The epidemiology of chronic obstructive pulmonary disease in the UK: spatial and temporal variations

Thesis submitted to the Faculty of Medicine of the University of London for the degree of Doctor of Philosophy

Anna Louise Hansell
Division of Primary Care and Public Health Sciences
Imperial College London, St Mary's Campus
2004

Abstract

Introduction

The UK has experienced high rates of chronic obstructive pulmonary disease (COPD) with marked spatial variability, for at least the past century. This PhD investigated hypotheses that factors other than smoking have had important influences on both temporal and spatial variations on COPD mortality.

Methods

Age-period-cohort analyses were used to examine time trends in COPD mortality rates 1950-99 in conurbation and non-conurbation areas of England & Wales in relation to the 1956 Clean Air Act.

Spatial analyses were conducted for UK districts in 1981-99. As smoking information was limited, a Bayesian shared component model was used to estimate COPD risks that were independent of risks shared with lung cancer (a proxy for cumulative smoking). Linear regression of log COPD risks was used to find associations with selected risk factors.

Results

COPD mortality rates have generally declined since 1950, with major falls occurring between the late 1950s to late 1970s that were larger in conurbation than non-conurbation areas. Falls were located as period (year on year) effects, independent of age and cohort effects, thought to capture much of the impact of smoking. Results were most consistent with contemporary declines in particulate air pollution, but changes in cigarette composition and improvements in COPD treatments may have contributed.

In multiple regression analyses for 1981-99, district COPD mortality risks were higher with higher deprivation and ambient SO_2. Significant associations with temperature and rainfall were complex, showing interactions with deprivation and each other, but suggested effects of cold and damp. COPD risks were also higher in districts with high pneumoconiosis deaths in males and low fruit and vegetable consumption in females.

Conclusion

Factors other than smoking, particularly air pollution, but also diet, occupational exposures and climate, exert important influences on COPD. These factors could be targeted in public health attempts to reduce COPD mortality.

Declaration

The work in this thesis is my own and where any material could be construed as the work of others, it is fully cited and referenced and/or full acknowledgement is given.

I had the original idea for the study, developed the study and conducted all analyses except as noted below. The final format of the study design was developed following input from Prof David Strachan and Prof Paul Elliott at the MPhil upgrade. The statistical methodology to be used was chosen in discussion with my statistical supervisors, Dr Nicky Best and Prof Leonhard Held (who was a supervisor on this PhD until 2003, when he moved back to Germany).

Part of the temporal analyses represents collaborative work with Prof Held and Herr Ralf Breuninger from the Institut für Statistik, Ludwig-Maximilians-Universität, München. Specifically the statistical methodology relating to classical age-period-cohort analyses incorporating an interaction term with degree of urbanisation was developed by Prof Held. Statistical analyses were conducted by Ralf Breuninger, supervised by Prof Held and using data supplied by me. I interpreted the analyses, supervised by Prof Held, and I also performed additional analyses such as linear regression of relative risks against year to find average changes in period effects and graphs of results.

The annual average air pollution data for individual monitoring stations were supplied by colleagues within my department, using data originally derived from the UK National Air Quality Information Archive. Maps in Figure 4-1 and Figure 8-2 were produced by colleagues, using data supplied by me. The extracts from the Taylor Nelson Sofres (TNS) dataset were provided by Giri Madhavan and Jennifer Hurley, who were advised by Claire Robertson and myself. I conducted the TNS data cleaning and the validation analyses.

Acknowledgments and thanks to:

Julian Seymour for patience, understanding and cooking when needed most.

My supervisors Paul Aylin, Nicky Best and Leonhard Held for their continued support and encouragement, regular supervision meetings and comments on drafts.

David Strachan, St George's Hospital and Paul Elliott for their helpful comments at my MPhil upgrade, which helped to refine the thesis into a scientific enquiry.

David Briggs for advice on air pollution exposure and deprivation measures.

Daniela Fecht, Danielle Vienneau, Jon Mitchell and especially Kees de Hoogh for advice and technical assistance with datasets used in spatial analyses, GIS and mapping issues.

Eric Johnson, without whose computing support this PhD would not have been possible. Linh Nguyen for additional computing support.

Annie Jones for information on historical antibiotic policies for COPD.

Alison Loader, AEA Technology Environment for advice on historical air pollution datasets and provision of paper records for data prior to 1961.

Guy Marks for advice on changes in mortality coding over time.

Giri Madhavan and Jennifer Hurley for extracting the TNS dataset. Claire Robertson for advice on nutrition and help on defining a fruit & vegetable variable.

Table of contents

ABSTRACT **2**
DECLARATION **3**
ACKNOWLEDGMENTS AND THANKS TO: **4**
TABLE OF CONTENTS **5**
LIST OF TABLES **9**
LIST OF FIGURES **12**
LIST OF ABBREVIATIONS **15**
SECTION I **17**
CHAPTER 1. INTRODUCTION **18**
OVERVIEW OF CHAPTER 1 18
INTRODUCTION 18
CURRENT IMPORTANCE OF COPD MORBIDITY AND MORTALITY IN THE UK 19
HYPOTHESES TO BE ADDRESSED BY THIS PHD 20
Hypotheses relating to time trends in COPD mortality *20*
Hypotheses relating to spatial variations in COPD mortality *22*
STRUCTURE OF THIS PHD DOCUMENT 22
CHAPTER 2. LITERATURE REVIEW **24**
OVERVIEW OF CHAPTER 2 24
Key findings *24*
AIMS AND METHODS OF THE LITERATURE REVIEW 26
Aims *26*
Methods used to identify literature *26*
DEFINING COPD 27
Terminology *27*
Clinical and epidemiological definitions of COPD *28*
ICD codes for COPD and asthma *31*
Relationship between asthma, COPD and obstructive lung disease (OLD) *31*
RISK FACTORS FOR COPD 33
Lung function and natural history of COPD *33*
Genetic factors *35*
Smoking *36*
Occupational exposures *41*
Air pollution *43*
Diet *45*
Climate *46*
Early life factors *47*
Asthma and airways hyperresponsiveness *48*
Socio-economic status *50*
Urban environment *51*
Gender *51*
Risk factors with acute effects on COPD mortality *52*
CHOICE OF METHODOLOGY, VARIABLES AND DATA 52
Methodology *52*
Variables and data *58*

CHAPTER 3. GENERAL DATA QUALITY AND METHODOLOGY ISSUES..65

OVERVIEW OF CHAPTER 3 65
Key points 65
MORTALITY DATA 67
Coding issues 67
Use of COPD mortality as an outcome measure 69
AIR POLLUTION DATA 70
TAYLOR NELSON SOFRES (TNS) DATA ON FRUIT & VEGETABLE PURCHASES 72
Exploring the validity of a fruit & vegetable purchase variable from the TNS Super Panel database 73
Comparisons of the TNS fruit & vegetable data with other national dietary survey data 76
Is the TNS valid for use in a geographical analysis? 81
SOCIO-ECONOMIC STATUS OF AREA 82
INTRODUCTION TO BAYESIAN INFERENCE 83

SECTION II 87

CHAPTER 4. TEMPORAL VARIATIONS IN COPD MORTALITY IN ENGLAND & WALES IN RELATION TO THE 1956 CLEAN AIR ACT: METHODS 88

OVERVIEW OF CHAPTER 4 88
DATA 90
Mortality and population data 90
Air pollution data 96
DESCRIPTIVE ANALYSES 100
ANALYTICAL ANALYSES 100
Age-period-cohort analyses of COPD and lung cancer mortality 100

CHAPTER 5. TEMPORAL VARIATIONS IN COPD MORTALITY IN ENGLAND & WALES IN RELATION TO THE 1956 CLEAN AIR ACT: RESULTS 109

OVERVIEW OF CHAPTER 5 109
TRENDS IN AIR POLLUTION 112
Air pollution levels prior to 1950 affecting older cohorts 112
The 1956 Clean Air Act 114
National smoke and SO_2 air pollution from 1950 onwards 115
Regional levels of air pollution 116
Air pollution episodes identified from the literature 120
Trends from the 13 monitoring stations recording from the 1950s to at least the 1980s 120
DESCRIPTIVE ANALYSES OF MORTALITY DATA 129
Numbers of deaths and population 129
Age-stratified rates by area 129
Age-standardised rates by area 136
BAYESIAN AGE-PERIOD-COHORT (BAMP) ANALYSES BY AREA 143
Period effects for COPD mortality in relation to hypotheses, artefacts and other influences 144
Period effects for lung cancer 146
Cohort effects for COPD and lung cancer mortality 146
CLASSICAL AGE-PERIOD-COHORT ANALYSES 158
Quantifying relative risks of period parameters relative to non-conurbations 160
COMPARISON OF AIR POLLUTION TRENDS WITH MORTALITY TRENDS 167

SECTION III **170**

CHAPTER 6. THE ROLE OF NON-SMOKING FACTORS IN SPATIAL VARIATIONS IN COPD MORTALITY IN GREAT BRITAIN 1981-1999: METHODS **171**

OVERVIEW OF CHAPTER 6 171
Key points *171*
DATA 173
Geographical resolution *173*
Mortality datasets 1981-1999 ..... *174*
Population datasets 1981-1999 ..... *174*
Deprivation index for 1991 ..... *175*
Air pollution datasets: PM_{10} and SO_2 concentrations 1996 ..... *179*
Meteorological data: average minimum temperature and rainfall for 1981-1999 ..... *182*
Fruit and vegetable purchases using TNS 1991-2000 ..... *183*
DESCRIPTIVE ANALYSES 184
SHARED COMPONENT ANALYSES 184
LINEAR REGRESSION 188

CHAPTER 7. THE ROLE OF NON-SMOKING FACTORS IN SPATIAL VARIATIONS IN COPD MORTALITY IN GREAT BRITAIN 1981-1999: RESULTS **193**

OVERVIEW OF CHAPTER 7 193
DESCRIPTIVE ANALYSES 196
COPD and lung cancer mortality *196*
Deprivation *201*
Ambient air pollution *201*
Meteorological variables – minimum temperature and rain *202*
Pneumoconiosis and asbestos-related disease mortality *203*
Average district fruit & vegetable purchases from the TNS Super Panel 1991-2000 ..... *206*
SHARED-COMPONENT ANALYSES FOR COPD AND LUNG CANCER MORTALITY 207
Proportion of variances that are explained by shared and disease-specific components and their spatial structuring *208*
Residual association of log independent COPD risk (COPDspec) from shared component analyses with lung cancer SMRs *209*
Spatial pattern of smoothed relative risks for COPD and lung cancer mortality 209
Independent COPD risks from shared component analyses *210*
Shared (COPD and lung cancer) risks from shared component analyses *215*
Independent lung cancer risks from shared component analyses *218*
LINEAR REGRESSION 221
Descriptive plots of the association of explanatory variables with COPDspec ... *221*
Spearman rank correlations of explanatory variables *224*
Univariate linear regression *224*
Multiple linear regression *227*
Secondary analyses *230*

SECTION IV **238**

CHAPTER 8. DISCUSSION **239**

OVERVIEW AND SUMMARY OF CHAPTER 8 239
METHODOLOGICAL ISSUES COMMON TO BOTH TEMPORAL AND SPATIAL ANALYSES .. 242

Relevance of COPD mortality as an outcome 242
Diagnostic labelling and coding variations in mortality 243
Using lung cancer mortality as a surrogate for cumulative smoking 245
Bias and confounding – general issues 248
TEMPORAL ANALYSES 250
Consistency of results with other age-period-cohort analyses 250
Agreement of results with original hypotheses 251
Can changes in air pollution explain observed patterns in COPD mortality trends and period effects? 253
Can other factors can explain the observed mortality trends and period effects in COPD mortality? 267
Other results from temporal analyses – cohort effects for COPD and lung cancer, period effects for lung cancer 277
Methodological issues in temporal analyses 281
Bias in temporal analyses 289
Conclusions from temporal analyses 290
SPATIAL ANALYSES 291
Findings in relation to original hypothesis 291
Methodological issues 292
Findings of shared component analyses 296
Linear regression associations of exposure variables with independent COPD risks 299
Implications 313
Conclusions from spatial analyses 316

CHAPTER 9. CONCLUSIONS 317

OVERVIEW OF CHAPTER 9 317
CONDUCTING POPULATION LEVEL TEMPORAL AND SPATIAL EPIDEMIOLOGICAL STUDIES 317
CONCLUSIONS FROM TEMPORAL ANALYSES 318
CONCLUSIONS FROM SPATIAL ANALYSES 319
IMPORTANCE OF CONSIDERING BOTH SPATIAL AND TEMPORAL VARIATIONS IN COPD MORTALITY 320

REFERENCE LIST 323

APPENDIX 345

LIST OF TABLES 346
LIST OF FIGURES 347

List of Tables

Table 1-1 UK death rate per 100,000 from all causes and from various tobacco-related diseases in 1999 20

Table 2-1 Definitions of COPD and indications of severity 30

Table 2-2 ICD codes used for obstructive lung diseases in selected mortality analyses 31

Table 2-3 Classical picture of differences between asthma and COPD 32

Table 2-4 Numbers of deaths attributed to smoking (% of all deaths) in the UK, indirectly estimated from vital statistics 36

Table 2-5 Selected recent UK cohorts with information on COPD 55

Table 3-1 Effect of change in interpretation of WHO Rule 3, by age, 1984 and of discontinuation medical enquires 69

Table 3-2 Descriptive statistics of TNS 1991-2000 dataset by household 74

Table 3-3 Fruit and vegetables brought into the house for consumption in grammes per person per week using TNS data 1991-2000 with differing exclusion criteria compared with National Food Surveys for 1992, 1997 and 2000 78

Table 3-4 Number of households recording fruit & vegetables brought into the house for consumption in the TNS super panel 1991-2000 and in the National Food Surveys for 1992, 1997 and 2000 78

Table 3-5 Average regional purchases of fruit and vegetables by region for the TNS 1991-2000 and NFS 1997, ranked with highest region first 79

Table 4-1 Codes used to define COPD and lung cancer in time trend analyses 94

Table 4-2 Codes used to define influenza in time trend analyses 95

Table 4-3 Classification of UK air pollution monitoring stations as supplied by AEAT 99

Table 4-4 Numbers of cohorts with no deaths and therefore removed from classical Poisson age-period-cohort analyses (total number of cohorts = 110) ... 107

Table 5-1 Sulphur dioxide concentrations in English towns and cities published in 1895 112

Table 5-2 Sulphur content of coals, figures published in 1918 114

Table 5-3 Levels of SO_2 and black smoke in conurbations and Greater London areas relative to measured levels in non-conurbations for 1958/9 and 1970/71, using smoothed sulphur dioxide and black smoke levels (in $\mu g/m^3$) from Warren Spring data for regions and derived area averages or actual levels 119

Table 5-4 Notable episodes of high pollution in London from 1950 onwards 120

Table 5-5 Descriptive data from 13 monitoring sites with long-term readings (restricted to years with >330 days of data available) 122

Table 5-6 Numbers of COPD and lung cancer deaths and population for ages 15 years and older in different areas of England & Wales 1950-99 129

Table 5-7 Percentage difference between age-standardised (to 1999 population) rates in conurbations or Greater London and rates in non-conurbations averaged over different time periods 136

Table 5-8 BAMP analyses: number of second order differences for period parameters (n=48) and cohort parameters (n=108) where 5-95%, 10-90%, 15-85% or 25-75% credible intervals around parameters do not cross zero. RW2 analyses and RW1 analyses 154

Table 5-9 Expected timing of chronic air pollution effects, air pollution episodes, artefacts, and influenza peaks in England & Wales (flu) related to evidence for period effects for COPD mortality in those years using 2nd order differences from BAMP RW2 analyses. 155

Table 5-10 Timing of (all) period effects for lung cancer in conurbations (C), Greater London (GL) and non-conurbations (non-C) in males (m) and females(f) using BAMP analyses 156

Table 5-11 Timing of (all) cohort effects for COPD and Lung Cancer using 2nd order differences from BAMP analyses 157

Table 5-12 Poisson regression model results from maximum likelihood analyses 159

Table 5-13 Absolute changes in average relative risk of COPD period effects for conurbations vs. non-conurbations and Greater London vs. non-conurbations from classical age-period-cohort models assuming shared age and cohort but varying period effects in the different areas (ApC model), assuming a linear trend 166

Table 5-14 Absolute changes in average relative risk of lung cancer cohort effects for conurbations vs. non-conurbations and Greater London vs. non-conurbations from classical age-period-cohort models assuming shared age and period but varying cohort effects (APc model) 166

Table 5-15 Comparison of relative risk for period effects for COPD mortality for conurbations and Greater London for 1958 and 1970 relative to non-conurbations with black smoke and SO_2 concentrations 169

Table 6-1 (Product moment) correlation and Spearman rank correlation coefficients for black smoke and SO2 data between 1981 & 1996 readings and 1996 & 1999 readings (n = number of stations) 180

Table 6-2 Groupings of exposure variables used in linear regression analyses ... 189

Table 6-3 Excluded districts from linear regression and reason for exclusion 191

Table 7-1 Summed counts for COPD and lung cancer deaths and population for ages 45+ years by district for Great Britain 1981-1999 196

Table 7-2 Descriptive statistics for COPD and lung cancer SMRs 197

Table 7-3 Relationship of district level deprivation with COPD SMRs for 1981-1999 in Great Britain 201

Table 7-4 PM_{10} and SO_2 ranges per quintile 202

Table 7-5 Descriptive statistics for average daily minimum district temperature and district rainfall 1981-1999 203

Table 7-6 Counts for pneumoconiosis and asbestos-related disease deaths and population for ages 45+ years by district for Great Britain 1981-99 ... 204

Table 7-7 Extreme (>1000) district-level SMRs for pneumoconioses and asbestos-related deaths 205

Table 7-8 Number of districts in SMR groupings for pneumoconioses and asbestos-related disease mortality 1981-99 in ages 45+ years 206

Table 7-9 Descriptive statistics for district-level TNS dataset variables 207

Table 7-10 Proportion of variance explained by shared component or spatially structured shared or disease-specific component: mean and 95% credible intervals from 100,000 samples 208

Table 7-11 Descriptive statistics for SMRs, smoothed relative risks and independent and shared risks from shared component analyses for COPD and lung cancer mortality 210

Table 7-12 Results of single variable linear regression of the natural log independent COPD risk (COPDspec) against explanatory variables .. 226

Table 7-13 Results of final model for multiple linear regression of ln independent COPD risk (COPDspec) against explanatory variables 228

Table 7-14 Percentage change in independent COPD risk in the final model and in the final model excluding deprivation 231

List of Figures

Figure 4-1 Map of England & Wales showing location of Metropolitan Counties and Greater London (1991 boundaries) 97

Figure 4-2 Possible shapes of age, period or cohort parameter curves from age-period-cohort analyses (label given to change point identified) 104

Figure 5-1 Smoke concentrations and number of days with fog per year in London 113

Figure 5-2 Average urban concentrations of smoke and emissions from coal combustion from industrial and domestic sources 1960-1985 115

Figure 5-3 Regional trends in annual average black smoke and sulphur dioxide concentrations 1958-1971 in England & Wales 118

Figure 5-4 Black smoke readings from long-running monitoring sites in conurbation areas 123

Figure 5-5 Sulphur dioxide readings from long-running monitoring sites in conurbation areas 126

Figure 5-6 COPD mortality rates 1950-1999 for each area in England & Wales by age-group 132

Figure 5-7 Lung cancer mortality rates 1950-1999 for each area in England & Wales by age-group 134

Figure 5-8 Directly age-standardised (to 1999 England & Wales population) mortality rates for COPD ages 15+ years in conurbations excluding Greater London, Greater London and non-conurbation areas and influenza in England & Wales for 1950-99 139

Figure 5-9 Directly age-standardised (to 1999 England & Wales population) mortality rates for lung cancer in ages 15+ in conurbations (excluding Greater London), Greater London and non-conurbation areas for 1950-99 140

Figure 5-10 Ratio of COPD: lung cancer (age-standardised) mortality rates in conurbations (excluding Greater London), Greater London and non-conurbations areas for 1950-99 141

Figure 5-11 Mortality rates for influenza by age for ages 45+ years in England & Wales 1950-99 142

Figure 5-12 BAMP COPD age -period-cohort RW1 analyses 148

Figure 5-13 BAMP lung cancer age-period-cohort RW1 analyses 151

Figure 5-14 Relative risk (95% confidence intervals) for period effects for COPD in conurbations and Greater London relative to non-conurbations, using a classical age-period-cohort quasi-likelihood model with interaction terms assuming age and cohort effects are shared across areas (males top two graphs, females below) 162

Figure 5-15 Relative risk (95% confidence intervals) for cohort effects for COPD relative to non-conurbations, using a classical age-period-cohort quasi-likelihood model with interaction terms assuming age and period effects are shared across areas (males top two graphs, females below, first 5 and last 20 cohorts omitted) 162

Figure 5-16 Relative risk for period effect for COPD for conurbations and Greater London compared with non-conurbations, using a classical age-period-cohort model with interaction terms assuming age and cohort effects are shared across areas. Upper graph males, lower graph females. 163

Figure 5-17 Relative differences between age-standardised rates for COPD mortality for conurbations (excluding Greater London) and Greater London compared with non-conurbations, 1950-1999. Upper graph males, lower graph females 164

Figure 5-18 Relative risk for cohort effects for lung cancer for conurbations vs. non-conurbations and Greater London vs. non-conurbations, using a classical age-period-cohort model with interaction terms assuming age and period effects are shared across areas. Graphs have been smoothed using a five cohort moving average. Upper graph males, lower graph females. 165

Figure 6-1 Wards with COPD SMRs statistically significantly higher than 100 by Carstairs quintile (5 = most deprived) 177

Figure 6-2 Districts with COPD SMRs statistically significantly higher than 100 by deprivation index quintile (5 = most population living in deprived areas) 177

Figure 6-3 Wards with COPD SMRs statistically significantly lower than 100 by Carstairs quintile (5 = most deprived) 178

Figure 6-4 Districts with COPD SMRs statistically significantly lower than 100 by deprivation index quintile (5 = most population living in deprived areas).... 178

Figure 6-5 Scatterplots of annual mean black smoke in $\mu g/m^3$ and annual mean SO_2 in ppb from continuous-running monitoring stations for 1981 & 1996 and 1996 & 1999 181

Figure 6-6 Scatterplot of rankings for black smoke and for SO_2 in 1996 for 59 stations with no missing values 1981-1999 182

Figure 7-1 Scatterplots for COPD and lung cancer SMRs for districts in Great Britain 1981-1999 198

Figure 7-2 Map of COPD SMRs for UK districts 1981-1999 199

Figure 7-3 Map of Lung cancer SMRs for UK districts 1981-1999 200

Figure 7-4 Plots of log COPD SMRs against log lung cancer SMRs (above) and of COPDspec, the log independent COPD risk, against log lung cancer SMRs (below – note different x and y scales) for males and females 209

Figure 7-5 Map of posterior mean district independent COPD risks for males (left), map of the probability that posterior mean risk is greater than 1 (centre), map of the probability that posterior mean risk is less than 1 (right) 213

Figure 7-6 Map of posterior mean district independent COPD risks for females (left), map of the probability that posterior mean risk is greater than 1 (centre), map of the probability that posterior mean risk is less than 1 (right) 214

Figure 7-7 Map of posterior mean district shared COPD and lung cancer risks for males (left), map of the probability that posterior mean risk is greater than 1 (centre), map of the probability that posterior mean risk is less than 1 (right) 216

Figure 7-8 Map of posterior mean district shared COPD and lung cancer risks for females (left), map of the probability that posterior mean risk is greater than 1 (centre), map of the probability that posterior mean risk is less than 1 (right) 217

Figure 7-9 Map of posterior mean district independent lung cancer risks for males (left), map of the probability that posterior mean risk is greater than 1 (centre), map of the probability that posterior mean risk is less than 1 (right) 219

Figure 7-10 Map of posterior mean district independent lung cancer risks for females (left), map of the probability that posterior mean risk is greater than 1 (centre), map of the probability that posterior mean risk is less than 1 (right) 220

Figure 7-11 Box (showing interquartile range) and whisker plots* of the association of explanatory variables with the ln independent COPD risk 222

Figure 7-12 Percentage change in relative risk for independent COPD risk factor for quintiles of minimum temperature and rainfall compared with first quintile (coldest and driest respectively) 229

Figure 7-13 Plot of residuals against fitted values for the final linear regression models 230

Figure 7-14 Percentage change in independent COPD risk per quintile increase in ambient SO_2 concentration by deprivation quintile 233

Figure 7-15 Percentage change in independent COPD risk with increasing average minimum temperature (quintiles compared with quintile 1 – lowest temperature) for males by deprivation quintile 234

Figure 7-16 Percentage change in independent COPD risk with asbestos-related disease SMRs <100 or >100 (compared with SMRs non-significantly different from 100) for males by deprivation quintile 235

Figure 7-17 Percentage change in independent COPD risk with increasing rainfall by temperature quintile – males 236

Figure 8-1 Tobacco cigarette consumption (tonnes weight) in Great Britain 1905-1987 and number of deaths from lung cancer and COPD 1950-1999 in England & Wales for ages 15+ years 258

Figure 8-2 Map highlighting 22 counties in England & Wales where most (96%) of deaths in miners occurred 1979-80 and 1982-90 307

List of abbreviations

AEAT	AEA Technology Environment (a commercial company responsible for maintaining the UK Air Quality Archive)
ATS	American Thoracic Society
BTS	British Thoracic Society
CAO	Chronic airflow obstruction
COLD	Chronic obstructive lung disease
COPD	Chronic obstructive pulmonary disease
CT	Computerised tomography
DETR	Department of the Environment, Transport and the Regions
ECRHS	European Community Respiratory Health Survey
ERS	European Respiratory Society
ETS	Environmental tobacco smoke
FEV_1	Forced expiratory volume in one second
$FEF_{(25-75)}$	Forced expiratory flow between 25% and 75% of FVC
FVC	Forced vital capacity
GOLD	Global Obstructive Lung Disease Initiative
GP	General Practice
GPRD	General Practice Research Database
HSE	Health Survey for England
HSE95	1995 Health Survey for England
ICD	International Classification of Disease
ICD9	International Classification of Disease version 9
IMD	Index of Multiple Deprivation
MARS	European Union Monitoring Agriculture with Remote Sensing Unit
MRC	Medical Research Council
NETCEN	AEA Technology's National Environmental Technology Centre
NHANES III	Third National Health and Nutrition Examination Survey
NICE	National Institute of Clinical Excellence
NDNS	National Diet and Nutrition Survey
NFS	National Food Survey
OLD	Obstructive Lung Disease
ONS	Office for National Statistics
OPCS	Office of Population Censuses and Surveys
PEF	Peak expiratory flow
PFR	Pneumoconiosis Field Research
PMR	Proportional Mortality Ratio
RR	relative risk
SAHSU	Small Area Health Statistics Unit
SMR	Standardised Mortality Ratio
TAC	Tobacco Advisory Council
TNS	Taylor Nelson Sofres
TSP	Total Suspended Particulates
UK	United Kingdom
USA	United States of America

The trite observation that familiarity breeds contempt is essentially true with regard to the outlook upon chronic bronchitis; those afflicted are inclined to accept the complaint as inevitable, as something troublesome but not serious. Those called upon to treat it do not find it sufficiently interesting to study closely....Yet bronchitis stands high not only in the list of causes of death, but also among causes of invalidity.

> *Edgar Collis. The general and occupational prevalence of bronchitis and its relation to other respiratory diseases. Journal of Industrial Hygiene 1923;5:264-276*

...cigarette smoking is now the most important cause of chronic obstructive lung disease, with air pollution and occupational exposure to dust fumes also making a contribution, especially in cigarette smokers. The effect of geographical differences in climate is uncertain...That other unidentified aetiological factors are also concerned is shown by the wide prevalence of, and high mortality from, what was diagnosed as 'chronic bronchitis' in Britain long before cigarette smoking became a common habit (Collis 1923) and by the largely unexplained social-class gradient of this, which ... is not just due to differences in smoking habits.

> *Charles Fletcher, Richard Peto, Cecily Tinker, Frank E Speizer. The Natural History of Chronic Bronchitis and Emphysema. Oxford University Press, Oxford, 1976*

Section I

The epidemiology of chronic obstructive pulmonary disease in the UK: spatial and temporal variations

Chapter 1 Introduction

Chapter 2 Literature search

Chapter 3 General data quality and methodology issues

Chapter 1. Introduction

Overview of Chapter 1

Chapter 1 introduces the main themes of this PhD. The first paragraphs relate to the public health importance of COPD in an international context and are followed by facts and figures relating to COPD prevalence, morbidity and mortality in the UK. The hypotheses to be investigated in this PhD are then set out. Finally, there is a short description of the way in which the PhD has been structured.

Introduction

Chronic obstructive pulmonary disease (COPD) is an important medical problem worldwide, responsible for a total of 2.7 million deaths worldwide in 2002.[1] The World Health Report 2002[1] placed COPD as the third leading cause of death and disability adjusted life years (DALYs) in adults aged 60 years and over (after ischaemic heart disease and cerebrovascular disease) and the 10th leading cause of mortality in adults aged 15-59 years. The number of deaths from COPD is predicted to double worldwide over the next 20 years,[2] following trends in tobacco-smoking, the major risk factor.

The 'English disease'

Even prior to widespread uptake of smoking in the late 19th and early 20th century, bronchitis was recognised to have a high prevalence in Great Britain, such that it sometimes referred to as the 'English disease'.[3] Higher rates of mortality in Britain compared with other countries have been recognised for many years[4,5] and are not solely due to diagnostic variations.[3,5] The most recent study of international mortality trends to include "chronic airways obstruction" (International Classification of Disease version 9 or ICD9 code 496) as well as bronchitis and emphysema relates to1984.[6] This showed that the highest mortality from COPD (ICD9 490-496) was in Great Britain, Eastern Europe and Australasia, with intermediate rates in Western Europe and low rates in Southern Europe. Only Romania and the German Democratic Republic (Eastern Germany) had higher rates in men and only Romania and New Zealand had higher rates in women. However, a more recent comparison with the United States suggests that the excess 'English Disease' may have disappeared by the 1990s.[4]

Current importance of COPD morbidity and mortality in the UK

COPD exerts a considerable cost in terms of health impairment and costs related to healthcare use. An epidemiological review of lower respiratory tract disease published in 1994[7] suggested that approximately 5% of men and 3% of women of middle age (age range not defined) in England have impaired respiratory function together with a productive cough. This is consistent with recently published data for the UK from the European Community Respiratory Health Survey (ECRHS),[8] which found a prevalence of mild COPD of 2.3% and moderate COPD of 1% in adults aged 20-44 years[8] in East Anglia, an area with low COPD Standardised Mortality Ratios (SMRs). A review in 2003[9] only found two UK studies relating to prevalence in older individuals, both relating to single primary care practices,[10-12] one using a clinical diagnosis of COPD[10,11] and the other using spirometry.[12] In the latter study,[12] 9.9% of adults aged 60-75 years had COPD defined as an FEV_1 lower than the fifth centile and a bronchodilator reversibility test.

The annual UK prevalence rates for COPD consultations in primary care in the 1990s were between 1-2%,[13] with a gradual increase occurring over 1990-1997 in both sexes. COPD and asthma accounted for approximately 175,000 hospital admissions in 1999/2000 in England,[14] a rise from 160,000 per year in 1991-4.[15] The 1999/2000 figures correspond to 1.6% of all admissions (1 in 8 of all medical admissions[16]) and 2.6% of bed-days.[14] Most of these admissions occurred in the elderly (>65 years).

Mortality from obstructive lung disease (ICD9 490-492, 496) was the underlying cause of death for just under 5% of all deaths in 1993-1999.[17] The UK mortality rate from obstructive lung disease in 1999 was 67.8 per 100,000 for males and 54.4 per 100,000 for females (Table 1-1, page 20). Mortality from COPD and bronchitis has shown consistent geographical variation over the last 50 years, with highest SMRs in the industrialised areas of north-west England (West Midlands, Mersey, Manchester, South Yorkshire), South Wales and central Scotland and lowest SMRs in the rural areas of East Anglia, South-West England and mid Wales.[18-22]

The costs to the NHS in England of COPD diagnosis and management have been variably estimated at approximately £500 million per year in 1998[24] and £846 million, or £1,154 per patient per year, in 1996.[25] Pharmaceutical costs for COPD and allied

conditions were estimated as 11% of the total expenditures for prescription medications in 1996.[25] Indirect costs were estimated at £600 million for attendance and disability living allowance and £1.5 billion to employers for work absence and related productivity.[24]

Table 1-1 UK death rate per 100,000 from all causes and from various tobacco-related diseases in 1999

	All causes	Ischaemic heart disease (ICD9 410-414)	Stroke (ICD9 430-438)	All cancers (ICD9 140-239)	Respiratory cancer (ICD9 160-163, 212)	Chronic obstructive lung disease (ICD9 490-496)
Males	1155.9	278.5	91.7	307.3	86.3	67.8
Females	1242.1	223.4	152.4	280.7	50.4	54.4

Source: Office for National Statistics[23]

Hypotheses to be addressed by this PhD

Smoking is the acknowledged major risk factor for COPD, but COPD mortality trends in England & Wales over the 20th century were not consistent with smoking patterns[26,27] unlike those for lung cancer mortality.[28] Also, smoking patterns do not fully account for geographical variations in COPD. Over the last 50 years, smoking variations have not fully explained the higher COPD rates in UK cities[5] nor the higher rates in more northern versus more southern areas of England.[21] Urban-rural differences in bronchitis deaths were apparent even before the start of the smoking epidemic: crude death rates in 1891-1900 were approximately a third higher in urban than rural counties in both men and women.[29]

A number of explanations have been suggested to explain these discordances between COPD and smoking time trends and geographical variations, but relatively few studies have been conducted. This PhD attempts to explore non-smoking related aetiological factors in COPD in Great Britain, using COPD mortality as a measure of COPD burden and lung cancer as a proxy measure for smoking.

