	<ul style="list-style-type: none"> • Prevalence of diagnosed COPD from outpatient visits in the United States. • Prevalence of diagnosed COPD from hospital discharges in the United States. • Proportion of total COPD cases that are currently diagnosed in the United States. • Estimates of total COPD prevalence (including undiagnosed) in Western Europe.

(Holt, Zhang et al. 2011)		Spatial cluster analysis and Bayesian hierarchical spatial modelling	<ul style="list-style-type: none"> • COPD hospitalisation data from Medicare claims for 1995-2006 for each Health Service Area in the United States. 	<ul style="list-style-type: none"> • Map detailing COPD hospitalisation rates per 1000 enrollees among Medicare beneficiaries by state and by Health Service Area in the United States. • Map of COPD hospitalisation relative risk exceeding 1 attributable to regional or local environmental factors.
(IHME 2010)	DISMOD 3	Burden estimation model	<ul style="list-style-type: none"> • Many inputs from international data. • COPD mortality was modelled using cause of death ensemble modelling (CODEm) for those countries that did not have enough mortality data to use as input data. • DALYs were modelled separately in a multi-national collaboration. 	<ul style="list-style-type: none"> • Global ranking of diseases according to disability adjusted life years. • Global mortality from COPD for men and women in 2010. • Years of life lost globally due to COPD in 2010.

Table 5.3 Quality of reporting framework scores for modelling studies

Study	Aim /4	Transparency /3	Data /1	Data Modelling /3	Sensitivity Analysis /2	Internal Validity /1	Calibration /1	Policy Makers Involved /1	Predictive Validity /3	Discussion of Limitations /1	Total Score /20
(Pichon-Riviere, Augustovski et al. 2011)	4	2	1	3	1	1	1	1	3	0.5	17.5
(Feenstra, van Genugten et al. 2001)	4	3	1	3	2	0	0	1	2	1	17
(Hoogendoorn, Rutten-van Molken et al. 2005)	4	2	1	3	2	0	1	1	2	1	17
(Peabody, Schau et al. 2005)	4	3	1	3	2	0	0	0	3	0	16
(Shibuya, Mathers et al. 2001; Barendregt, Van Oortmarssen et al. 2003)	4	3	1	2	0	1	1	1	2	1	16

(Atsou, Chouaid et al. 2011)	4	3	1	3	2	0	0	0	2	1	16
(Kulik, Nusselder et al. 2012; Lhachimi, Nusselder et al. 2012)	4	3	1	3	1	1	0	1	0	1	15
(Najafzadeh, Marra et al. 2012)	4	3	1	3	2	0.5	0	0	0	1	14.5
(Perera, Armstrong et al. 2012)	3	2	1	3	2	0	0	0	2	1	14
(Hoogendoorn, Rutten-van Molken et al. 2011)	4	2	1	3	2	0	0	0	0	2	14
(Hurley and Matthews 2007)	4	3	1	3	2	0	0	0	0	1	14
(Erbas, Ullah et al. 2012)	4	1	1	3	2	0	0	0	2	0.5	13.5

(Stanciole, Ortegon et al. 2012)	3	3	1	3	2	0	0	1	0	0	13
(van Genugten, Hoogenveen et al. 2003)	4	2	1	3	1	0	0	1	0	0.5	12.5
(Lopez, Shibuya et al. 2006)	4	2	1	1	0	0	0	0	3	1	12
(Bronnum-Hansen and Juel 2000)	2	3	1	3	0	0	0	1	2	0	12
(Chapmann, Mannino et al. 2006)	4	2	1	3	2	0	0	0	0	0	12
(Pham, Ozasa et al. 2012)	3	2	1	3	0	0	0	0	0	0.5	9.5
(Nacul, Soljak et al. 2007)	2	3	1	3	0	0	0	0	0	0	9
(Stang, Lydick et al. 2000)	3	1	1	2	0	0	0	0	2	0	9
(Holt, Zhang et al. 2011)	4	0	1	2	0	0	0	0	0	1	8

Table 5.4 Reported outcome of each modelling study

Study	Model	Country	Main outcome – prevalence and burden	Additional outcome - costs
(Pichon-Riviere, Augustovski et al. 2011)	Monte Carlo Microsimulation model	Latin America	The International Society for Pharmacoeconomics and Outcomes Research criteria for model development and reporting were followed. Internal validation was performed by inputting null and extreme values to detect inconsistencies and programming errors. Calibration was performed using Argentinian data for all data excluding COPD data (as national statistics were thought to underestimate mortality due to COPD and so this would have been counter-productive). External validation was performed using the statistics from the PLATINO Burden of Obstructive Lung Disease survey of five South American cities which were found to have a good fit to the modelled statistics.	Cost results were not presented.
(Feenstra, van Genugten et al. 2001)	The chronic disease model. A Dynamic population model	Netherlands	The result from the model showed that there would be an increase in COPD disease burden between 1994 and 2015, in particular in women. By 2015, changes in the population alone would lead to an expected rise in prevalence to 33 per 1000 (men) and 25 per 1000 (women), 27 per 1000 total. This was largely as a result of past smoking behaviour and aging of the population. All efforts to decrease smoking prevalence by smoking cessation campaigns would only show an effect on disease burden in the long term. In the medium term in order to impact the disease burden, public health associations were recommended to focus on the burden from exacerbations.	For a projection that assumed changes in smoking behaviour, it was found that the total costs increased approximately 90% over the period 1994 to 2015. Healthcare costs increased more for women than for men in line with the projected prevalence increases.

(Hoogendoorn, Rutten-van Molken et al. 2005)	Dynamic population model	Netherlands	A projection between 2000 and 2025 demonstrated an increase in prevalence from 24 to 33 per 1000 for men and from 15 to 27 per 1000 inhabitants for women. Absolute number of deaths from COPD increased from 15000 to 23000 in 2025 in men and from 8000 to 16000 in women.	Total related healthcare costs in 2000 were estimated to be 280 million Euros. In 2025 this is estimated to grow to £495 million Euros. An 80% increase when costs per patient are kept constant.
(Peabody, Schau et al. 2005)	Prevalence estimation model	Spain, Norway, Nepal and Poland	Model projections were calculated for Spain (6.2%), Norway (6.3%), Nepal (11.1%) and Poland (6.7%) and were validated by comparison with survey obtained prevalence data. Model projections were not statistically different from collected survey prevalence data for any of the four comparison countries. This model was then used to project prevalence in 12 Asia-Pacific countries.	No cost estimates.
(Shibuya, Mathers et al. 2001; Barendregt, Van Oortmarssen et al. 2003)	DISMOD II	Worldwide	Estimates of incidence, prevalence and mortality of COPD for different WHO regions of the world are derived from available data on mortality and compared with estimates from other national burden of disease studies and other published data. Regions including China and India accounted for 52 and 11 per cent respectively of the global total in 2000 despite a smaller prevalence than in other regions where risk factors were more prevalent – this is because of these regions' large populations. In Africa where female smoking prevalence is low the male to female ratios of years lost to disability were higher than in other regions. The COPD global total disease burden from disabling states accounted for 2.7% of global total disease burden using 1990 weights. Other comparisons of the regional proportions of the global COPD burden were made.	No cost estimates.

(Atsou, Chouaid et al. 2011)	Multistate Markov	England	COPD patients remaining continuous smokers till death had a mean remaining lifespan of 15.6 lifeyears worth 8.47 QALY. However, if they became sustained quitters then they would gain 2.73 lifeyears and 1.225 QALY.	Considering the patients' lifespans the additional disease-related costs of continuous smoking versus sustained abstinence was £1661 per patient. During a ten year period, the 1.4 million COPD smokers in England resulted in total disease related costs of £1.657 billion.
(Kulik, Nusselder et al. 2012; Lhachimi, Nusselder et al. 2012)	Dynamo-HIA, Markov based multi-state	UK	An example of the total elimination of smoking in the UK projected 25 years into the future with the whole population consisting of never smokers and no uptake of smoking: the prevalence of COPD halved from 1.7% to 0.5%. Prevalence of ischaemic heart disease and stroke fall, prevalence of lung cancer, oesophageal cancer and oral cancer all decrease. Prevalence of diabetes, breast cancer and colorectal cancer all increase due to larger number of surviving individuals at risk of such chronic diseases.	No cost estimates.
(Najafzadeh, Marra et al. 2012)	Dynamic population model	Canada	In terms of demographics alone, the population over the age of 40 in Canada will increase from 17 million to 24 million by 2035. In terms of COPD, the model projects that the number of COPD patients will increase from 3.45 million in 2011 to 5.83 million in 2035. In terms of strategies for managing this increase, the most effective strategy was to improve prevention of exacerbations in order to reduce costs and burden.	The annual societal cost of COPD in Canada was \$4.52 billion in 2011 and will be \$3.61 billion (\$7.33 billion undiscounted) per year in 2035. Over the 25 years from 2011 to 2035 COPD will cost Canadian society \$101.4 billion (\$147.5 billion undiscounted) and 12.9 million QALYs lost (19.0 million undiscounted).

(Perera, Armstrong et al. 2012)	Generalised linear model	USA	1,254,703 hospitalisations in the USA in 2006. Average length of stay was 5.9 days. Inpatient mortality was 4.3%. After controlling for variables in the models several co-morbidities were associated with both higher costs and higher mortality including acute MI, congestive heart failure, cerebrovascular disease, lung cancer, cardiac arrhythmias, pulmonary circulation disorders and weight loss.	Acute exacerbations of COPD overall aggregate cost \$ 11.98 billion in 2006, adjusted to \$14.05 billion in 2010. Costs were higher for non-invasive ventilation cases by a factor of 1.36 and for invasive ventilation by a factor of 1.84.
(Hoogendoorn, Rutten-van Molken et al. 2011)	Stochastic Dynamic Population Model	Netherlands	The cost effectiveness of three different COPD intervention strategies was modelled. These were a pharmacologic intervention, a smoking cessation intervention and a pulmonary rehabilitation intervention. Total exacerbations avoided were 77,700 for the pharmacologic intervention, 90 for increased smoking cessation and 0 for community rehabilitation.	Savings in COPD-related healthcare costs were 37.9 million Euros for the pharmacotherapy, -0.9 million Euros for the smoking cessation and 0 (break-even) for the community based pulmonary rehabilitation.
(Hurley and Matthews 2007)	The Quit Benefits Model, a Markov cycle tree model	Australia	Overall 40 per every 1000 who quit smoking individuals will be spared a diagnosis of acute MI/COPD/lung cancer or stroke in the first ten years following quitting with an estimated saving of 47 life years and 75 QALYs.	For every 1000 males chosen at random from the reference population who quit smoking there is an average saving of \$408,000 in healthcare costs associated with the four diseases (MI, COPD, lung cancer or stroke) \$328,000 for females.
(Erbas, Ullah et al. 2012)	Australia forecasting model	Australia	Trends in COPD mortality rates among men in Australia have been declining since 1970 and the rates are expected to continue to decline for the next 20 years. Likewise the trends in COPD mortality rates among women are expected to decline but at a lower rate than that of men.	No cost estimates.

(Stanciole , Ortegon et al. 2012)	POPMOD	Sub Saharan Africa and South East Asia	The model considered both asthma and COPD. Inhaled bronchodilators averted 58 DALYs per million population in Africa and 370 DALYs in SE Asia. This was based on the use of long term anticholinergic bronchodilator reducing COPD associated disability by up to 97% as measured by St George’s Respiratory Questionnaire. (This would seem to be very optimistic).	For COPD the most cost-effective intervention was influenza vaccination – incremental cost \$4,590 per DALY averted. COPD had higher costs in SE Asia than in Africa because of the high number of patients who could benefit from treatment in that region.
(van Genugten, Hoogenveen et al. 2003)	Dynamic model of antismoking interventions	Netherlands	Several scenarios of antismoking measures were considered. The reference scenario, with future smoking prevalence based on trend extrapolation resulted in the prevalence of smoking men declining from 35% in 1995 to 29% in 2020 and 28% in 2050. For women there is a decrease from 27% in 1995 to 23% in 2050. Other scenarios included stopping young people from starting and alternatively increased quit rates amongst established smokers. Taxing tobacco also reduced smoking prevalence.	If the Netherlands were able to increase the rate of smokers quitting, this model estimates that up to 80 million Euros for males and 100 million Euros for females, could be saved. Avoided costs were even greater if people could be prevented from starting but there was a long time lag before benefits were seen. The tax scenario behaves very similarly to the don’t start scenario.
(Lopez, Shibuya et al. 2006)	Bayesian Age-period-cohort model	England and Wales	COPD death rates in males aged over 45 will continue the decline of recent decades falling by 24% by 2009 from a 1999 baseline. In females projections suggested fluctuations in COPD death rates and wide credible intervals with 2% higher rates in 2009. A fall in rates was strongly suggested for males in their 60s and 70s and females in their 60s. However, for other age groups 90% credible intervals encompassed both a fall and rise in rates.	No cost estimates.

(Bronnum-Hansen and Juel 2000)	PREVENT	Denmark	The PREVENT model estimated that 1052 men and 864 women died a smoking-attributable death in 1993. Overall 33% of deaths in men and 23% in women were attributable to smoking. The authors validated their model with comparison to the Peto model.	No cost estimates.
(Chapman, Mannino et al. 2006)	The BOLD (Burden of Obstructive Lung Disease) Health Economic Model	Norway and Iceland	The model was used to project the prevalence of COPD among males and females over the age 40 over 10 years, i.e. from 2005 to 2015. After 10 years the prevalence of COPD stage I + will be 22% in Iceland and 24% in Norway.	The current overall annual cost in Euros for COPD patients (2005) in Iceland was 12 million rising to 13 million in 10 years' time. For Norway the annual cost was 141 million rising to 154 million in 10 years.
(Pham, Ozasa et al. 2012)	Age-Period-Cohort model	Japan	From 1950 to 2004 COPD mortality substantially decreased, from 71.3 per 100 000 to 19.7 in men and from 41.7 to 4.3 in women. The greatest decreases were observed from 1959 to 1956 when the age standardised rate decrease by half, the downward trends then slowed. The effects of age, time period and birth cohort were expressed in terms of relative risk The age effects increase with age in both sexes. The period effects initially rapidly decline then start increasing in the 1960s. In men period effects tended to increase in recent years, whereas they continue to decrease in women. Birth cohort effects initially increase up to the 1880-1889 cohort then decrease continuously until the 1960-1964 cohort.	No cost estimates.

(Nacul, Soljak et al. 2007)	Health Survey for England model logistic regression analysis of risk factors	England	<p>The average prevalence in over 45s varies fourfold with the highest values in men in deprived urban areas and lowest in women in wealthy rural areas.</p> <p>Comparison of model prevalence estimates gives an indication of undiagnosed subjects and unmet need. The prevalence of COPD in 2004 according to General Practice Quality and Outcomes Framework figures was only 1.4%, translating as 600 000 cases of COPD (nearly half) remaining undiagnosed.</p>	No cost estimates.
(Stang, Lydick et al. 2000)	COPD smoking prevalence model	Western Europe	This study's main outcome is that only a small proportion of people with COPD have a spirometric diagnosis and therefore an effort to increase case-finding is desirable. Estimates were a population of 3.0 million people with COPD in the UK, 2.7 million in Germany, 2.6 million in Italy and 2.63 million in France and 1.5 million in Spain.	No cost estimates.
(Holt, Zhang et al. 2011)	Spatial cluster analysis and Bayesian hierarchical spatial modelling	USA	The Bayesian spatial mixture model showed that 73% of the variability of the COPD hospitalisation relative risk was attributable to the spatially structured effects. Thus regionalised influences on the social and physical environment had more impact on COPD hospitalisation rates than local social and physical environmental factors. The specific factors were not identified in the report.	No cost estimates.
(Murray, Vos et al. 2012)	DISMOD 3	Worldwide	The global DALY rank for COPD has fallen from 6 th to 9 th . This is as a result of a decrease in indoor air pollution in India and China despite an increase in cumulative cigarette smoking worldwide. Global mortality from COPD for men and women, all ages was estimated as 2,899,900 in 2010. Years of life lost globally due to COPD in 2010 for both sexes all ages was estimated as 47,359,000 years down from 58,186,000 in 1990.	No cost estimates.

5.3.2 Narrative results

This section reports a narrative version of the results, namely by summarising the results of the studies in a form as follows: the seven models reported in the following nine papers (Feenstra, van Genugten et al. 2001; Shibuya, Mathers et al. 2001; Barendregt, Van Oortmarssen et al. 2003; Hoogendoorn, Rutten-van Molken et al. 2005; Peabody, Schau et al. 2005; Atsou, Chouaid et al. 2011; Pichon-Riviere, Augustovski et al. 2011; Kulik, Nusselder et al. 2012; Lhachimi, Nusselder et al. 2012) that have been awarded the most points in the quality of reporting assessment were ranked in that order and are presented here. Each modelling procedure is described, then, as these models were reported in a high quality fashion, a comment may be made on the underlying quality of the model; as enough details have been included to draw conclusions as to how well structured and designed the model was. The main focus of quality was whether sensitivity analysis had been reported and what were the results of this. I also considered whether other features of quality assurance were present, such as debugging and external validation of results.

5.3.2.1 Pichon-Riviere smoking burden model (Pichon-Riviere, Augustovski et al. 2011)

This model was designed to meet the needs of 68 decision makers who had been surveyed across seven countries in Latin America (i.e. Argentina, Bolivia, Brazil, Chile, Colombia, Mexico and Peru). It considered heart disease, cerebrovascular disease, COPD, pneumonia/influenza, lung cancer and nine other neoplasms. The model used country specific data sources where possible. However, as a result of the lack of local data, estimates of incidence were often derived from mortality data using the WHO DISMOD II methodology.(Shibuya, Mathers et al. 2001) The baseline risk in non-smokers was then calculated based on the age-, sex- and country-specific smoking prevalence as well as disease specific smoking relative risk. The parameters for smokers and former smokers were then calculated from this baseline risk. The model was a first order Monte Carlo (probabilistic) micro-simulation of individual subjects incorporating the natural history, costs and quality of life of the above diseases. A functioning version of the model was constructed, validated and calibrated using data from Argentina.

The Argentinian version of this model underwent extensive internal validation with debugging by the inputting of null and extreme values and checks for inconsistencies. The model was then calibrated by comparing general mortality and all age- and sex-specific death rates predicted by the model with local health statistics. COPD was an exception from this process as COPD mortality was agreed to be underestimated in national statistics.

Equations were modified to improve fit to the reference values. Lethality and survival rates were estimated from local and international studies and also used to calibrate the model by visual exploration of observed and expected curves to confirm a good fit. After calibration, as had been expected, correlation between predicted and observed results was better among high incidence events.

External validation for age- and sex-specific COPD predicted prevalence was performed by comparison with the results from the Latin American Project for the Investigation of Obstructive Lung Disease, a population-based survey carried out in five Latin American cities.(Menezes, Lopez et al. 2009) The model underestimated the level of COPD in comparison to the survey, however, the results were within 5% for each age group. This model's report followed the International Society Pharmacoeconomics and Outcomes Research guidelines for model development and reporting (Weinstein, O'Brien et al. 2003) and therefore included extensive description of preliminary consultation with policy and decision makers and detail regarding debugging, internal validation, calibration and external validation procedures.

5.3.2.2 Feenstra chronic disease model (Feenstra, van Genugten et al. 2001)

This model was designed to describe the population with COPD in the Netherlands. A dynamic multistate life table model combined information on the demography of the Dutch population and their smoking behaviour. The number of new cases of COPD each year were calculated from the incidence rates of COPD for smokers and former smokers in combination with COPD prevalence data. Independently modelled modules exist for mortality from lung cancer, stroke, coronary heart disease and asthma. The model is dynamic in that each year's total incorporated calculations for birth, migration and mortality.

One way sensitivity analysis was performed to determine the effect of changes in inputs of incidence and excess mortality rates on the output prevalence rates. It was found that a 20% change in incidence input rates resulted in a 17% change in prevalence rates and a 20% change in excess mortality input resulted in a 4% change in estimated prevalence. Therefore the model was more sensitive to changes in incidence rates than changes in excess mortality rates. However, overall it was felt that the model was relatively stable when challenged with changes in input rates.

5.3.2.3 Hoogendoorn COPD model 2005 (Hoogendoorn, Rutten-van Molken et al. 2005)

This model was developed from the Feenstra chronic disease model (Feenstra, van Genugten et al. 2001) by the addition of the four Global initiative for Obstructive Lung Disease (GOLD) severity stages for COPD into the model to provide estimates for the Dutch population. The COPD population is divided into mild, moderate, severe and very severe COPD subpopulations and also by smoking status: never-smoker, smoker or former smoker. Annual incidence and mortality rates are applied to each age, sex and smoking subpopulation. The model exhibits the Markov property, in that the model has no memory for the previous years' conditions of each patient at the individual level. Costs are calculated by multiplying annual cost data for each severity stage by the number of patients in each stage.

Sensitivity analyses for this model were conducted by changing the input severity distribution of the COPD prevalence so that it was assumed firstly that COPD patients began with less severe COPD (i.e. mild and moderate cases) next it was assumed that the distribution changed to include a higher proportion of severe and very severe cases. The next sensitivity analysis to be conducted changed the severity distribution of the incident COPD cases to first less severe and then more severe. Further sensitivity analyses were conducted varying the rate of lung function decline and changing the impact on lung function decline of stopping smoking from an improvement to a null impact. The results of the sensitivity analyses were that all prevalence results were within a range of 5% of the projections of the base case. The results were most sensitive to the distribution of the severity of COPD at incidence. Cost projections were more sensitive than prevalence projections due to the difference in costs for different COPD severity stages.

5.3.2.4 Peabody prevalence estimation model (Peabody, Schau et al. 2005)

This model used six key risk factors for COPD identified and quantified from the literature: smoking, age, gender, indoor air pollution, outdoor air pollution and occupational exposure to airborne particles. They assumed no cases under the age of 30 years, and a smoking prevalence of 15%. Based on the literature it was assumed that there was an exposure of non-smoking population to occupational airborne particles of 1.9% and a prevalence of COPD in urban populations of 1.4%. In the model, population was distributed between urban and rural populations and geographical distribution was linked to environmental exposure. There was also an input for national socio-economic development based on World Bank figures. COPD prevalence was then estimated for four countries and compared with survey data for external

validation. Then COPD prevalences were estimated for an additional 12 countries (Bangladesh, Brazil, China, Denmark, Korea, Mexico, Philippines, Russia, Thailand, Turkey, UK and the US).

Sensitivity analysis was performed on hypothetical populations with differences in their rural/urban population distribution and high/low income split. The model was externally validated by comparing its results to survey results for Nepal, Norway, Poland and Spain. The predictions were not statistically different from the survey results for any of the countries suggesting that the model is a useful prediction tool.(Peabody, Schau et al. 2005) The authors concluded that the model appeared robust and capable of providing reliable projections although further validation is desirable, especially in developing countries where data may be scarcer.

5.3.2.5 WHO burden of disease model DISMOD II 2001 (Shibuya, Mathers et al. 2001; Barendregt, Van Oortmarsen et al. 2003)

DISMOD (Disease Model) II used the WHO world regions to calculate the global burden of COPD in terms of years lived with disability and mortality in terms of years of life lost. The model was based on a regression model including a measure of smoking impact and an air pollution variable to take into account proportions of households in each region that use indoor biofuels and age and sex dummy variables which reflected variability in the exposure data for different regions. The input data to the model was COPD related mortality rates and total mortality rates per WHO region. These data were used to generate an equation which could be solved for the prevalence and incidence of COPD.

No sensitivity analysis using DISMOD II was undertaken. However, the predictive validity of the model was checked by comparing the estimated relative risks for death from COPD for the America and Pacific regions with recent burden of disease analyses from the USA and Australia. Published prevalence data based on spirometry in the WHO regions were also compared to the DISMOD II outputs and found to be consistent with the exception of estimates from the Global Burden of Disease 1990 estimates due to improved methodology since then.(Murray 1994)

5.3.2.6 Atsou smoking burden model (Atsou, Chouaid et al. 2011)

This modelled the severity stages of COPD: mild, moderate, severe and very severe. Transition probabilities could be altered for modelled patients moving up these severity stages as their disease progressed. The model aimed to report the impact of smoking cessation on a COPD patient's life expectancy in terms of individual health gains. The model

then performs a health economic assessment of the impact of various cost-effectiveness of smoking cessation programmes in England. This was a Markov-type model, but used individual simulation in place of the more usual Markov cohort simulation to assess both expected values and variability.

The sensitivity analyses that were conducted involved changing the transition rates from one disease severity stage to the next, the mortality and exacerbation rates and the costs of COPD management, different discounting rates and different smoking cessation rates. For the scenarios investigating the costs, QALYs and cost-effectiveness of the smoking cessation programmes, the effects of different input costs for the programme and different quit rates, were investigated. The results were sensitive to transition rates from one disease stage to the next and to an increase in mortality rates; however, they were not very sensitive to an exacerbation-free rate of 15% greater than in the reference case. But then when it was suggested that ex-smokers experience fewer exacerbations than smokers (probability of an exacerbation-free year 30% higher in ex-smokers than current smokers) there were monetary and healthcare gains. The model was not very sensitive to changes in disease management costs. Overall, extensive sensitivity analysis was undertaken so that the model could be used for a wide variety of smoking cessation scenarios.

5.3.2.7 DYNAMO-HIA model (Kulik, Nusselder et al. 2012; Lhachimi, Nusselder et al. 2012)

This was a multi-state Markov-type and partial micro-simulation model designed for European Union-wide use. Data from most European Union countries were collected for nine diseases: COPD, diabetes, ischaemic heart disease, stroke, lung cancer, oral cancer, oesophageal cancer, colorectal cancer and breast cancer. The model was designed so that member government policy makers could firstly quantify the development of risk factor exposure over time and secondly quantify the impact of health interventions on the population, such as smoking cessation initiatives. This would be represented in the model as a risk factor reduction resulting in lower prevalence and hence mortality. The risk factors considered in the model were body mass index, alcohol and tobacco consumption. The model is publicly available online at www.dynamo-hia.eu. The output of the model is in terms of the modelled cohort, most often the input country or region, giving population pyramids and survival rates and life table data including life expectancy. As the inputs can be updated, the model can be modified.

The model can be set for sensitivity analysis as required, either one way or multi-way. The authors argued that gathering data for predictive validity assessment is embedded with

uncertainty and therefore projecting future disease patterns is always compromised. They argue that their strength is in comparing alternative scenarios for the future as the same baseline data can be used in each scenario. They chose to incorporate trend-free data into the model as past trends are not always indicative of future trends, so, in this way, the model is also a compromise (Lhachimi, Nusselder et al. 2012)

5.4 Selection of model for further study

In conducting the systematic review, it became apparent that three of the 22 models were very closely related: (Feenstra, van Genugten et al. 2001; Hoogendoorn, Rutten-van Molken et al. 2005; Hoogendoorn, Rutten-van Molken et al. 2011). On further enquiry I discovered that the Dutch Model in the Feenstra et al study (Feenstra, van Genugten et al. 2001) was the original model and was updated firstly to the (Hoogendoorn, Rutten-van Molken et al. 2005) model then to the (Hoogendoorn, Rutten-van Molken et al. 2011) model. The team which had produced these three models work closely together in the Netherlands. The three models had scored 17, 17 and 14 respectively on the quality of reporting scale. The reason for the slightly lower score of the 2011 model was that some of the quality elements had been reported more fully in the 2001 and 2005 manuscripts and so the 2011 manuscript did not receive the points for these elements. On reviewing the models I noted the extensive sensitivity analysis which had been undertaken and the high quality of the models. This Dutch team clearly held a dominant position in the modelling of COPD and had the longest history of experience in this area.

I therefore decided to contact the Dutch team with a view to collaboration. They offered that I could use their latest model along with English and Scottish routine data. In addition, the model could be run in Mathematica (Wolfram 2013) which was a readily available computer programme. Another advantage of this choice was that, unlike other models, the outputs were not only limited to current estimates but also could be used to project estimates of the prevalence and burden of COPD in England and Scotland in the future.

5.5 Conclusion to Chapter 5

Models which aim to predict the prevalence and burden of COPD have widely differing structures and include Markov models, demographic models and risk factor models. In this chapter I have described an extensive and systematic search for models and the process I undertook in order to appraise their quality of reporting. I then listed the features of all 22 models in tables including their main characteristics and the inputs and outputs required for the models. Then I presented a grid with quality of reporting scores. This was followed by a

summary table of the outcomes of each model in context. The seven models which scored most highly among the quality of reporting were then assessed for their underlying quality, with a focus on sensitivity analysis. A narrative summary of these seven models was given.

Finally, I observed that three of the models were closely related and were produced by the same team in the Netherlands. These three models had all obtained reasonably high scores on the quality of reporting scale. It became clear that the Dutch team had much experience in developing and running COPD population models. An approach for collaboration with the team in the Netherlands was successful and so their most recent model was chosen to produce projections of the prevalence and burden of COPD in England and Scotland in the next phase of the PhD project. This modelling phase is discussed in Chapters 6 and 7.

Chapter 6: Model description and data requirements

6.1 Introduction

The previous chapter reported a systematic review to find models for calculating the prevalence and burden of COPD. Chapter 5 concluded with the selection of related models for further study. (Feenstra, van Genugten et al. 2001; Hoogendoorn, Rutten-van Molken et al. 2005; Hoogendoorn, Rutten-van Molken et al. 2011). The Dutch team who had produced these models in Rotterdam agreed to collaborate and to share their most up to date model (as of October 2013) with me for calculating projections of prevalence and burden of COPD for England and Scotland from 2011 to 2030.

This chapter has two main parts: firstly, a description of the Dutch Model; this is then followed by an overview of the data required to run the model. The data section continues with an explanation of the origins of the data for each component of the Dutch Model. The governance processes which were gone through to obtain the data permissions are also described alongside each type of data.

In Chapter 7, the results of the Dutch Model are presented for prevalence, healthcare costs and mortality, and finally the sensitivity analyses are presented. These results represent original estimates for the prevalence, costs and mortality of COPD in England and Scotland as estimated for 2011 to 2030.

6.2 Dutch Model description

The Dutch model, as discovered through the systematic review of Chapter 5, had evolved over the years from a dynamic epidemiological projection model (Feenstra, van Genugten et al. 2001) to a model which incorporated the distribution of severity within the COPD population and also estimates of numbers and costs of exacerbations. The 2011 incarnation of the model also could use stochastic techniques (i.e. mathematical techniques involving running the model through a certain high number of repetitions thus giving an overview of the most likely outcomes) to provide uncertainty analysis for the model (Hoogendoorn, Rutten-van Molken et al. 2011). Uncertainty was modelled for selected parameters (see Chapter 7 and Appendix 5 for a fuller discussion of uncertainty in the model). Subsequently the most up to date version of this model (supplied in October 2013) is referred to as “The Dutch Model.”

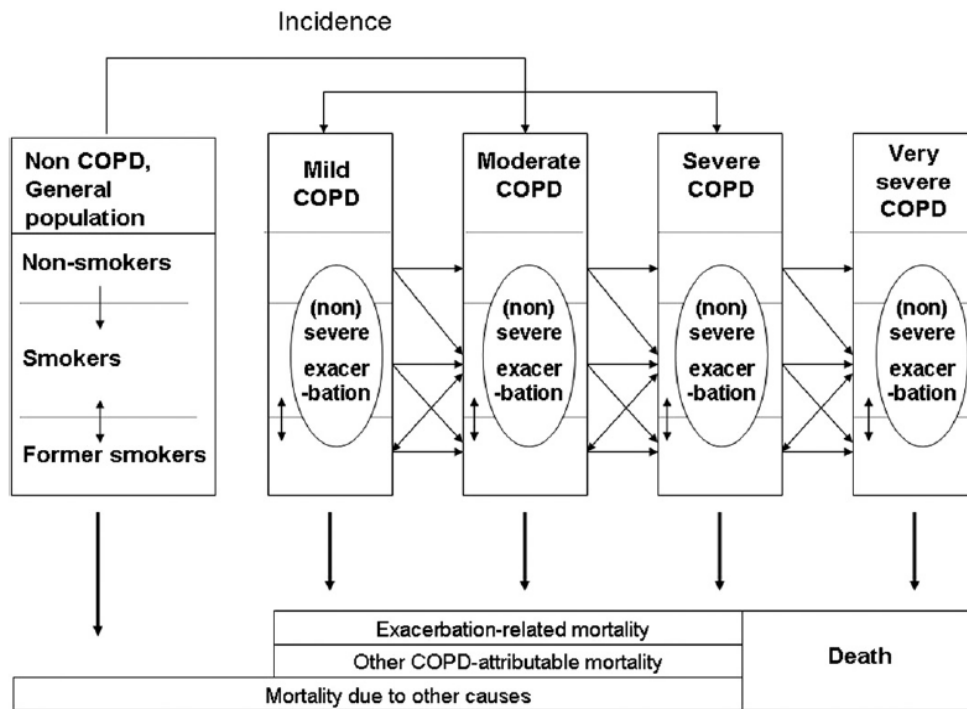
The Dutch Model is a Markov type multi-state cohort model where members of the population are first characterised as either having COPD or not having COPD. The

prevalence of COPD included four COPD severity stages based on the 2011 GOLD classification as illustrated in Figure 6.1. The prevalence was distributed over the three smoking classes (i.e. smoker, non-smoker and former smokers) using the numbers of smokers in each smoking class and the relative risks of smokers and former smokers of having COPD.

COPD cohorts were modelled progressing through the stages of severity: mild, moderate, severe and very severe according to transition probabilities that were related primarily to lung function decline. The lung function decline was modelled as the annual decline in FEV1% predicted depending on age, sex, smoking status and FEV1% predicted. Exacerbations accelerate this decline slightly. Outcomes were death from COPD or other causes or survival to 2030. This annual decline was modelled using a random effects model based on 5,000 patients with longitudinal spirometry information in the Lung Health Study.(Scanlon, Connett et al. 2000) Transitions between smoking status categories, based on proportions of people starting, stopping and restarting smoking, were also modelled.

Difference equations described the change of state variables over time, as a result of transitions from one state to the other. The cycle length was one year. The two most important state variables in the model were: 1) probability values per COPD severity stage (including not having COPD as a special stage), per smoking class, per age class and sex; and 2) coefficients describing the distribution of FEV1% predicted within each COPD severity stage per smoking class. This latter distribution was interpreted as the mean distribution over both sexes and all ages. The severity distribution of COPD prevalence at the start of the modelling process was based on Dutch GP data.(Van Weel, Smith et al. 2000; Wijnhoven, Kriegsman et al. 2001). The distribution of COPD severity stages among the incident cases was calculated to preserve this severity distribution of COPD prevalence in the first year of the model.

Figure 6.1 The Dutch Model. Reproduced with permission (Hoogendoorn, Rutten-van Molken et al. 2011)



The number of COPD exacerbations at each stage of severity is also important for calculating the transition probabilities between severity categories. (Hoogendoorn, Feenstra et al. 2010; Hoogendoorn, Hoogenveen et al. 2011) A severe exacerbation was defined as an exacerbation necessitating a hospital admission. A moderate exacerbation was defined as an acute exacerbation of COPD necessitating a visit to a GP for an increase in symptoms. (Hoogendoorn, Rutten-van Molken et al. 2010)

The costs of COPD were calculated from the costs of maintenance in each severity stage plus the costs of exacerbations. In the original model, these costs were distinct for age and sex groups with costs of maintenance therapy increasing with age and costs for women being higher than for men. (Hoogendoorn, Rutten-van Molken et al. 2005) However, the data available for the UK adaptations of the model did not include such detail – only giving differential costs for different COPD severity stages. This will be explained further in the section on data sources later in this chapter.

6.3 Modelling data input overview

There was a long process of ensuring that the data with which the model would be run were suitable. The plan was to re-run the Dutch Model with English and Scottish data, to produce respective estimates for each country. The final data sources that were used for each of the

model inputs are summarised in Table 6.1. The base year for the model was selected to be 2011 so, whenever possible, estimates for that year were obtained.

Table 6.1 Model inputs and respective data sources for England and Scotland

Data Input	English Source	Scottish Source
1 year age and sex distribution of the general population in 2011	Office for National Statistics(ONS 2011)	General Register Office for Scotland(GROS 2012)
Incidence of COPD	Clinical Practice Research Datalink(CPRD 2011)	Lothian COPD Cohort (see below)
Prevalence of COPD	Clinical Practice Research Datalink(CPRD 2011)	Scottish Practice Team Information Database (see below)
Smoking prevalence	The General Lifestyle Survey for England 2011(ONS 2011) and Health Survey for England 2010(ESRC 2011)	Scottish Health Survey 2011(Bradshaw, Bromley et al. 2011) and Scottish Schools Adolescent Lifestyle and Substance Use Survey 2010(Black, Eunson et al. 2010)
Start smoking rates	Avon Longitudinal Study of Parents and Children(Bristol University 1991)	Avon Longitudinal Study of Parents and Children(Bristol University 1991)
Stop smoking rates	Smoking Toolkit Survey(West and Brown 2013)	Smoking Toolkit Survey(West and Brown 2013)
Restart smoking rates	Original Netherlands data(Hoogendoorn, Rutten-van Molken et al. 2005)	Original Netherlands data(Hoogendoorn, Rutten-van Molken et al. 2005)
Age and sex specific relative risks of smokers and non-smokers to develop COPD	Original Netherlands data(Hoogendoorn, Rutten-van Molken et al. 2005)	Original Netherlands data(Hoogendoorn, Rutten-van Molken et al. 2005)

Modelling lung function decline	Original Netherlands data(Hoogendoorn, Rutten-van Molken et al. 2005)	Original Netherlands data(Hoogendoorn, Rutten-van Molken et al. 2005)
COPD related maintenance costs	Microcosting from UK indacaterol study(Price, Asukai et al. 2013)	Microcosting from UK indacaterol study(Price, Asukai et al. 2013)
COPD Excess Mortality	Calculated with relative risks from CPRD and English total mortality and prevalence	Calculated with relative risks from CPRD and Scottish total mortality and prevalence
Severity distribution of COPD	Original Netherlands data(Hoogendoorn, Rutten-van Molken et al. 2005)	Original Netherlands data(Hoogendoorn, Rutten-van Molken et al. 2005)

6.4 Model data inputs and sources

6.4.1 General population data

6.4.1.1 England

Demographic details of the English population in 2011 were obtained from the Office for National Statistics (ONS). These included the number of males and females at each age in one year age bands to age 100.

6.4.1.2 Scotland

Demographic details of the Scottish population in 2011 were obtained from the General Register Office for Scotland (GROS) mid-year population estimates. These included the number of males and females at each age in one year age bands to age 100.

6.4.2 Incidence of COPD from general practice data

6.4.2.1 England

The Clinical Practice Research Datalink CPRD is the English NHS observational data and interventional research service, jointly funded by the NHS National Institute for Health Research (NIHR) and the Medicines and Healthcare products Regulatory Agency (MHRA). CPRD extracts data from electronic health records in primary care (see Chapter 3, Table 3.1). My colleague, Dr Daniel Kotz, and two of my supervisors, Dr Colin Simpson, and Prof

Aziz Sheikh, had formerly approached CPRD to extract a COPD patient dataset based on general practice data and linked to mortality data (see CPRD protocol Appendix 1.1). They had used these data to develop an individual prognostic model for COPD based on GP data.(Kotz, Simpson et al. 2014) With permission from CPRD, I was able to extend the analysis of these data to include provision of incidence and prevalence estimates for COPD in primary care (see Appendix 1.2 for amendment to protocol). The data were drawn from a broad and representative sample of general practices throughout England, Scotland and Wales, however they were coded for geographical region and so incidence and prevalence of COPD could be estimated for England. These were requested in one year age and sex groups. The risk of identification and the likelihood of uneven data in such narrow age categories was lower than in Scotland because of the larger population size.

The incidence of COPD in English patients was calculated using the CPRD resource as above. As this COPD cohort were derived from data which had linked mortality data, only those patients could be used as source whose practices had provided data to participate in the linkage. This limited the source population to 4,780,887 patients.

The source population for the cohort was further limited to those that have at least five years of “up to standard” follow-up on the CPRD GOLD data extract (see Table 3.1 which describes the sources of this GP database, the CPRD GOLD database is the highest quality CPRD database, there is also a CPRD Silver database which contains less high quality data). This gave a total of 2,655,917 patients. “Up to standard” data essentially involves ensuring that the data can be used for research purposes by examining for gaps as a result of problems in coding when patients move practices or die.

The numerator and denominator for the incidence were provided, as documented in Appendix 1.3.

6.4.2.1.1 Permissions required for use of the CPRD data

The following permissions were required to use CPRD data in the modelling.

1. University of Edinburgh, local level 1 ethical approval, this was covered by my supervisor who confirmed that the data would not be distributed or identifiable. (See Appendix 2.1)
2. Permission to use the data as in the protocol amendment was applied to from the (Independent Scientific Advisory Committee) ISAC committee of CPRD, who oversee all applications to CPRD for data use – see Appendices 1.4 and 1.5.

6.4.2.2 Scotland

There were several options for measuring the incidence of COPD in Scotland. As incidence is number of new cases of a condition presenting in a population, it was decided to take advantage of the Lothian COPD Cohort database. One of my supervisors (Professor Wild) was a co-applicant on a project to bring together primary care data on COPD for Lothian. Following review of the Chief Scientist's Office (CSO) application for this grant, we approached the Principal Investigator, Dr Rachel Hardie, with a plan to use the cohort data to fill the incidence data field. Dr Leonie Hunter was my liaison researcher at Lothian Health Board.

The Lothian COPD Cohort database brought together a list of all (approximately 7000) COPD patients in the primary care records of the Lothian region of Scotland. Data extraction had been undertaken from the 72 general practices in Lothian that had agreed to data sharing for the original project (overall there are 126 general practices in Lothian). Of these, all practices agreed that their data could be used for my further work.

The following Read Codes (version 2) were extracted from the records to form the COPD Lothian cohort. These Read Codes were selected as those specified in the Quality and Outcomes Framework Business Rules.(HSCIC 2013) Originally the research team for the above project extracted data for spirometry, smoking and co-morbidities; but I only used the data to estimate the incidence of COPD.

Table 6.2 COPD Read Codes used to identify the Lothian COPD Cohort

Read Code	Read Code description
H3...	Chronic obstructive pulmonary disease
H31..	Chronic bronchitis
H310.	Simple chronic bronchitis
H3100	Chronic catarrhal bronchitis
H310z	Simple chronic bronchitis NOS
H311.	Mucopurulent chronic bronchitis
H3110	Purulent chronic bronchitis
H3111	Fetid chronic bronchitis
H311z	Mucopurulent chronic bronchitis NOS
H312.	Obstructive chronic bronchitis
H3120	Chronic asthmatic bronchitis

H3121	Emphysematous bronchitis
H3123	Bronchiolitis obliterans
H312z	Obstructive chronic bronchitis NOS
H313.	Mixed simple and mucopurulent chronic bronchitis
H31y.	Other chronic bronchitis
H31y1	Chronic tracheobronchitis
H31yz	Other chronic bronchitis NOS
H31z.	Chronic bronchitis NOS
H32..	Emphysema
H320.	Chronic bullous emphysema
H3200	Segmental bullous emphysema
H3201	Zonal bullous emphysema
H3202	Giant bullous emphysema
H3203	Bullous emphysema with collapse
H320z	Chronic bullous emphysema NOS
H321.	Panlobular emphysema
H322.	Centrilobular emphysema
H32y.	Other emphysema
H32y0	Acute vesicular emphysema
H32y1	Atrophic (senile) emphysema
H32y2	MacLeod's unilateral emphysema
H32yz	Other emphysema NOS
H32z.	Emphysema NOS
H36..	Mild chronic obstructive pulmonary disease
H37..	Moderate chronic obstructive pulmonary disease
H38..	Severe chronic obstructive pulmonary disease
H39..	Very severe chronic obstructive pulmonary disease
H3y..	Other specified chronic obstructive airways disease
H3y0.	Chronic obstructive pulmonary disease with acute lower respiratory infection
H3y1.	Chronic obstructive pulmonary disease with acute exacerbation, unspecified
H3z..	Chronic obstructive airways disease NOS

The Lothian Cohort database fulfilled the requirement for estimating incidence because it was made up of new diagnoses of COPD with the Read Codes as shown in Table 6.2. The incidence in the Lothian Cohort Database was calculated as follows:

1. Numerator=all new diagnoses of COPD identified using specified Read Codes between 2000 and 2008 grouped by sex, five year age bands and Scottish Index of Multiple Deprivation (SIMD) quintile (based on the 2009 SIMD code);
2. Denominator=practice population at 2010 multiplied by 8 to reflect the 8 years in the cohort, and grouped by sex, 5 year age band and SIMD quintile (again based on the 2009 SIMD code). Inclusion of SIMD allows rates in each deprivation quintile to be considered separately.
3. Incidence rate was calculated by dividing numerator by denominator to give incidence per 100,000 population per year. This was the incidence rate in the Lothian COPD Cohort only.

In order to calculate the incidence rate in Scotland as a whole it was necessary to adjust the estimates because the practices in the original cohort hailed from Lothian and as such were not representative of Scotland as a whole. The Lothian population was, on average, more affluent than the Scottish population and has a different age and sex structure. This correction was done as follows.

1. The first step was to obtain Scottish population numbers for each deprivation quintile. This was done by accessing SIMD online (GovernmentStatistics 2009) which gave the population per 2009 decile and summing every two deciles together to obtain numbers of patients per SIMD quintile for the Scottish population by five year age category and sex.
2. Then incidence by age, sex and SIMD quintile for the Lothian COPD cohort was applied to Scottish population estimates to give estimates of number of COPD patients in each stratum. These were summed over SIMD quintiles to give the total number with COPD in each five year age and sex group.(A)
3. The total Scottish population in each age and sex group was calculated by summing the data over the five SIMD quintiles.(B)
4. Finally, the overall Scottish incidence of COPD (now weighted for national distribution of deprivation) was calculated for each age and sex group by dividing A/B.

6.4.2.2.1 Permissions required for use of the Lothian COPD Cohort Data

The following permissions/waivers had to be obtained before it would be possible to use the Lothian COPD Cohort Data in the modelling:

1. National Health Service Ethical approval waiver, on the understanding that the data were anonymised without risk of identification of participants and so did not require NHS ethical approval. See Appendix 3.1.
2. University of Edinburgh, local level 1 ethical approval, was submitted to the local committee confirming that the data would not be distributed or identifiable. See Appendix 2.1.
3. Privacy Advisory Committee (PAC) at Information Services Division (ISD) approval waiver. This special health board has a data collection and analysis function for the NHS. However, where data are going outside the NHS they approve the transfer and review the risks of identification. In my case, permission had already been obtained for the original research project regarding the Lothian COPD Cohort it was just a matter of obtaining additional approval for additional analysis. See email from Janet Murray, Appendix 3.2 and PAC Application, Appendix 3.3.
4. Caldicott Guardian review, this is a review under the terms of the data protection act. The Caldicott guardian for Lothian Health Board was Dr Allison McCallum and to start with she was very conservative in her advice – however, this was due to a lack of clarity on my part as to the precise use and whereabouts of the data at all times. I wrote again to the Caldicott Guardian to confirm that the only data requested from the Lothian COPD Cohort would be the incidence and that the only data taken to Rotterdam would be the calculated, aggregated Scottish level estimates. See Appendices 4.1 to 4.5. Finally, permission was forthcoming as long as the data were transported on an encrypted laptop.

6.4.3 Prevalence of COPD

6.4.3.1 England

Prevalence of COPD was calculated from the same COPD CPRD GOLD database extract as incidence, see Section 6.3.2.1. Appendix 1.3 contains details of the calculation undertaken by CPRD to provide prevalence. The results of this calculation are shown in table 6.3.

Table 6.3 Prevalence of COPD in England from Clinical Practice Research Datalink 2011.

Prevalence rate per 100,000 cases (95% CI)
--

Age	males	females
35	92.0 (47.6 , 160.7)	152.7 (93.3 , 235.8)
36	124.7 (72.7 , 199.7)	66.9 (30.6 , 126.9)
37	139.4 (85.2 , 215.3)	163.9 (103.9 , 245.9)
38	246.3 (173.4 , 339.4)	183.2 (120.8 , 266.6)
39	244.2 (173.7 , 333.8)	160.7 (104.0 , 237.2)
40	190.9 (130.6 , 269.5)	236.4 (168.1 , 323.2)
41	239.9 (171.4 , 326.7)	361.6 (275.3 , 466.4)
42	312.6 (234.2 , 408.9)	345.9 (262.0 , 448.1)
43	457.9 (362.5 , 570.6)	298.2 (221.3 , 393.1)
44	396.3 (308.4 , 501.6)	488.3 (389.5 , 604.5)
45	412.1 (322.4 , 518.9)	428.7 (337.2 , 537.4)
46	484.1 (387.7 , 597.1)	599.8 (490.6 , 726.1)
47	624.2 (514.4 , 750.4)	618.2 (507.6 , 745.8)
48	597.9 (489.1 , 723.9)	731.5 (609.4 , 870.9)
49	849.7 (717.9 , 998.7)	969.1 (825.6 , 1130.3)
50	787.4 (658.8 , 933.8)	1061.6 (909.3 , 1232.1)
51	985.8 (839.8 , 1149.8)	1075.2 (920.9 , 1247.9)
52	1206.2 (1041.2 , 1389.9)	1408.9 (1227.7 , 1609.4)
53	1502.0 (1316.1 , 1706.8)	1393.4 (1212.6 , 1593.7)
54	1470.1 (1283.9 , 1675.8)	1669.6 (1468.2 , 1890.8)
55	1689.8 (1483.2 , 1917.2)	1751.2 (1541.7 , 1981.3)
56	2019.8 (1791.3 , 2269.4)	2366.7 (2116.4 , 2638.5)
57	2278.0 (2033.7 , 2543.5)	2396.8 (2145.9 , 2669.0)
58	2451.5 (2197.8 , 2726.4)	2681.2 (2415.3 , 2968.4)
59	3040.2 (2756.0 , 3345.7)	2760.1 (2487.5 , 3054.6)
60	2959.3 (2674.3 , 3266.4)	3418.8 (3112.2 , 3747.5)
61	3702.1 (3384.0 , 4041.9)	3472.9 (3168.2 , 3799.0)
62	3947.1 (3622.0 , 4293.5)	3702.7 (3392.1 , 4034.0)
63	4109.5 (3785.3 , 4454.0)	3858.7 (3547.8 , 4189.5)
64	4732.4 (4400.8 , 5082.3)	4274.2 (3963.2 , 4603.1)
65	5121.9 (4760.9 , 5503.0)	4547.6 (4211.9 , 4902.9)
66	6037.8 (5600.6 , 6500.1)	4925.2 (4535.4 , 5339.5)
67	6286.9 (5858.5 , 6738.4)	5673.2 (5272.3 , 6096.5)
68	7180.7 (6695.8 , 7691.4)	5646.0 (5227.4 , 6089.1)
69	7589.7 (7075.3 , 8131.6)	5839.5 (5401.4 , 6303.6)
70	7473.9 (6925.4 , 8054.3)	6810.5 (6301.9 , 7349.2)
71	8378.5 (7792.2 , 8997.3)	6519.3 (6022.8 , 7045.8)
72	8704.3 (8108.5 , 9332.2)	7241.5 (6728.8 , 7783.0)
73	9778.5 (9144.1 , 10445.4)	7168.4 (6647.5 , 7719.3)
74	9398.4 (8759.5 , 10071.6)	7579.6 (7037.4 , 8152.4)
75	9158.3 (8510.7 , 9842.2)	7440.4 (6893.2 , 8019.6)
76	10352.1 (9640.5 , 11102.3)	7706.0 (7143.2 , 8301.3)

77	10050.6 (9337.0 , 10804.2)	8147.5 (7557.1 , 8771.7)
78	10517.6 (9751.2 , 11328.2)	7821.2 (7215.7 , 8464.0)
79	10317.8 (9552.4 , 11128.3)	7706.6 (7113.7 , 8335.7)
80	10994.9 (10173.1 , 11865.3)	8286.5 (7661.2 , 8949.3)
81	11872.4 (10997.3 , 12798.6)	8541.3 (7900.6 , 9220.2)
82	11408.1 (10505.3 , 12367.8)	8156.7 (7509.0 , 8845.3)
83	12755.2 (11739.0 , 13835.9)	8489.7 (7807.0 , 9216.2)
84	12132.6 (11074.6 , 13264.5)	7477.9 (6807.2 , 8196.8)
85	11844.6 (10769.5 , 12997.9)	8149.2 (7422.2 , 8928.2)

6.4.3.1 Scotland

It was decided to use data from the Information Services Division's (ISD) Practice Team Information (PTI) database to estimate COPD prevalence for Scotland. The source of this database is described in Chapter 3 Table 3.1. The COPD Read Code Group is a standard grouping used by ISD Scotland for producing statistics about COPD. It contains the following codes for COPD and related conditions. Unfortunately, although these codes substantially overlap, they are not identical to those used for the Lothian COPD Cohort Database.

Table 6.4 COPD Read Code Grouping for Practice Team Information Database

Read Code	Read Code description
66YB.	Chronic obstructive pulmonary disease monitoring
66YB0	Chronic obstructive pulmonary disease 3 monthly review
66YB1	Chronic obstructive pulmonary disease 6 monthly review
66Yd.	COPD accident and emergency attendance since last visit
66YD.	Chronic obstructive pulmonary disease monitoring due
66Ye.	Emergency COPD admission since last appointment
66Yf.	Number of COPD exacerbations in past year
66Yg.	Chronic obstructive pulmonary disease disturbs sleep
66Yh.	Chronic obstructive pulmonary disease does not disturb sleep
66Yi.	Multiple COPD emergency hospital admissions
66YI.	COPD self-management plan given
66YL.	Chronic obstructive pulmonary disease follow-up
66YM.	Chronic obstructive pulmonary disease annual review

66YS.	Chronic obstructive pulmonary disease monitoring by nurse
66YT.	Chronic obstructive pulmonary disease monitoring by doctor
8BMa0	Chronic obstructive pulmonary disease medication optimisation
8CeD.	Preferred place of care for next exacerbation of COPD
8CMV.	Has chronic obstructive pulmonary disease care plan
8CR1.	Chronic obstructive pulmonary disease clinical management plan
H0614	Obliterating fibrous bronchiolitis
H3...	Chronic obstructive pulmonary disease
H30..	Bronchitis unspecified
H300.	Tracheobronchitis NOS
H301.	Laryngotracheobronchitis
H302.	Wheezy bronchitis
H30z.	Bronchitis NOS
H31..	Chronic bronchitis
H310.	Simple chronic bronchitis
H3100	Chronic catarrhal bronchitis
H3101	Smokers' cough
H310z	Simple chronic bronchitis NOS
H311.	Mucopurulent chronic bronchitis
H3110	Purulent chronic bronchitis
H3111	Fetid chronic bronchitis
H311z	Mucopurulent chronic bronchitis NOS
H312.	Obstructive chronic bronchitis
H3120	Chronic asthmatic bronchitis
H3121	Emphysematous bronchitis
H3122	Acute exacerbation of chronic obstructive airways disease
H3123	Bronchiolitis obliterans
H312z	Obstructive chronic bronchitis NOS
H313.	Mixed simple and mucopurulent chronic bronchitis
H31y.	Other chronic bronchitis
H31y0	Chronic tracheitis
H31y1	Chronic tracheobronchitis

H31yz	Other chronic bronchitis NOS
H31z.	Chronic bronchitis NOS
H32..	Emphysema
H320.	Chronic bullous emphysema
H3200	Segmental bullous emphysema
H3201	Zonal bullous emphysema
H3202	Giant bullous emphysema
H3203	Bullous emphysema with collapse
H320z	Chronic bullous emphysema NOS
H321.	Panlobular emphysema
H322.	Centrilobular emphysema
H32y.	Other emphysema
H32y0	Acute vesicular emphysema
H32y1	Atrophic (senile) emphysema
H32y2	MacLeod's unilateral emphysema
H32yz	Other emphysema NOS
H32z.	Emphysema NOS
H36..	Mild chronic obstructive pulmonary disease
H37..	Moderate chronic obstructive pulmonary disease
H38..	Severe chronic obstructive pulmonary disease
H39..	Very severe chronic obstructive pulmonary disease
H3A..	End stage chronic obstructive airways disease
H3y..	Other specified chronic obstructive airways disease
H3y0.	Chronic obstructive pulmonary disease with acute lower respiratory infection
H3y1.	Chronic obstructive pulmonary disease with acute exacerbation, unspecified
H3z..	Chronic obstructive airways disease NOS
Hyu3.	[X]Chronic lower respiratory diseases
Hyu30	[X]Other emphysema
Hyu31	[X]Other specified chronic obstructive pulmonary disease

The numbers of people consulting a GP or practice nurse at least once for a COPD Read Code in the time period 1 January 2011 to 31 March 2012 were extracted from the database. This number of visits was thought to be a reasonable proxy for prevalence because, as a

result of the Quality and Outcomes Framework (QOF) in the General Practice General Medical Services Contract 2003: people with COPD were put on a register by their General Practitioner and invited to attend at least once a year for a review.(GMC 2003) Not all the visits occur within a year period, however, and so a 15 month period was used. The number of visits was stratified by SIMD deprivation for the PTI population then mapped onto the Scottish population according to SIMD. This was to enable a correction for the different level of deprivation among the 6% of the Scottish population that fall within the PTI and the level of deprivation in the total Scottish population. Finally, the populations were aggregated into sex and five year age groupings. This was done by summing the numbers of cases across each five year age grouping for each sex and then dividing this by the sex specific population for each 5 year age grouping to produce prevalence for each five year age group. The original intention had been to use sex and one year age categories, however, the numbers of people varied widely from year to year and could have resulted in unwarranted and unrepresentative fluctuations when modelled and so it was decided to use five year intervals.

Table 6.5 COPD prevalence for Scotland

Age group (years)	Male	Female
0-4	0.002	0.002
5-9	0.000	0.001
10-14	0.001	0.000
15-19	0.000	0.001
20-24	0.000	0.000
25-29	0.001	0.001
30-34	0.001	0.002
35-39	0.002	0.002
40-44	0.003	0.004
45-49	0.008	0.010
50-54	0.016	0.021

55-59	0.028	0.030
60-64	0.049	0.053
65-69	0.057	0.070
70-74	0.094	0.090
75-79	0.097	0.084
80-84	0.094	0.080
85+	0.073	0.046

6.4.4 Proportions of the population who smoke

6.4.4.1 England

The General Lifestyle Survey for England 2011(ONS 2011) describes the proportion of smokers in fairly broad age categories. However, when compared with the data for Scotland and the original Dutch Model data it appears that the smoking prevalence in each age group were comparable and therefore I accepted that the General Lifestyle Survey for England estimates were valid. I had initially looked at using more granular, one year age group data from the Smoking Toolkit Study (West and Brown 2013), however, in older age brackets it appeared that too few people had completed questionnaires and there were some unusual patterns in the data (e.g. 50% or greater rates at every age over 75) therefore, I could not accept that the data were valid and decided not to use this source.

Table 6.6 England smoking proportions from General Lifestyle Survey Data 2011

From General Lifestyle Survey				
Smoking prevalence in England in 2011 %				
	Male		Female	
age	current smoker	former smoker	current smoker	former smoker
16-19	18	2	19	4
20-24	30	10	28	7

25-35	26	15	21	18
35-49	25	22	23	22
50-59	20	27	18	24
60 and over	14	45	12	30

Never smoker proportions were calculated as 100%- (current smoker + former smoker).

For younger smokers, the Health Survey for England 2010(HSE 2010) was the most recently published source of prevalence estimates and this was retrieved via the Economic and Social Data Service online portal.(ESRC 2011) Only one estimate for the age group for which data had been missing was available, for children age 8-15 years, but sex specific estimates were not available. These were used to estimate the percentage of never, current and former smokers for the age group 10-14 years.

Table 6.7 Health Survey for England smoking prevalence among 8-15 year olds

Health Survey for England 2010 from Economic and Social Data Service (ESRC 2011)		
Male and female combined		
Never	Current	Occasional
0.974	0.017	0.002

6.4.4.2 Scotland

The data for Scotland in terms of the proportion of the population by sex and age who were smokers, former smokers and who have never smoked was another important input to the model. These data were published in the Scottish Health Survey 2011 online report, page 119.(Bradshaw, Bromley et al. 2011) As the age groups were different to the age groupings in the model I contacted the Scottish Health Survey and Rosalia Munoz-Arroyo responded with the data in the correct age groupings down to age 16 for 2011.(Munoz-Arroyo 2011)

Table 6.8 Scottish Health Survey 2011, proportion of smokers by age and sex

Scottish Health Survey 2011(Bradshaw, Bromley et al. 2011)						
	Males			Females		
Age group (years)	Never Smoker	Current smoker	Former Smoker	Never Smoker	Current smoker	Former Smoker
0-4	1	0	0	1	0	0
5-9	1	0	0	1	0	0
10-14	0.9	0.03	0.07	0.79	0.03	0.07
15-19	0.8	0.19	0.02	0.81	0.15	0.04
20-24	0.65	0.32	0.03	0.57	0.35	0.08
25-29	0.55	0.37	0.07	0.63	0.26	0.1
30-34	0.53	0.31	0.16	0.59	0.24	0.17
35-39	0.55	0.31	0.14	0.56	0.28	0.16
40-44	0.52	0.26	0.21	0.6	0.22	0.18
45-49	0.54	0.23	0.23	0.6	0.25	0.15
50-54	0.45	0.27	0.28	0.54	0.25	0.21
55-59	0.52	0.19	0.29	0.46	0.3	0.25
60-64	0.46	0.25	0.28	0.48	0.24	0.28
65-69	0.35	0.17	0.48	0.5	0.18	0.31
70-74	0.43	0.12	0.45	0.55	0.11	0.34
75-79	0.41	0.07	0.52	0.56	0.09	0.35
80-84	0.39	0.11	0.5	0.62	0.07	0.31
85+	0.49	0.07	0.45	0.67	0.06	0.27

For the 10-14 age group, data came from the Scottish Schools Adolescent Lifestyle and Substance Use Survey (SALSUS) which was accessed online, these data were available for 13 year olds by sex in 2010 when the most recent survey was undertaken.(Black, Eunson et al. 2010)

6.4.5 Start smoking proportions

It was desirable to find a dataset with age and sex related proportions of people starting smoking in England and Scotland in any one year. Unfortunately, I was unable to find a current dataset. However, the Avon Longitudinal Cohort Study in 1991/2 collected this information on their cohort in Bristol and so I decided I would have to extrapolate these proportions to both the English and Scottish populations (BristolUniversity 1991). I did not standardise for deprivation as it was unclear which measure could be taken as representative for the cohort to standardise to the total populations. In addition, as the cohort was small relative to the total populations, any errors during standardisation would be amplified to the total population.

6.4.6 Stop smoking proportions

These were calculated from smoking quit proportions as collected by the Smoking Toolkit Survey, University College London.(West and Brown 2013) Dr Jamie Brown and Professor Robert West kindly shared their data from the questionnaires. The numbers quitting were stratified by one year age intervals and sex. Then these data were converted into a proportion using the denominator of number of people in each age/sex strata that had been surveyed. Such a proportion can effectively be used as a probability of stopping smoking at that age.

Although these data were collected in England, I decided that because they were current data from a large sample and so likely to have high validity, they would be representative enough for Scotland as well.

6.4.7 Restart smoking proportions

No source was found of current restart proportions for smoking quitters in Scotland or England. Therefore, the original data used in the Dutch Model which came from a survey conducted annually in the Netherlands(STIVORO 2003), were used in the English and Scottish projections. It was thought that this would be reasonably valid as overall smoking prevalences for the three countries are comparable. 37% of males and 31% of females aged

25-64 smoked in 2012 in the Netherlands, compared with 33% of UK males and 29% of UK females in this age group).(Zatonski, Przewozniak et al. 2012)

6.4.8 Relative risks of smokers and non-smokers

The relative risks for smokers to receive a diagnosis of COPD were the same risks as were used in the original model. These came from a report by the US Surgeon General(SurgeonGeneral 2004) and a Dutch report(Van Oers 2002). I thought that such rates could reasonably be transferred from the Dutch and American contexts because the risk of getting COPD is intrinsic to the biological mechanisms of the development of COPD from cigarette smoking and therefore does not change with geography. (There may be underlying genetic susceptibility issues in different populations in emerging research – however, as yet, there is insufficient data available to vary this input on this basis.)

6.4.9 Lung function decline rate

The Dutch Model included a lung function decline rate which had been calculated from a random effects model using data from 5000 Dutch COPD patients. The lung function decline rate was set at the same rate as had been calculated for a previous incarnation of the Dutch Model.(Hoogendoorn, Rutten-van Molken et al. 2005) The lung function decline rate was then varied in the sensitivity analyses to see how sensitive the model is to changes in this parameter.

6.4.10 COPD-related maintenance costs

The Dutch Model included a table of costs from a previous Dutch study.(Hoogendoorn, Feenstra et al. 2006) However, the numbers in this table show some unusual patterns.

Table 6.9 Original COPD direct cost input data for modelling (Euros) from Dutch Model report (Hoogendoorn, Rutten-van Molken et al. 2010)

	Mild COPD	Moderate COPD	Severe COPD	Very severe COPD
Males:				
45-49	47	58	65	96
50-54	32	39	44	65
55-59	31	39	43	65
60-64	72	89	99	148
65-69	135	167	187	277
70-74	197	245	273	406
75-79	346	430	480	712
80-84	344	428	477	708
85+	659	820	913	1356
Females:				
45-49	220	273	305	452
50-54	270	335	374	555
55-59	261	324	361	536
60-64	263	327	364	541
65-69	326	405	452	671
70-74	292	364	405	602
75-79	371	462	514	764
80-84	407	507	564	838
85+	831	1034	1152	1711

These are the costs for maintenance care of COPD in the community, so this includes GP and nurse visits and drugs. The costs of exacerbations including hospital admissions were modelled separately.

Firstly, it is not clear why the costs for women were so much greater than the costs for men. It is also not clear why the costs varied four-fold for women and over ten-fold for men for the different age groups when one compares caring for the youngest age group compared with the oldest age group.

The Rotterdam team encouraged me to find equivalent data by age and sex for the maintenance costs of COPD in Scotland and England. The “best estimate” scenario for generating these costs would be to undertake a micro-costing exercise. This is a skilled undertaking involving first the identification of typical costs for each age and sex band within the population. For example, the costs of drugs and GP visits and nurse visits would

need to be identified. Then the typical drug doses/ prescription numbers and the typical numbers of visits per time period would need to be counted. This often involves the use of detailed cohort study data. Indeed, one of the Lothian COPD Cohort's future aims is to pursue a grant for a micro-costing exercise of this nature. The results of this exercise were not to be available until beyond the submission date for this PhD, and so it would not be feasible to undertake a micro-costing exercise to generate COPD maintenance cost data as part of the PhD.

There remained several alternatives for generating these cost data. One option was to update the costs from the Dutch study using the retail price index. However, as a result of the problems discussed above it seemed unlikely that the distribution of these Dutch data could be transferred to an equivalent Scottish or English population.

Another option was to use the available costs from an Audit Scotland report to provide Scottish data from 2004/5 without the correction for age and sex and to use the upper and lower estimates for the different regions in Scotland as a guide for setting boundaries for sensitivity analysis.(Matthew 2007) However, the Audit Scotland costs were not divided by age or by severity and so did not provide useful variability for the modelling process.

I reviewed NICE technology appraisal cost-effectiveness analysis for the new drug Roflumilast.(NICE 2012) However, this only modelled severe and very severe COPD and so the costs for more moderate disease states are not included.

Finally, I consulted a source of micro-costing evidence from the cost utility analysis of Indacaterol.(Price, Asukai et al. 2013) This is very detailed evidence giving median and uncertainty range per item. However, I required the total costs per average patient and so I contacted the authors of the paper to see if they could share their breakdown of total maintenance costs per severity grouping and possibly by age. Unfortunately they had no information on costs by age; however, the cost data per item was published along with average annual usage statistics per severity level of COPD that came either from analysis of a large COPD patient database or from expert opinion via a Delphi-like process. This allowed me to multiply cost by usage per average patient of each severity level and total up a cost per severity level, as shown in Table 6.9.

Table 6.10 Unit costing from indacaterol study data converted to total cost per severity level (Price, Asukai et al. 2013)

Item	Cost £	Mild use	Cost mild £	Mod use	Cost mod £	Severe use	Cost severe £	Very severe use	Cost very severe £
flu vaccine	14.2	0.73	10.366	0.73	10.366	0.73	10.366	0.73	10.366
pneumovacc	46.75	0.69	32.2575	0.69	32.2575	0.69	32.2575	0.69	32.2575
Theophylline (no. of scripts)	3.43	0.26	0.8918	0.32	1.0976	0.73	2.5039	1.63	5.5909
Mucolytics	21.85	0.35	7.6475	0.4	8.74	0.8	17.48	2.05	44.7925
oral corticosteroids	8.79	0.88	7.7352	0.96	8.4384	1.7	14.943	2.7	23.733
short acting beta agonists	5.98	3.74	22.3652	4.65	27.807	6.87	41.0826	9.78	58.4844
Inhaled corticosteroids	12.12	0.89	10.7868	0.81	9.8172	0.71	8.6052	0.62	7.5144
short acting antimuscarinics	10.24	0.59	6.0416	0.65	6.656	0.91	9.3184	1.19	12.1856
leukotriene receptor antagonists	31.77	0	0	0	0	0.37	11.7549	0	0
Pulmonary rehabilitation	1017.27	0.02	20.3454	0.03	30.5181	0.06	61.0362	0.09	91.5543
GP visits	36	15.05	541.8	15.76	567.36	16.2	583.2	16.16	581.76
Outpatient respiratory specialist visit	134.61	0	0	0	0	2	269.22	4	538.44
Spirometry	51.38	1	51.38	2	102.76	2	102.76	4	205.52
Total			711.617		805.8178		1164.528		1612.199

These costs did not include costs of tiotropium as that drug was also investigated in the study or of home oxygen, as the study had found such costs hard to source.(Price, Asukai et al. 2013)

In order to validate these costs I could triangulate with the Audit Scotland costs mentioned above. This report put the range of costs for providing services to patients with COPD at £988 in NHS Tayside and £1222 in NHS Grampian. However, these Audit Scotland costs were largely made up from inpatient costs which were not investigated in the indacaterol study except for exacerbations. The exacerbations costs are considered later in this section.

Another source of costing evidence that came to my attention was a cost-utility analysis conducted for tiotropium.(Hettle, Wouters et al. 2012) Here, costs for each disease severity state had been estimated by a Delphi panel consisting of four general practitioners and four secondary care consultants. Unit costs were obtained from the Personal Social Services Research Unit and NHS National Tariff 2009/2010 (as are standard to be used for micro-costing) and usage was estimated by the Delphi panel. The problem arose that there were no costs for the mild COPD category.

Table 6.11 Costs of managing COPD in the UK 2011 from tiotropium study.(Hettle, Wouters et al. 2012)

GOLD Stage	Monthly cost of managing COPD in the UK 2011 (£)	Annual cost of managing COPD in the UK 2011(£)
Moderate COPD	39.50	474.00
Severe COPD	82.88	994.56
Very Severe COPD	136.87	1642.44

This costing also did not include the costs of exacerbations. This costing can therefore be used to validate the indacaterol costing for very severe COPD because the cost estimates from the two studies are similar. However, the indacaterol costings for mild, moderate and severe COPD are not so comparable. It seems most likely that the tiotropium study costs are an underestimate because they were produced by expert opinion and the indacaterol study used database evidence.

In the end, I decided to use the indacaterol calculated costs per year for each severity level of COPD for the maintenance costs of COPD because this was the only source with an estimate for all four severity stages of COPD and was also the most transparent formulation of costs.

In terms of costs pertaining to exacerbations I used the tiotropium study’s Delphi panel’s estimates for exacerbations in Scotland and England.(Hettle, Wouters et al. 2012) They calculated that the cost of a moderate exacerbation would be £118 and the cost of a severe exacerbation, i.e. one requiring hospital admission, was £3726 in England per event and £3329 in Scotland per event.

6.4.11 Mean utility scores by COPD severity stage

Utility scores are a way of estimating the reduced quality of life as a result of having COPD as described in Chapter 4. A person with a utility score of 1.0 has perfect health. Where utility scores decrease, this reflects worsening health. Utility scores are used to calculate the numbers of QALYs lost as a result of people having COPD and thus the burden of this disease. The utility scores for COPD in the original Dutch Model were used for these calculations in the English and Scottish versions as no published UK-based utility weightings could be found. (Hoogendoorn, Rutten-van Molken et al. 2010)

Table 6.12 Utility scores for Dutch Model (Hoogendoorn, Rutten-van Molken et al. 2010)

GOLD Stage	Mean Utility Score (SD)
Mild COPD	0.8971(0.1117)
Moderate COPD	0.7551(0.2747)
Severe COPD	0.7481(0.2991)
Very severe COPD	0.5493(0.3129)

6.4.12 Excess mortality

The Dutch Model used a slightly complicated manipulation to calculate the excess mortality. In the model all-cause mortality among COPD patients was divided into “excess mortality” and “mortality from other causes”, where “excess mortality” was defined as the difference in mortality between COPD patients and the general population which includes the increased risk of dying from other smoking-related diseases.

Definitions:

M_1 = All cause mortality of COPD patients

M_0 = All cause mortality in those without COPD

R= relative risk of death from COPD

p=Prevalence

$R=M_1/M_0$ therefore $M_1= M_0R$

Total population mortality= mortality in those with COPD + mortality without COPD

$$= M_1p+M_0(1-p)$$

Substituting for M_1 $=M_0Rp+M_0(1-p)$

Factorising for M_0 $=M_0\{Rp+(1-p)\}$

Rearranging $M_0=\text{total population mortality}/\{Rp+(1-p)\}$

Then “Excess mortality” $=(R-1)*M_0$ where M_0 is defined as above.

In the 2005 Dutch Model, in order to calculate this excess mortality, data from the original GPRD (UK data) were used with Poisson regression to model age and sex related relative-risks for COPD. These were then substituted into the formulae with the Dutch total mortality from National Registers to obtain Dutch excess mortality.

A similar calculation had been undertaken in 2007 to obtain UK COPD excess mortality for the DYNAMO-HIA project.(Lhachimi, Nusselder et al. 2012) These excess mortality figures were obtained to be used as a sensitivity analysis.

However, I also decided to generate de novo estimates of the relative risks of death from COPD by age and sex using the available CPRD case-control database. This was because the original relative risks of death in the original model were over 10 years old and I believed that an update would more closely reflect the current risk of death from COPD given current treatments. A time to event analysis was undertaken using Cox-regression to generate these relative risks then the excess mortality by age and sex was calculated by substitution again into the above formulae with the prevalence data from CPRD (England) and PTI (Scotland), respectively.

The CPRD case-control database was derived from CPRD data which included “up to standard follow up” for at least five years for inclusion in the cohort. COPD patients had been identified using relevant Read Codes (see Appendix 1.1 original protocol for CPRD COPD case-control database for a list of the relevant Read Codes). Then this file had been linked by a unique patient identifier number to a mortality file which included the date of death. The person-time each individual contributed to the cohort was calculated and an analysis conducted from time of inclusion to time of either event or censoring (when data ceased to be submitted to CPRD for that specific practice).

The results for the coefficients in the modelling equation are shown in Table 6.12. The parameters were age (AGEA), sex (GENDERA) and a marker as to whether the patient had COPD or not and interaction terms (AGEGENDER, AGE COPD, GENDER COPD, GENDER COPD AGE). The outcome was death over the total time for which data were available. The coefficient with the most influence on whether or not a patient died was whether the patient had a diagnosis of COPD (odds ratio 9.4).

Table 6.13 Parameters for Cox Regression Model of COPD patient survival

	B	SE	Wald	Df	Sig.	Exp(B)
GENDERA	-.894	.228	15.352	1	.000	.409
AGEA	.054	.002	835.708	1	.000	1.055
AGEGENDER	.008	.003	8.946	1	.003	1.008
AGE COPD	-.018	.002	62.812	1	.000	.982
GENDER COPD	.458	.277	2.739	1	.098	1.581
GENDER COPD AGE	-.005	.003	2.231	1	.135	.995
Marker	2.245	.185	146.713	1	.000	9.441

These coefficients were then used to calculate relative risks for patients in the middle of each five year age bracket by multiplying out the model for every x. Where x was a patient of a specific age and gender (gender=1 for females and gender =0 for males) and COPD marker (1=COPD, 0=no COPD).

$$\text{Hazard at age and sex } x = e^{(B_{\text{Marker}x} * B_{\text{GENDER}Ax} * B_{\text{AGE}Ax} * B_{\text{AGEGENDER}x} * B_{\text{AGE COPD}x} * B_{\text{GENDER COPD}x} * B_{\text{GENDER COPD AGE}})}$$

Relative risk was calculated by finding the risk (or hazard) at age and sex x with COPD then dividing this by the risk at age and sex x without COPD.

Table 6.14 Age specific relative risk of dying per year if patient has COPD compared to an age and sex matched control as obtained from CPRD data extract

Age	Male	female
	relative risk	relative risk
40-44	4.39	5.62
45-49	4.01	5.01
50-54	3.67	4.46
55-59	3.35	3.98
60-64	3.06	3.54
65-69	2.80	3.16
70-74	2.56	2.82
75-79	2.34	2.51
80-84	2.14	2.24
85+	1.95	1.99

Then these relative risks were combined with the total mortality (all-cause mortality) from the ONS(England) for 2011 and the GROS(Scotland) for 2011 according to the above formulae and each 1 year age and sex interval in order to generate the excess mortality. I used the modelled excess mortality from these smoothed relative risks for England and Scotland base cases (i.e. run of the model without sensitivity analysis). Then the UK modelled excess mortality as used in the DYNAMO-HIA study as a sensitivity analysis.

6.5 Strengths and limitations of datasets

Table 6.15 Strengths and limitations of datasets

Dataset	Strengths	Limitations
Clinical Practice Research Datalink(CPRD 2011)	Largest available dataset from primary care in England	Routine data so quite “messy” (some miscodes, some terms included in free text so not captured by codes) Also see Chapter 3 section on validity for the representativeness of this dataset compared to the national population. Charged for work done to produce estimates.

Lothian COPD Cohort	COPD specific dataset, only available incidence data, not charged for use.	Also routine data. Only derived from Lothian, so needed SIMD corrections to produce estimates for whole of Scotland.
Scottish Practice Team Information Database	Able to give COPD estimates specific to Scotland, cheaper than CPRD.	Also routine data. Uses consultation rates as proxies for prevalence. Does not include whole Scottish population (only 6% representative sample). Charged for work done to produce estimates.
The General Lifestyle Survey for England 2011(ONS 2011)	Deliberate sampling of representative sample, largest survey available.	Research quality data, therefore high quality, 15,000 interviews, therefore small sample compared to population.
Avon Longitudinal Study of Parents and Children(BristolUniversity 1991)	Only source of start smoking rates available in UK for males and females (parents of the children who were born).	Research grade data. Based on 14,541 pregnancies, so relatively small sample compared with national population.
Smoking Toolkit Survey(West and Brown 2013)	Only source of up to date quit smoking rates in UK.	Large sample size (139000 smokers and non smokers) is still a relatively small section of the national population.
Scottish Health Survey 2011(Bradshaw, Bromley et al. 2011)	Largest survey available in Scotland, data easier to interpret than Primary Care Record data.	Research grade data. 7245 Adult interviews is a small sample compared with the whole population of Scotland.

6.6 Conclusion to Chapter 6

This chapter has described the selected Dutch Model. It is a Markov multi-state model and models a population cohort progressing through their lifetimes with and without COPD. This

description was followed by an initial overview and then a detailed description of the data inputs to the Dutch Model required for English and Scottish versions of the model. The governance and permissions required for specific datasets were described.

The next chapter contains the results of the modelling process and the sensitivity analyses.

Chapter 7: Modelling process and results

7.1 Introduction

The previous chapter introduced the Dutch Model and gave details of the data inputs required to run this model for England and Scotland. This chapter will describe the modelling process then detail the results in terms of population, costs and mortality for the deterministically modelled base case scenario and the sensitivity analyses. Then the uncertainty intervals for the results will be discussed and presented.

7.2 Modelling process

The Dutch Model was set to run in Wolfram Mathematica version 9 (Wolfram 2013) on a Windows 7 laptop. This is a mathematical programme that enables computational analysis and modelling with large datasets. The code for the Dutch Model was loaded from the Rotterdam team's source onto the laptop. Then the programme was instructed to seek the relevant input data files from the laptop's directories.

The base case model was run using a deterministic process with the data inputs as detailed in Chapter 6. The model was run once with English data for the English base case and once with Scottish data for the Scottish base case.

Sensitivity analyses were then performed. These enabled me to see the impact of making changes to the input of the model on the output of the model. It is important to do this in order to understand the sensitivity of the outputs to changes in the inputs as this sensitivity is an effect of the Model's intrinsic structure. The 10 sensitivity analyses shown in Table 7.1 were run for both England and Scotland.

These analyses were all run as one way sensitivity analyses where one input parameter was changed in one direction (i.e. either increase or decrease) at a time.

Table 7.1 One way sensitivity analyses for Dutch Model projections from 2011 to 2030, all run for both England and Scotland.

Sensitivity analysis	Parameter	Change
1	Prevalence	Increased by 10%
2	Prevalence	Decreased by 10%
3	Incidence	Increased by 10%
4	Incidence	Decreased by 10%
5	Mean annual decline in lung function	Increased by 10%
6	Mean annual decline in lung function	Decreased by 10%
7	Severity distribution of baseline prevalence	More severe cases
8	Severity distribution of baseline prevalence	Less severe cases
9	Use DYNAMO-HIA data for excess mortality	N/A
10	Smoking quit rate increased to 10% at each age	N/A

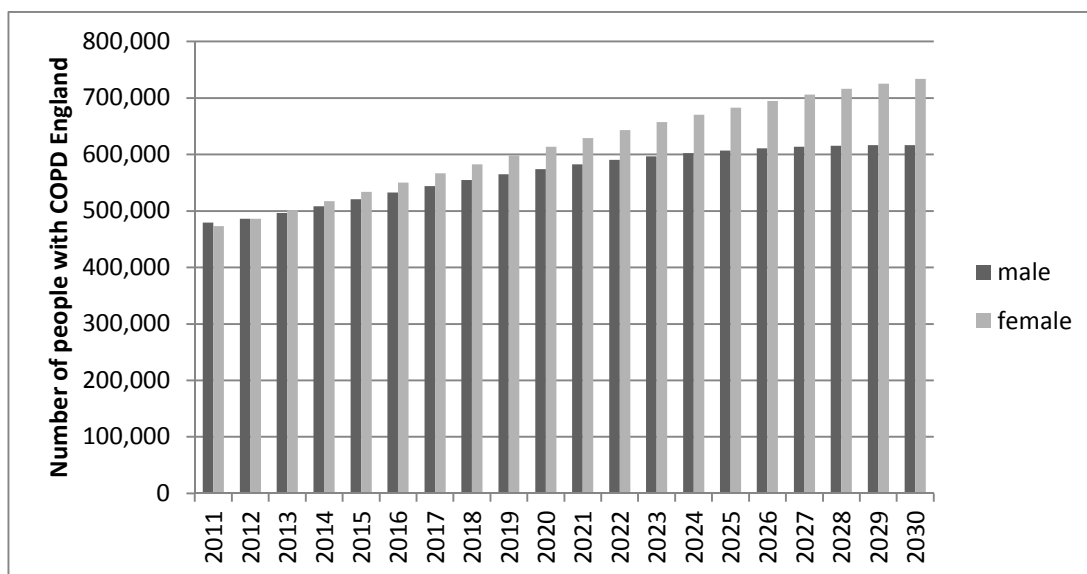
7.3 Results of modelling base case scenario

7.3.1 England: COPD population results

The population with COPD was calculated in life-years. The number of life-years in any one year can be taken as the mid-year (June) number of people with COPD for that year. The populations for England and Scotland were given to the nearest thousand.

Using the Dutch Model, I estimated that in 2011 there were a total of 952,000 people with diagnosed COPD in England (480,000 males and 473,000 females, 274,000 of whom were current smokers). The first year calculation (2011) was a straightforward demographic calculation where the prevalences for COPD were mapped onto the population in each one year age group for males and females. The modelling element only becomes functional in years beyond the first year where increases in COPD population are dictated by smoking risk and COPD incidence year on year. The COPD population increased to 1.35 million people by 2030 of whom there were 616,000 males and 734,000 females (419,000 smokers), see Figure 7.1 for the modelled population over this period.

Figure 7.1 Modelled population of males and females with COPD in England



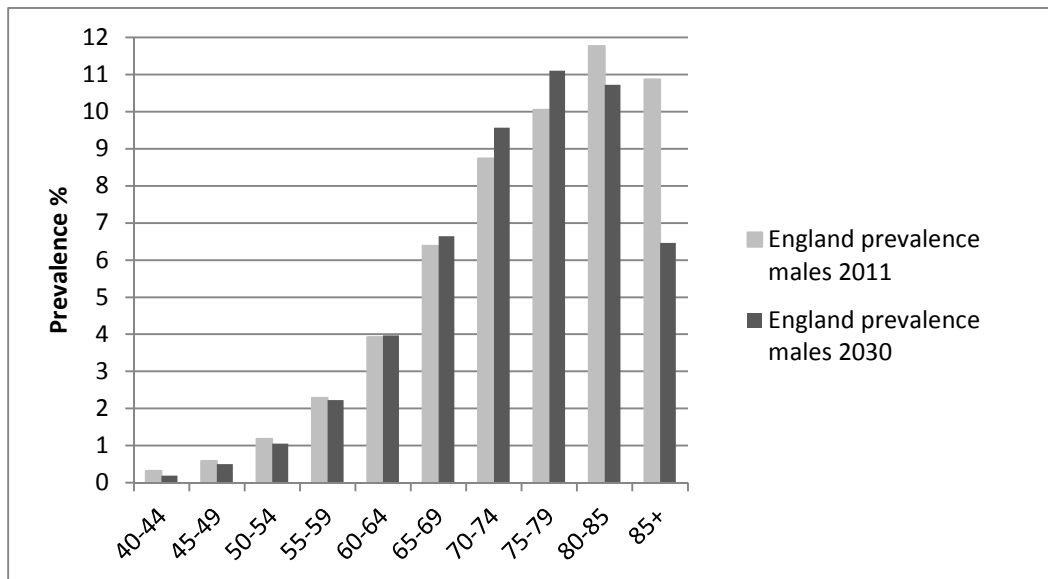
7.3.2 England: overall prevalence results

The model projected that the overall prevalence of diagnosed COPD among over 40 year olds would rise from 3.6% in 2011 to 4.5% in 2030. This involved an increase from 3.8% in 2011 among males over 40 to 4.3% in 2030. The change in COPD prevalence among females over 40 years of age was from 3.4% in 2011 to 4.6% in 2030. This represented a 0.5% increase among males and a 1.2% increase among females in the over 40 years age group.

7.3.3 England: age-related prevalence

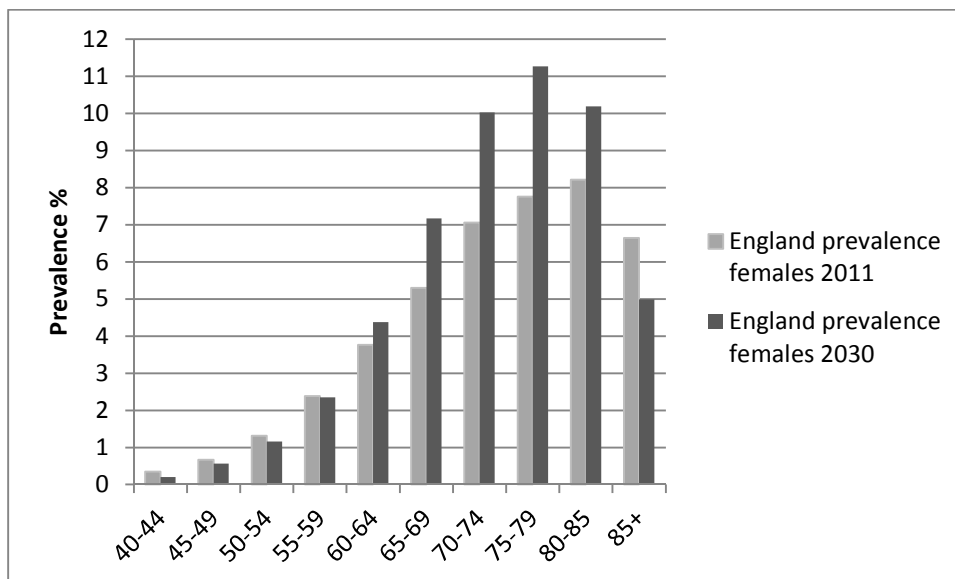
Figures 7.2 and 7.3 show the age-specific prevalence, as modelled by the Dutch Model, of males and females in England and how it is anticipated to change between 2011 and 2030. Note that the y-axis is comparable across both figures and that the units are percentage of population with COPD in that age group.

Figure 7.2 Age-specific prevalence of males with COPD in England



It can be seen that in most age categories in males the age-specific prevalence was relatively stable. In fact, in the oldest age category there may be a reduction in prevalence of COPD between 2011 and 2030.

Figure 7.3 Age-specific prevalence of females with COPD in England



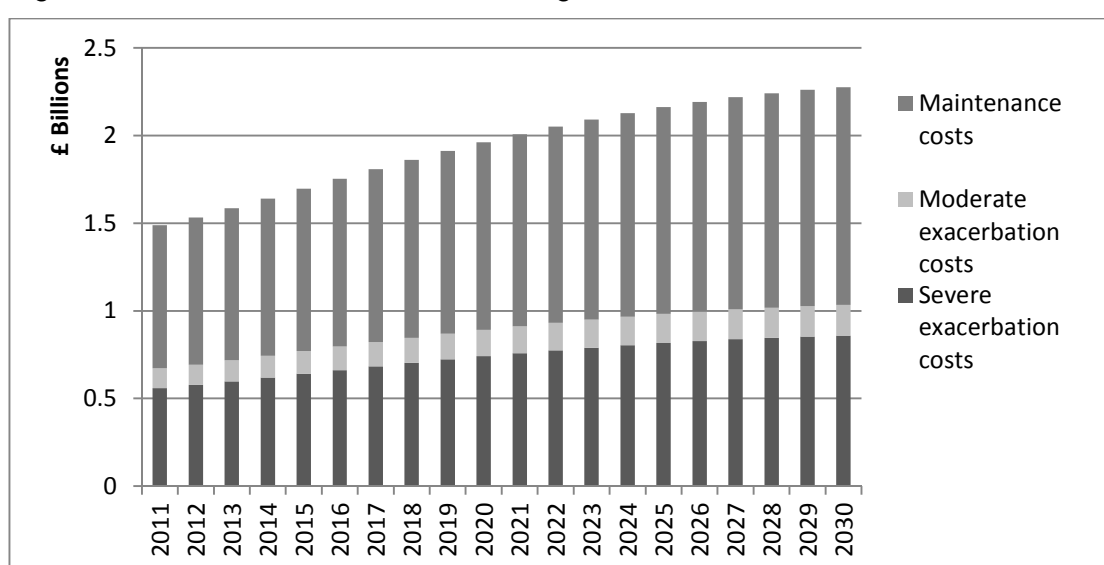
In England in 2011, the age specific prevalences were mostly lower among females than among males (see Figure 7.2). However, the prevalences were projected to be comparable with that seen in males by 2030.

7.3.4 England: direct healthcare costs from COPD

The total annual costs of COPD in England were projected to increase from £1.489 billion in 2011 to £2.275 billion in 2030. These costs are in terms of 2011 costs without the application of any economic trends (no annual increase applied for inflation).

These costs were made up of maintenance costs: that increased from £818 million in 2011 to £1.241 billion in 2030, and costs of caring for moderate and severe COPD exacerbations, as shown in Figure 7.4. A moderate exacerbation involved contact with the doctor for deterioration in symptoms, a severe exacerbation involved a hospital admission by definition.

Figure 7.4 Modelled costs for COPD in England

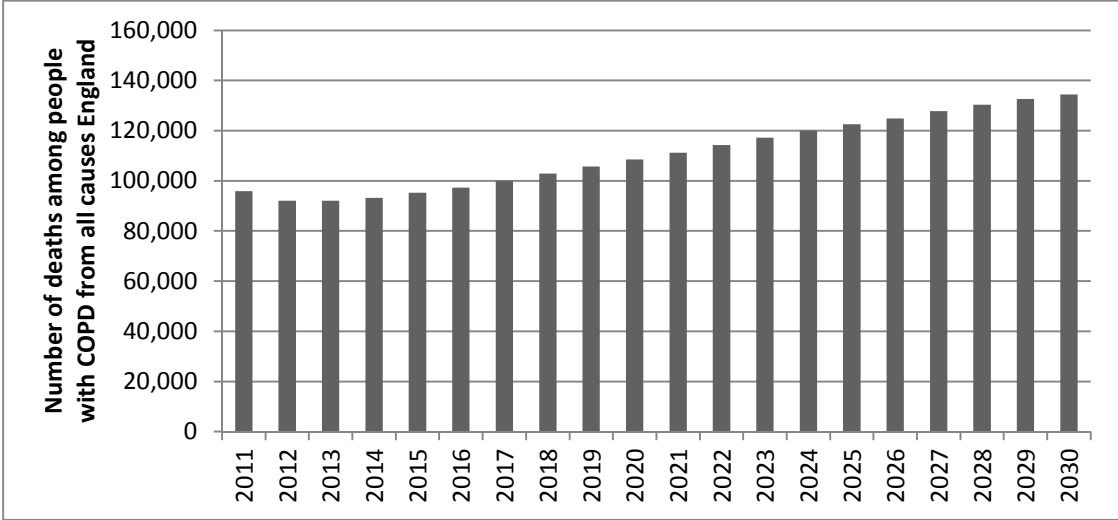


7.3.5 England: number of deaths

The number of deaths among people with COPD was modelled to be 96,000 in 2011, then to fall slightly over the next few years as the model stabilised, then gradually increase again to 134,000 people with COPD in 2030, see Figure 7.5. These reflect estimates of all-cause mortality among people with COPD and therefore these figures cannot be equated with COPD recorded on the death certificate. These estimates come from the increased relative risk of dying with COPD due to COPD or other diseases which was extracted by Cox regression from the CPRD database and used to calculate the excess mortality. The excess mortality was then combined with the prevalence and total mortality to give the number of

deaths of COPD patients from all causes. This process was more fully explained in Section 6.4.12.

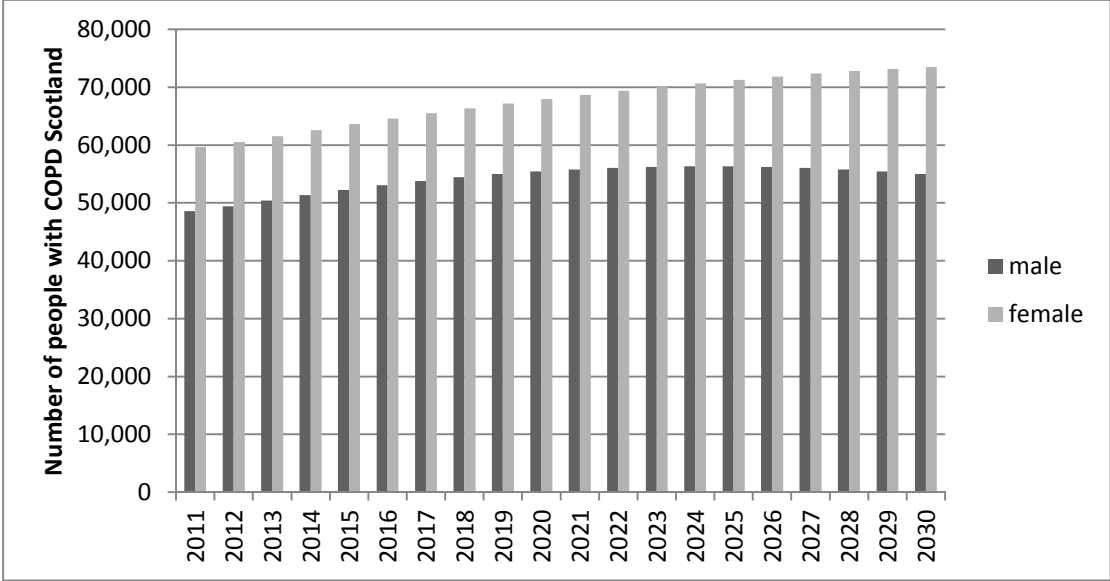
Figure 7.5 Number of deaths among people with COPD in England from all causes



7.3.6 Scotland: COPD population

In Scotland, in 2011, there were estimated to be a total of 108,000 people with COPD (49,000 males and 60,000 females, 33,000 of whom were smokers). This was projected to increase to 128,000 people by 2030, of whom there were 55,000 males and 73,000 females (39,000 smokers), see Figure 7.6.

Figure 7.6 Modelled population of males and females with COPD in Scotland



7.3.7 Scotland: overall prevalence results

The prediction for Scotland was that the prevalence among the over 40 year olds was 4.0% in 2011 and that this would rise to 4.5% by 2030. This involved an increase from 4.1% to 4.8% among females over 40; however, among males over 40, the prevalence was more stable at 3.9% increasing to 4.1%.

7.3.8 Scotland: age-specific prevalence

Figures 7.7 and 7.8 show the age-specific prevalence, as modelled by the Dutch Model, of males and females in Scotland and how it changes between 2011 and 2030. Note that the y-axis is comparable across both figures and that the units are percentage of population with COPD in that age group.

Figure 7.7 Age-specific prevalence of males with COPD in Scotland

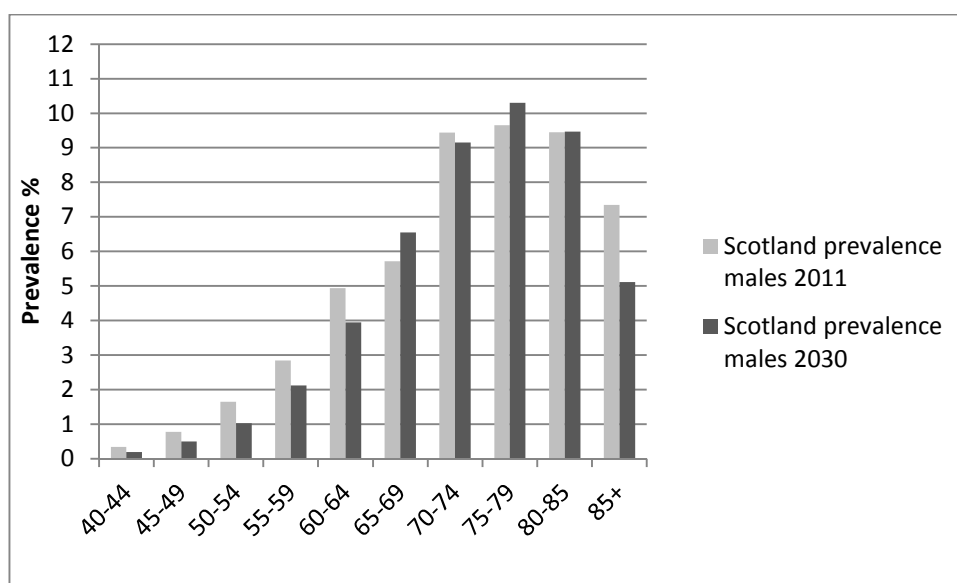


Figure 7.7 demonstrates that the prevalence of COPD in every age group in males is projected to remain roughly stable in Scotland with some decrease in the oldest age group. The prevalence was projected to be, in fact, falling among some of the younger age groups (40-64 years).

Figure 7.8 Age-specific prevalence of females with COPD in Scotland

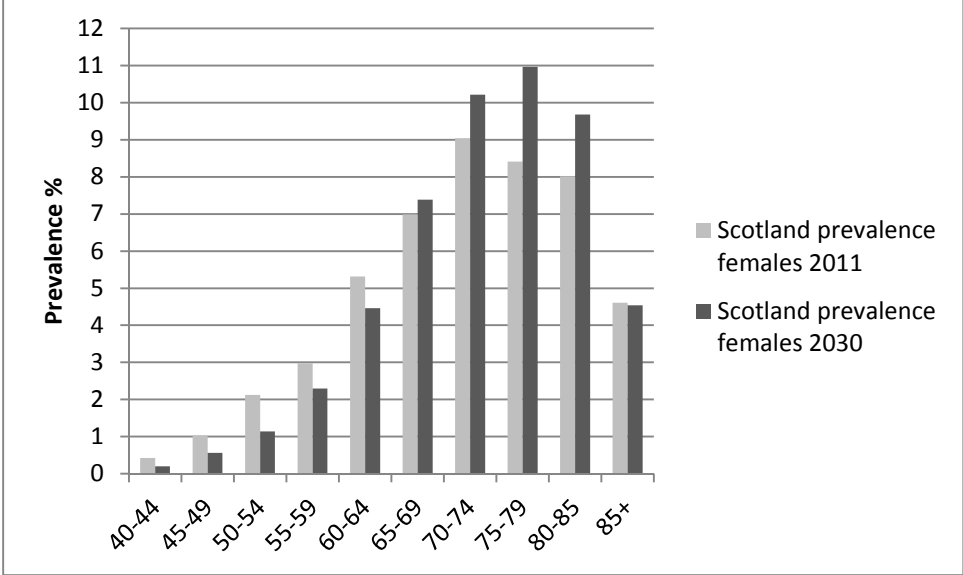


Figure 7.8 shows that the prevalence of COPD among female younger age groups is estimated to fall between 2011 and 2030 in Scotland. However, among the older female age groups, the prevalence was projected to increase, even above the level of prevalence seen for males in those age groups.

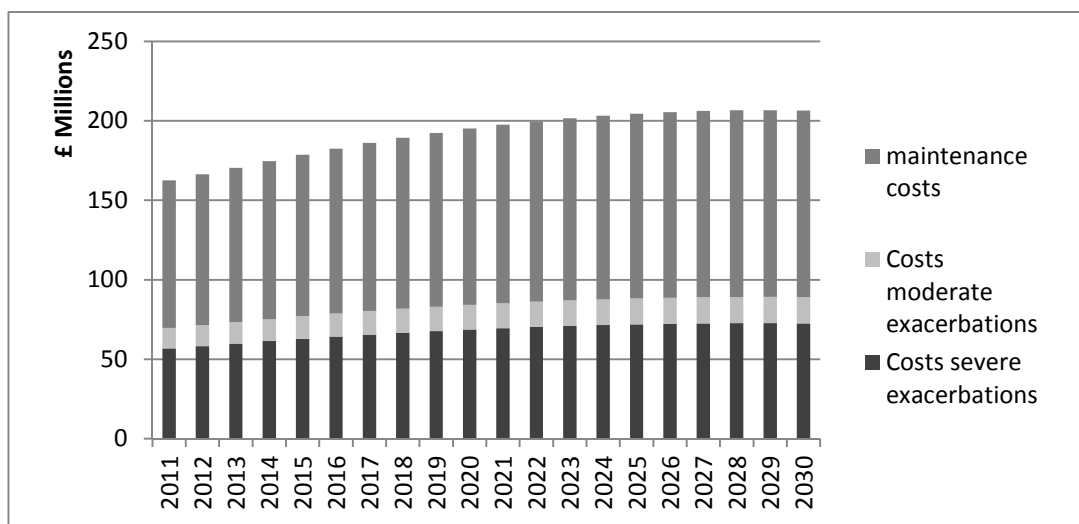
7.3.9 Scotland: direct healthcare costs from COPD

The total annual direct healthcare costs of COPD in Scotland were estimated to increase from £163 million in 2011 to £207 million in 2030. These costs are, again, in terms of 2011 costs without the application of any economic trends.