Hypotheses relating to time trends in COPD mortality

Air pollution has been suggested as a potential important influence on COPD by a number of authors.[5,20,21,26] Fogs were also considered an important factor in COPD mortality and morbidity in earlier years,[30-34] probably because they were closely

associated with high air pollution levels.[32] There have been marked changes in UK air pollution levels over the last 100 years, particularly particulates and sulphur dioxide. A general improvement in black smoke and sulphur dioxide levels started around 1900.[35,36] The Clean Air Act of 1956 was followed by a more marked improvement in air pollution in the late 1950s and early 1960s that showed some levelling off in the mid 1970s.[37] By the 1990s, total emissions of black smoke had reduced to about one fifth of their level in 1950.[26]

Marks and Burney in their review of COPD mortality in the England & Wales in the 20th Century[26] postulated that these changes in outdoor air pollution could potentially explain some of the observed geographical and urban variations and longer term time trends. Time trends in COPD and lung cancer mortality in England & Wales show strong age, period and cohort effects.[26,27] Period effects relate to factors common to all age-groups at a particular time period (which would include changes in air pollution levels and treatment advances), while cohort effects relate to factors more common in people born at a particular point in time, such as lifetime patterns in smoking. It is possible that air pollution exerts both short-term (period) effects and longer-term (period and cohort) effects on COPD. Short-term changes in particulate air pollution have been associated with short-term effects on respiratory mortality,[38,39] while long term means of community particulate and sulphate air pollution have been associated with bronchitis and bronchitic symptoms[40] and respiratory mortality.[41]
If air pollution were an important influence in COPD mortality in the UK, then population-level trends in COPD mortality would be expected to reflect air pollution changes following the 1956 Clean Air Act, particularly in more urban areas where levels were originally highest and where the greatest declines were seen.[37] Lung cancer trends, on the other hand, would be expected to largely reflect trends in smoking, as the attributable risk from air pollution is small.[42] These expectations were used to formulate the following *a priori* hypotheses with respect to temporal variations in COPD mortality:

1. ***Overall rates of COPD mortality would be higher in more urban than more rural areas from 1950 onwards. Alternatively, COPD rates would be higher in more urban areas until the late 1950s after which they would converge towards those in more rural areas.***

2. ***Declines in COPD mortality rates would been seen in both urban and rural areas starting from the late 1950s and this would be located as a period effect in age-period-cohort analyses. Lung cancer would not show the same period effects.***
3. ***Age effects in an age-period-cohort analyses of rates 1950 onwards would be similar in all areas.***
4. ***Cohort effects in age-period-cohort analyses would be similar in COPD and lung cancer reflecting lifetime smoking patterns. There might be an additional COPD cohort effect in those born 1935-60, related to progressive reductions in air pollution exposure in childhood.***

Hypotheses relating to spatial variations in COPD mortality

Aside from differences in smoking prevalence and air pollution, a number of risk factors have been suggested as responsible for spatial variations in COPD mortality in the UK[21] including temperature, diet and some unidentified aspect of urbanisation. Differing distribution of industries such as coal mining[43] may also be important. A major difficulty in such analyses is that data on smoking, the major determinant of COPD, are generally not available at suitable geographical resolution and any available data usually provide measures of smoking prevalence when the variable of most interest is cumulative smoking exposure.[44,45] One way of estimating of cumulative smoking is to assume lung cancer mortality is a suitable proxy.[46] This led to the following hypothesis with respect to spatial variations in COPD mortality:

5. ***The spatial distribution of areas where patterns of lung cancer and COPD mortality are discordant reflects the distribution of risk factors other than smoking that influence COPD development and/or mortality .***

Structure of this PhD document

To facilitate the understanding of the work done in the course of this PhD, this document has been laid out in four sections:
Section I: Introduction (chapter 1), literature search (chapter 2) and general data quality and methodology issues (chapter 3)

Section II: Temporal variations in COPD mortality in England & Wales in relation to the 1956 Clean Air Act: Methods (chapter 4) and Results (chapter 5)
Section III: The role of non-smoking factors in spatial variations in COPD mortality in Great Britain 1981-1999: Method (chapter 6) and Results (chapter 7)
Section IV: Discussion (chapter 8) and Conclusions (chapter 9)

Section II relates to the investigation of hypotheses 1 to 4 – mortality trends in relation to trends in air pollution. Section III relates to the investigation of hypothesis 5, that different patterns of non-smoking risk factors will explain discordant patterns of COPD and lung cancer mortality.

Each chapter starts with an overview and summary of the key points or findings.

Chapter 2. Literature review

Overview of Chapter 2

This chapter contains findings from a literature review used to refine the research questions and to identify appropriate risk factors, datasets and methodology to investigate the main hypotheses.

Key findings

- The term COPD is a broad diagnostic label, which may include a number of conditions causing airflow limitation such as chronic bronchitis, emphysema and asthma. Terms are often used interchangeably by doctors, but chronic bronchitis or emphysema may be used more in men and smokers, and asthma in women.
- Tobacco smoking is the major risk factor for COPD with cumulative effects on lung function at all stages of life. 70-90% COPD deaths can be attributed to smoking depending on age and time period. Components of tobacco smoke causing COPD have not been clearly established – unlike the association of tar with lung cancer.
- Factors other than smoking implicated in COPD development are
    - Genetic factors of which the only currently established factor is α_1-antitrypsin deficiency accounting for <1% of cases
    - Occupational exposures to dusts, gases and chemical fumes. Coal mining has been best studied, permitting estimates of the dose-response relationship between dust exposure and COPD development.
    - Ambient air pollution, especially particulates and SO_2.
    - Diet, particularly fruit and vegetable or fish consumption with associations for fruit and vegetables more consistently seen than those for fish.
    - Climate. Cold and damp were considered important in the 1950s. While little investigated since then, two analyses from the 1990s[21,47] supported a role for climate.
    - Early life factors, including low birthweight, pneumonia and possibly bronchitis in early life.
    - Long-standing asthma and airways hyperresponsiveness.

- Deprivation, the effect of which persists even after adjustment for higher smoking rates.
- Urban environment.
- Gender. COPD mortality is higher in men, which may relate to differential diagnostic labelling as well as lower exposure to risk factors in women.

- Population-level methodology was chosen to investigate the PhD hypotheses because:
  - Population-effects were being studied
  - Improvements in age-period-cohort and spatial methodology permitted clearer and easier identification of trends
  - A wide range of outcome and risk factor routine data were available
  - Setting up and analysing a new individual-level study on the scale required was beyond the scope of this PhD
  - Limitations in existing British cohort studies in terms of age (too young), size (small numbers of COPD outcomes), breadth of risk factor information or generalisability (not population-based) made analysis of existing individual-level data unrealistic to address the PhD hypotheses
- COPD mortality was chosen as the outcome for analyses, using an inclusive definition of COPD. Both temporal and spatial analyses used lung cancer mortality as a proxy for cumulative smoking and considered ambient black smoke and SO2. Analyses were conducted separately for males and females.
- Other risk factor variables chosen for the spatial analyses were
  - Mortality from pneumoconioses and from asbestos-related diseases as markers for areas with high concentrations of industries with exposure to dust and fumes
  - Data on fruit and vegetables purchases from a market research database
  - Temperature and rainfall
  - Area-level deprivation using Carstairs index

Aims and methods of the literature review

Aims

A review of the literature on COPD was undertaken with the following aims:

(i) To clarify the definition of COPD and its relation to asthma to inform the definition of COPD used in the analyses

(ii) To identify known and possible risk factors for COPD (development and mortality) to define explanatory variables for spatial analyses

(iii) To identify previous epidemiological analyses relevant to the UK to refine the research questions

(iv) To determine the appropriate variables, data and methodology to investigate the main hypotheses

Given the breadth and depth of the available literature, a systematic review was not undertaken, but an attempt was made to locate important reviews and key studies.

Methods used to identify literature

A number of literature searches were conducted using BIDS Embase from 1980 to January 2004 and Medline 1966 to January 2004. The search terms for COPD were 'Chronic obstructive lung disease' in Embase and 'Lung diseases, obstructive' in Medline. These include all synonyms for COPD; the Medline term also includes asthma. Initially review articles were located by using the subheadings of 'epidemiology' or 'etiology' and limited to review articles. More specific searches were then conducted in certain areas. For example, articles on geographical variations in COPD used the search strategy ('Lung diseases, obstructive' AND ('small-area analysis' OR 'geography'(exploded) OR 'geographical variations'(textword) OR 'cross-cultural comparison' OR 'residence characteristics')). The effect of smoking on COPD mortality was investigated using searches for COPD and 'smoking' AND 'mortality'. Specific searches were also used to locate publications on some cohorts by using alternative names of the cohort as a textword (e.g. 'Boyd Orr' and 'Carnegie' for the Boyd Orr cohort).

Further journal articles and books were identified through citations of the papers read, recommendations from colleagues and supervisors, through attendance at respiratory and epidemiological conferences and from my own files. Periodic hand searches were

conducted of issues of Thorax, BMJ, The Lancet, the European Respiratory Journal and the American Journal of Respiratory & Critical Care Medicine. A weekly email alert (Amadeo http://www.amadeo.com/) was used to identify new articles on obstructive lung diseases from selected respiratory journals.

Defining COPD

Terminology

A number of synonyms for obstructive lung disease are in current usage. The term 'chronic obstructive pulmonary disease' (COPD) was first introduced in the late 1950s to cover conditions previously described as chronic bronchitis or emphysema.[28] Currently, COPD is used internationally to include any or all of the following:[48]

- Emphysema
- Chronic bronchitis
- Chronic airflow limitation (CAL)
- Chronic airflow obstruction (CAO)
- Chronic airways obstruction (CAO)
- Non-reversible obstructive airways disease (NROAD)
- Chronic obstructive airways disease (COAD)
- Chronic obstructive lung disease (COLD)
- Some cases of chronic asthma

However, sometimes these terms also include asthma. For example, asthma is included in the group "chronic obstructive pulmonary disease and allied conditions" in ICD9 (International Classification of Disease, version 9) codes (ICD9 490-496).

Obstructive lung disease (OLD) diagnoses (asthma, chronic bronchitis, emphysema, COPD) may be used interchangeably, with around 20% of patients with an obstructive lung disease ending up with more than one OLD diagnosis: approximately 17% of people in a community survey in the USA and the 19% of primary care patients in the UK[49] had been given more than one OLD diagnosis and 22% of patients visiting an emergency room in Spain[50] had a different OLD diagnosis on their second visit.

Clinical and epidemiological definitions of COPD

Several clinical diagnostic definitions of COPD are available (Table 2-1, page 30) and they are generally lengthy. The British Thoracic Society (BTS) guidelines has several pages relating to diagnosis starting with the following:

> *'a chronic slowly progressive disorder characterised by airflow obstruction (reduced FEV_1 and FEV_1/FVC ratio that does not change markedly over several months). Most of the lung function impairment is fixed, although some reversibility can be produced by bronchodilator (or other) therapy. Thus a diagnosis of COPD in clinical practice requires:*
>
> - *a history of chronic progressive symptoms (cough and/or wheeze and/or breathlessness);*
> - *objective evidence of airways obstruction, ideally by spirometric testing, that does not return to normal with treatment.'* BTS 1997 [48]

The guidelines remark that a firm diagnosis can only be made by with spirometric confirmation of airway obstruction and that a cigarette smoking history of more than 20 pack-years is usually obtained (where one pack year is the equivalent of smoking 20 cigarettes per day for one year). They also state that COPD arises from varying combinations of airway disease and pulmonary emphysema, that differentiating severe COPD from chronic severe asthma may be difficult and that COPD does not conventionally include other specific conditions causing airways obstruction such as cystic fibrosis, bronchiectasis, or bronchiolitis obliterans.[48]

Definitions of COPD have also been produced by the European Respiratory Society (ERS),[51] American Thoracic Society (ATS),[52] by the Global Obstructive Lung Disease Initiative (GOLD)[53] and most recently included in national clinical management guidelines from the National Institute for Clinical Excellence in England[54] (Table 2-1, page 30). GOLD[55] argued against a rigid definition of COPD and its relationship with other obstructive airways diseases until there was a clearer understanding of causal mechanisms. Generally,[56] guidelines use the ratio of FEV_1 to FVC to diagnose COPD and percentage of predicted FEV_1 to measure disease severity (Table 2-1, page 30).

One of the most widely used epidemiological definition of COPD is the MRC definition of chronic bronchitis:

> '*chronic cough and sputum production for at least three months of two consecutive years in the absence of other diseases recognised to cause sputum production*' MRC 1965[57]

However, these symptoms do not necessarily signify the presence of airways obstruction or a diagnosis of COPD.[48] Epidemiological studies are increasingly using varying cut-offs in lung function tests, that can usually be related to one of the guidelines in current use (Table 2-1, page 30), for example below a certain percentage for FEV_1 or the ratio of FEV_1/FVC or those below two standard deviations of norms for age and height.[58]

Use of different definitions for COPD can lead to more than two-fold differences in assessments of the population prevalence of COPD.[59] GOLD stage II (see Table 2-1, page 30) was found to be closest to measures of self-reported prevalence of COPD in a survey of over 13,000 American individuals as part of the Third National Health and Nutrition Examination Survey (NHANES III) conducted in 1988-1994,[59] while other spirometric definitions gave higher prevalences. Further the differences between prevalences obtained from different definitions varied by age.[59] Variations in prevalence using different definitions were found to be greater when assessing mild than moderate or severe disease.[60] However, a clinical diagnosis of an obstructive lung disease will usually indicate moderate or severe disease, i.e. that which needs clinical attention.[58] A small study in Crete following 67 COPD patients over 18 months suggested that the GOLD and ERS severity scales correlated better with number of hospitalisations than the ATS or BTS measures.[61] Despite the emphasis on lung function measurements in COPD definitions, a substantial proportion of the population (9.3% of those aged over 50 years in the American NHANES III survey[49]) have objective airflow obstruction without any clinical respiratory diagnosis.

Table 2-1 Definitions of COPD and indications of severity

Guideline	Organisational Sponsor	Definition of COPD	Severity of COPD
NICE 2004[54]	National Collaborating Centre for Chronic Conditions, UK	*'A diagnosis of COPD should be considered in patients over the age of 35 who have a risk factor (generally smoking and who present with one or more of the following symptoms:* • *exertional breathlessness* • *chronic cough* • *regular sputum production* • *frequent winter "bronchitis"* • *wheeze* *... Spirometry is fundamental to making a diagnosis of COPD and a confident diagnosis of COPD can only be made with spirometry. A diagnosis of airflow obstruction can be made if the $FEV_1/FVC < 0.70$ (i.e. 70%) and FEV_1 <80% predicted.'*	*'A true assessment of severity should include assessment of the degree of airflow obstruction and disability, the frequency of exacerbations and the following prognostic factors:* • *FEV_1* o *Mild airflow obstruction* *FEV_1 50-80% predicted* o *Moderate airflow obstruction* *FEV_1 50-80% predicted* o *Sever airflow obstruction* *FEV_1 <30% predicted* • *T_LCO* • *breathlessness (MRC scale)* • *health status* • *exercise capacity* • *body mass index (BMI)* • *partial pressure of oxygen in arterial blood (PaO_2)* • *cor pulmonale'*
GOLD 2001[55,62]	National Heart, Lung and Blood Institute, USA World Health Organization	*'A diagnosis of COPD should be considered in any individual who presents characteristic symptoms and a history of exposure to risk factors for the disease, especially cigarette smoking... The diagnosis should be confirmed by spirometry'* where *'..all FEV_1 values refer to a postboronchodilator FEV_1'.* Stage III (previously called IIA) *'$FEV_1/FVC < 0.70$ and FEV_1 <80% predicted is often used as a threshold for identifying COPD'.*[59]	Stage 0: At risk. Chronic symptoms, exposure to risk factors, normal spirometry Stage 1: Mild. $FEV_1/FVC \geq 70\%$. $FEV_1 \geq 80\%$. With/without symptoms Stage II. Moderate. $FEV_1/FVC<70\%$, $50\% \leq FEV_1<80\%$, with/without symptoms Stage III: Severe. $FEV_1/FVC<70\%$, $30\% \leq FEV_1<50\%$, with/without symptoms Stage IV: Very severe. $FEV_1/FVC<70\%$, $FEV_1<30\%$ or $FEV_1<50\%$ predicted plus chronic respiratory failure
BTS 1997[48]	British Thoracic Society	*'a chronic slowly progressive disorder characterised by airflow obstruction (reduced FEV_1 and FEV_1/FVC ratio that does not change markedly over several months). Most of the lung function impairment is fixed, although some reversibility can be produced by bronchodilator (or other) therapy. Thus a diagnosis of COPD in clinical practice requires:* *a history of chronic progressive symptoms (cough and/or wheeze and/or breathlessness); objective evidence of airways obstruction, ideally by spirometric testing, that does not return to normal with treatment.'*	Mild FEV_1 60-80% predicted Moderate FEV_1 40-59% predicted Severe FEV_1 <40% predicted
ATS 1995[52]	American Thoracic Society	*'characterised by airflow obstruction caused by chronic bronchitis, emphysema, or both. Emphysema is defined as abnormal permanent enlargement of the air spaces distal to the terminal bronchioles, accompanied by destruction of their walls and without obvious fibrosis. Chronic bronchitis is defined as chronic cough, mucus production, or both, for at least three months for at least two successive years where other causes of chronic cough have been excluded.'*	Staged on the basis of airflow obstruction using the 1991 ATS statement on lung function testing[63] Stage I $FEV_1 \geq 50\%$ predicted Stage II FEV_1 35-49% predicted Stage III FEV_1 <35% predicted
ERS 1995[51]	European Respiratory Society	No succinct prose definition given in the original document,[51] but often interpreted[59] as: *'FEV_1 <88% predicted in men or <89% predicted in women (i.e. >1.64 residual standard deviation below predicted value)'*	Mild $FEV_1 \geq 70\%$ predicted Moderate FEV_1 50-69% predicted Severe FEV_1 <50% predicted

ICD codes for COPD and asthma

Different researchers have chosen different ICD codes and used varying terms when investigating obstructive lung diseases (Table 2-2). The full list of ICD codes for chronic obstructive lung diseases including asthma from ICD1 to ICD9 are listed in Table A-1 in the Appendix (page 348).

Table 2-2 ICD codes used for obstructive lung diseases in selected mortality analyses

Disease term Paper	ICD5 codes	ICD6 codes	ICD7 codes	ICD8 codes	ICD9 codes
COPD analyses					
COPD* Marks 1997 [26]	106b,106c,113	501-502,526	501-502,526,527.1	490-492,518,519.8	490-492,494,496
COLD Lee 1990[27]	114e,106b,106c	501-502,527.0,527.2	501-502,527.0,527.2	490-491,519	491,496,518-519
Emphysema Lee 1990[27]	113	527.1	527.1	492	492
COLD Doll 1997[28]		502,527	502,527	491-492,519	491-492,496,519
COPD Hospers 1999[64]				490-492,518	490-492,494,496
COPD Thomason 2000[65]				491-493,519.8	
Asthma analyses					
Asthma Marks 1997*[26]	112(1),112(6)	241	241	493	493.0,493.1,493.9

COPD = chronic obstructive pulmonary disease, COLD = chronic obstructive lung disease
*Codes from ICD1 to ICD4 not shown

Relationship between asthma, COPD and obstructive lung disease (OLD)

Asthma and COPD have been described as a complex of diseases, which have airflow limitation in common.[66] While there are some similarities between the two diseases in terms of epidemiology, risk factors and lung function, there are also marked differences in terms of markers of airway inflammation, lung volumes, diffusion capacity and airway morphology in-between exacerbations (Table 2-3, page 32).[67] The pathological differences between asthma and COPD in both airways morphology in-between exacerbations and loss of integrity of the lung parenchyma (seen mainly in COPD) are considered to offer the firmest evidence that asthma and COPD are different diseases.[67] It can be difficult to distinguish between the two conditions, particularly in later life[68] and it is possible that some individuals have both.[67]

Diagnostic labelling

Different doctors may differ in their use of COPD or asthma as a diagnostic label for the similar symptoms, as noted previously (page 27). Choice of clinical diagnosis may be influenced by the gender and age of the patient, with a COPD diagnosis more likely in

men and in the elderly and asthma more likely in women.[50,69] Studies of death certificates[70] suggest that asthma deaths may be recorded as due to COPD or cardiovascular causes in males, while COPD deaths may be recorded as due to asthma in females.

Table 2-3 Classical picture of differences between asthma and COPD

Characteristics	Asthma	COPD
Lung function		
Airflow limitation[48,52]	Reversible	Fixed
Bronchial hyperresponsiveness	Present	Less marked or absent
Lung volumes	Hyperinflation less common	Hyperinflation more common
Clinical characteristics		
Disease pattern	Episodic	Slowly progressive
Epidemiology		
Countries	Higher prevalence in affluent countries	Prevalent in affluent and developing countries
Smoking	Passive smoking may be important in children	Prevalence reflects past smoking prevalence
Age range most affected	Children and young adults	Late middle age onward
Risk factors		
Established environmental risk factors	Occupational sensitizers Allergen exposure	Cigarette smoking Some occupational exposures
Established host risk factors	Atopy Gender	α_1-antitrypsin deficiency
Likely risk factors	Small size at birth Respiratory infections Drugs and food additives	Low socio-economic status Environmental tobacco smoke Air pollution Bronchial hyperresponsiveness
Putative risk factors	Smoking Air pollution Diet	Diet Genetic predisposition
Airway morphology		
Structural changes	No emphysema	Emphysema
Surface epithelium	Fragility or loss	Metaplasia
Reticular basement membrane	Thickened and hyaline	Normal
Cellular infiltrate (in-between exacerbations)	Lymphocytes CD3+, CD4+, CD25+, IL-2R+ Marked eosinophilia	Lymphocytes CD3+, CD8+, CD68+, CD25+, VLA-1+, LA-DR+ Mild eosinophilia
Predominant cytokines	IL4 and IL-5 (Th-2 type in allergic and non-allergic asthma) IL16 (allergic asthma)	GM-CSF protein IL-4 sometimes present
Induced sputum (in-between exacerbations)	Higher percentage of eosoinophils than COPD	Higher percentage of neutrophils than asthma

Source: adapted from Magnussen 1998[67]

Asthma accounts for the largest proportion of OLD morbidity and was found to compose 79.4% of OLD in a large American community survey and 50.3% of patients in general practice in the UK.[49] However, the proportion of asthma in OLD prevalence decreases markedly with age, with current asthma forming only 42% of OLD diagnoses in UK patients aged ≥50 years.[49] Asthma forms a much smaller proportion of COPD

deaths, constituting 6.9% of OLD deaths in a study of deaths in England and Wales 1993-1999.[17]

Risk factors for COPD

As with all diseases, genotype, environmental exposures and gene-environment interactions are involved in the pathogenesis and progression of COPD. It is generally considered a complex polygenic disease.[71] More is currently known about environmental risk factors, of which cigarette smoking is by far the most important,[71] than genetic factors. Occupational exposures to dusts and gases (e.g. mining) are also established risk factors. Environmental tobacco smoke, long term exposure to air pollution, deprivation, diet, climate, bronchial hyperresponsiveness and reduced lung growth in fetal and early life have also been identified as possible risk factors and these are all discussed in detail below.

Lung function and natural history of COPD

A key feature of COPD is a progressive, irreversible decline in lung function,[58] usually measured using FEV_1. A consideration of the natural growth and decline in lung function is relevant to any discussion of risk factors, which may have differing impacts depending on the stage of life at which they act.

Natural growth and decline of FEV1

The natural course of FEV_1 is of growth in childhood and adolescence, a plateau phase in early adulthood,[72] followed by a decline, which accelerates with ageing.[73] Lung function growth generally slows from about age 20 and ceases in the mid[74] to late[75] 20s and slightly later in males than females.[74] The exact timing of the onset of plateau phase or decline in healthy young adults is disputed and lung function may continue to rise to age of 25 years or into the early 30s.[73]

FEV_1 level will therefore be determined by

(i) The maximally attained level of lung function in early adulthood

(ii) The duration of the plateau phase and age at onset of decline

(iii) The rate of decline of lung function

Low levels of FEV_1 in older life and therefore risk of COPD may result from failure to attain a good level of lung function in early adulthood, or a shortened plateau phase or a faster than normal decline of lung function, or a combination of these factors.[73] Rijcken and Weiss[76] considered that maximally attained level of FEV_1 was the most important predictor of reduced lung function in later life, after adjustment for smoking, respiratory symptoms, eosinophilia and airway hyperresponsiveness.

Both cohort and period effects are seen in population lung function measures, with higher lung function in young adults observed now than 40 years ago.[73] The effect of this has been quantified for vital capacity at 5ml/year or higher.[77,78] This is partly a cohort effect due to due to increases in height in younger cohorts as a result of nutritional or environmental changes.[79] Period effects related to learning effects of individuals in cohort studies and to improvements in spirometry techniques and apparatus may also result in apparent higher lung function in more recent years. Period effects in the 24 year follow-up study of Vlaardingen-Vlagtwedde in the Netherlands resulted in an average increases in FEV_1 over the course of the study of 250ml for men and 219ml for women.[78]

Low FEV_1 levels are associated with higher risk of death[80] and of death from COPD in later life.[65] In a nested case-control study using the Whitehall cohort,[65] a 10% reduction in FEV_1 in middle age was associated with a doubling in risk of death from COPD in later life.

Relationship between symptoms and COPD

Two studies[81,82] have assessed the relationship between various questionnaire assessed symptoms (including the MRC scale) and lung function, finding only a moderate association with levels of FEV_1 (correlation coefficients of around 0.4). The relationship between questionnaire assessments of COPD symptoms in individuals and later mortality is unclear. However, a study[83] assessing geographical correlations between COPD symptoms from the 1995 Health Survey for England and COPD general practice (GP) consultations, hospital admissions and mortality found good correlations (Spearman rank correlation coefficients ~ +0.8) between symptoms and the other databases, suggesting that areas with high symptom prevalence experience high morbidity and high mortality from COPD.

There has been much debate about the relationship between symptoms such as mucus hypersecretion, a key feature of the chronic bronchitic presentation of COPD, and mortality from obstructive lung diseases.[84] While there is no clear relationship between symptoms and extent of FEV_1 decline (and therefore COPD),[58] several [85-87] but not all[88] studies have found an association between mucus hypersecretion and COPD mortality, which is independent of, but much weaker than, the association with FEV_1.

Genetic factors

Various differing observations have suggested a role for genetic influences on COPD development e.g. an increased familial risk of airflow obstruction seen in smoking siblings of patients with severe COPD,[89] observations that only a proportion of smokers develop COPD and marked ethnic differences in COPD prevalence that cannot be fully accounted for by differences in smoking patterns.[71] Despite many studies of single candidate genes, the only current established genetic factor remains α_1-antitrypsin (α_1-AT) deficiency.[90] α_1-AT is the major antiprotease in the lungs that neutralises neutrophil elastase activity, thereby limiting damage to the lung during inflammation. Homozygotes for the Z variant of α_1-AT have low levels of α_1-AT (10% of normal) and develop severe emphysema, even if non-smokers. However, they only account for 1-2% of all patients with emphysema[71] and this genetic variant is predominantly seen in Caucasian Northern Europeans. There is no good evidence that heterozygotes for this condition, or those with other variants of α_1-AT gene or its expression are at increased risk of COPD.[71]

There is also no consistent evidence that genetic variations in other antiproteases, proteases, detoxifying enzymes, enzymes involved in the generation of reactive oxygen species, cytokine or immunoglobulin production, blood group or HLA status give an increased risk of developing COPD.[71] Mutations in the cystic fibrosis transmembrane regulator have been linked with development of bronchiectasis, but not with chronic bronchitis or other obstructive lung diseases.[71]

It seems likely that, with the exception of α_1-AT deficiency, COPD is a genetically complex disease.[90] Studies of lung function have consistently shown small but statistically significant correlations between related family members and moderate to good correlations between monozygotic twins (that are higher than dizygotic twins) and

suggest that lung function is controlled by many independent genes.[91] Several recent studies have found associations between combinations of polymorphisms and lung function decline or emphysematous changes on chest computed tomography (CT) e.g. polymorphisms of enzymes that activate or detoxify tobacco smoke such as epoxide hydrolase and hemoxygenase-1,[92] antioxidant gene polymorphisms,[93] or a combination of α_1-AT deficiency with certain haplotypes of microsomal epoxide hydrolase.[94]

A review published in 1999[91] suggested that 20-40% of FEV_1 variability could be explained by genetic variation, with a possible increase in genetic variance in smokers. However, it is not possible to completely separate genetic and environmental variance, particularly if there is an interaction between genotype and environment, so these figures should be interpreted with caution. Gene-environment interactions have yet to be studied in any detail.[95]

Smoking

Smoking is the acknowledged major risk factor for COPD. Peto et al,[46] using indirect estimates from vital statistics, estimated that approximately 75% of COPD mortality in the UK was related to smoking between 1950-2000 and that the percentage had fallen over time in men and been rising in females (Table 2-4).[96] Other authors have suggested figures between 70%-90% of COPD deaths attributable to smoking.[97-99]

Table 2-4 Numbers of deaths attributed to smoking (% of all deaths) in the UK, indirectly estimated from vital statistics

Year	Males	Females
1995		
Age 35-69	2,654 (76%)	2,212 (75%)
Age 70+	9,387 (77%)	6,170 (76%)
1990		
Age 35-69	3,765 (79%)	2,406 (74%)
Age 70+	11,754 (79%)	5,854 (74%)
1985		
Age 35-69	4,997 (82%)	2,476 (72%)
Age 70+	13,501 (81%)	5,324 (69%)
1975		
Age 35-69	7,344 (84%)	1,970 (65%)
Age 70+	12,237 (81%)	3,316 (53%)

To be conservative, no deaths before age 35 years were attributed to smoking.
Source: Peto et al, 1994[96].

The method used by Peto[46] involved using lung cancer mortality to estimate the smoking rates by age-group, determining the relative excess mortality expected from COPD in that population (summing over age-groups) using the risk estimate associated with smoking derived from a six-year follow-up study of over 1 million Americans, then halving the excess risk to be conservative and allow for confounders. The halving has little impact on percentage excess where the tobacco-attributable mortality is high as in COPD; also this method applied to the USA came up with similar numbers of deaths attributed to smoking as did a report by the US Surgeon-General that incorporated data on the prevalence smoking. A refinement of this method by Ezzati and Lopez[100] suggested that halving the excess risk was overly conservative to allow for confounders and reduced the correction factor from 50% to 30%.

Cigarette smoke is a mixture of about 4,000 substances. There is limited evidence about the constituents likely to induce COPD but possible components include nicotine, irritants such as acrolein and free radicals.[101] Tar exposure, which is associated with lung cancer risk, does not seem to be associated with airflow obstruction.[101]

Effect of smoking on lung function

Smoking affects

(i) the maximally attained level of lung function

(ii) the age of onset and duration of the plateau phase of lung function

(iii) the rate of decline of lung function

Passive tobacco exposure in utero and early childhood

Passive exposure (usually to parental smoking) affects lung growth in utero and childhood resulting in lower maximum lung function in early adulthood. A systematic review of the effects of parental smoking[102] suggested that the effects were small but statistically significant, probably related to a combination of smoke exposure in utero, neonatal life and later childhood . A meta-analysis of 21 surveys of school-children[102] found that the mean percentage reduction in FEV_1 in children whose parents smoked was 1.4%. Maternal smoking appeared to be more important than paternal smoking in all studies (except for a study in Shanghai, where women almost never smoke) and there was a dose-response effect, with greater effects on lung function in studies where mothers smoked more. A meta-analysis of nine studies[102] suggested there was a greater effect in boys, but this was not statistically significant (the pooled effect estimate for

exposure to passive smoke was a 2.1% (95% CI 1.5% to 2.8%) reduction in FEV1 in boys and a 1.3% (95% CI 0.6% to 2.0%) reduction in girls). A follow-up of 19,000 individuals in the European Community Respiratory Health Survey demonstrated that the reduction of lung function resulting from passive tobacco exposure in childhood persists into adulthood.[103] Additionally, there may be a synergistic effect between childhood exposure and personal smoking in adult life.[104]

Active smoking in adolescence

Active smoking, even of relatively small numbers of cigarettes, in adolescence can slow lung growth resulting in lower maximum lung function in early adulthood, independent of maternal smoking.[105] In a large follow-up study of lung function in children aged 10-18 as part of the Harvard Six Cities studies,[106] lower growth of the order of approximately 1% per year in FEV_1 and $FEF_{(25-75)}$ was seen in girls smoking five or more cigarettes per day; the effect was smaller in boys and not statistically significant. Another study with eight years of follow-up of 669 subjects, suggested that adolescents who begin smoking at age 15 would have a FEV_1 level approximately 92% of that of non-smokers after five years of smoking,[105] with similar size effects in males and females.

The effect of smoking on lung function may be underestimated due to the 'healthy smoker effect': children with healthier lungs (non-asthmatics or those with larger lungs) may experience less discomfort from smoking and therefore persist with smoking. Supporting this, adolescent smokers in the Harvard Six Cities studies[106] were observed to have higher levels of FVC (forced vital capacity) in the year before starting smoking than non-smokers (although growth of FVC subsequently became lower than in non-smokers).