In 2011, COPD maintenance costs in Scotland were estimated to be £93 million and this was projected to rise to £117 million in 2030.

The composition of total costs in terms of maintenance costs and costs of caring for moderate and severe exacerbations is shown for Scotland in Figure 7.9.

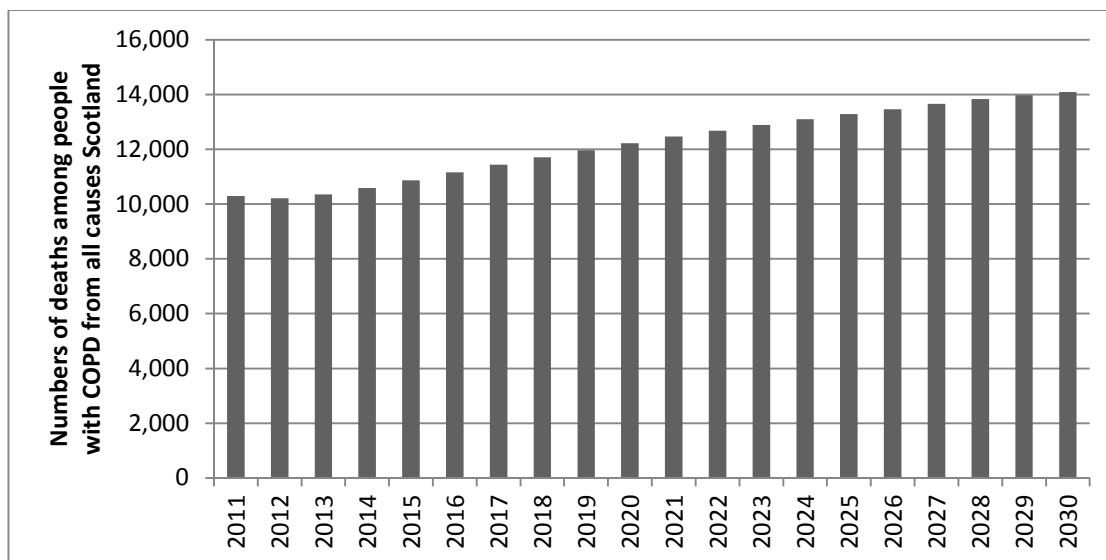
Figure 7.9 Modelled costs for COPD in Scotland



7.3.10 Scotland: number of deaths results

Deaths among COPD patients for Scotland were projected to be stable over the next 3-4 years from 2011 then rise steadily. In 2030 there were projected to be 14,000 deaths among people with COPD from all causes, having risen from 10,000 in 2011.

Figure 7.10 Number of deaths among people with COPD in Scotland from all causes

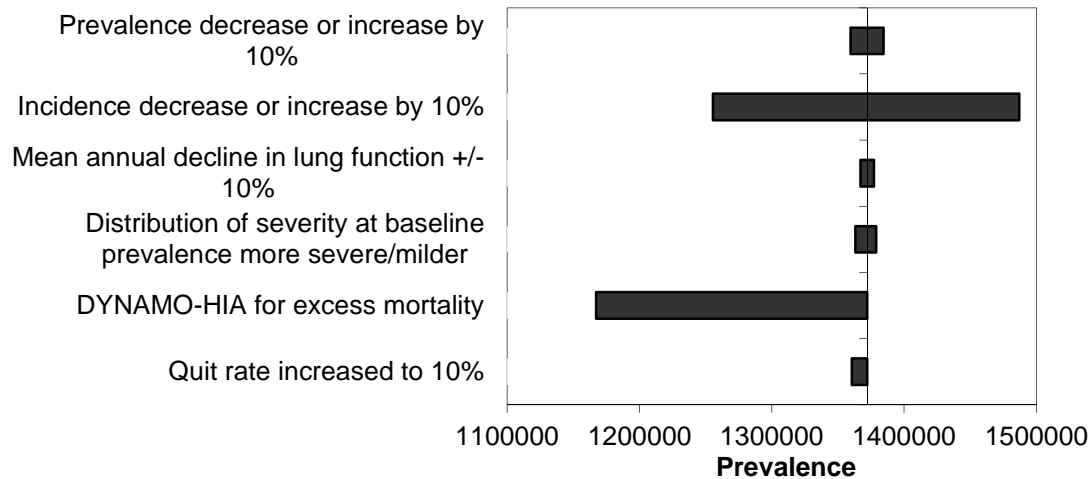


7.4 Results of sensitivity analyses

7.4.1 Sensitivity analyses for England

The following Tornado diagram (Figure 7.11) shows the results of the modelling multiple one-way sensitivity analyses on the projected number of people diagnosed with COPD in 2030, in England.

Figure 7.11 Sensitivity analyses for England, vertical line represents base case



This Tornado diagram (Figure 7.11) shows the results of the sensitivity analyses. Increasing and decreasing prevalence by 10% had only a limited effect on the number of people with COPD in England in 2030. However, increasing and decreasing incidence by 10% had much more impact on the number of people with COPD in England in 2030. The model was, therefore, much more sensitive to changes in incidence than prevalence. The sensitivity of the model to changes in the mean annual decline in lung function and to changes in the distribution of severity of COPD at baseline was much less than the sensitivity of the model to changes in incidence and slightly less than the sensitivity of the model to changes in prevalence.

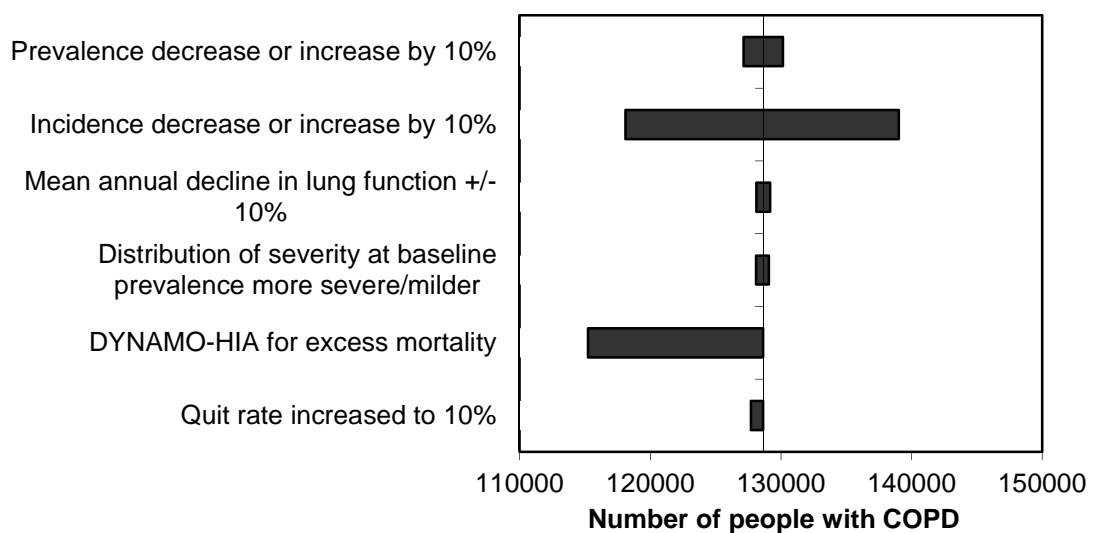
Using the DYNAMO-HIA estimates for excess mortality greatly increased the mortality of people with COPD resulting in many fewer people with COPD in 2030. This is not surprising as the excess mortality estimates for COPD from DYNAMO-HIA were much higher than the excess mortality estimates that I derived from the CPRD data. The reasons for this difference are not clear but may be because these estimates are considerably older (from DYNAMO-HIA data gathered in the 1990s) when prognosis with COPD was worse due to lack of treatment. The prognosis also appeared to be worse because fewer milder cases were diagnosed as COPD than in the current figures. Milder cases live longer affecting the excess mortality figures from the current figures.

Increasing the quit rate of smokers to 10% within each one year age category resulted in a slight decrease in the numbers with COPD due to fewer smokers, therefore fewer people developing COPD. However, the model demonstrated that overall numbers of COPD are not very sensitive to an increase in quit rate to 10% at all ages. This is because in some age and sex groups (e.g. females of reproductive age) quit rates are almost 10% already, so it is not a great increase in quit rate for all age groups.

7.4.2 Sensitivity analyses for Scotland

Figure 7.12 shows the results of the modelling multiple one-way sensitivity analyses on the projected number of people diagnosed with COPD in 2030, in Scotland.

Figure 7.12 Sensitivity analyses for Scotland, vertical line represents base case



The sensitivity analyses for Scotland show a similar pattern to those for England, with the model being more sensitive to changes in incidence than prevalence and not very sensitive to changes in the mean annual rate of decline in lung function or changes in the distribution of severity of the baseline prevalence of COPD.

Again, the DYNAMO-HIA estimates for excess mortality resulted in many more people estimated to be dying with COPD than the estimates from the CPRD dataset (see Chapter 6). This resulted in a much reduced population with COPD in 2030.

There was only a very small effect of increasing the quit rate for smoking to 10% in each one year age category. This is because smoking quit rates are approaching 10% in many age categories already according to the data collected by the Smoking Toolkit Study (see Chapter 6). However, the model did demonstrate that increasing the smoking quit rate would have the

desired public health effect of reducing the overall number of people with COPD. The fact that the results are consistent with this effect helps to validate the model.

7.5 Uncertainty limits for results

Uncertainty limits for the results were calculated by my colleagues in Rotterdam. I helped in an intermediate step by preparing data. The method of doing this was the same as the method that they had used to produce uncertainty estimates in their 2011 paper.(Hoogendoorn, Rutten-van Molken et al. 2011)

The first part of calculating uncertainty limits involved using the raw data for incidence, prevalence, and excess mortality to generate 200 smoothed curves for each of incidence, prevalence and excess mortality for both England and Scotland which were consistent with the probability distributions of the original data. This minimises the impact of the “messiness” of raw gathered data on the uncertainty estimates. That is: some chance findings may not fit the underlying pattern and would have a disproportionate effect when the next step in developing uncertainty intervals is taken. This method helps to clarify the research data’s underlying pattern. The method for doing this is explained in Appendix 5 by Dr Rudolf Hoogenveen, the mathematician who did this work.

This generated 200 potential input files for each parameter of incidence, prevalence and excess mortality. The model was then set to run 1000 times picking randomly from these input files. The 1000 results from the Model’s runs were then arranged in order, and the values at 2.5% and 97.5% were recorded because these values contain 95% of the predicted values of the results from the Model. These “uncertainty intervals” are shown in the table below:

Table 7.2 Uncertainty intervals for outputs from the Dutch Model for England

Parameters	2011 lower uncertainty interval	2011 upper uncertainty interval	2030 lower uncertainty interval	2030 upper uncertainty interval
COPD population males	473,558	486,236	511,405	629,904
COPD population females	467,302	479,518	605,105	778,566
Costs	£1.18 billion	£2.50 billion	£1.85 billion	£3.08 billion
Number of deaths	92,504	128,511	126,350	133,431

Table 7.3 Uncertainty intervals for outputs from the Dutch Model for Scotland

Parameters	2011 lower uncertainty interval	2011 upper uncertainty interval	2030 lower uncertainty interval	2030 upper uncertainty interval
COPD population males	46,186	49,559	46,265	57,872
COPD population females	56,947	60,801	66,647	78,562
Costs	£128 million	£268 million	£165 million	£274 million
Number of deaths	9,000	12,281	13,418	14,548

These uncertainty intervals are all quite broad, in particular for costs, where the upper limit for costs in 2011 overlaps with the lower limit for costs in 2030, thus giving the theoretical possibility that costs could decrease, although this seems unlikely given the direction of all other parameters.

7.5.1 Point estimates and overall prevalence

Up until now, I have been reporting point estimates from the deterministic base case and sensitivity analyses which were modelled initially by me in Rotterdam. However, with the

benefit of the smoothed base case and uncertainty estimates it is possible to give a better point estimate, taking this information into account. Therefore, in Table 7.4, I report the mean of the 1000 runs for the uncertainty calculation as the point estimate along with the 2.5 and 97.5 percentile estimates. Table 7.4 also shows the prevalence results from the modelling study with the overall projected prevalence of COPD in England and Scotland calculated from the ONS population projections (ONS 2012).

Table 7.4 Overall numbers of people with COPD, the population size and prevalence of COPD in 2011 and 2030 in England and Scotland (uncertainty intervals from Tables 7.2 and 7.3, with probabilistic mean)

	England	Scotland
Males with COPD in 2011 (all ages)	479,423	47,891
Total male population in 2011 (Census 2011)	26,133,162	2,548,200
Prevalence (%)	1.8(1.8-1.9)	1.9(1.8-1.9)
Males with COPD in 2030 (all ages)	595,274	53,370
Total male population in 2030 (from ONS 2012 projections(ONS 2012))	30,076,000	2,789,000
Prevalence (%)	2.0(1.7-2.1)	1.9(1.7-2.1)
Females with COPD in 2011 (all ages)	472,812	58,843
Total female population in 2011 (Census 2011)	26,974,007	2,706,600
Prevalence (%)	1.8(1.7-1.8)	2.2(2.1-2.3)
Females with COPD in 2030 (all ages)	728,465	71,834
Total female population in 2030(ONS 2012)	30,334,000	2,901,000
Prevalence(%)	2.4(2.0-2.6)	2.5(2.1-2.7)

7.6 Conclusion to Chapter 7

This chapter has described the modelling and sensitivity analyses undertaken for this PhD. The results of the modelling were given separately for England and Scotland. The dynamic population projection results were presented, followed by the prevalence, the costs and the number of deaths. The results of the sensitivity analyses for England and Scotland are detailed in Tornado diagrams. Finally, the uncertainty limits for the major results are presented. The strengths and limitations of this work, the interpretation of the results, and their implications for policy, practice and further research will be discussed in the concluding chapter of this thesis.

Chapter 8: Discussion

8.1 Introduction

In this final chapter I draw together and interpret the findings from this PhD. Firstly, I summarise the principal findings of the systematic review and modelling studies, then I discuss their strengths and weaknesses. I next consider the findings of the PhD in relation to other published work. I explore the interpretation of the PhD followed by a discussion of its implications for clinicians and policymakers. Finally, I consider the implications of the PhD in terms of what future relevant research needs to be done.

8.2 Statement of principal findings

8.2.1 Principal findings from the systematic review

This is the first work to systematically review models for calculating the prevalence and burden of COPD. I identified 22 relevant models (Bronnum-Hansen and Juel 2000; Stang, Lydick et al. 2000; Feenstra, van Genugten et al. 2001; Shibuya, Mathers et al. 2001; van Genugten, Hoogenveen et al. 2003; Hoogendoorn, Rutten-van Molken et al. 2005; Peabody, Schau et al. 2005; Chapmann, Mannino et al. 2006; Lopez, Shibuya et al. 2006; Hurley and Matthews 2007; Nacul, Soljak et al. 2007; IHME 2010; Atsou, Chouaid et al. 2011; Holt, Zhang et al. 2011; Hoogendoorn, Rutten-van Molken et al. 2011; Pichon-Riviere, Augustovski et al. 2011; Erbas, Ullah et al. 2012; Lhachimi, Nusselder et al. 2012; Najafzadeh, Marra et al. 2012; Perera, Armstrong et al. 2012; Pham, Ozasa et al. 2012; Stanciole, Ortegon et al. 2012). These models were very varied and context specific with some specifying the burden of COPD in DALYs and QALYs and others using population and cost measures. The models were scored for their quality of reporting. Seven models, described in the following nine papers (Feenstra, van Genugten et al. 2001; Shibuya, Mathers et al. 2001; Barendregt, Van Oortmarssen et al. 2003; Hoogendoorn, Rutten-van Molken et al. 2005; Peabody, Schau et al. 2005; Atsou, Chouaid et al. 2011; Pichon-Riviere, Augustovski et al. 2011; Kulik, Nusselder et al. 2012; Lhachimi, Nusselder et al. 2012) scored highly on the quality of reporting framework. (McLean, Wild et al. 2013) Their high quality of reporting enabled an assessment to be made of the actual quality of modelling within these seven models.

Three studies (Feenstra, van Genugten et al. 2001; Hoogendoorn, Rutten-van Molken et al. 2005; Hoogendoorn, Rutten-van Molken et al. 2011) were closely related as versions of the same model and I had the opportunity to collaborate with the Dutch team who had produced

these studies and use their most up to date model for projecting the prevalence and burden of COPD in England and Scotland. The model used routine data from primary care health records including incidence and prevalence rates of COPD as described in Chapter 6; estimates are therefore based on diagnosed COPD.

8.2.2 Principal findings of modelling study

The Dutch Model demonstrated an increase in the population with COPD in England and Scotland between 2011 and 2030. In England, the model projected that the overall prevalence of diagnosed COPD among the over 40 year olds would rise from 3.6% in 2011 to 4.5% in 2030. In Scotland, the estimated prevalence among the over 40 year olds was 4.0% in 2011 and this was projected to rise to 4.5% by 2030. This increase in prevalence was chiefly due to an increase in the numbers of older people with COPD, in particular females over 65 years of age.

Using the probabilistic means and uncertainty intervals, the total annual direct healthcare costs of COPD in England were projected to increase from £1.60 billion (95% uncertainty interval 1.18-2.5) in 2011 to £2.35 billion (1.85-3.08) in 2030. And in Scotland, costs were projected to increase from £170 million (128-268) in 2011 to £210 million (165-274) in 2030. These costs are in terms of 2011 costs without the application of any economic trends (i.e. no annual increase applied for inflation).

Again using probabilistic means and uncertainty intervals, the number of deaths among people with COPD in England was modelled as 99,181 (92,504-128,511) in 2011, increasing to 129,428 (126,350-133,431) in 2030. In Scotland there were modelled to be 9,734 (9,000-12,281) deaths in 2011, increasing to 13,896 (13,418-14,548) in 2030.

The Dutch Model suggested a 39% increase in the number of people with COPD in England and a 17% increase in Scotland between 2011 and 2030. It modelled a 30% increase in deaths among people with COPD in England and a 43% increase in Scotland. Overall, there was an estimated 46% increase in the direct healthcare costs required to care for people with COPD in England and a 23% increase in Scotland between 2011 and 2030.

8.3 Comparison of modelling results from England with results from Scotland

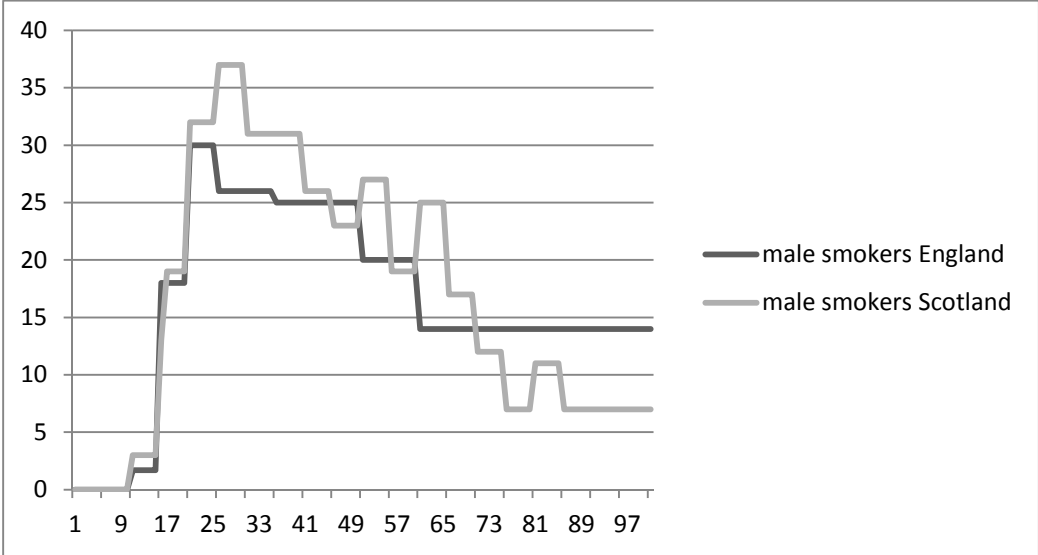
There are interesting differences between the results of the modelling for England and Scotland. Both England and Scotland prevalence results reveal that there will be a predominance of females with COPD by 2030. However, in comparing figures 7.1 and 7.6, it

can be seen that the predominance of females with COPD appears to be established in Scotland, however, it is still emerging in England. There is some evidence that this may be a real difference from the smoking patterns in the two countries, however, it must also be explored whether the difference is artefactual, due to differences in the input data for the two countries.

8.3.1 Female COPD predominance, differences between England and Scotland, evidence from smoking habits

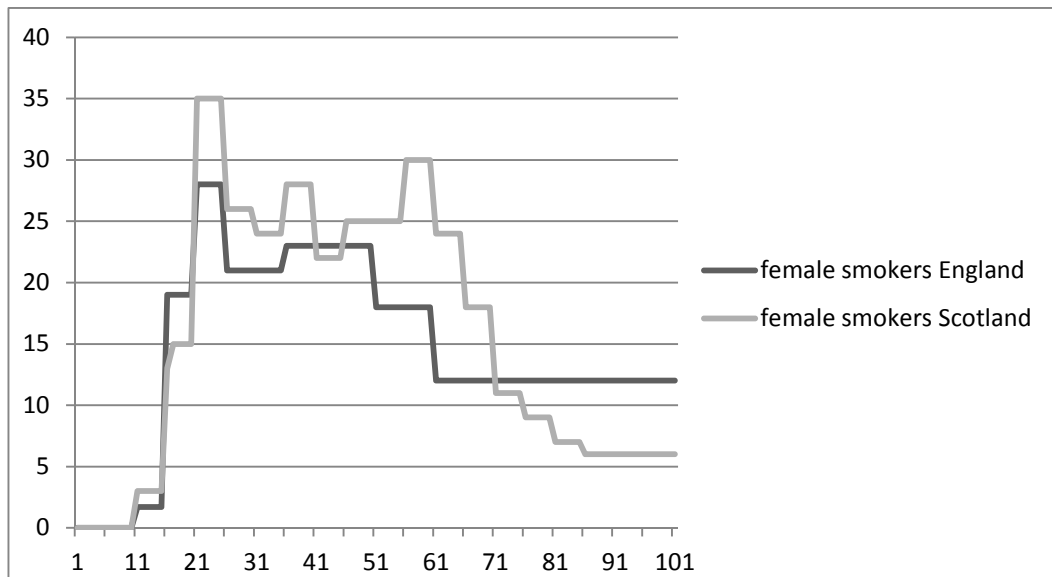
Comparing the smoking habits of males and females in England and Scotland by graphing the prevalence of smokers from tables 6.5,6.6 and 6.7, a pattern emerges. In figure 8.1, it can be seen that males from the two countries have an approximately similar prevalence of smoking at different ages, with the overall prevalence slightly higher in Scotland, especially at younger ages.

Figure 8.1 Age related prevalence of smoking in England and Scotland in males



This is different to figure 8.2, which shows the prevalence of smoking in England and Scotland in females. This demonstrates a markedly greater prevalence among females in Scotland compared with females in England in the key age group for the development of COPD of 45 to 75 year olds. As this age group will be composed of established smokers, this suggests that the COPD female predominance is already established in Scotland and is partly a real finding.

Figure 8.2 Age related prevalence of smoking in England and Scotland in females



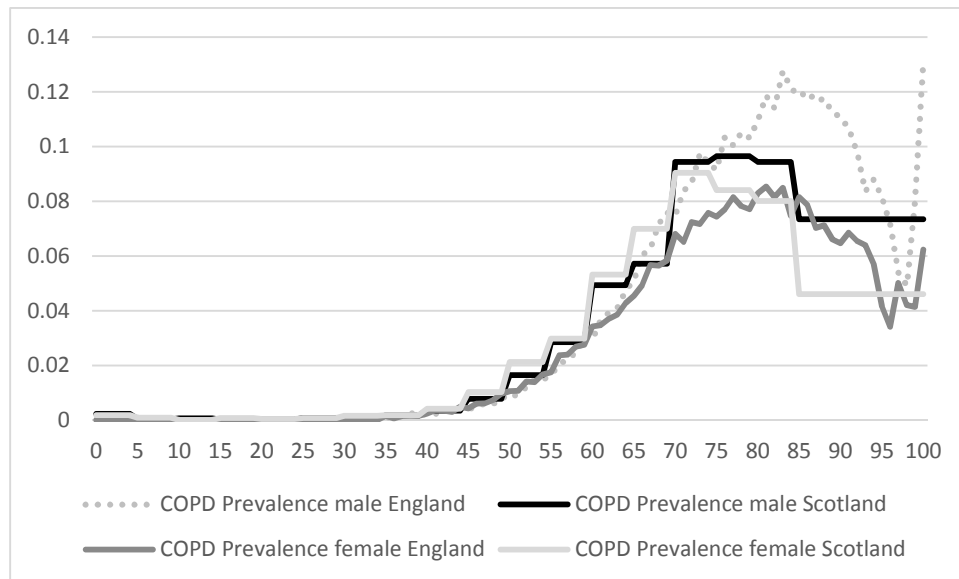
8.3.2 Female COPD predominance, differences between England and Scotland, differences in baseline input prevalence

Before it can be concluded that the early predominance of females in Scotland compared with England is a real difference, it is important to review any evidence that it may in fact be an artefactual difference. Due to the differences in data used in the modelling process.

Whereas the above smoking data for both countries came from broadly equivalent survey sources, the General Lifestyle Survey in England and the Scottish Health Survey in Scotland, the data used for the prevalence at the baseline of the model was quite different between England and Scotland. CPRD data was used for England and PTI consultation data was used for Scotland. The fact that the Scottish PTI dataset was based on consultation rates may bias this data. The reason for this is that there is a body of evidence which shows that females are heavier users of healthcare services than men. (Hunt, Ford et al. 1999; Bertakis, Azari et al. 2000; Wang, Hunt et al. 2013) This may result in the calculation of a higher prevalence for COPD among females in Scotland compared with the CPRD England data.

The reason for using PTI data rather than CPRD data for the prevalence in Scotland was down to cost restrictions. It should also be noted that there were some slight differences in the Read codes used to extract data from PTI and from CPRD, (see table 6.3 and Appendix 1.1). In addition, the definition used from CPRD was lifetime point prevalence, whereas in PTI it was a 15 month consultation prevalence. In order to see the impact using these different data may have had, the following graph compares the baseline prevalence rates calculated from CPRD data in England and PTI data in Scotland. See figure 8.3.

Figure 8.3 Age related prevalence at baseline in England and Scotland



It can be seen that in England the CPRD data clearly shows a higher prevalence of COPD among males at baseline, than among English females. Whereas in Scotland, up until the age of 70 the prevalence is higher among females than males. This could clearly be artefactual due to the data being based on consultation rates. Therefore one of the limitations of this modelling study is that the baseline 2011 results for Scotland may be biased due to the source of the prevalence data. However, as the model continues to run the impact of this bias will be attenuated as a result of the effects of the input data for incidence having greater impact as shown in the sensitivity analysis. See 7.4.1 and 7.4.2.

8.3.3 Differences between England and Scotland, differences in relative increases over 20 years

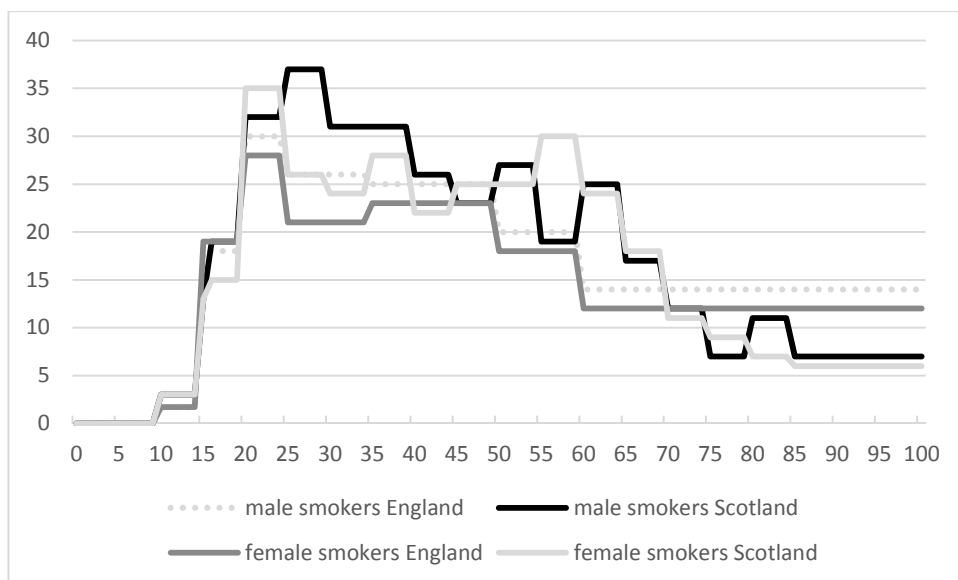
The above section compared the differences in prevalence at baseline, however, over the course of running the model comparisons can be made in the absolute increases of males and females with COPD. There appears to be something strange happening when you consider these absolute and relative increases over the time to 2030. The relative increase for English males and females is 24% and 54% respectively. The relative increase for males and females in Scotland is 11% and 22% respectively. See table 8.1, these differences need to be explained.

Table 8.1 Absolute and relative differences in increases in the COPD population

	2011	2030	Absolute increase	Increase as % of 2011 value
England males	479423	595274	115851	24
England females	472812	728465	255653	54
Scotland males	47891	53370	5479	11
Scotland females	58843	71834	12991	22

Initially I thought that these differences could be caused by higher smoking rates among English females, however, the opposite is true, English females have the lowest smoking rates and Scottish males have the highest rates. See figure 8.4. The effects of these smoking rates are that females in England experience lung function decline and COPD related mortality later on in their life than males in Scotland. Males in Scotland, in contrast die younger as a result of their heavy smoking and so are not prevalent with COPD for as long. It should also be noted that among the age groups over 70 there are fewer people within each demographic category and so the effect on the COPD absolute numbers comes from the largest demographic categories age 30-65.

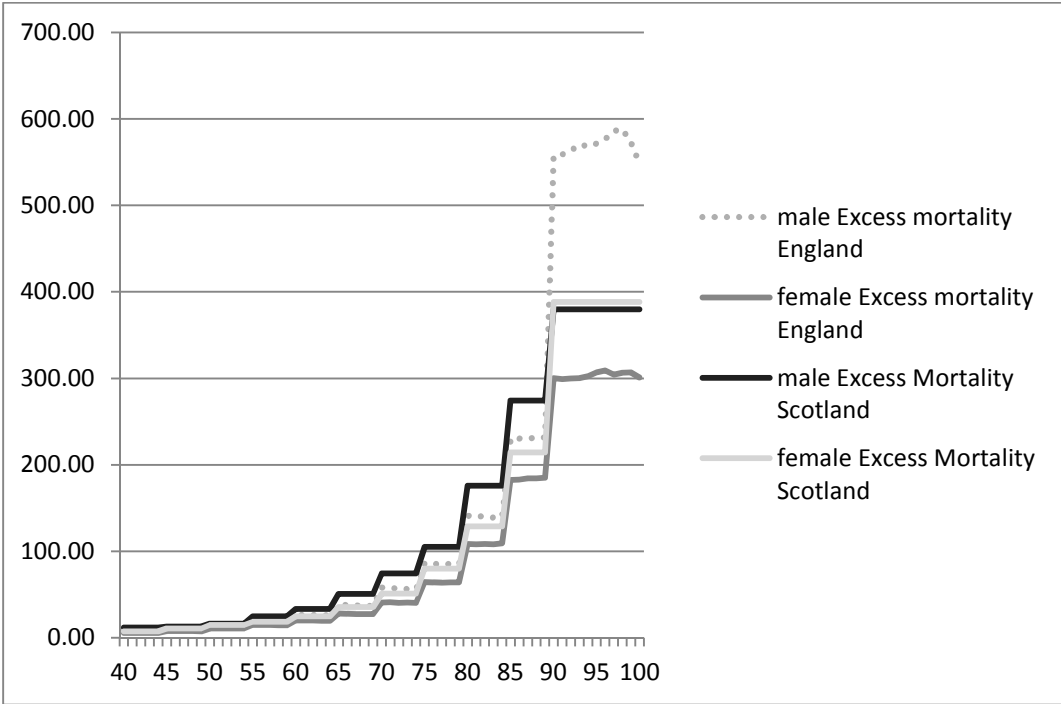
Figure 8.4 Age related smoking prevalence in England and Scotland (a combination of figures 8.1 and 8.2)



The impact of this pattern of smoking is enhanced in the model's projections by the impact of the COPD related excess mortality. English females have the lowest rate of excess mortality (they have the lowest rate of smoking) and Scottish males have the highest rate of

excess mortality (due to their high rate of smoking). The different levels of excess mortality are compared in figure 8.5.

Figure 8.5 Age related COPD excess mortality per 1000 population.



As the sensitivity analyses showed that the model was most sensitive to the input of excess mortality in terms of the output of absolute numbers of population, it is, therefore, not surprising that the pattern of high excess mortality among males in Scotland and low excess mortality among females in England results in the above respective increases in absolute numbers of population. (i.e. 54% for females in England and 11% for males in Scotland). Note that although the pattern of excess mortality for those over 90 changes in the above graph this has little effect on the absolute number with COPD because there are few people alive in this age group.

8.4 Strengths and limitations

8.4.1 Strengths and limitations of the systematic review

The strengths of the systematic review include that a protocol was developed and registered in advance and that I used a highly sensitive search strategy. I searched Medline, Embase, CAB Abstracts, World Health Organization (WHO) Library and Information Services (WHOLIS – library catalogue of books and reports), WHO Regional Indexes (AIM (AFRO,

LILACS), (AMR/PAHO), IMEMR (EMRO), IMSEAR (SEARO), WPRIM (WPRO)) and used a modified search strategy to identify reports from the WHO home website and Google. I updated the searches in November 2013. There were no language restrictions employed when undertaking searches. The selection of studies and data extraction was checked by a second person as per established systematic review methods. Furthermore, a formal approach to assessing reporting quality was used.(McLean, Wild et al. 2013)

The limitations of the systematic review include that by preliminary assessment of the quality of reporting of studies I may have excluded a study that was of high quality, but not appropriately reported. There is some potential for this to have happened as I could not find a suitable report of the WHO Burden model DISMOD 3 (IHME 2010) to assess its quality.

In addition, I did not undertake an in-depth critical review of the 22 models identified. This limitation is in part due to modellers not fully disclosing their modelling techniques in the reporting papers. Also, I did not have the mathematical skill set to fully evaluate the mathematical properties of the models.

I would also like to comment on the limitations of using a quality of reporting scoring checklist. This type of scoring checklist which assigns a numerical score to different elements of the study was previously often used by systematic reviewers including those who worked for the Cochrane collaboration. However, more recently raw numerical scores have fallen out of Cochrane's favour due to their perception that it is not logical to "sum up" technically diverse elements of a study in this way and also that such summing allows high performance in one area to have the potential to disguise fatal deficiencies in another. Cochrane replaced numerical sum scores with a system of scoring separate categories for potential risk of bias and then giving an overall summary impression for the study as to whether it was at high, moderate or low risk of bias. This system involves a loss of granularity of information when the categories are summarised. As the purpose of using a scoring system in this thesis was to identify the "best" study, I would argue that maintaining a numerical aspect to the categorisation was helpful in enabling me to discriminate between studies. However, I do concede that the logic of summing scores for separate technically diverse elements is, at best, somewhat fuzzy. Therefore the use of the quality scoring system in this thesis should be seen as a compromise methodology. However, I would argue that Cochrane's current risk of bias scores have not negated completely the need to summarise across diverse technical elements and still maintain a degree of fuzziness in the process, and so I am in good company in having to choose a compromise.

8.4.2 Strengths and limitations of modelling study

The strengths of the modelling study include the careful selection of the Dutch Model. Having the opportunity to work alongside the team that developed the model enabled me to acquire a deep understanding of the model. The high validity of the data sources used to enter into the model is an additional strength. (See Chapter 6 discussion of the strengths and weaknesses of the original data sources). The model was also thoroughly tested using sensitivity analyses, which can also be regarded as a strength.

In addition, another strength is the fact that both modelled COPD prevalence for England and Scotland in 2030 among the over 40 year olds were anticipated to be 4.5% despite slightly different starting points. These similar results can be taken as cross-validation of the modelling process because in these two countries with similar prevalences of smoking and demographic distributions, the model, run on these two separate but similar sets of data, produces similar results.

Limitations of the modelling study include that no further predictive validation was possible for the population or mortality data. Ideally, comparisons would be made with another source of projections. However, projections for England (ERPHO, Walford et al. 2011) which might have provided a comparison are for rates of diagnosed and undiagnosed COPD, as described in Chapter 2. The alternative is to compare the projections with true count data as they are gathered, year upon year, as time goes on.

It is also important to consider the validity of the data that were employed in the modelling study. In Chapter 6, I described each of the data sources and the efforts that were made to select data sources of high validity. In Section 3.3.1, I detailed the relevant validation work done on the CPRD database by Campbell et al. (Campbell, Dedman et al. 2013) which indicated that overall the CPRD database may contain a group that is older than the general UK population. The implications of this difference for the modelling study is that prevalence of COPD may be higher among the older CPRD database cohort than among the general UK population, leading to an over-estimate of the English prevalence of COPD by the model. In addition, there are implications for the mortality figures. If mortality is lower among the CPRD cohort including among COPD patients in the CPRD cohort than among the whole population then the modelling of the excess mortality will result in a lower estimate of the excess mortality due to COPD and a consequently higher prevalence of COPD from the model. However, the differences in the ages of the CPRD GOLD database compared to the whole population found by Campbell et al. were relatively minor in proportion.

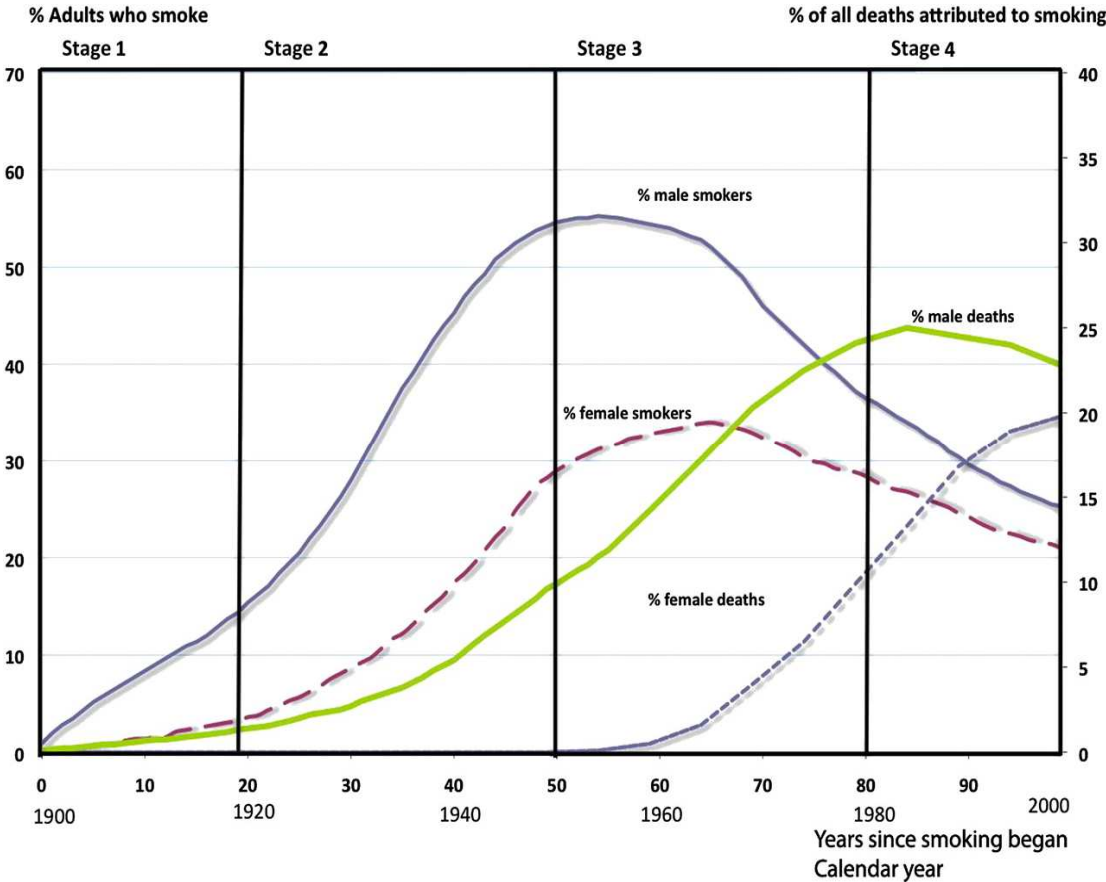
Another limitation to be highlighted is the inevitable compromises that occur when expressing real-world populations as models. Representing the COPD population of one or more countries in the form of a Markov model results in an inevitable compromise on information. The model represents a caricature of disease progression and although the Dutch model is complex and even includes parameters for some patients to recover some lung function if they quit smoking it cannot fully represent the real-world situation. One of the areas where most information is compromised in this model is the characterisation of COPD severity on the basis of airflow limitation alone. As mentioned in chapter 2, the 2013 GOLD rubric for classification of severity of COPD has been amended to include a consideration of symptoms and exacerbations in addition to airflow limitation. (GOLD 2013) These GOLD guidelines have also highlighted the need to consider co-morbidities when classifying the severity of people with COPD. However, such updating can result in simply shifting the compromises to different areas, for example, it has been shown that although the 2013 GOLD classification is better than the 2007 classification for predicting exacerbations, it is worse at predicting mortality and lung function decline. (Goossens, Leimer et al. 2014) It is inevitable that all models are compromises, however, when carefully used, with the appropriate caveats they remain useful tools.

8.5 Interpretation of the results of this study in the context of previous research

8.5.1 The shape of the smoking epidemic

A schematic diagram describing the smoking epidemic in developed countries was originally suggested in 1994. (Lopez, Collishaw et al. 1994) This diagram was updated in 2012 (Figure 8.1) and shows the epidemic's shape (Thun, Peto et al. 2012). Stage 1 is the beginning of the epidemic when smoking prevalence is rising among males, but smoking has, as yet, caused relatively few deaths. Stage 2 shows a rapid rise in smoking prevalence among men to a peak of 40-80%. Also during Stage 2 is the start of the increase in smoking among women and the start of the increase in smoking attributable mortality which still only accounts for 10% of male deaths at the end of this stage. Stage 3 sees a flattening in smoking prevalence among men and a convergence of male and female smoking prevalence. Mortality due to cigarette smoking rises from 10% to approximately 30% of all deaths, mostly in men. In Stage 4, although paradoxically the prevalence of smoking is decreasing in both sexes, the mortality attributable to smoking continues to rise for some time. This mortality peaks at about a third of all male deaths (more among middle-aged men) and a smaller percentage of all female deaths before beginning to decline.

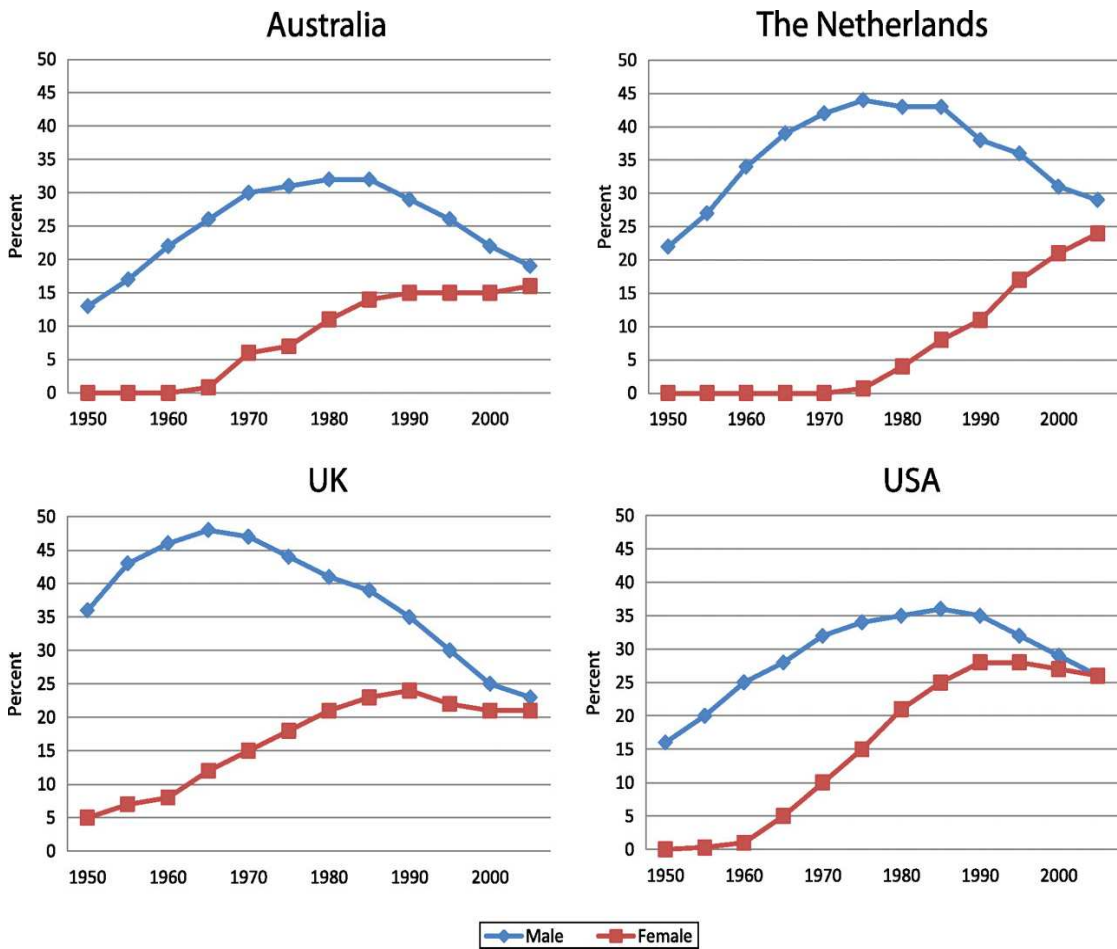
Figure 8.6 Smoking epidemic diagram for high-income countries. Reproduced with permission (Thun, Peto et al. 2012)



The methods used to develop this diagram used estimates of smoking attributed mortality in economically-developed countries in an indirect approach, where the lung cancer death rates were used as a guide to the smoking attributed fractions of the deaths from various other diseases. If in an economically-developed country, lung cancer rates are high then smoking is probably responsible – not only for these deaths but also for many other deaths. The updated diagram (see Figure 8.2) emphasised the age group of 35-69 years because the proportion of all deaths attributable to smoking is higher in that age group than at older ages and death in this age group is considered premature.

The patterns of smoking attributable deaths in four high income countries are shown below:

Figure 8.7 Smoking-attributable deaths at ages 35-69 estimated indirectly from lung cancer rates and expressed as a percentage of all deaths. Reproduced with permission (Thun, Peto et al. 2012)



In the UK, smoking attributable mortality among men in 1970 was 47% at the peak of the epidemic curve. It then fell to 22% in 2009 and continues to fall. The epidemic among women is significantly later than among men in all four countries. Smoking prevalence has decreased among women, but more slowly than among men. Therefore, smoking attributable deaths have not decreased substantially among women in any country yet, and are still increasing rapidly in the Netherlands as shown. This figure also shows that the male and female smoking attributed proportions of all deaths at ages 35-69 have converged. This is in line with the earlier convergence of the male and female prevalences of smoking. In the Netherlands the male and female smoking attributed proportions of deaths at ages 35-69 years will cross over in the early 2010s. The smoking epidemic among females lagged behind males by 20-30 years in the UK and the USA. Cultural norms changed with World War II and women became freer to smoke. (Centers_for_disease_control_and_prevention 2001)

This description illustrates how there is a long delay between a large increase in smoking in a particular generation of young adults and the full increase in smoking attributable mortality when that generation becomes middle aged and elderly. The updated schematic diagram separates out the epidemic for males and females with the following stages. Stage 1: the prevalence of cigarette smoking begins to rise in the gender of interest. Stage 2: the prevalence of smoking increases rapidly but until it levels off and cigarette smoking only accounts for a limited proportion of premature deaths. Stage 3: the paradoxical stage in which the smoking prevalence is stable or decreasing, however the smoking attributable mortality is increasing rapidly. Stage 3 ends when this mortality peaks. Then, in Stage 4, prevalence and then, many years later, smoking attributed mortality are expected to both fall - this has not yet been experienced by any population.

This schematic diagram can be used to put the future projections from the Dutch Model of COPD prevalence and mortality for England and Scotland in context.

The predictions from the Dutch Model fit broadly into this characterisation of the smoking epidemic. The model projects that over the next 20 years there will be an increase in the numbers of older people with COPD, in particular the numbers of older women (see age-specific prevalence diagrams for England and Scotland in Chapter 7). This is as a result of the time lag from the peak in the prevalence of smoking (and also a high prevalence of ex-smokers) to getting COPD. In addition, people are living longer with COPD and increasing survival also contributes to higher prevalence figures.

8.5.2 Comparison with UK data from the Global Burden of Disease Study 2010

The main measure used to quantify burden of disease in the Global Burden of Disease (GBD) Study is the Disability Adjusted Life Year (DALY).(Murray, Richards et al. 2013) (Page 1009) The absolute numbers of DALYs for all ages in the UK, contributed by COPD in 1990 was 714,000 (95% uncertainty interval 594,000-866,000) and in 2010 was 707,000 (573,000-875,000), a fall of 1.0% (Page 1009 of GBD Lancet report). When this decrease is standardised by population, a decrease of 8.5% change in DALYs becomes apparent, from 1,249,000 (1,038,000-1,513,000) per 100,000 in 1990 to 1,142,000 (942,000-1,413,000) in 2010. The GBD authors concluded that this represented falling rates of tobacco attributable burden for both men and women as the UK has a more advanced tobacco epidemic than most high-income nations.(Murray, Richards et al. 2013)

The apparent inconsistency is that the GBD study states that the burden of COPD fell (by 8.5% in standardised rates) between 1990 and 2010 in DALYs in the UK (Murray, Richards et al. 2013), whereas, my work suggests that the prevalence, costs and deaths will increase from 2011 to 2030.

There is one area where the findings of the GBD and my study are comparable. The major decrease of the GBD is among Years of Life Lost (YLL) due to COPD in the age group 20-54 (Murray, Richards et al. 2013). My findings of falling age specific prevalence among the younger age groups as reported in Chapter 7 are consistent with this finding from the GBD. It should, however, be remembered that prevalence is not the same as YLL.