Active smoking in late adolescence and early adulthood

Active smoking shortens duration of the plateau phase in early adulthood and brings forward the age at which FEV_1 starts to decline.[74,75]

Active smoking in adulthood

Lung function declines with age and the rate of decline increases with age.[73] The natural average rate of decline of FEV_1 is approximately 30ml per year. Smoking increases the rate of decline in moderate to heavy smokers by at least 15ml[73] and some

studies have suggested that the decline in smokers is double that of non-smokers at 60ml per year[44,107]. There are a subgroup of approximately 15-20% of smokers that show an increased susceptibility to tobacco smoke with a faster rate of decline of FEV1 (70-120 ml per year).[44]

Passive smoking in adulthood

While the detrimental effect of environmental tobacco smoke (ETS) in childhood on decreased lung function is well established, the effect in adulthood is less clear. Coultas[98] in a review published in 1998 found that only half of 20 studies assessing the impact of ETS on lung function found an effect, but all four studies looking at COPD hospitalisation or mortality found an increased risk with ETS exposure. He concluded that passive smoking in adulthood was a plausible risk factor for COPD, but that the magnitude of the association was likely to be small and further studies were needed.

Effect of different types of smoke and quantity smoked on developing COPD

Quantity smoked

The effects of smoking on lung function decline show a dose-response effect, being higher with more pack years.[44,45,73] However, a wide variation in decline is seen among smokers of similar numbers of cigarettes.[84] The forty year follow-up of the British doctors study[108] suggested a dose-response effect of smoking on COPD mortality, with a two-fold difference between smokers of <15 and >25 cigarettes per day.

In middle-aged men, stopping smoking will reduce the rate of decline in FEV_1[109] to that seen in non-smokers[110] and also has beneficial effects on mortality.[111] However, the benefits of late quitting are less clear. Reducing or quitting smoking in older individuals (>55 years[109] or >45 years[112]) may have no effect on FEV_1 decline while two large cohorts studies have shown that death from COPD actually increased for 5-10 years after quitting[110] – presumably in those with advanced disease.

Different types of smoking – plain or filter cigarettes, pipes, cigars

There have been marked changes in the form in which tobacco has been smoked over the last century, which may be relevant to trends in COPD mortality. From the late 19th century to 1940, manufactured cigarettes became more popular than pipes or cigars.[113] Cigarettes have also changed over time from plain to filter tipped and tar contents have reduced.

All types of smoke – from cigarettes, pipes or cigars – have been found to accelerate the decline in FEV_1.[73] Unlike phlegm production, there appears to be no significant difference in decline in FEV_1 between smoking plain or filter[86,114] and high tar or low tar cigarettes.[84,114] Other studies have shown no cross-sectional difference in lung function between smokers of manufactured cigarettes and leaf tobacco (smoked in pipes or hand-rolled cigarettes),[115] while a six-months randomised controlled trial showed no effect of tar reduction on respiratory symptoms or peak expiratory flow rates.[116] However, the type of cigarette may influence mortality. A 1981 review by Lee[117] quoted data from the American Cancer Society where emphysema mortality was found to be 30-40% lower in low tar cigarette smokers compared with those using higher tar cigarettes and data from a Tobacco Research Council study in North East England where chronic bronchitis mortality was found to be 40-60% lower in those smoking filter compared with plain cigarettes.

Cigar or cheroot smokers and pipe smokers who inhale were found to have greater declines in FEV_1 than cigarette smokers in a five year follow-up of the Copenhagen City heart study,[86] possibly because this group had very high tobacco consumption. Pipe smokers who did not inhale had a decline in FEV_1 similar to that of non-smokers.[86] However, those smoking pipes and cigars in the British doctors 40 year follow-up[108] had COPD mortality rates that were higher than never smokers and similar to those of former cigarette smokers.

Smoking is not the only influence on COPD development

Strachan (commenting on data from the 1946 birth cohort) stated that

> '... even if smoking were entirely eliminated, a substantial burden of chronic respiratory morbidity would remain. Respiratory symptoms and illness are not uncommon in lifelong non-smokers, as illustrated by data from the British 1946 cohort...' [118]

Additionally, other factors may be needed in combination with smoking to produce clinically important disease.

That factors other than smoking are important in COPD development is supported by different types of studies:

- Not all smokers develop clinical COPD. Many authors state that 10-20% of smokers develop COPD.[119] This figure is sometimes referenced to the cohort of London Transport and Post Office workers followed from 1961-1969 by Fletcher and colleagues,[44] although this percentage is inferred rather than directly stated in the original study. Barnes[71] qualified the figure as 10-20% of Caucasian smokers developing COPD, with lower rates in other ethnic groups. A more recent Swedish cohort study suggested that 50% of smokers may develop COPD[119] and this may because the cohort included older smokers than the West London cohort.[44]
- Some non-smokers develop COPD. A study of 13,000 lifelong non-smokers participating in the US Health and Nutrition Examination Surveys (NHANES) found that 4% of men and 5% of women reported physician-diagnosed chronic obstructive pulmonary disease.[120]
- British COPD mortality time trends in the 20th Century were not consistent with trends in smoking patterns.[26]
- Spatial variations in COPD mortality within England[21] with excess mortality in northern areas could not be fully explained by smoking variations – smoking may have explained as little as 50% of the excess risk.
- Cigarette smoking does not explain international differences in COPD. Comparisons of lung function published in 1971[5] found that British men had poorer lung function than men in America and Norway after controlling for smoking. Smoking did not explain international differences in COPD mortality for 31 countries from 1979-1888[121] – however, the authors comment that data issues may have been at least partly responsible.

Occupational exposures

Occupational exposures to dusts, gases and chemical fumes have all been associated with an increased risk of COPD.[122] Coal mining has been the most extensively studied, but occupational exposures to other types of dusts and some chemical agents may also be involved. There is likely to be considerable variation in individual responses to dust exposure.[43,123] Coggon and Newman Taylor[43] suggested that part of this may be related to variations in diet or in exposure to respiratory infections during infancy, which might alter the response of the lung to dust. Also, the influence of occupational dust exposure on lung function, COPD and chronic bronchitis may be substantially underestimated

due to bias and confounding.[124] Biases operating include selection bias (loss to follow-up of workers leaving the industry – the 'healthy worker effect') and non-differential misclassification bias tending to underestimate the true association. Confounding due to other exposures, social class and incomplete control for smoking will have had variable effects on the effect estimates.

Coal Mining

The relationship between dust exposure in coal and other mines and both reductions in lung function and excesses of mortality from COPD has been clearly established.[43]

A dose-response relationship was seen between increasing exposure to dust from mining and reductions in FEV_1, even in the absence of bronchial symptoms in smokers and non-smokers and whether progressive massive fibrosis was present or absent.[43] The effects of coal mining on lung function decline studied in the UK have been shown to be additive to those of smoking,[43] but results from some other types of mines e.g. South African gold mines suggest different modes of action e.g. an interaction between effects of gold mine dust and smoking.[124] A study[125] with moderate to heavy dust exposure suggested that the loss of lung function relating to smoking in British coal miners over an 11 year period was approximately three times that related to dust exposure (122ml loss in FEV_1 vs. 42ml).

A re-analysis by Oxman et al in 1993[124] of an analysis of surveys of British coal miners by Marine et al,[126] suggested that following a lifetime of moderate exposure to dust, 0.45% of non-smokers and 0.74% of smokers would be expected to develop chronic bronchitis attributable to dust. A review by Coggan & Newman Taylor[43] of studies in England, Wales and the USA suggested that miners have an excess mortality from COPD, which is of the order of 20-40% higher than in the general population. Some studies reviewed also examined lung cancer mortality and did not find an excess, suggesting that differences in smoking rates were not responsible for the excess. A study of deaths in miners in South Wales in 1950-1970 (the area with some of the highest proportional mortality ratios or PMRs for pneumoconiosis) suggested that COPD deaths were not related to the presence or severity of pneumoconiosis.[127] Part of the excess may be due to increased likelihood of doctors to diagnose or to use this on the death certificate, but this is more likely to occur with deaths from 1992 onwards,

when Industrial Injuries Disablement Benefit was first made available for miners developing COPD.[123]

Other mineral and organic dusts

Silica exposure in occupations other than mining[128,129] and dust from rock blasting and drilling[129,130] have been associated with accelerated decline of FEV_1 in men in their 30s and 40s. Estimates of the effects vary – a Norwegian population cohort with self-reported exposure to silica suggested a decline in FEV_1 in those exposed equivalent to 60% of that associated with smoking.[128] Declines in FEV_1 in tunnel workers exposed to respirable dust and α-quartz were 140-180% those of non-exposed smokers. Exposure to asbestos dust has been implicated in accelerated lung function decline by some authors,[131,132] both in mining asbestos (with rock dust exposure) and in manufacturing processes,[132] but asbestos exposure does not feature as an accepted risk factor for COPD mortality. Organic dust exposures implicated in an increased decline of lung function include cotton dusts (byssinosis), grain dusts, wood dusts[131] (including paper mills)[133] and those associated with animal husbandry, especially pig farming.[134] However, more studies are needed to clarify the exposure-response relationships and host susceptibility factors.[131]

Chemical agents and fumes

There is debate as to whether chemical agents known to cause occupational asthma, such as isocyanates, cause COPD.[133] Exposure to welding fumes has been linked with chronic productive cough and a small excess of asthma, but not convincingly with COPD.[133]

Air pollution

Mechanism of effects of particulates

It has been proposed that antioxidant stress is the mechanism of action by which particulate air pollution has effects on COPD, increasing the already increased oxidant burden in the lungs of COPD patients, together with upregulation of transcription factors and pro-inflammatory gene expression.[135] These proinflammatory effects may interact with infections such as adenovirus, increasing the risk of an exacerbation. However, other factors involved may include number and size and surface area of particles deposited in the lungs, together with surface chemistry.[136]

Acute effects

While short-term changes in particulate air pollution are clearly associated with short-term effects on respiratory mortality,[38,39] few studies have specifically looked at COPD mortality, possibly because of smaller numbers. Of the few studies that have been conducted, short-term increases in particulate and SO_2 air pollution were found to be associated with increases in COPD mortality[137] and increases in COPD emergency hospital admissions[138] in several countries in Western Europe (in the latter study NO_2 and ozone were also seen to have an effect).

Chronic effects

Sunyer reviewed the relationship between outdoor air pollution and COPD in 2001 and concluded that 'study of the chronic effects of air pollution in COPD is incomplete'.[139]

Individual-level studies

Most of the information comes from cross-sectional studies with large potential for bias and few longitudinal cohort studies have specifically looked at this issue.[139] Reviews by Sunyer[139] and Ackermann-Liebrich[140] concluded that:

(i) Most but not all[141] cross-sectional studies of lung function levels had found that higher outdoor air pollution levels were associated with small reductions in FEV_1 and FVC

(ii) Higher air pollution levels were associated with decreased growth of FEV_1 in children. Three cohort studies were located showing effects from higher particulates and SO_2 in Poland, higher particulates in Brazil and higher ozone in Austria & Germany

(iii) A greater decline of FEV_1 was seen in adults in a large Californian cohort[142] who lived in an area with high levels of particulates, sulphates and nitrogen oxides (the only adult cohort study located) in the review.[139]

(iv) Associations were also found between living in areas with higher particulate and sulphate air pollution levels and increased reporting of respiratory symptoms (e.g. The Study on Air Pollution and Lung Diseases in Adults (SAPALDIA)[40]). However, different studies have implicated different pollutants, usually one or more of sulphur dioxide, nitrogen dioxide, particulate matter and ozone.[58]

A large cohort study to examine the long term effects of ambient air pollution on decline of lung function is currently in progress, involving follow-up of adults from the European Community Respiratory Health Survey (ECHRS) in 29 European centres.[143]

One of the few individual-level studies to look at COPD mortality found that childhood exposure to smoke and sulphur dioxide air pollution in the UK was not associated with deaths from chronic obstructive pulmonary disease (COPD) or lung cancer in adults born in the 1930s and dying in 1994-96.[144] This unpublished study[144] used a case-control design and assessed exposure using a four-point index of air pollution derived from coal consumption. Confounders adjusted for included social class, place of residence in childhood and the 1990s at individual level and infant mortality at district level. However, the study did not examine cumulative effects of exposure in adult life and the exposure misclassification may have made any effect difficult to detect.

Ecological studies

A number of ecological studies of associations between air pollution and mortality have been conducted in the UK, but these have examined respiratory rather than COPD mortality, presumably as numbers of events were larger. Broad scale (county-level) studies conducted in the UK found associations between rates of respiratory disease and average air pollution levels in the 1950s[145,146] and 1960s[147] but not in the 1970s.[148] A more recent currently unpublished UK small-area study found associations between respiratory mortality 1982-1998 and cumulative air pollution levels in the preceding 4-16 years in over 400 wards containing an air pollution monitoring station throughout this time period.[149] This study included adjustments for ward-level deprivation, which may have partially adjusted for smoking.[150]

Diet

Diet may be an important risk factor for COPD. There are plausible mechanisms by which fruit and fish may protect against lung damage. Fruit is high in antioxidant vitamins (e.g. vitamins A, C and E) and nutrients (such as selenium and flavonoids), which may be protective against oxidative lung damage.[151,152] COPD patients and smokers have been shown to have increased markers of oxidative stress.[153,154] Oxidative stress in the lungs from tobacco smoke and other agents such as diesel exhaust particles[152] leads to injury, inflammation and epithelial permeability through a variety of mechanisms,[153] including inactivation of antiprotease agents such as α_1-AT.

Antioxidants such as vitamin C are expressed in lung epithelial lining fluid[155] and may modify the adverse effects of cigarette smoke.[156] Fish oils are thought to have anti-inflammatory effects through influences on arachidonic acid metabolism.[157] However, the role of diet is not clear-cut. Biological markers of oxidative stress have not been demonstrably correlated with lung function in individual level studies.[154] Any observed effects of diet may be a result of confounding by other aspects of a healthy lifestyle.[157]

A number of epidemiological studies have supported a role for fruit and vegetable or fish consumption or related nutrients; in some cases a marked effect. A longitudinal analysis of the EPIC-Norfolk cohort[158] found that the rate of decline of lung function in smokers was two-fold greater in those in the lowest quartile of vitamin C than those in the three higher quartiles after adjusting for pack years smoked. Cross-sectional studies in adults have shown associations between reduced consumption of fresh fruit or fish[159-161] and low plasma vitamin C levels[158] and reductions in lung function. An ecological study of COPD mortality rates in 16 cohorts in the Seven Countries Study found that baseline mean fruit consumption explained about 57% and fish consumption 10% of the variance in 25 year COPD mortality.[162] This study found that the effect of specific nutrients in the diet (vitamins C and E and selenium) were not statistically significant (although protective), which may be related to the methodology used or suggest that a number of nutritional components are necessary for protective effects. Not all observational studies have found a relationship between fish intake and clinically manifest respiratory disease, possibly because of low fish intake in some of the populations studied.[157]

Associations between high intakes of sodium[163,164] and low intakes of magnesium and zinc[164] on asthma and bronchitis symptoms have also been reported, but conflicting results have been seen, particularly with sodium[165] and more epidemiological studies are needed to confirm the relationships.[157]

Climate

Climate – in particular fogs – were considered an important factor in development and exacerbations of chronic bronchitis in earlier years.[31,33,34] Reid[31] reported that in London in the 1940s, fogs (visibility less than 1100 yards) affected 8-10 days per month between October to March on average. Since the marked effects of fogs on respiratory disease[31] are now considered to be related to the associated high levels of air pollution,

any possible independent effect of fogs or humidity has not been considered in recent years. Nandram et al[166] investigated deaths from COPD (ICD9 490-6) in health services areas in the USA and found that death rates were higher in areas where rainfall was low. However, this corresponded to areas of very low rainfall in the western states, and may have been confounded by other potential risk factors such as high altitude and living in an area remote from medical facilities.

An analysis by Law and Morris[21] suggested that approximately 25% of the differences in COPD mortality across England & Wales by latitude could be attributed to temperature differences. Blane et al[47] found a relationship between poorer climate (the sum of Z-scores of average rainfall and temperature 1961-1990) and lung function (deviation from expected FEV_1/FVC) that was independent of smoking status and social class. However, the analysis suggested that the interaction between poorer climate (in more northern areas) and housing quality may be a more important factor than climate in explaining UK variations in lung function.

Early life factors

Barker[167] in particular has highlighted the contribution of impaired lung growth in early life from nutrition or respiratory infections to an increased risk of development of COPD.

The first component of this is thought to be undernutrition in mid-late gestation leading to impaired airway development. This would explain the associations observed between low birthweight in 1911-1930 and chronic bronchitis deaths.[168] Associations between low birthweight and low FEV_1 have been demonstrated in both children – from the 1958 birth cohort examined at ages 5-11 years[169] – and adults e.g. men aged 59-70 years born in Hertfordshire in 1920-30[168] and men and women of mean age 47 years born in Mysore, India in 1934-53.[170] However, no association between low birthweight and lower lung function in later adult life (mean age 58 years) was seen in a cohort study of 239 Scottish adults.[171]

The second component of impaired lung growth in early life relates to respiratory infection during early childhood. Cohort studies have fairly consistently found an association between pneumonia in infancy and early childhood to be associated with reductions in lung function in childhood,[172] mid adult life[173,174] and later adult

life[168,171,175] (for details see Table 2-5, pages 55-56). The association with reduced lung function appears to be specific to pneumonia as associations have not been seen with whooping cough[171,173-175] (except in the Hertfordshire cohort,[168] where an effect was seen for whooping cough at age<1 yr but not age 1-4 years), with measles,[168,171] nor with bronchitis[171,175] (except in the Hertfordshire cohort again,[168] where an effect was seen for bronchitis at age<1 yr but not age 1-4 years). These findings from cohort studies are consistent with routine data analyses by Marks and Burney,[26] where temporal trends in mortality due to measles, influenza and whooping cough in children aged less than five could not be used to explain trends in COPD mortality in later life.

The importance of the association between pneumonia in early life or bronchitis in infancy and later COPD is unclear. The size of the reduction in lung function related to pneumonia in early life is small: a mean reduction of 6% in FEV_1 at age 34 years was seen in the 1958 birth cohort. The attributable risk from respiratory infections (bronchitis, bronchiolitis and pneumonia) before the age of 2 years was only 2.9% for low peak expiratory flow rate at age 36 years, 10.7% for wheeze and asthma at age 36 years and 16.5% for phlegm at age 25 and 36 years in the 36 year follow-up of the UK 1946 birth cohort.[176] However, Barker and Osmond[177] found a strong geographical correlation between death rates from chronic bronchitis and emphysema in adults aged 35-74 years in 1968-78 and infant mortality rates from bronchitis and pneumonia in 1921-1925. The study inferred that a threefold variation in bronchiolitis in infancy might be associated with a twofold variation in mortality from chronic bronchitis if the association were causal.[178]

However, the relationship between respiratory infections in childhood and lower adult lung function may be confounded by maternal smoking, inadequate adjustment for social class or overcrowding, birthweight or gestational age.[179] Also, the association may be of reverse causation, with susceptibility to respiratory infections in early life accompanying chronic conditions such as asthma that are also associated with reductions in lung function.[179]

Asthma and airways hyperresponsiveness

There is some evidence that asthma and bronchial hyperresponsiveness may be risk factors for later development of COPD.[58]

Patients with a long duration of asthma have been found to have worse lung function than non-asthmatics, whether smokers or non-smokers.[180,181] A large Danish population cohort study[182] following approximately 14,000 individuals over 17 years found that the asthmatics had poorer survival and a substantially higher risk of later death from COPD. Poorer survival was predominantly mediated through reduced FEV_1 and was seen in never-smokers as well as smokers.[182] The relationship between COPD and asthma may be due to under-treatment of asthma leading to airways remodelling and fixed airflow obstruction.[58] There is concern that under-treatment may be responsible for the observed increased decline of lung function in asthma.[58,181,183] This is of particular public health significance to the UK as there was a well-documented rise in asthma prevalence between the 1950s and early 1990s[184-186] and evidence of under-diagnosis and treatment in at least the earlier part of this time period[187,188] (but a suggestion of over-treatment or inappropriate treatment by the end of the 1990s[189]). The implication of this for the UK is that the expected reduction in COPD following declines in smoking rates may not be fully realised.

Airway hyperresponsiveness appears to be associated with an accelerated longitudinal decline of lung function in both cross-sectional and longitudinal studies[76,190] and also with later death from COPD.[191] Although atopy, which includes allergic asthma, predisposes to airways hyperresponsiveness,[190] the association between hyperresponsiveness and lung function decline appears to be independent of atopy – and also of smoking and symptom status.[76] It is likely that airways hyperresponsiveness in the UK increased in parallel with the rises in atopy and asthma. However, only one UK study[192] has specifically looked at population changes in airway responsiveness, measured by a fall in peak expiratory flow (PEF) after exercise, over this time. The study found a small increase in the proportion of children with a 15% drop in PEF (6.7% in 1973 and 7.7% in 1988), but the validity of this type of test to measure airway responsiveness has been criticised due to poor reproducibility.[193]

There is no clear evidence that atopy (as assessed by skin testing and serum IgE concentration) is a risk factor for irreversible airflow obstruction in persons without asthma,[190] but some recent studies have suggested an association independent of bronchial hyperresponsiveness.[194]

Socio-economic status

Lower social class is a strong risk factor for COPD mortality at both individual[195] and area-level.[196] An ecological study of ward-level mortality in England in 1981-1985[196] found linear increases in premature mortality from smoking-related diseases including COPD with increasing deprivation (as well as regional differences at similar levels of deprivation). There are also strong associations between socio-economic class and indices relevant to COPD (lung function, symptoms, mortality, diagnosis).[195]

Although there is a strong social class gradient in smoking, with higher rates seen in lower social classes,[197] smoking is not the only explanation for the observed social class effects. A social class gradient for chronic bronchitis mortality was observed before there were any large variations in smoking prevalence by class.[27] Prescott and Vestbo[195] reviewed 14 studies (three from the UK[34,80,198]) with individual level data detailing the effects of social class. The social class effect persisted after adjustment for smoking, which was available in 10 of the 14 studies. The only UK study in this review to adjust for smoking[34] suggested a difference of approximately 10% in FEV_1 (adjusted for age and height) between manual and non-manual classes in Bath and in Caerphilly, independent of smoking or working as a miner.

Results from the Copenhagen City Heart Study[199] suggested that the difference in FEV_1 between the high socio-economic index (long education and medium or high income) and low socio-economic index (short education and low income) after adjustment for smoking was between 200-400 ml depending on age and sex (males had higher differences). In never smokers, the difference in FEV_1 between low and high socio-economic index was equivalent to approximately 10 years of ageing.[199]

The mechanisms involved in the relationship with social class are likely to be multi-factorial[195] including lower lung growth in utero (e.g. through maternal smoking), environmental exposures (e.g. to tobacco smoke and air pollution) in childhood, higher risk of childhood respiratory infections, higher levels of smoking, occupation (e.g. higher exposure to dust, gases or chemical fumes is seen in manual as opposed to non-manual workers[200]), poorer housing and a diet with fewer protective antioxidants found in fresh fruit and vegetables starting in childhood.[201] These factors will operate

throughout the life course,[28,202] and may result in a life-time accumulation of hazard exposures to those from lower socio-economic class.[203]

Urban environment

High SMRs for chronic bronchitis were noted in the mid 1980s in industrial areas of south Wales, northern England and London, with low SMRs in rural areas of England and Wales.[20] An ecological analysis by Strachan et al[15] of English mortality data, emergency hospital admissions and GP patient contacts in the early 1990s and symptoms from the Health Survey for England for 1995 (HSE95) found a clear urban-rural gradient. Standardised event (i.e. mortality, hospital admission, GP contact or symptom) ratios were lower in rural areas for COPD in all data sources, except for GP contacts where conurbations were significantly higher than the national average. Individual level data from the HSE95 also showed a small, but statistically significant excess of symptoms of cough or phlegm in those living in households which the interviewer judged to be in an urban setting (prevalence in urban setting 6.6%, rural setting 5.0%, $p<0.001$). Studies of migrants to the USA from Norway and England published in the 1960s found that prevalence of chronic bronchitis was higher in those born in urban as opposed to rural areas.[5]

The 'urban' effect is likely to have a number of components. These include higher smoking levels in urban areas and differences in socio-economic factors. In the analysis of English data in the 1990s by Strachan et al,[15] adjustment for smoking and social class using individual level HSE95 data attenuated the urban rural gradient in symptoms, but rural and mixed areas continued to have lower SERs than the average while urban and conurbation areas had higher SERs. Other components of the 'urban' effect may be related to higher levels of air pollution and differences in lifestyle and occupational exposures.

Gender

Differences in diagnostic labelling between men and women by physicians have already been noted with COPD more likely to be recorded as asthma in women.[50,69,70] After a diagnosis of COPD has been made, men have been found to have a higher COPD fatality rate on follow-up,[58] while women have been found to have a higher (all-cause) mortality rate after a diagnosis of asthma.[182] This may be related to differential

diagnostic labelling. However, symptom reporting may vary between the sexes.[204] While women are more likely to report shortness of breath for any FEV_1 deficit than men, they are less likely to report phlegm production and more likely to swallow phlegm than men,[204] which may affect chronic bronchitis diagnoses.

Women differ in both exposures and in susceptibility to COPD risk factors.[204] Women are less likely to be employed in occupations or have hobbies where there is exposure to mineral dusts than men, but are more likely to be exposed to environmental tobacco smoke, cooking fumes (e.g. NO_2 from gas cookers) and cleaning fluids in the home. Given similar exposures, girls and women appear to be more susceptible to the deleterious effects of smoking,[58,205] with greater effects on lung growth[106] and lower lung function in adulthood,[112,206] more bronchial hyper-responsiveness[207] and higher risk of admission to hospital.[208]

Risk factors with acute effects on COPD mortality

Two risk factors with acute effects on COPD mortality were identified:

i. Acute, usually infective, exacerbations of COPD that are severe enough to require admission to hospital are associated with a high risk of death, with 10-30% dying during the admission.[58]
ii. Air pollution episodes have been recognised to lead to acute increases in COPD mortality since the Meuse valley episode in France in 1930 and the London smog of 1952.[140] This was discussed earlier under air pollution (page 43).

Choice of methodology, variables and data

The hypotheses to be explored in this PhD set out in the introduction relate to time trends and spatial variations in COPD. It was decided to conduct ecological and age-period-cohort analyses using routine data for the reasons that are set out below.

Methodology

Population-level analyses

The focus of this PhD was on population time trends and population-level spatial variations in COPD mortality, with the intention to investigate hypotheses of public

health relevance. These population-level effects are arguably best investigated at population level. Certainly it would not be possible currently to investigate time trends in COPD mortality over the last 50 years using existing UK cohort studies and it would be difficult to investigate spatial variations across the whole of the UK using existing survey data (see paragraphs below). While there are drawbacks to population-level types of studies, especially in relation to confounding and the ecological fallacy,[209] the last decade has seen advances in methodological understanding and techniques relating to geographical studies.[210] New methodology and software developed to identify age, period and cohort trends[211,212] has allowed clearer and easier identification of age, period and cohort trends than those demonstrated by previous authors.[26,27] Advances in computing capabilities have allowed the increasing use of sophisticated statistical models using Bayesian techniques,[211,213,214] which are able to smooth fluctuations due to small numbers by 'borrowing' information from neighboring categories and combine information collected at different spatial levels. A newly developed Bayesian shared component model[215] has aided geographical analyses by allowing the partitioning of shared and disease-specific spatial variation in components of closely related diseases. This model therefore could be utilised to investigate the spatial hypothesis looking at areas discordant for COPD and lung cancer mortality.

The literature search suggested that most risk factors for COPD identified appeared to have additive rather than multiplicative effects i.e. acted independently rather than interacting. No interaction had been demonstrated between occupational exposure, smoking or air pollution and respiratory symptoms or spirometry results.[200] The only exception was the interaction between climate and housing variables[47] in relation to lung function parameters (not COPD). This suggested a limited role for *a priori* interaction terms to be included in the analysis.

Cohort, case-control and cross-sectional studies

It was beyond the scope of this PhD to set up and analyse a new study on the scale required to investigate the hypotheses in question. An alternative might have been to reanalyse data from existing studies. Potentially, data from a cohort study could have been used to investigate time trends and data from a cohort, case-control or cross-sectional study could have been used to investigate spatial variations, but limitations in existing cohorts and cross-sectional studies (discussed next) did not make this a realistic option.

Limitations in existing British cohort studies in relation to COPD

The main limitation in most of the British cohort studies located for the study of COPD (Table 2-5, pages 55-56) is that cohort members are not old enough for appreciable numbers to have developed the disease. COPD principally affects the elderly, with most deaths from the disease occurring over 75 years.[65] For example, one of the oldest cohorts with good information about risk factors is the 1946 birth cohort,[216] which is still too young for appreciable numbers to have developed the disease. Since COPD can be considered to affect approximately 5% of the population in terms of prevalence and mortality,[7] even where cohorts are old enough, the numbers involved may be too small for detailed analyses e.g. the Boyd Orr cohort[217] or West Coast of Scotland cohort.[218] Members of the Longitudinal Study cohort[219] are old enough to have died from COPD, but the cohort was enrolled in 1971 – after the greatest reductions in air pollution, and therefore hypothesised greatest reductions in COPD mortality, had occurred.

One possibility to circumvent the problem of cohorts being too young, is to use an early marker of COPD within existing cohort studies. Lung function either as a single measure or increased rate of decline would seem the most promising early marker of COPD. Data from the first Whitehall cohort[65] suggested that a 10% reduction in FEV1, FVC, FEV1/FVC ratio, FEF_{20-75} and FEF_{60-80} was associated with an approximate doubling of the risk of death from COPD, with the most consistent association in multi-variate analyses seen with FEV1 and FEF_{20-75} (which are closely correlated). Another analysis of the first Whitehall cohort at 15 years follow-up[220] suggested that 70% of the male chronic bronchitis deaths occurred in men with low FEV_1 in middle-age at baseline. There are a number of problems, however, with the use of lung function as an early marker of COPD. Some of the problems with spirometric definitions of COPD have been discussed above in the section on definitions of COPD (page 28). For example, different definitions give different prevalences of COPD, patients may be asymptomatic with mild or moderate COPD solely defined by lung function parameters[62] and cut-offs may not correlate well with clinical measures such as hospitalisations.[61] Low FEV_1 is not specific for COPD mortality and is also associated with higher risks of death from ischaemic heart disease and stroke.[80,220,221 80] Only a proportion of all patients with spirometric diagnoses of COPD are reported as dying from the disease (23% in the Tucson study[222]). Finally, not all UK cohorts have lung function measures (e.g. the Longitudinal Study, the British Doctors cohort).

To avoid the problems of numbers with COPD from individual cohorts being too small, information could be pooled from several studies. However, a study by Tang et al[223] combining four cohorts (over 56,000 men aged >35 years) only resulted in 283 deaths from COPD (ICD8 491-2,519) after an average follow-up of 13 years. Also combining cohorts will not solve other problems such as lack of information on risk factors or generalisibility issues due to selection criteria for participants.