The main thing to realise is that the GBD and the Dutch Model represent very different approaches and time periods and so cannot easily be compared. In addition, the GBD used a spirometric definition of COPD as its input data for prevalence including data from the Health Survey for England whereas the Dutch Model uses a definition of physician diagnosed COPD.(Gupta 2014)

One explanation to be considered is that the GBD and my work are not actually inconsistent. The GBD represents findings from 1990 to 2010, therefore, it may be that during this period there were high levels of undiagnosed COPD and low levels of diagnosed COPD with the COPD that was diagnosed being among the more severe cases. Therefore this would be calculated to represent a high burden in terms of DALYs (this would be the case if the work done to inform the DALY allocation was done in diagnosed COPD rather than in spirometrically defined COPD). This burden would fall with falling smoking prevalence. In contrast, my work which focuses on diagnosed COPD shows a rising prevalence of COPD because, from 2011 to 2030, more mild cases of COPD are being diagnosed and people are living longer with their COPD, therefore being prevalent for longer. However, the burden in terms of DALYs for this period would not actually be so high because the cases represent the milder end of the COPD spectrum.

8.5.3 The lag effect of smoking: ex-smokers developing COPD

The paradox that needs addressing in these results is why the prevalence of COPD is still modelled to be increasing despite the widely observed fall in smoking prevalence in Scotland and England. A similar paradox was seen in the Netherlands (Pride 2006). This is because people who have given up smoking still are at risk of developing COPD (see Table 8.1).

Table 8.2 Relative risks for ex-smokers of developing COPD compared to non-smokers (Van Oers 2002; SurgeonGeneral 2004)

Age	Male ex-smoker relative risk of developing COPD*	Female ex-smoker relative risk of developing COPD*
55-59	9.6	7.7
60-64	10.3	8.3
65-69	9.5	8.2
70-74	8.5	6.5

*No confidence intervals were available.

Many of these ex-smokers become part of the older age classes where incidence of COPD is high. The relative risk of developing COPD after smoking cessation does not decline as fast as it does for lung cancer. The rate of FEV1 decline is faster in smokers than ex-smokers, however, there is evidence from the Lung Health Study that the rate of decline in FEV1 in ex-smokers reverts to values similar to those in healthy never-smokers within a year of stopping smoking.(Scanlon, Connett et al. 2000) However, note that this is still a decline, there is little recovery in lung function (i.e. there is no increase in FEV1 with stopping smoking.) The effect of this decline at normal “healthy” rates in ex-smokers who quit smoking with some established impairment is to delay rather than prevent death from or with COPD.(Pride 2001) This delay in death thereby increases the population with established COPD.

8.5.4 Comparison with results from the Netherlands

The Dutch Model was used to calculate projections of COPD prevalence from 2000 to 2025 in the Netherlands. It projected that COPD patients increased from 188,000 to 270,000 for males and from 117,000 to 224,000 for females (Hoogendoorn, Rutten-van Molken et al. 2005). The prevalence estimates are shown in the following Table 8.2.

Table 8.3 Overall projected prevalence of COPD in 2025 in three countries

	England	Scotland	Netherlands
Males with COPD in 2025	582,226 (506,432-612,210)	54,676 (47,795-58,828)	270,000
Total male population in 2025 (from ONS 2012 projections(ONS 2012)	29,144,000	2,728,000	-
Prevalence(%)	2.0(1.7-2.1)	2.0(1.8-2.2)	3.3
Females with COPD in 2025	665,107 (563,240-705,747)	69,209 (59,203-74,904)	224,000
Total female population in 2025(ONS 2012)	29,463,000	2,857,000	-
Prevalence(%)	2.3(1.9-2.4)	2.4(2.1-2.6)	2.7

Table 8.2 shows that in 2025 the overall prevalence of COPD among males (all ages) is projected to be 2.0% in both England and Scotland, whereas in the Netherlands it is projected to be 3.3%. The overall prevalence among females (all ages) is projected to be England 2.3% and Scotland 2.4%; but not as high as for females in the Netherlands 2.7%.

In this context, estimates for Scotland and England appear very similar but estimates for the Netherlands are different. It may be an artefactual difference due to the inputs to the Netherlands model being sourced before 2000 and therefore reflecting higher smoking prevalences. The later data sourcing for England and Scotland may result in lower smoking prevalences - as in European Union countries smoking prevalences have been falling. However, there is some evidence that the smoking prevalence in the Netherlands is higher than in the UK (37% of males and 31% of females aged 25-64 smoked in 2012 in the Netherlands, compared with 33% of UK males and 29% of UK females in this age group). (Zatonski, Przewozniak et al. 2012) In this case, these findings may represent a real difference.

An alternative explanation is that perhaps patterns of diagnosis are different in the Netherlands from England and Scotland. It may, therefore, be the case that although undiagnosed prevalence rates might be similar across the three countries, the Netherlands are more efficient at diagnosing COPD. Therefore these figures would show a greater prevalence of diagnosed COPD in the Netherlands than in England and Scotland.

8.6 Interpreting this research: implications for clinicians and policymakers

8.6.1 Five, 10 and 20 year projections

Table 8.3 summarises the results from Chapter 7 (this time using the deterministically modelled results rather than the probabilistic results with uncertainty intervals as it was not possible to calculate probabilistic results for each year). In England, the COPD population was modelled to increase by 13% from 2014 to 2019 and by 24% from 2014 to 2024. In Scotland the increases were modelled to be 7% and 11% respectively. Deaths among people with COPD occur among about 10% of the COPD population each year and this is not projected to change greatly over the 20 year period; except for Scotland in 2030 when deaths are expected to affect 11% of the population with COPD. The UK government is already aware that there is likely to be an increase in the number of COPD patients dying in the future and so is planning evidence-based action to help try and prevent this as far as possible.(Dept_of_Health 2014) This forms part of the aims of the UK as a signatory to the World Health Organization’s Framework Convention on Tobacco Control.(FCA 2015) This WHO public health treaty encourages signatories to introduce smoke free legislation, stop the promotion of tobacco, make tobacco less affordable through tax, regulate tobacco products effectively and help tobacco users to quit.

Table 8.4 Predictions from the Dutch Model

England	2011	2014	2019	2024	2030
COPD population	952,000	1,026,000	1,163,000	1,273,000	1,350,000
COPD deaths	96,000	93,000	106,000	120,000	134,000
COPD cost (billions)	1.49	1.64	1.91	2.13	2.28
Scotland					
COPD population	108,000	114,000	122,000	127,000	128,000
COPD deaths	10,000	11,000	12,000	13,000	14,000
COPD cost (billions)	0.18	0.19	0.21	0.22	0.21

8.6.2 Implications of the age of the COPD population

The results presented in Figures 7.2 and 7.3 show the age-specific prevalence of COPD in England by sex, comparing predictions for 2011 with 2030. These show that the majority of increase in overall prevalence of COPD is as a result of an increase in females age 65 to 85. A similar pattern is seen for Scotland in Figures 7.7 and 7.8. These women will suffer

disabilities in proportion to their COPD and this will have important implications for the shape of the health and social care services that they require. Both my work and the GBD imply that disability is not so much prevented as postponed, which means that there will be a demographic bulge of frail elderly people with COPD requiring health and social care. In addition, these people with COPD are likely to have co-morbid conditions resulting in very complex needs. There is already evidence of a trend among elderly patients towards multi-morbidity and highly complex medical needs.(Barnett, Mercer et al. 2012) Traditional health and social services are already stretched beyond capacity dealing with such frail elderly people. There is policy interest in the role of technology to provide care services that were once provided by staff, this being driven by a shortage of skilled staff and a drive to keep costs down. Technological solutions to providing care services are already being trialled in some settings (e.g. to deliver intelligent context-specific medication reminders so that an elderly person is prompted to take medication when patients are at home.)(Hayes, Cobbinah et al. 2009) Other developments include “smart homes” for elderly people to enable independent living as long as possible, with a particular focus on falls prevention.(Haux, Hein et al. 2014) The role of telehealth in caring for people with COPD is also being clarified, a key problem is how best to predict COPD exacerbations from the data gathered from telehealth monitoring and further work is ongoing in this area.(Pinnock, Hanley et al. 2013)

8.6.3 Implications in terms of NHS spending for England

According to the ONS (ONS 2011) the NHS spent £142 billion in 2011. The deterministically modelled costs for COPD in England £1.490 billion in 2011 would be 1% of UK NHS spending for that year.

Some predictive validation is possible for the cost data. Table 8.4 shows the direct costs predicted for COPD from the model for the years 2011-2013 and the actual costs estimated from Primary Care Trust (PCT) returns for the financial years 2010/11, 2011/12 and 2012/13.(Dept_of_Health 2012; Dept_of_Health 2013)

Table 8.5 Comparison of English cost data from the Dutch Model with PCT returns

Year	2010	2011	2012	2013
Modelled cost (£billions)	Not modelled	1.49 (95% UI 1.17 to 2.46)	1.53	1.58
Financial year	2010/11	2011/12	2012/13	2013/14
Estimated cost from PCT returns (£billions) “obstructive airways disease”	0.70	0.75	0.80	Not yet available

These PCT returns are the best data available to guide how much NHS England is spending on different conditions, however, there are some issues with the quality of the returns data. In the PCT returns, costs are categorised firstly by speciality (i.e. “Respiratory”) and include all primary care, hospital inpatient and outpatient costs. They do not include Department of Health and Strategic Health Authority expenditure. Therefore, they are an estimate of direct costs similar to the direct costs calculated in this PhD and so can be used for validation. However, the PCT returns data encountered problems classifying “Respiratory” spending into subcategories: it was not always possible for subcategories to be assigned. In 2011/12, “Respiratory” expenditure was £4.41 billion, this was sub-categorised as £0.75 billion “Obstructive Airways Disease”, £1.01 billion “Asthma” and £2.65 billion “Other”. This £2.65 billion “Other” is not exclusive, therefore, it is likely to include unclassified spending on COPD. (Lung cancer is considered under “Cancer” rather than “Respiratory”). “Other” expenditure in 2010/11 was £2.58 billion and in 2012/2013 was £2.85 billion so these are not insignificant funds and may well account for the difference between the subcategory and the cost estimates from this PhD. In other words, my belief is that my PhD has not underestimated the direct healthcare costs of COPD, simply the remaining COPD costs have been classified as “Other” in the PCT Returns and so I think the official NHS England PCT returns vastly underestimate true COPD related costs.

8.6.4 Implications in terms of NHS spending for Scotland

Table 8.5 shows data from a costing exercise at Information Services Division (ISD) to try to establish how much was spent on COPD by the NHS in Scotland in each of the three years shown. The estimated costs represent hospital inpatient and day-case costs for COPD. Therefore the equivalent subgroup of costs which were modelled would be the severe

exacerbation costs, which were defined as those COPD exacerbations that required hospital management.

Table 8.6 Comparison of Scottish cost data from the Dutch Model for severe exacerbations with ISD cost data for hospital admissions

Year	2010	2011	2012	2013
Modelled cost of exacerbations (£millions)	Not modelled	56.8	58.3	59.8
Financial year	2010/11	2011/12	2012/13	2013/14
Estimated cost from ISD (£millions)	50.0	55.3	57.9	Not yet available

These estimates show that in the financial year 2011/12, Scotland was estimated to have £55.3 million on COPD inpatient and day-case care by ISD. The Dutch Model estimated expenditure on severe exacerbations to be £56.8 million in 2011. Relevant estimates for 2012/13 were £57.9 million and £58.3 million respectively. Therefore, it can be seen that these ISD figures closely validate the Dutch Model's predictions for severe exacerbations.

8.6.5 Implications in terms of COPD inpatient care

The Dutch Model provided projections in terms of exacerbations, both moderate and severe exacerbations (with a severe exacerbation requiring an admission to hospital by definition.)

Table 8.7 Estimated numbers of exacerbations as predicted by modelling for England and Scotland for selected years

England	2011	2014	2019	2024	2030
Severe exacerbations	150166	166027	193980	215725	230532
Multiplier for severe exacerbations	1	1.11	1.29	1.44	1.54
Moderate exacerbations	957334	1060303	1241200	1381835	1477448
Scotland					
Severe exacerbations	17069	18425	20327	21452	21804
Multiplier for severe exacerbations	1	1.08	1.19	1.26	1.28
Moderate exacerbations	108816	117658	130047	137380	139686

Table 8.6 shows that the model projects an 11% increase in severe exacerbations in England by 2014, becoming a 29% increase by 2019 and a 54% increase by 2030. The equivalent values for Scotland are 8% by 2014, 19% by 2019 and 28% by 2030. The annual exacerbation rates were dependent on the GOLD severity class of the patients.

These figures mean that additional hospital beds and hospital staffing will have to be found to manage this increased number of severe COPD exacerbations. Alternatives to managing COPD exacerbations in hospital will also have to be found, such as hospital at home programmes. In addition, improved detection and treatment of moderate exacerbations should help prevent severe exacerbations. The programme of influenza and pneumococcal vaccination for the COPD population also has a role in preventing exacerbations.(Menon, Gurnani et al. 2008; Cimen, Unlu et al. 2014) Maximising such initiatives will hopefully prevent the predicted growth in COPD patients requiring care for exacerbations.

8.6.6 Policy developments in COPD

The model helps to validate the recent emphasis that policymakers have put on smoking cessation.(Dept_of_Health 2014) The sensitivity analysis that considered how COPD prevalence would be affected if there were a 10% smoking quit rate achieved among current smokers at all ages, predicted a decrease in the COPD population from 1,350,000 to 1,340,000 in 2030, that is 10,000 fewer people having COPD. In Scotland the decrease was 500 fewer people with COPD than if smoking prevalence remained constant. This demonstrates that the increase in COPD can be limited if the government increases investment in effective stop-smoking services.

Other possible areas which may address the increasing clinical and economic burdens of COPD were identified by the British Thoracic Society and the Primary Care Respiratory Society in their document “Improving and Integrating Respiratory Services” IMPRESS (Williams, Baxter et al. 2012), these include the following:

1. Under use of smoking cessation.
2. Rationalising oxygen prescribing.
3. Reducing inhaler waste.
4. Overuse of high dose inhaled corticosteroids and/or combination products, such as seretide.
5. Overuse of enteric-coated prednisolone tablets for patients with COPD instead of uncoated prednisolone tablets – there is a difference in the drug tariff.

6. Overuse of hospital beds (COPD length of stay) and underuse of hospital respiratory specialist care.
7. Underuse of referral to pulmonary rehabilitation.
8. Under use of psychological support.
9. Overuse of outpatients for common conditions.
10. Under-coordination with social care.

The Managed Clinical Network (MCN) for COPD in Lothian has made some progress in addressing these policy areas.(Hewitt 2014) Their strategy has included the following elements:

1. Promotion of pulmonary rehabilitation.
2. Creation of a community respiratory team.
3. Specific investment identified for COPD.
4. Targeting training of GP practices on therapy.
5. Creation of self-management plans for COPD.
6. Updating of website resources for all professionals who care for patients with COPD.

The results to autumn 2014 in Lothian include the following:

1. Admissions for COPD exacerbations have reduced.
2. Length of stay for COPD exacerbations have reduced.
3. The total bed days for COPD have reduced.
4. 300 frequent admission patients have been identified.
5. A multi-disciplinary team has been created and meets weekly.
6. A psychologist has assessed 90 patients.
7. Repeat admissions have reduced, but not yet by the target amount.
8. The admitting unit now advises GPs to contact the community respiratory team where appropriate rather than admit.
9. The community respiratory team is now available seven days a week including early evenings.
10. Use of seretide has stabilised and costs have been contained.
11. Self-management plans are widely used by GPs and practice nurses.(Hewitt 2014)

8.7 Unanswered questions and future research

8.7.1 The COPD population and socio-economic status

It would be interesting to know how the COPD population is distributed across the socio-economic groups of the populations in England and Scotland. This is of interest with regard to health inequalities as smoking prevalence is higher among lower socio-economic groups than among higher socio-economic groups (West 2012) and premature all-cause mortality is higher among lower socio-economic groups.(ScottishGovernment 2012) Further research would help unpick the role of COPD in premature mortality.

8.7.2 Alternative modelling strategies

It may be interesting to use trends as inputs to the Dutch Model. It would be possible to use regression models to project the trend for incidence or prevalence or start and stop smoking rates over the years and have different inputs for each year. Although this is theoretically possible it would add significantly to the Model's complexity and may produce outputs which are not very different from those demonstrated in this version, especially, as the sensitivity analysis demonstrated that the level of sensitivity of the model to changes in input prevalence and smoking quit rate was not very sensitive. However, the changes would be more dramatic if trends in incidence were used, instead of point incidence as inputs, because the model is much more sensitive to changes in the input of incidence than to other inputs. This is another potential avenue for further research.

Additionally, it may be possible to use some or all of the inputs I used in this model with alternative high quality models from the systematic review, e.g. the Dynamo-HIA model(Lhachimi, Nusselder et al. 2012). As long as the inputs were for the same time frame and population, comparison of the outputs would then allow cross-validation of these models. Alternatively, some of the models from the systematic review may require subtly different inputs to enable their models to run and allow an output comparison.

8.7.3 Further work that may require new models

COPD often prevents its sufferers from working and therefore results in lost economic output.(HealthcareCommission 2006; NICE 2010) This is an indirect cost of COPD. New health economic models are required in order to quantify the lost productivity for which this accounts.

Similarly, COPD patients will require social care in addition to healthcare, in particular as they become older and frailer and suffer with co-morbidities. New models will help to

quantify the additional costs of this social care. Such costs are likely to be substantial given the projected age profile of the COPD population (See Section 8.5.2).

The Dutch Model does not take into account many of the factors that research is uncovering as being related to the development of COPD. The importance of prematurity, childhood respiratory infection, asthma and occupational lung disease in influencing risk of COPD all provide possible avenues for future research into alternative models for projecting the future patterns of COPD.(Barker, Godfrey et al. 1991; Galobardes, McCarron et al. 2008; Baines, Hsu et al. 2013; Toren and Jarvholm 2014)

8.7.4 Validation of projections with real data

As time goes on, it would be useful to validate these projections with real data. The increasing scope of routine data that is collected may eventually allow closer validation of the projection for the number of people with COPD. It may prove more difficult to validate the number of deaths from all causes among people with COPD – however this could still be possible with some large primary care routine datasets for example that from CPRD, THIN QResearch or SAIL (as described in Chapter 3).

In terms of validating the costs of COPD (see Section 8.5.2), the Dutch Model's projections appeared to be perhaps overestimates for NHS England; however, it is more likely that the NHS England PCT returns data underestimates the costs of COPD. Where available, costs for Scotland have been shown to closely validate the expenditure on severe exacerbations.

8.7.5 Re-working the Dutch Model

The Dutch Model is based on the progression of COPD through the mild, moderate, severe and very severe disease stages as related to FEV1 decline. However, the more we learn about COPD as a disease, it is increasingly suggested that having a linear basis for decline in lung function is an oversimplification. The ECLIPSE study, as described in Chapter 2, found that lung function decline in COPD is far more variable than previously thought and that exacerbations play an important role in non-linear decline in lung function.(Vestbo, Edwards et al. 2011; Vestbo, Agusti et al. 2014) There is a future need to rework the Dutch Model to enable modelling of more complex, non-linear scenarios.

In addition, with increased computing power it may be possible in the future to model individual patient trajectories rather than modelling in cohorts. This would blur the lines between population models and individual prognostic models.

8.7.6 The definition of COPD

The Dutch Model considered only those patients who have a recorded diagnosis of COPD. Future work could usefully consider the current prevalence estimates from Nacul et al (Nacul, Soljak et al. 2007) based on COPD as identified from spirometry that includes both diagnosed and undiagnosed COPD. These prevalence estimates could then be fed into the Dutch Model to project the future total population of COPD regardless of the proportion of the COPD that is diagnosed. However, using such a technique has implications for calculating cost estimates – as it may be difficult to establish a cost of treating “undiagnosed” COPD.

8.8 Conclusion to thesis

I have contributed carefully undertaken projections of the increase in the COPD population and the number of deaths among people with diagnosed COPD for England and Scotland from 2011 to 2030. I also include estimates of the direct healthcare costs for caring for this group of patients. I used a high quality modelling tool, the Dutch Model, and worked in close association with its original developers.

I believe these to be the best estimates of future COPD prevalence and costs available to guide English and Scottish policymakers. Sensitivity analyses have shown how the Model responds to changes in inputs, in that it is robust to changes in prevalence estimates but sensitive to changes in incidence estimates and mortality estimates. Some of the cost projections may be over-estimates; however, some information is missing from the NHS England PCT returns costs that were used for comparative validation.

In contrast to the highly cited GBD Study (Murray, Richards et al. 2013), my estimates project that there will be an increase in the number of people with COPD from 2011 to 2030 as opposed to the decrease in the number of COPD related DALYs from 1990 to 2010 that they showed. This difference is more due to the different measures used and the different ways of modelling burden employed in the two studies than to the different time period. I believe that my modelling is more transparent in comparison to the Global Burden of Disease and that my projections are, thus, likely to prove more helpful to policymakers.

My results suggest that governments should do more to support effective quit smoking services, especially among women, to reduce future burden of COPD. In addition, there is a need to plan to address higher prevalence of COPD among age groups over 65 years of age as part of the wider approach to managing increasing prevalence of multi-morbidity.

In order to validate and update this model there is a need to monitor the incidence, prevalence, mortality and healthcare costs of COPD in England and Scotland. Routine data from primary care needs to be collected as it has a valuable role in this process. There are sources of high quality data in England and Scotland which may be exploited in the future for similar projects.

Appendix 1.1 CPRD Original protocol

GPRD data protocol

Predicting the risk of chronic obstructive pulmonary disease: a nested case-control study

Aziz Sheikh^{1*}, Daniel Kotz^{1,2}, Colin R. Simpson²

¹University of Edinburgh, Centre for Population Health Sciences

²Maastricht University, Department of General Practice

**Principal Investigator. Correspondence: Prof. Aziz Sheikh, Allergy and Respiratory Research Group, Centre for Population Health Sciences, University of Edinburgh, Edinburgh EH8 9AG, phone: 0131 6514151, e-mail: Aziz.sheikh@ed.ac.uk*

Lay summary of research

Chronic obstructive pulmonary disease (COPD) is a major cause of chronic disease and death and results in an economic and social burden that is both substantial and increasing. Cigarette smoking is the most important risk factor of COPD. Other factors play a role in the development and progression of the disease as well, but our knowledge about these risk factors is limited.

We therefore wish to use the General Practice Research Database (GPRD) to identify all possible risk factors that are related to COPD. We will do this by undertaking a study in which we compare a random selection of 50,000 registered patients who have a clinician-recorded diagnosis of COPD with a random selection of 50,000 patients who are not recorded with such a diagnosis.

We will assess which factors from the patient's electronic health record are associated with the development of COPD. In the group of patients with COPD, we will also assess which factors are associated with death within 10 years of diagnosis.

Results from this study would enable the identification of potentially modifiable risk factors at an early stage thereby possibly preventing progression of disease and premature death.

Objectives, specific aims, and rationale

Objectives:

- The primary objective is to utilise the GPRD for developing a model that integrates smoking and other risk factors from the patient's electronic health record to predict the risk of developing COPD.
- The secondary objective is to utilise the GPRD for assessing smoking and other risk factors in patients with a recorded diagnosis of COPD to predict the 10-year-risk of death from all causes and from circulatory or respiratory disease.

Specific aims:

- The primary aim is to assess which factors, in addition to smoking, predict a first recorded diagnosis of COPD in patients aged 35 years or older.
- The secondary aim is to assess which factors, in addition to smoking, predict the 10-year-risk of death from all causes, from circulatory disease and from respiratory disease in patients with a recorded diagnosis of COPD.

Rationale:

- The efficiency of early treatment of modifiable risk factors could be enhanced if individuals at high risk of COPD could be identified at an earlier stage.
- Furthermore, the identification of risk factors of premature death in patients with a recorded diagnosis of COPD could help to identify improved treatment strategies in this group of patients.

Background

Chronic obstructive pulmonary disease (COPD) is a respiratory disease which is

characterised by progressive airflow limitation that is not fully reversible.(Global Initiative for Chronic Obstructive Lung Disease 2008) COPD is a major cause of chronic morbidity and mortality throughout the world and results in an economic and social burden that is both substantial and increasing.(Global Initiative for Chronic Obstructive Lung Disease 2008) The current worldwide prevalence of COPD is estimated to be 10%.(Buist, McBurnie et al. 2007) Projections indicate that COPD will become the third largest cause of death worldwide by the year 2020.(Murray and Lopez 1997) COPD patients not only die prematurely, they also have a decreased quality of life due to impaired physical, mental, and social functioning.(Heijmans, Spreeuwenberg et al. 2005)

COPD is a largely preventable disease.(Global Initiative for Chronic Obstructive Lung Disease 2008) The most important risk factor for its development and progression is cigarette smoking.(Global Initiative for Chronic Obstructive Lung Disease 2008) Smokers with COPD experience an accelerated decline in lung function, which cannot be fully reversed.(Fletcher and Peto 1977) Notably, almost all patients with COPD have a history of cigarette smoking, but only a minority of smokers will develop COPD; the estimated overall percentage varies between 15-50% of smokers.(Lundback, Lindberg et al. 2003) This large variability indicates the role of other risk factors in the aetiology and pathogenesis of the disease. Our appreciation of these risk factors is however as yet limited as much of the current knowledge is based on cross-sectional association studies rather than analytical studies that would allow causal relationships to be more reliably ascertained.(Global Initiative for Chronic Obstructive Lung Disease 2008) These other risk factors regarded to be associated with COPD include respiratory infections, age, sex, nutrition, and socioeconomic factors.(Mannino and Buist 2007; Global Initiative for Chronic Obstructive Lung Disease 2008)

Given the characteristics of COPD, it is reasonable to assume that the morbidity and mortality from COPD could be reduced in the future if individuals at high risk of developing the disease could be identified, and if a subsequent intervention could reduce modifiable risk factors. Concerning the first requirement, no tool currently exists to identify individuals at high risk of developing COPD. Available tools are aimed at case-finding individuals with established COPD (see for example (van Schayck, Halbert et al. 2005; Price, Tinkelman et al. 2006)).

In the GPRD, patient information (including diagnostic and disease-related data and smoking status) are well recorded. The database offers the possibility for an assessment of the association between established and novel risk factors and COPD diagnosis. Integrating these risk factors into one model would allow a valid and accurate prediction of COPD risk in individual patients screened in primary care.

Study type

The study will be hypothesis generating. The association between cigarette smoking and the development of COPD is well established. However, knowledge about other risk factors in the aetiology of the disease is limited.

Study design

We will use a nested case-control design. A random sample of 50,000 patients older than 35 years who have a recorded diagnosis of COPD will be selected as cases. For each case, one control patient with the same year of birth will be randomly selected from the same general practice (i.e., a total of 50,000 controls). Potential risk factors from the electronic health records will be compared between cases and controls.

The analysis of risk factors of mortality will be restricted to the group of 50,000 patients with a recorded diagnosis of COPD. These patients will be linked to Office for National Statistics (ONS) mortality data to assess risk factors of 10-year mortality from all causes and from circulatory disease (ICD-10 I00-I99) and respiratory diseases (ICD-10 J00-J99).

Study population

A random sample of 50,000 cases of COPD will be drawn from the GPRD.

Inclusion criteria:

- a recorded diagnosis of COPD

- age ≥ 35 years (this lower limit has been chosen because risk prediction of COPD at earlier age is unlikely to be feasible)

Exclusion criteria:

- less than one year of complete data in the medical record.

Selection of comparison group(s) or controls

For each case, one control patient will be randomly selected matched by year of birth and general practice (i.e., a total of 50,000 controls).

Inclusion criteria:

- no recorded diagnosis of COPD
- age ≥ 35 years

Exclusion criteria:

- less than one year of complete data in the medical record

Sample size/power calculation

In order to achieve a sufficient sample size for the statistical analyses of all potential risk factors, the calculation of the sample size should be based on a potential risk factor that is less frequent in the population of patients registered in GPRD. One of the potential risk factors of COPD morbidity we want to assess is a history of acute respiratory infection (ARI). In order to provide an estimate of the percentage of patients over 35 years that have been exposed to a clinician-recorded episode of ARI we interrogated the primary care database of the Primary Care Clinical Informatics Unit (PCCIU) of the University of Aberdeen. This database includes the electronic health records of patients from more than 300 Scottish primary care practices. The lifetime prevalence of a recorded diagnosis of ARI in patients over 35 years of age that had was 5%.

We based our sample size calculation on the following parameters:

- minimum odds ratio to detect = 1.1
- percentage exposed with ARI among controls = 5%
- $\alpha = 0.05$

- power = 0.9
- number of controls per case = 1 (1:1 matched study design)

This resulted in a total required sample size of 93,290 patients. Thus, the 100,000 patient records that we request from the GPRD database would be sufficient to statistically detect even a small effect of a rather infrequent risk factor of COPD.

Exposure, outcomes, and covariates

The GPRD will be used as the main source of data for this study. We will use the complete time period of the database as our observational period (1990 to current).

Exposure variables for COPD:

- sex
- socioeconomic status / deprivation score (Carstairs deprivation categories)
- ethnic category (Read Codes (9i and below, 9S and below)
- tobacco consumption (Read Codes 137 and below)
- smoking status (Read Codes 13p and below)
- smoking cessation (Read Codes 6791, 67A3, 67H1, 8CAL, 8h7i, 8HTK, 9N2k, 9OO and below)
- body mass index (Read Codes 229, 22A, 22K and below)
- acute respiratory infections (Read Codes H0 and below)
- respiratory symptoms:
 - cough (Read Codes 171, R06z, A33z)
 - wheezing and difficulty breathing (Read Codes 173 and below)
- occupational exposure to risk factor (Read Codes 0 (i.e., occupation), ZV4C)
- family history of COPD (Read Codes 12D1 and 12DZ)
- pulmonary tuberculosis (Read Codes A11, A1z, 65P2)
- physical activity (Read Codes 138 and below)
- genes: alpha-1 antitrypsin deficiency

Exposure variables for mortality (in addition to the above mentioned variables):

Prescription of COPD pharmacotherapy listed in the National Institute for Health and Clinical Excellence (NICE) guideline(NICE 2010):

- short- and long-acting bronchodilators
- short-acting beta₂-agonists
- short-acting anticholinergics
- theophylline
- inhaled and oral corticosteroids

Common pathway, complicating, and coexisting morbidities(Global Initiative for Chronic Obstructive Lung Disease 2008):

- asthma (Read Codes H33 and below)
- ischaemic heart disease (Read Codes G3 and below)
- hypertension (Read Codes G2 and below)
- heart failure (Read Codes G58 and below)
- anxiety and depression (Read Codes E200 and below, E204, E2B)

Outcome variables for COPD:

Diagnosis of COPD based on one or more of the following Read Codes (case definition):

- H3: COPD
- H31 and below: chronic bronchitis
- H32 and below: emphysema
- H36: mild COPD
- H37: moderate COPD
- H38: severe COPD
- H3y and below: other specified COPD
- H3z: COPD not otherwise specified

Emergency admission due to COPD (Read Code 8H2R)

Lung function measurements:

- pre- and post-bronchodilator Forced Expiratory Volume in one second (FEV1)
- pre- and post-bronchodilator Forced Vital Capacity (FVC)

Oxygen treatment (Read Code 6639)

Outcome variables for mortality:

- all cause mortality

- mortality from circulatory diseases (ICD-10 I00-I99)
- mortality from respiratory diseases (ICD-10 J00-J99)

Data/statistical analysis

For the primary analysis, a "baseline date" will be assigned to all cases and controls, i.e. the latest of the following dates: 35th birthday or the date of registration with the practice. For each potential risk factor of COPD diagnosis, we will use the entry of the risk factor in the electronic health record at the date which is closest to the baseline date. The association between potential risk factors and COPD will be analysed using a multiple logistic regression model in which COPD diagnosis (case vs. control) will be regressed on all potential risk factors. Risk factors that are not associated with COPD diagnosis (based on a p-value of >0.05) will be removed from the model one by one (i.e., backward stepwise logistic regression).

For the secondary analysis, a "COPD diagnosis date" will be assigned to all cases, i.e. the date of the first recorded diagnosis of COPD confirmed by spirometry including reversibility testing (where available). For each potential risk factor of mortality, we will use the entry which is closest to the COPD diagnosis date. The association between potential risk factors and mortality will be analysed using three separate multiple logistic regression models in which the 10-year-risk of: (1) all cause mortality, (2) mortality from circulatory diseases (ICD-10 I00-I99), and (3) mortality from respiratory diseases (ICD-10 J00-J99) will be regressed on all potential risk factors. Risk factors that are not associated with mortality (based on a p-value of >0.05) will be removed from the model one by one (i.e., backward stepwise logistic regression).

In both analyses, we will assess the level of missing data and will use multiple imputation (MI) techniques to replace missing values where necessary. MI is a statistical technique that allows patients with incomplete data to still be included in analyses, thereby making full use of all available data. This will increase the power of the analyses and may produce models that are more statistically reliable and applicable within clinical practice.(Clark and Altman 2003)

All analyses will be carried out by Prof. Aziz Sheikh in collaboration with Dr. Daniel Kotz and Dr. Colin Simpson.

Patient or user group involvement

The Edinburgh Allergy & Respiratory Research Group has its own Consumer Involvement Group, which will advise on this project. Results of the study will be fed back to the British Lung Foundation.

Limitations of the study design, data sources, and analytic methods

Our case definition is dependent on a clinician-recorded diagnosis. Thus, we have to assume that the absence of a recorded diagnosis of COPD is equivalent to the person not having the disease, which is not always the case and may lead to some degree of misclassification. This problem is amplified by the fact that COPD is not easily diagnosed; a proper diagnosis can only be made by spirometry with reversibility testing. Furthermore, the quality of recording of disease has probably varied over time and has improved in the last years which may bias the association between risk factors and COPD.

This study is based on routinely collected primary care data. An important disadvantage of this approach is that not all potential risk factors associated with the development of COPD are recorded in GPRD. Furthermore, the level of exposure to some of the risk factors (i.e., the dose-response relationship) is insufficiently recorded. This will result in a reduction of variation in COPD morbidity and mortality that can be explained by the model. On the other hand, using routinely recorded data for the development of the model maximises the usability for primary care clinicians who want to identify patients at high risk of developing COPD.

Other potential limitations are inherent in the study design; a case-control study is commonly more vulnerable to bias (in particular confounding) than a prospective design.

Plans for disseminating and communicating study results

We plan to disseminate our results via publication in an international peer reviewed journal in the field of respiratory research (e.g., European Respiratory Journal). We also plan to submit an abstract to the annual congress of the European Respiratory Society. Any publication will be in accordance with ICMJE and STROBE guidelines (www.equator-network.org).

Ethical review

The GPRD Group has obtained ethical approval from a Multi-centre Research Ethics Committee (MREC) for all purely observational research using GPRD data.

National Research Register (NRR)

This study has not yet been registered at the NRR, but will be once the study is approved.

References

- 1 Global Initiative for Chronic Obstructive Lung Disease. Global strategy for the diagnosis, management, and prevention of chronic pulmonary disease. Updated 2008. <http://www.goldcopd.com> [accessed 1 December 2009]: NHLBI/WHO, 2008
- 2 Buist AS, McBurnie MA, Vollmer WM, et al. International variation in the prevalence of COPD (the BOLD Study): a population-based prevalence study. *Lancet* 2007; 370:741-750
- 3 Murray CJ, Lopez AD. Alternative projections of mortality and disability by cause 1990-2020: Global Burden of Disease Study. *Lancet* 1997; 349:1498-1504
- 4 Heijmans MJWM, Spreeuwenberg P, Rijken PM. Monitor Zorg- en Leefsituatie van mensen met astma en mensen met COPD. Trends en ontwikkelingen over de periode 2001 - 2004. [Trends in the living condition and healthcare of people with asthma and COPD 2001 - 2004]. Utrecht: NIVEL, 2005

- 5 Fletcher C, Peto R. The natural history of chronic airflow obstruction. *BMJ* 1977; 1:1645-1648
- 6 Lundback B, Lindberg A, Lindstrom M, et al. Not 15 but 50% of smokers develop COPD?--Report from the Obstructive Lung Disease in Northern Sweden Studies. *Respir Med* 2003; 97:115-122
- 7 Mannino DM, Buist AS. Global burden of COPD: risk factors, prevalence, and future trends. *Lancet* 2007; 370:765-773
- 8 van Schayck CP, Halbert RJ, Nordyke RJ, et al. Comparison of existing symptom-based questionnaires for identifying COPD in the general practice setting. *Respirology* 2005; 10:323-333
- 9 Price DB, Tinkelman DG, Halbert RJ, et al. Symptom-based questionnaire for identifying chronic obstructive pulmonary disease in smokers. *Respiration* 2006; 73:285-295
- 10 Chronic obstructive pulmonary disease. Management of chronic obstructive pulmonary disease in adults in primary and secondary care. London: National Institute for Health and Clinical Excellence (NICE), 2004
- 11 Clark TG, Altman DG. Developing a prognostic model in the presence of missing data: an ovarian cancer case study. *J Clin Epidemiol* 2003; 56:28-37

Appendix 1.2 CPRD Amendment protocol

Predicting the risk of chronic obstructive pulmonary disease: a nested case-control study

Protocol Amendment: predicting the future disease burden of chronic obstructive pulmonary disease

Aziz Sheikh^{1*}, Daniel Kotz^{1,2}, Colin R. Simpson¹, Susannah McLean¹, Sarah Wild¹, Maureen Rutten-van Mölken³, Martine Hoogendoorn³.

¹University of Edinburgh, Centre for Population Health Sciences

²Maastricht University, Department of General Practice

³Erasmus University, Rotterdam, Institute for Medical Technology Assessment

**Principal Investigator. Correspondence: Prof. Aziz Sheikh, Allergy and Respiratory Research Group, Centre for Population Health Sciences, University of Edinburgh, Edinburgh EH8 9AG, phone: 0131 6514151, e-mail: Aziz.sheikh@ed.ac.uk*

Data was made available under MRC licence agreement no additional funding available for amendment to protocol, no additional data required, just additional analysis of dataset.

Lay summary of research

Chronic obstructive pulmonary disease (COPD) is a major cause of chronic disease and death and results in an economic and social burden that is both substantial and increasing. Cigarette smoking is the most important risk factor of COPD. Other factors play a role in the development and progression of the disease as well, but our knowledge about these risk factors is limited.

We plan to use the General Practice Research Database (GPRD) to identify all possible risk factors that are related to COPD. We will do this by undertaking a study in which we compare a random selection of 50,000 registered patients who have a clinician-recorded diagnosis of COPD with a random selection of 50,000 patients who are not recorded with such a diagnosis. We will assess which factors from the patient's electronic health record are associated with the development of COPD. In the group of patients with COPD, we will also assess which factors are associated with death within 10 years of diagnosis. Results from this study would enable the identification of potentially modifiable risk factors at an early stage thereby possibly preventing progression of disease and premature death.

In addition

We would use specific measures of the disease burden of COPD derived from this database to feed into a Markov type model (Hoogendoorn, Rutten-van Molken et al. 2005; Hoogendoorn, Rutten-van Molken et al. 2011) developed by Erasmus University in order to predict the costs and burden of COPD patients 5 and 10 years in the future. Such predictions would help healthcare planners ensure that there were adequate resources made available for caring for people with COPD.

Amendment objective, aims and rationale

Objective

The amendment objective is to quantify disease burden derived from this UK cohort and then to compare with similar disease burdens computed from Dutch and Scottish cohorts.

Specific aims:

The amendment aim is to predict disease burden from COPD in 5 and 10 years time using a pre-existing model(Hoogendoorn, Rutten-van Molken et al. 2011) developed by Erasmus University, Rotterdam, using parameters derived from this database.

These predictions can then be compared with predictions made by other means or in other countries.

Sensitivity analysis to the parameters derived from the database will also be undertaken

Rationale:

Quantification and prediction of the disease burden in the future will better enable planning of healthcare resources.

Background

Chronic obstructive pulmonary disease (COPD) is a respiratory disease which is characterised by progressive airflow limitation that is not fully reversible.(GOLD 2010) COPD is a major cause of chronic morbidity and mortality throughout the world and results with a large economic and social burden. The 2010 World Health Organization Global Burden of Disease study lists COPD as the third largest cause of death worldwide.(Lozzano, Naghavi et al. 2012) COPD patients not only die prematurely, they also have a decreased quality of life due to impaired physical, mental, and social functioning.(Murray, Vos et al. 2012)

The most important risk factor for COPD development and progression is cigarette smoking.(GOLD 2010) Smokers with COPD experience an accelerated decline in lung function.(Peto, Speizer et al. 1983)

As populations are ageing, the prevalence of long term conditions such as COPD is increasing. This means that more resources are required to provide healthcare for the COPD population.(DoH 2010) To help plan these resources a dynamic COPD population model has been developed by Erasmus University.(Hoogendoorn, Rutten-van Molken et al. 2005; Hoogendoorn, Rutten-Van Molken et al. 2011) This model includes provision for calculating

how many people will develop COPD and at what severity if they continue to smoke or stop smoking later in life. It also includes provision for modelling the healthcare needs of these people from data on the rate of exacerbations and costs of healthcare.(Hoogendoorn, Rutten-Van Molken et al. 2011)

In order to plan healthcare for people with COPD in the future, these calculations as to the size of the disease burden are useful. In addition, the effects of interventions such as smoking cessation campaigns may be computed.

The CPRD will allow calculation of UK-specific parameters for, lung function decline, mortality and exacerbations which can be fed into the model. Then predictions can be made on this basis. The sensitivity of the predictions to these inputs will also be tested.

Study type

Providing parameters for an existing model. Hypothesis generating in this context.

Study design Methodology

A cohort of 50,000 cases in the case-control database who have COPD will be used to generate parameters to feed into an established Markov Model, by Erasmus University.

Study population

A random sample of 50,000 cases of COPD has been drawn from the CPRD.

Inclusion criteria:

a recorded diagnosis of COPD

age ≥ 35 years (this lower limit has been chosen because risk prediction of COPD at earlier age is unlikely to be feasible)

Exclusion criteria:

less than one year of complete data in the medical record.

Sample Size

The original Erasmus study used a cohort of only 5887 patients (the lung health study 2000) to model the annual decline of FEV1% in the following model:

Annual decline FEV1% predicted~ intercept +year+smoking status+sex+age+ baseline FEV1 + interaction effects

With this much larger cohort of 50,000 patients the annual decline of FEV1 would be ten times more powerfully modelled.

Exposures, Outcomes and covariates:

This following table demonstrates the inputs to the Erasmus model and a * indicates the items which will be calculated using this CPRD data.

Ref	Data title	Scottish Run	UK wide run
A1	General Population Data	National Records Scotland	Office National Statistics (England and Wales)
A2	Proportions of population that smoke, former smoker and never smoker by age and gender	Scottish Health Survey, over 16. Under 16 SSALSUS	Age over 16, Smoking toolkit study; (England) Under 16 General Lifestyle Survey, (England)
A3a	GP prevalence of COPD, age and gender specific	Data from information services division Scotland	*1

A3b	GP incidence of COPD age and gender specific	Lothian cohort data with practice populations to calculate crude incidence	*2
A4	Relative risks of smokers and nonsmokers	Use Rotterdam data,	Use Rotterdam data,
A5	Age and sec specific start stop and restart probabilities for smoking	Use Rotterdam data,	Use Rotterdam data,
A6	Random effects model to predict lung function decline	Recalculate model using COPD Lothian data	*3
A7	COPD related maintenance costs	Use Rotterdam data	Use Rotterdam data
A8	Mean utility scores by COPD severity stage	Use Rotterdam data	Use Rotterdam data
A9	All cause mortality, mortality attributable to COPD mortality and mortality from other causes, by age and sex (per 1000 COPD patients) (i.e. what do COPD patients die of?)	Lothian cohort data will allow to form a case fatality profile.	*4

*1 In order to calculate the prevalence of COPD from the cohort of 50000 patients, a denominator of the total population of the practices from which the 50000 were drawn ideally, per year or at a fixed point in time.

*2 In order to calculate the incidence rate of COPD, it would have to be established whether the 50,000 cases were incident cases and over what time period they were discovered. This may be possible using the second half of the database with the first half as a “look-back

period” to eliminate previous diagnoses. The denominator would then be the total person time for the practice populations which contributed these cases.

*3 see above

*4 As these 50,000 patients with a recorded diagnosis of COPD are linked to Office for National Statistics (ONS) mortality, it will be possible to develop an age stratified case fatality profile of the population. A profile of age at death, sex and whether death was due to COPD or due to another cause will also be an input to the Erasmus model.

Data/statistical analysis

The parameters below would also be modelled/calculated using the COPD cohort.

A	Rate of Exacerbation by COPD stage	Individual patient data calculation, potentially a fresh model, as the CPRD is linked to HSE for exacerbations the rate of exacerbations at different times of year, by sex and smoking status can be modelled.
B	Case fatality rate	Develop age stratified case fatality profile of either Lothian Cohort or CPRD from linked death certificate data rather than modelling
C	Decline in FEV1 due to exacerbation	Calculate using CPRD or Lothian cohort (Average lung function decline per year/no. of exacerbations per year)= average lung function decline per exacerbation

All analyses will be carried out by a collaboration of Prof. Aziz Sheikh, Dr Susannah Mclean, Dr. Daniel Kotz, Dr. Colin Simpson, Prof Sarah Wild, Dr Martine Hoogendoorn and Prof Maureen Rutten-van Mólken.

Patient or user group involvement

The Edinburgh Allergy & Respiratory Research Group has its own Consumer Involvement Group, which will advise on this project.

Limitations of the study design, data sources, and analytic methods

The case definition is dependent on a clinician-recorded diagnosis. Thus, we have to assume that the absence of a recorded diagnosis of COPD is equivalent to the person not having the disease, which is not always the case and may lead to some degree of misclassification. This problem is amplified by the fact that COPD is not easily diagnosed; a proper diagnosis can only be made by spirometry with reversibility testing.(Holleman and Simel 1995) There is also debate regarding the precise spirometry threshold to use.(Celli, Halbert et al. 2003; Roberts, Farber et al. 2006; Garcia-Rio, Soriano et al. 2011)

This study is based on routinely collected primary care data.(Jones, Dickson-Spillmann et al. 2008) These vary in their accuracy and completeness and techniques may have to be used to account for missing data, for example, data imputation – a standard statistical technique where missing data re estimated from the statistical distributions and other features of the available data for the same parameter.

Another limitation is the complexity of the Markov model. As these models often build up layers of assumptions, they become vulnerable to criticism due to accumulation of errors. However, using this dataset would be a breakthrough in terms of validating the Erasmus model as it has not so far been trialled with the data of a country other than the Netherlands. Also as the Netherlands and the UK have similar provision of healthcare, interesting comparisons will, therefore, be made and critiqued.

Plans for disseminating and communicating study results

We plan to disseminate our results via publication in an international peer reviewed journal in the field of respiratory research (e.g., European Respiratory Journal). We also plan to submit an abstract to the annual congress of the European Respiratory Society. Any publication will be in accordance with ICMJE guidelines.

Ethical review

The GPRD Group has obtained ethical approval from a Multi-centre Research Ethics Committee (MREC) for all purely observational research using GPRD data.

References

1. Hoogendoorn M, Rutten-van Molken MPMH, Hoogenveen RT, van Genugten MLL, Buist AS, Wouters EFM, et al. A dynamic population model of disease progression in COPD. *Eur Respir J* 2005;26(2):223-33.
2. Hoogendoorn M, Rutten-van Molken MPMH, Hoogenveen RT, Al MJ, Feenstra TL. Developing and applying a stochastic dynamic population model for chronic obstructive pulmonary disease. *Value Health* 2011;14(8):1039-47.
3. GOLD. Global Initiative for Chronic Obstructive Lung Disease: Pocket guide to COPD diagnosis management and prevention. 2010.
4. Lozano R, Naghavi M, Foreman K, Lim S, et al. E. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012;380(9859):2095-128. doi: 10.1016/S0140-6736(12)61728-0.
5. Murray CJ, Vos T, Lozano R, Naghavi M, et al. E. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012;380(9859):2197-223. doi: 10.1016/S0140-6736(12)61689-4.
6. Peto R, Speizer FE, Cochrane AL. The relevance in adults of air-flow obstruction, but not of mucus hypersecretion, to mortality from chronic lung disease. Results from 20 years of prospective observation. *Am Rev Respir Dis*. 1983;128(3):491-500.
7. DoH. Improving the health and wellbeing of people with long term conditions: Department of Health, 2010.

8. Hoogendoorn M, Rutten-Van Molken MPMH, Hoogenveen R, Maiwenn J, Feenstra TL. Developing and applying a stochastic dynamic population model for chronic obstructive pulmonary disease. *Value Health*. 2011;14(8):1039-47. Epub 2011 Sep 22.
9. Holleman DR, Jr., Simel DL. Does the clinical examination predict airflow limitation? *Jama* 1995;273(4):313-9.
10. 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(1):200-6.
11. Garcia-Rio F, Soriano JB, Miravittles M, Munoz L, Duran-Tauleria E, Sanchez G, et al. Overdiagnosing subjects with COPD using the 0.7 fixed ratio: correlation with a poor health-related quality of life. *Chest* 2011;139(5):1072-80.
12. Celli BR, Halbert RJ, Isonaka S, Schau B. Population impact of different definitions of airway obstruction. *Eur Respir J* 2003;22(2):268-73.
13. Jones RC, Dickson-Spillmann M, Mather MJ, Marks D, Shackell BS. Accuracy of diagnostic registers and management of chronic obstructive pulmonary disease: the Devon primary care audit. *Respir Res* 2008;9:62.

Appendix 1.3 CPRD agreement for provision of incidence and prevalence data



Date: 11 September 2013

Predicting the risk of chronic obstructive pulmonary disease: a nested case-control study

Prepared by: Helen Strongman, CPRD

Distribution: Susannah Maclean, University of Edinburgh

Objective

To estimate the incidence and point prevalence of COPD in the general population in 2011.

Source Population

The extraction population will comprise of all patients in CPRD GOLD from the CPRD build used in the original research; April 2012 (n=13,519,941).

Practice Criteria

The study cohort will be drawn from of all up-to-standard English practices that were participating in the linkage programme when this data was extracted (6,989,019).

Study Population

The study population will consist of all patients in eligible practices who:

- Are acceptable (n=6,112,231)
- Are male or female (n=6,112,056)
- Were alive and registered in an up-to-standard practice on and for at least 18 months prior to 01/01/2011 (n=2,152,744)
- Were aged at least 35 in 2011 (2011 – birthyear \geq 35) (n=1,346,902)
- Have not been diagnosed with COPD at an unspecified date (nn=1,346,895)

The age of the patient will be the calculated by subtracting the year of birth of the patient from 2011.

The following criteria that were used to define the cases and controls in the original study will not be applied:

- have a valid postcode/SOA (for socioeconomic status)
- had at least five years of up-to-standard follow up data in the CPRD GOLD database.

Study design

Aggregated data based on cohort analysis

Point prevalence of COPD

Denominator

Denominator data will include all patients in the study population.

Numerator

The numerator will be all patients in the denominator population who have a first record of COPD at any time in their patient record prior to 01/01/2011.

Point prevalence calculation

The point prevalence of COPD will be calculated by dividing the number of patients in the numerator by the number of patients in the denominator. Prevalence data will be presented per 100,000 patients with 95% CI calculated using the Poisson distribution. Results will be stratified by gender and age. The data tables will include the numerator and the denominator.

Incidence of COPD

Numerator

The number of patients with a first mention of COPD in 2011 whilst registered at an up-to-standard practice.

Denominator

Denominator data will consist of the sum of the person years of up-to-standard follow-up in 2011 of all at risk patients in the study population.

The start of observation will be the 01 January 2011. The end of observation will be defined as the minimum of the practice last collection date, the patient transfer out date, the CPRD derived death date, the first record of COPD (index date) or 31 December 2011.

Incidence rate calculation

The incidence rate of COPD in 2011 will be calculated by dividing the number of incident cases by total person time between 01/1/2011 and 31/12/2011 in the denominator population. Incidence data will be presented per 100,000 patient years with 95% CI calculated using the Poisson distribution. Incidence rate data will be stratified by gender and age. The data tables will include the numerator and the denominator.

Note on small cell counts

Due to the possibility of deductive disclosure, CPRD policy is that no cell should contain < 5 events when data is reported.

Appendix 1.4 Covering letter to CPRD ISAC for approval

Centre for Population Health
Studies
University of Edinburgh
Medical Quad
Teviot Place
Edinburgh
EH8 9AG

Tel 0131 650 9242
Email
Susannah.mclean@ed.ac.uk

Dear ISAC,

Re: Amendment to CPRD protocol “Predicting the risk of COPD a nested case-control study”

Our research group led by Prof Aziz Shiekh, previously submitted a protocol for data for the above study under the MRC funding scheme. We received the data and Daniel Kotz has recently commenced analysis of it.

We would now like to submit an amendment to this protocol to enable further analysis of this dataset by a wider team in order to calculate some parameters for a COPD health economics model.

This is a linked project because it covers a different aspect of projection in COPD. The primary aspect was a plan to use a nested case control study to predict outcomes in COPD. This secondary aspect is to consider the health economics needs of a simulated population with COPD. The model is a well established health economics model developed by Erasmus University, the Netherlands. Data from the “cases” will be used as a cohort in order to

populate certain parameters of the model including the rate of decline of FEV1 with time and with exacerbations. Please see protocol for further details.

The protocol also includes a proposal to include in this analysis an estimation of both the prevalence and incidence of COPD in the UK dataset, for which some additional data will be required. Whether or not we finalise this request for additional data will depend on the costing as some of the data may have already been extracted as part of the preliminary work-up to the data that we previously received.

With the above exception, this protocol amendment is essentially an application for permission to extend and publish further related analysis on the same data which we already possess and will not require further work by CPRD, as such we anticipate that no further costs will be incurred.

Thank you for considering this proposal.

Yours faithfully

Susannah McLean

On behalf of the research team at University of Edinburgh, Centre for Population health studies.

Appendix 1.5 ISAC Application form

ISAC APPLICATION FORM: PROTOCOLS FOR RESEARCH USING THE GPRD DATA AMENDMENT TO PROTOCOL

ISAC use only:		IMPORTANT
Protocol Number	If you have any queries, please contact ISAC Secretariat:
Date submitted	ISAC@gprd.com

1. Study Title Predicting the risk of chronic obstructive pulmonary disease: a nested case-control study, amendment.
--

2. Does this protocol describe a purely observational study using GPRD data (this may include the review of anonymised free text)?
Yes <input checked="" type="checkbox"/> No <input type="checkbox"/>

3. Does this protocol also seek access to data held under the GPRD Data Linkage Scheme?
Yes <input checked="" type="checkbox"/> No <input type="checkbox"/>

4. If you are seeking access to data held under the GPRD Data Linkage Scheme, please select the source/s of linked data being requested.
<input type="checkbox"/> Hospital Episode Statistic [†] <input type="checkbox"/> Cancer Registry Data * <input type="checkbox"/> MINAP*

Townsend Score
 ONS Mortality Data
 Other: *(please specify)*

** Please note that access to these data sources are not covered under the GPRD-MRC Scheme.*

† Please note that only limited access to this data source is covered under the GPRD-MRC Scheme.

5. If you are seeking access to data held under the GPRD Data Linkage Scheme, have you discussed your request with a member of the Research team?

Yes No

If *No*, please contact the GPRD Research Team to discuss your requirements before submitting your application.

6. Does this protocol involve requesting any additional information from GPs?

Yes No

If yes, please indicate what will be required:

Completion of questionnaires by the GP* Yes No

Provision of anonymised records (e.g. hospital discharge summaries) Yes No

Other (please describe)

**any questionnaire for completion by GPs needs to be approved by ISAC before being sent out for completion.*

GUIDANCE ON ANSWERING QUESTIONS 4-6:

These questions must be completed by all applicants. You should note the following:

(i) If you have answered NO to question 2, may need to seek separate ethics approval from an NHS Research Ethics Committee for this study. The ISAC will provide advice on whether this may be needed.

(ii) If you have answered YES answered to question 2 above and you will be using data obtained from the GPRD Group at the MHRA, this study does not require separate ethics approval from an NHS Research Ethics Committee.

If you will be using data obtained from EPIC, you will need to consult the data provider regarding their arrangements for obtaining ethics approval for the study.

NB: Answering YES to question 2 means that the answers to questions 7-9 should all be NO. If any of the answers below are YES please review your answer to question 2 as it should be NO.

7. Has this protocol been peer reviewed by another Committee? Yes <input type="checkbox"/> No <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>
<i>If yes, please state in your protocol the name of the reviewing Committee/s and provide an outline of the review process and outcome on final review.</i>

8. Does the study involve linking to patient <i>identifiable</i> data from other sources? Yes <input type="checkbox"/> No <input checked="" type="checkbox"/>
--

9. Does this study require contact with patients in order for them to complete a questionnaire? Yes <input type="checkbox"/> No <input checked="" type="checkbox"/>
--

10. Does this study require contact with patients in order to collect a sample? Yes <input type="checkbox"/> No <input checked="" type="checkbox"/> If yes, please state what will be collected

11. Type of Study (<i>please tick one box below</i>)
--

Adverse Drug Reaction	<input type="checkbox"/>	Drug Use	<input type="checkbox"/>	Disease Epidemiology	<input checked="" type="checkbox"/>
Pharmacoeconomic	<input type="checkbox"/>	Drug Effectiveness	<input type="checkbox"/>	Other	<input type="checkbox"/>

12. Data source *(please tick one box below)*

GPRD Division at MHRA	<input type="checkbox"/>	Other	<input type="checkbox"/>
		<i>(please specify)</i>	

Full Feature on-line access

Ad hoc dataset

MRC dataset

Other commissioned study

13. Financial Sponsor of study

MRC* Pharmaceutical Industry *(please specify)*

Government / NHS *(please specify)* Other *(please specify)*

None

* Tick this box if you wish to access GPRD data under the MRC licence with GPRD. It is expected that if you use the MRC licence, no other direct commercial/public sector funding for this study will be sought/has been applied for or is in place. If funding is in place, but does not cover the use or extraction of GPRD data, please tick the boxes for relevant funding sources (including MRC) and provide details in the protocol of why funding under the MRC licence is required.

14. Is the study intended for
 Publication in peer reviewed journals Presentation at scientific conference

Presentation at company/institutional meetings Other

15. Principal Investigator (full name, job title, organisation & e-mail address for correspondence regarding this protocol)

Prof. Aziz Sheikh, Professor of Primary Care Research & Development, University of Edinburgh,
phone: 0131 6514151, e-mail: Aziz.sheikh@ed.ac.uk

16. Affiliation (full address)

Allergy and Respiratory Research Group, Allergy and Respiratory Research Group, Centre for Population Health Sciences - General Practice Section, University of Edinburgh, Doorway 3, Teviot Place, Edinburgh EH8 9AG

17. Type of Institution (*please tick one box below*)

Academia	<input checked="" type="checkbox"/>	Research Service Provider	<input type="checkbox"/>	Pharmaceutical Industry	<input type="checkbox"/>
NHS	<input type="checkbox"/>	Government Departments	<input type="checkbox"/>	Others	<input type="checkbox"/>

18. Experience/expertise available

Please complete the following questions to indicate the experience/expertise available within the team of researchers actively involved in the proposed research, including analysis of data and interpretation of results

	Previous GPRD Studies	Publications using GPRD data
None	<input type="checkbox"/>	<input type="checkbox"/>
1-3	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
> 3	<input type="checkbox"/>	<input type="checkbox"/>

	Yes	No
Is statistical expertise available within the research team?	<input checked="" type="checkbox"/>	<input type="checkbox"/>

If yes, please outline level of experience

Expertise in applied statistics is available from the research group of Epidemiology and Statistics which is part of the Centre for Population Health Sciences

Is experience of handling large data sets (>1 million records)
available within the research team?

If yes, please outline level of experience

Prof. Aziz Sheikh and Dr. Colin Simpson have long standing experience from working with large data sets. They have published several papers using data from GPRD, QRESEARCH, and the Primary Care Clinical Informatics Unit of the University of Aberdeen.

Is UK primary care experience available within the research team?

If yes, please outline level of experience

Prof. Aziz Sheikh and Dr. Colin Simpson are experienced researchers in the field of UK primary care.

19. Other collaborators (if applicable: *please list names and affiliations of all collaborators*)

Dr. Daniel Kotz, CAPHRI School for Public Health and Primary Care, Maastricht University Medical Centre

Dr. Colin Simpson, Centre for Population Health Sciences, University of Edinburgh

Prof. Sarah Wild, Centre for Population Health Sciences, University of Edinburgh

Dr Susannah McLean, Centre for Population health Sciences University of Edinburgh

Prof Rutten van Molken, Erasmus University, Rotterdam

Dr Martine Hoogendoorn, Erasmus University, Rotterdam

20. Protocol's Author (if different from PI)

Susannah Mclean for amendment to protocol

Protocol content checklist

In order to help ensure that protocols submitted for review contain adequate information for protocol evaluation, ISAC have produced instructions on the content of protocols for research using GPRD data. These instructions are available on the GPRD website (www.gprd.com/ISAC). All protocols using GPRD data which are submitted for review by ISAC must contain information on the areas detailed in the instructions. IF you do not feel that a specific area required by ISAC is relevant for your protocol, you will need to justify this decision to ISAC.

Applicants must complete the checklist below to confirm that the protocol being submitted includes all the areas required by ISAC, or to provide justification where a required area is not considered to be relevant for a specific protocol. Protocols will not be circulated to ISAC for review until the checklist has been completed by the applicant.

Please note, your protocol will be returned to you if you do not complete this checklist, or if you answer ‘no’ and fail to include justification for the omission of any required area.

Required area	Included in protocol?		If no, reason for omission
	Yes	No	
<i>Lay Summary (max.200 words)</i>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
<i>Objective, specific aims and rationale</i>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
<i>Background</i>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
<i>Study Type</i>			
<i>Hypothesis Generating</i>		<input type="checkbox"/>	

<i>Hypothesis Testing</i>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
<i>Study Design Methodology</i>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
<i>Study population, including estimate of expected number of relevant patients in the GPRD</i>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
<i>Sample size/ power calculation</i>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
<i>Selection of comparison group(s) or controls</i>		<input checked="" type="checkbox"/>	<i>No controls required for this analysis</i>
<i>Exposures, outcomes and covariates</i>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
<i>Data/ Statistical Analysis</i>			
<i>Hypothesis Generating</i>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
<i>Hypothesis Testing</i>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
<i>Patient/ user group involvement [†]</i>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
<i>Limitations of the study design, data sources and analytic methods</i>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
<i>Plans for disseminating and communicating study results</i>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	

[†] *It is expected that many studies will benefit from the involvement of patient or user groups in their planning and refinement, and/or in the interpretation of the results and plans for further work. This is particularly, but not exclusively true of studies with interests in the impact on quality of life. Please indicate whether or not you intend to engage patients in any of the ways mentioned above.*

*ISAC strongly recommends that researchers using GPRD consider registering as a NRR data provider in order that others engaged in research within the UK can be made aware of current works. The **National Research Register (NRR)** is a register of ongoing and recently completed research projects funded by, or of interest to, the United Kingdom's National Health Service. Information on the NRR is available on www.nrr.nhs.uk.*

Please Note: Registration with the NRR is entirely voluntary and will not replace information on ISAC approved protocols that are published in summary minutes or in the ISAC annual report.

Appendix 1.6 ISAC Protocol approval

ISAC EVALUATION OF PROTOCOLS FOR RESEARCH INVOLVING CPRD DATA

FEED-BACK TO APPLICANTS

CONFIDENTIAL <i>by</i>	
<i>e-mail</i>	
PROTOCOL NO:	10_084RA
PROTOCOL TITLE:	Predicting the risk of chronic obstructive pulmonary disease: a nested case-control study
APPLICANT:	Prof. Aziz Sheikh, Professor of Primary Care Research & Development, Allergy and Respiratory Research Group, Allergy and Respiratory Research Group, Centre for Population Health Sciences - General Practice Section, University of Edinburgh
APPROVED <input checked="" type="checkbox"/>	APPROVED SUBJECT TO MINOR AMENDMENT (resubmission not required) <input type="checkbox"/>
	REVISION/ RESUBMISSION REQUESTED <input type="checkbox"/>
INSTRUCTIONS: <i>Please include your response/s to the Reviewer's feedback below <u>only</u> if you are required to Revise/ Resubmit your protocol.</i> <i>Protocols with an outcome of 'Approved' or 'Approved subject to minor amendments' <u>do not</u> require resubmission to the ISAC</i>	
REVIEWER COMMENTS: The proposed amendment to Protocol 10-084RA is approved.	
DATE OF ISAC FEEDBACK:	05 April 2013

REJECTED

DATE OF APPLICANT FEEDBACK:	
------------------------------------	--

Appendix 2.1 Level 1 Ethics approval from UoE

University of Edinburgh,
Centre for Population Health Sciences
RESEARCH ETHICS SUBGROUP

Self-Audit Checklist for Level 1 Ethical Review for PGR projects

See Intra website for further information: <http://www.cphs.mvm.ed.ac.uk/intra/research/ethicalReview.php>

NOTE to student: Completion of this form should be under the oversight of your supervisor. A good strategy would be to complete a draft as best you can, then discuss with your supervisor before completing a final copy for your supervisor to sign.

Proposed Project (State research question and topic area, and briefly describe method/ data. Specify also countries in which data will be collected.):

Research Aim: To develop population projections for the future disease burden (in 5 & 10 years from 2011) from chronic obstructive pulmonary disease (COPD) in Scotland and England.

Methods: This will be done using secondary analysis of routinely collected large datasets of anonymised data. There will be a computer & economic modelling process.

1. Bringing the University into disrepute

Is there any aspect of the proposed research which might bring the University into disrepute?

YES/ NO

2. Data protection and consent

Are there any issues of DATA PROTECTION or CONSENT which are NOT adequately dealt with via established procedures?

YES/ NO

These include well-established sets of undertakings. For example, a 'No' answer is justified only if:

- There is compliance with the University of Edinburgh's Data Protection procedures (see www.recordsmanagement.ed.ac.uk);
- Respondents give consent regarding the collection, storage and, if appropriate, archiving and destruction of data;
- Identifying information (eg consent forms) is held separately from data;
- There is Caldicott Guardian approval for (or approval will be obtained prior to) obtaining/ analysing NHS patient-data.
- There are no other special issues arising about confidentiality/consent.

3. Study participants

Will a study researcher be in direct contact with participants to collect data, whether face-to-face, or by telephone, electronic means or post, or by observation? (eg interviews, focus groups, questionnaires, assessments)

YES/ NO

4. Duty to disseminate research findings

Are there issues which will prevent all relevant stakeholders* having access to a clear, understandable and accurate summary of the research findings if they wish?

YES/ NO

* If, and only if, you answered 'yes' to 3 above, 'stakeholders' includes participants in the research

5. Moral issues and Researcher/Institutional Conflicts of Interest

Are there any SPECIAL MORAL ISSUES/CONFLICTS OF INTEREST?

YES/ NO

- An example of conflict of interest for a researcher would be a financial or non-financial benefit for him/herself or for a relative or friend.
- Particular moral issues or concerns could arise, for example where the purposes of research are concealed, where respondents are unable to provide informed consent, or where research findings could impinge negatively/ differentially upon the interests of participants.
- Where there is a dual relationship between researcher and participant (eg where research is undertaken by practitioners so that the participant might be unclear as to the distinction between 'care' and research)

6. Potential physical or psychological harm, discomfort or stress

- (a) Is there a FORSEEABLE POTENTIAL for PSYCHOLOGICAL HARM or STRESS for participants?
- (b) Is there a FORSEEABLE POTENTIAL for PHYSICAL HARM or DISCOMFORT for participants?
- (c) Is there a FORSEEABLE RISK to the researcher?

YES/ NO
 YES/ NO
 YES/ NO

Examples of issues/ topics that have the potential to cause psychological harm, discomfort or distress and should lead you to answer 'yes' to this question include, but are not limited to: relationship breakdown; bullying; bereavement; mental health difficulties; trauma / PTSD; violence or sexual violence; physical, sexual or emotional abuse in either children or adults.

7. Vulnerable participants

Are any of the participants or interviewees in the research considered to be vulnerable? e.g. children and young people under age of 16, people who are in custody or care, marginalised/stigmatised groups

YES/ NO

8. Protection of research subject confidentiality

Are there any issues of CONFIDENTIALITY which are NOT adequately handled by normal tenets of confidentiality for academic research?

YES/ NO

These include well-established sets of undertakings that should be agreed with collaborating and participating individuals/organisations. For example, a 'No' answer is justified only if:

- (a) There will be no attribution of individual responses;
- (b) Individuals (and, where appropriate, organisations) are anonymised in stored data, publications and presentation;
- (c) There has been specific agreement with respondents regarding feedback to collaborators and publication.

Overall assessment

- If every answer above is a definite NO, the self-audit has been conducted and confirms the **ABSENCE OF REASONABLY FORESEEABLE ETHICAL RISKS** – please tick box
This means that regarding this study, as currently self-audited, no further ethical review actions are required within CPHS. However, if in the coming weeks/months there is any change to the research plan envisaged now (and outlined above), the study should be re-audited against a Level 1 form, because it may be that the change made negates the absence of ethical risks signed off here.
- If one or more answers are YES, then risks have been identified and prior to commencing any data collection **formal ethical review is required** - either:
 - ~ by NHS REC (NB copy of ethics application and decision letter to be sent to CPHS Ethics); or
 - ~ if not to be formally reviewed by NHS REC, then CPHS level 2/3 ethical review required.
[If either of 5 or 7 are answered 'yes' then almost certainly level 3 is required.]



Two copies of this form should be taken for inclusion in the final dissertation and the original should be returned to the CPHS Ethics administrator.

S C McLean
Student Name

Aziz Sheikh
Supervisor Name

S C McLean
Student Signature

Aziz Sheikh
Supervisor Signature *

*** NOTE to supervisor:** *The CPHS Ethics Subgroup will not check this form (the light touch Level 1 form means we have insufficient detail to do so). By counter-signing this check-list as truly warranting all 'No' answers, you are taking responsibility, on behalf of CPHS and UoE, that the research proposed truly poses no potential ethical risks. Therefore, if there is any doubt on any issue, it would be a wise precaution to mark it as 'uncertain' and contact the Ethics Subgroup as to whether a level 2 form might be required as well. (See Intra Ethics website – URL at top of form)*

Appendix 3.1 Ethics waiver for use of Lothian COPD cohort database

South East Scotland Research Ethics Service

Waverley Gate
2-4 Waterloo Place
Edinburgh
EH1 3EG



Name: Susannah McLean
Address: Doorway 1
Medical Quad
Teviot Place
Edinburgh
EH8 9AG

Date: 28/03/2013
Your Ref:
Our Ref: NR/1303AB19
Enquiries to: Alex Bailey
Direct Line: 0131 465 5679
Email: alex.bailey@nhslothian.scot.nhs.uk

Dear Susannah,

Project Title: Modeling the prevalence and burden of COPD in Scotland and projections for the future

You have sought advice from the South East Scotland Research Ethics Service on the above project. This has been considered by the Scientific Officer and you are advised that, based on the submitted documentation (email correspondence, Proposal Lothian Database for ethics 22032013, Ethics application reply to alex bailey), it does not need NHS ethical review under the terms of the Governance Arrangements for Research Ethics Committees (A Harmonised Edition). The advice is based on the following:

- *The project is a study that will utilise previously-collected routine clinical data that is anonymised to the researcher*

If the study is considered as research you may require ethical approval as outlined in The Research Governance Framework for Health and Community Care. You may wish to contact your employer or professional body to arrange this.

For studies that are not research and will be conducted within the NHS you should contact the relevant local Quality Improvement Team(s) who will inform you of the governance procedures required before the study commences.

This letter should not be interpreted as giving a form of ethical approval or any endorsement of the project, but it may be provided to a journal or other body as evidence that NHS ethical approval is not required. However, if you, your sponsor/funder or any NHS organisation feels that the project requires ethical review by an NHS REC, please write setting out your reasons and we will be pleased to consider further. You should retain a copy of this letter with your project file as evidence that you have sought advice from the South East Scotland Research Ethics Service.

Yours sincerely,

A handwritten signature in black ink that reads 'Alex Bailey'.

Alex Bailey
Scientific Officer
South East Scotland Research Ethics Service

Appendix 3.2 Email approval for use of the Lothian COPD Cohort database from Privacy Advisory Committee (PAC) representative Janet Murray

Dear Susannah

Thank you for this information regarding the modelling for which you will use the Lothian data. I am sorry that I overlooked your response. I have not had a reminder since then.

I do not think that there is a risk of identifying individuals from the aggregate data you will produce and agree that it is acceptable for you to analyse the Lothian linked dataset.

With best wishes

Janet

Dr Janet Murray

Consultant CPHM

Information Services Division

Gyle Square

1 South Gyle Crescent

Edinburgh EH12 9EB

0131 275 6954

PA Sian Bennett

T: 0131 275 6293 E: sian.bennett@nhs.net

-----Original Message-----

From: Susannah McLean [<mailto:Susannah.McLean@ed.ac.uk>]

Sent: 16 April 2013 11:14

To: Murray Janet (NATIONAL SERVICES SCOTLAND)

Cc: Campbell Fiona (NATIONAL SERVICES SCOTLAND); PAC (NATIONAL SERVICES SCOTLAND)

Subject: Re: Amendment to pac Application 49/10

Dear Janet,

Thank you for your reply, with regards the first point, that is good news.

With regards the second I have attached a more detailed working document

which outlines the origin of every parameter we hope to include in the

model. the parameters that will originate from the Lothian cohort can be

seen in the summary tables in pages 2 and 3 are described in detail

later in the document. they are

A3b. incidence of COPD in Lothian/Scotland

A6. a random effects model to predict lung function decline.

A9. a case fatality profile by age and sex

A A rate of exacerbation model

B case fatality profile, very similar to A9

C decline in FEV1 due to exacerbation, calculated using results from previous models.

I hope this helps, the document has further details on the models but I

plan to have done the detailed work on the model before I go to Rotterdam. Then simply travel with the aggregated data.

Please let me know if I can provide further clarification.

Yours sincerely,

Susannah

On 11/04/2013 15:16, Murray Janet (NATIONAL SERVICES SCOTLAND) wrote:

> Dear Susannah and Fiona

>

> Thank you for your updated application Susannah. I have read it with the original PAC application and note the two changes to the original application which you list in your email below. I do not think that this application requires full assessment by the committee. I think I may be able to make a decision alone, depending on the extent of data to be sent to Rotterdam.

>

> 1 The additional use of the data is epidemiological research into COPD. Although the focus is different from the aim of the original application, I am content that the purpose is in line with that for

which the database was created. I note that Dr Rachel Hardie who as responsible for its creation is involved with this work.

>

> 2 The application indicates that aggregate data will be supplied to the University of Rotterdam and will be stored on a laptop there. Depending on the level of detail required there may be a risk of identifying individuals even from an aggregated database. So that I can consider the level of risk, and whether particular sensitive information would be revealed, please would you let me know all of the variables which will be used to create the aggregate data? Depending on the level of risk, I may have to discuss with our privacy and security advisors.

>

> With best wishes

>

> Janet

>

> Dr Janet Murray

> Consultant CPHM

> Information Services Division

> Gyle Square

> 1 South Gyle Crescent

> Edinburgh EH12 9EB

>

> 0131 275 6954

>

> PA Sian Bennett

>

> T: 0131 275 6293 E: sian.bennett@nhs.net

>

>

> -----Original Message-----

> From: Susannah McLean [<mailto:Susannah.McLean@ed.ac.uk>]

> Sent: 10 April 2013 10:24

> To: Campbell Fiona (NATIONAL SERVICES SCOTLAND)

> Subject: Re: Amendment to pac Application 49/10

>

> Dear Fiona,

> The original application created the Lothian COPD Cohort which is
> currently undergoing the approved analysis. This application is to
> approve an extension to this analysis which involves an economic
model

> from Erasmus University and therefore involves taking aggregated
data to

> Rotterdam for use in the model. Data from the Lothian COPD Cohort
on

> lung function decline, mortality and exacerbations will be used to
adapt

> the model for use in Lothian/Scotland to predict the size of the
future

> local and national COPD disease burden.

> The fact that we extend analysis and take data to the Netherlands,
we

> thought we'd need a new PAC application.

> Thank you

> Susannah

>

> On 10/04/2013 09:26, Campbell Fiona (NATIONAL SERVICES SCOTLAND)
wrote:

>> Hi Susannah,

>>

>> Thanks for the amendment to your original PAC form.

>>

>> To make it easier for us, could you explain in a couple of
paragraphs what is additional to your previous PAC application? It's
not that clear from the form you have submitted.

>>

>> Also, given the time that has lapsed since your original
application can you justify why it's an amendment and not a new
application.

>>

>> Thanks for this. It will help us to deal with your application as
promptly as possible.

>> Regards,

>> Fiona.

>>

>> -----

>> Fiona Campbell

>> Information Consultant

>> ISD Scotland

>> Gyle Square

>> 1 South Gyle Crescent

>> Edinburgh, EH12 9EB

>> Tel: 0131 275 6923

>> Email: Fiona.Campbell2@nhs.net

>>

>> -----Original Message-----

>> From: Susannah McLean [<mailto:Susannah.McLean@ed.ac.uk>]

>> Sent: 08 April 2013 15:40

>> To: PAC (NATIONAL SERVICES SCOTLAND)

>> Subject: Amendment to pac Application 49/10

>>

>> Dear PAC,

>> Please find attached a PAC application for a study which has already been approved and an amendment request on the pilot form dated 8th April 2013.

>> Please could you consider this request for an amendment to the already approved study protocol.

>> Please contact me if you require further information.

>> Thank you very much

>> Susannah

>> --

>> Dr Susannah McLean

>> Doorway 1

>> Medical Quad

>> Teviot Place

>> Edinburgh

>> EH8 9AG

>>

>> Tel: 0131-650-9242

>>

>> The University of Edinburgh is a charitable body, registered in Scotland, with registration number SC005336.

>>

>>

>>

>>

>> This message may contain confidential information. If you are not the intended recipient please inform the

>> sender that you have received the message in error before deleting it.

>> Please do not disclose, copy or distribute information in this e-mail or take any action in reliance on its contents:

>> to do so is strictly prohibited and may be unlawful.

>>

>> Thank you for your co-operation.

>>

>> NHSmail is the secure email and directory service available for all NHS staff in England and Scotland

>> NHSmail is approved for exchanging patient data and other sensitive information with NHSmail and GSi recipients

>> NHSmail provides an email address for your career in the NHS and can be accessed anywhere

>>

>>

>>
>
> --
> Dr Susannah McLean
> Doorway 1
> Medical Quad
> Teviot Place
> Edinburgh
> EH8 9AG
>
> Tel: 0131-650-9242
>
> The University of Edinburgh is a charitable body, registered in
> Scotland, with registration number SC005336.
>
>
>

>
> This message may contain confidential information. If you are not
the intended recipient please inform the
> sender that you have received the message in error before deleting
it.

> Please do not disclose, copy or distribute information in this e-mail or take any action in reliance on its contents:

> to do so is strictly prohibited and may be unlawful.

>

> Thank you for your co-operation.

>

> NHSmail is the secure email and directory service available for all NHS staff in England and Scotland

> NHSmail is approved for exchanging patient data and other sensitive information with NHSmail and GSi recipients

> NHSmail provides an email address for your career in the NHS and can be accessed anywhere

>

>

>

--

Dr Susannah McLean

Doorway 1

Medical Quad

Teviot Place

Edinburgh

EH8 9AG

Tel: 0131-650-9242

The University of Edinburgh is a charitable body, registered in Scotland, with registration number SC005336.

This message may contain confidential information. If you are not the intended recipient please inform the sender that you have received the message in error before deleting it.

Please do not disclose, copy or distribute information in this e-mail or take any action in reliance on its contents:

to do so is strictly prohibited and may be unlawful.

Thank you for your co-operation.

NHSmial is the secure email and directory service available for all NHS staff in England and Scotland

NHSmial is approved for exchanging patient data and other sensitive information with NHSmial and GSi recipients

NHSmial provides an email address for your career in the NHS and can be accessed anywhere

Appendix 3.3 Privacy Advisory Committee (PAC)

Application for Lothian COPD Cohort Database

NHS National Services Scotland Privacy Advisory Committee

Application for Privacy Advisory Committee Approval

THE PRIVACY ADVISORY COMMITTEE (PAC) is an advisory committee to NHS National Services Scotland (NSS) and the Registrar General.

The PAC advises on the correct balance between protecting personal data and making data available for research, audit and other important uses and ensures that any information releases are carefully controlled.

PAC views are sought in relation to any request for access to information that would involve the release of data under the control of NSS or National Records of Scotland (NRS) that are, or have the potential to be, person-identifiable, and in respect of any new record linkages.

Where the data you wish to access is not in the control of NSS or NRS you may need to consider applying to another advisory body - see 'Appendix One: Advisory Bodies within NHSScotland'.

Assistance with PAC

NSS Research Coordinators are a specialist team of principal and senior information analysts based within the Information Services Division (ISD) who will help you to define

requirements for linked datasets and any analyses required. The NSS Research Coordinator will explain what information is available and assist you deciding what variables would be useful for your study. They will also advise when you may require to approach another advisory body.

When to Complete a Privacy Advisory Application Form

The PAC Application Form must be completed for data requests that involve:

- (a) access to identifiable or potentially identifiable information;
- (b) circumstances where NSS or National Records of Scotland (NRS) have indicated their intention to
 - seek guidance from PAC; and/or
- (c) record linkage of previously unlinked datasets involving data from more than one Health Board.

Our aim is to make data for research available through the Scottish Informatics Programme (SHIP) infrastructure. All applicants to PAC will be expected to use the SHIP National Safe Haven as a means of accessing data except in very exceptional circumstances.

Scottish Informatics Programme (SHIP)

SHIP is a Scotland-wide research platform for the collation, management, dissemination and analysis of Electronic Patient Records (EPRs). The programme brings together the Universities of Dundee, Edinburgh, Glasgow and St Andrews with the Information Services Division (ISD) of NSS.

The SHIP programme will provide a platform for Scottish record linkage that will drive EPR research throughout the UK and abroad. The SHIP National Safe Haven is the analytical platform for this. This includes the following.

- Provision of a record linkage service where personal identifiers are kept separate from the payload/content data.
- Provision of a secure environment for researchers to analyse anonymised patient level or summarised records.
- Provision of a Secure File Transfer service to support the transmission of data between data providers and researchers.

Anyone wishing to access data through the SHIP National Safe Haven requires to be registered as a SHIP Approved Researcher or to have recently completed an Information Governance training course.

Further Advice on National Records Service and NHS Central Registry

To find out more about NRS data including those relating to the NHS Central Registry visit <http://www.gro-scotland.gov.uk/national-health-service-central-register/index.html> or e-mail:

dumf-uhb.NHSCR-Scotland-Medical-Research@nhs.net

For All Other Requests

For all other requests or advice please contact the Research Coordinators at nss.ISDMedicalRecordLinkage@nhs.net

Application Checklist

Before you submit your application, you should include the following items and ensure that the application has been signed by the appropriate individuals and then save the final document as a PDF to be e-mailed to the PAC mailbox at:

nss.pac@nhs.net.

Your application should be typed, not handwritten.

Items to support application

Where applicable, you should include the following.

- Study protocol
- Information provided to study participants and/or the wider public
- Participant consent forms
- Draft correspondence (if the data generated through your study will be used to contact any individuals)
- Evidence of ethical approval
- Evidence of approval from other Data Controllers e.g. CHIAG
- Local Information Governance / security policies and procedures
- The list of variables you require in the file for analysis
- Details of each individual accessing the data (where there are more than five allowed on this form).

Please ensure that your application has been signed by the:

- main study contact;
- study information custodian (for studies requesting the release of data outside the SHIP National Safe Haven).

You must also ensure that individuals named on the form have read and approved this submission.

You must save this form as a PDF and only send the PDF version.

Application for Privacy Advisory Committee Approval

Application Title	Modelling the disease burden of the COPD population
Date Submitted	March 2013
NSS Research Co-ordinator Name	Philip Johnson
NSS Information Request Number (Your NSS Research Coordinator will provide this information)	This an amendment to PAC application 49/10

Please Note

- **The information contained in this application form will be regarded as confidential as it goes through the scrutiny process but it is the duty of applicants to point out any information within the application that they consider to be particularly sensitive, confidential or commercially sensitive.**

- **As NHS National Services Scotland is a public authority, it is subject to Freedom of Information (Scotland) Act.**
- **PAC applications are kept by NHS National Services Scotland for a minimum period of 15 years from date of application or date of the last linkage undertaken in relation to the application.**

Application Part One: People Involved

The names of all the individuals involved in and responsible for the design and analysis of the study should be included here. It is expected that those who have access to the data supplied by NSS have adequate and regular updated knowledge and skills in the secure and confidential handling of health data.

- You must ensure that anyone required to access the data provided should be a SHIP Approved Researcher or should be able to demonstrate skills and knowledge of/having successfully recently completed approved training in Information Governance.
- You must ensure that all staff taking part in this study must have appropriate contracts in place containing clauses that clearly identify their duties and responsibilities for confidentiality, data protection and data security.
- You do not need to include clerical and secretarial support staff.

Head of Department responsible for project/study or the Principal Investigator <i>This should be the name of the person with overall responsibility for the study and who will take responsibility for it.</i>	
Title	Prof
Forename	Aziz
Surname	Sheikh
Position	Professor of Research and Development in Primary Care
Qualifications	MBChB, MD, MRCPCH, FRCGP
Professional Registration Number (if applicable) e.g. General Medical Council (GMC)	

Organisation Name	University of Edinburgh	
Address	Doorway 3, Medical Quad, Centre for Population Health Sciences, Teviot Place, Edinburgh	
Postcode	EH8 9AG	
Telephone number	0131-650-8101	
Email	Aziz.sheikh@ed.ac.uk	
Please complete following questions if you will access the individual level data requested in this application		
Are you a SHIP Approved Researcher	Yes	
What was the date when SHIP Approved Researcher status was conferred		

If not a SHIP Approved Researcher, please provide details of most recent Information Governance training undertaken.	Name of Course: Graham Laurie's online SHIP ethics course	
	Institution University of Edinburgh	
	Date attended	

Main Contact <i>Researcher responsible for day to day running of project (to whom all correspondence will be sent) - if different from the above.</i>	
Title	Dr
Forename	Susannah
Surname	McLean

Position	PhD Student	
Qualifications	MBBS, MSc, MRCP	
Professional Registration Number (if applicable) e.g. GMC	6052508	
Organisation Name	University of Edinburgh	
Address	Doorway 1, Medical Quad, Centre for Population Health Sciences, Teviot Place, Edinburgh	
Postcode	EH8 9AG	
Telephone number	0131-650-9242	
Email	Susannah.mclean@ed.ac.uk	
Are you a SHIP Approved Researcher	Yes	
What was the date when SHIP Approved Researcher status was conferred		
If not a SHIP Approved Researcher, please provide details of most recent Information Governance training undertaken.	Name of Course “ ADLS Safe researcher Course”	
	Institution ADLS (Administrative Data and linked service)	
	Date attended May 2011	

Information Custodian *The Information Custodian is the person taking responsibility for safeguarding the confidentiality of the data. This is likely to be the Head of Department responsible for the project. An Information Custodian is required if the SHIP National Safe Haven will not be used e.g. where data is requested from NRS*

Title	Prof	
Forename	Aziz	
Surname	Sheikh	
Position	Professor of primary care research and development	
Qualifications	MBChB, MD, MRCPCH, FRCGP	
Professional Registration Number (if applicable) e.g. GMC	4016942	
Organisation Name	University of Edinburgh	
Address	Doorway 3, Medical Quad, Centre for Population Health Sciences, Teviot Place, Edinburgh	
Postcode	EH8 9AG	
Telephone number	0131-650-8101	
Email	Aziz.sheikh@ed.ac.uk	
Are you a SHIP Approved Researcher	Yes	
What was the date when SHIP Approved Researcher status was conferred	20.3.2012	
If not a SHIP Approved Researcher, please provide details of most recent Information Governance training undertaken.	Name of Course Graham Laurie's online SHIP Ethics course	
	Organisation SHIP	
	Date attended 20.3.2012	

Please list the name(s) of each person who will access the individual level data requested in this application. There is space here to provide details for three names. If there are more than three names, please append the additional information with your application.

Access Individual Level Data – 1	
Title	Dr
Forename	Susannah
Surname	McLean
Position	PhD Student
Qualifications	MBBS, MSc, MRCP
Professional Registration Number (if applicable) e.g. GMC	6052508
Organisation Name	University of Edinburgh
Address	Doorway 1, Medical Quad, Centre for Population Health Sciences, Teviot Place, Edinburgh
Postcode	EH8 9AG
Telephone number	0131-650-9242

Email	Susannah.mclean@ed.ac.uk	
Are you a SHIP Approved Researcher	Yes	
What was the date when SHIP Approved Researcher status was conferred	May 2011	

If not a SHIP Approved Researcher, please provide details of most recent Information Governance training undertaken.	Name of Course ADLS safe researcher
	Organisation ADLS
	Date attended May 2011

Access Individual Level Data – 2	
Title	Prof
Forename	Sarah
Surname	Wild
Position	PhD Supervisor, Professor of Epidemiology
Qualifications	MB BChir FRCP FFPH
Professional Registration Number (if applicable) e.g. GMC	3133196
Organisation Name	University of Edinburgh
Address	Doorway 1, Medical Quad, Centre for Population Health Sciences, Teviot Place, Edinburgh
Postcode	EH8 9AG
Telephone number	0131-651-1630
Email	Sarah.Wild@ed.ac.uk
Are you a SHIP Approved Researcher	Yes
What was the date when SHIP Approved Researcher status was conferred	May 2011

If not a SHIP Approved Researcher, please provide details of most recent Information Governance training undertaken.	Name of Course ADLS safe researcher
	Organisation ADLS
	Date attended May 2011

Access Individual Level Data – 3	
Title	Dr
Forename	Colin
Surname	Simpson
Position	PhD Supervisor, Reader in Epidemiology
Qualifications	BSc PhD
Professional Registration Number (if applicable) e.g. GMC	NA
Organisation Name	University of Edinburgh
Address	Doorway 3, Medical Quad, Centre for Population Health Sciences, Teviot Place, Edinburgh
Postcode	Eh8 9AG
Telephone number	0131 651 4151
Email	c.simpson@ed.ac.uk
Are you a SHIP Approved Researcher	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No
What was the date when SHIP Approved Researcher status was conferred	
If not a SHIP Approved Researcher, please provide details of most recent Information Governance training undertaken.	Name of Course
	Organisation
	Date attended

Please list the names of any people not listed above that have had significant input into the design and content of this study, or will be involved in the study hereafter

Other People Involved	
Name (Forename/Surname)	Organisation
Aziz Sheikh	University of Edinburgh
Leonie Hunter	Lothian Health Board
Rachel Hardie	Lothian Health Board

Please list up to three publications which you have produced or been involved in that demonstrate experience in the use of administrative data for research.

Journal Name	Title of publication	Date of publication	Author(s) and Organisation
NA			

Confirmation

Please confirm that the data supplied by NSS will not be passed to anyone other than those individuals identified on this form. X *J. C. McLan*

Application Part Two: Study Overview

In order to help the PAC assess your application, you are required to provide an overview of your study. The following section should be completed with information accessible and comprehensible to a lay reader, expressing any acronyms in full.

You should include your study protocol with your application.

What is the background to the study?

Chronic obstructive pulmonary disease (COPD) is a respiratory disease which is characterised by progressive airflow limitation that is not fully reversible. COPD is a major cause of chronic morbidity and mortality throughout the world and results in an economic and social burden that is both substantial and increasing. The 2010 World Health Organization Global Burden of Disease study lists COPD as the third largest cause of death worldwide. COPD patients not only die prematurely, they also have a decreased quality of life due to impaired physical, mental, and social functioning.

The most important risk factor for COPD development and progression is cigarette smoking. Smokers with COPD experience an accelerated decline in lung function.

As populations are ageing, the prevalence of long term conditions such as COPD is increasing. This means that more resources are required to provide healthcare for the COPD population. In order to accurately plan these resources a dynamic COPD population model has been developed by Erasmus University, Rotterdam, the Netherlands. This model includes provision for calculating how many people will develop COPD and at what severity if they continue to smoke or stop smoking later in life. It also includes provision for modelling the healthcare needs of these people from data on the rate of exacerbations and costs of healthcare. No equivalent model exists for Scotland.

In order to plan healthcare for people with COPD in the future, these calculations as to the size of the disease burden are useful. In addition, the effects of interventions such as smoking cessation campaigns may be computed.

Why is the study needed?
In order to better plan the care costs and provisions of people with COPD in Scotland in the future.
What are the aims and objectives of the study?
Objective <ul style="list-style-type: none"> ▪ The objective is to quantify COPD burden derived in Scotland and then to compare with similar disease burdens computed from Dutch and other UK cohorts. <p>Specific aims:</p> <ul style="list-style-type: none"> ▪ The aim is to predict disease burden from COPD in 5 and 10 years time using a pre-existing model developed by Erasmus University, Rotterdam, using parameters derived from this database.
Give a brief outline of the study design and data sources involved.
Study Design <p>A modelling study involving generating specific parameters based on the Lothian COPD Cohort database and inputting them into model from Erasmus University to better estimate the disease burden from COPD in the future in Lothian and Scotland. Comparisons will be made with model outputs generated from other sources of data, e.g. Dutch and other UK data.</p> <p>Data Source</p> <p>Lothian COPD Cohort Database</p>
Please describe your study sample (inclusion/exclusion criteria e.g. involvement in trial/survey, health event, relevant date range, requirement for a matched control cohort, etc).
The Lothian COPD Cohort is an existing incidence cohort database of patients with a COPD code in participating Lothian primary care practices.

Please indicate whether this study has any implications for sensitive groups of vulnerable populations (see Appendix Two for details).

NA

Please describe envisaged benefits of your study either to patients or the wider public.

Better planning of COPD healthcare resources.

Does this study have any of the following commercial aims?

To be used by your organisation as part of a product or service to be sold by you to your customers

No

To be used as part of a commercially funded research project

No

To be used by your organisation as part of a product or service given freely by you to your customers

No

To be incorporated into a publication which will be subsequently distributed free of charge

Yes

Please specify any additional intended usage for this data

None

Does this application relate to previous PAC applications?

Yes

If yes, please provide the PAC reference number(s)	This is an amendment to 49/11	
Do you have funding in place for your study?	Yes	
If yes, please provide the name of the funding body	Susannah Mclean is funded by a University Scholarship	
Do you intend to use the SHIP National Safe Haven? Scottish-based researchers will be expected to use the National SHIP safe haven. By using SHIP safe haven the applicant doesn't need to complete the IG section		No – as the data are already removed from the NHS.
If no, please explain your reason why below.		
This project is an extension to an existing project and data are already available on a secure University of Edinburgh Server.		

Application Part Three: Data Requests

This section of the form requests further detail regarding the use of data to meet the objectives of the study.

3.1 NRS Data

Does your study involve NRS data? If no, go to section NSS Data?	Yes	
Does your study require use of NHS Central Registry (NHSCR) as a sampling frame for study controls?		No
Does your study involve flagging of individuals on the NHSCR? If yes, please answer the following questions.		No
Is the flagging of individuals:		
▪ To help trace and contact individuals throughout the UK?	Yes	No
▪ To be informed of fact and cause of death?	Yes	No
▪ To be informed of cancer registrations?	Yes	No
▪ To be informed of emigrations prospectively and retrospectively?	Yes	No
▪ To be provided with safe haven analysis?	Yes	No
▪ For population analysis?	Yes	No
▪ For household analysis?	Yes	No
Does your study require the provision of any other service from NRS? If so detail below.		No

3.2 NSS Data

Does your study involve NSS data?	Yes	
If yes, please identify the dataset(s) involved below.		
▪ SMR00 - Outpatients		
▪ SMR01 - Inpatients and Daycases	Yes	
▪ SMR02 - Maternity		
▪ SMR04 - Mental Health		
▪ SMR06 - Cancer Registration		
▪ SMR11/SBR - Neonatal/Scottish Birth Record		
▪ CHSP-PS/CHSP-S/SIRS - Child Health Surveillance		
▪ A&E - Accident and Emergency		
▪ PIS - Prescribing Information		
▪ National Audits and Disease Registries e.g. Surgical Mortality, Renal Registry		
▪ Birth, Stillbirth or Death Records (NRS)	Death records	
▪ Other (please list below)		
These dataset have been already linked into the Lothian COPD Cohort database		

Indicate by ticking all the box (es) that apply whether the information provided by NSS Scotland will be used to make direct contact with the following.				
	Make Contact	By Letter	By Telephone	Other method -

				please specify
▪ Hospital consultants				
▪ Other hospital staff				
▪ General Practitioners				
▪ Study members or patients				
▪ Relatives of study members or patients - please specify				
▪ Some other party - please specify				
Does your application request NSS to facilitate communication with individuals in the study sample?				No
Does your study involve use of NSS Data as a sampling frame?				No
Will all analysis be done in NSS by NSS staff e.g. you only require aggregate output.				No

3.3 Non-NRS/NSS Datasets

Does your study involve linkage to non-NRS/NSS data? If yes, you must provide information on each dataset. If no go to next section.	Yes	
--	-----	--

Please complete for each of the non-NSS datasets that you will provide to ISD e.g.

- Data held by GPs

This linkage has already been performed

There is space here to provide details for three non-NRS/NSS datasets. If there are more than three names, please append the additional information with your application.

Non NRS / NSS Dataset - 1		
What is the name of the dataset?	Lothian COPD Cohort Database is already in existence, please see PAC application 49/11	
The purpose for which it was collected		
Describe the content of the dataset		
The time period to which it pertains		
The identifying variables which will be provided to enable linkage (please tick all that apply)	Forename	
	Middle name	
	Surname	
	CHI Number	

	Postcode	
	Date of Birth	
	Gender	
	UK NHS Birth Registration Number	
	Other (please specify below):	

Non NRS/NSS Dataset - 2		
What is the name of the dataset?		
The purpose for which it was collected.		
Describe the content of the dataset.		
The time period to which it pertains.		
The identifying variables which will be provided to enable linkage (please tick all that apply)	Forename	
	Middle name	
	Surname	
	CHI Number	
	Postcode	
	Date of Birth	
	Gender	
	UK NHS Birth Registration Number	

	Other (please specify below):

Non NRS/NSS Dataset - 3		
What is the name of the dataset?		
The purpose for which it was collected.		
Describe the content of the dataset.		
The time period to which it pertains.		
The identifying variables which will be provided to enable linkage (please tick all that apply)	Forename	
	Middle name	
	Surname	
	CHI Number	
	Postcode	
	Date of Birth	
	Gender	
	UK NHS Birth Registration Number	
	Other (please specify below):	

3.4 Output File for Analysis

The risk of disclosure of individuals increases with increased level of detail contained in the dataset. Only variables required to meet study objectives should be requested.

Identifiable data includes variables such as name and date of birth. In general, access to personal identifiers will not be provided. Exceptional requests for access may be considered taking account of Information Governance principles.

Do you require any patient identifiers in the output file for analysis?	Yes	
---	-----	--

If yes, please identify which patient identifiers you require in the output file for analysis	
Forename	
Surname	
CHI Number	
NHS Number	
Postcode (full or part)	
Date of Birth	Yes
Gender	Yes
Other (please list	
Please provide justification on why you need these patient identifiers. <i>This information can only be provided where a clear need is shown.</i>	

Male and female and age to give the profile of people with COPD

Data without identifiable variables may retain the potential to identify an individual. This increases with the level of detail included, particularly where the denominator population is small as when rare conditions or low level geographies are involved.

Our Research Co-ordinators will be able to advise whether ‘derived variables’ may be provided e.g. length of stay with month of admission rather than dates of admission and discharge. Output files may include a study index number where recognition of individual records is necessary.

Please identify which, if any, of the following variables are required.	
Dates:	
▪ Full dates	
▪ Year and Month	Yes
Geographical variables:	
▪ NHS Board area	
▪ Local Authority area	
▪ Datazone	
▪ Postcode district	
▪ Information regarding rare conditions (please list the conditions)	
Please provide justification on why you need these variables. <i>Variables can only be provided where a clear need is shown.</i>	
To derive age and time with COPD	

3.5 Other Variables

Please list all the other variables (not already listed) which will be required, identifying which variables come from which data source.

Variable	Data Source
Lothian COPD Cohort Database is already in existence, please see PAC application 49/11	

3.6 Duration of the Study

What is the expected duration of the study?
18 months

3.7 Updates and Retention

Does your study require access to a regular update of data?		No
If yes, please explain the reason for this.		

Please specify the frequency of updates your study will require.
How long will you be keeping the data (including the updates) after the study is complete?
5 years in the Edinburgh University Data Archive
Please provide justification on why you need to retain the data for that length of time.
To complete publications from the study

Application Part Four: Permissions to Use Data

For each dataset not under the control of NRS or NSS, you must seek authority to access those data.

- An application to the CHI Advisory Group (CHIAG) is necessary where the study requires access to information from the CHI dataset.
- An application to the NHSScotland Caldicott Guardian Forum is necessary where the study requires access to information datasets held by multiple NHS Boards.
- Approval from the National Research Ethics Service should be sought in the following situations:
 - Where the study involves linkage of a research dataset to another dataset.
 - Where the study involves use of identifiable data.
 - Where study involves use of highly disclosive data e.g. information regarding rare conditions or at the level of small geography.

Evidence that use is authorised by the Data Controller(s). <i>This may include authorisation from CHIAG, NHS Caldicott Guardians Forum or other. This can be attached to your application.</i>
Permission will be sought from the NHS Caldicott Guardian, We have a waiver from NHS ethics and we will apply to the University Internal ethics procedure.
Please describe the methods used to inform study participants and/or the wider public regarding the use of their data in this way.
GP practices that gather patient data have information in the practice waiting room regarding data use.
Please provide copies of the information provided to study participants and/or the wider public regarding the use of their data in this way. <i>This can be attached to your application.</i>
NA

Please provide the consent form for research studies or surveys. <i>This can be attached to your application.</i>
NA
Where no consent for proposed use has been obtained from data subjects, please provide justification below explaining why there is use without consent <i>e.g. please explain why consent has not been obtained and explain how this proposed use relates to the original purpose of data collection.</i>
It is believed that this study is in the wider public interest. Namely to improve the prediction of the burden from COPD in a country and therefore patients will benefit from the study output. In addition, all data required for this study are non-identifiable, thereby protecting the participants and only aggregate data will be required for the model.
Have any members of the public/lay people been involved in the study design? Please provide information.
No

Application Part Five: Information Governance

NSS must ensure that any data approved for release to the applicant will be adequately protected against inappropriate access and use during the term of the study and on completion of the study the data will be securely disposed of.