Table 2-5 Selected recent UK cohorts with information on COPD

Cohort	Description of cohort	Respiratory findings	Comments
1946 birth cohort (MRC National Survey of Health & Development) Wadsworth et al[216] Mann et al[176]	5362 individuals born between 3-9 March 1946. Single, legitimate births: all from non-manual and agricultural families and random selection of one in four from manual classes	Follow up to age 36[176] found that poor home environment (overcrowding, manual class, lacking amenities), parental bronchitis and atmospheric pollution before age 2, childhood lower respiratory illness and later smoking were best predictors of adult lower respiratory tract problems (bronchitis, pneumonia or other chest illness since age 20, positive response to MRC questions on phlegm production, or any adult reports of chest wheeze/whistle)	Not old enough for appreciable numbers to have developed COPD. Most recent follow-up to age 43, but respiratory info not published yet.
British doctors cohort Doll et al[108]	34,439 male and 6,194 female doctors on British medical register in 1951 (response rates: 69% male and 60% female). Questionnaire surveys in 1951, 1957, 1966, 1972, 1978, 1990 with detailed smoking information. Follow-up for 40 years for death notification (flagged at central register).	At 40 year follow-up of male doctors, there had been 542 deaths from chronic obstructive lung disease (chronic bronchitis and emphysema). COPD death rates per 100,000 per year were 10 in non-smokers and 225 in current smokers of 25 or more cigarettes per day, with a highly significant dose-response effect related to the number of cigarettes currently smoked.	Old enough for appreciable numbers to have developed COPD and detailed information on smoking habits, but no information on early life risk factors. Professional group - may limit transferability of magnitude of smoking-related attributable risk to general population due to different lifestyle and occupational exposures.
Boyd Orr cohort Blane et al[217]	1352 families in 16 locations in Britain in 1937-9 interviewed including retrospective data collection 85% of children aged 5-14 years surveyed in 1937-9 were traced and flagged at central registry for death registration 294 of these children followed up at ages 65-75 including life-grid interview & spirometry.	Lung function regression analyses have not been published yet, but were presented at a seminar in Oct 2000 by Dr Blane: In follow-up at ages 65-75, lower lung function was associated only with adult but not childhood factors: – in men with social class (p=0.02), cigarette smoking (p=0.02) and occupational exposures to dust and fumes (p=0.02) – in women with exposure to domestic labour (domestic labour index used coal fires as sole means of heating & no. of people in household with double weighting for children < 10y)	Older age follow-up only includes 294 individuals in total, limiting the power of any analyses. These 294 are only ~10% of the original sample – validation studies suggest less deprived individuals probably over-represented.
Derbyshire cohort[175]	650 men and women born 1917-1922 in six districts in Derbyshire, with information from health visitors records for ages 0-2 years, and follow-up and spirometry at ages 67-74 years. Results relate to 330 men and 288 with satisfactory spirometry.	Pneumonia before age 2 years was associated with a lower mean FEV_1 of 650ml in men and 190 ml (but not statistically significant) in women. This result was unaltered by adjustment for smoking. Bronchitis, measles and whooping cough were no associated with differences in mean FEV_1.	There was no association between overcrowding at birth, own or parental social class, parental smoking or working as a coal miner.
GPRD COPD cohort Soriano et al[13]	78,172 patients with a diagnosis of COPD in the General Practice Research Database (GPRD) in 1990-1997	50,714 (45.9% women) were incident cases. Prevalence in women rose from 0.8% to 1.4% and in men from 1.4% in 1990 to 1.7% in 1997. Men had higher rates of mortality than women at all levels of severity.	No information on early life factors. Smoking information is available in the GPRD, but of questionable validity, probably related to under-recording.[15]

Table 2-5 (continued)

Cohort	Description of cohort	Respiratory findings	Comments
Hertfordshire cohort[168]	5718 men born in Hertfordshire in 1920-30, with information from health visitors records for ages 0-5 years. Spirometry results from 825 of 1157 men still living in Herts. in late 1980s.	55 of original 5718 had died from COPD by late 1980s, with a weak but non-significant trend of falling SMR with increasing birthweight. Lower birthweight was associated with worse adult lung function (at mean age 64), as was bronchitis, pneumonia or whooping cough at age <1 year but not measles. Bronchitis or pneumonia was associated with a 170 ml reduction in adult FEV_1 and whooping cough with a 220 ml reduction.	Small numbers of deaths from COPD. All results independent of smoking and social class. Results on whooping cough not replicated in other studies. Some but not all other studies have found an association between birthweight.
MacKenzie cohort[171]	Original birth cohort of 1525 born in St Andrews with health information available from the MacKenzie Institute for ages 0-14. 1070 sample selected for tracing, analysis performed on 239 men and women who had lung function measured at mean age 58 years.	An association was found with reduced FEV_1 and FVC in adulthood and pneumonia or bronchitis before the age of 2, but not with birthweight. Asthma or wheeze at age 2 years or older and cough after age 5 were also associated with reduced FEV_1. Associations were independent of parents' and own smoking.	Very small numbers of COPD deaths. Of the 1070 in sample selected for tracing, 183 had died but 107 deaths had occurred in those under five years of age. Only five deaths from COPD had occurred in those aged 40+.
ONS Longitudinal Study Moser & Goldblatt[219]	1% sample of the 1971 England & Wales Census, with individual level census information linked to death and cancer registrations	An analysis of occupational mortality from the 1971 Census in women followed up to 1977 suggested high levels of mortality from respiratory disease among charwomen and cleaners.	No published analyses specifically concerning respiratory mortality. No information on health related behaviours or health status (such as lung function)
Renfrew & Paisley cohort Davey Smith et al[218]	Population survey in 1972-6 of 6961 men & 7991women aged 45-64, including smoking habits and FEV_1. Follow-up over 15 years for death notification (flagged at central register).	2,133 male deaths and 1,492 female deaths from all causes. Both area and individual level measures of deprivation were associated with worse lung function and higher prevalence of MRC bronchitis at baseline and with all cause and cardiovascular mortality during follow-up.	No published papers refer to COPD mortality; some consider respiratory disease mortality. No information on early life factors.
West Coast of Scotland cohort Davey Smith et al[224]	Survey in 27 workplaces in 1970-73 in the west of Scotland, including information on smoking, lung function and father's social class. Follow-up over 21 years of 5645 men aged 35-64 at initial survey for death notification (flagged at central register).	1,580 deaths, of which 105 were from 'respiratory disease' – no definition available, presumably non-lung cancer (from which there were 185 deaths). Men with fathers of manual social class had a non-significant higher relative rate of death from respiratory disease, RR 1.6 (0.88-2.90) after adjustment for adult socio-economic status, smoking, body mass index and FEV_1 score – suggestive that both childhood circumstances and adult factors are important.	No published papers for this cohort refer specifically to COPD mortality, presumably because numbers are too small. No information on early life factors other than father's social class.
Whitehall study Ebi-Kryston[220] Thomson & Strachan[65]	19,018 male civil servants aged 40-74 (most <65 years) examined 1967-9 including FEV_1. Follow-up for death notification (flagged at central register) over 15 years[220] and 20 years[65]	At 15 year follow-up[220] 76 men had died from COPD (ICD8 490-2) – most (64) from chronic bronchitis (ICD8 491). Analyses suggested that the 6% of men with FEV_1 <65% of predicted for age and height accounted for 70% of the chronic bronchitis mortality. At 20 year follow-up[65] 170 men had died from COPD (ICD9 491-3,519.8); most were aged 75-80 when they died.	No information on early life factors. Spirograms are only now available for 143 of the 170 who had died of COPD by 20 year follow-up and for 143 controls – this limits potential for further aetiological studies.
Whitehall II study Marmot et al[225]	10,308 persons (6,895 men, 3,413 women) working in London government offices aged 35-55 at 1st survey in 1985-8. Phase 2 1989-90, phase 3 1991-3 (average 5.3 years follow-up)	828 cases of self-reported (MRC definition) chronic bronchitis at start. 287 male and 138 female incident cases during follow-up. Grade of employment in adulthood rather than father's social class was associated with chronic bronchitis morbidity.	No information on early life factors other than father's social class.

Many UK cohorts are not population-based (e.g. the British doctors cohort, Whitehall cohorts) or they relate to limited geographical areas (e.g. West coast of Scotland, Renfrew & Paisley), making them unsuitable for examination of spatial differences in COPD morbidity or mortality. Only the 1946 cohort, GPRD cohort and ONS Longitudinal study have sufficient geographical coverage to examine the known regional variations in COPD, but analyses would be limited by small numbers (1946 cohort) or lack of information on risk factors (GPRD cohort and ONS Longitudinal Study).

Most papers from these cohorts have been published on cardiovascular disease and limited information is available on respiratory disease. Even where respiratory disease analyses have been conducted, they can be slow to be published: the most recent respiratory findings for the 1946 cohort were published in 1992,[176] which only give information on respiratory function to age 36, while analyses of the Boyd Orr cohort relating lung function to early and later life factors have been conducted but not published (personal communication, Dr David Blane).

Case-control studies

Very few case-control studies have been conducted investigating COPD epidemiology worldwide. Given that cohort studies are generally held to be better at investigating exposures,[122] it is perhaps surprising that case-control studies have been mainly used to investigate exposures, such as passive smoking,[226] air pollution,[144,227,228] occupation[229] and diet.[230] Most of the UK studies have been nested within cohort studies.[65,144,230]

Cross-sectional surveys

No UK cross-sectional studies contain enough information to be able to look at time trends over the last 50 years. Even if there were suitable studies, repeated cross-sectional studies may give a biased view of lung function, due to a survivor effect – those with worse lung function will die earlier and therefore not contribute data at an older age.[44] Conversely, Dockery and Brunekreef have defended the use of repeated cross-sectional studies to look at air pollution effects, arguing that during a cohort study with a long follow-up period the effects of known confounders such as age and biases from differential drop-outs will be much larger than the expected effects of air pollution, while heterogeneity in exposures will tend to decrease due to averaging over time, thus reducing the power to detect effects.[231]

Regular detailed national health surveys have been conducted from the 1990s onwards in the UK and could potentially be used to examine spatial variations in COPD. Such surveys contained information on self-reported disease, symptoms and prevalence and in some cases spirometry. The Health Surveys for England in 1995 and 1996[232] and Scottish Health Surveys in 1995[233] and 1998[234] contained both spirometry results and questions on COPD symptoms based on the MRC questionnaire. No spirometry data were available from Welsh health surveys.

As discussed in definitions of COPD (page 28) and in '*Limitations in existing British cohort studies in relation to COPD*' (page 54), the use of spirometry data as a measure of clinically relevant COPD is not ideal. A practical limitation to using spirometry as outcome measures from English and Scottish surveys is that data are not generally available below regional level. Another issue to be considered is that cross-sectional surveys conducted in young adulthood may demonstrate only small lung function differences[235] and the number of older individuals by region and sex is relatively small. For example, combining the 1995 and 1996 Health Surveys for England gave approximately 250 males and 250-350 females per region aged 65+ years. Inclusion of large numbers of younger adults may be the reason why no substantial differences in mean FEV_1 were seen between Scotland in 1998 and England or northern England in 1996,[234] or this may be because grouping Scotland together obscures higher mortality in central Scotland with lower levels in the Highlands and islands.[22]

Using and combining surveys with information on symptoms alone was not considered justified as symptoms of chronic bronchitis may be a poor predictor of later COPD.[81] Another limitation of using survey data is the availability of risk factor information. The surveys generally have information on social class, smoking and some aspects of diet, but not other potentially important exposures such as air pollution. It would not be difficult to link survey data with other datasets as place of residence of participating individuals is not released due to confidentiality issues.

Variables and data

A major advantage in using population-level data was that a wide range of outcome and risk factor variables were available using routine data.

Mortality endpoint

Mortality was chosen as the endpoint of interest for a number of reasons. Mortality from COPD is of public health importance, as it forms approximately 5% of all deaths.[17] Temporal and geographical[83] variations in mortality should also reflect patterns of severe disease likely to require medical attention such as frequent hospitalisations and primary care visits. Mortality data in the UK are readily available and are generally of good quality.[236] While diagnostic and coding issues occur, these are well recognised and documented.[26,237] Most importantly for these analyses, the format of available mortality data readily permitted investigation of temporal trends and spatial variations. It was possible to obtain mortality data for England and Wales from 1950 onwards with reasonably consistent coding and geographical boundaries relating to three large groupings: conurbations, smaller cities and towns and less urban or rural areas. It was also possible to obtain mortality data for England, Wales and Scotland to small area level for 1981 onwards that could be grouped to the required spatial aggregation for spatial analyses. Since COPD accounts for approximately 5% of deaths, aggregating over time allowed substantial counts of deaths.

Mortality data were preferred to hospital data because they are available for a much longer time period, the quality is less variable and there have been fewer changes in collection procedures. Also, interpretation of spatial analyses of hospital data at sub-regional levels is complex due to local variations in factors that are difficult to adjust for. These include supply of hospital beds, admissions policies, ease of access and referral patterns of local general practitioners.[238]

COPD definition – including asthma and using an age cut-off

An inclusive definition of COPD was used to minimise the effects of spatial variations in diagnostic labelling and in diagnostic changes over time.[239] For example, death certificates show a gradual change from using chronic bronchitis in the 1970s to chronic airflow obstruction in the 1980s[65] and 1990s.[239] For the time trend analyses covering the last 50 years, a relatively consistent diagnostic group over time could be obtained by combining diagnoses of bronchitis, emphysema and chronic airflow obstruction. For the spatial analyses, which only considered data from 1981 onwards, a more inclusive diagnosis was possible, including chronic bronchitis, emphysema, chronic airways obstruction and asthma. This allowed to some extent for the diagnostic confusion

between COPD and asthma in the elderly and women. Since COPD is mainly a disease of older individuals, an age cut-off was used to help to distinguish between patients likely to have COPD and those likely to have asthma. Studies of the effect of inhaled steroids on FEV_1 decline have used a lower cut-off age of 40 years.[107,240] A study in Barcelona looking at patients attending emergency rooms for asthma or COPD observed that before age 45 asthma rather than COPD was diagnosed in 70% of males and 95% of females, while after the age of 45 asthma rather than COPD was diagnosed in 8% of males and 40% of females.[50] An age cut-off of 45 years would therefore be likely to be fairly sensitive and specific for males, but could be less so for females.

Choice of risk factors variables used in both time trend and spatial analyses

The time trend analyses were chiefly concerned with COPD mortality trends in relation to trends in air pollution, but smoking needed to be accounted for as one of the most important risk factors for COPD.[28] The spatial analyses were concerned with a range of risk factors but again, smoking needed to be considered.

Smoking

Components of smoke likely to cause airflow obstruction have not been definitively identified,[101] but there is no clear dose-response relationship of lung function decline with cumulative tar consumption.[241] This suggested that direct measures of smoking should be used in analyses.

The most comprehensive collation of direct measures of smoking levels in the UK up to the end of the 1980s is found in the 'UK Smoking Statistics' publications.[113,242] These chiefly relate to prevalence of smoking or tobacco sales. Data come from the Tobacco Advisory Council (TAC), which commissioned annual surveys on smoking from a market research company between 1948-1987, after which it was disbanded and from the General Household Surveys. These are ongoing annual lifestyle surveys that started in 1971, with individuals selected at random from the Electoral Register. Information on smoking was available in GHS surveys from 1972-1976 and every second year thereafter. More information on recent smoking can be obtained from the Health Surveys for England, which started in 1991, also selecting a random population sample. Where comparisons between surveys can be made, slight differences in results can be seen, probably because of the different methods used.[243] None of these sources of information on smoking provide data compatible with spatial units required for the

analyses presented here, being only available at national level or, at best, government region level – and where presented at regional level not broken down by age group.

It is likely, as for lung cancer, that the most important measure in development of COPD is that of cumulative smoking exposure.[108,241] It was therefore decided to use lung cancer mortality as an indirect estimate of smoking trends, following other authors analysing routine data[46] and assuming a reasonably similar lag period for COPD to that seen with lung cancer.

Air pollution

The literature search found that surprisingly few studies had specifically looked at COPD mortality in relation to either acute or long-term air pollution exposures. In the absence of clear information on the most important pollutants, dose-response, length of exposure required and lag periods, the choice of air pollution variables was made on the practical grounds of what was available.

Black smoke and sulphur dioxide levels from individual monitoring stations using similar measurement techniques were available from the mid 1950s from AEA Technology Environment (AEAT). Also, a number of previously published analyses of data from the 1950s onwards were available.[37,244-246] These were used to provide qualitative information in trends in air pollution over the last 50 years. For periods in the early 1950s, domestic coal consumption has been used as a proxy for air pollution by several analyses[144,176] using the categorical Douglas-Waller index[247] based on data from the former Warren Spring Laboratory. A copy of this index was not located, but is unclear whether it would have added to the qualitative information about trends in air pollution.

For the spatial analyses covering 1981-1999, interpolated air pollution data containing estimated background levels (rather than data for individual stations) were required. These were only available for the UK for a single year – 1996 – within the time period of analysis.

Choice of additional risk factor variables used in spatial analyses

Genetics

Information on genetic factors could not be included in the analyses because most genetic variations involved in COPD have yet to be definitively identified. Established genetic risk factors for COPD are associated with 1-2% of emphysema cases and some cases of bronchiectasis, therefore forming only a very small proportion of all COPD cases and therefore would not explain spatial variations in COPD mortality. It is possible that as yet unidentified genetic risk factors or gene-environment interactions may make a large contribution to spatial variations. However, even if these were already known it would not be possible to investigate them in this analysis as there are currently no genetic databases covering the UK.

Occupation

Occupational influences of industries with exposure to dusts and fumes (primarily mining and heavy industries) might be expected to have a non-trivial impact on the prevalence of COPD in an area. As certain occupations and therefore their associated mortalities are concentrated in certain areas, it is possible to include this information in a spatial analysis. For example, Coggon et al,[248] looking at mortality by occupation in 1979-1990 in England found that almost all (96%) deaths in miners were concentrated in 22 counties. Geographical analyses suggest that mesothelioma deaths, which are a marker for asbestos exposure, are concentrated in areas with dockyards and major ports such as Tyneside, Portsmouth and Plymouth and in east London.[249] It was therefore decided to use mortality from pneumoconioses and asbestos-related diseases as a marker for areas with concentrations of industries with exposure to dusts and fumes.

Diet

The literature search suggested that diet variations may be important in both individual and spatial variations in COPD mortality and that information on fruit and/or fish consumption would be preferred over average micronutrient content of food.[162] The National Food Surveys (NFS) give information on household food purchases and has been conducted annually from 1940.[250] The National Diet and Nutrition Survey (NDNS) was conducted in 1986-7[251] and again in 2000-1.[252] More limited information on diet is available from national health surveys such as the Health Surveys for England[232] and Scottish Health Survey.[234] However, these data were not readily

available at suitable spatial level and suffer from relatively small sample sizes (6,000 households surveyed for a single week in the NFS,[250] 1,200 individuals in the NDNS 1986-7[251] and 1,700 in the NDNS 2000-1,[252] cross-sectional data on 15,000 – 20,000 individuals each in the English[232] and Scottish Health Surveys[234]). It was therefore decided to use data from commercially available database on food purchases, the Taylor Nelson Sofres (TNS) database. This had recently been evaluated against the NFS[253] and the authors concluded that '*the TNS database should be considered as a potentially viable research database for estimating national dietary trends*'.

Climate

The literature search suggested that cold and damp could be relevant to COPD mortality. Data on climatic variables were readily available for individual monitoring stations from the Meteorological Office. However, interpolated data were found to be freely available from the European Union Monitoring Agriculture with Remote Sensing Unit (MARS)[254] and being used in a European Union air pollution interpolation project within the host department.

Early life factors

Only pneumonia and possibly bronchitis in early life were associated with lower lung function in adulthood,[168,171,175] but their importance as risk factors for later death from COPD is unclear. It was decided not to include this as a variable in the analyses given the difficulties in obtaining spatial data for the early 20th century, especially to boundaries consistent with those in the 1980s and 1990s, and the biases introduced by migration (no linkage is possible between British routine datasets such as birth and mortality data).

Asthma and airways hyperresponsiveness

Asthma was already included in definition of the outcome measure (COPD mortality) so could not be used as a risk factor. There is no national data on airways hyperresponsiveness that could be used in a geographical analysis of COPD in Britain.

Socio-economic factors

Socio-economic status is an independent risk factor for COPD and would need to be included into any spatial analysis. It was decided to use the Carstairs index[255] at area level as the socio-economic indicator. Data on occupation (and therefore social class)

can be derived from death certificates, but they are incomplete, especially in women and after retirement (generally age 60 onwards).[256]

Urban environment

It is likely that higher risks of COPD from living in an urban environment are related to differences in smoking levels, socio-economic class, occupational exposures and air pollution, all included in the spatial analyses. It was decided not include an urban co-variate (e.g. population density) as this may have over-controlled for other risk factors.

Gender

Because of the differences in exposures to potential risk factors for COPD, susceptibility to risk factors and diagnostic labelling of COPD in women, analyses were conducted separately by sex or adjusted for sex where combined.

Chapter 3. General data quality and methodology issues

Overview of Chapter 3

This chapter covers general issues with respect to quality of the data used in the temporal and spatial analyses and briefly introduces the concepts underpinning a Bayesian approach to statistical analysis. Specific issues are dealt with in later chapters.

Key points

Mortality data

- Mortality data in the UK are nearly 100% complete and almost all have been certified by a medical practitioner or coroner. They are coded using International Classification of Disease (ICD) coding.
- Time trend analyses covering 1950-99 needed to take account of potential effects of changes in ICD coding versions and the 'rule 3' coding interpretation 1984-1992 in England & Wales. ICD9 was used throughout the time period of spatial analyses, but may have been affected by 'rule 3'.
- Many death certificates mention COPD as a contributing cause to the death where it is not the underlying cause of death. However, analyses[17] suggest that – at least for time trends – that trends in deaths with any mention of COPD on the death certificate parallel trends for any mention of COPD.
- Geographical variations in COPD mortality are closely correlated with COPD morbidity measures, so associations of exposure variables with mortality would be expected to give indications of factors affecting COPD morbidity.

Data on COPD risk factors

- Air pollution measures from individual monitoring stations are available for black smoke and SO_2 going back to the 1950s. While measuring techniques have remained the same, the pollution sources have changed and this may affect both the accuracy of the measures and the toxicity, particularly for particulate matter.
- The Taylor Nelson Sofres (TNS) super panel survey contains information on food purchases and, by inference, intakes in the UK from 1991 onwards. It is

available at high geographical resolution, but has not been widely used for epidemiological studies. Comparisons of fruit & vegetable purchases with information from the long-standing National Food Surveys and the National Diet and Nutrition Survey were conducted. These suggested under-estimation in the TNS, but also highlighted possible biases in the comparison surveys. As there was no reason to believe that the under-estimation in the TNS varied geographically, it was judged reasonable to use the TNS data to estimate relative differences in consumption.

Bayesian inference

- The analyses presented in future chapters make use of both frequentist and Bayesian statistical techniques.
- Most epidemiological analyses currently published use statistical techniques based on a frequentist framework. Probabilities in classical analyses should technically be interpreted as frequencies observed in a long run of repeated frequencies and do not take account of prior scientific knowledge or beliefs about the parameter of interest.
- In Bayesian inference, the end results are expressed as the probability of a parameter given the observed data and the statistical model specified takes at least some account of prior scientific knowledge. Prior distributions are specified for unknown parameters of interest (θ), based on scientific knowledge. The likelihood of the observed data (X) given these parameters, denoted as $p(X|\theta)$, relates all variables into a 'full probability model' summarising current knowledge. The posterior distribution $p(\theta \mid X)$ is obtained using Bayes' theorem to obtain probability distributions for unobserved quantities of interest, conditional on the observed data. This is implemented through repeated iterative sampling.

Mortality data

The key outcome used in both temporal and spatial analyses was COPD mortality, with additional information on mortality from lung cancer, influenza, asbestos-related diseases and pneumoconioses. Mortality data in the UK are nearly 100% complete and almost all registered deaths have been certified by a medical practitioner or coroner.[236] As less than 10% of deaths certified by doctors in England & Wales include information from a post-mortem,[236] this may also lead to some inaccuracies in the certified cause of death.

Up to 1992 deaths in England & Wales relate to the year in which the death was registered but from 1993 onwards they relate to the year in which the deaths occurred. However, mortality data used in this analysis relating to 1981 onwards relate to the year of occurrence as data were obtained from the Small Area Health Statistics Unit (SAHSU) in the host department, which has converted data to year of occurrence. These differences involve small numbers of deaths and should not affect the analyses described, but may result in minor discrepancies between totals quoted in this PhD and official figures.

Coding issues

ICD coding revisions

Mortality data are coded using International Classification of Disease (ICD) coding. Versions of this change over time. In the UK, ICD version 6 (ICD6) was used in mortality coding from 1950-57, ICD7 from 1958-67, ICD8 from 1968-78, ICD9 from 1979-2000 and ICD10 from 2001 onwards. These changes would not have affected spatial analyses as these were conducted for 1981-99 with ICD9 in use throughout this period, but could potentially have affected temporal analyses as these were conducted from 1950-1999. Bridge coding exercises are conducted in the year before the change in coding version is implemented using both the new and the old versions to estimate the numbers of deaths coded to each condition. As a result bridge coding exercise conversion factors can be obtained, allowing comparisons between the two different coding versions.

Selection of underlying cause of death

Automated coding to determine the underlying cause of death using the United States automatic coding software was introduced in England and Wales in 1993[236] and in 1996 in Scotland.[257] Prior to this coding was conducted by nosologists (coders experienced in the classification of diseases). Authors conducting death certificate validation studies prior to this have commented on the variable application of coding rules to determine the underlying cause of death by the registering officers.[258,259] Few analyses on the validity of UK death certificates in obstructive lung disease have been conducted in recent years, but these suggest that a label of asthma may be preferentially applied to women with COPD and that some asthma deaths are recorded as COPD and vice versa.[258-260] This should be minimised by the choice of an inclusive definitions of COPD.

Another potential influence is the interpretation of WHO Rule 3 in assigning underlying cause of death. From 1984-1992, England & Wales used a different interpretation of this rule from that used in Scotland (and other countries).[257] In the broad interpretation adopted by England & Wales, 11 conditions were considered to be terminal including pneumonia (of any type), pulmonary embolism, venous thrombosis and embolism, cardiac or hepatic failure and cardiac arrest. If another major conditions was recorded elsewhere on the certificate, rule 3 would be used to select that condition as the underlying cause of death.[236] The chief effects were on the 11 terminal conditions involved and other conditions such as diabetes mellitus, epilepsy and rheumatoid arthritis.[261] Published analyses related to the effect of changing the interpretation in 1993, but the effects in 1993 appear to be generally in the opposite direction and of a similar magnitude to those in 1984 when this interpretation was introduced.[261]

Malignant neoplasms were little affected by the rule-3 change with an estimated 1-2% increase in numbers of deaths due to the change[261] and a 0.3% increase for influenza deaths (Table 3-1, page 69). However, there were non-negligible effects on COPD with an estimated 7% increase for COPD (ICD9 490-496) and effects 1-2% higher on the elderly (Table 3-1, page 69). The effect on pneumoconioses and asbestos-related diseases were not published. Since two of three codes used for asbestos-related diseases were neoplasms (of pleura and peritoneum) it is likely that these were little affected. The impact on pneumoconioses is unclear, but if large one would have expected them to have been mentioned in the ONS review of the impact of rule 3.[261]

Coincident with the introduction of automated coding and the changes in rule 3, OPCS stopped medical enquiries. Here, further information would be sought from the certifying doctor on conditions reported on the death certificate that could only be coded to broad categories (e.g. senility without mention of psychosis ICD9 797) to see if a more definite disease contributed to the death. This would have had only small effects on the conditions considered in these analyses (Table 3-1).

Table 3-1 Effect of change in interpretation of WHO Rule 3, by age, 1984 and of discontinuation medical enquires

(The % amount by which 1993 and 1994 figures would have altered if the 1984 changes in Rule 3 interpretation had continued to be used or if medical enquiries had continued to be used.)

ICD9 code	Condition	% effect of rule 3 interpretation	% effect of rule 3 interpretation by age (where available)		% effect of medical enquiries
		All ages	Under 75 years	75+ years	
162	Malignant neoplasm of trachea, bronchus and lung	1	1	2	1
487	Influenza	0.3 male 0.3 female			2.3 male 1.3 female
490-2	Bronchitis and emphysema	6	5	7	
490-493	Bronchitis, emphysema and asthma	6.1 male 6.1 female			0.4 male 0.4 female
496	Chronic airways obstruction	8	6	10	
490-496	Chronic obstructive pulmonary disease	7.1 male 7.1 female			0.3 male 0.3 female

Source: Mortality statistics 1996, DH2 no.21[261]. Tables B and C, pages xxv to xxxiii

Use of COPD mortality as an outcome measure

COPD as an indicator for COPD burden

The diagnoses on death certificates will reflect the disease incidence, the case fatality rate and the current diagnostic fashion (for example, whether obstructive airways disease in an elderly person is labelled as asthma or as chronic obstructive pulmonary disease).[262] Diagnostic fashion could be partly accounted for by using as inclusive a definition of COPD as possible.

Relationship between underlying cause of death and any mention of COPD

The mortality analyses presented have used underlying cause of death, as mention of COPD in other areas of the death certificate (multiple-cause coding) was not available nationally until 1993. Multiple cause coding analyses of both England & Wales for 1993-1999[17] and the Oxford record linkage study for 1979-1998 suggested that COPD on the death certificate coded as the underlying cause formed approximately 60% of all mentions on the death certificate. This contrasts with lung cancer, which is usually

coded as the underlying cause when present on the death certificate.[239] However, both the Oxford[239] and national[17] studies suggested that time trends in any mention of COPD on the death certificate closely paralleled those for COPD as the underlying cause of death.

Relationship between COPD in life and mention on the death certificate

Not all patients with OLD in life will have OLD mentioned on the death certificate – for example, because it did not contribute to the final illness or because of poor recording of co-morbidities. The Tucson Epidemiologic Study of Airways Obstructive Diseases[222] suggested that patients with spirometric evidence of airways disease in life only had airways disease reported on the death certificate in 13% (9 of 71 cases), but that the percentage was much higher (77%) in cases with moderate to severe disease (patients with $FEV_1\%<50$). A twenty-two year follow-up of 5,542 adults in the first National Health and Nutrition Examination Survey (NHANES I) found that 47.7% of patients with severe COPD at baseline had COPD listed on the death certificate and 23.1% had COPD as the underlying cause of death.[263] However, COPD diagnosis was based at spirometry at baseline only – percentages would almost certainly have been higher if information on COPD progression during follow-up had been available.

Relation between geographical variations in COPD mortality, morbidity and prevalence

Geographical variations in COPD mortality are likely to reflect the variations in COPD prevalence and morbidity. A study examining the relationship between COPD mortality, hospital admissions, general practice consultations and symptoms in 1991-1995 in England found a good spatial correlation at health authority level,[15] implying that areas with high COPD mortality would also have high levels of hospital admissions and GP consultations. This suggests that geographical associations environmental exposures in the spatial analyses with mortality would also be seen with morbidity.

Air pollution data

The chief pollutants considered in this PhD are black smoke or particulates and sulphur dioxide as these have been major constituents of air pollution over a long period of time, with documented chronic effects on health in the USA.[41,264,265] One advantage of choosing these pollutants is that exposure measures are available over a very long time

span and methodology used for recording black smoke and sulphur dioxide in the UK has not changed over 55 years of recording. However, differences in the composition of air pollution may affect the comparability of results over time.

Particles arise from both natural sources and man-made sources. Examples of natural sources include sea salt, erosion of soil and rocks by wind, forest fires, pollen grains and fungal spores. Sources of man-made particles include combustion sources (e.g. power stations, vehicle emissions), industrial processes and working of soil and rock. Particles may be primary – released directly into the air – or secondary, formed in the atmosphere by the chemical reaction of gases. It follows that areas with different types of land use (e.g. industrial versus farming) will have qualitative differences in the types of particles to which people are exposed. Particles from different sources are likely to have different biological effects. For example, PM_{10} particles from an urban source (Cardiff) were found to damage DNA *in vitro* and to be 100 times more bioreactive than diesel exhaust particles (a major component of PM_{10}) alone, which in turn were more reactive than pure carbon.[152] However, outside of special studies, information is usually limited to the amount of particulate matter in the air.

The most commonly used measures of particulates used in current epidemiological research are PM_{10} and $PM_{2.5}$. These respectively relate to amount of fine particles with average diameter less than 10μm (more strictly particles that pass through a size selective inlet with 50% efficiency cut-off at 10μm aerodynamic diameter) and to ultrafine particles with average diameter or 2.5μm. Fine particles <10μm can penetrate into the airways; particles < 2.5μm can penetrate deeply into the lung. These measures were not widely available prior to the 1990s.

Black smoke is another measure of particulate matter in the air. The advantage of this measure is that measurements are available nationally going back to the 1950s and earlier for selected monitoring stations. It was designed chiefly to measure particles formed from combustion of fossil fuels, especially coal. The measuring technique involves drawing air at a specified rate through filter paper. The resulting blackness of the filter paper is determined using a reflectance measure.[246] The measures are calibrated on a scale originally developed for coal combustion that predominantly resulted in black particles. The major source of particulates now is vehicle emissions rather than coal burning and a substantial proportion of particles in present day

emissions are lighter in colour than those from coal combustion.[266] As a result, black smoke measures may underestimate current particulate levels – although the level of underestimation is changing as an increasing of vehicles use diesel fuel, which produces black combustion particles readily detected using the black smoke method.

Current measurements of particulates deal only with the mass and not with the qualitative composition.[246] The chemical composition of particles has changed over time, with present-day particles in urban areas dominated by combustion engine exhaust particles rather than the coal combustion particles of the 1950s and 1960s.[266] The situation is further complicated by the adsorption of chemical species onto the surface of primary particles and the formation of secondary particles such as ammonium sulphate and ammonium nitrate in the atmosphere.[244]

Sulphur dioxide is primarily formed as a result of combustion processes; concentrations in urban areas chiefly reflect the extent to which coal and fuel oil is burnt.[266] Vehicles are generally not a significant source of SO_2 emissions, although some SO_2 is emitted from diesel vehicles (lorries, buses and some cars). SO_2 measures are conducted using the 'net acidity' method, more descriptively termed the twenty-four hour 'bubbler'.[266] Filtered air is 'bubbled' through a dilute solution of hydrogen peroxide for 24 hours. Any SO_2 reacts with the hydrogen peroxide to form sulphuric acid and analyses are performed to determine the amount of acid present. Although sulphur dioxide is the major source of acid in the air, this measure really is a measure of air acidity rather than sulphur dioxide per se. While SO_2 was the major component of acidity in previous years, some of the acidity will now originate from nitrous oxides from vehicle emissions. The measure will also be affected by any alkali in the air, which may be important in some rural areas.

Taylor Nelson Sofres (TNS) data on fruit & vegetable purchases

Dietary data relating to fruit & vegetables were used in the spatial analyses. The dataset used was the Taylor Nelson Sofres (TNS) Super Panel database, which was the only database with wide enough geographical coverage to permit a spatial analysis of the UK.

TNS are a market research company who run an ongoing panel study of over 10,000 households per year[267] to provide information on food and household items brought into the home, chiefly used for information on consumer purchase preferences for retailers. Recruitment uses stratified sampling so that households are representative of the population structure of Great Britain. All participating households are supplied with a hand-held barcode scanner. All foods purchased and brought home for consumption are barcode scanned. Items sold loose (e.g. some fruits, vegetables, bakery items) are scanned from a booklet of relevant barcodes then weighed and, additionally, price is entered. Quality control procedures include regular monitoring of data capture, panel continuity and bar-code matching and identification. There is an acknowledgement that declines are seen in the recording of smaller size shopping trips over time,[267] but TNS estimate that about 70% of total household food intake is captured in their database.[253] Separate smaller panel studies are also conducted to provide information on foods consumed outside the home and impulse food purchases, but these were not available for this study.

The TNS Super Panel database has only been used in epidemiological research once before, when it was investigated as a potential database to track novel food consumption in the UK.[253] Prior validation was conducted via a comparison of macronutrient intake with the National Food Survey (NFS)[253] and a comparison with the National Diet and Nutrition Survey (NDNS) is underway (personal communication, Claire Robertson). As this database has not been used to estimate national or sub-national fruit and vegetable consumption previously, comparisons with the NFS and NDNS were conducted to explore the validity of using the database for this purpose in these PhD analyses.

Exploring the validity of a fruit & vegetable purchase variable from the TNS Super Panel database

Description of the TNS database

The TNS Super Panel (household intake) database contained information on food purchases for 1991-2000 made by 33,177 households, with a median of three persons per household (Table 3-2, page 74). Households entered and left the survey over time, but the database contained a total of 4,550,087 weeks of purchase data, with a mean of 137 weeks and a median of 64 weeks in the study per household (Table 3-2, page 74).

Purchases were not recorded every week by every household, but the median was 46 weeks in which food was recorded as being purchased – i.e. nearly a year of data. However, 17% of households (5763) were in the study for less than 12 weeks. Additionally, many households had large numbers of missing weeks – 13,171 (40%) of households had more than one out of every five weeks missing purchase data and 7,842 (24%) had more than one out of every three weeks with missing purchase data. The reasons for zero purchase weeks were not recorded and are likely to reflect a combination of supermarket shopping patterns (chiefly recorded by the panel), inconsistent recording, and holidays or other absences from home.

Geographical coverage was good, with data for 453 districts out of a possible 459 (using 1991 district boundaries), exclusions being mainly in remote areas (Isles of Scilly, Skye and Lochalsh, Sutherland, Orkney, Shetland, Western Isles).

Table 3-2 Descriptive statistics of TNS 1991-2000 dataset by household

	Mean	Median	Min	Max	10th percentile	25th percentile	90th percentile
Number of weeks in study	137	64	1	522	6	19	435
Number of weeks in which purchases were recorded	111	46	1	522	4	14	353
Weeks in study / purchase weeks	1.4	1.2	1	41.5	1	1	2
Households per district	73	58	1	573	15	28	142
Households per district where purchases were recorded on average at least two out of every three weeks.	57	45	1	433	11	21	106
Households per district where purchases were recorded on average at least 4 out of 5 weeks.	44	35	1	337	8	16	88
Average no. persons per household	3.1	3	1	12	1	2	5

Creation of a fruit and vegetable variable

With the help of a nutritionist, a variable representing fruit and vegetable purchases was created. This was expressed in terms of gram intake per day in line with most nutritional studies and also as portions per day using the Food Standards Agency[268] definition of a portion as 80g (but 125ml – approx 125g – for fruit juice). There is no standard nutritional definition of what items should be included in fruit and vegetables and different studies have used different definitions. No consensus was found from previous studies as to which fruit or vegetables confer greatest benefit in prevention of respiratory disease. For example, studies have found benefits from fresh fruit consumption,[269,270] fruit and fruit juice consumed in winter,[159] salads and green

vegetables[270] and sub-groups of solid fruit(apples, pears), citrus fruits and other fruits (soft, preserved and juices).[162] The definitions used are seldom given and sub-group analyses vary so the fruit and vegetable grouping used in this study was chosen to be broad.

The fruit & vegetable variable consisted of the summed weight purchased of fresh and prepared fruit and vegetables, including legumes, fruit juice, frozen and canned items and prepared fruit salads. Food purchases representing 176,724 different food items (some differing only by weight or quantity) are pre-coded in the TNS database into one of 186 groupings of similar types of products called validation fields. To simplify coding and extraction of the data the pre-set validation field groupings e.g. "fresh soft fruit", "fresh vegetables – beans etc." were used rather than individual food items. These represented 24 out of the 186 validation field groupings. Potatoes were excluded. Foods where fruit or vegetables were an ingredient – chiefly prepared foods such as tomato soup, baked beans, muesli, fruit cakes, baby foods etc. – were also excluded, representing a further 77 validation food groupings. Due to the use of validation fields rather than individual food items, a small number of eligible foods were excluded. For example, baked beans (containing legumes and tomatoes) were excluded as the weight purchased could not easily be separated from other types of items (e.g. canned ravioli). Exclusion of these items will have resulted in an under-estimate of total fruit and vegetable purchases.

The variable created relates to purchases not to food actually consumed and does not take account of wastage due to food preparation, spoilage or leftovers. This will result in an over-estimate if purchases are used to estimate fruit & vegetable consumption. Very few studies are available to quantify wastage, but one study from 1980[271] suggested that approximately 5-6% of foods may be wasted and a figure of 10% is used by the National Food Surveys when calculating nutrient intake.[272] Food wastage may vary by household composition, with larger households wasting less per person than smaller households and adults wasting more in absolute terms than children.[271] These differences have not been taken into account in the analyses.

Comparisons of the TNS fruit & vegetable data with other national dietary survey data

Comparisons with the National Food Survey (NFS)

Methods

Comparisons were made with the National Food Survey (NFS) . The NFS uses a stratified random sample of approximately 6,000 households per year. A questionnaire and food diary are compiled for a one week period – a rolling sample is used to attempt to capture seasonal variations. Both surveys look at the weight of food brought into the home for home consumption and do not adjust for age and sex of household members.

Variables compared were average TNS weights of fruit and vegetables purchased per person per week for 1991-2000 and NFS average weights brought into the household for consumption per person per week. As the National Food Survey groupings had changed slightly over time, it was not possible to combine across years so comparisons were made for 1992, 1997 and 2000. Items included in 'fruit and vegetables' were not clearly described in the NFS publications, but excluded potatoes. Weights for purchases of single food items such as pears, apples, citrus fruits, mushrooms, root vegetables, fruit juices were also extracted from the TNS for comparison with the NFS in the expectation that these would be less affected by coding differences.

NFS data relate to a single week in which food data were recorded. TNS data was averaged over many weeks of purchases, but most households had a number of zero purchase weeks. Some of these may have been holiday or business absences, but households with a large number of these may have been households less likely to be compliant with the survey, particularly where weights for loose fruit and vegetables and non-barcoded items from non-supermarket retailers had to be recorded. Because of this, comparisons were also made between NFS data and TNS data excluding households with more than one in five or one in three weeks with zero purchases. Additionally comparisons were made between NFS data and TNS data, where it was assumed that TNS purchases in zero purchase weeks would have been similar to the weeks for which data were recorded.

The number of households recording each food item was also calculated. For the NFS, this was not published but could be inferred by multiplying the published percentage of

households consuming this type of food during the purchase week by the published total number of households taking part in the survey.

Results

The initial comparisons of the TNS data (with no exclusions and no weightings) compared with the NFS (Table 3-3, page 78, sixth column) suggested that the TNS for total fruit and vegetables was about 60% of the weights suggested by the NFS. Some of the short-fall between the TNS and NFS may have been due to coding differences. However, specific food items that might be less prone to suffer from coding differences still suggested a short-fall in the TNS of 20% to 50% that of the NFS (Table 3-3, page 78).

Some of the short-fall between the TNS and NFS may have been due to the inclusion of zero purchase weeks – weighting resulted in an increase of around 15% in the comparisons with the NFS (Table 3-3, page 78, column 5). Excluding households with large numbers of zero purchase weeks also increased the percentage comparison with the NFS by about 15% as compared with no exclusions or weighting alone. Combining weighting with exclusions increased the percentages further (Table 3-3, page 78, column 2) putting the TNS weights for total fruit and vegetables at about 75% of that in the NFS.

The percentage of households recording fruit and vegetable consumption/purchases showed large differences between the NFS and TNS (Table 3-4, page 78). Only 79% of households in the NFS had recorded any fruit and vegetable consumption during their week in the study, compared with 100% of households in the TNS at some point during their time in the super panel. For single food items or food groups e.g. citrus fruits or mushrooms, the percentage of households recording this item in the NFS was low (10% to < 50%) and much lower than in the TNS (60% to 90% depending on food group) (Table 3-4, page 78).

Table 3-3 Fruit and vegetables brought into the house for consumption in grammes per person per week using TNS data 1991-2000 with differing exclusion criteria compared with National Food Surveys for 1992, 1997 and 2000

	TNS Super Panel database						National Food Survey		
	TNS weighting for zero purchase weeks and excluding those with >1/5 weeks with zero purchases (% NFS 1997)	**TNS with no weighting for zero purchase weeks but excluding those with >1/5 weeks with zero purchases (%NFS2000)(%NFS1997)(%NFS1992)**	TNS weighting for zero purchase weeks and excluding those with >1/3 weeks zero purchase weeks (% NFS 1997)	TNS weighting for zero purchase weeks (% NFS 1997)	Original TNS dataset – including zero purchase weeks and no weighting (% NFS 1997)		NFS 2000[273]	NFS 1997[250]	NFS 1992[274] (published in oz, but converted to g by factor of 28.35)
Total fruit & veg	1611 (76%)	**1497 (68%) (70%) (67%)**	1558 (73%)	1509 (71%)	1226 (58%)		2197	2130 (2076 if 39g 'other veg products' and 15g nuts are excluded)	2222
Fresh fruit juice	226 (82%)	**210 (69%) (76%) (105%)**	225 (81%)	225 (81%)	179 (65%)		303(ml)	277	221
Fresh apples	135 (78%)	**125 (69%) (72%) (67%)**	128 (74%)	119 (69%)	99 (57%)		180	173	187
Fresh pears	35 (76%)	**32 (70%) (70%) (89%)**	32 (70%)	30 (65%)	25 (54%)		46	46	36
Fresh citrus fruit	94 (70%)	**87 (56%) (64%) (66%)**	88 (64%)	83 (61%)	68 (50%)		155	137	132
Fresh mushrooms	29 (81%)	**27 (75%) (75%) (84%)**	28 (78%)	28 (78%)	22 (61%)		36	36	32
Root vegetables	178 (109%)	**165 (92%) (101%) (98%)**	170 (104%)	160 (98%)	132 (81%)		180	163	169

Total fruit and vegetables for TNS dataset includes fruit juice and 'chilled fruit/veg'. For NFS, this is sum of "total vegetables excl. potatoes and potato products" and total fruit
Fresh citrus fruit was the combination of "oranges" and "other citrus fruit" in the NFS

Table 3-4 Number of households recording fruit & vegetables brought into the house for consumption in the TNS super panel 1991-2000 and in the National Food Surveys for 1992, 1997 and 2000

	TNS excluding those if >1/5 weeks zero purchases	**TNS excluding those with >1/5 weeks with zero purchases**	TNS excluding those with >1/3 weeks zero purchase weeks	Original TNS dataset		NFS 2000[273]	NFS 1997[250]	NFS 1992[274]
Total number of households in study	n=20006	n=20006	n= 25695	n=33177		n=5974	n=6065	n=7556
Total fruit & veg	19907 (100%)	19907 (100%)	25578 (100%)	33035 (100%)		5078 (85%*)	4731 (78%)	5969 (79%)
Fresh fruit juice	16582 (83%)	16582 (83%)	21404 (83%)	27683 (83%)		1852 (31%)	1820 (30%)	2116 (28%)
Fresh apples	17592 (88%)	17592 (88%)	22523 (88%)	28856 (87%)		2390 (40%)	2608 (43%)	3476 (46%)
Fresh pears	12853 (64%)	12853 (64%)	16310 (63%)	20171 (61%)		836 (14%)	910 (15%)	1058 (14%)
Fresh citrus fruit	16781 (84%)	16781 (84%)	21511 (84%)	27426 (83%)		1255 (21%**)	788 (13%)	1360 (18%)
Fresh mushrooms	15804 (79%)	15804 (79%)	20304 (79%)	26005 (78%)		1613 (27%)	1698 (28%)	2116 (28%)
Root vegetables	18191 (91%)	18191 (91%)	23364 (91%)	29974 (90%)		538 (9%)	607 (10%)	756 (10%)

*Not supplied combined. 85% is the higher % of 'total fruit' (79%) and 'total other vegetables' (85%) **Not supplied combined. 21% is the higher % of 'oranges' (11%) and 'other citrus fruit' (21%)

Regional comparisons with the National Food Survey 1997

Regional comparisons of TNS data with the National Food Surveys are difficult as regions are defined differently. However, a comparison was attempted using the NFS 1997. This suggested a reasonably similar regional pattern, with highest consumption in the southern part of England and lowest in the north of England and Scotland. In the NFS consumption was 22% higher in the South-East than in Scotland, in the TNS it was 27% higher. However, there were some differences with the TNS in rankings for regions in the middle, particularly Wales and East Anglia (Table 3-5).

Table 3-5 Average regional purchases of fruit and vegetables by region for the TNS 1991-2000 and NFS 1997, ranked with highest region first

TNS fruit, vegetable and fruit juice purchases 1991-2000					National Food Survey 1997		
Region name	Average district purchase of fruit/vegetable/fruit juice in g per week	Percentage of maximum region purchases of fruit/vegetable/fruit juice in g per week	Average district purchases of portions of fruit/vegetables (80g = 1 portion) and fruit juice (125g = 1 portion) per person per day		Government office region name	Average fruit & vegetables (excl. potatoes) brought into the house for consumption in g per person per week	Percentage of maximum region average fruit & vegetables (excl. potatoes) brought into the house for consumption
South East	1617	100%	2.7		South-East	2586	100%
South West	1547	96%	2.6		Greater London	2583	100%
East Anglia	1403	87%	2.4		South-West	2558	99%
West Midlands (Region)	1403	87%	2.4		Wales	2505	97%
East Midlands	1394	86%	2.4		Yorkshire and the Humber	2397	93%
Yorkshire & Humberside	1386	86%	2.4		East Midlands	2358	91%
North	1317	81%	2.2		West Midlands	2291	89%
Wales	1305	81%	2.2		Eastern (East Anglia)	2266	88%
North West	1289	80%	2.2		North-West and Merseyside	2221	86%
Scotland	1274	79%	2.1		North-East	2200	85%
					Scotland	2121	82%

Reasons for differences between TNS and NFS fruit & vegetable amounts

Both TNS and NFS related to food brought into the house for consumption so should be reasonably comparable, although it is not clear whether NFS households allow for wastage. There were differences in the way that fruit & vegetables were defined, which may have led to an under-estimate in the TNS (but this may be offset if wastage is allowed for in the NFS but not TNS).

The TNS households were found to have lower weights for total fruit & vegetables and for specific items than the NFS. However, comparisons were consistent with the TNS

estimate that about 70% of total household food intake is captured in their database[253] and with a previous analysis[253] comparing macronutrient intake in the TNS with the NFS, where total energy intakes per person was approximately 75% of those recorded in the NFS.

Excluding TNS households with large numbers of missing weeks, or weighting to allow for missing weeks decreased the differences between TNS and NFS. Another factor to be considered is that the TNS had larger numbers of families with children than in the NFS, which would tend to lower TNS averages (as children eat less).

Some entries in the TNS had food codes but no weights entered and hence could not be included in the estimates. These could represent loose vegetables or garden produce, which could also lower averages. However, on further analysis these were a small percentage of the total and were unlikely to have to have produced the larger difference seen. For example, for apples only 416 of approx. 29,000 households with recorded apple purchases had scanned in apples but had occasional missing weights.

However, some of the differences may be due to the methodology used in the NFS, which may have resulted in some bias. For example, the NFS survey was for a single week and may therefore have suffered from small numbers of purchases or bias from different eating habits during the week of the survey. The TNS relied on more than one week of data (median 46 weeks in which purchases were actually made) and may be more representative than the NFS figures, particularly for specific food items based on small numbers of consuming households (Table 3-4, page 78). The comparison of percentages of households consuming fruit or vegetable items raised questions about the way in which the NFS averages had been calculated – whether over all households or all consuming households. The methodology is not clearly set out in NFS publications and standard deviations are not presented. If averages were calculated over all households, for certain food items some NFS households must have had very high levels of consumptions. For example, for pears the NFS average for 1997 and 2000 was 50g. If only 20% of households are consuming, the average for consuming households in the NFS must have been 250g (if the denominator is all households). In the TNS 1991-2000, purchases were 25-35g depending on adjustments based on 60% of households consuming so the consuming households were consuming on average approx 40-50g.

Comparison of TNS fruit & vegetable totals with the National Diet and Nutrition Survey (NDNS) 2000-2001

The National Diet and Nutrition Survey was conducted in 2000-2001 in 1,724 UK adults aged 19-64 years.[252] In this survey individuals are asked to weight whatever they consumed (incomplete answers are excluded, eating out is included unlike the NFS) and there is a post-survey dietary interview. There are minor differences in the way in which fruit & vegetables are defined and the NDNS does not include children, which would tend to increase observed consumption per person compared with the TNS (which does include children). In published analyses, results are only given for males and females separately.[252] Results are likely to be lower than for the TNS as they are for food actually consumed and, unlike the TNS, have already taken account of wastage.

A preliminary analysis of the NDNS data (personal communication Jennifer Hurley), using 80g as an indicator of a portion, suggested that the average consumption in the NDNS was approximately 2.5 portions per person per day (excluding fruit juice and composites such as fruit or vegetables forming part of ready meals, apple pies etc). In comparison, the TNS data 1991-2000 suggested 2.3 portions per person per day *purchased*, excluding households with more than one in five weeks of zero purchases and excluding fruit juice). If all the dataset was used, there were 1.8 portions per day in the TNS. This still suggests a short-fall between the TNS and NDNS, but much less marked than that seen with the NFS.

Is the TNS valid for use in a geographical analysis?

Comparisons of the TNS with the NFS, which should have the most similar methodology, and with the NDNS, which surveys consumption (i.e. wastage has been taken account of) suggested that the TNS had values 10-30% lower when looking at total fruit and vegetables (NFS and NDNS) or individual food items. However, regional comparisons using the TNS and NFS gave reasonably similar results – at least for lowest and highest regions.

There are problems with making these comparisons as each survey was carried out in different ways and the definitions of fruit & vegetables differed. This may be partly responsible for the differences. Additionally, those taking part in a national nutrition survey may be biased towards those more likely to eat fruit & vegetables and those

participating may have eaten more healthily during the week of the survey such that the single weeks surveyed in the NFS and NDNS may be unrepresentative of population consumption. There are also issues with the TNS data including under-reporting of purchases. Also, the full details of the quality control mechanisms in the TNS, a commercially available database, are not published.

The most crucial point for this geographical analysis is whether the differences between the TNS and more widely used surveys or between the TNS and 'true' consumption are likely to vary geographically. This is probably unlikely as the main variation would be expected to be due to small sample numbers and to individual household variations, with under-reporting likely to occur in a random fashion across the country. This is partially supported by the previous analysis of macronutrient intake,[253] which, although found to be 75% of that recorded in the NFS, had a similar geographical variation. Systematic differences at an area level might occur if induction into the TNS and quality control is carried out more thoroughly in some areas than others (for example, because of a more motivated co-ordinator), but it seems unlikely these would vary with COPD mortality. It therefore seems reasonable to use the TNS data to estimate spatial distribution of fruit and vegetable consumption, subject to the usual limitations of inference for ecological studies.

Socio-economic status of area

Carstairs score was chosen as the indicator of area-level deprivation used in the spatial analyses. This is a composite index based on an unweighted combination of four standardised Census variables: proportion of male unemployment, proportion of people in households without access to a car, proportion of people in social classes IV and V and proportion of people living in overcrowded households.[255] It is calculated at enumeration district (average 400 population) and ward level (average 5500 population). Composite deprivation indices are generally preferred to single item measures as the latter are more prone to bias (e.g. overcrowding may be more common in urban areas and thus a less good measure of deprivation if used as a single item).

A number of composite deprivation indices are available such as Carstairs, Townsend or Jarman scores. These generally show high correlations ($r > 0.8$) between each other[275] and with all-cause mortality ($r = 0.7$), but lower correlations with all-cause mortality in

older individuals (>65 years, r = 0.5) and with hospital admissions. The Jarman (UPA) score shows weaker correlation with health events than with health service use.[255]

Individual level information on socio-economic status of individuals was not available for this study, but deprivation indices should reflect the average levels of deprivation of individuals in the area and may provide additional information. While some studies have found few effects of area-level deprivation once individual-level deprivation has been taken into account,[276] other studies have suggested that there are contextual effects of area-level deprivation that are additional to or interact with those of individual socio-economic status.[277] These contextual effects may be due to unmeasured individual effects, but could also represent physical or social characteristics of areas where poorer people live.

Ward-level Carstairs score has been found to be strongly associated with individual-level smoking status (current smoker or non-smoker).[150] Lung cancer mortality will be used as an indicator of cumulative (past) smoking in the spatial analyses as area-level smoking data are not available, which will affect the interpretation of any association between deprivation and area-level COPD mortality.

Introduction to Bayesian inference

Bayesian statistical methods were chosen for some of the temporal and spatial analyses because they allow the incorporation of prior 'structural' assumptions about the temporal and/or spatial similarities (i.e. smoothness) of the risk parameters in the model.

Frequentist analyses

Most epidemiological analyses currently published use statistical techniques based on a frequentist framework (also referred to as 'classical' analyses in this PhD). Probabilities in classical analyses should technically be interpreted as frequencies observed in a long run of repeated frequencies. Parameters are considered as fixed, non-random quantities and probability statements concern the data. The 95% confidence intervals using classical techniques technically have a rather non-intuitive interpretation that for 95% of confidence intervals obtained from repeated sampling from the population, these limits will include the 'true' population value (but it is not known which 95% confidence intervals).[278] However, results from frequentist analyses

are often –strictly incorrectly – interpreted in a Bayesian way: that there is a 95% probability that the population mean lies within the confidence interval.[278] Assumptions in classical analyses may not be made explicit – for example, assumptions made about the distribution of the data – and the interpretation of the results will usually be based on existing scientific knowledge that has rarely been incorporated into the statistical formulation of the model.

Bayesian inference

In a Bayesian analysis, the end results will be expressed as a probability statement: the probability of a value or distribution of values for a parameter (or unobserved data) given the observed data, and the prior assumptions based on scientific judgment incorporated into the statistical (full probability) model. In a Bayesian analysis, it is correct to interpret 95% confidence intervals as being 95% confident that the 'true' population mean (or parameter value etc. depending what the model is estimating) lies within those confidence intervals. The choice of statistical cut-offs, referred to as credible intervals in Bayesian analyses rather than confidence intervals, is more varied in reports of Bayesian analyses (80%, 90%, 95%) as appropriate to the analysis, compared with the usual 95% cut-offs of frequentist analyses. However, cut-offs of 80% are common reflecting that prior beliefs have to some extent been incorporated into the model as well as the role of chance – only the role of chance is evaluated in probability statements from frequentist analyses.

Bayesian analyses have three main steps[279]:

(i) Setting up a full probability model, which is a joint probability distribution for observable and unobservable quantities, based on existing scientific knowledge – i.e. at least some prior assumptions can be incorporated into the statistical model.

(ii) Conditioning on observed data – calculating and interpreting the posterior distribution, which is the conditional probability distribution of the unobserved quantities of interest, given the data

(iii) Evaluating the fit of the model and the implications of the resulting posterior distributions – this includes evaluating the effect of modelling assumptions and considering whether results are reasonable.

The full probability model includes observable parameters (the data) and model parameters (unknowns – depending on the problem, these may include statistical parameters, missing data, mismeasured data).

- The prior distribution $p(\theta)$ expresses information or uncertainty available at the start of the study about unknown quantities (variables) by means of a specified probability distribution.
- The likelihood $p(X|\theta)$ relates all variables into a 'full probability model' summarising current knowledge.
- The posterior distribution $p(\theta \mid X)$ is obtained using Bayes' theorem to obtain probability distributions for unobserved quantities of interest, conditional on the observed data.

Bayes' theorem

Bayes' theorem is a mathematical theorem attributed to an 18th century clergyman, the Rev Thomas Bayes (1702-1761), which gives the conditional probability distribution of a random variable A given variable B in terms of the conditional probability distribution of variable B given A and the marginal probability of A alone. This can be written as

$$p(A|B) = \frac{p(B \mid A)p(A)}{p(B)} \qquad \textbf{Equation 3-1}$$

Using Bayes' theorem, it is possible to show that the posterior distribution is proportional to the product of the prior distribution and the likelihood. Using the notation given at the top of this page, this is written as:

$$p(\theta \mid X) \propto p(\theta)\, p(X|\theta) \qquad \textbf{Equation 3-2}$$

The joint probability distribution of unobserved parameters θ and the observed data X can be written as a product of the (specified) prior distribution of the unobserved parameters $p(\theta)$, and the probability distribution of the observed data X given the unobserved parameter $p(X|\theta)$:

$$p(\theta,X) = p(\theta)\, p(X|\theta) \qquad \textbf{Equation 3-3}$$

Bayes' theorem as set out in Equation 3.1 is used to infer the result of interest, the posterior distribution $p(\theta \mid X)$ (the probability distribution of the unobserved parameters of interest, given the observed data). Substituting θ for A and X for B in Equation 3.1 gives

$$p(\theta \mid X) = \frac{p(X \mid \theta)p(\theta)}{p(X)} \qquad \textbf{Equation 3-4}$$

where the probability distribution of X (for continuous X) is given by

$$p(X) = \int p(\theta)p(X \mid \theta)\, d\theta \qquad \textbf{Equation 3-5}$$

(Where θ is categorical, Equation 3.5 becomes the sum over all possible values of θ). Therefore, the full expression of interest can be written:

$$p(\theta \mid X) = \frac{p(\theta)p(X \mid \theta)}{\int p(\theta)p(X \mid \theta)\, d\theta} \propto p(\theta)p(X \mid \theta) \qquad \textbf{Equation 3-6}$$

As the marginal probability distribution of the observed data $p(X)$ in Equation 3.5 does not depend on the parameter θ, it can be thought of as a constant and can therefore be absorbed into the proportionality constant on the right hand side of Equation 3.6.

Bayesian analyses in practice

In most realistic models, the posterior distribution $p(\theta \mid X)$ cannot be evaluated directly, but using Bayes' theorem and information about the prior distributions of parameters and the observed data, it can be derived through repeated iterative sampling. Analyses typically use many thousand samples to provide a valid estimate of the sample mean and range, making use of high speed computing.

With iterative sampling, the choice of starting parameters is important – if these are relatively close to the 'true' posterior distribution, a steady state is usually reached quite quickly, with subsequent sample values oscillating within a similar range and not drifting off. If starting values are far from the 'true' posterior distribution, it may take such a long time to reach steady state that for practical purposes steady state is not reached. Posterior distribution values are estimated from samples taken after steady state is reached.

The most commonly used Bayesian software is WinBUGS[280] developed jointly by the MRC Biostastistics Unit, Cambridge and Imperial College London, and is used in the spatial analyses of this PhD. The temporal analyses made use of purpose-written Bayesian software for age-period-cohort analyses called BAMP.[281]

Section II

Temporal variations in COPD mortality in England & Wales in relation to the 1956 Clean Air Act

Chapter 4 Methods

Chapter 5 Results

Chapter 4. Temporal variations in COPD mortality in England & Wales in relation to the 1956 Clean Air Act: Methods

Overview of Chapter 4

This part of the PhD aimed to investigate time trends in COPD mortality and air pollution levels in urban and less urban areas of England & Wales over the second half of the twentieth century in relation to the 1956 Clean Air Act, using lung cancer mortality to infer levels of cigarette smoking. This chapter describes the methods used to investigate hypotheses 1-4 (see Section I, Chapter 1).

Creation of datasets

- Data to the level of aggregation needed were available for England & Wales.
- COPD and lung cancer mortality and population counts for 1950-1999 were obtained from Office for National Statistics (previously the Office for Populations, Censuses and Surveys and the Registrar General's Office). Data were obtained for conurbations, Greater London and non-conurbations.
- Inconsistencies in the data potentially leading to artefactual changes in rates were identified. These chiefly related to changes in geographical boundaries and to coding e.g. ICD version changes, variations in published data over time.
- Information on air pollution trends over time were obtained from various published sources. Descriptive analyses were conducted on black smoke and SO_2 data from 13 long-running air pollution monitoring sites that had been in operation and using similar measuring techniques between 1956 to 1985.

Statistical analyses

- Descriptive analyses of mortality data were conducted.

- A Bayesian age-period-cohort method was used to identify timing points of age, period and cohort effects. The timings were compared with those predicted according to the initial hypotheses and by potential artefactual influences on rates such as ICD coding changes.
- A modified classical age-period-cohort Poisson analysis with an interaction term for degree of urbanisation was used to quantify relative risks of period and cohort effects in conurbations and Greater London compared with non-conurbations.

Data

Mortality and population data

Numbers of deaths from COPD and home population figures were obtained for 1950-1999 for England and Wales by ten year age band from 15-24 to 75+ years, split into three area types: (i) metropolitan counties excluding Greater London ('conurbations'), (ii) Greater London and (iii) areas outside conurbations or Greater London ('non-conurbations'). All data originated from the Office for National Statistics (previously called the Office for Populations, Censuses and Surveys), but data for 1950-1980 were obtained from published volumes, while data electronically held within the Dept of Epidemiology & Public Health of Imperial College London were used for 1981-1999. Numbers of deaths from lung cancer were obtained for the same geographical areas and time periods, as a proxy for cumulative smoking levels.

Data from published Registrar General for 1950-73 and Office of Populations, Censuses and Surveys DH5 series for 1974-80 were manually double-entered for COPD, lung cancer and population by area, age-group and sex for each year. As a secondary check, graphs of rates by age and sex for each condition were examined and any outliers or anomalies were checked with original source data and corrections made if necessary. Numbers of deaths for non-metropolitan counties were available directly in published data for 1950-54 but inferred by subtraction of metropolitan county data from England & Wales published data for 1955-80.

Total numbers of deaths from influenza were also obtained for England & Wales (but not sub-nationally) to be able to identify peak influenza years, as high influenza activity is known to be associated with increases in death attributed to COPD[282] or respiratory disease.[283]

While the choice of International Classification of Disease (ICD) coding, geographical units and age-groups could be flexible for the electronically held data (1981 onwards), there were limitations to the published information available for geographical areas from 1950-80 and there were a number of changes to formats over time that may have affected consistency of the data. These are briefly described next.

Bridge-coding conversion factors

The ICD code versions used were ICD6 (1950-57), ICD7 (1958-67), ICD8 (1968-78) and ICD9 (1979-1999). Data were not adjusted for ICD changes using bridge-coding factors derived from dual coding in the year before a change for a number of reasons:

(i) While these factors may be useful in the years immediately following a change, their validity over long periods of time has not been established and may result in distortions to the data.

(ii) Descriptive analyses did not show evidence of step changes in rates in these years suggestive of an artefactual change in rates

(iii) Age-specific bridge-coding conversion factors were small (see Appendix, Table A-2, page 349) and did not suggest large changes in rates across change years.

(iv) These bridge-coding factors are derived from national data and may differ for the geographical areas used here.

(v) Age-specific bridge-coding conversion factors were sought but not available from ONS for ICD8 to ICD9 changes.

Rule 3 coding change 1984-1992

As discussed in chapter 3, the introduction of a broader interpretation of rule 3 in England & Wales between 1984-1992 was estimated to have caused an increase of around 7% in COPD mortality, but to have had little effect on neoplasms. Data were not adjusted for the estimated changes for similar reasons as those listed for ICD coding changes. For example, factors were produced for effects on 1993 and 1994 and it may not have been valid to simply apply a factor of –7% over 1984-1992. Also, age-specific factors were not available but there was some evidence of heterogeneity with greater impacts in the elderly (Table 3.1, page 69) and factors may have varied by area. Again, there was no obvious step change in rates for these years.

Inconsistencies in coding over time

The ICD codes used to define COPD and lung cancer are given in Table 4-1 (page 94) and those for influenza in Table 4-2 (page 95). Reasonably but not exactly consistent coding was available over the 50 year period of analysis. COPD was defined as 'bronchitis' in earlier years and as chronic bronchitis or emphysema but excluding

asthma in later years. Bronchitis deaths for 1950-57 (ICD6) and 1958-67 (ICD7) included deaths coded to 500 'Acute bronchitis'. This gave a potential for an artefactual rate change in 1968. It was not possible to add 'acute bronchitis' deaths in all subsequent years (1968-1999) because this information was not available in published volumes for 1968-80 by the geographical areas used in this analysis. Adding acute bronchitis deaths for 1981 onwards was not logical as this gave the potential for a second artefactual rate change. It was not possible to remove 'acute bronchitis' deaths for all of 1950-67 because sub-national information on numbers of these deaths were not available in published volumes for 1950-57. Removing 'acute bronchitis' from 1958 onwards was not desirable as it could have led to artefactual changes in rates at the time of greatest interest with respect to air pollution levels.

Codes for 1979-1980 included ICD9 494 (Bronchiectasis) and 495 (Extrinsic allergic alveolitis) but in these years represented a small proportion of deaths (2.1% of male and 4.5% of female). Bronchiectasis codes (ICD6 & ICD7 526, ICD8 518) had not previously been included in published figures for 'bronchitis' and extrinsic allergic alveolitis did not exist in ICD coding prior to ICD9 (introduced in 1979). It was therefore considered inappropriate to include bronchiectasis and extrinsic allergic alveolitis from 1981 onwards. Removing these two conditions for 1979-80 was problematic as it could only have been done by interpolating England & Wales data and making assumptions about the distribution of these deaths by area – when these could have varied by area according to the hypotheses in question.