- This section does not require to be completed for studies in which data will be accessed only through the SHIP National Safe Haven. Please complete this section for all other applications i.e. studies requiring release of data to the applicant.
- Researchers from Scottish institutions will be expected to access NSS data using the SHIP National Safe Haven. Any alternative to this will require justification.

- You may need to consult your organisation's Information Governance Lead and IT service supplier when completing this part of the form.

5.1 Local Information Governance Policies and Procedures

Your organisation(s) should have Information Governance policies and procedures for each location and that these are accessible and used by all staff on this study. Please attach copies of local Information Governance policy/policies for each location to this application or provide the URL if the policies are available online.

URL(s) if policies are available online:

<http://www.ed.ac.uk/schools-departments/records-management-section/data-protection/guidance-policies/research/research>

5.2 Information Governance Incident Reporting

Your organisation(s) should have Information Governance incident reporting procedures for each location and that these are accessible and used by all staff on this study.

NSS should be notified immediately of any information governance breaches that have occurred involving NSS supplied data during this study. Please confirm you will

notify nss.pac@nhs.net of any such incidents X



5.3 Data Protection Registration

Please provide the Data Protection Registration Number of each of the organisation (s) where data will be held.

Organisation(s) Name/Data Storage location	Data Protection Registration Number
University of Edinburgh	Z6426984

5.4 ISO 27001

If your organisation(s) have adopted ISO27001 - Information Security - Security Techniques - Information Security Management Systems, please provide your certification number.

Organisation Name / Data Storage location	ISO 27001 Certification Number
University of Edinburgh	We do not have this certificate

5.5 NSS Data Transfer and Storage Policies

NSS requires that all sensitive and person identifiable data is encrypted during transfer and whilst stored on mobile data storage devices and desktop and laptop computers.

- NSS prefers that any data provided is stored on secure networked drives as part of a secure managed server. If mobile data storage devices have to be used, you must implement adequate protection against device loss or theft, unauthorised interception and access .
- Where NSS supplied data is being stored on mobile data storage devices (for example but not limited to: USB ‘sticks’ and USB data storage drives, desktop or laptop computer) these devices must be fully encrypted to FIPS 140-2/CESG CAPS certified level of security protection.

- Where devices cannot be encrypted (for example: CDs, DVDs) then the data must be encrypted to FIPS 140-2/CESG CAPS certified level of security prior to storage on the mobile data storage device.
- NSS uses NHSmail email services for regular communications and the transfer of person identifiable data however cannot send or receive password encrypted attachments using these email services. NSS can only send sensitive and person identifiable data to other users using NHSmail services (either NHSmail accounts or the secure file transfer service). Please discuss methods with the Research Coordinator.
- There are risks associated with using any email services for the transfer of data including sending the communication to the wrong email address. Please ensure that the NHSmail email address you provide is the correct address to be used. NSS will only send data to individual user (e.g. named) email addresses and not to generic email addresses.

Please confirm that you have read and understood the details regarding NSS Data Transfer and Storage Policies by ticking here X

5.6 Data Storage

Please provide details on how and where you will store the data supplied by NSS. If data is being stored in more than one location then this section needs to be clearly completed for each location.

At what location(s) will data be stored? <i>Please list.</i>	Information services , University of Edinburgh.	
Will any data be stored outside of Scotland?	Yes	
If yes, please state the location outside Scotland where data will be stored.	Aggregated summary data will be used in Rotterdam to facilitate the running of the modelling.	

<i>Specific considerations will apply where data is stored outside of the European Union.</i>	
---	--

5.7 Storing of Data

This section relates to where data supplied by NSS will be stored.

Storage Device	Please tick all that apply and specify for each, the location at which the data will be stored. *delete as appropriate.	
	Confirm	Location
Networked server disk drive	X	University of Edinburgh
Networked desktop PC*/ laptop*	X (laptop)	Aggregated data at University of Rotterdam
Standalone desktop PC*/ laptop*		
Mobile device		

Storage Format	Please tick all that apply and specify for each, the location at which the data will be stored.	
	Confirm	Location
Database		
Oracle database		
Microsoft Access database		
Microsoft SQL server		
IBM DB2		
MySQL		

Flat file e.g Excel spreadsheet, comma delimited file.	X	University of Edinburgh and Aggregated data at University of Rotterdam
---	---	--

5.8 Backup

Please confirm that your back up schedule is subject to the same security as the data provided by NSS for each location(s) that the data will be stored.

If not please provide details of your backup here:

5.9 Other Encryption or Anonymisation Procedures

NSS Data Transfer and Storage Policies require that all sensitive and person identifiable data is encrypted during transfer and whilst stored on mobile data storage devices and desktop and laptop computers to the standards outlined earlier. Please provide details of any other encryption or anonymisation procedures that may be used and at what stage.

Any other encryption or anonymisation procedures used	At what stage
Password protected encryption and transfer in a tamper proof bag	Transfer of data from Lothian Health Board to Edinburgh University, further transfer of aggregated summary data from Edinburgh University to Rotterdam

5.10 Data Transfer - In

If you are providing NSS with a copy of data for linkage and/or analytical purposes please detail how this data will be transferred and what security, complying with NSS Policy, will be used to protect the data from interception and inappropriate access.

- Please note that by sending sensitive or personal data you will be responsible for ensuring the data is adequately protected against inappropriate access and tampering during the transfer. Data must not be sent via fax services.

- NSS uses NHSmail services for regular communications and the transfer of person identifiable data. NSS cannot send or receive password encrypted attachments using email services. NSS can send sensitive and person identifiable data to other users using NHS mail services (either NHSmail accounts or the secure file transfer service).
- NSS may use the same mode of data transfer however still reserves the right to use alternative mechanisms and security measures for the protection of NSS data. This can be discussed with your NSS SHIP Research Coordinator.

Please specify	
Mobile data storage device e.g. CD, USB, data stick.	Encrypted hard drive
FTP URL	
sFTP URL	
Email address from which data will be sent	
Other data transfer method	

The NSS Research coordinator dealing with your application will discuss requirement for ` and/or passwords for the data transfer and/or encryption process. You should not provide this information to PAC either via this form or to the PAC e-mail address.

5.11 User Access

Please provide details on user access and account management policies that you have in place to limit or prevent inappropriate access to the data supplied by NSS.

Will those accessing data, access it through individual or shared accounts?	Individual	Shared
Are 'complex' passwords (a mixture of alpha, numeric, upper/lower case, special characters) used on all accounts?	Yes	No
How often are users required to change their passwords?	Monthly	
	Quarterly	
	Bi-annually	d
	Annually	X

	Other <i>please specify</i>	
Are procedures in place to regularly review user access to sensitive and potentially identifiable personal data?	Yes	
Are procedures in place to revoke user access to sensitive and potentially identifiable personal data when the user no longer requires this access?	Yes	
Will the data be accessed by staff working off site e.g. staff working from home?	Yes	
If yes, please detail how this access will be secured		
Via University of Edinburgh's Secure VPN service.		
Please provide any additional details of how data provided by NSS will be protected from unauthorised access.		
Data are stored on a Unix file system which is mounted on shared access systems. Full access control list support is available on the system and is configured to be restrictive by default b the system administrators .		

5.12 Hardware Security

<p>Please describe the physical security arrangements for the location where the data is to be stored e.g. this could be your computer department if the data is stored on a networked server, or may be where the PC/laptop holding the data is physically located.</p>
<p>The university servers are stored in secure machine rooms. These are locked, and accessible only to authorised personnel using swipe card access. Backups are held in separate machine rooms in different sites, subject to the same access restrictions.</p>

Please describe the physical security arrangements for the location where the data is to be processed e.g. this is where your PC/laptop is located or wherever you are accessing the data from.
Office with Locked Doors and restricted access to building after work hours.
Please detail any protection that is implemented against the introduction of malicious software (e.g. computer viruses) in the areas where the data will be stored and processed.
Anti Virus and malware scans. Firewall protection. Regular software patching. Regular monitoring of network trafficking and filtering.

Do your hardware replacement agreement(s) address how data are handled when hardware under warranty fails?		No
If yes, would the hardware be returned to the supplier if there was a fault(s)?		No
Please explain below how your organisation(s) dispose of hardware that they no longer require, that are faulty or covered by warranty.		

Data is always securely wiped and then disposed of in secure manner at the end of its life. Any equipment returned under warranty is securely wiped also.

If the data is being held in long-term archive(s) please explain how this data will be secured against further unauthorised access.

All Data archived on archive NAS boxes. Password restricted using AD and stored in Secure locked rooms and cabinet.

Who will have data management responsibilities for the data whilst in archive(s)?

Prof Aziz Sheikh

What procedures are in place to retrieve the data from the archive(s)?

Requests can be placed to the computing support team to retrieve any data.

5.13 Data Retention and Disposal

Data should not be kept any longer than is necessary.

Please give details of your data retention policy for each of the organisations(s) holding the data, including any back-up copies.

Data will be retained as long as required to complete publications and PhD thesis. This is possibly up to 5 years.

Please give details of how the data, and any back-up copies, will be securely disposed of at the appropriate time by each of the organisation(s) holding the data.

Hard drives and USB sticks will be securely wiped using appropriate software. Information on Network drives will be deleted and written over

Data are stored in such a way that it would be difficult to reconstruct it from an individual drive if that failed and was returned to the supplier for replacement (all data is striped over a large number of volumes, each of which is made up of several disks which are themselves striped over several drives, thus splitting the data up into very small components and dispersing it over many drives).

Application Part Six: Declaration

I DECLARE THAT this application is accurate, and that any health data made accessible to me, should it be successful, will be used for no other purpose, and in no other way, than as described above.

I UNDERSTAND THAT NHS National Services Scotland will refuse any future applications by me, or my employing or sponsoring organisation, should I use any health data made accessible to me for any other purpose or in any other way than that described above.

I CERTIFY THAT all staff who have access to health data are aware of the requirements of confidentiality and understand that its breach (e.g. disclosure of confidential information to a person not authorised to receive it) constitutes grounds for disciplinary action, which might result in dismissal.


I UNDERSTAND THAT NHS National Services Scotland will refuse any future applications by me, or my employing or sponsoring organisation, should I use any health data made accessible to me for any other purpose or in any other way than that described above.

I CERTIFY THAT all staff who have access to health data are aware of the requirements of confidentiality and understand that its breach (e.g. disclosure of confidential information to a person not authorised to receive it) constitutes grounds for disciplinary action, which might result in dismissal.

I GUARANTEE THAT no publication will appear in any form in which an individual may be identified unless the written permission of that individual has been obtained, and that I

will follow the ISD Statistical Disclosure Control Protocol when planning publications involving the data requested.

To be signed by the applicant

Applicant Signature: 	Date: 8.4.2013
Name (in Capitals): SUSANNAH MCLEAN	

To be signed by the Information Custodian named in Part One where the Information Custodian is not the applicant.

I DECLARE THAT (the applicant named above) is a bona fide worker engaged in a reputable project and that the data he/she asks for can be entrusted to him/her in the knowledge that he/she will conscientiously discharge his/her obligations, including in regard to confidentiality of the data, as stated in the declaration above.

Information Custodian Signature: Aziz Sheikh	Date:08.04.2013
Name (in Capitals): AZIZ SHEIKH	

Appendix 4.1 Request for Caldicott Guardian approval for file transfer to the University of Edinburgh

20th March, 2012

Dear Alison

Re amendment to Caldicott application ref JMS/J1/10115

We are seeking approval for an amendment to our original Caldicott application for project title: Chronic Obstructive Pulmonary Disease in Lothian: a cohort study to identify modifiable risk factors in primary care that are associated with COPD emergency admission. Caldicott approval for this project was originally granted in August 2010 (ref: JMS/J1/10115).

The amendment is to cover the transfer of linked pseudonymised COPD data from NHS Lothian to University of Edinburgh to allow statisticians based there to use advanced statistical tests to carry out further analysis. This work is being funded by the CSO. The aims of the work are to identify modifiable risk factors and interventions that reduce the risk of hospital admission and repeat admission. This project will extend the current work, taking account of time varying covariates and the interdependence of exacerbations occurring within the same patient that were beyond the scope of the initial analyses.

The data will be transferred to the University and secured at the University using a strict protocol that has been agreed with yourself as Caldicott Guardian. This is detailed below. It should be noted that only copies of files are being transferred; the database will still remain in the control of NHS Lothian.

Approval is currently being sought from PAC for this data transfer. NHS Lothian R&D consider it to be a minor amendment to the original application and have requested that we inform MREC of this once Caldicott approval is obtained.

Dr Chris Weir, Associate Director (Statistics), MRC Hub for Trials Methodology Research, Centre for Population Health Sciences, University of Edinburgh Medical School, and co-applicant for the CSO project, will be overseeing the analyses. A further named statistician will be working on the database. We will be able to forward you the details of this person once they are appointed. Dr Sarah Wild, co-applicant on the HSRU funded project, and collaborator on CSO project will be acting as data intermediary at University of Edinburgh. We confirm that access to the data at the University will be restricted to these named individuals and that they have/will have received appropriate data governance training before being given access to the data.

Protocol for file transfer from NHS Lothian to University of Edinburgh:

1. Appropriate approvals obtained for the transfer and access to data from PAC and NHS Lothian Caldicott guardian.
2. A COPD research project specific folder from which all analyses will be performed has been set up at Edinburgh University on a secure university server. Access will be limited to data intermediary (Dr Sarah Wild) and named analysts (Dr Chris Weir and statistician name tbc). The server is regularly backed-up.
3. Data files will be transferred using NHS net email from Dr Leonie Hunter to Dr Sarah Wild, who will upload the files to the research project folder (it is not anticipated that file size will be a problem for this transfer but if necessary larger files will be transferred by FTP file transfer).
4. University policies on records management, retention and disposal of records and data protection <http://www.recordsmanagement.ed.ac.uk/> will be adhered to.

For the record, the university servers are stored in secure machine rooms. These are locked, and accessible only to authorised personnel using swipe card access. Backups are held in separate machine rooms in different sites, subject to the same access restrictions. These machine rooms have fire suppressants, security alarms, air conditioning (including humidity control). Data are stored on a Unix file system which is mounted on shared access systems. Full access control list support is available on the system, and this is configured to be restrictive by default by the system administrators. Data are stored in such a way that it would be difficult to reconstruct it from an individual drive if that failed and was returned to the supplier for replacement (all data is striped over a large number of volumes, each of which is made up of several disks which are themselves striped over several drives, thus splitting the data up into very small components and dispersing it over many drives).

We would be grateful if you could confirm as Caldicott guardian that you allow transfer of and access to the database as described above for these further analyses.

Yours sincerely,

Dr Rachel Hardie

Principal Investigator

Consultant in Public Health Medicine

Appendix 4.2 Request for Caldicott Guardian approval for PhD data use

20th March, 2012

Dear Professor McCallum

Re Further amendment to Caldicott application ref JMS/J1/10115

I am seeking approval for a further amendment to an original Caldicott application for project title: Chronic Obstructive Pulmonary Disease in Lothian: a cohort study to identify modifiable risk factors in primary care that are associated with COPD emergency admission. Caldicott approval for this project was originally granted in August 2010 (ref: JMS/J1/10115). A subsequent amendment was approved on 20th March 2012 allowing the data to be transferred and stored on a University of Edinburgh server.

This further request for amendment of the project would enable further analysis to be conducted on the data by a PhD student at the University of Edinburgh, Susannah McLean, who is supervised by Professor Aziz Sheikh, Professor Sarah Wild and Dr Colin Simpson. whose project involves adapting a Dutch model to predict the population burden and costs of COPD in Scotland and Lothian in 5 and 10 years time. Following approval from you as Caldicott Guardian, PAC and the ethics committee I would use selected linked anonymised data on the server to enable me to:

1. estimate incidence of COPD by age, sex and SIMD using population data for participating practices in order to provide estimates for Scotland as a whole.
2. describe patterns of lung function over time and their relation to exacerbation
3. Estimate annual age and sex specific mortality among a cohort of people with COPD
4. Estimate costs and QALYs generated by COPD burden

The project would involve me taking these aggregated data to Erasmus University, Rotterdam, the Netherlands to use the statistical program *Mathematica* to run the model with the support of the Dutch team. Aggregated data for other inputs to the model will be obtained from a variety of sources (please

see attached table for more details) and patient identifiable data will not be used in this modeling process. I propose taking the aggregated data to Rotterdam on an encrypted hard drive.

Please let me know if you require any further information or have recommendations before you can approve the request and I look forward to hearing from you

Yours sincerely,

Dr Susannah McLean MB BS MRCP (on behalf of supervisors)

PhD student

Summary Data Table

Data Item	Data title	Scottish/Lothian Run
1	General Population Data	National Records Scotland
2	Proportions of population that smoke, former smoker and never smoker by age and gender	Scottish Health Survey
3a	GP prevalence of COPD, age and gender specific	Information services division prevalence data
3b	GP incidence of COPD age and gender specific	Lothian cohort data and practice populations
4	Relative risks of smokers and nonsmokers	Use Rotterdam data
5	Age and sec specific start stop and restart probabilities for smoking	Use Rotterdam data
6	Random effects model to predict lung function decline	Recalculate model using COPD Lothian data
7	COPD related maintenance costs	Use Rotterdam data
8	Mean utility scores by COPD severity stage	Use Rotterdam data
9	All-cause mortality, mortality attributable to COPD mortality and mortality from other causes, by age and sex (per 1000 COPD patients) (i.e. what do COPD patients die of?)	Lothian cohort data

Calculated parameters table

Parameter	Name	Alternative source

A	Rate of Exacerbation by COPD stage	Individual patient data calculation, potentially a fresh model using Lothian cohort data
B	Case fatality rate	Develop age stratified case fatality profile of either Lothian Cohort from linked death certificate data rather than modelling
C	Decline in FEV1 due to exacerbation	Calculate using Lothian cohort (Average lung function decline per year/no. of exacerbations per year)= average lung function decline per exacerbation

Appendix 4.3 Initial reply from Caldicott Guardian May 2013

Sent on behalf of Professor Alison McCallum

Dear Susannah

Thank you for your letter of March 20, 2013. It came in when I was on leave so I am sorry about the delay in getting back to you. If these are de-identified tabulated data then they should still be transferred by secure file transfer protocol from Edinburgh to Rotterdam.

Transferring data in person by encrypted hard drive in a locked case would be a last resort.

Whichever approach is used, a data processing agreement with Erasmus will be required and will need to be checked by the lawyers for Erasmus and Central Legal Office. Erasmus will have a standard document in English.

A full list of variables will also be required and all of the documentation will obviously have to be signed by Sarah and Aziz.

Kind regards

Alison

Denise Foley

Caldicott Administrator

Tel 0131 465 (3)5452

Lothian NHS Board

Waverley Gate

2-4 Waterloo Place

Edinburgh

EH1 3EG

Please note that I am not in the office on a Thursday or Friday

Appendix 4.4 Clarification of data use for PhD to Caldicott

Dear Professor McCallum

Re Revised application for amendment to Caldicott application ref JMS/J1/10115

Thank you for considering my request to use data from the Lothian COPD cohort to estimate COPD burden in Scotland and the UK using a model developed in the Netherlands in my PhD project. Based on my initial request you recommended obtaining a legal agreement for data sharing between Erasmus and Edinburgh University.

Since I submitted my request to you I have clarified the data I would like to use from the Lothian COPD cohort. I would simply like to request estimates of COPD incidence and 95% confidence intervals by sex, age in 5 year age groups from 35 -85+ years and Scottish SIMD quintile to allow me to generate estimates of COPD incidence for Scotland as a whole by applying the incidence estimates to Scottish population estimates. Leonie Hunter from Lothian is able to calculate Lothian incidence estimates using data from the cohort and participating practice population data provided to her by Bill Ramsay. I will then take the Scottish estimates to the Netherlands to use in the model.

Examples of the data formats are given below.

Thank you for considering this revised application which, given the aggregated nature of the data that will be used, presumably will not require the legal agreements necessary for use of individual level data. Please contact me if you require any further information.

Yours sincerely,

Susannah Mclean

Student of Sarah Wild, Aziz Sheikh and Colin Simpson

Lothian Cohort format (to be used to generate Scottish estimates)

sex	ageband	scsimd2009 quintiles	Incidence per		
			100000 population	Lower 95%CI	Upper 95%CI
F	35-39	1			
F	35-39	2			
F	35-39	3			
F	35-39	4			
F	35-39	5			

Scotland data format (to be used in Dutch model)

sex	Age band	Incidence per		
		1000 population	Lower 95%CI	Upper 95%CI
F	35-39			
F	40-44			
F	45-49			
F	50-54			

F	55-59
F	60-64
F	65-69
F	70-74
F	75-79
F	80-84
F	85+
M	35-39
M	40-44
M	45-49
M	50-54
M	55-59
M	60-64
M	65-69
M	70-74
M	75-79
M	80-84
M	85+

Appendix 4.5 Caldicott Guardian Approval August 2014

Dear Susannah

I have heard back from Professor Alison McCallum with regards to the amendment to Caldicott application, ref 10115. This amendment has been approved with the following caveat.

'If the tabulated data needs to be transferred using a mobile device then evidence of compliance with the relevant circular from 2012 is required. This means using an encrypted laptop and NHS encrypted memory stick.'

Best wishes

Denise

Denise Foley

Caldicott Administrator

Tel 0131 465 (3)5452

Lothian NHS Board

Waverley Gate

2-4 Waterloo Place

Edinburgh

EH1 3EG

Appendix 5 Description of method of generating random incidence and prevalence curves for COPD

Rudolf Hoogenveen,

25 March 2014

Introduction

The research question was to generate age-curves of incidence and prevalence probability values for England and Scotland, based on age-specific data, and taking account of uncertainty. Uncertainty can be defined in different ways here. I have assessed the uncertainty of the population mean values over age, specified by gender.

Method

I have applied the same method for all data, i.e. initial prevalence probability values for England and Scotland, and incidence rates for England. All data available were specified by gender and age, the latter being different between England and Scotland data. I stuck to the age-specifications applied in the data.

The calculation steps applied were:

- I calculated a smooth curve through the age-specific data, using the R-routine `smooth.spline` (spline-based method, R Version 3.0.1). The smoothing parameter used was `spar0`, and the value was selected to optimize both fit and smoothness. For values, see below. I used the mid-points of the age-classes as the independent variable.
- Next, I applied the same procedure for each generated random curve. That means, I randomly drew event (prevalence or incident) numbers for each age-class. In case of prevalence, I drew from binomial distributions with parameters sample size and observed prevalence probability numbers, i.e. for each age-class. In case of

incidence, I drew from Poisson distributions with parameter observed event number. Then again, I calculated a smooth curve through the randomly drawn age-specific data using the same smoothing parameter value.

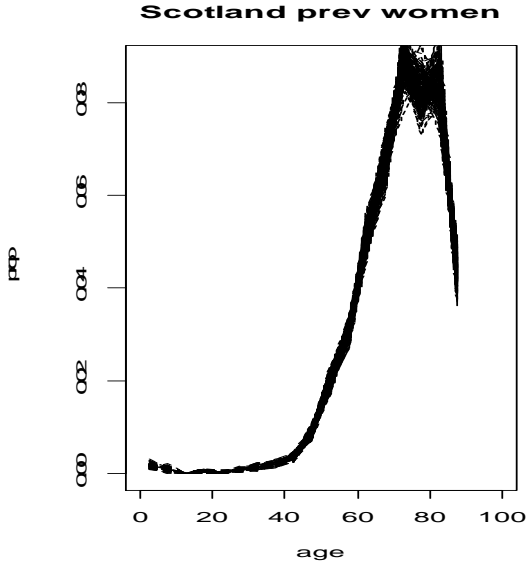
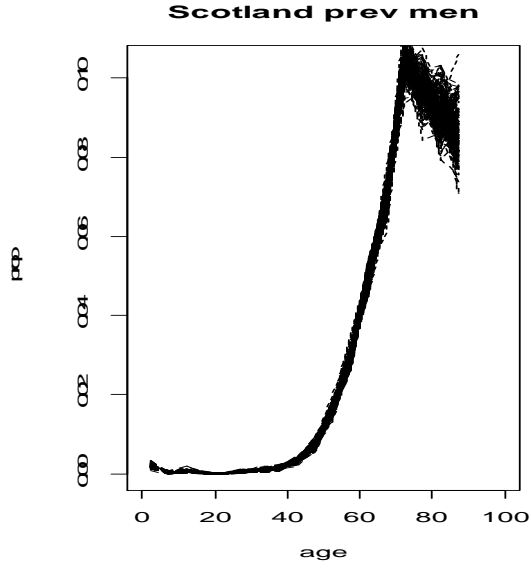
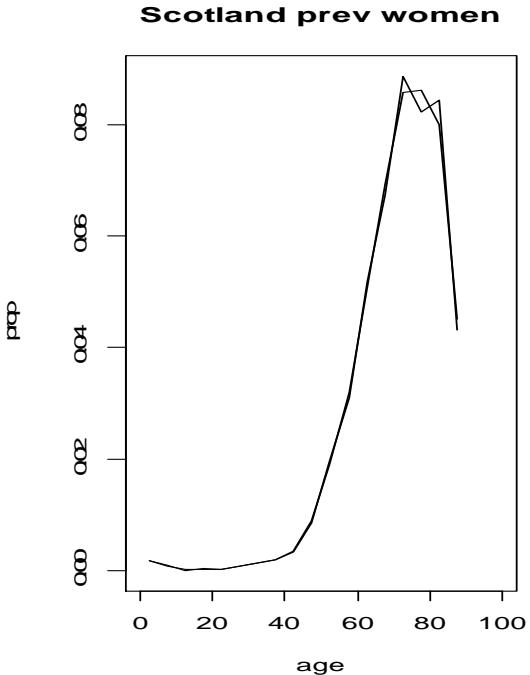
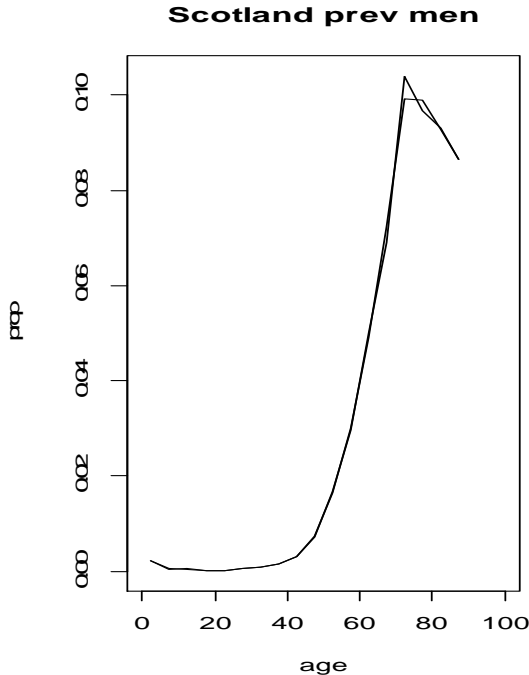
Smoothing parameter values used: Scotland: $\text{spar0} = .25$, England: $\text{spar0} = .6$. The number of randomly drawn curves was $\text{nMC} = 200$.

For the highest ages I applied the following procedure. For Scotland, I assumed the mid-point of the highest age-class (85+) being 87.5 years. For England, I aggregated all numbers for ages 100+, with mid-point age 102 years.

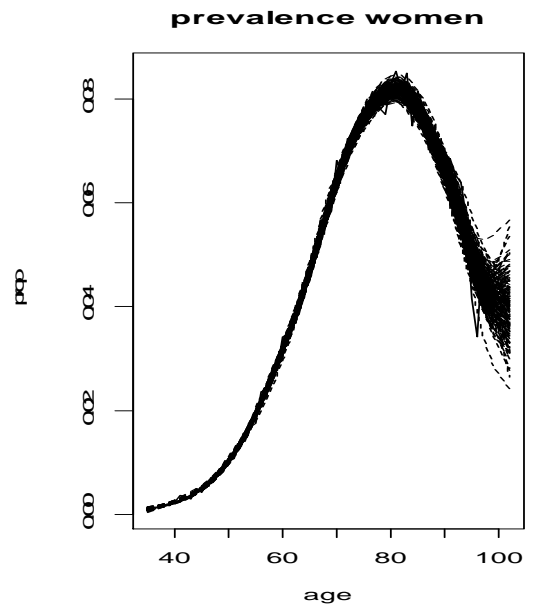
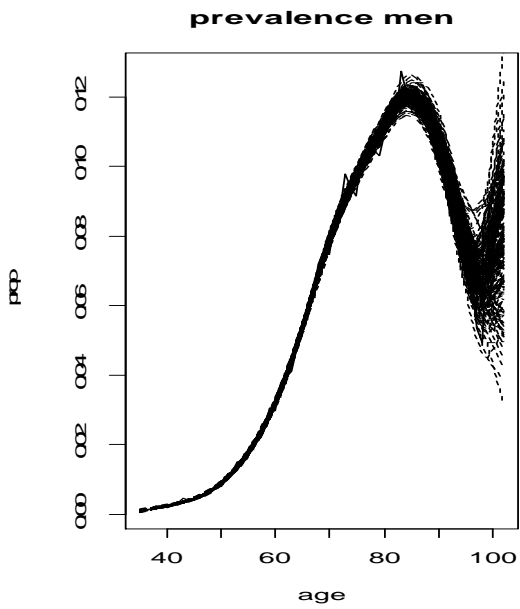
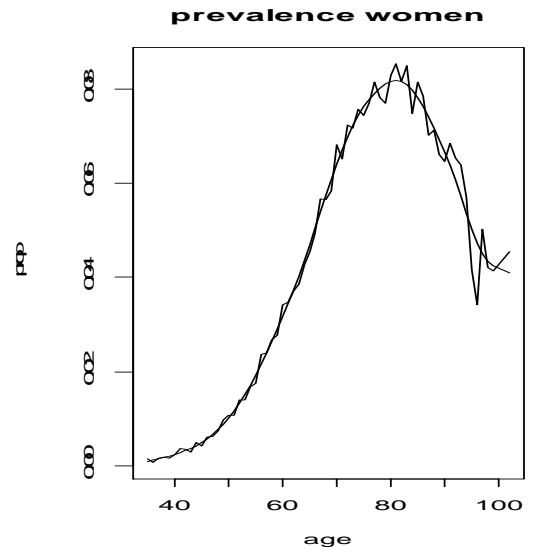
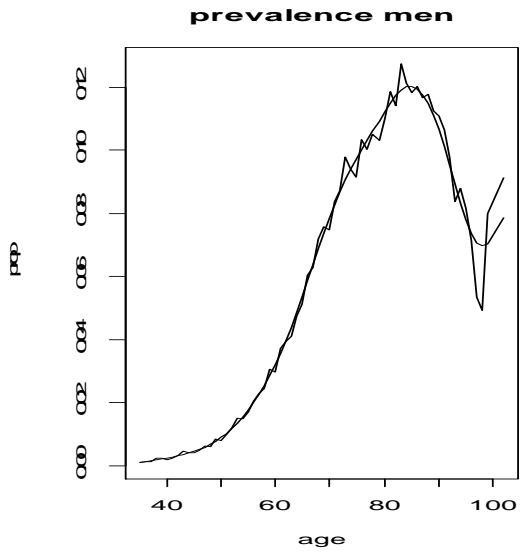
Results

For each parameter (incidence, prevalence, country) I present below graphical results without and with randomly drawn curves.

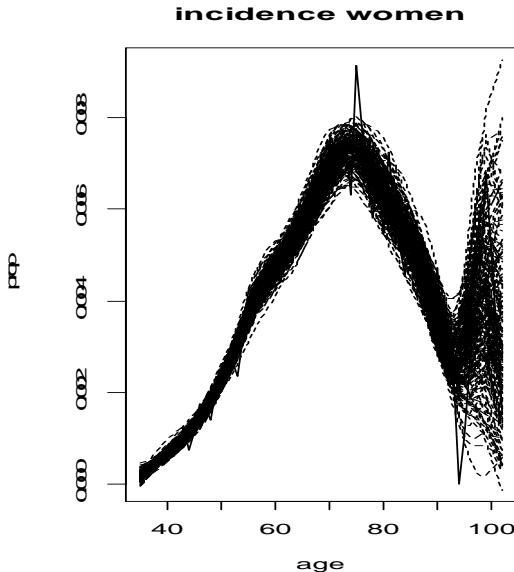
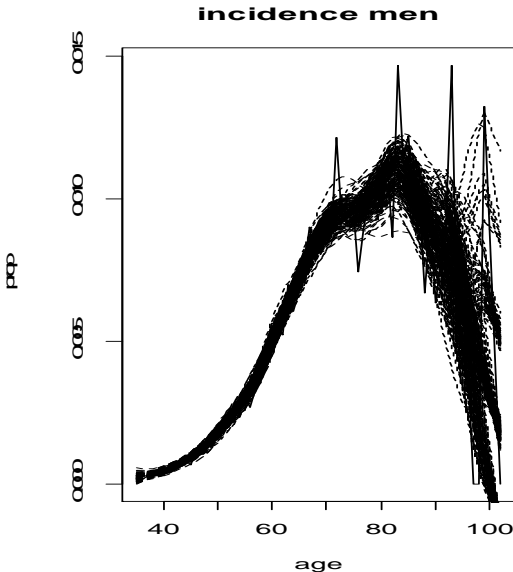
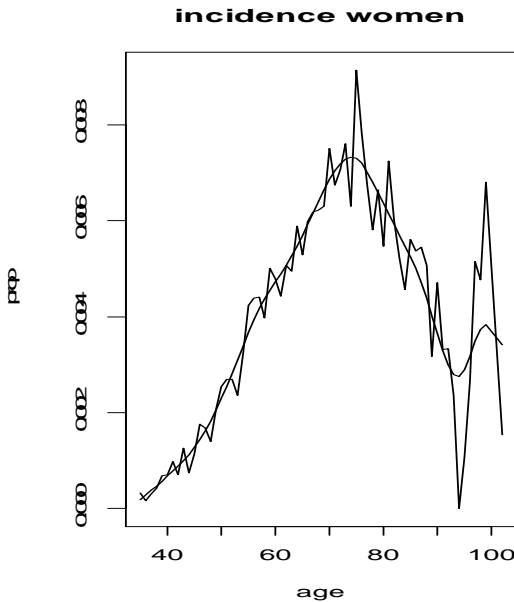
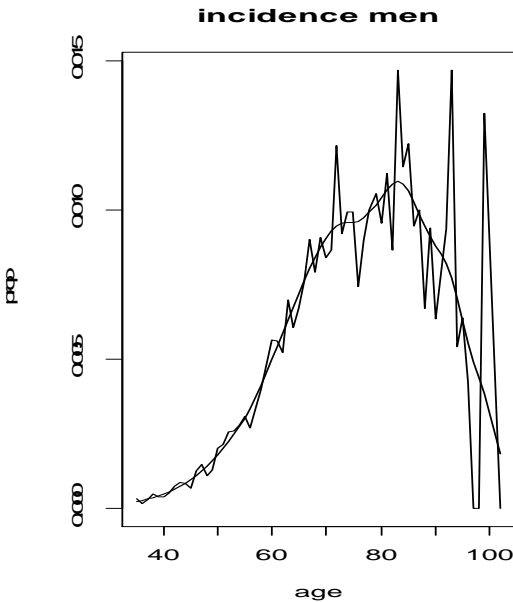
Prevalence, Scotland



Prevalence, England



Incidence, England



Conclusion & discussion

This is a first solution to the research question stated above. There are some smaller issues to be taken care of, e.g. the smoothing procedure sometimes generates negative probability

values. The results for excess mortality probability values are still to come. I have presented graphical results only here. Numerical values are of course available.

Appendix 6 Contribution to science

Publications

Published: McLean, S., S. H. Wild, C. R. Simpson, A. Sheikh. (2013). "Models for estimating projections for the prevalence and disease burden of chronic obstructive pulmonary disease (COPD): systematic review protocol." *Prim Care Respir J* **22**(2): PS8-21.

Accepted: McLean, S, V. Barbour, S. H Wild, C. R Simpson, A. Sheikh. Models for estimating projections for disease prevalence and burden: a systematic review focusing on chronic obstructive pulmonary disease. *The Journal of Health Services Research and Policy*

Policy

My data were used in a series of briefings that Lothian Public Health supplied to the Strategic Planning Group as part of the formulation of the NHS Lothian Strategic Plan: <http://www.nhslothian.scot.nhs.uk/OurOrganisation/OurHealthOurCareOurFuture/SupportingDocuments/OurHealthOurCareOurFuture/Appendix%201%20-%2020population%20and%20disease%20projections.pdf>

Talks

4th March 2014 Presentation to School of Population Health Sciences, University of Edinburgh, Teviot Place, Edinburgh "Chronic Obstructive Pulmonary Disease; Modelling prevalence in England and Scotland"

31st March 2014 Continuing professional development in public health, presentation at the Scottish Government, St Andrew's House, Edinburgh, "Chronic Obstructive Pulmonary Disease; Modelling prevalence in England and Scotland"

25th April 2014 Presentation to NHS Lothian COPD Managed Care Network, Effectivity and Productivity Meeting, Little France, Edinburgh "Chronic Obstructive Pulmonary Disease; Modelling prevalence in England and Scotland"

Poster

9th September 2014 European Respiratory Society Conference, Munich, "Burden of chronic obstructive pulmonary disease in Scotland and England"

Glossary

Birthrate a summary rate based on the number of live births in a population over a given period, usually 1 year.(Porta 2014)

$$\text{Birthrate} = \frac{\text{Number of live births to residents in an area in a calendar year}}{\text{Average or midyear population in the area in that year}} * 1000$$

Difference equation an equation expressing a functional relationship of one or more independent variables, one or more functions dependent on these variables, and successive differences of these functions.(McGraw-Hill 2003)

Disability adjusted life year (DALY) DALYs are the opposite of QALYs (see QALYs below) DALYs are a measure of health gap. On the DALY scale, 1.0 is total disability and 0 is the target of zero disability. The DALYs are calculated from the measure of disability on the DALY scale multiplied by the number of life years at that level of disability.(Gold, Stevenson et al. 2002)

Functional data analysis is a branch of statistics that analyses data providing information about curves, surfaces or anything else that varies over a continuum. The continuum is often time, but may also be spatial location, wavelength, probability, etc.

Incidence the rate at which new events occurs in a population. The numerator is the number of new events that occur in a defined period. The denominator is the person-time at risk of the event during this period.

$$\text{Incidence} = \frac{\text{Number of new events in specified period}}{\text{Average number of persons exposed to risk during this period}} * 10^n$$

Strictly, this ratio is neither a rate nor a proportion but is instead the rate multiplied by the length of the specified period. If the period is a year the ratio is called the **annual incidence rate**. The average size of the population is often the estimated population at the midpoint of the time period. If the number of new cases during a specified period is divided by the sum of the person time units at risk for all persons during the period the result is the **person-time incidence rate**.(Porta 2014)

Model a system of postulates, data and inferences presented as a mathematical description of an entity or state of affairs.(Merriam-Webster 2013)

Prevalence A measure of occurrence of disease or risk factor. The total number of individuals who have the condition or risk factor at a particular time, divided by the population at risk of having the condition at that time, or midway through the period. It is a proportion not a rate. **Point Prevalence** is the proportion of individuals who have the condition at a specified point in time. (Porta 2014)

Quality adjusted life year (QALY) QALYs rely on a generic quantification of health status. The health status is a functional level or “utility”. This health status is quantified by assuming that the person without illness is unimpaired or has 1.0 utility and that a person in state 0 is dead. QALYs are calculated from the utility or reduced utility as a result of a health condition, multiplied by the time spent at this utility, the life years.(Zeckhauser and Shepard 1976)

Spatial modelling refers to a particular form of disaggregation, in which an area is divided into a number (often a large number) of similar units: typically grid squares or polygons. The model may be linked to a geographical information system (GIS) for data input and display. (simulistics 2011)

Statistical mediation study a type of statistical study in which the influence of one variable on the second variable is determined by the effects of a third intermediate variable.

Stochastic process a process, which is usually a temporal sequence, which incorporates an some element of randomness. (Porta 2014)

References

- Appleton, S., P. Poole, et al. (2006). "Long-acting beta2-agonists for poorly reversible chronic obstructive pulmonary disease." Cochrane Database Syst Rev(3): CD001104.
- Atsou, K., C. Chouaid, et al. (2011). "Simulation-based estimates of effectiveness and cost-effectiveness of smoking cessation in patients with chronic obstructive pulmonary disease." PLoS One **6**(9): e24870.
- Atsou, K., C. Chouaid, et al. (2011). "Variability of the chronic obstructive pulmonary disease key epidemiological data in Europe: systematic review." BMC Med **9**: 7.
- Baines, K. J., A. C. Hsu, et al. (2013). "Novel immune genes associated with excessive inflammatory and antiviral responses to rhinovirus in COPD." Respir Res. **14:15**.(doi): 10.1186/1465-9921-1114-1115.
- Bakke, P. S., V. Baste, et al. (1991). "Prevalence of obstructive lung disease in a general population: relation to occupational title and exposure to some airborne agents." Thorax **46**(12): 863-870.
- Barendregt, J. J., G. J. Van Oortmarssen, et al. (2003). "A generic model for the assessment of disease epidemiology: the computational basis of DisMod II." Popul Health Metr **1**(1): 4.
- Barker, D. J. P., K. M. Godfrey, et al. (1991). "Relation of birth weight and childhood respiratory infection to adult lung function and death from chronic obstructive airways disease." Bmj **21**(303): 671-675.
- Barnett, K., S. W. Mercer, et al. (2012). "Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study." Lancet **380**(9836): 37-43.
- Barry, S. J., E. Dinnett, et al. (2013). "Are routinely collected NHS administrative records suitable for endpoint identification in clinical trials? Evidence from the West of Scotland Coronary Prevention Study." PLoS One **8**(9): e75379.
- Bednarek, M., J. Maciejewski, et al. (2008). "Prevalence, severity and underdiagnosis of COPD in the primary care setting." Thorax **63**(5): 402-407.
- Bertakis, K. D., R. Azari, et al. (2000). "Gender differences in the utilization of health care services." J Fam Pract **49**(2): 147-152.
- Black, C., J. Eunson, et al. (2010). "Scottish Schools Adolescent Lifestyle and Substance Use Survey (SALSUS) National Report." 43.
- Blanc, P. D., A. M. Menezes, et al. (2009). "Occupational exposures and COPD: an ecological analysis of international data." Eur Respir J **33**(2): 298-304.
- BLF (2007). *Invisible Lives: Chronic Obstructive Pulmonary Disease (COPD) finding the missing millions*, British Lung Foundation.
- Bobrow, M. (2013). "Balancing privacy with public benefit." Nature **500**(7461): 123.
- Borland, S. (2014). NHS delays plan to harvest your details: Victory for the Mail as database is shelved for six months. Mail Online.
- Bradshaw, P., C. Bromley, et al. (2011). *Scottish Health Survey, Scottish Government. I: Adults*.
- Briggs, A. and M. Sculpher (1998). "An introduction to Markov modelling for economic evaluation." Pharmacoeconomics **13**(4): 397-409.
- BristolUniversity. (1991). "Avon Longitudinal Study of Parents and Children." Retrieved 2013, from <http://www.bristol.ac.uk/alspac/>.
- Bronnum-Hansen, H. and K. Juel (2000). "Estimating mortality due to cigarette smoking: two methods, same result." Epidemiology **11**(4): 422-426.
- Brooks, R., R. Rabin, et al. (2003). *The measurement and valuation of health status using EQ-5D: a European perspective; evidence from the EuroQol BIOMED Research Programme*. Dordrecht, the Netherlands.

- Brotons, B., J. A. Perez, et al. (1994). "[The prevalence of chronic obstructive pulmonary disease and asthma. A cross-sectional study]." *Arch Bronconeumol* **30**(3): 149-152.
- Brownstein, J. S., S. N. Murphy, et al. (2010). "Rapid identification of myocardial infarction risk associated with diabetes medications using electronic medical records." *Diabetes Care* **33**(3): 526-531.
- Brownstein, J. S., M. Sordo, et al. (2007). "The tell-tale heart: population-based surveillance reveals an association of rofecoxib and celecoxib with myocardial infarction." *PLoS One* **2**(9): e840.
- Brusselle, G., G. Joos, et al. (2011). "New insights into the immunology of chronic obstructive pulmonary disease." *Lancet*. **378**(9795): 1015-1026.
- Bruyneel, M., V. Jacob, et al. (2011). "Hoover's sign is a predictor of airflow obstruction severity and is not related to hyperinflation in chronic obstructive pulmonary disease." *Eur J Intern Med* **22**(6): e115-118.
- BTS (2006). *The Burden of Lung Disease*. n. Edition, British Thoracic Society.
- Buist, A. S., M. A. McBurnie, et al. (2007). "International variation in the prevalence of COPD (the BOLD Study): a population-based prevalence study." *Lancet* **370**(9589): 741-750.
- Calik-Kutukcu, E., S. Savci, et al. (2014). "A comparison of muscle strength and endurance, exercise capacity, fatigue perception and quality of life in patients with chronic obstructive pulmonary disease and healthy subjects: a cross-sectional study." *BMC Pulm Med* **14**: 6.
- Calverley, P. M., J. A. Anderson, et al. (2007). "Salmeterol and fluticasone propionate and survival in chronic obstructive pulmonary disease." *N Engl J Med* **356**(8): 775-789.
- Calverley, P. M. and J. A. Wedzicha (2007). "Chronic obstructive pulmonary disease past, present and future." *Thorax* **62**(12): 1026-1027.
- Campbell, J., D. J. Dedman, et al. (2013). "Is the CPRD GOLD Population Comparable to the U.K. Population?" *Pharmacoepidemiology and drug safety* **22**(s1): A 567.
- Cazzola, M., E. Puxeddu, et al. (2011). "The prevalence of asthma and COPD in Italy: a practice-based study." *Respiratory Medicine* **105**(3): 386-391.
- Celli, B. R., R. J. Halbert, et al. (2003). "Population impact of different definitions of airway obstruction." *Eur Respir J* **22**(2): 268-273.
- Centers_for_disease_control_and_prevention (2001). *Women and Smoking: A report of the Surgeon General, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Helath Promotion, Office on Smoking and Health*.
- Cerveri, I., S. Accordini, et al. (2003). "Chronic cough and phlegm in young adults." *Eur Respir J* **22**(3): 413-417.
- Cerveri, I., S. Accordini, et al. (2001). "Variations in the prevalence across countries of chronic bronchitis and smoking habits in young adults." *Eur Respir J* **18**(1): 85-92.
- Chapmann, K. R., D. M. Mannino, et al. (2006). "Epidemiology and costs of chronic obstructive pulmonary disease." *European Respiratory Journal* **27**: 188-207.
- Cheek, B. (2014). "Spirometry." Retrieved 6.1.2015, 2015, from <http://www.gp-training.net/protocol/respiratory/copd/spirometry.htm>.
- Cimen, P., M. Unlu, et al. (2014). "Should Patients With COPD Be Vaccinated?" *Respir Care*.
- Clark, T. G. and D. G. Altman (2003). "Developing a prognostic model in the presence of missing data: an ovarian cancer case study." *J Clin Epidemiol* **56**(1): 28-37.
- Cleland, J., M. Moffat, et al. (2012). "A qualitative study of stakeholder views of a community-based anticipatory care service for patients with COPD." *Prim Care Respir J* **21**(3): 255-260.
- Colhoun, H. M., S. J. Livingstone, et al. (2012). "Hospitalised hip fracture risk with rosiglitazone and pioglitazone use compared with other glucose-lowering drugs." *Diabetologia* **55**(11): 2929-2937.

- ConnectingForHealth. (2013). "Read Codes." from <http://www.connectingforhealth.nhs.uk/systemsandservices/data/uktc/readcodes/index.html#1>.
- Cook, D. G., D. P. Strachan, et al. (1998). "Health effects of passive smoking. 9. Parental smoking and spirometric indices in children." *Thorax* **53**(10): 884-893.
- Coory, M., D. Youlden, et al. (2002). "Interpretation of hospital-specific outcome measures based on routine data." *Australian Health Review* **25**(4): 69-72.
- CPRD, C. p. r. d. (2011). Incidence of COPD in GOLD dataset.
- CPRD, C. p. r. d. (2011). Prevalence of COPD in GOLD dataset.
- Cresswell, K., M. Ali, et al. (2011). The Long and Winding Road... An independent evaluation of the implementation and adoption of the national health service care records service in secondary care in England, Connecting for health, University of Birmingham.
- Cresswell, K., Z. Morrison, et al. (2012). ""There are too many, but never enough": qualitative case study investigating routine coding of clinical information in depression." *PLoS One* **7**(8): e43831.
- Cunningham, S., R. McAlpine, et al. (2011). "Using web technology to support population-based diabetes care." *J Diabetes Sci Technol* **5**(3): 523-534.
- Curtis, L. H., J. Brown, et al. (2014). "Four health data networks illustrate the potential for a shared national multipurpose big-data network." *Health Aff (Millwood)* **33**(7): 1178-1186.
- de Lusignan, S. (2005). "The barriers to clinical coding in general practice: a literature review." *Med Inform Internet Med* **30**(2): 89-97.
- de Lusignan, S., C. Pearce, et al. (2011). "What are the barriers to conducting international research using routinely collected primary care data?" *Stud Health Technol Inform* **165**: 135-140.
- de Lusignan, S. and C. van Weel (2006). "The use of routinely collected computer data for research in primary care: Opportunities and challenges." *Family Practice* **23**(2): 253-263.
- de Marco, R., G. Pesce, et al. (2013). "The coexistence of asthma and chronic obstructive pulmonary disease (COPD): prevalence and risk factors in young, middle-aged and elderly people from the general population." *PLoS One* **8**(5): e62985.
- de Torres, J. P., C. Casanova, et al. (2014). "Prognostic evaluation of COPD patients: GOLD 2011 versus BODE and the COPD comorbidity index COTE." *Thorax* **69**(9): 799-804.
- Dept_of_Health (1997). The Caldicott Report, NHS.
- Dept_of_Health. (2012). "2003-2004 to 2010-2011 Programme Budgeting Data." Retrieved 5th Sept 2014, from <https://www.gov.uk/government/publications/2003-04-to-2010-11-programme-budgeting-data>.
- Dept_of_Health (2013). Information: To share or not to share, Caldicott Information Governance Review.
- Dept_of_Health. (2013). "Programme Budgeting Data." Retrieved 5th Sept 2014, from <http://www.pcc-cic.org.uk/external?url=http://www.england.nhs.uk/resources/resources-for-cggs/prog-budgeting/>.
- Dept_of_Health (2014). Living well for longer: National support for local action to reduce premature avoidable mortality.
- Dickinson, J. A., M. Meaker, et al. (1999). "Screening older patients for obstructive airways disease in a semi-rural practice." *Thorax* **54**(6): 501-505.
- DoH (2010). Improving the health and wellbeing of people with long term conditions, Department of Health.
- Doll, R., R. Peto, et al. (2004). "Mortality in relation to smoking: 50 years' observations on male British doctors." *Bmj* **328**(7455): 1519.

- Dugas, M., M. Lange, et al. (2010). "Routine data from hospital information systems can support patient recruitment for clinical studies." *Clinical Trials* **7**(2): 183-189.
- Editorial (2007). "Beyond the lungs - a new view of COPD." *The Lancet* **370**(713).
- Editorial (2011). "Variable loss of lung function in COPD." *N Engl J Med*. **365**(13): 1246-1247. Epub 2011 Sep 1226.
- EHIIntelligence (2012). Market Forecast Report for English NHS IT Market.
- Elkin, P. L., B. E. Trusko, et al. (2010). "Secondary use of clinical data." *Studies in Health Technology and Informatics* **155**: 14-29.
- EMIS. (2015). Retrieved 6.1.2015, 2015, from <http://www.emis-online.com/company-profile>.
- Erbas, B., S. Ullah, et al. (2012). "Forecasts of COPD mortality in Australia: 2006-2025." *BMC Med Res Methodol* **12**: 17.
- ERPHO, H. Walford, et al. (2011). COPD Prevalence Modelling Briefing Document, East of England Public Health Observatory.
- ESRC, E. a. S. R. C. (2011). "UK Data service." from <http://www.esds.ac.uk/newRegistration/newLogin.asp>.
- Ezzati, M. and A. D. Lopez (2003). "Measuring the accumulated hazards of smoking: global and regional estimates for 2000." *Tob Control* **12**(1): 79-85.
- FCA. (2015). "WHO Framework Convention on Tobacco Control." Retrieved 12.1.2015, from <http://www.fctc.org/about-fca/tobacco-control-treaty/latest-ratifications/parties-ratifications-accessions>.
- Feenstra, T. L., M. L. van Genugten, et al. (2001). "The impact of aging and smoking on the future burden of chronic obstructive pulmonary disease: a model analysis in the Netherlands." *Am J Respir Crit Care Med* **164**(4): 590-596.
- Fernando, B., D. Kalra, et al. (2012). "Benefits and risks of structuring and/or coding the presenting patient history in the electronic health record: systematic review." *BMJ Qual Saf* **21**(4): 337-346.
- Fletcher, C. and R. Peto (1977). "The natural history of chronic airflow obstruction." *Br Med J* **1**(6077): 1645-1648.
- Fletcher, C. and R. Peto (1977). "The natural history of chronic airflow obstruction." *BMJ* **1**: 1645-1648.
- Fletcher, C. H. (1978). "Terminology in chronic obstructive lung diseases." *J Epidemiol Community Health* **32**(4): 282-288.
- Ford, D. V., K. H. Jones, et al. (2009). "The SAIL Databank: Building a national architecture for e-health research and evaluation." *BMC Health Services Research* **9**(157).
- Freund, K. M., A. J. Belanger, et al. (1993). "The health risks of smoking. The Framingham Study: 34 years of follow-up." *Ann Epidemiol* **3**(4): 417-424.
- Galobardes, B., P. McCarron, et al. (2008). "Association between early life history of respiratory disease and morbidity and mortality in adulthood." *Thorax* **63**(5): 423-429.
- Ganesan, S., A. T. Comstock, et al. (2013). "Barrier function of airway tract epithelium." *Tissue Barriers* **1**(4): e24997.
- Garcha, D., S. J. Thurston, et al. (2012). "Changes in prevalence and load of airway bacteria using quantitative PCR in stable and exacerbated COPD." *Thorax* **3**: 3.
- Garcia-Rio, F., J. B. Soriano, et al. (2011). "Overdiagnosing subjects with COPD using the 0.7 fixed ratio: correlation with a poor health-related quality of life." *Chest* **139**(5): 1072-1080.
- Giacomini, M., D. DeJean, et al. (2012). "Experiences of Living and Dying With COPD: A Systematic Review and Synthesis of the Qualitative Empirical Literature." *Ont Health Technol Assess Ser* **12**(13): 1-47.
- Gibson, P. G. and J. L. Simpson (2009). "The overlap syndrome of asthma and COPD: what are its features and how important is it?" *Thorax* **64**(8): 728-735.

- Global Initiative for Chronic Obstructive Lung Disease (2008). Global strategy for the diagnosis, management, and prevention of chronic pulmonary disease. Updated 2008. <http://www.goldcopd.com> [accessed 1 December 2009], NHLBI/WHO.
- Global Burden of Disease (2012). "Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010." *Lancet*. **380**(9859): 2095-2128. doi: 10.1016/S0140-6736(2012)61728-61720.
- GMC (2003). The standard general medical services contract.
- GOLD (2010). "Global Initiative for Chronic Obstructive Lung Disease: Pocket guide to COPD diagnosis management and prevention."
- GOLD (2010). Pocket Guide to COPD Diagnosis, Management and Prevention, Global Initiative for Chronic Obstructive Lung Disease.
- GOLD (2013). Pocket guide to COPD diagnosis, management and prevention; A Guide for Health Care Professionals.
- Gold, M. R., D. Stevenson, et al. (2002). "HALYs and QALYs and DALYs, Oh My: Similarities and differences in summary measures of population health." *Annual Review of Public Health* **23**: 115-134.
- Goossens, L. M., I. Leimer, et al. (2014). "Does the 2013 GOLD classification improve the ability to predict lung function decline, exacerbations and mortality: a post-hoc analysis of the 4-year UPLIFT trial." *BMC Pulm Med* **14**: 163.
- Government Statistics. (2009). "Scottish Index of Multiple Deprivation." from <http://www.scotland.gov.uk/Topics/Statistics/SIMD>.
- Grant, A., J. Ure, et al. (2013). "Acceptability and perceived barriers and facilitators to creating a national research register to enable 'direct to patient' enrolment into research: the Scottish Health Research Register (SHARE)." *BMC Health Serv Res* **13**: 422.
- GROS. (2012). "General Register Office for Scotland Vital Events." from <http://www.gro-scotland.gov.uk/statistics/theme/vital-events/index.html>.
- Gupta, R. (2014). Telephone call.
- Haddow, G., A. Bruce, et al. (2011). "'Nothing is really safe': A focus group study on the processes of anonymizing and sharing of health data for research purposes." *Journal of Evaluation in Clinical Practice* **17**(6): 1140-1146.
- Haddow, G., G. Laurie, et al. (2007). "Tackling community concerns about commercialisation and genetic research: a modest interdisciplinary proposal." *Soc Sci Med* **64**(2): 272-282.
- Hansell, A., J. Hollowell, et al. (2003). "Validity and interpretation of mortality, health service and survey data on COPD and asthma in England." *European Respiratory Journal* **21**(2): 279-286.
- Hansell, A. L., J. A. Walk, et al. (2003). "What do chronic obstructive pulmonary disease patients die from? A multiple cause coding analysis." *Eur Respir J* **22**(5): 809-814.
- Hansen, J. G., L. Pedersen, et al. (2008). "The Prevalence of chronic obstructive pulmonary disease among Danes aged 45-84 years: population-based study." *COPD* **5**(6): 347-352.
- Hasselgren, M., M. Arne, et al. (2001). "Estimated prevalences of respiratory symptoms, asthma and chronic obstructive pulmonary disease related to detection rate in primary health care." *Scand J Prim Health Care* **19**(1): 54-57.
- Haux, R., A. Hein, et al. (2014). "Information and communication technologies for promoting and sustaining quality of life, health and self-sufficiency in ageing societies--outcomes of the Lower Saxony Research Network Design of Environments for Ageing (GAL)." *Inform Health Soc Care* **39**(3-4): 166-187.
- Hayes, T. L., K. Cobbinah, et al. (2009). "A study of medication-taking and unobtrusive, intelligent reminding." *Telemed J E Health* **15**(8): 770-776.

- Healthcare Commission (2006). Clearing the air: A national study of chronic obstructive pulmonary disease, The Healthcare Commission.
- Hedman, J., J. Kaprio, et al. (1999). "Prevalence of asthma, aspirin intolerance, nasal polyposis and chronic obstructive pulmonary disease in a population-based study." Int J Epidemiol **28**(4): 717-722.
- Heijmans, M. J. W. M., P. Spreeuwenberg, et al. (2005). Monitor Zorg- en Leefsituatie van mensen met astma en mensen met COPD. Trends en ontwikkelingen over de periode 2001 - 2004. [Trends in the living condition and health care of people with asthma and COPD 2001 - 2004]. Utrecht, NIVEL.
- Herfs, M., P. Hubert, et al. (2012). "Proinflammatory cytokines induce bronchial hyperplasia and squamous metaplasia in smokers: implications for chronic obstructive pulmonary disease therapy." Am J Respir Cell Mol Biol **47**(1): 67-79.
- Hettle, R., H. Wouters, et al. (2012). "Cost-utility analysis of tiotropium versus usual care in patients with COPD in the UK and Belgium." Respiratory Medicine **106**: 1733-1733.
- Hewitt, N. (2014). Lothian Policy on COPD.
- Hill, E. M., E. L. Turner, et al. (2013). "'Let's get the best quality research we can': public awareness and acceptance of consent to use existing data in health research: a systematic review and qualitative study." BMC medical research methodology **13**: 72.
- Hill, K., R. S. Goldstein, et al. (2010). "Prevalence and underdiagnosis of chronic obstructive pulmonary disease among patients at risk in primary care." CMAJ Canadian Medical Association Journal **182**(7): 673-678.
- HL7. (2015). "Health Level Seven International." Retrieved 7.1.2015, from <http://www.hl7.org/>.
- HMTreasury (2010). Spending Review 2010, HM Treasury.
- Holleman, D. R., Jr. and D. L. Simel (1995). "Does the clinical examination predict airflow limitation?" Jama **273**(4): 313-319.
- Holt, J. B., X. Zhang, et al. (2011). "Geographic disparities in chronic obstructive pulmonary disease (COPD) hospitalization among Medicare beneficiaries in the United States." International Journal of Copd **6**: 321-328.
- Hoogendoorn, M., T. L. Feenstra, et al. (2010). "Association between lung function and exacerbation frequency in patients with COPD." International Journal of Copd **5**: 435-444.
- Hoogendoorn, M., T. L. Feenstra, et al. (2006). "[Projection of future resource use and the costs of asthma and COPD in the Netherlands]." Ned Tijdschr Geneesk. **150**(2): 1243-1250.
- Hoogendoorn, M., R. T. Hoogenveen, et al. (2011). "Case fatality of COPD exacerbations: a meta-analysis and statistical modelling approach." Eur Respir J **37**(3): 508-515.
- Hoogendoorn, M., M. P. Rutten-van Molken, et al. (2010). Comparing the cost-effectiveness of a wide range of COPD interventions using a stochastic, dynamic, population model for COPD, Institute for Medical Technology Assessment, Erasmus University, Rotterdam, The Netherlands.
- Hoogendoorn, M., M. P. Rutten-van Molken, et al. (2011). "Developing and applying a stochastic dynamic population model for chronic obstructive pulmonary disease." Value Health **14**(8): 1039-1047.
- Hoogendoorn, M., M. P. Rutten-van Molken, et al. (2005). "A dynamic population model of disease progression in COPD." Eur Respir J **26**(2): 223-233.
- Hoogendoorn, M., M. P. M. H. Rutten-Van Molken, et al. (2011). "Developing and applying a stochastic dynamic population model for chronic obstructive pulmonary disease." Value Health. **14**(8): 1039-1047. Epub 2011 Sep 1022.

- Hoogendoorn, M., M. P. M. H. Rutten-van Molken, et al. (2011). "Developing and applying a stochastic dynamic population model for chronic obstructive pulmonary disease." *Value in Health* **14**(8): 1039-1047.
- Hoogendoorn, M., M. P. M. H. Rutten-van Molken, et al. (2005). "A dynamic population model of disease progression in COPD." *European Respiratory Journal* **26**(2): 223-233.
- Hooper, R., P. Burney, et al. (2012). "Risk factors for COPD spirometrically defined from the lower limit of normal in the BOLD project." *Eur Respir J* **39**(6): 1343-1353.
- HSCIC. (2014). "Hospital Episode Statistics." from <http://www.hscic.gov.uk/hes>.
- HSCIC, Q. b. r. t. (2013). "Data and Business Rules - Chronic Obstructive Pulmonary Disease (COPD) Indicator Set." Retrieved 12 November, 2014, from http://cdn.pcc-cic.org.uk/sites/default/files/articles/attachments/copd_ruleset_v27.1_0.pdf.pagespeed.ce.ljG9-ecs5n.pdf.
- HSE (2010). "The Health Survey for England."
- Huchon, G. J., A. Vergnenegre, et al. (2002). "Chronic bronchitis among French adults: high prevalence and underdiagnosis." *Eur Respir J* **20**(4): 806-812.
- Hunt, K., G. Ford, et al. (1999). "Are women more ready to consult than men? Gender differences in family practitioner consultation for common chronic conditions." *J Health Serv Res Policy* **4**(2): 96-100.
- Hurley, S. F. and J. P. Matthews (2007). "The Quit Benefits Model: a Markov model for assessing the health benefits and health care cost savings of quitting smoking." *Cost Eff Resour Alloc* **5**: 2.
- IHME. (2010). "DISMOD 3." from <http://winthrop.ihme.washington.edu/>.
- Information_Commissioner's_Office (2012). Anonymisation: managing data protection risk Code of Practice, The Information Commission.
- INPS4. (2015). "Vision." Retrieved 6.1.2015, from <http://www.inps4.co.uk/vision/vision-gp-practices>.
- ISD. (2014). "National Data Catalogue." from <http://www.ndc.scot.nhs.uk/National-Datasets/data.asp?SubID=2>.
- Jaen, A., A. Ferrer, et al. (1999). "[Prevalence of chronic bronchitis, asthma and airflow limitation in an urban-industrial area of Catalonia]." *Arch Bronconeumol* **35**(3): 122-128.
- John, G., K. Kohse, et al. (2014). "The composition of cigarette smoke determines inflammatory cell recruitment to the lung in COPD mouse models." *Clin Sci (Lond)* **126**(3): 207-221.
- Jones, R. C., M. Dickson-Spillmann, et al. (2008). "Accuracy of diagnostic registers and management of chronic obstructive pulmonary disease: the Devon primary care audit." *Respir Res* **9**: 62.
- Jones, R. C., G. C. Donaldson, et al. (2009). "Derivation and validation of a composite index of severity in chronic obstructive pulmonary disease: the DOSE Index." *American Journal of Respiratory & Critical Care Medicine* **180**(12): 1189-1195.
- Jurafsky, D. and J. H. Martin (2009). *Speech and language processing : an introduction to natural language processing, computational linguistics, and speech recognition*
- Kalra, D., B. Fernando, et al. (2012). "A review of the empirical evidence of the value of structuring and coding of clinical information within electronic health records for direct patient care." *Inform Prim Care* **20**(3): 171-180.
- Kauppi, P., H. Kupiainen, et al. (2011). "Overlap syndrome of asthma and COPD predicts low quality of life." *J Asthma* **48**(3): 279-285.
- Kennon, B., G. P. Leese, et al. (2012). "Reduced incidence of lower-extremity amputations in people with diabetes in Scotland: a nationwide study." *Diabetes Care* **35**(12): 2588-2590.

- Kesten, S., B. Celli, et al. (2011). "Adverse health consequences in COPD patients with rapid decline in FEV1--evidence from the UPLIFT trial." Respir Res **12**: 129.
- Kew, K. M., S. Dias, et al. (2014). "Long-acting inhaled therapy (beta-agonists, anticholinergics and steroids) for COPD: a network meta-analysis." Cochrane Database Syst Rev **3**: CD010844.
- Kew, K. M. and A. Seniukovich (2014). "Inhaled steroids and risk of pneumonia for chronic obstructive pulmonary disease." Cochrane Database Syst Rev **3**: CD010115.
- Klarman, H. E., J. O. Francis, et al. (1968). "Cost effectiveness analysis applied to the treatment of chronic renal disease." Medical Care **6**(1): 48-54.
- Kohane, I. S., S. E. Churchill, et al. (2012). "A translational engine at the national scale: informatics for integrating biology and the bedside." J Am Med Inform Assoc **19**(2): 181-185.
- Kohansal, R., P. Martinez-Camblor, et al. (2009). "The Natural History of Chronic Airflow Obstruction Revisited." American Journal of Respiratory and Critical Care Medicine **180**(1): 3-10.
- Kotz, D., C. R. Simpson, et al. (2014). "Development and validation of a model to predict the 10-year risk of general practitioner-recorded COPD." NPJ Prim Care Respir Med **24**: 14011.
- Kulik, M. C., W. J. Nusselder, et al. (2012). "Comparison of tobacco control scenarios: quantifying estimates of long-term health impact using the DYNAMO-HIA modeling tool." PLoS One **7**(2): e32363.
- Lehman, R. (2004). "How long can I go on like this? Dying from cardiorespiratory disease." Br J Gen Pract **54**(509): 892-893.
- Lemyze, M. and F. Bart (2011). "Hoover sign." CMAJ Canadian Medical Association Journal **183**(2): E133.
- Lewis, G. (2014). "Better data means better Care." from <http://www.england.nhs.uk/2014/01/15/geraint-lewis/>.
- Lhachimi, S., W. J. Nusselder, et al. (2012). "DYNAMO-HIA--a Dynamic Modeling tool for generic Health Impact Assessments." PLoS One **7**(5): e33317. doi: 33310.31371/journal.pone.0033317. Epub 0032012 May 0033310.
- Lopez, A. D., N. E. Collishaw, et al. (1994). "A descriptive model of the cigarette epidemic in developed countries." Tobacco Control **3**(3): 242-247.
- Lopez, A. D., K. Shibuya, et al. (2006). "Chronic obstructive pulmonary disease: Current burden and future projections." European Respiratory Journal **27**: 397-412.
- Lozano, R., M. Naghavi, et al. (2012). "Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010." Lancet. **380**(9859): 2095-2128. doi: 2010.1016/S0140-6736(2012)61728-61720.
- Lundback, B., A. Lindberg, et al. (2003). "Not 15 but 50% of smokers develop COPD?-- Report from the Obstructive Lung Disease in Northern Sweden Studies." Respir Med. **97**(2): 115-122.
- Lundback, B., A. Lindberg, et al. (2003). "Not 15 but 50% of smokers develop COPD?-- Report from the Obstructive Lung Disease in Northern Sweden Studies." Respir Med **97**(2): 115-122.
- Lundback, B., L. Nystrom, et al. (1991). "Obstructive lung disease in northern Sweden: respiratory symptoms assessed in a postal survey." Eur Respir J **4**(3): 257-266.
- MacNee, W. (2009). "Accelerated lung aging: a novel pathogenic mechanism of chronic obstructive pulmonary disease (COPD)." Biochem Soc Trans **37**(Pt 4): 819-823.
- MacNee, W. and R. M. Tudor (2009). "New paradigms in the pathogenesis of chronic obstructive pulmonary disease I." Proc Am Thorac Soc **6**(6): 527-531.
- Malek, M. (2001). Implementing QALYs, Bandolier.
- Maltais, F., M. Decramer, et al. (2014). "An official American Thoracic Society/European Respiratory Society statement: update on limb muscle dysfunction in chronic

- obstructive pulmonary disease." American Journal of Respiratory & Critical Care Medicine **189**(9): e15-62.
- Mannino, D. M. and A. S. Buist (2007). "Global burden of COPD: risk factors, prevalence, and future trends." Lancet **370**(9589): 765-773.
- Mannino, D. M., R. A. Etzel, et al. (1993). "Do the medical history and physical examination predict low lung function?" Archives of Internal Medicine **153**(16): 1892-1897.
- Mannino, D. M., A. Sonia Buist, et al. (2007). "Chronic obstructive pulmonary disease in the older adult: what defines abnormal lung function?" Thorax **62**(3): 237-241.
- Marco Jordan, L., J. C. Martin Berra, et al. (1998). "[Chronic obstructive lung disease in the general population. An epidemiologic study performed in Guipuzcoa]." Arch Bronconeumol **34**(1): 23-27.
- Mathers, C. D. and D. Loncar (2006). "Projections of global mortality and burden of disease from 2002 to 2030." PLoS Med **3**(11): e442.
- Matthew, J. (2007). Managing long-term conditions, Audit Scotland.
- Matza, L. S., K. S. Boye, et al. (2014). "Impact of caregiver and parenting status on time trade-off and standard gamble utility scores for health state descriptions." Health Qual Life Outcomes **12**(1): 48.
- McCabe, C., K. Claxton, et al. (2008). "The NICE cost-effectiveness threshold: what it is and what that means." Pharmacoeconomics **26**(9): 733-744.
- McDonald, V. M., I. Higgins, et al. (2013). "Managing older patients with coexistent asthma and chronic obstructive pulmonary disease: diagnostic and therapeutic challenges." Drugs Aging **30**(1): 1-17.
- McGraw-Hill. (2003). "McGraw-Hill Dictionary of Scientific & Technical Terms,." 6th. Retrieved 23rd March 2015, from <http://encyclopedia2.thefreedictionary.com/difference+equation>.
- McLean, S., U. Nurmatov, et al. (2011). "Telehealthcare for chronic obstructive pulmonary disease." Cochrane Database Syst Rev(7): CD007718.
- McLean, S., U. Nurmatov, et al. (2012). "Telehealthcare for chronic obstructive pulmonary disease: Cochrane Review and meta-analysis." Br J Gen Pract **62**(604): e739-749.
- McLean, S., S. H. Wild, et al. (2013). "Models for estimating projections for the prevalence and disease burden of chronic obstructive pulmonary disease (COPD): systematic review protocol." Prim Care Respir J **22**(2): PS8-21.
- Menezes, A. M., M. V. Lopez, et al. (2009). "Prevalence of smoking and incidence of initiation in the Latin American adult population: the PLATINO study." BMC Public Health **9**: 151.
- Menezes, A. M., R. Perez-Padilla, et al. (2008). "Worldwide burden of COPD in high- and low-income countries. Part II. Burden of chronic obstructive lung disease in Latin America: the PLATINO study." Int J Tuberc Lung Dis **12**(7): 709-712.
- Menon, B., M. Gurnani, et al. (2008). "Comparison of outpatient visits and hospitalisations, in patients with chronic obstructive pulmonary disease, before and after influenza vaccination." Int J Clin Pract **62**(4): 593-598.
- Merriam-Webster. (2013). "Dictionary and Thesaurus." from <http://www.merriam-webster.com/>.
- Merriam and Webster. (2014). from <http://www.merriam-webster.com/dictionary/data%20mining>.
- MICROS. (2015). "Torex Synergy." from <http://www.micros-retail.co.uk/>.
- Miravittles, M., J. B. Soriano, et al. (2009). "Prevalence of COPD in Spain: impact of undiagnosed COPD on quality of life and daily life activities." Thorax **64**(10): 863-868.
- Mohamed Hoesein, F. A. A., P. Zanen, et al. "Lower limit of normal or FEV1/FVC < 0.70 in diagnosing COPD: an evidence-based review." Respiratory Medicine **105**(6): 907-915.

- Montnemery, P., E. Adelroth, et al. (1998). "Prevalence of obstructive lung diseases and respiratory symptoms in southern Sweden." *Respiratory Medicine* **92**(12): 1337-1345.
- Montnemery, R., U. Nihlen, et al. (2006). "Obstructive airways diseases, smoking and use of inhaled corticosteroids in southern Sweden in 1992 and 2000." *Int J Tuberc Lung Dis* **10**(5): 490-498.
- Morrison, Z., B. Fernando, et al. (2012). An evaluation of different levels of structuring withing the clinical record, NHS connecting for health evaluation programme.
- Morrison, Z., B. Fernando, et al. (2013). "National evaluation of the benefits and risks of greater structuring and coding of the electronic health record: exploratory qualitative investigation." *J Am Med Inform Assoc*.
- Munoz-Arroyo, R. (2011). The Scottish Health Survey.
- Murray, C. J. and A. D. Lopez (1997). "Alternative projections of mortality and disability by cause 1990-2020: Global Burden of Disease Study." *Lancet* **349**(9064): 1498-1504.
- Murray, C. J., M. A. Richards, et al. (2013). "UK health performance: findings of the Global Burden of Disease Study 2010." *Lancet* **381**(9871): 997-1020.
- Murray, C. J., T. Vos, et al. (2012). "Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010." *Lancet*. **380**(9859): 2197-2223. doi: 2110.1016/S0140-6736(2112)61689-61684.
- Murray, C. J., T. Vos, et al. (2012). "Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010." *Lancet* **380**(9859): 2197-2223.
- Murray, C. J. L. (1994). "Quantifying the burden of disease; the technical basis for disability adjusted life years." *Bulletin of the World Health Organization* **72**(3): 429-445.
- Murray, C. J. L. and A. D. Lopez (1994). "Quantifying disability data, methods and results." *Bulletin of the World Health Organization* **72**(1994): 481-494.
- Murray, S. A., M. Kendall, et al. (2005). "Illness trajectories and palliative care." *Bmj* **330**(7498): 1007-1011.
- Murtagh, E., L. Heaney, et al. (2005). "Prevalence of obstructive lung disease in a general population sample: the NICECOPD study." *Eur J Epidemiol* **20**(5): 443-453.
- Nacul, L. C., M. Soljak, et al. (2007). "Model for estimating the population prevalence of chronic obstructive pulmonary disease: cross sectional data from the Health Survey for England." *Popul Health Metr* **5**: 8.
- Najafzadeh, M., C. A. Marra, et al. (2012). "Future impact of various interventions on the burden of COPD in Canada: a dynamic population model." *PLoS One* **7**(10): e46746.
- NHS_England. (2014). "Care.Data." Retrieved November 2014, from <http://www.england.nhs.uk/ourwork/tsd/care-data/>.
- NICE (2010). Chronic obstructive pulmonary disease: management of chronic obstructive pulmonary disease in adults in primary and secondary care. . London, National Clinical Guideline Centre.
- NICE (2012). Roflumilast for the management of severe chronic obstructive pulmonary disease. *NICE Technology Appraisal Guidance 244*, National Institute for Clinical Excellence: 9.
- Niewoehner, D. E., K. Rice, et al. (2005). "Prevention of exacerbations of chronic obstructive pulmonary disease with tiotropium, a once-daily inhaled anticholinergic bronchodilator: a randomized trial." *Annals of Internal Medicine* **143**(5): 317-326.
- Norman, R., B. Barnes, et al. (2007). "Estimating the burden of disease attributable to indoor air pollution from household use of solid fuels in South Africa in 2000." *S Afr Med J* **97**(8 Pt 2): 764-771.
- OECD. (2013). "Health Update June 2013 No 17." Retrieved 30th July, 2013, from <http://www.oecd.org/els/health-systems/issueno17june2013.htm>.

- Olsen, J. A. (2009). Principles in Health Economics and Policy, Oxford.
- ONS, O. o. N. S. (2011). English Population by single year of age and sex.
- ONS, O. o. N. S. (2011). General Lifestyle Survey, Smoking prevalence chapter 1.
- ONS, O. o. N. S. (2011). "NHS expenditure 2011." Retrieved 8th Sept 2014, from <http://www.ons.gov.uk/ons/rel/psa/expenditure-on-healthcare-in-the-uk/2011/art-expenditure-on-healthcare-in-the-uk-2011.html#tab-Total-expenditure-on-healthcare-in-the-UK--1997-2011>.
- ONS, O. o. N. S. (2012). "Population projections." 2014, from <http://www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%3A77-318453>.
- OPSI (1998). The Data Protection Act, Office of Public Sector Information.
- Park, H. Y., A. Churg, et al. (2013). "Club cell protein 16 and disease progression in chronic obstructive pulmonary disease." American Journal of Respiratory & Critical Care Medicine **188**(12): 1413-1419.
- Peabody, J. W., B. Schau, et al. (2005). "COPD: A prevalence estimation model." Respirology **10**: 594-602.
- Pena, V. S., M. Miravittles, et al. (2000). "Geographic variations in prevalence and underdiagnosis of COPD: results of the IBERPOC multicentre epidemiological study." Chest **118**(4): 981-989.
- Perera, P. N., E. P. Armstrong, et al. (2012). "Acute exacerbations of COPD in the United States: inpatient burden and predictors of costs and mortality." COPD **9**(2): 131-141.
- Peto, R. and C. Fletcher (1977). "The natural history of chronic airflow obstruction." Br Med J **1**(6077): 1645-1648.
- Peto, R., F. E. Speizer, et al. (1983). "The relevance in adults of air-flow obstruction, but not of mucus hypersecretion, to mortality from chronic lung disease. Results from 20 years of prospective observation." Am Rev Respir Dis **128**(3): 491-500.
- Pham, T. M., K. Ozasa, et al. (2012). "Age-period-cohort analysis of chronic obstructive pulmonary disease mortality in Japan, 1950-2004." J Epidemiol **22**(4): 302-307.
- Philips, Z., L. Bojke, et al. (2006). "Good practice guidelines for decision-analytic modelling in health technology assessment: a review and consolidation of quality assessment." Pharmacoeconomics **24**(4): 355-371.
- Pichon-Riviere, A., F. Augustovski, et al. (2011). "Development and validation of a microsimulation economic model to evaluate the disease burden associated with smoking and the cost-effectiveness of tobacco control interventions in Latin America." Value in Health **14**(5 SUPPL.): S51-S59.
- Pillai, S. G., D. Ge, et al. (2009). "A Genome-Wide Association Study in Chronic Obstructive Pulmonary Disease (COPD): Identification of Two Major Susceptibility Loci." PLoS Genet **5**(3).
- Pinnock, H., J. Hanley, et al. (2013). "Effectiveness of telemonitoring integrated into existing clinical services on hospital admission for exacerbation of chronic obstructive pulmonary disease: researcher blind, multicentre, randomised controlled trial." Bmj **347**: f6070.
- Pinnock, H., M. Kendall, et al. (2011). "Living and dying with severe chronic obstructive pulmonary disease: multi-perspective longitudinal qualitative study." Bmj **342**: d142.
- Porta, M. (2014). A dictionary of epidemiology. Oxford, Oxford University Press.
- Price, D., Y. Asukai, et al. (2013). "A UK-based costs utility analysis of Indacaterol, A Once Daily Maintenance Bronchodilator for Patients with COPD, using real world evidence on resource use." Applied Health Economics Health Policy **11**: 259-274.
- Price, D. B., D. G. Tinkelman, et al. (2006). "Symptom-based questionnaire for identifying chronic obstructive pulmonary disease in smokers." Respiration **73**: 285-295.

- Price, D. B., B. P. Yawn, et al. (2010). "Improving the differential diagnosis of chronic obstructive pulmonary disease in primary care." Mayo Clinic Proceedings **85**(12): 1122-1129.
- Pride, N. B. (2001). "Smoking cessation: effects on symptoms, spirometry and future trends in COPD." Thorax **56** Suppl 2: ii7-10.
- Pride, N. B. (2006). "Projections of COPD in males in The Netherlands." Eur Respir J **27**(1): 240-241; author reply 241-242.
- Prokosch, H. U. and T. Ganslandt (2009). "Perspectives for medical informatics. Reusing the electronic medical record for clinical research." Methods Inf Med **48**(1): 38-44.
- PROSPERO. (2013). "Register of Systematic Reviews." from <http://www.crd.york.ac.uk/prospero/>.
- Puhan, M., M. Scharplatz, et al. (2009). "Pulmonary rehabilitation following exacerbations of chronic obstructive pulmonary disease." Cochrane Database Syst Rev(1): CD005305.
- Puhan, M. A., J. Garcia-Aymerich, et al. (2009). "Expansion of the prognostic assessment of patients with chronic obstructive pulmonary disease: the updated BODE index and the ADO index." Lancet **374**(9691): 704-711.
- Puhan, M. A., E. Gimeno-Santos, et al. (2011). "Pulmonary rehabilitation following exacerbations of chronic obstructive pulmonary disease." Cochrane Database Syst Rev(10): CD005305.
- Qaseem, A., T. J. Wilt, et al. (2011). "Diagnosis and management of stable chronic obstructive pulmonary disease: a clinical practice guideline update from the American College of Physicians, American College of Chest Physicians, American Thoracic Society, and European Respiratory Society." Annals of Internal Medicine **155**(3): 179-191.
- Quanjer PH, Tammeling GJ, et al. (1993). "Lung volumes and forced ventilatory flows. Report Working Party Standardization of Lung Function Tests, European Community for Steel and Coal. Official Statement of the European Respiratory Society." European Respiratory Journal Supplement **16**: 5-40.
- Raftery, J., P. Roderick, et al. (2005). "Potential use of routine databases in health technology assessment." Health Technology Assessment (Winchester, England) **9**(20): 1-92, iii-iv.
- Renwick, D. S. and M. J. Connolly (1996). "Prevalence and treatment of chronic airways obstruction in adults over the age of 45." Thorax **51**(2): 164-168.
- Roberts, S. D., M. O. Farber, et al. (2006). "FEV1/FVC ratio of 70% misclassifies patients with obstruction at the extremes of age." Chest **130**(1): 200-206.
- Robertson, A., D. W. Bates, et al. (2011). "The rise and fall of England's National Programme for IT." J R Soc Med **104**(11): 434-435.
- Robertson, A., K. Cresswell, et al. (2010). "Implementation and adoption of nationwide electronic health records in secondary care in England: qualitative analysis of interim results from a prospective national evaluation." Bmj **341**: c4564.
- Roche, N., F. Dalmay, et al. (2008). "Impact of chronic airflow obstruction in a working population." Eur Respir J **31**(6): 1227-1233.
- Rona, R. J., M. C. Gulliford, et al. (1993). "Effects of prematurity and intrauterine growth on respiratory health and lung function in childhood." Bmj **306**(6881): 817-820.
- Salpeter, S. R., N. S. Buckley, et al. (2006). "Meta-analysis: anticholinergics, but not beta-agonists, reduce severe exacerbations and respiratory mortality in COPD." J Gen Intern Med **21**(10): 1011-1019.
- Scanlon, P. D., J. E. Connett, et al. (2000). "Smoking cessation and lung function in mild-to-moderate chronic obstructive pulmonary disease. The Lung Health Study." American Journal of Respiratory & Critical Care Medicine **161**(2 Pt 1): 381-390.
- Schirnhofner, L., B. Lamprecht, et al. (2007). "COPD prevalence in Salzburg, Austria: results from the Burden of Obstructive Lung Disease (BOLD) Study." Chest **131**(1): 29-36.

- Science_and_Technology_Committee (2009). House of Lords Science and Technology Committee, 2nd Report of Session 2008-9: Genomic Medicine. London: para 6.15.
- ScottishGovernment (2012). Long-term Monitoring of Health Inequalities: Headline Indicators.
- Sethi, N. and G. T. Laurie (2013). "Delivering proportionate governance in the era of eHealth: Making linkage and privacy work together." *Med Law Int* **13**(2-3): 168-204.
- Shahab, L., M. J. Jarvis, et al. (2006). "Prevalence, diagnosis and relation to tobacco dependence of chronic obstructive pulmonary disease in a nationally representative population sample." *Thorax* **61**(12): 1043-1047.
- Sheikh, A., T. Cornford, et al. (2011). "Implementation and adoption of nationwide electronic health records in secondary care in England: final qualitative results from prospective national evaluation in "early adopter" hospitals." *Bmj* **343**: d6054.
- Sheikh, A., A. Jha, et al. (2014). "Adoption of electronic health records in UK hospitals: lessons from the USA." *Lancet* **384**(9937): 8-9.
- Shibuya, K., C. D. Mathers, et al. (2001). Chronic Obstructive Pulmonary Disease, Consistent Estimates of Incidence, Prevalence and Mortality by WHO Region., World Health Organization.
- Shibuya, K., C. D. Mathers, et al. (2001). Chronic Obstructive Pulmonary Disease (COPD): Consistent Estimates of Incidence Prevalence and Mortality by WHO Region. *Global Programme on Evidence for Health Policy*, World Health Organization.
- Sichletidis, L., I. Tsiotsios, et al. (2005). "Prevalence of chronic obstructive pulmonary disease and rhinitis in northern Greece." *Respiration* **72**(3): 270-277.
- Siedlinski, M., D. Tingley, et al. (2013). "Dissecting direct and indirect genetic effects on chronic obstructive pulmonary disease (COPD) susceptibility." *Hum Genet.* **132**(4): 431-441. doi: 410.1007/s00439-00012-01262-00433. Epub 02013 Jan 00439.
- simulistics. (2011). "spatial modelling." Retrieved 23rd March, 2015, from <http://www.simulistics.com/tour/spatialmodels.htm>.
- Sinha, S., P. K. Myint, et al. (2008). "Accuracy of death certification and hospital record linkage for identification of incident stroke." *BMC medical research methodology* **8**: 74.
- Soo, M., L. M. Robertson, et al. (2014). "Approaches to ascertaining comorbidity information: validation of routine hospital episode data with clinician-based case note review." *BMC Res Notes* **7**(1): 253.
- Soriano, J. B., K. J. Davis, et al. (2003). "The proportional Venn diagram of obstructive lung disease: two approximations from the United States and the United Kingdom." *Chest* **124**(2): 474-481.
- Sorlie, P., E. Lakatos, et al. (1987). "Influence of cigarette smoking on lung function at baseline and at follow-up in 14 years: the Framingham Study." *J Chronic Dis* **40**(9): 849-856.
- SPHO-PTI. (2012). "Practice Team Information Programme." Retrieved 20th September, 2012, from <http://www.scotpho.org.uk/health-wellbeing-and-disease/chronic-obstructive-pulmonary-disease-copd/data/primary-care-data>.
- SPHO. (2012). "Scottish Public Health Observatory." Retrieved 20 September, 2012, from <http://www.isdscotland.org/Health-Topics/General-Practice/Quality-And-Outcomes-Framework/2010-11/Register-and-prevalence-data.asp>.
- Stallberg, B., C. Janson, et al. (2013). "Management, morbidity and mortality of COPD during an 11-year period: an observational retrospective epidemiological register study in Sweden (PATHOS)." *Prim Care Respir J.*
- Stanciole, A. E., M. Ortegon, et al. (2012). "Cost effectiveness of strategies to combat chronic obstructive pulmonary disease and asthma in sub-Saharan Africa and South East Asia: mathematical modelling study." *Bmj* **344**: e608.

- Stang, P., E. Lydick, et al. (2000). "The prevalence of COPD: using smoking rates to estimate disease frequency in the general population." *Chest* **117**(5 Suppl 2): 354S-359S.
- Stausberg, J., N. Lehmann, et al. (2008). "Reliability of diagnoses coding with ICD-10." *International Journal of Medical Informatics* **77**(1): 50-57.
- Stavem, K., L. Sandvik, et al. (2006). "Can global initiative for Chronic Obstructive Lung Disease stage 0 provide prognostic information on long-term mortality in men?" *Chest* **130**(2): 318-325.
- Sterne, J. A., I. R. White, et al. (2009). "Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls." *Bmj* **338**: b2393.
- STIVORO (2003). *Roken, de harde feiten: Volwassenen*. Den Haag.
- Strachan, D. P., V. Seagroatt, et al. (1994). "Chest illness in infancy and chronic respiratory disease in later life: an analysis by month of birth." *Int J Epidemiol* **23**(5): 1060-1068.
- Strachan, D. P. and A. Sheikh (2004). *A life course approach to respiratory and allergic diseases. A life course approach to chronic disease epidemiology*. D. Kuh and Y. Ben-Shlomo. Oxford, Oxford University Press: 240-259.
- SurgeonGeneral (2004). *The health consequences of smoking: a report of the Surgeon General Atlanta, US Department of Health and Human Services*
Centers for Disease Control and Prevention
National Center for Chronic Diseases Prevention and Health Promotion
Office on Smoking and Health.
- Thiru, K., S. De Lusignan, et al. (2002). "Paperless practices: A report from a research network." *Informatics in Primary Care* **10**: 89-94.
- Thuesen, L., L. O. Jensen, et al. (2013). "Event detection using population-based health care databases in randomized clinical trials: a novel research tool in interventional cardiology." *Clin Epidemiol* **5**: 357-361.
- Thun, M., R. Peto, et al. (2012). "Stages of the cigarette epidemic on entering its second century." *Tob Control* **21**(2): 96-101.
- Tinkelman, D. G., D. Price, et al. (2007). "COPD screening efforts in primary care: what is the yield?" *Prim Care Respir J* **16**(1): 41-48.
- Toren, K. and B. Jarvholm (2014). "Effect of occupational exposure to vapors, gases, dusts, and fumes on COPD mortality risk among Swedish construction workers: a longitudinal cohort study." *Chest* **145**(5): 992-997.
- TPP-Systmone. (2015). Retrieved 6.1.2015, from <http://www.tpp-uk.com/>.
- Treweek, S., P. Lockhart, et al. (2013). "Methods to improve recruitment to randomised controlled trials: Cochrane systematic review and meta-analysis." *BMJ Open* **3**(2).
- Tzanakis, N., U. Anagnostopoulou, et al. (2004). "Prevalence of COPD in Greece." *Chest* **125**(3): 892-900.
- Unal, B., S. Capewell, et al. (2006). "Coronary heart disease policy models: a systematic review." *BMC Public Health* **6**: 213.
- van der Meer, R. M., E. J. Wagena, et al. (2003). "Smoking cessation for chronic obstructive pulmonary disease." *Cochrane Database Syst Rev*(2): CD002999.
- van der Molen, T., M. Miravitlles, et al. (2013). "COPD management: role of symptom assessment in routine clinical practice." *International Journal of Copd* **8**: 461-471.
- van Durme, Y. M., K. M. Verhamme, et al. (2009). "Prevalence, incidence, and lifetime risk for the development of COPD in the elderly: the Rotterdam study." *Chest* **135**(2): 368-377.
- van Genugten, M. L., R. T. Hoogenveen, et al. (2003). "Future burden and costs of smoking-related disease in the Netherlands: a dynamic modeling approach." *Value Health* **6**(4): 494-499.

- Van Oers, J. A. (2002). *Gezondheid op koers? Volksgezondheid Toekomst Verkenning*. Bilthoven, The Netherlands, National Institute of Public Health and the Environment (RIVM).
- van Schayck, C. P., R. J. Halbert, et al. (2005). "Comparison of existing symptom-based questionnaires for identifying COPD in the general practice setting." *Respirology* **10**(3): 323-333.
- Van Weel, C., H. Smith, et al. (2000). "Family practice research networks. Experiences from 3 countries." *Journal of Family Practice* **49**(938-943).
- Vestbo, J., A. Agusti, et al. (2014). "Should we view COPD Differently after ECLIPSE? A Clinical Perspective from the Study Team." *American Journal of Respiratory & Critical Care Medicine*.
- Vestbo, J., L. D. Edwards, et al. (2011). "Changes in forced expiratory volume in 1 second over time in COPD." *N Engl J Med* **365**(13): 1184-1192.
- Viegi, G., M. Pedreschi, et al. (2000). "Prevalence of airways obstruction in a general population: European Respiratory Society vs American Thoracic Society definition." *Chest* **117**(5 Suppl 2): 339S-345S.
- Vollenweider, D. J., H. Jarrett, et al. (2012). "Antibiotics for exacerbations of chronic obstructive pulmonary disease." *Cochrane Database Syst Rev* **12**: CD010257.
- Wang, Y., K. Hunt, et al. (2013). "Do men consult less than women? An analysis of routinely collected UK general practice data." *BMJ Open* **3**(8): e003320.
- Ware, J. E., Jr. and C. D. Sherbourne (1992). "The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection." *Med Care* **30**(6): 473-483.
- Wedzicha, J. and G. C. Donaldson (2012). "Natural history of successive COPD exacerbations." *Thorax* **2**: 2.
- Weinstein, M. C. (1989). "Methodologic issues in policy modeling for cardiovascular disease." *J Am Coll Cardiol* **14**(3 Suppl A): 38A-43A.
- Weinstein, M. C., B. O'Brien, et al. (2003). "Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices--Modeling Studies." *Value Health* **6**(1): 9-17.
- Weinstein, M. C., E. L. Toy, et al. (2001). "Modeling for health care and other policy decisions: uses, roles, and validity." *Value Health* **4**(5): 348-361.
- Welsh, E. J., C. J. Cates, et al. (2013). "Combination inhaled steroid and long-acting beta2-agonist versus tiotropium for chronic obstructive pulmonary disease." *Cochrane Database Syst Rev* **5**: CD007891.
- West, R. (2012). "Impact of the 'smoking ban' on smoking prevalence in England." Retrieved 9th Sept 2014, from <http://www.google.co.uk/url?sa=t&rct=j&q=&esrc=s&frm=1&source=web&cd=4&ved=0CDgQFjAD&url=http%3A%2F%2Fwww.smokinginengland.info%2Fdownloadfile%2F%3Ftype%3Dsts-documents%26src%3D12&ei=YBwPVJ64NunA7Ab8qoDoDg&usg=AFQjCNFdjDLvyW-2jZIKYTqnq8saHSNmzA>.
- West, R. and J. Brown. (2013). "Smoking in England." *The Smoking Toolkit Study*, from <http://www.smokinginengland.info/>.
- Westwood, M., J. Bourbeau, et al. (2011). "Relationship between FEV1 change and patient-reported outcomes in randomised trials of inhaled bronchodilators for stable COPD: a systematic review." *Respir Res* **12**: 40.
- WHO. (2010). "International Classification of Diseases - 10: ICD-10." from <http://apps.who.int/classifications/icd10/browse/2010/en>.
- Wijnhoven, H. A., D. M. Kriegsman, et al. (2001). "Determinants of difference dimensions of disease severity in asthma and COPD: pulmonary function and health related quality of life." *Chest* **119**(1034-1042).
- Williams, J. G., W. Y. Cheung, et al. (2003). "Can randomised trials rely on existing electronic data? A feasibility study to explore the value of routine data in health

- technology assessment." Health Technology Assessment (Winchester, England) **7**(26): iii, v-x, 1-117.
- Williams, S., N. Baxter, et al. (2012). IMPRESS Improving and Integrating Respiratory Services.
- Williams, W. G. and J. M. Morgan (1995). "The clinician-information interface." Medinfo **8 Pt 1**: 801-805.
- Wilt, T. J., D. Niewoehner, et al. (2007). "Management of stable chronic obstructive pulmonary disease: a systematic review for a clinical practice guideline." Annals of Internal Medicine **147**(9): 639-653.
- Wise, L., J. Parkinson, et al. (2009). "New approaches to drug safety: a pharmacovigilance tool kit." Nat Rev Drug Discov **8**(10): 779-782.
- Wolfram. (2013). "Mathematica Version 9." from <http://www.wolfram.com/mathematica/>.
- Wood, A. M. (2011). "Re-evaluating COPD risk: just a matter of definition." American Journal of Respiratory & Critical Care Medicine **183**(7): 837-838.
- Wyatt, J. and D. Spiegelhalter (1991). "Field trials of medical decision-aids: potential problems and solutions." Proc Annu Symp Comput Appl Med Care: 3-7.
- Zatonski, W., K. Przewozniak, et al. (2012). "Tobacco smoking in countries of the European Union." Annals of Agricultural and Environmental Medicine **19**(2): 181-192.
- Zeckhauser, R. and D. Shepard (1976). "Where now for saving lives." Law and contemporary problems **40**(4): 5-45.

PROTOCOL SUMMARY

Models for estimating projections for the prevalence and disease burden of chronic obstructive pulmonary disease (COPD): systematic review protocol

*Susannah McLean¹, Sarah H Wild², Colin R Simpson¹, Aziz Sheikh¹

¹ Allergy and Respiratory Research Group, Centre for Population Health Sciences, The University of Edinburgh, Edinburgh, UK

² Centre for Population Health Sciences, The University of Edinburgh, Edinburgh, UK

Received 20th March 2013; revised 16th April 2013; accepted 24th April 2013; online 30th May 2013

© 2013 Primary Care Respiratory Society UK. All rights reserved.

S McLean *et al.* *Prim Care Respir J* 2013; 22(2): PS8-PS21

<http://dx.doi.org/10.4104/pcrj.2013.00048>

Keywords COPD, epidemiology, modelling

Introduction

Policymakers and governments must decide their healthcare priorities on the basis of the best healthcare intelligence available to them. Recent interest has increasingly focused on the global implications of an increasing and elderly population with long-term conditions.¹⁻³ The most recent figures from the Global Burden of Disease Study 2010 show that the third top global cause of death was chronic obstructive pulmonary disease (COPD),⁴ rising from fourth place in 1990.⁵ It is predominantly caused by cigarette smoking and leads to lung airflow limitation, cough, excessive sputum production, and breathlessness. People with COPD can suffer from substantial disability as the condition progresses.⁶ A pressing challenge for governments is how best to project the future trend in the prevalence and burden of COPD in order to plan adequate health and social care for those affected by this condition within the scope of limited resources. Governments should ideally be planning for COPD on two levels: (1) they should consider how to manage resources to care and treat people who are already affected by COPD; and (2) how to prevent a greater increase in the burden from COPD by minimising the continuing smoking epidemic.

In order to make such calculations, governments and other healthcare providers need to draw on epidemiological models. Merriam-Webster's dictionary defines a 'model' as 'a system of postulates, data, and inferences presented as a mathematical description of an entity or state of affairs'. This is a useful starting point when considering the role of models in epidemiology. Most models are explanatory in nature and describe the relationships between different parameters. The focus of this study is on models which help to project future epidemiological trends and patterns in

populations with COPD. Governments and policymakers have access to many models, but a review is required to appraise the published COPD models to aid selection between them.

Various features of COPD present a particular challenge to mathematical and epidemiological modelling, including the many different definitions of a COPD diagnosis and its overlap with a diagnosis of asthma. Although COPD is most clearly attributable to cigarette smoking, there is debate over how best to classify non-smokers who develop COPD with the immunological and pathological features of COPD as a result of exposure to occupational dusts and gases or recurrent chest infections. In addition, there is uncertainty as to the correct classification of older non-smoking adults who have evidence of lung cell remodelling including squamous metaplasia following chronic inflammation due to long-term asthma. Such older adults have often lost the reversibility in their airways obstruction and demonstrate spirometry which is consistent with the thresholds for COPD.⁷⁻⁹

According to the Global Initiative for Chronic Obstructive Lung Disease (GOLD), the diagnosis of COPD is characterised by an obstructive lung defect with forced expiratory volume in one second to forced vital capacity (FEV₁/FVC) ratio <0.7.¹⁰ Controversy regarding this threshold also complicates decisions of precisely which population to include in modelling. Lung function decreases with age, so a proportion of elderly people (age 75+) who have never smoked still fit these criteria for COPD. Some doctors reasonably argue that such elderly people really have normal lung function for their age and that medicalisation of the elderly should be avoided.⁷ An alternative threshold of the lower limit of normal for FEV₁/FVC has been proposed with a decreasing threshold according

* **Corresponding author:** Dr Susannah McLean, Allergy and Respiratory Research Group, Centre for Population Health Sciences, The University of Edinburgh, Edinburgh EH8 9DX, UK. Tel: +44 (0)131 650 9242 E-mail: Susannah.mclean@ed.ac.uk

to age by percentile. The bottom 5% of FEV₁/FVC measurements for whichever total population being measured would be considered abnormal in the older age group.⁹ However, no up-to-date large standardised population database currently exists to validate such a measure. The nearest is the use of the European Coal and Steel Workers Population to provide percent predicted FEV₁ values; however, this population was standardised over 20 years ago and is based on a working white European population without ethnic minorities.¹¹⁻¹³ Similarly, younger people (age 30–40 years) with larger FVC values and greater respiratory reserve may already have sustained COPD-type damage to their lungs before they reach the <0.7 ratio threshold, so at this end of the age range there is a risk of under-diagnosis of COPD.¹³

The debate regarding the diagnosis of COPD is more than just a debate over spirometry thresholds. As many developing countries do not have access to spirometry or even to a reliable power supply, the usefulness of such diagnostic thresholds is limited. It has been proposed that COPD may also be diagnosed on history and clinical features. However, studies have shown that using clinical indicators of pulmonary function to diagnose COPD missed many participants who had low lung function and airways obstruction, especially in current smokers.¹⁴ Therefore, in many countries the current situation has evolved where COPD is diagnosed from physician opinion without corroborating evidence from spirometry, resulting in a significant overlap between a diagnosis of COPD and a diagnosis of asthma.