The effect of including bronchiectasis and extrinsic allergic alveolitis in 1979-80 and acute bronchitis before 1968 was examined by inspecting rates and (for acute bronchitis) referring to published bridge-coding factors derived from dual coding in the ICD change year of 1968. This suggested that the impact of these coding issues on the analysis was likely to be minimal, but that inclusion of acute bronchitis could potentially result a small period effect around 1968.

Agebands

Data were obtained by ten-year agebands for 15-24 to 75+ years. The upper ageband used was 75+ years as the 75-84 and 85+ agebands were not reported separately until 1977. Sub-national data were reported in 20 year agebands for ages 25-44 and 45-64 years for 1950-57 and for 25-44 years in 1958-62. To obtain data for conurbation and

non-conurbation areas in 10 year age-bands for these time periods, all England & Wales data that were available by 10 year age-bands for corresponding ICD codes were used to interpolate regional data assuming that age breakdowns would be similar in all areas of the country.

Other issues

Deaths from 1950-80 related to the year of registration, while deaths 1981 onwards related to the year of occurrence, but this was not expected to affect the present analyses. Industrial action was taken by Registration Officers in 1981-82. This had effects on the reliability of coding of occupational data for deaths in 1981[284] and on deaths attributed to injury and poisoning, but should not have affected present analyses.

Issues with population denominator data

Population figures for 1980, obtained from two different ONS source volumes could not be reconciled with those for 1979 and 1981 so were interpolated from adjacent years.

Table 4-1 Codes used to define COPD and lung cancer in time trend analyses

Year	Code in OPCS Registrar General/ONS DH5 series published volumes	ICD codes	Description
COPD			
1950-57	B32 Bronchitis	ICD6 500-502	500 Acute bronchitis 501 Bronchitis unqualified 502 Chronic bronchitis
1958-67	B32 Bronchitis	ICD7 500-502	As ICD6
1968-78	B33 (Pt) Bronchitis	ICD8 490-492	490 Bronchitis, unqualified 491 Chronic bronchitis 492 Emphysema
1979-80	F323 – F325. Bronchitis, emphysema and asthma minus F323.2 Asthma	ICD9 490-492, 494-496	490 Bronchitis not specified as acute or chronic 491 Chronic bronchitis 492 Emphysema 494 Bronchiectasis 495 Extrinsic allergic alveolitis 496 Chronic airways obstruction, not elsewhere classified
1981-1999	-	ICD9 490-492 & 496	As 1979-80
Lung cancer			
1950-57	B18 Lung cancer	ICD6 162 & 163	162 Malignant neoplasm of trachea, and of bronchus and lung specified as primary 163 Malignant neoplasm of lung and of bronchus, unspecified as to whether primary or secondary
1958-67	B18 Lung cancer	ICD7 162 & 163	162 Malignant neoplasm of bronchus and trachea, and of lung (primary) 163 Malignant neoplasm of lung, unspecified as to whether primary or secondary
1968-73	B18 Lung cancer	ICD8 162	162 Malignant neoplasm of trachea, bronchus and lung
1974-78	B18 Lung and bronchus cancer	ICD8 162	As 1968-73
1979-80	F101. Trachea, bronchus and lung	ICD9 162	As ICD8
1981-1999	-	ICD9 162	As ICD8

Table 4-2 Codes used to define influenza in time trend analyses

Year	ICD codes	Description
1950-57	ICD6 480-483	480 Influenza with pneumonia 481 Influenza with other respiratory manifestations, and influenza unqualified 482 Influenza with digestive manifestations, but without respiratory symptoms 483 Influenza with nervous manifestations, but without digestive or respiratory symptoms
1958-67	ICD7 480-483	As ICD6 above
1968-78	ICD8 470-474 & 482.1	470 Influenza unqualified 471 Influenza with pneumonia 472 Influenza with other respiratory manifestations 473 Influenza with digestive manifestations 474 Influenza with nervous manifestations 482.1 Other bacterial pneumonia: Haemophilus influenzae
1979-96	ICD9 482.2 & 487	482.2 Pneumonia due to Influenza 487 Influenza
1997-99	ICD9 487*	487 Influenza

*Totals for ICD9 482.2 "Other bacterial pneumonia, due to Haemophilus influenzae" were not available in published volumes as these only record causes for which there were 10 or more deaths for either sex

Geographical boundaries for mortality and population data

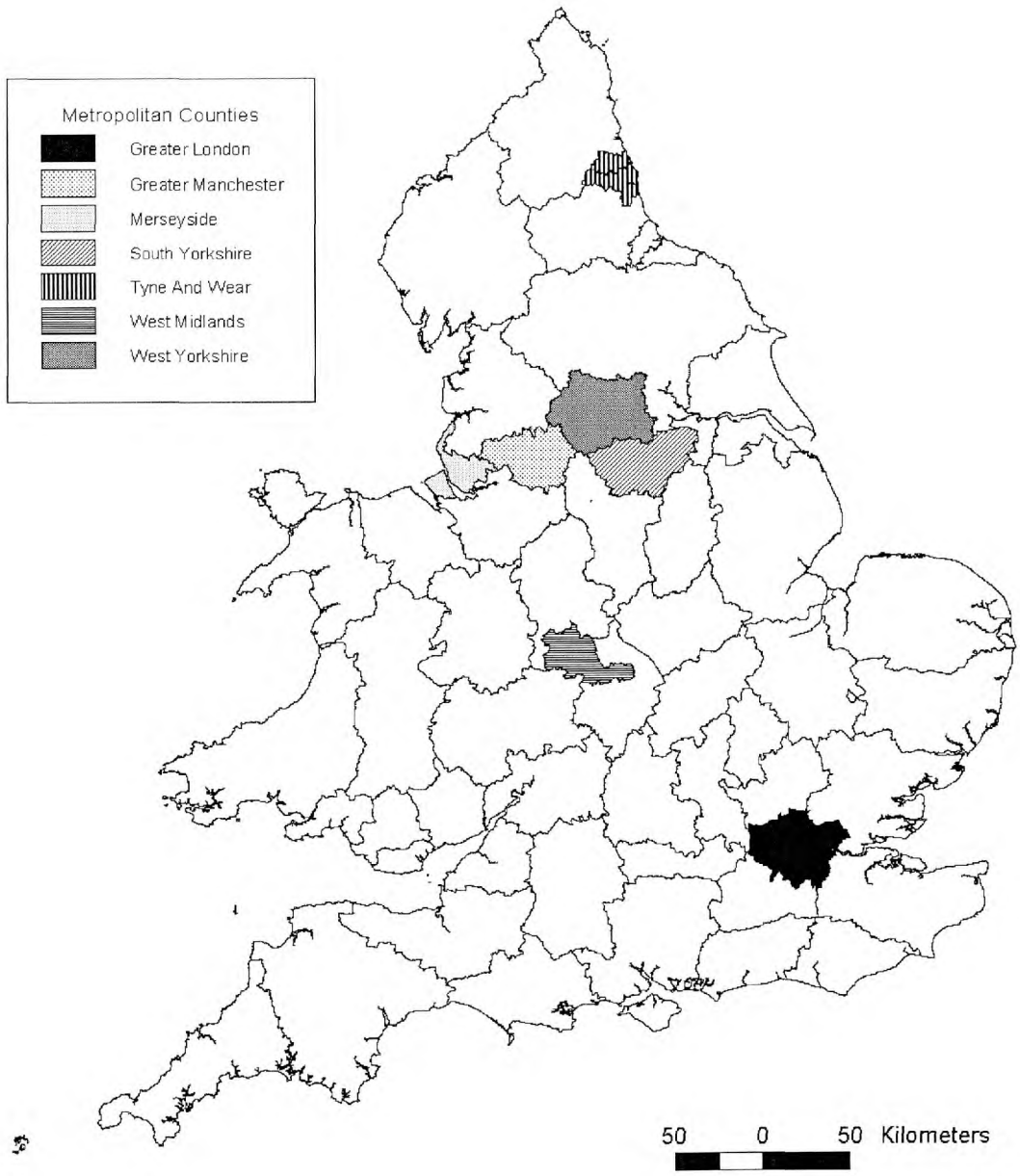
For 1950-1980, a dataset was constructed to give boundaries as homogeneous as possible given the limitations of the published data – data prior to 1980 were not available to small area level, which would have allowed definition of similar boundaries. The conurbation area comprised the metropolitan counties of Merseyside, South-East Lancashire, Tyneside, West Midlands, West Yorkshire (Figure 4-1, page 97). The main change to these boundaries occurred in 1974, when all these areas were enlarged. Since the analyses used rates rather than numbers, this was considered likely to have a limited impact on interpretations. South Yorkshire was also designated a metropolitan county in 1974, but was not included as a conurbation for the purposes of consistency in geographical boundaries over time. Boundary and resulting population changes prior to 1974 were detailed in Registrar General volumes for each year and were all minor (not shown). Comparisons of geographical boundaries for 1981 and 1991 boundaries using Geographical Information Systems (GIS) software demonstrated that, although there were changes *within* metropolitan counties and Greater London, the outer boundaries did not change between 1981 and 1991.

Air pollution data

Information on air pollution episodes and time trends from published literature

Information on general air pollution trends were obtained from published literature identified during the literature search, from the references of such articles and books identified, from the Imperial College Library and through colleagues. Searches concentrated on black smoke and sulphur dioxide (SO_2) as comparable data were available for most of the period 1950-99.[246] Similar literature search methods were used to identify years with important black smoke and SO_2 air pollution episodes possibly resulting in period effects in age-period-cohort analyses.

Figure 4-1 Map of England & Wales showing location of Metropolitan Counties and Greater London (1991 boundaries)



Regional data on air pollution

Data for annual average black smoke and SO_2 levels for 1958-1971 by region, collated by Warren Spring laboratory[37] were obtained. These data consisted of valid averages from approximately 100 country sites and 900 urban sites, covering 59 of 62 (95%) large towns (population >100,000) but only a quarter of smaller towns (population 5,000 to 20,000). Smoke and sulphur dioxide were measured by the British standard method.[37] This involved drawing a sample of air through a filter paper for a 24 hour period, measuring the staining by a reflectometer and using a calibration curve to give the smoke concentration in $\mu g/m^3$. Diesel smoke is darker, weight for weight, than 'standard' smoke, while smoke with contamination by cement or chalk dust could be lighter in colour. Air passing the smoke filter was then bubbled through dilute hydrogen peroxide, oxidising sulphur dioxide to sulphuric acid, the amount of which in $\mu g/m^3$ was determined by titration to pH 4.5. This method actually measured net gaseous acidity rather than SO_2 per se. In most urban areas, most of the acidity was due to sulphur dioxide. However, in the countryside ammonia neutralised some of the acidity, giving lower apparent levels of acidity.

Air pollution data from long-running monitoring sites

Data for annual mean black smoke and sulphur dioxide levels were obtained from individual long-running monitoring stations. Information on location, opening and closing dates of monitoring stations was obtained from AEA Technology Environment (AEAT). Black smoke and SO_2 data for stations operating prior to 1962 were obtained from published tables supplied by AEAT. Black smoke and SO_2 data for stations operating 1961-1996 were obtained from the Dept of Epidemiology & Public Health of Imperial College London. These departmental data had been previously downloaded from the Air Quality Archive (internet site maintained by AEAT), but excluded a small number of non-residential sites classified as industrial areas or open countryside. Black smoke data measurements were expressed in $\mu g/m^3$. Sulphur dioxide data from AEAT for the two years prior to April 1958 were expressed in $mg/100m^3$ and were multiplied by the specified[285] factor of 28.6 to convert them to $\mu g/m^3$. Departmental data were expressed in ppb and were multiplied by a factor of 2.66 as specified on the AEAT website (www.airquality.co.uk/archive/kb.php?action=showpost&question_id=38) to convert them into $\mu g/m^3$.

The environment of the monitoring sites were classified as 'A' residential high density, 'B' residential medium density, 'R' rural community, 'C' industrial, 'D' commercial/town centre or 'O' open country. These classifications had additional sub-divisions and an 'E' suffix if a smoke control area (Table 4-3). The criteria by which classifications were made was not specified.

Table 4-3 Classification of UK air pollution monitoring stations as supplied by AEAT

Site classification	Description
A1	Residential area with high-density housing (probably terraced),or with medium-density housing in multiple occupation, in either case surrounded by other built-up areas.
A2	Predominantly A1, but interspersed with some industrial undertakings.
A3	Residential area with high-density housing or medium-density housing in multiple occupation surrounded by, or interspersed with, other areas with low potential air pollution output (parks, fields, coast).
B1	Residential area with medium-density housing, typically an inner suburb or housing estate, surrounded by other built-up areas.
B2	Predominantly B1, but interspersed with some industrial undertakings.
B3	Residential area with medium-density housing surrounded by or interspersed with areas with low potential air pollution output (parks, fields, coast), or any residential area with low-density housing.
C1	Industrial area without domestic premises.
C2	Industrial area interspersed with domestic premises of high density or in multiple occupation.
D1	Commercial area or one with predominantly central heating.
D2	Town centre with limited commercial area, possibly mixed with old residential housing and/or minor industry.
E	Smoke control area or smokeless zone (the letter to be added to the primary classification, e.g. A1/E).
R	Rural community.
O1	Open country but not entirely without source(s) of pollution, e.g. airfields.
O2	Completely open country; no sources within at least 400 metres.
X	Unclassified site, or mixed area.

Only three stations (Darton 1, Elland 2, Leeds 4) all within Yorkshire-Humberside region had recorded continuously from 1950 to 1996. Fifteen stations had started recording in 1956 or earlier and were still recording in 1985 or later years. However, two non-residential sites (Lytham St Annes 1 – O1 or open countryside – and St Helens 12 – C1 or industrial area) had been dropped from departmental data, leaving 13 stations with data for at least 1956 to 1985 .

Data related to financial year but for simplicity have been labelled as calendar year (e.g. data for April 1960 to March 1961 have been labelled 1960). Published tables prior to 1962 listed annual averages only if 12 months of data were available. For data from 1962 onwards, daily readings and the number of days on which readings were recorded was available so annual means could be calculated.

Descriptive analyses

Descriptive analyses of air pollution data

Descriptive analyses of air pollution data were conducted. To obtain a semi-quantitative impression of the difference between conurbations and other areas, linear regression lines were fitted through the regional values, available for 1958-9 to 1970-71. Fitted values for 1958-9 and 1970-1 for each region were then calculated using these regression lines and results were averaged over the four conurbation regions and five non-conurbation regions. The levels of air pollution in conurbations and Greater London relative to non-conurbations for 1958-9 and 1970-1 were then calculated.

Descriptive statistics on annual means of air pollution readings were conducted for the long-running air pollution monitoring sites. As the visual impression of trends was very similar when using all years or when restricting data points to years with >330 days of readings, plots show data for all available years, regardless of number of days in which readings were recorded during the year.

Descriptive analyses of mortality data

Mortality rates for conurbations, Greater London and non-conurbations were examined for males and females separately throughout, initially stratified by age, then age-standardised to the 1999 population of England & Wales. Age-standardised rates were further examined by looking at the absolute and relative differences between conurbations and Greater London with respect to non-conurbations. Finally plots of the ratio of COPD: lung cancer by area were examined.

Analytical analyses

Age-period-cohort analyses of COPD and lung cancer mortality

Age-period-cohort analyses of COPD and lung cancer mortality in conurbations, Greater London and non-conurbations were conducted using both classical (frequentist) and Bayesian statistical estimation methods. The starting point for age-period-cohort analyses of mortality rates over time is the assumption that the observed rates result

from a combination of age, period and cohort trends. Period effects represent factors likely to affect people of all ages at a particular period in time e.g. treatment advances, influenza epidemics or air pollution episodes, while cohort effects represent factors more common in people born at a particular point in time e.g. smoking habits.

Age, period and cohort parameters cannot be independently estimated, as cohort is a deterministic function of age and period i.e. all individuals in the same age and period category will automatically be in the same cohort category. Since cohort is defined as a linear function of age and period, there are infinitely many linear transformations of the age, period and cohort parameters that would give the same overall estimate of the rate underlying the observed counts in each age-period category. For this reason, differences in magnitude of linear trends in the estimated age, period or cohort effects cannot be interpreted as they could be the result of arbitrary linear transformations of the parameters and the solution set will be dependent on the model constraints chosen – usually by arbitrarily fixing one of the parameters (e.g. assuming a pattern for age on biological grounds). This problem affects both classical[286] and Bayesian[211] age-period cohort models and is referred to as the non-identifiability problem. However, non-linear trends i.e. changes in the rate of changes, are intepretable and it is possible to identify whether change points occur when rates are changing from an increase to a decrease (concave down) or vice versa (concave up). It is also possible to interpret the differences between parameters (e.g. period effects) for different geographical areas.

Classical age-period-cohort analyses

Age-period-cohort analyses assume an underlying relationship between the observed number of cases and age, period and cohort parameters of the form

$$\mathbf{Y_{ijk} \sim Poisson\ (\mu.\ \lambda_{ijk})} \qquad \textbf{Equation 4-1}$$

where Y_{ijk} is the number of cases for age-group i and cohort k in period j

$i = 1, \ldots, I$

$j = 1, \ldots, J$

$k = 1, \ldots, K$

μ is the baseline rate

and λ_{ijk} is the rate for age-group i and cohort k in period j

(Strictly speaking, k is defined by i and j so is redundant in the notation, but it has been left in for comprehension).

Classical analyses using maximum likelihood estimation can be conducted using a Poisson regression model. Assuming the relationship outlined in Equation 4-1 above, the Poisson model has the form

$$\log(\lambda_{ijk}) = \mu + \theta_i + \phi_j + \psi_k \qquad \textbf{Equation 4-2}$$

where

μ is the baseline rate

θ_i is the age parameter for age-group i

ϕ_j is the period parameter for period j

ψ_k is the cohort parameter for cohort k.

Limitations in mortality data by geographical area in published volumes meant that numbers of deaths were only available by 10 year age bands. This would only have given five (10 year) periods for mortality severely limiting the ability to investigate the timing of period or cohort change points. Because of this a classical age-period-cohort analysis for the conurbations areas and non-conurbations was not conducted.

Bayesian age-period-cohort analyses by area to identify timing of period and cohort effects

A recently developed Bayesian age-period-cohort model,[211] was used to investigate the timing of period and cohort effects for COPD and lung cancer for each area (conurbations, Greater London and non-conurbations) and for men and women separately. The model was implemented in BAMP software.[281] Unlike classical[286] and some Bayesian models,[287] BAMP incorporates an *a priori* assumption that consecutive cohort parameters are similar through the use of smoothing priors to model the dependence between consecutive cohort parameters (it also assumes that consecutive age parameters and consecutive period parameters are similar). This is a particular advantage if wanting to incorporate age-groups and periods on different time grids – for example, with mortality data supplied by five year agebands and one year periods – as this results in an even greater overlap of birth cohorts than with frequentist models, so that the same data contribute to many cohort categories. In this case, it is inappropriate to use an estimation method that assumes that cohort parameters are independent.

This analysis used seven ten-year age-groups and 50 one-year periods from 1950-1999, associated with 110 overlapping birth cohorts of 11 years length with the first cohort 1865-1875 and the last cohort 1974-1984.

As well as age, period and cohort parameters, the model also incorporates a term for unobserved observation-specific covariates, often referred to as overdispersion in the frequentist literature.[211] This ensures a good fit to the data.

The Bayesian age-period-cohort model was used with a Random Walk 1 (RW1) and Random Walk 2 (RW2) smoothing constraints. RW1 uses first order differences of age, period or cohort parameters, favouring solution parameters with constancy i.e. it penalises departures from a constant value for consecutive effects solutions and assumes a smoothness of trends. This allows age, period and cohort trends to be examined visually. The RW2 constraint uses second order differences of age, period or cohort parameters and this model penalises deviations from a linear trend. Visual plots using RW2 constraints are less meaningful because adjacent parameters are not constrained in the same way as RW1, but RW2 may be more appropriate to examine change-points.

Within the Bayesian framework information from both prior beliefs and from the data are combined to give a posterior distribution, from which actual values for parameters can be derived though repeated iterative sampling. Specified values for the hyperprior settings (these are the parameters defining the shape and scale of the prior distributions) of unknown precision parameters were chosen to be uninformative,[211] allowing the data to strongly influence the posterior distribution. The first 2,000 samples representing 'burn-in' for the model were discarded and quoted results are based on a further 100,000 samples. Credible intervals (analogous to confidence intervals in frequentist analyses) can be automatically provided during this process based on the appropriate percentiles (e.g. 5%, 95%) of the set of sampled values for the parameters of interest.

Identification of change points for period and cohort effects

Change points for period effects were located by examining second order differences for analyses conducted using a Random Walk 2 (RW2) constraint and identifying the years where the 5%-95% credible intervals did not encompass zero (i.e. differences were all negative indicating a peak effect or all positive indicating a change from a fall to a rise) in relation to the original hypotheses. Using BAMP, change points may represent either

a local peak or trough or a marked change in slope (see Figure 4-2, (a) to (d), below). The current formulation of BAMP does not enable a distinction to be made between period effects relating to short-term fluctuations and those related to longer-term underlying trends (see Figure 4-2 (e)). Moreover, long-term underlying trends causing very gradual changes, may fail to register as statistically significant changes. To increase the sensitivity of detection of change points, when 5-95% centiles did not indicate an effect, other levels of statistical certainty were examined e.g. 25%-75% credible intervals not encompassing zero. Visual examination for potential change points of period parameters using RW1 analyses were also conducted, which also allowed visual assessment of gradual changes as in Figure 4-2, (e) below.

Figure 4-2 Possible shapes of age, period or cohort parameter curves from age-period-cohort analyses (label given to change point identified)

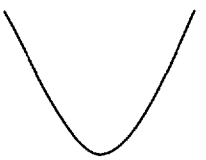
(a) Rise to fall (maximum)



(b) Rise with levelling off (maximum)



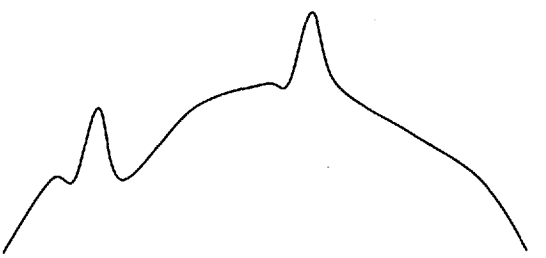
(c) Fall to rise (minimum)



(d) Fall with levelling off (minimum)



(e) Complex – long-term rise then fall with superimposed acute rises and falls (Acute rises and falls identified as maxima. Long-range rise may be visible in plots but may not be identified as statistically significant if change is gradual as here, minima associated with first acute peak may not be statistically significant if small as here).



Cohort effects were identified in a similar fashion to those for period effects. Where cohort effects were significant for a number of consecutive (and overlapping cohorts), this was considered to form a single cohort effect and the centre year(s) of all the cohorts was taken as the midpoint. Interpretations of change points and corresponding levels of statistical certainty for period effects were made in relation to prior hypotheses relating to the Clean Air Act, artefacts related to coding and boundary changes, and other effects such as years with high influenza mortality in England & Wales and air pollution episodes identified from the literature.

Classical age-period-cohort analyses incorporating an interaction term with degree of urbanisation and age and period effects on different grids

This section relating to classical age-period-cohort analyses incorporating an interaction term with degree of urbanisation represents collaborative work with Professor Leonhard Held and Herr Ralf Breuninger in the Institut für Statistik, Ludwig-Maximilians-Universität, München. The statistical methodology was developed by Professor Held and statistical analyses were conducted by Ralf Breuninger[288] supervised by Leonhard Held using data supplied by Anna Hansell. The interpretation of analyses was by Anna Hansell, supervised by Leonhard Held. This included additional analyses such as graphs and linear regression of relative risks against year to find average changes in period effects.

The statistical model developed by Professor Held and Ralf Breuninger was adapted from the usual classical age-period-cohort model, which has age and period on similar time grids, to be able to use age in 10 year bands but yearly mortality data (period). Development of a Bayesian model incorporating an interaction term with degree of urbanisation was planned, but was not completed in time for results to be incorporated into this PhD.

These analyses were used to attempt to quantify the relative changes in period or cohort parameters in conurbations and Greater London relative to non-conurbations. One of the original hypotheses suggested that period effects would be different between the more urban and less urban areas, while age effects would be similar. To investigate whether age, period or cohort parameters were shared across each geographical area, classical age-period-cohort analyses were conducted introducing an interaction term

relating to degree of urbanisation where specified age, period and/or cohort effects were allowed to vary by geographical area. Here the underlying relationship was assumed to fit the form

$$\mathbf{Y_{ijku} \sim Poisson\ (\mu\ .\ \lambda_{ijku})} \qquad \textbf{Equation 4-3}$$

where u is the level of urbanisation as a factor

and $u = 1$ (conurbation), 2 (Greater London) or 3 (non-conurbations)

with other notation as above in Equation 4-1, page 101.

Various log-linear models for the rate parameter were then specified, assuming that either none or some combination of the age, period and cohort effects varied by degree of urbanisation. For example, the Poisson model assuming varying period effects but similar age and cohort effects in each urban area was of the form

$$\mathbf{\log\ (\lambda_{i\ j\ k\ u})} = \mu + \theta_i + \phi_{ju} + \psi_k \qquad \textbf{Equation 4-4}$$

where

$\log\ (\lambda_{i\ ju\ k})$ is the log rate for age-group i and cohort k in period j and urbanisation level u

μ is the baseline rate

θ_i is the age parameter for age-group i

ϕ_{ju} is the interaction term between the period parameter for period j and the urbanisation level u

ψ_k is the cohort parameter for cohort k

Similarly, the Poisson model assuming varying age effects but similar period and cohort effects in each urban area was of the form

$$\lambda_{ijk\ u} = \mu + \theta_{iu} + \phi_j + \psi_k \qquad \textbf{Equation 4-5}$$

The Poisson model assuming varying cohort effects but similar age and period effects in each urban area was of the form

$$\lambda_{ijk\ u} = \mu + \theta_i + \phi_j + \psi_{ku} \qquad \textbf{Equation 4-6}$$

and so on.

The statistical model was adapted from the usual classical age-period-cohort model, which has age and period on similar time grids, to be able to use age in 10 year bands

but yearly mortality data (period). To do this, for each geographical area, cohorts where no deaths had occurred in any of the possible combinations of age and period had to be removed (as this would imply that the log rate was minus infinity). This resulted in the removal of less than 5 (out of 110) cohorts where cohort effects were assumed to be shared by all three types of urbanisation level (i.e. pooled cohort effects), but 16-27 cohorts where cohort effects were estimated independently for each area (Table 4-4). Another problem with this model was that, unlike the Bayesian analyses, overlapping cohort parameters are treated as independent by the classical estimation method.

Table 4-4 Numbers of cohorts with no deaths and therefore removed from classical Poisson age-period-cohort analyses (total number of cohorts = 110)

	Numbers of cohorts with no deaths in models assuming shared cohort effects across areas (therefore pooled numbers of deaths).	Numbers of cohorts with no deaths in models estimating cohort effects separately for each area
COPD – male	0	16
COPD – female	2	26
Lung cancer – male	2	18
Lung cancer – female	4	27

It was not possible to identify the age, period and cohort parameters themselves due to both the identifiability problem and because of artefactual regular periodicity resulting from the different grids used for agebands and periods. However, it was possible to draw inferences about the effect of differing assumptions of shared age, period and/or cohort effects on fit of the model by examining deviance and Akaikes information criteria (AIC). Bayesian inference criteria (BIC), an alternative measure of fit, was also examined for comparison.

Log-linear Poisson regression maximum likelihood analyses were conducted assuming no effects were similar in all geographical areas, one of age, period or cohort effects were similar (shared), two of the three effects were shared and all effects were shared. Using a model assuming that age and cohort effects were shared (according to the original hypothesis), it was possible to identify and interpret the relative difference between the urban areas and the non-conurbations for period effects by subtracting the estimated period parameters for non-conurbations (urbanisation level 3) from that for conurbations (urbanisation level 1) and from that for Greater London (urbanisation level 2), mathematically expressed as:

$$\left(\hat{\phi}_{j1} - \hat{\phi}_{j3}\right) \quad and \quad \left(\hat{\phi}_{j2} - \hat{\phi}_{j3}\right)$$

The exponential of these values were taken to examine the variation in relative risk between the urban areas and the non-conurbations over the period 1950-99. A similar process was conducted looking at cohort effects for conurbations and Greater London as compared with non-conurbations, assuming age and period effects were shared across areas.

Although the above models could be estimated using maximum-likelihood analyses, these did not take account of overdispersion or extra-Poisson variation evident in the data (the variance for death counts was generally much larger than the mean). Not taking account of the overdispersion would have resulted in falsely narrow confidence intervals and exaggerated statistical significance. To produce confidence intervals for the relative risks adjusted for overdispersion, additional analyses were conducted using Wedderburn's quasi-likelihood method as set out by Tango and Kurashin,[289] which assumes that

$$\mathbf{var(Y) = \sigma^2 E(Y)} \qquad \textbf{Equation 4-7}$$

where

Y is the observed number of counts

E(Y) is the expected number of counts (equivalent to λ_{ijku} in Equation 4.3, page 106)

and

σ^2 is the overdispersion factor.

Under this assumption, the quasi-likelihood is identical to the Poisson log-likelihood resulting in the same point estimates for parameters regardless of the value of σ^2. The variances (and therefore standard errors used to produce confidence intervals) are then adjusted by multiplying them by the overdispersion factor.

As part of the interpretation (conducted by Anna Hansell), the average change in period effect over the time period of the analysis was calculated through linear regression of the relative risk for each area against time and these results were compared with information on air pollution trends.

Chapter 5. Temporal variations in COPD mortality in England & Wales in relation to the 1956 Clean Air Act: Results

Overview of Chapter 5

This chapter contains the results from the temporal analyses section of the PhD.

Key findings

Timing of changes in COPD mortality rates in relation to air pollution trends

- Age-standardised COPD mortality rates suggested a decline in all areas (conurbations, Greater London and non-conurbations), which was most marked between the late 1950s and late 1970s. This was located to a period effect in Bayesian age-period-cohort (BAMP) analyses with evidence of timing of onset most consistent for 1962-63. However, some areas also had effects located to the late 1950s and it was not possible using the BAMP analyses to distinguish between period effects relating to year-to-year variations (e.g. influenza activity, air pollution episodes) and those related to longer term trends. Results were most consistent for a levelling off of the decline in 1977-78.
- The trends in COPD mortality and the timing of these period effects showed close similarities with trends for black smoke concentrations in the same years, but less similarity with trends in SO_2.
- In contrast with COPD, no clear period effects were located for lung cancer in BAMP analyses.

Relative differences in COPD mortality rates and period effects in different areas and in relation to air pollution trends

- As well as a general decline in COPD mortality rates, age-standardised COPD mortality rates showed a convergence of rates in conurbations and Greater

London towards rates in non-conurbations, with most of this occurring between 1950 to the mid 1970s.

- Classical age-period-cohort models assuming different period but shared age and cohort effects across areas were used to look at relative risks of the period effects for COPD mortality comparing conurbations and Greater London with non-conurbations. Convergence was confirmed for males in conurbations and Greater London and females in conurbations during 1957-78. However, females in Greater London showed a greater convergence towards non-conurbations after 1979 than before.
- There were similarities in relative risks of the period effect for COPD mortality (for conurbations and Greater London relative to non-conurbations from classical age-period-cohort analyses) and the relative differences between averaged black smoke concentrations in conurbations and Greater London with respect to non-conurbations. Similarities were also seen with SO_2, but for conurbations (relative to non-conurbations) only.

Cohort patterns

- The timing of cohort effect patterns identified using BAMP analyses were similar for COPD and lung cancer in all areas, but differed by sex consistent with expected effects from lifetime smoking patterns. BAMP located diffuse peaks in cohort effects (highest mortality throughout life) in male cohorts centred around 1899-1903 and in female cohorts around 1925-1929.
- An additional, but only weakly significant, cohort effect in men born in cohorts centred on 1925-1926 (at the same time as the cohort effect in females) was identified in non-conurbations. A possible second cohort effect for COPD had been predicted for the 1940s and 1950s related to falling air pollution levels in childhood. However, there was only weak and inconsistent evidence for this in the 1950s, which was only seen in women.
- There was some suggestion that the magnitude of cohort effects varied by area (reflecting differences in smoking patterns between conurbations, Greater London and non-conurbations). Age-stratified analyses for lung cancer mortality suggested a cohort pattern with falls in rates occurring earlier in Greater London than in non-conurbations. The classical age-period-cohort analyses found:

- the best-fitting model for lung cancer when only one parameter was allowed to vary was allowing different cohort effects in different areas
- the relative risk of the cohort effect for lung cancer for males and females suggested a marked fall in relative risk for Greater London with respect to non-conurbations, but a fall followed by a rise in conurbations relative to non-conurbations

Trends in air pollution

Results from the air pollution data are discussed first, as these were subsequently used in some of the analytical analyses.

Air pollution levels prior to 1950 affecting older cohorts

Air pollution levels prior to 1950 are not just of historical interest as high exposures could have resulted on effects in members of older cohorts (members of the oldest cohort in this analysis were born 1865-1875). Early estimates of SO_2 air pollution concentrations in English towns and cities suggested high levels in cities (Table 5-1).

Table 5-1 Sulphur dioxide concentrations in English towns and cities published in 1895

Location	Sulphur dioxide $\mu g/m^3$ (as SO_2)
Manchester	2930
London	2180
Buxton	1950
Didsbury	1746
Blackpool	620

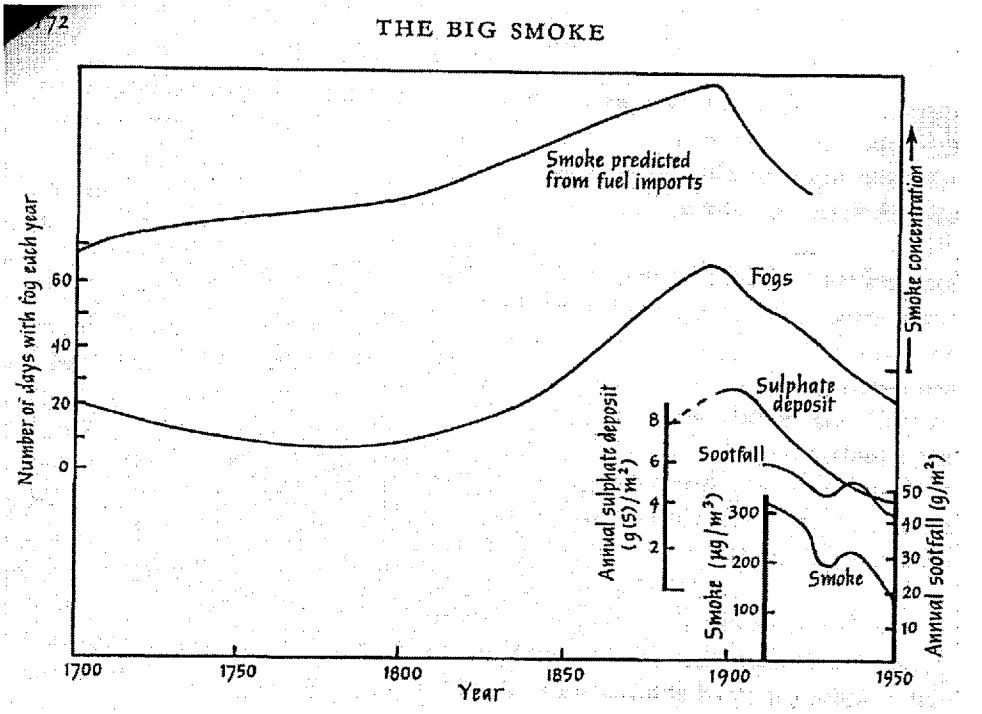
Source: Table 7.6 in Brimblecombe 1987,[35] page 155

Most of the available information on historical air pollution trends relates to London, but reports of soot-fall from English health authorities in 1914[290] suggested that Birmingham and northern cities such as Liverpool, Paisley and Newcastle were dirtier than London.

By Victorian times, London was well known for its fogs, which were a combination of climatic conditions and air pollution. Fogs in London had increased in frequency between 1750 and 1890 then began to decline[35]. The fogs of 1873 resulted in 700 more deaths than expected in London at that time of year. Over the next 20 years 'great fogs' recurred – these settled over London for many days due to stagnant conditions, allowing high concentrations of pollutants and increased rates of mortality.[35] By 1886-90 there were 70-80 foggy days per year. Fogs were so thick that houses and shops had to be lit during the day and there were occasional drownings as people inadvertently walked into the Thames or canals.[35] It is also suggested that the lack of light may have been partially responsible for the increase in prevalence of rickets. Some fogs were coloured yellow or orange and left oily deposits on windows.[35] This was probably due to tar substances derived from the coal. For example, after the fogs of 1891, tar was found to

be deposited on the glasshouse at Kew[291] and an oily yellow film covered the roads after the fogs of 9-12 December 1924. While fogs decreased in frequency by the 1890s, annual sulphate deposits in London peaked around 1900 (Figure 5-1 below).

Figure 5-1 Smoke concentrations and number of days with fog per year in London



Source: Brimblecombe P. The Big Smoke[35]

Much of the air pollution in London and presumably other conurbations related to the widespread burning of cheap coal containing relatively high amounts of sulphur, both industrially and in the home on an open hearth for heating and cooking. The English were not keen to change the low temperature burning hearth for the more efficient (therefore less polluting) high temperature stoves used on the continent, partly because the climate was not as cold so less heat was needed and partly because of the attachment to a welcoming open fire in the living area. There was little incentive for industries to spend money to reduce emissions and much air pollution legislation was weak. However, the sulphur content of coal varied depending on the source. Sulphur content in coal from northern England was relatively high and certainly higher than Scottish coal (Table 5-2, page 114).

Table 5-2 Sulphur content of coals, figures published in 1918

Origin	Type	% sulphur
Notts. slack	bituminous	0.45
Lancashire	bituminous	1.38
Yorkshire	bituminous	1.20
Durham	bituminous	1.00
Scotland	anthracite	0.10
South Wales	anthracite	~0.7
East Indian	semi-bituminous	2.50

Source: Table 4.1 in Brimblecombe 1987,[35] page 66

The 1956 Clean Air Act

The 1956 Clean Air Act[292] was the political response to the major air pollution episode in London in December 1952, that lead to at least 4,000[293] and up to 12,000[294] deaths. The Act introduced smoke control areas, controlled chimney heights and prohibited emissions of dark smoke from chimneys. It was held responsible for sharp falls in smoke emissions from industry between 1955 (when the Act was in discussion) and the early 1960s and longer term improvements in domestic emissions.[37] Control of domestic emissions was implemented through Local Authorities in the worst affected areas (the 'Black Areas' as named by the Beaver committee – The Report of the Committee on Air Pollution, Chairman Sir Hugh Beaver in 1954), through the establishment of Smoke Control Areas where the only solid fuels permitted were smokeless fuels such as coke or anthracite.[37] However, air pollution remained appreciably high in the 1960s. For example, in 1962 it was noted[295] that on the hills to the east of Manchester

> 'at times the smoke can be seen crossing the summits in a shallow layer at ground level, and a fresh fall of snow may be heavily soiled in 2 hours'.

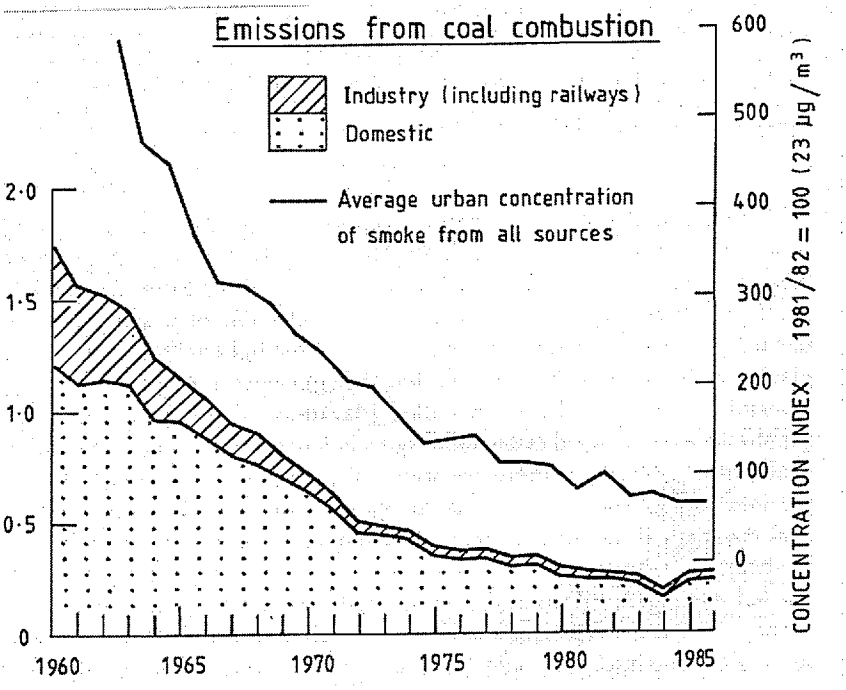
The 1956 Act applied to smoke but not sulphur dioxide. However, smoke control measures indirectly encouraged the reduction of sulphur dioxide levels by encouraging the use of oil, electricity and gas (which was by the 1950s rapidly changing to sulphur-free natural gas, rather than gas produced from coal and coke).[296] Smokeless fuels only resulted in a minimal reduction in sulphur dioxide emissions. The first general legislation to limit emissions of sulphur dioxide was not into force until 1972 when the City of London (Various Powers) Act prohibited the use of fuel oil containing more than 1 per cent sulphur in the city.[37] Virtually all premises had converted to low sulphur fuels by 1983 and this was reflected in a decline in sulphur dioxide levels in the City.[296]

Another factor reducing ambient sulphur dioxide concentrations was the closure of urban industrial sources such as coal-burning electricity generating stations[296] or their relocation to rural areas.

National smoke and SO_2 air pollution from 1950 onwards

Most of the published data on time trends in ambient concentrations of air pollutants relate to 1960 onwards. Urban black smoke ambient concentrations in the UK fell from an annual average of 180 μg/m^3 in 1960 to 12 μg/m^3 in the 1990s[244] with most of the fall occurring prior to 1975 (Figure 5-2). Similar figures are quoted by the Quality of Urban Air Review Group (QUARG),[245] with a fall in average urban smoke concentrations across the UK from 150 μg/m^3 in 1962/3 to an average of 11 μg/m^3 in 1993/4 in the Basic Urban Network, with most of the fall occurring by the mid to late 1970s. Data for emissions are available for earlier years than those for concentrations, but trends in concentrations were similar to those of emissions.[37] National total smoke emissions fell by an average of 65% between 1954 and 1970 and this is likely to be representative of emission trends in urban areas, where 80% of the population and most industry was concentrated in the 1970s.[37]

Figure 5-2 Average urban concentrations of smoke and emissions from coal combustion from industrial and domestic sources 1960-1985



Source: Harrison RM. Pollution: Causes, effects & control. 2nd edition. 1990[244] *Figure 5, page 135*

While SO_2 *emissions* rose between 1950-1962 and were steady to the early 1970s,[37] *concentrations* fell from the mid 1960s[266] related to the combination of decreases in domestic SO_2 emissions and the siting of new power stations and other emission sources outside towns. On average, concentrations in urban areas decreased by about 29% between 1961 and 1971.[37] An examination[266] of levels in central Lincoln showed a fall in concentrations in the city from 150 $\mu g/m^3$ in 1960 to 30 $\mu g/m^3$ in 1989/90, the latter levels being similar to those experienced in rural Lincolnshire over most of the time period. Because of rural siting of power stations, some rural areas now have slightly higher SO_2 concentrations than city areas.[246] Concentrations of SO_2 in industrial and residential sites in Greater London had converged to low levels by 1984/5 according to a review published in 1987.[296]

Regional levels of air pollution

Greater London 1933 to 1985

Laxen and Thompson[296] reviewed annual mean SO_2 concentrations in Greater London using data from the Department of Scientific and Industrial Research (1933-1960), the Warren Spring Laboratory reports of the National Survey (1959-1984) and a Greater London Council report for 1984/5. From 1963/4 there was a clear downward trend in SO_2 in all areas (four sites in the city, six sites in central London and 13 sites in the outer zone). Initially levels were highest in the City of London (approximately 300 μ/m^3), intermediate in central areas and lower in outer areas (approximately 200 μ/m^3), but had converged by 1984/5 to around 40 μ/m^3. Data prior to 1963/4 were only presented for sites around County Hall (central London) and no clear trend was seen there in 1931/2 to 1963/4. Annual mean SO_2 levels in central London were generally 300-400 μ/m^3 for 1938/9 to 1963/4 and slightly higher in the early 1930s.[296]

Highest daily SO_2 concentrations in Greater London were also examined for 1958/9 to 1984/5[296] and showed a decline following a peak in the winter of 1962/3, (presumably the December 1962 air pollution episode[297]) when maximum 24-hour sulphur dioxide level reached 4500 μ/m^3, levels similar to those seen in the 1952 smog. Highest daily winter values were approximately double the summer values in 1964/5, but the seasonal difference had virtually disappeared by 1984/5.[296]

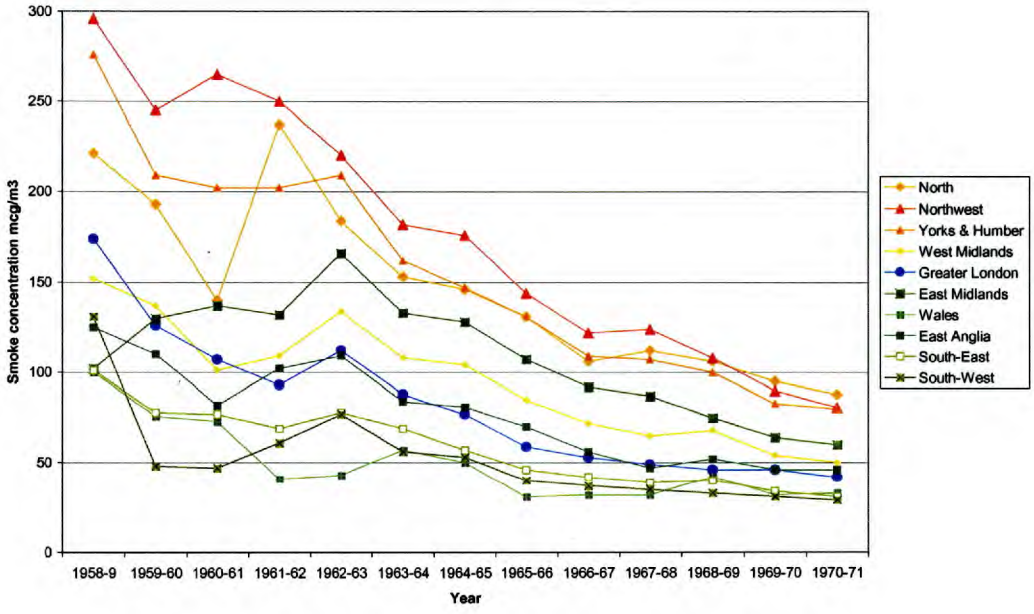
Analysis of regional air pollution data 1958-1971

Warren Spring survey data[37] covering 1958-1971 suggested that smoke and SO_2 concentrations fell in all regions in the late 1950s (and into the early 1960s in non-conurbation regions), rose slightly to a peak in 1962-63 then declined throughout the decade in most areas, consistent with the data for Greater London.[296] Figure 5-3 (page 118), created using the Warren Spring data, shows that conurbation regions (those containing conurbations as defined in this PhD) generally had the highest levels, while non-conurbations had the lowest levels. Greater London levels were intermediate for smoke but one of the highest for sulphur dioxide, being just lower than North-West and Yorkshire & Humberside until 1966-67 and the highest thereafter (Figure 5-3, page 118). The figures for smoke concentrations are consistent with data for coal consumption[37] (not shown).

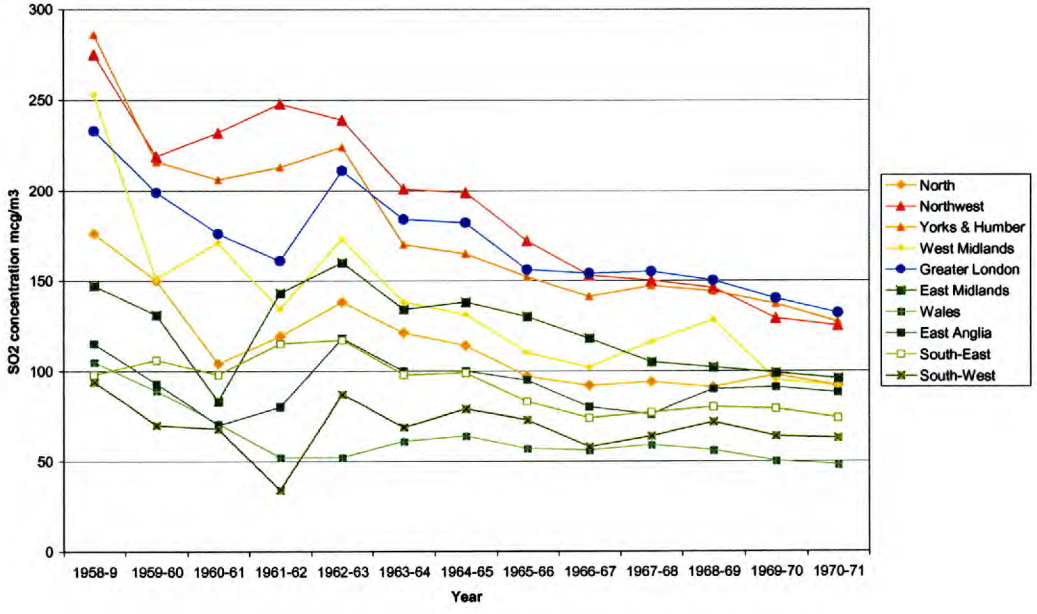
Results of linear regression conducted using these Warren Spring regional data to find average trends 1958-1971 are shown in Table 5-3 (page 119). Regression analyses of averages for conurbations and non-conurbations were then conducted by combining data for all regions containing conurbations and all regions not containing conurbations or Greater London. Such averages obviously have to be treated with caution as it is not known how representative the sites chosen are of air pollution in the country and combining the sites to make conurbation and non-conurbation averages may not be appropriate, but in the absence of more detailed information these figures give some indication of possible differences in air pollution concentrations between the conurbation, Greater London and non-conurbation areas.

Figure 5-3 Regional trends in annual average black smoke and sulphur dioxide concentrations 1958-1971 in England & Wales

Black Smoke



Sulphur dioxide



Source: National Survey of Air Pollution 1961-71, Warren Spring Laboratory[37]

Table 5-3 Levels of SO_2 and black smoke in conurbations and Greater London areas relative to measured levels in non-conurbations for 1958/9 and 1970/71, using smoothed sulphur dioxide and black smoke levels (in $\mu g/m^3$) from Warren Spring data for regions and derived area averages or actual levels

	CONURBATIONS					GREATER LONDON	NON-CONURBATIONS							
	North	Northwest	Yorkshire & Humberside	West Midlands	Average conurbations*		East Midlands	Wales	East Anglia	South-East	South-West	Average non-conurbations*	Conurbation level relative to non-conurbations	Greater London level relative to non-conurbations
SO_2 linear regression trend lines														
r^2 for trend	0.67	0.91	0.83	0.66	N/A	0.72	0.30	0.55	0.08	0.63	0.06	N/A		
1958/9	147.5	263.6	244.3	192.1	169.5	210.2	141.5	82.0	98.1	110.8	74.3	101.3	**1.67**	**2.07**
1970/1	81.1	119.2	113.8	83.9	79.6	133.4	102.5	44.3	85.9	73.5	63.4	73.9	**1.08**	**1.80**
SO_2 using actual values for non-conurbations because of low r^2 values seen in regression trend lines														
1958/9							147	105	115	98	94	111.8	**1.52**	**1.88**
1970/1							96	48	88	74	63	73.8	**1.08**	**1.81**
Black smoke linear regression trend lines														
r^2 for trend	0.76	0.96	0.93	0.88	N/A	0.86	0.58	0.67	0.87	0.92	0.53	N/A		
1958/9	212.3	286.1	245.2	142.7	177.2	139.0	146.2	76.5	117.2	90.6	82.7	102.6	**1.73**	**1.35**
1970/1	81.9	68.2	65.2	47.9	52.6	25.9	71.1	22.3	38.2	26.3	21.6	35.9	**1.47**	**0.72**
Black smoke using actual values for Greater London because of higher levels in 1958/9														
1958/9						174								**1.70**
1970/1						42								**1.17**

*Averages are the averages of the smoothed regional levels taken from regression trend lines

Air pollution episodes identified from the literature

Only information on episodes in Greater London could be located.[297,298] There were high pollution episodes in London in 1952, 1956, 1957, 1959, 1962 and 1991 but the estimated excess deaths were greatest in the 1952 and 1956 episodes (Table 5-4). In the 1952 episode, approximately 1,000 of the 4,000 excess deaths in the two weeks during and after the fog were due to bronchitis,[293] which gives an approximate 10% increase to the year's total mortality for Greater London (see column 4 in Table 5-4) relating to those two weeks alone. The chief pollutants in the 1991 episode[298] were nitrogen dioxide and black smoke, with peak levels over 3 times higher than the usual December levels. However, SO_2 levels were not raised.

Table 5-4 Notable episodes of high pollution in London from 1950 onwards

Year	Days	Estimated excess total deaths* (all causes)	Number of COPD deaths in ages 15+ years (males+females) in that year	Maximum 24-hour black smoke level ($\mu g/m^3$)	Maximum 24-hour sulphur dioxide level ($\mu g/m^3$)
1952	December 5-8	4,000	10,913	4460	3830
1956	January 3-6	1,000	10,257	2830	1430
1957	December 2-5	750	9365	2417	3335
1959	January 26-31	250	10,020	1723	1850
1962	December 3-7	700	10,071	3144	3834
1975	December	100-200	7854	546	994
1991	December 12-18	101 to 178	7356	148	112

*comparison with preceding week for 1952-62, no details given for 1975, comparison with control areas for 1991
Source: Holland et al, 1979[297] for 1952-75, Anderson et al, 1995[298] for 1991

Trends from the 13 monitoring stations recording from the 1950s to at least the 1980s

Trends in annual average readings for each monitoring station with long-running readings showed a decline in black smoke levelling off in the mid to late 1970s (Figures 5-4a, 5-4b & 5-4c, pages 123-125), and a decline levelling off in the mid 1980s for SO_2 (Figures 5-5a, 5-5b & 5-5c, pages 126-128). Peak annual averages of black smoke and SO_2 were seen in the 1950s and early 1960s, with the exception of two medium or low density housing monitoring sites for SO_2 (Table 5-5, page 122). The highest averages were seen in commercial sites (code D1). The worst site for both black smoke and SO_2 was Leeds 4, classified as in a conurbation area for this PhD analysis. The next highest averages were seen in high density housing areas located in Greater London and South Yorkshire (South Yorkshire was only designated a conurbation in 1974 and has therefore been classed as a non-conurbation area in this analysis). Because of the small

number of sites in different environments, it is difficult to determine from these data whether conurbations or Greater London experienced worse air pollution. However, the two non-conurbation sites excluding South Yorkshire did appear to have lower peak values than conurbations or Greater London. Lowest annual averages for all sites occurred in the late 1980s and 1990s for both pollutants (Table 5-5, page 122), with all types of sites declining to comparably low levels.

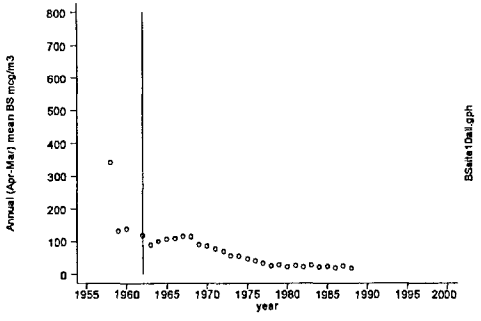
Table 5-5 Descriptive data from 13 monitoring sites with long-term readings (restricted to years with >330 days of data available)

Site name	County	Site code	Type of site	Years when recording BS	Number of years with >330 days recorded for BS (max 43)	Number of years with >330 days recorded for SO_2 (max 43)	Annual mean black smoke in $\mu g/m^3$ Lowest reading and years above Highest reading and years below	Annual mean SO_2 in $\mu g/m^3$ Lowest reading and years above Highest reading and years below
Conurbations								
Bradford 6	West Yorkshire	D1/E	Commercial area / Smoke control area	1954-1996	20	19	16 in 1985, 1990, 1991,1993,1994 340 in 1955	18 in 1994 406 in 1958
Elland 2	West Yorkshire	D2	Town centre	1954-1996	37	35	12 in 1988 233 in 1962	16 in 1990 231 in 1964
Leeds 4	West Yorkshire	D1/E	Commercial area/ Smoke control area	1954-1987	30	30	19 in 1983, 1986, 1987 820 in 1956	30 in 1980 429 in 1955
Newcastle Upon Tyne 17	Tyne & Wear	B3	Residential: medium density + parks, fields OR low density housing	1958-1988	26	26	17 in 1986, 1988 340 in 1958	46 in 1987 173 in 1972
Greater London								
Deptford 1	Greater London	A1/E	Residential: high density housing / Smoke control area	1955-1986	28	28	20 in 1984, 1985 400 in 1958	37 in 1985 257 in 1955
Ealing 3	Greater London	B2	Residential: medium density housing + industry	1958-1987	17	17	13 in 1977, 1978 144 in 1958	26 in 1985, 1986 231 in 1964
Hackney 4	Greater London	A1	Residential: high density housing	1955-1990	30	30	17 in 1987 370 in 1955	33 in 1984,1988 229 in 1955-57
Lambeth 7	Greater London	A1	Residential: high density housing	1955-1996	29	29	14 in 1985, 1986, 1990,1993 400 in 1955	31 in 1985 257 in 1955, 1956
Southwark 3	Greater London	D1	Commercial area	1954-1986	20	19	17 in 1986 530 in 1955	32 in 1986 343 in 1955
St Marylebone 6	Greater London	D1	Commercial area	1955-1987	22	22	17 in 1987 480 in 1955	39 in 1987 332 in 1958
Non-conurbation								
Dartford 6	Kent	D2	Town centre	1958-1985	23	23	21 in 1982 109 in 1958	48 in 1984 177 in 1962
Darton 1	South Yorkshire	B2	Residential: medium density housing + industry	1954-1987	15	14	24 in 1985, 1986 205 in 1962	54 in 1975, 1987 91 in 1981
Sheffield 40	South Yorkshire	A2	Residential: high density housing + industry	1955-1991	12	12	17 in 1987 350 in 1955	44 in 1987, 1991 278 in 1958

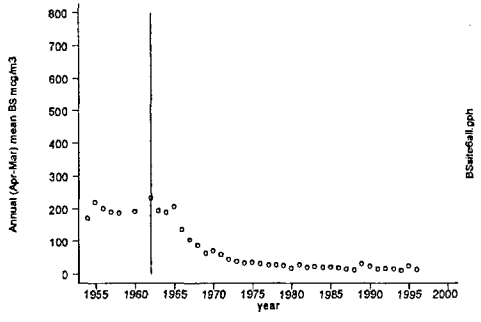
Figure 5-4a Black smoke readings from long-running monitoring sites in conurbation areas

Data prior to 1962 (marked by a vertical line) were obtained from published tables supplied by AEAT where data were only published if 12 months of data were available. Data 1962 onwards are annual means from daily data from the Data Quality Archive (ignoring missing days).

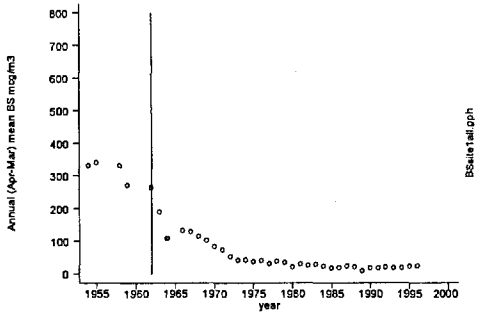
Newcastle on Tyne 17, B3
Low density housing



Elland 2, D2
Town centre



Bradford 6, D1/E
Commercial area



Leeds 4, D1/E
Commercial area

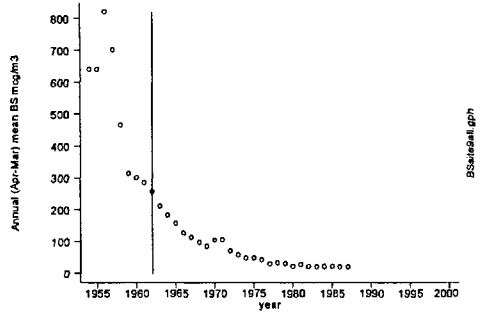
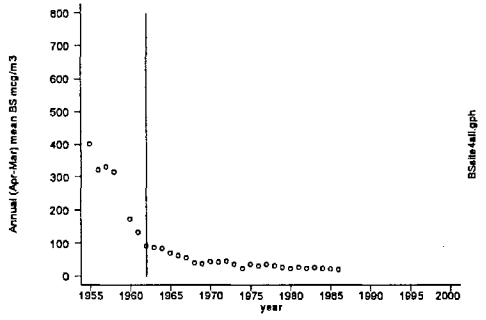


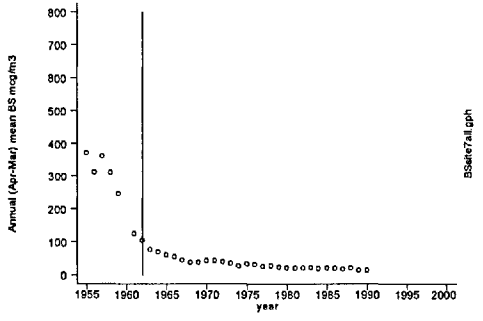
Figure 5-4b Black smoke readings from long-running monitoring sites in Greater London

Data prior to 1962 (marked by a vertical line) were obtained from published tables supplied by AEAT where data were only published if 12 months of data were available. Data 1962 onwards are annual means from daily data from the Data Quality Archive (ignoring missing days).

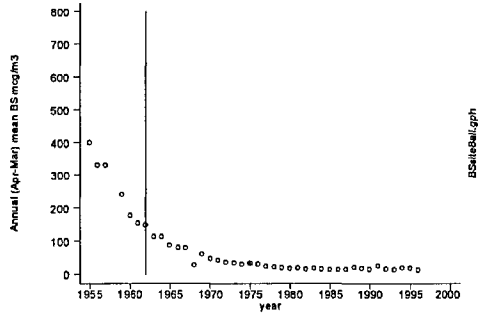
Deptford 1, A1/E
High density housing



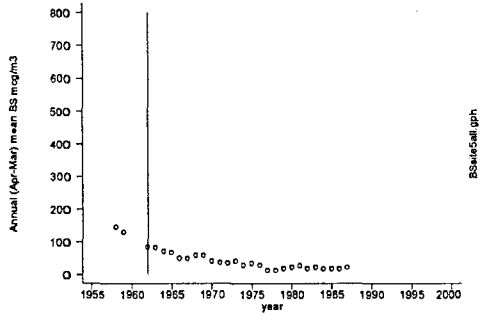
Hackney 4, A1
High density housing



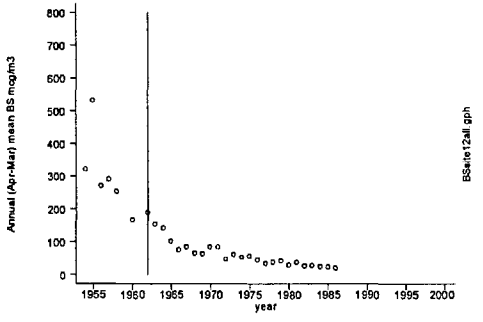
Lambeth 7, A1
High density housing



Ealing 3, B2.
Medium density housing + industry



Southwark 3, D1
Commercial area



St Marylebone 6, D1
Commercial area

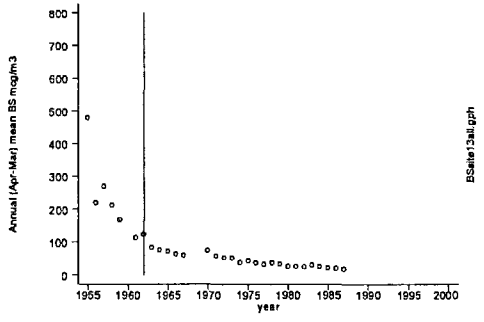
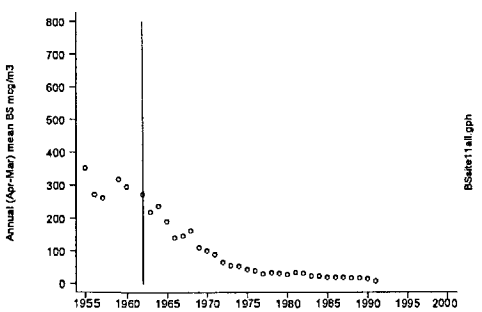
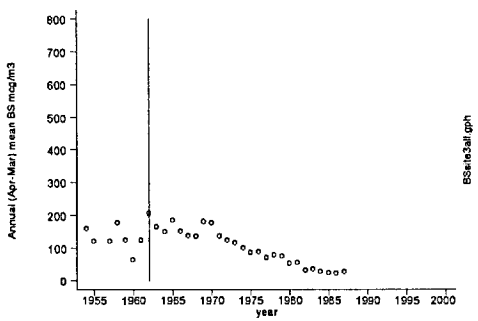


Figure 5-4c Black smoke readings from long-running monitoring sites in non-conurbation areas
Data prior to 1962 (marked by a vertical line) were obtained from published tables supplied by AEAT where data were only published if 12 months of data were available. Data 1962 onwards are annual means from daily data from the Data Quality Archive (ignoring missing days).