It seems likely that classifications in the future will evolve as the role of host susceptibility is increasingly understood in terms of genetic and epigenetic features. Several candidate genes related to COPD have been identified.¹⁵ In addition, the science of epigenetics helps to explain how DNA transcription has been activated or suppressed by DNA methylation, acetylation, or other mechanisms in response to predominantly prenatal and early life environmental influences.¹⁶ The result of such switching on or off of DNA transcription is to determine the host's response to noxious stimuli including cigarette smoke. Increased understanding of these factors is helping to unravel the mysteries of why some life-long smokers are virtually unaffected by their habit while others have severe COPD. Estimates as to the prevalence of COPD among smokers aged >45 years vary from 15% to 50% according to the criteria used for diagnosis.^{17,18}

Modelling COPD is also challenged by the key feature of exacerbations. An exacerbation may be triggered by increased bacterial or viral load in the lungs which induce an aggressive immune response and associated clinical features.¹⁹⁻²¹ Associated with a greater frequency of exacerbations is higher morbidity, due to faster disease progression in terms of loss of lung function, and also mortality.²¹

An additional challenge is the level of mathematical sophistication within each model. Ideally, a researcher with considerable statistical skill would be available to check the algorithms that drive each model and so provide a full appraisal of the quality of each model. In the absence of this ideal, it was decided to appraise the quality of reporting of each model as a proxy for the

model's mathematical quality. Taking these challenges into account, it will be necessary to describe a degree of context with each model in order that it can be applied in an appropriate setting. This will help subsequent researchers to understand the necessary caveats to include when describing the results from each model.

Objectives

To identify all available models for estimating projections of COPD prevalence and burden, and to assess the quality of reporting of each model in its key publication.

Methods

A search strategy has been developed using search terms to cover the three concepts of 'modelling', 'disease burden', and 'chronic obstructive pulmonary disease' (see Appendix 1 for full details). Searches will be conducted in the following electronic databases: MEDLINE, EMBASE, CAB Abstracts, World Health Organization (WHO) Library and Information Services (WHOLIS – library catalogue of books and reports), WHO Regional Indexes (AIM (AFRO), LILACS (AMRO/PAHO), IMEMR (EMRO), IMSEAR (SEARO), WPRIM (WPRO)), and a modified search strategy will be used to identify reports from the WHO home website and Google. Searches will be for both published and unpublished modelling studies from 1980 (when modelling methods first began to be widely used) to 2013. Two authors will independently review the studies against the inclusion criteria and make a decision as to whether the study is suitable. Disagreements will be resolved by discussion and, if this is not possible, a third reviewer will arbitrate.

Inclusion criteria

Any modelling study which uses demographic and epidemiological data to project the prevalence and disease burden will be included. The included projected outcomes which are of interest are one or more of: incidence, prevalence and mortality, and disease burden. With regard to 'disease burden', the outcomes of interest can be considered from the individual's point of view, from the point of view of the healthcare system, and from the point of view of broader society. For the purposes of this review, the focus is on the perspective of the healthcare system. Other perspectives are valid; however, different instruments are used to measure them and the purpose of this study is to guide policymakers who will focus on the healthcare system perspective. Quality-adjusted life years (QALYs) and disability-adjusted life years (DALYs) are often used to measure and quantify the burden to the individual of the morbidity they are suffering. Treatments are assigned a cost per restored QALY, and this is an important measure used in cost-effectiveness studies. However, the scope of this study is more limited in order to avoid confusion of perspectives. Some of the studies included may discuss QALYs and DALYs, but they have not been chosen as primary disease burden outcomes for this review. Instead, we will concentrate on primary care visits, emergency department visits, hospital admissions, and COPD treatment costs.

Exclusion criteria

There will be no exclusions on the basis of language of the report. Studies which are population-based surveys of prevalence without

modelling will be excluded as there has recently been a systematic review of such studies.²² 'Models' will be excluded if they describe animals, cell lines, clinical series, or estimates of individual risk (such as individual prognostic models). Decision analytical models or decision support models will be excluded where they refer to clinical decision-making for individuals rather than populations. Models that compare one intervention with another intervention will also be excluded, as the aim is accurately to project the baseline outcomes so it is premature to take into account the effect of interventions. Also excluded will be regression models which start with a COPD population and 'back-calculate' the prevalence or burden using regression to quantify risk factors, as this follows a different logic from that of projection modelling.

Participants

The source population for the model may be from anywhere in the world. The model will pertain to adult populations aged >40 years as it is usually not appropriate to diagnose COPD in younger people.¹⁰ COPD may be diagnosed by physician, spirometry, or by questionnaire. Other assumptions regarding the diagnosis of COPD will be evaluated in the context of the model.

Data extraction

The data will be extracted by one author and checked by a second. Data will be extracted using a pre-piloted data extraction form. The following identification details will be extracted for each model: author and email address, year, institution, and funding source. These data will be followed by: the purpose of the model, model title, model type, model setting, time period, and population (age, sex and country). Also extracted will be: inputs to the model, source of input data, details of processing of the model, outcomes for COPD (incidence, prevalence, mortality, GP visits, emergency department visits, hospitalisations, treatment costs), model output/results, details of the model's availability, any comparisons with other studies, social and economic policy implications of model outcomes, and future research recommendations. In this way, the data extraction form aims to encompass a comprehensive picture of the model.

Quality appraisal framework

Ideally, a quality appraisal of the actual modelling process would be undertaken. However, this requires significant statistical technical expertise. A pragmatic decision has therefore been made to quality appraise the reporting of the models rather than the actual modelling process for those that have full published reports. In order to do this, a quality of reporting framework has been designed following review of key guidelines as to good practice in modelling.²³⁻²⁶ A scoring mechanism was devised in collaboration with Simon Capewell of Liverpool University²⁷ to weight the importance of the different elements required to produce a relevant high-quality model (see Appendix 2).

Strategy for data synthesis

The study will be the unit of analysis. Models will be described and classified. A detailed critical narrative synthesis of the highest scoring models will be undertaken. Where the models are not available, we will write to the model authors for further clarification. No subgroup analysis is planned.

Handling editor David Bellamy

Conflicts of interest The authors declare that they have no conflicts of interest in relation to this protocol. AS is Joint Editor-in-Chief of the *PCRJ*, but was not involved in the editorial review of, nor the decision to publish, this protocol.

Contributorship SM drafted the article with oversight from CS SW and AS. AS and SM conceived the project as part of SM's PhD.

Funding SM is funded by the University of Edinburgh's Principal's Career Development PhD Scholarship.

Protocol registration A shortened version of this protocol has been registered online in the PROSPERO University of York database: Systematic review of models for estimation and future projecting of the prevalence and the disease burden of chronic obstructive pulmonary disease (COPD), PROSPERO 2012:CRD42012002623, available from http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42012002623

References

1. Lim S, Vos T, Flaxman A, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012;**380**(9859):2224-60. [http://dx.doi.org/10.1016/S0140-6736\(12\)61766-8](http://dx.doi.org/10.1016/S0140-6736(12)61766-8)
2. Murray CJ, Vos T, Lozano R, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012;**380**(9859):2197-223. [http://dx.doi.org/10.1016/S0140-6736\(12\)61689-4](http://dx.doi.org/10.1016/S0140-6736(12)61689-4)
3. Department of Health. Improving the health and wellbeing of people with long term conditions: London: Department of Health, 2010.
4. Lozano R, Naghavi M, Foreman K, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012;**380**(9859):2095-128. [http://dx.doi.org/10.1016/S0140-6736\(12\)61728-0](http://dx.doi.org/10.1016/S0140-6736(12)61728-0)
5. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med* 2006;**3**(11):e442. <http://dx.doi.org/10.1371/journal.pmed.0030442>
6. Pauwels R, Rabe K. Burden and clinical features of chronic obstructive pulmonary disease (COPD). *Lancet* 2004;**364**(9434):613-20. [http://dx.doi.org/10.1016/S0140-6736\(04\)16855-4](http://dx.doi.org/10.1016/S0140-6736(04)16855-4)
7. Mannino DM, Buist SA, Vollmer WM. Chronic obstructive pulmonary disease in the older adult: what defines abnormal lung function? *Thorax* 2007;**62**(3):237-41. <http://dx.doi.org/10.1136/thx.2006.068379>
8. Price D, Freeman D, Cleland J, Kaplan A, Cerasoli F. Earlier diagnosis and earlier treatment of COPD in primary care. *Prim Care Respir J* 2010;**20**(1):15-22. <http://dx.doi.org/10.4104/pcrj.2010.00060>
9. Mohamed Hoessein FAA, Zanen P, Lammers J-WJ. Lower limit of normal or FEV1/FVC <0.70 in diagnosing COPD: an evidence-based review. *Respir Med* 2011;**105**(6):907-15. <http://dx.doi.org/10.1016/j.rmed.2011.01.008>
10. GOLD. Global Initiative for Chronic Obstructive Lung Disease: Pocket guide to COPD diagnosis, management and prevention. 2010. <http://www.goldcopd.org/guidelines-pocket-guide-to-copd-diagnosis.html>
11. Miller MR, Quanjer PH, Swanney MP, Ruppel G, Enright PL. Interpreting lung function data using 80% predicted and fixed thresholds misclassifies more than 20% of patients. *Chest* 2011;**139**(1):52-9. <http://dx.doi.org/10.1378/chest.10-0189>
12. Quanjer PH, Enright PL, Miller MR, et al. The need to change the method for defining mild airway obstruction. *Eur Respir J* 2011;**37**(3):720-2. <http://dx.doi.org/10.1183/09031936.00135110>
13. Swanney MP, Ruppel G, Enright PL, et al. Using the lower limit of normal for the FEV1/FVC ratio reduces the misclassification of airway obstruction. *Thorax* 2008;**63**(12):1046-51. <http://dx.doi.org/10.1136/thx.2008.098483>
14. Mannino D, Etzel RA, Flanders WD. Do the medical history and physical examination predict low lung function? *Arch Intern Med* 1993;**153**(16):1892-7. <http://dx.doi.org/10.1001/archinte.1993.00410160052004>
15. Bakke PS, Zhu G, Gulsvik A, et al. Candidate genes for COPD in two large data sets. *Eur Respir J* 2011;**37**(2):255-63.

- <http://dx.doi.org/10.1183/09031936.00091709>
16. Sakao S, Tatsumi K. The importance of epigenetics in the development of chronic obstructive pulmonary disease. *Respirology* 2011;**16**(7):1056-63. <http://dx.doi.org/10.1111/j.1440-843.2011.02032.x>.
 17. Lundback B, Lindberg A, Lindstrom M, *et al.* Not 15 but 50% of smokers develop COPD? Report from the Obstructive Lung Disease in Northern Sweden Studies. *Respir Med* 2003;**97**(2):115-22. <http://dx.doi.org/10.1053/rmed.2003.1446>
 18. Barker DJP, Godfrey KM, Fall C, Osmond C, Winter PD, Shaheen SO. Relation of birth weight and childhood respiratory infection to adult lung function and death from chronic obstructive airways disease. *BMJ* 1991;**21**(303):671-5. <http://dx.doi.org/10.1136/bmj.303.6804.671>
 19. Brusselle G, Joos G, Bracke KR. New insights into the immunology of chronic obstructive pulmonary disease. *Lancet* 2011;**378**(9795):1015-26. [http://dx.doi.org/10.1016/S0140-6736\(11\)60988-4](http://dx.doi.org/10.1016/S0140-6736(11)60988-4).
 20. Garcha D, Thurston S, Patel A, *et al.* Changes in prevalence and load of airway bacteria using quantitative PCR in stable and exacerbated COPD. *Thorax* 2012;**67**(12):1075-80. <http://dx.doi.org/10.136/thoraxjnl-2012-201924>.
 21. Wedizcha JA, Donaldson GC. Natural history of successive COPD exacerbations. *Thorax* 2012;**67**(11):935-6. <http://dx.doi.org/10.1136/thoraxjnl-2012-202087>
 22. Atsou K, Chouaid C, Hejblum G. Variability of the chronic obstructive pulmonary disease key epidemiological data in Europe: systematic review. *BMC Med* 2011;**9**:7. <http://dx.doi.org/10.1186/1741-7015-9-7>
 23. Weinstein MC, Toy EL, Sandberg EA, *et al.* Modeling for health care and other policy decisions: uses, roles, and validity. *Value Health* 2001;**4**(5):348-61. <http://dx.doi.org/10.1046/j.1524-4733.2001.45061.x>
 24. Weinstein MC, O'Brien B, Hornberger J, *et al.* Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices—Modeling Studies. *Value Health* 2003;**6**(1):9-17. <http://dx.doi.org/10.1046/j.1524-4733.2003.00234.x>
 25. Phillips 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):355-71. <http://dx.doi.org/10.2165/00019053-200624040-00006>
 26. Weinstein MC. Methodologic issues in policy modeling for cardiovascular disease. *J Am Coll Cardiol* 1989;**14**(3 Suppl A):38-43A. [http://dx.doi.org/10.1016/0735-1097\(89\)90160-5](http://dx.doi.org/10.1016/0735-1097(89)90160-5)
 27. Unal B, Capewell S, Critchley JA. Coronary heart disease policy models: a systematic review. *BMC Public Health* 2006;**6**:213. <http://dx.doi.org/10.1186/1471-2458-6-213>

Available online at <http://www.thepcrj.org>

Appendix 1: Search Strategy

Medline copd and burden model - 1946 to week 4 March 2012

animals/

2 humans/

3 1 not (1 and 2)

4 2 not 3

5 lung diseases, obstructive/

6 exp pulmonary disease,chronic obstructive/

7 emphysem*.mp.

8 (chronic* adj3 bronchiti*).mp.

9 (obstruct* adj3 (pulmonary or lung* or airway* or airflow* or bronch* or respirat*)).mp.

10 COPD.mp.

11 COAD.mp.

12 COBD.mp.

13 AECB.mp.

(exacerbation* adj3 bronchiti*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, protocol supplementary concept, rare disease supplementary concept, unique identifier]

15 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14

16 4 and 15

17 prevalence/

18 Incidence/

19 "cost of illness"/ or forecasting/ or "quality of life"/

20 "burden of illness".mp.

21 quality-adjusted life years/ or models, statistical/ or monte carlo method/

22 Health Care Rationing/ or "disability adjusted life years".mp.

23 "Cause of Death"/

24 Hospitalization/

25 house calls/ or office visits/ or "referral and consultation"/

Appendix 1: Search Strategy

26 16 and (17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25)

27 (model.mp. or modelling.mp.)

28 26 and 27

Embase copd and burden model

1. exp ANIMAL/

2. Nonhuman/

3. Human/

4. 1 or 2

5. 3 not 4

6. Chronic obstructive lung disease/

7. Obstructive airway disease/

8. chronic bronchitis/

9. lung emphysema/

10. (Chronic\$ adj3 bronchiti\$).mp.

11. (obstruct\$ adj3 (pulmonary or lung\$ or airway\$ or airflow\$ or bronch\$ or respirat\$)).mp.

12. COPD.mp.

13. COAD.mp.

14. COBD.mp.

15. AECB.mp.

16. (Acute exacerbation adj3 chronic bronchitis).mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]

17. 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16

18. 5 and 17

19. prevalence/

20. incidence/

21. "cost of illness"/ or "health care cost"/

22. mortality/

23. "burden of disease".mp.

24. quality adjusted life year/ or "quality of life"/

25. "disability adjusted life year".mp.

26. morbidity/

27. 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26

28. 18 and 27

29 (model.mp. or modelling.mp.)

30 28 and 29

CAB abstracts

COPD and burden model

Appendix 1: Search Strategy

1. Animal.mp. [mp=abstract, title, original title, broad terms, heading words]
2. animal.mp.
3. human diseases.sh.
4. 3 not 2
5. chronic obstructive pulmonary disease.sh.
6. (chronic adj3 bronchit\$).mp. [mp=abstract, title, original title, broad terms, heading words]
7. pulmonary emphysema/
8. chronic obstructive lung disease.mp.
9. (obstruct\$ adj3 (pulmonary or lung\$ or airway\$ or airflow\$ or bronch\$ or respirat\$)).mp.
10. COPD.mp.
11. COAD.mp.
12. COBD.mp.
13. AECB.mp.
14. (exacerbat\$ adj3 bronchi\$).mp. [mp=abstract, title, original title, broad terms, heading words]
15. bronchitis.sh.
16. 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15
17. 3 and 16
18. prevalent.mp. or disease prevalence.sh.
19. incidence.sh.
20. "burden of disease".mp.
21. economic impact.sh.
22. "causes of death"/
23. morbidity/
24. health services/
25. "house call".mp.
26. health care costs/
27. "cost benefit analysis"/
28. 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27
29. 17 and 28

- 30 (model.mp. or modelling.mp.)

- 31 29 and 30

Appendix 1: Search Strategy

WHOLIS (World Health Organization Library Information Services)

"chronic obstructive pulmonary disease" and prevalence - 0 results

("chronic obstructive pulmonary disease" or bronchitis or emphysema) and prevalence – 0

("chronic obstructive pulmonary disease" or bronchitis or emphysema) and (prevalence or incidence or mortality or morbidity)

Global Health Library Regional Indexes

("chronic obstructive pulmonary disease" or bronchitis or emphysema) and (prevalence or incidence or mortality or morbidity)

AIM (AFRO),

LILACS (AMRO/PAHO),

IMEMR (EMRO),

IMSEAR (SEARO),

WPRIM (WPRO)

Appendix 2: Data extraction and quality of reporting

Reference	
Author and email address	
Year	
Institution	
Funding source	
Purpose of model	
Model Title	
Model Type	
Model setting, time period and population (age, sex and country).	
Inputs to model	
Details of model's processing (including algorithm)	
Model output/ results	
Model availability	
Comparisons with other studies	
Social and economic policy implications of model outcomes	
Future research recommendations	
Any other comments	

Risk Factors Included - not a risk factor study

Risk Factor	Tick if included	Form (Con/Cat)	Describe intervention
Smoking			
Indoor air pollution			
Outdoor air pollution			
Socioeconomic deprivation			
Nutrition			
Other			
Pulmonary rehabilitation			
Smoking cessation			
Bronchodilator			
Corticosteroid			
Other treatment			

Appendix 2: Data extraction and quality of reporting

Disease categories included – please tick

		NICE 12 2004	ATS/ERS 2004	GOLD 2010 ⁶	NICE 101 update 2011 ³⁰	Please Tick Category	
Post-bronchodilator FEV1/FVC	FEV1 % predicted	Severity of airflow obstruction					
< 0.7	≥ 80%		Mild	Stage 1 – Mild	Stage 1 – Mild*		
	50-79%	Mild	Moderate	Stage 2 – Moderate	Stage 2 – Moderate		
	30-49%	Moderate	Severe	Stage 3 – Severe	Stage 3 – Severe		
	< 30%	Severe	Very Severe	Stage 4 – Very Severe**	Stage 4 – Very Severe**		

* symptoms should be present to diagnose COPD in people with NICE 2011 mild airflow obstruction.

**or FEV1<50% with respiratory failure

Appendix 2: Data extraction and quality of reporting

Outcomes studied

Outcomes studied		
What is the prevalence of COPD per region?	Prevalence rate	
What is the incidence of COPD per region?	New cases per thousand person years	
What is the COPD disease-specific mortality?	COPD-related mortality	
What is the COPD disease-specific burden to the individual?	Disability/quality adjusted life years lived with mild/moderate/severe COPD Monetary cost of COPD healthcare to the individual	
What is the COPD disease-specific burden to the healthcare system?	Average annual GP visits Average annual emergency dept visits Average annual hospital admissions per patient for COPD exacerbations Average annual readmissions per patient (measure of effectiveness of treatment)	
What is the COPD-specific burden to society	Cost of healthcare cumulative loss of earnings by patients cumulative loss of time at work/study carers burden	
Other		
What is the "main outcome" of the study in the author's words		

Appendix 2: Data extraction and quality of reporting

QUALITY OF REPORTING ASSESSMENT

Model purpose and aim

Statement of the question which the model is trying to answer	
Perspective of model	
Time horizon of model	
Model type	

Transparency

Transparency	Not available	Available
Illustrations/examples		
Assumptions		
Model availability for reader		

Data Input: - not much detail given

Type of data	Source	Comment on quality (sample size and response rate for surveys etc.)	Limitations
Population data			
Mortality data/rate			
Morbidity data/rate			
Treatment uptake			
Risk factor prevalence/trends			
Treatment effectiveness			
Risk factor change effectiveness /Betas			
Costs			
Health utilities			

Data modelling

Discussion of model's derivation	
Assumptions documented and justified	
Model consistent with accepted techniques of statistics and epidemiology	

Appendix 2: Data extraction and quality of reporting

Data Incorporation:

Deterministic methodology	
Probabilistic methodology	

Sensitivity analysis

Were sensitivity analysis carried out (Y/N)	
Were 95% CI for RRS used for sensitivity analyses	
Which analyses	
Was the discussion of sensitivity analyses	Poor Reasonable Good
Please tick	

Internal validation

Was there evidence that the model had undergone debugging	
Was there evidence that the model had been calibrated	
How was the model calibrated? (describe)	

Was the predictive validity of the model tested? (Y/N).....

How was the predictive validity of the model checked? (Describe)	
How was the validity quantified?	e.g. % explained

Potential Limitations

Potential Limitations	Not Reported	Reported	Discussed	Method refined
Assumptions				
Confounding				
Lag times				
Competing causes				

Involvement of policy makers, planners and decision makers in model:

Who was involved?	
How and at what stage were they involved?	
Will policy makers, planners and decision makers have an opportunity to respond to the results of the study?	

Other comments on the study:

--

Appendix 2: Data extraction and quality of reporting

Overall summary

purpose and aim 4	
Transparency 3	
Data 1	
Data modelling 3	
Sensitivity analysis 2	
Internal validity 1	
Calibration 1	
Involvement of policymakers 1	
predictive validity 3	
Discussion of limitations 1	
Overall mark /20	

Models for estimating projections for disease prevalence and burden: a systematic review focusing on chronic obstructive pulmonary disease

Susannah McLean¹, Victoria Barbour², Sarah Wild³,
Colin Simpson⁴ and Aziz Sheikh⁵

Journal of Health Services Research &
Policy
0(0) 1–8
© The Author(s) 2015
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/1355819615579232
jhsrp.rsmjournals.com


Abstract

Objective: Epidemiological models for estimating the prevalence and burden of disease inform health policy and service planning decisions. Our aim was to describe the challenges in evaluating such models using the example of epidemiological models for chronic obstructive pulmonary disease (COPD).

Methods: Two reviewers searched Medline, Embase, CAB Abstracts and World Health Organization (WHO) Databases from 1980 to November 2013 for epidemiological models of COPD prevalence and burden. Two reviewers extracted data and assessed the quality of the studies. We then undertook a descriptive and narrative synthesis of data.

Results: We identified 22 models employing a variety of techniques to calculate the prevalence and/or burden of COPD. Models calculated prevalence and/or mortality or other facet of disease burden using demographics and risk factors or trends, Markov-type modelling and microsimulation modelling. The six models which scored highly on the quality framework were: the Peabody model, which generated estimates of COPD prevalence; the WHO DISMOD II model which produced burden estimates in terms of disability adjusted life years with COPD and life years lost to COPD; the Atsou model which gave the life expectancy gains of individual smokers who quit smoking and associated costs; two Dutch COPD models which produced estimates of mortality and health care costs related to COPD; and the Pichon–Riviere model which gave the costs and cost effectiveness of smoking quit programmes.

Conclusions: The field of chronic disease modelling is burgeoning. As a result, policy makers need to understand how to interpret epidemiological models and their data sources.

Keywords

chronic obstructive pulmonary disease, epidemiology, systematic review

Introduction

Governments need accurate and timely information in order to be able to plan for the health care needs of their population. Over recent years, the challenges of ageing populations with a high prevalence of chronic diseases have been widely discussed.¹ While epidemiological modelling has traditionally been used to describe patterns of infectious diseases, chronic disease epidemiological modelling is relatively new. It combines elements of mathematical and health economic disease modelling with clinical data to estimate current and future prevalence and disease burden. However, as models proliferate, decision makers will need to choose which will best suit their purposes.

¹PhD Student, Allergy and Respiratory Research Group, Centre for Population Health Sciences, The University of Edinburgh, UK

²Research Fellow, Allergy and Respiratory Research Group, Centre for Population Health Sciences, The University of Edinburgh, UK

³Professor of Epidemiology, Centre for Population Health Sciences, The University of Edinburgh, UK

⁴Senior Lecturer, Allergy and Respiratory Research Group, Centre for Population Health Sciences, The University of Edinburgh, UK

⁵Professor of Primary Care Research and Development, Allergy and Respiratory Research Group, Centre for Population Health Sciences, The University of Edinburgh, UK

Corresponding author:

Susannah McLean, Allergy and Respiratory Research Group, Centre for Population Health Sciences, The University of Edinburgh, Doorway 1, Medical Quad, Teviot Place, Edinburgh, EH8 9AG, UK.
Email: Susannah.mclean@ed.ac.uk

Key challenges in the evaluation of an epidemiological model

First, it is pertinent to consider as precisely as possible the definition of the disease that is being modelled. Diseases may be defined according to the presence of risk factors or according to diagnostic tests.

Second, data sources should be clearly documented so that they can be consulted and assessed by other researchers for their validity. This is sometimes poorly done in modelling; however, it is important that other researchers can decide for themselves whether the sources of data used for the modelling are the most representative available and substitute alternative sources of data into the model where appropriate to allow cross validation of the model.

Third, the method of modelling should be documented transparently so that future researchers can understand what analysis has taken place. The 2010 global burden of disease² did not make clear how they had processed their data. The earlier version of the global burden of disease had an online supplement and additional publication^{3,4} detailing how they had calculated prevalence from mortality figures; however, this was lacking for the 2010 version.

Fourth, modelling is a relatively young science and there are a variety of different techniques being used. The Markov model is one of the most commonly used, and many models in this systematic review describe themselves as Markov models or Markov-type models. Such models originate in the discipline of health economics but also give estimates of prevalence and burden. Essentially, Markov models describe a disease as a number of linked health states and a cohort moves through these states from health to deteriorating health to (eventually) death. Sometimes states can additionally describe recovery from disease. Data from clinical trials is used to inform the transition probabilities between the states. For an excellent introduction to Markov models please see Briggs and Sculpher.⁵

Fifth, the time frame for the model's operation needs to be considered. It should be considered whether the model is set to use past data to produce an estimate for current times and whether this done by projecting trends or by another method, or does it use present data and project into the future. The population for which the projection is made should also be considered. Populations as big as entire countries are often the frame of reference; however, this often pools estimates across urban and rural areas and other forms of geographical heterogeneity.

Sixth, sensitivity analysis should be carried out. This is where the inputs of the model are altered by a threshold amount, e.g. 10% and the model re-run to see the effect on the output. This allows the reviewer to thereby

evaluate how sensitive the model is to each of its inputs. Sensitivity analysis may be one-way where only one input parameter is changed at a time or multi-way where additional input parameters are changed simultaneously to evaluate the combined effect of these simultaneous changes on the output.

And finally, it would be possible to compare the results from each model with results from large prevalence surveys. Comparisons with surveys have to ensure that the model and the survey were aiming for the same target population prevalence or burden. In the absence of an appropriate survey with which to compare, more than one model for the same population can be used for cross validation.

Our aim was to review systematically the models available for one specific long-term condition – chronic obstructive pulmonary disease (COPD).

Methods

We developed and reported a detailed protocol for this work.⁶ We also registered the protocol in the PROSPERO database.⁷

Eligibility criteria

Any modelling study which used demographic and epidemiological data to estimate the prevalence and/or disease burden was eligible for inclusion. The outcomes of interest were incidence, prevalence, disease burden and mortality. For the purposes of this review, disease burden was considered from the perspective of the health system.

The types of models of interest included demographic models, microsimulation models and Markov-type models. Models were excluded if they described animal cell lines, clinical series or estimates of individual risk (such as individual prognostic models). Decision analytical models or decision support models were excluded where they referred to clinical decision-making for individuals. Models comparing one intervention with another intervention were also excluded as the aim was to estimate the prevalence, disease burden and mortality rather than to investigate the effectiveness of interventions.

Information sources and study selection

A search strategy was developed using search terms to include the concepts of 'modelling', 'disease burden' and 'chronic obstructive pulmonary disease' as has been fully described elsewhere in McLean et al.⁶ Searches were conducted in Medline, Embase, CAB Abstracts, World Health Organization (WHO) Library and Information Services (WHOLIS – library

catalogue of books and reports), WHO Regional Indexes (AIM (AFRO, LILACS, (AMR/PAHO)), IMEMR (EMRO), IMSEAR (SEARO), WPRIM (WPRO)). A modified search strategy was used to identify reports from the WHO home website and Google. Searches were for both published and unpublished modelling studies from 1980 (when modelling first began to be widely used^{8,9}) to November 2013. All studies were independently reviewed against the stated inclusion criteria by SM and VB and all disagreements were resolved by discussion between reviewers.

Data extraction

A piloted data extraction form was used by SM and checked by VB. The following data were extracted from each study: author and email address, year, institution and funding source, the purpose of the model, model title, model type, model setting, time period and population. Model inputs and source of input data details of processing of the model, were also extracted, along with COPD outcomes (incidence, prevalence, mortality, primary care visits, emergency department visits, hospitalizations, treatment costs), model output/ results, details of the model's availability, any comparisons with other studies, social and economic policy implications of model's output and future research recommendations.

Quality appraisal framework

We quality appraised the reporting of the models. A quality of reporting framework was designed following review of key guidelines as to good practice in modelling.^{10–13} A scoring mechanism was devised to weight the different elements required to produce a relevant and high quality model.¹⁴ This scoring framework was described in the protocol for this systematic review⁶ and considered the reporting of whether model development and the elements of calibration, internal and external validation and whether it was reported that policy and decision makers had had input into the development of the model. The quality of reporting scoring was up to a maximum of 20 points.

Synthesis of results

As the models had different purposes and were based in different settings we present a narrative synthesis of results as heterogeneity precluded any meaningful quantitative synthesis of results.

Results

Around 1726 studies were title-screened following duplicate deletion. About 157 titles and abstracts were

selected for full text review. Excluded were regression models which were quantifying the effect of risk factors on COPD rather than projecting prevalence and/or disease burden. Twenty-one models^{3,4,15–35} and DISMOD 3 (unpublished) were selected, as described in the PRISMA flow diagram in Figure 1.

Six models scored highly on the quality of reporting framework out of a maximum 20 points: Peabody's prevalence model²¹ (16), Atsou's smoking burden model³⁶ (16), WHO burden model as described by Shibuya⁴ (16), two Dutch models: Feenstra²¹ (17) and Hoogendoorn et al.¹⁷ (17) and the Pichon–Riviere smoking burden model³⁴ (17.5). These models are described below in detail; summaries of all models are found in Tables 1 to 4, available online. High reporting quality should ensure that adequate information is included in the report in order to draw some conclusions as to the underlying quality of the model. The main focus of quality was whether sensitivity analysis had been reported and what were the results and other features of quality assurance such as debugging and external validation of results.

Peabody prevalence estimation model

This model used six key risk factors for COPD identified and quantified from the literature: smoking, age, gender, indoor air pollution, outdoor air pollution and occupational exposure to airborne particles. They assumed no cases under the age of 30 and a smoking prevalence of 15%. Based on the literature it was assumed that there was an exposure of non-smoking population to occupational airborne particles of 1.9% and a prevalence of COPD in urban populations of 1.4%. In the model, population was distributed by percentage between urban and rural populations and linked to environmental exposure. There was also an input for national socio-economic development based on World Bank figures. COPD prevalence was then estimated for four countries and compared with survey data for external validation. Then COPD prevalences for an additional 12 countries were estimated.

Sensitivity analysis was performed on hypothetical populations with differences in their rural/urban population distribution and high/low income split. The model was externally validated by comparing its results to survey results for Nepal, Norway, Poland and Spain. The predictions were not statistically different from the survey results for any of the countries suggesting that the model is a useful prediction tool.

Atsou smoking burden model

This Markov-type model included the severity stages of COPD: mild, moderate, severe and very severe.

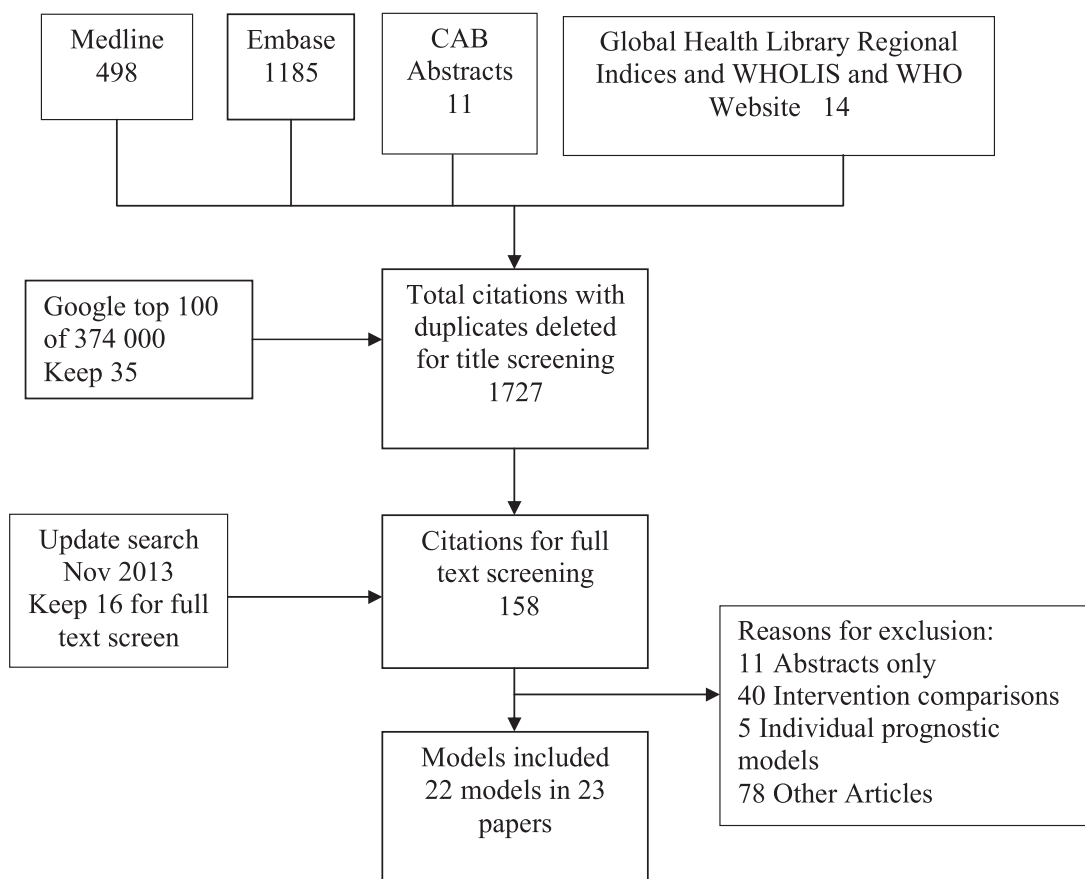


Figure 1. PRISMA diagram.

Transition probabilities could be altered for modelled patients moving up these severity stages as their disease progressed. The model aimed to report the impact of smoking cessation on a COPD patient's life expectancy in terms of individual health gains. The model includes a cost-effectiveness assessment of the impact of smoking cessation programmes in England.

The sensitivity analyses that were conducted involved changing the transition rates from one disease severity stage to the next, the mortality and exacerbation rates and the costs of COPD management, different discounting rates and different smoking cessation rates. For the scenarios investigating the costs, Quality Adjusted Life Years and cost effectiveness of the smoking cessation programmes; the effects of different input costs for the programme and different quit rates, were investigated. The results were sensitive to transition rates from one disease stage to the next and to an increase in mortality rates. When it was suggested that ex-smokers experience fewer exacerbations than smokers, there were monetary and health care gains. The model was not very sensitive to changes in disease management costs.

WHO burden of disease model DISMOD II

DISMOD II used the WHO world regions to calculate the burden of COPD in terms of years lived with disability and mortality in terms of years of life lost. The model was based on a regression model including a measure of smoking impact and an air pollution variable to take into account proportions of households in each region that use indoor biofuels and age and sex dummy variables which reflected variability in the exposure data for different regions. The input data to the model was the COPD-related mortality rates and total mortality rates per WHO region. These data were used to generate an equation which could be solved for the prevalence and incidence of COPD.

No sensitivity analysis using DISMOD II was undertaken. However, the predictive validity of the model was checked by comparing the estimated relative risks for death from COPD for the America and Pacific regions with recent burden of disease analyses from the United States and Australia. Published prevalence data based on spirometry in the WHO regions were also compared to the DISMOD II outputs and found

to be consistent with the exception of estimates from the Global burden of disease 1990 estimates due to improved methodology since then.³⁷

Feenstra chronic disease model

This model concentrated on the population with COPD in the Netherlands. A dynamic multistate life table model combined information on the demography of the Dutch population and smoking. The number of new cases of COPD each year was calculated from the incidence rates of COPD for smokers and former smokers in combination with prevalence data. Independently modelled modules exist for mortality from lung cancer, stroke, coronary heart disease and asthma. The model is dynamic in that each year's total incorporated calculations for birth, and mortality.

One way sensitivity analysis was performed to determine the effect of changes in inputs of incidence and excess mortality rates on the output prevalence rates. It was found that a 20% change in incidence input rates resulted in a 17% change in prevalence rates and a 20% change in excess mortality input resulted in a 4% change in prevalence output. Therefore, the model was more sensitive to changes in incidence rates.

Hoogendoorn COPD model

This Markov-type model was developed from the Feenstra chronic disease model¹⁶ by the addition of the four Global Initiative for Obstructive Lung Disease (GOLD) severity stages for COPD into the model. The COPD population is divided into mild, moderate, severe and very severe COPD subpopulations and also by smoking status: never-smoker, smoker or former smoker. Annual incidence and mortality rates are applied to each age, sex and smoking subpopulation. Costs are calculated by multiplying annual cost data for each severity stage by the number of patients in each stage.

Sensitivity analyses for this model were conducted by changing the input severity distribution of the COPD prevalence so that it was assumed firstly that COPD patients began with less severe COPD: more mild and moderate cases, then it was assumed that COPD patients began with more severe COPD: more severe and very severe cases. The next sensitivity analysis to be conducted was similarly changing the severity distribution of the incident COPD cases to first less severe and then more severe. Further sensitivity analyses were conducted varying the rate of lung function decline and changing the impact on lung function decline of stopping smoking from an improvement to a null impact. The results of the sensitivity analyses were that all prevalence results were within a range of 5% of the projections of the base case. The results were

most sensitive to the input of the severity distribution of the incidence. Cost projections were more sensitive than prevalence projections due to the difference in costs for different COPD severity stages.

Pichon–Riviere smoking burden model

This model met the needs of 68 decision makers who had been surveyed across seven countries in Latin America (i.e. Argentina, Bolivia, Brazil, Chile, Colombia, Mexico and Peru). It considered heart disease, cerebrovascular disease, COPD, pneumonia/influenza, lung cancer and nine other neoplasms. The model used country-specific data sources where possible. However, as a result of the lack of local data, estimates of incidence were often derived from mortality data using the WHO DISMOD II methodology.⁴ The baseline risk in non-smokers was then calculated based on the age-, sex- and country-specific smoking prevalence as well as disease-specific smoking relative risk. The parameters for smokers and former smokers were then calculated from this baseline risk. The model was a microsimulation of individual subjects incorporating the natural history, costs and quality of life of the above diseases. A functioning version of the model was constructed validated and calibrated using data from Argentina.

The Argentinian version of this model underwent extensive internal validation with debugging by the inputting of null and extreme values and checks for inconsistencies. The model was then calibrated by comparing general mortality and all age- and sex-specific death rates predicted by the model with local health statistics. COPD was excepted from this process as COPD mortality was agreed to be underestimated in national statistics. Equations were modified to improve fit to the reference values. Lethality and survival rates were estimated from local and international studies and also used to calibrate the model by visual exploration of observed and expected curves to confirm a good fit. After calibration, as had been expected, correlation between predicted and observed results was better among high incidence events.

External validation for age- and sex-specific COPD predicted prevalence was performed by comparison with the results from the Latin American Project for the Investigation of Obstructive Lung Disease, a population-based survey carried out in five Latin American cities.³⁸ The model underestimated the level of COPD in comparison to the survey; however, the results were within 5% for each age group. This model's report followed the International Society Pharmacoeconomics and Outcomes Research guidelines for model development¹¹ and reporting and therefore included extensive description of preliminary

consultation with policy and decision makers and detail regarding debugging, internal validation, calibration and external validation procedures.

Discussion

Main findings

We identified 22 models, of which six scored highly on the quality of reporting framework. These six were: the Peabody model²¹ generates credible estimates of COPD prevalence in specific countries round the world. In addition, the WHO DISMOD II model⁴ produces burden estimates in terms of disability adjusted life years with COPD and life years lost to COPD and the Atsou model³² gives the life expectancy gains of individual smokers who quit smoking and associated costs. Feenstra et al.¹⁶ and Hoogendoorn et al.¹⁷ models produce estimates of mortality and health care costs related to COPD. The Pichon–Riviere model³⁴ gives the costs and cost effectiveness of smoking quit programmes in the context of all tobacco-related diseases.

Strengths and limitations

This study involved a very broad and comprehensive search strategy including many world databases included in the WHO's library. A limitation of this study is that we did not consider models that studied the effects of interventions on COPD because it was decided that our first priority should be to establish the baseline burden of COPD and pharmacological interventions have already been systematically reviewed.³⁹

COPD models are chiefly focused on smoking as the main risk factor and any risk-reduction predictions made by these models were from consideration of smoking alone. As understanding of the pathogenesis of COPD increases the role of other factors in the development of COPD is better understood, the overwhelming focus on smoking may be considered to be a further limitation of these models.

A limitation of using the quality of reporting framework is that a high quality of model may be ignored because its reporting is not such high quality or lacks clarity. No report on the development of the model and its subsequent quality validation was found for COPD for DISMOD 3, this is the model that was used to calculate Global burden of Disease in the recent Lancet series.^{2,40}

Deep critical review of all the models in this paper was not possible as we did not have full access to all the models and their underlying mathematics. This is a potential limitation when reviewing any kind of

model, there is a need for transparency in publication and for a mathematical modelling skill-set to interpret the findings.

Interpretation of findings in the light of previous research

This is the first systematic review of modelling studies for COPD prevalence and burden. A review of chronic heart disease policy models¹⁴ commented that while reporting criteria are available for many study types this does not yet apply to modelling studies and this affects the quality of model reporting and consequently the underlying model quality as inferred by potential model users. In addition, it was highlighted that although models are heavily reliant on their data inputs, few models critique the quality of their data sources.

Implications for policy and research

Implications for COPD policy are not clear overall as the models included have many different purposes and were designed for different contexts, there is not yet a consensus on the structure of the optimum COPD model.

Implications for policymakers include that, as techniques of chronic disease modelling are increasingly being used and recognized, they need to be aware of how to interpret them and how to critique the datasets that are used as model inputs.

Implications for research on COPD models include that sensitivity analyses should be conducted in order to highlight the parameters that most affect the outputs and so demonstrate the internal validity of each model. Validation of the included COPD models is fertile ground for further research, in particular as more detailed and higher quality data become available. In addition the design of future COPD models to include other risk factors such as prematurity, low birthweight, asthma, tuberculosis and childhood respiratory infections needs further research.

In terms of models in general, Markov-type models cannot easily model differences between cohorts and so future work could be directed towards how to best represent such differences where they exist.

Conclusions

COPD epidemiological models have widely differing structures and include Markov models and microsimulation models. In general, the field of chronic disease modelling is burgeoning. As a result, policy makers need to understand how to interpret epidemiological models and their data sources.

Funding

SM is funded by the University of Edinburgh and an Edinburgh and Lothian Research Fund Grant. VB had no specific funding.

Acknowledgements

SW, CS and AS are all members of staff at the University of Edinburgh. AS has served as a research consultant to Admirall and Napp.

Professor Simon Capewell of Liverpool University helped to devise the quality of reporting scoring framework.

References

1. DoH. *Improving the health and wellbeing of people with long term conditions*. London: Department of Health, 2010.
2. Murray CJ, Vos T, Lozano R, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012; 380: 2197-2223.
3. Barendregt JJ, Van Oortmarssen GJ, Vos T, et al. A generic model for the assessment of disease epidemiology: the computational basis of DisMod II. *Popul Health Metr* 2003; 1: 4.
4. Shibuya K, Mathers CD and Lopez AD. *Chronic obstructive pulmonary disease, consistent estimates of incidence, prevalence and mortality by WHO region*. Geneva: World Health Organization, 2001.
5. Briggs A and Sculpher M. An introduction to Markov modelling for economic evaluation. *Pharmacoeconomics* 1998; 13: 397-409.
6. McLean S, Wild SH, Simpson CR, et al. Models for estimating projections for the prevalence and disease burden of chronic obstructive pulmonary disease (COPD): systematic review protocol. *Prim Care Respir J* 2013; 22: PS8-21.
7. PROSPERO. Register of Systematic Reviews. University of York; 2013. <http://www.crd.york.ac.uk/prospéro/>
8. Weinstein MC, Coxson PG, Williams LW, et al. Forecasting coronary heart disease incidence, mortality, and cost: the Coronary Heart Disease Policy Model. *Am J Public Health* 1987; 77: 1417-1426.
9. Gunning-Schepers L. The health benefits of prevention: a simulation approach. *Health Policy* 1989; 12: 1-255.
10. Weinstein MC, Toy EL, Sandberg EA, et al. Modeling for health care and other policy decisions: uses, roles, and validity. *Value Health* 2001; 4: 348-361.
11. Weinstein MC, O'Brien B, Hornberger J, et al. Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices-Modeling Studies. *Value Health* 2003; 6: 9-17.
12. Philips Z, Bojke L, Sculpher M, et al. Good practice guidelines for decision-analytic modelling in health technology assessment: a review and consolidation of quality assessment. *Pharmacoeconomics* 2006; 24: 355-371.
13. Weinstein MC. Methodologic issues in policy modeling for cardiovascular disease. *J Am Coll Cardiol* 1989; 14(3 Suppl A): 38A-43A.
14. Unal B, Capewell S and Critchley JA. Coronary heart disease policy models: a systematic review. *BMC Public Health* 2006; 6: 213.
15. Nacul LC, Soljak M and Meade T. Model for estimating the population prevalence of chronic obstructive pulmonary disease: cross sectional data from the Health Survey for England. *Popul Health Metr* 2007; 5: 8.
16. Feenstra TL, van Genugten ML, Hoogenveen RT, et al. The impact of aging and smoking on the future burden of chronic obstructive pulmonary disease: a model analysis in the Netherlands. *Am J Respir Crit Care Med* 2001; 164: 590-596.
17. Hoogendoorn M, Rutten-van Molken MP, Hoogenveen RT, et al. A dynamic population model of disease progression in COPD. *Eur Respir J* 2005; 26: 223-233.
18. Hoogendoorn M, Rutten-van Molken MP, Hoogenveen RT, et al. Developing and applying a stochastic dynamic population model for chronic obstructive pulmonary disease. *Value Health* 2011; 14: 1039-1047.
19. Lhachimi S, Nusselder WJ, Smit HA, et al. DYNAMO-HIA - a Dynamic Modeling tool for generic Health Impact Assessments. *PLoS One* 2012; 7(5): e33317.
20. Kulik MC, Nusselder WJ, Boshuizen HC, et al. Comparison of tobacco control scenarios: quantifying estimates of long-term health impact using the DYNAMO-HIA modeling tool. *PLoS One* 2012; 7: e32363.
21. Peabody JW, Schau B, Lopez-Vidriero M, et al. COPD: A prevalence estimation model. *Respirology* 2005; 10: 594-602.
22. Stang P, Lydick E, Silberman C, et al. The prevalence of COPD: using smoking rates to estimate disease frequency in the general population. *Chest* 2000; 117(5 Suppl 2): 354S-359S.
23. Pham TM, Ozasa K, Kubo T, et al. Age-period-cohort analysis of chronic obstructive pulmonary disease mortality in Japan, 1950-2004. *J Epidemiol* 2012; 22: 302-307.
24. Erbas B, Ullah S, Hyndman RJ, et al. Forecasts of COPD mortality in Australia: 2006-2025. *BMC Med Res Methodol* 2012; 12: 17.
25. Lopez AD, Shibuya K, Rao C, et al. Chronic obstructive pulmonary disease: current burden and future projections. *Eur Respir J* 2006; 27: 397-412.
26. Bronnum-Hansen H and Juel K. Estimating mortality due to cigarette smoking: two methods, same result. *Epidemiology* 2000; 11: 422-426.
27. Holt JB, Zhang X, Presley-Cantrell L, et al. Geographic disparities in chronic obstructive pulmonary disease (COPD) hospitalization among Medicare beneficiaries in the United States. *Int J Chron Obstruct Pulmon Dis* 2011; 6: 321-328.
28. Chapmann KR, Mannino DM, Soriano JB, et al. Epidemiology and costs of chronic obstructive pulmonary disease. *Eur Respir J* 2006; 27: 188-207.
29. Perera PN, Armstrong EP, Sherrill DL, et al. Acute exacerbations of COPD in the United States: inpatient

- burden and predictors of costs and mortality. *COPD* 2012; 9: 131–141.
30. Najafzadeh M, Marra CA, Lynd LD, et al. Future impact of various interventions on the burden of COPD in Canada: a dynamic population model. *PLoS One* 2012; 7: e46746.
 31. Stanciole AE, Ortegon M, Chisholm D, et al. Cost effectiveness of strategies to combat chronic obstructive pulmonary disease and asthma in sub-Saharan Africa and South East Asia: mathematical modelling study. *BMJ* 2012; 344: e608.
 32. Atsou K, Chouaid C and Hejblum G. Simulation-based estimates of effectiveness and cost-effectiveness of smoking cessation in patients with chronic obstructive pulmonary disease. *PLoS One* 2011; 6: e24870.
 33. Hurley SF and Matthews JP. The quit benefits model: a Markov model for assessing the health benefits and health care cost savings of quitting smoking. *Cost Eff Resour Alloc* 2007; 5: 2.
 34. Pichon-Riviere A, Augustovski F, Bardach A, et al. Development and validation of a microsimulation economic model to evaluate the disease burden associated with smoking and the cost-effectiveness of tobacco control interventions in Latin America. *Value Health* 2011; 14(5 Suppl.): S51–S9.
 35. van Genugten ML, Hoogenveen RT, Mulder I, et al. Future burden and costs of smoking-related disease in the Netherlands: a dynamic modeling approach. *Value Health* 2003; 6: 494–499.
 36. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012; 380(9859): 2224–2260.
 37. Murray CJL. Quantifying the burden of disease; the technical basis for disability adjusted life years. *Bull WHO* 1994; 72: 429–445.
 38. Menezes AM, Lopez MV, Hallal PC, et al. Prevalence of smoking and incidence of initiation in the Latin American adult population: the PLATINO study. *BMC Public Health* 2009; 9: 151.
 39. Rutten-van Molken MP and Goossens LM. Cost effectiveness of pharmacological maintenance treatment for chronic obstructive pulmonary disease: a review of the evidence and methodological issues. *Pharmacoeconomics* 2012; 30: 271–302.
 40. Murray CJ, Richards MA, Newton JN, et al. UK health performance: findings of the Global Burden of Disease Study 2010. *Lancet* 2013; 381: 997–1020.