Sheffield 40, A2
High density housing + industry



Darton 1, B2
Medium density housing + industry



Dartford 6, D2
Town centre

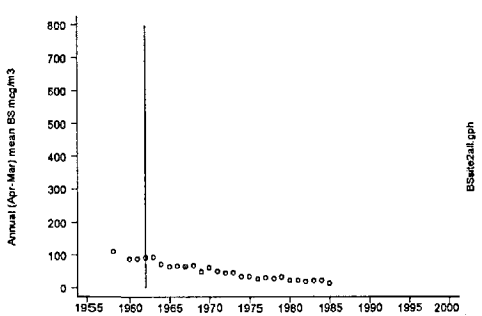
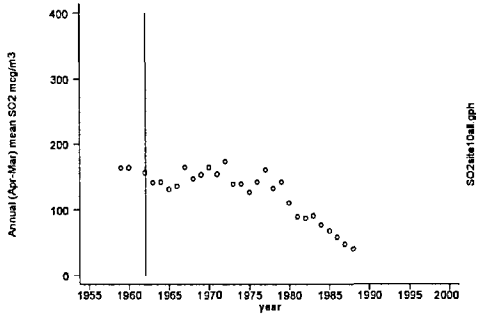


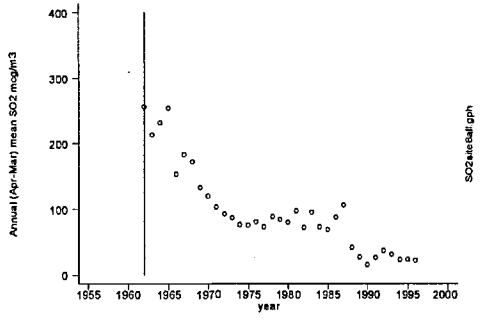
Figure 5-5a Sulphur dioxide readings from long-running monitoring sites in conurbation areas

Data prior to 1962 (marked by a vertical line) were obtained from published tables supplied by AEAT where data were only published if 12 months of data were available. Data 1962 onwards are annual means from daily data from the Data Quality Archive (ignoring missing days).

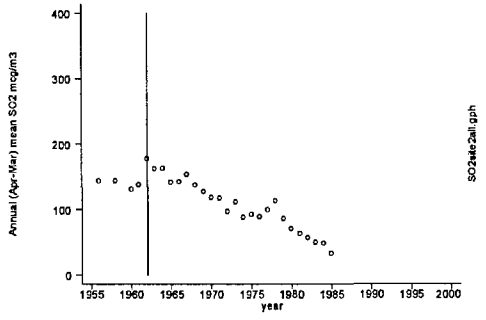
Newcastle on Tyne 17, B3
Low density housing



Elland 2, D2
Town centre



Bradford 6, D1/E
Commercial area



Leeds 4, D1/E
Commercial area

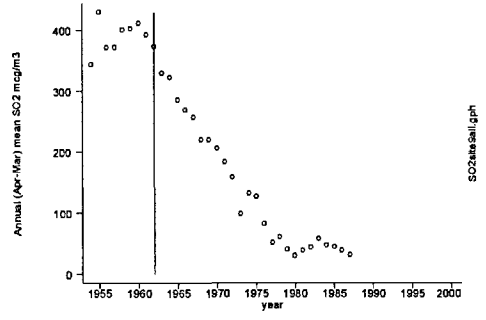
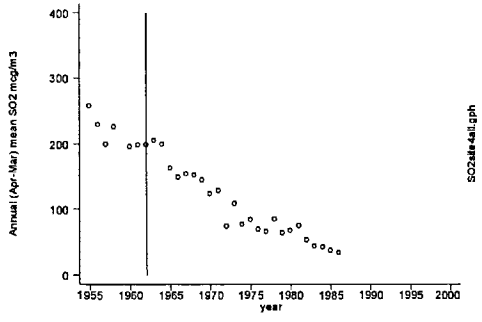
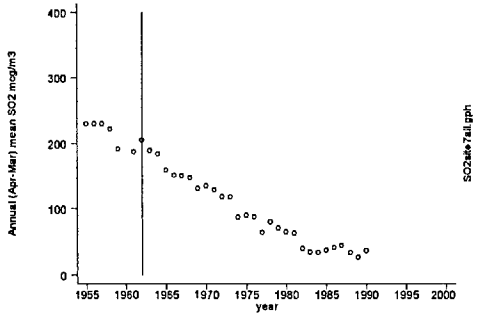


Figure 5-5b Sulphur dioxide readings from long-running monitoring sites in Greater London
Data prior to 1962 (marked by a vertical line) were obtained from published tables supplied by AEAT where data were only published if 12 months of data were available. Data 1962 onwards are annual means from daily data from the Data Quality Archive (ignoring missing days).

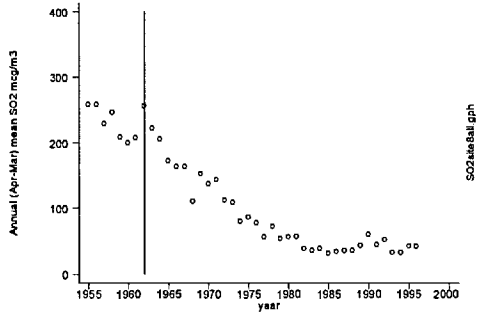
Deptford 1, A1/E
High density housing



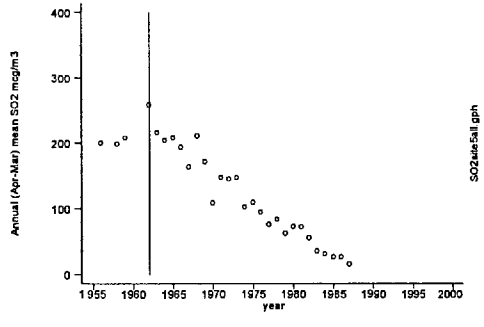
Hackney 4, A1
High density housing



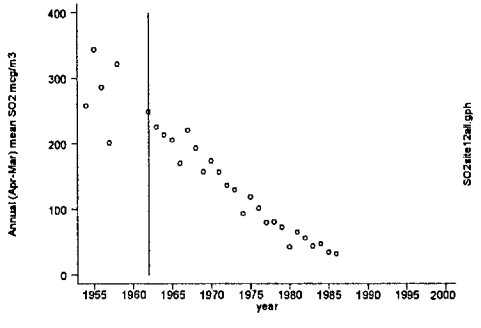
Lambeth 7, A1
High density housing



Ealing 3, B2.
Medium density housing + industry



Southwark 3, D1
Commercial area



St Marylebone 6, D1
Commercial area

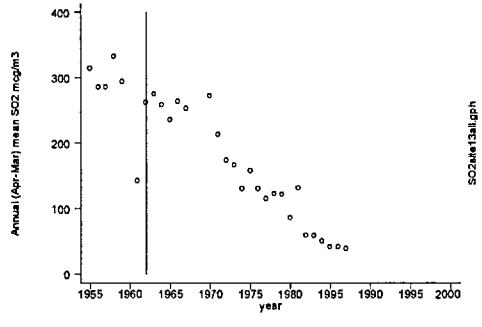
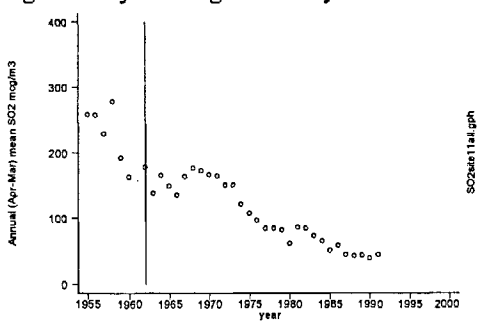


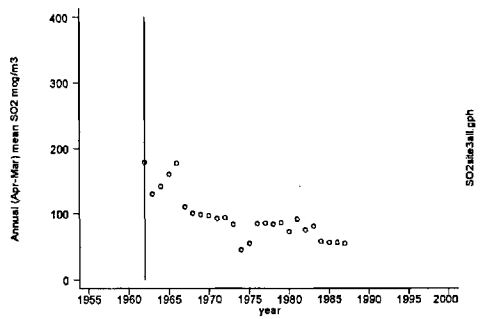
Figure 5-5c Sulphur dioxide readings from long-running monitoring sites in non-conurbation areas

Data prior to 1962 (marked by a vertical line) were obtained from published tables supplied by AEAT where data were only published if 12 months of data were available. Data 1962 onwards are annual means from daily data from the Data Quality Archive (ignoring missing days).

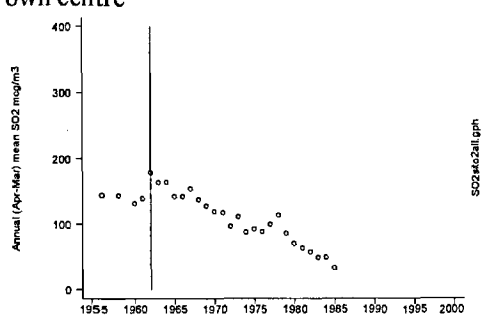
Sheffield 40, A2
High density housing + industry



Darton 1, B2
Medium density housing + industry



Dartford 6, D2
Town centre



Descriptive analyses of mortality data

Numbers of deaths and population

Over the time period 1950-99 there were a total of just under 1.4 million deaths from COPD and just over 1.4 million deaths from lung cancer in those aged 15 years and older (Table 5-6). The percentage of these deaths occurring in conurbations (23.5% COPD, 22.0% lung cancer) and Greater London (16.5% COPD, 17.1% lung cancer) was higher than the percentage of population living in these areas (19.1% and 15.6% respectively). The average number of deaths per year from COPD and lung cancer were approximately 6,000 each in conurbations, 4,500 each in Greater London and 16,000 and 17,000 respectively in non-conurbations. However, actual numbers varied markedly across years (Table 5-6) with highest numbers of COPD deaths in the 1950s and early 1960s but highest numbers of lung cancer deaths in the 1970s and 1980s.

Table 5-6 Numbers of COPD and lung cancer deaths and population for ages 15 years and older in different areas of England & Wales 1950-99

	Conurbations	Greater London	Non-conurbations	England & Wales
COPD deaths (% of England & Wales) 1950-99	321,983 (23.5%)	225,659 (16.5%)	820,127 (60.0%)	1,367,894 (100%)
Lung cancer deaths (% of England & Wales) 1950-99	312,459 (22.0%)	243,292 (17.1%)	866,446 (60.9%)	1,422,322 (100%)
Population (% of England & Wales) 1950-99	364 million (19.1%)	299 million (15.6%)	1,246 million (65.3%)	1,909 million (100%)
COPD				
Mean number (SD) of COPD deaths per year	6,440 (793.6)	4,513 (1,185.5)	16,402 (1,704.5)	27,358 (2,903)
Lowest number of deaths per year	5,342 in 1978	2,949 in 1998	13,201 in 1952	22,077 in 1977
Highest number of deaths per year	9,291 in 1951	7,532 in 1951	20,154 in 1963	36,366 in 1951
Lung cancer				
Mean number (SD) of lung cancer deaths per year	6,240 (1,716.9)	4,866 (835.8)	17,337 (4,941.1)	28,446 (6,988.0)
Lowest number of deaths per year	2,782 in 1950	3,115 in 1950	6,330 in 1950	12,238 in 1950
Highest number of deaths per year	8,198 in 1983	5,864 in 1972	22,420 in 1985	35,728 in 1985

Age-stratified rates by area

In males, there were declines in COPD mortality for all ages, while lung cancer mortality showed declines for younger age-groups but a rise then fall in older age-groups. COPD mortality rates by age in females showed declines in younger age-

groups, but a levelling off of the decline in ages 55-65 and rises in ages 65+ years after around 1975. Lung cancer rates by age in females showed patterns consistent with the known 1925 cohort effect. There were initial increases followed by declines spaced at 10 year intervals, except for ages 75+ where no decline was seen. This suggests that examination of age-standardised rates in descriptive analyses (below) were appropriate for COPD in males, but will have obscured some detail for COPD in females and lung cancer.

More detailed comparison of each age-sex stratum in different areas revealed the following:

COPD age-stratified mortality rates

COPD rates (Figures 5-6a, page 132 & 5-6b, page 133) were highest in conurbations throughout and generally lowest in non-conurbations, except for 1952, the year of the Greater London smog, when rates in Greater London were higher than conurbations for older age-groups (males aged 65+ years and females aged 75+ years). Rates in conurbations converged towards those of non-conurbations in the 1970s in males of all ages and in younger and older females (ages <55 years and >74 years). Rates became broadly similar in all areas in later years in younger age groups (<55 years in males and <45 years in females).

Rates in Greater London relative to conurbations and non-conurbations changed over time and by age-groups (Figure 5-6a, page 132 & 5-6b, page 133) with broadly similar patterns in males and females, suggesting complex interactions between area related risk factors and period factors (as males and females were both affected, changes occurred at similar times for different age-groups and graphs did not show an obvious cohort pattern). Initially, in the early 1950s, COPD rates for Greater London were intermediate between conurbations and non-conurbations for all age-groups in males and ages 45+ years in females. Rates in Greater London from the late 1950s and early 1960s onwards then became slightly *lower* than non-conurbations for younger age-groups in both sexes (ages 25-34 & 35-44 and also, in females, ages 45-54), similar in to non-conurbations in middle ages (45-54 years in males, 55-64 years in females) but slightly higher than non-conurbations for older age-groups (ages 55-64 in males, ages 65-74 years in females). Rates in the mid 1960s onwards were reasonably similar in non-conurbations and Greater London for all except the oldest age-groups (75+ years). For these oldest age-

groups, both males and females showed fairly similar rates in conurbations and Greater London but lower rates in non-conurbations throughout 1950-1999.

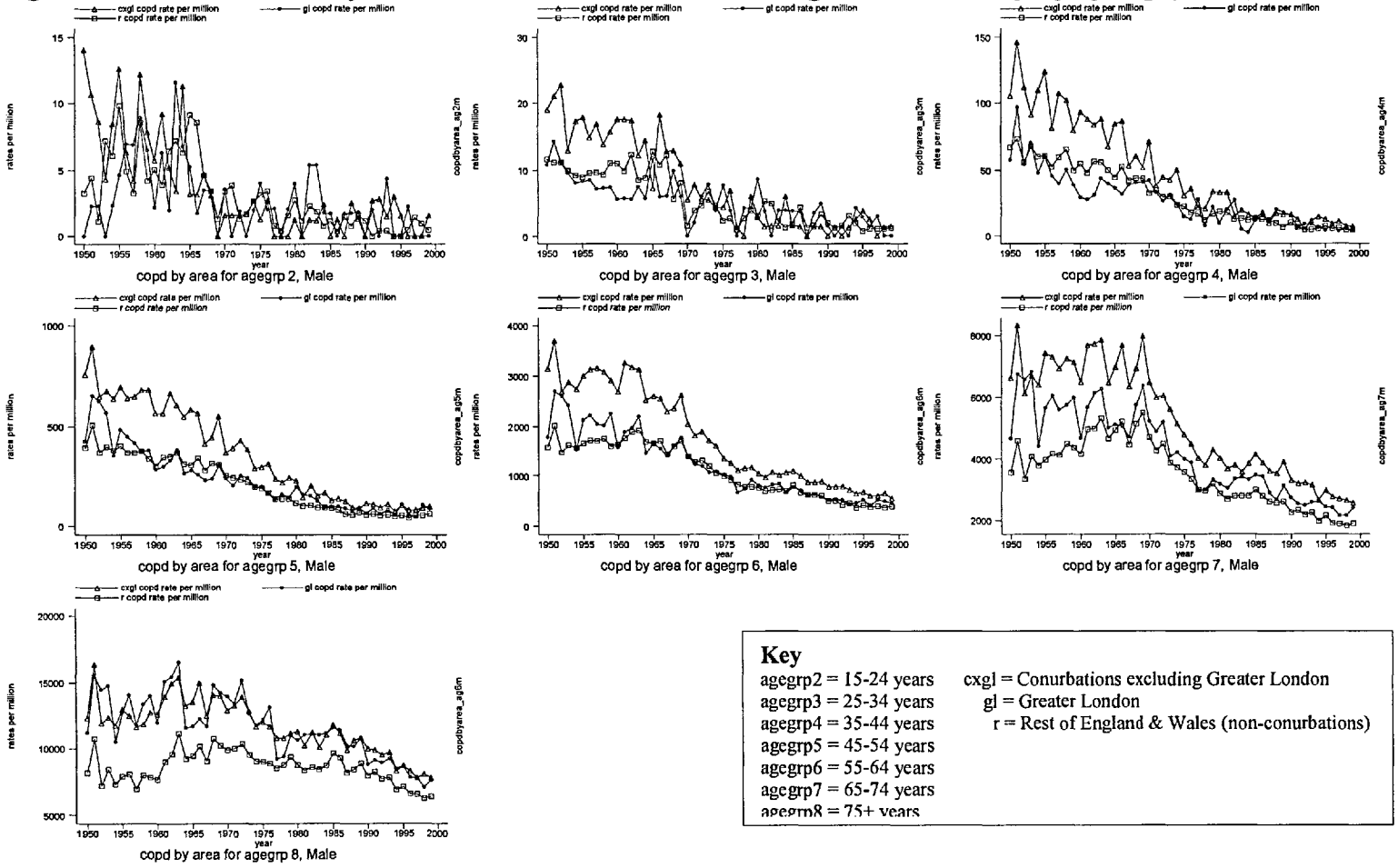
Lung cancer age-stratified mortality rates

For lung cancer, broadly similar patterns were visible in all areas consistent with cohort effects (higher mortality rates throughout life) for males born 1900-1915 (Figure 5-7a, page 134) and women born around 1925-30 (Figure 5-7b, page 135), but with minor differences in timing of cohort effects in different areas. Rates were lowest in non-conurbations throughout. The position of Greater London relative to conurbations and non-conurbations varied, suggestive of interactions between cohort and area-related risk factors. For both COPD and lung cancer, numbers were small for ages 15-24 and 25-34 years, with consequent year-to-year fluctuations, so were difficult to interpret.

There was some suggestion that the cohort effect for males in conurbations occurred five to ten years later than in Greater London, as peak rates for ages 45-54, 55-64, 65-74 and 75+ years were seen around 1955, 1965, 1970 and 1975 respectively in Greater London and around 1965, 1970, early 1970s and early 1980s in conurbations – consistent with cohort effects for those born around 1905 in Greater London and 1915 in conurbations. In contrast, the female cohort effect was seen at the same time in each area, with peak rates for each area seen around the late 1960s, 1975-77, 1988 and 1995 for ages 35-44, 45-54, 55-64 and 65-74 years respectively, while peak rates may not have been reached by the end of the analysis period for ages 75+ years. This was consistent with a cohort effect for those born around 1925-30.

In males (Figure 5-7a, page 134), lung cancer mortality rates in Greater London were initially either greater than those in conurbations (ages 65+ years) or similar to those in conurbations (ages <65 years) but became lower than those in conurbations about 10 years after the peak rates (1960, 1965, 1975 and the late 1980s for the oldest four age-groups) and fell swiftly towards those in non-conurbations. In females (Figure 5-7b, page 135), lung cancer mortality rates were also initially highest in Greater London. Rates in females became lower than those in conurbations about 10 years *before* the peak rates in females and 5-10 years later than in males (i.e. suggestive of a cohort effect for females and not suggestive of a period effect affecting both sexes) at around the mid 1950s, 1966-69, 1979, 1986 and 1994 for ages 35-44, 45-54, 55-64, 65-74 and 75+ years respectively.

Figure 5-6a COPD mortality rates 1950-1999 for each area in England & Wales by age-group (males)

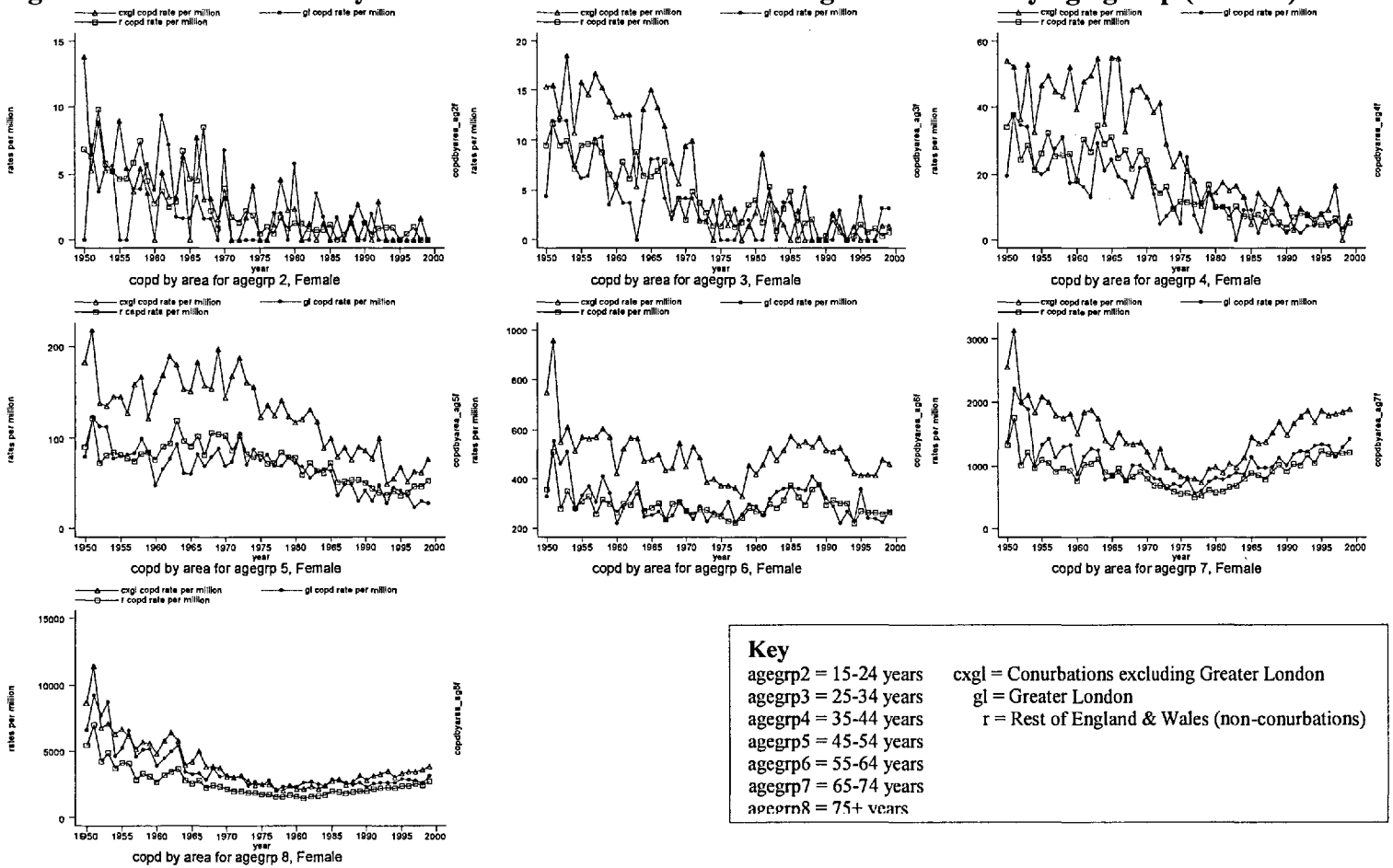


Key

agegrp2 = 15-24 years
agegrp3 = 25-34 years
agegrp4 = 35-44 years
agegrp5 = 45-54 years
agegrp6 = 55-64 years
agegrp7 = 65-74 years
agegrp8 = 75+ years

cxgl = Conurbations excluding Greater London
gl = Greater London
r = Rest of England & Wales (non-conurbations)

Figure 5-6b COPD mortality rates 1950-1999 for each area in England & Wales by age-group (females)

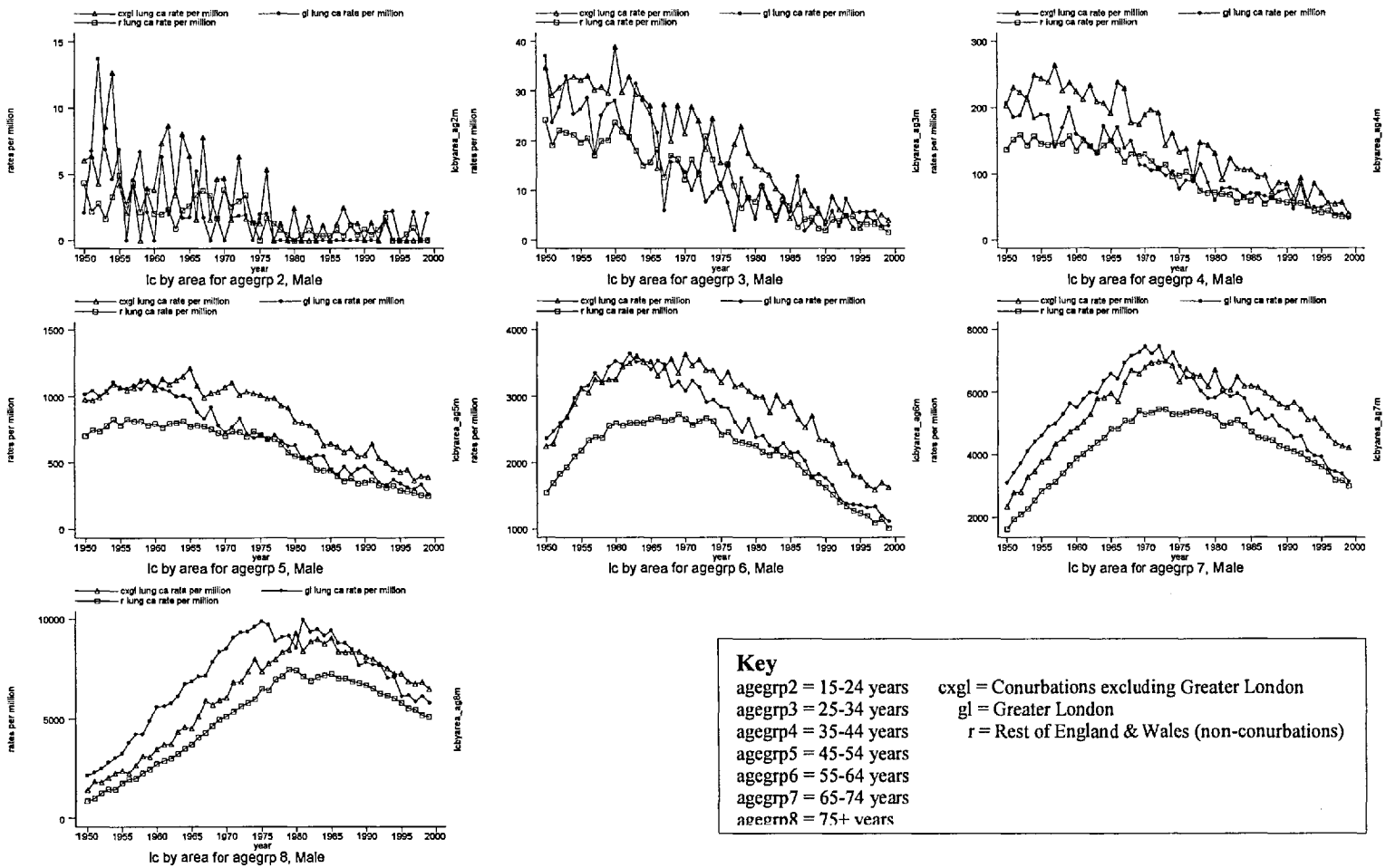


Key

agegrp2 = 15-24 years
agegrp3 = 25-34 years
agegrp4 = 35-44 years
agegrp5 = 45-54 years
agegrp6 = 55-64 years
agegrp7 = 65-74 years
agegrp8 = 75+ years

cxgl = Conurbations excluding Greater London
gl = Greater London
r = Rest of England & Wales (non-conurbations)

Figure 5-7a Lung cancer mortality rates 1950-1999 for each area in England & Wales by age-group (males)

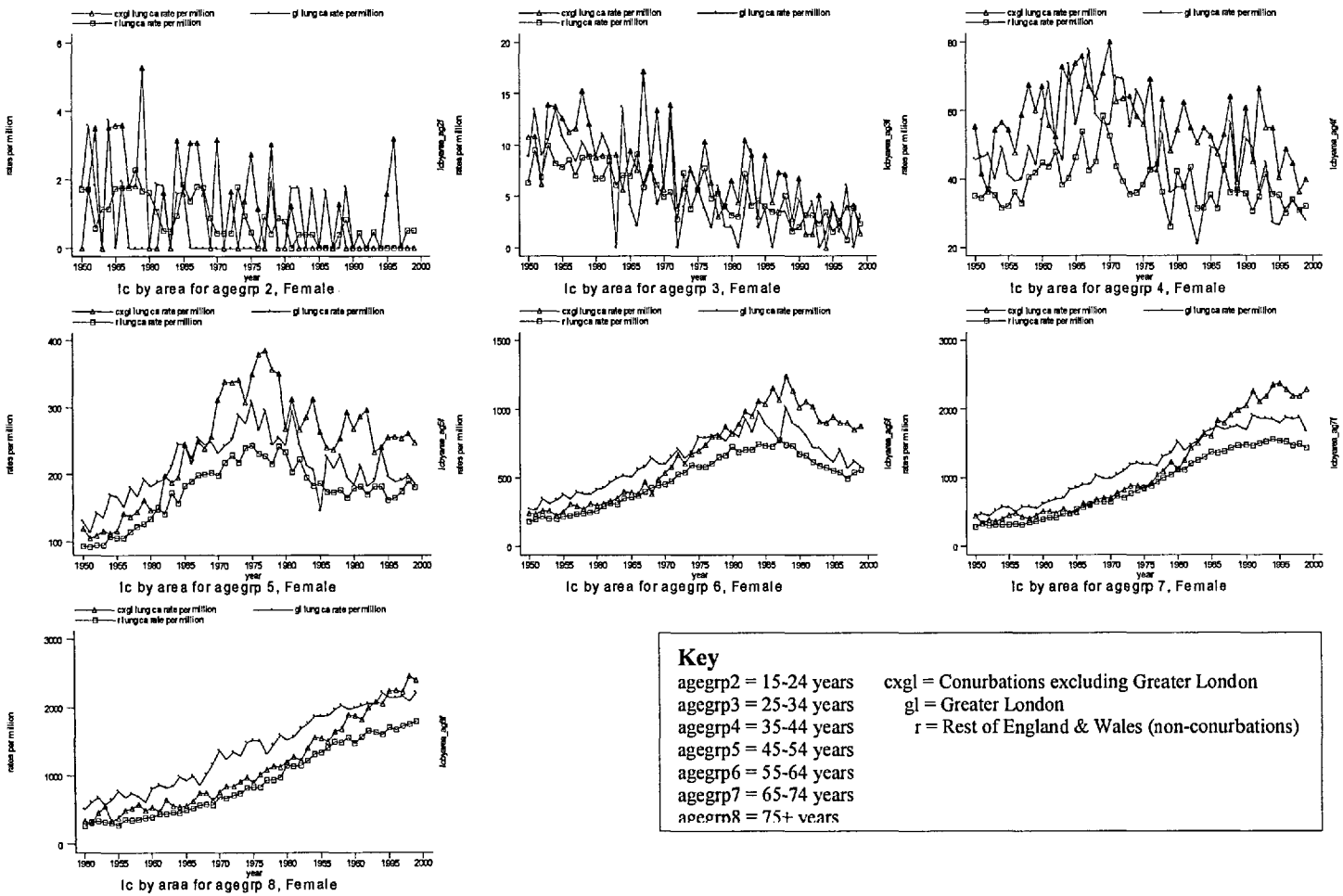


Key

agegrp2 = 15-24 years
agegrp3 = 25-34 years
agegrp4 = 35-44 years
agegrp5 = 45-54 years
agegrp6 = 55-64 years
agegrp7 = 65-74 years
agegrp8 = 75+ years

cxgl = Conurbations excluding Greater London
gl = Greater London
r = Rest of England & Wales (non-conurbations)

Figure 5-7b Lung cancer mortality rates 1950-1999 for each area in England & Wales by age-group (females)



Key

agegrp2 = 15-24 years
agegrp3 = 25-34 years
agegrp4 = 35-44 years
agegrp5 = 45-54 years
agegrp6 = 55-64 years
agegrp7 = 65-74 years
agegrp8 = 75+ years

cxgl = Conurbations excluding Greater London
gl = Greater London
r = Rest of England & Wales (non-conurbations)

Age-standardised rates by area

COPD mortality

Rates of COPD age-standardised to the 1999 population were highest in conurbations, intermediate in Greater London and lowest in non-conurbations over 1950-99 (see Figure 5-8, page 139). All areas showed peak rates in 1951 (an influenza epidemic year), with age-standardised rates in conurbations at 25.7 per 10,000 in men and 18.1 per 10,000 in women compared with 15.3 per 10,000 in males and 10.9 per 10,000 in females in non-conurbations. The 1951 peak was sustained in 1952 and 1953 in Greater London but not in other areas, which would correspond to the December 1952 smog, with high influenza levels plus late registrations of deaths accounting for the 1953 peak.

COPD mortality rates for both males and females suggested convergence towards rates in non-conurbations between 1950 to the mid 1970s (Figure 5-8, page 139). This was confirmed through inspection of graphs of absolute differences (not shown) and ratios of rates of conurbations: non-conurbations and Greater London: non-conurbations (Figure 5-10, page 141). However, some subsequent divergence of rates for females between conurbations and non-conurbations was seen in the 1980s for absolute (but not relative) differences, while the difference between Greater London and non-conurbations became smaller mainly because of a rise in rates in non-conurbations. Rates of COPD mortality were approximately two-thirds higher in conurbations than in non-conurbations for males and females in the early 1950s, falling to a third higher in males and just under 50% higher in females in the early 1980s with little change thereafter (Table 5-7).

Table 5-7 Percentage difference between age-standardised (to 1999 population) rates in conurbations or Greater London and rates in non-conurbations averaged over different time periods

	1950-54	1950-51,1953-55	1980-84	1995-99
Males				
Conurbations: non-conurbations	+67.5%	+67.8%	+31.0%	+31.0%
Greater London: non-conurbations	+49.7%	+41.8%	+23.1%	+19.1%
Females				
Conurbations: non-conurbations	+65.9%	+65.7%	+47.6%	+46.0%
Greater London: non-conurbations	+41.6%	+31.8%	+41.7%	+11.3%

COPD rates in Greater London were approximately 40% higher in men than in non-conurbations in the early 1950s excluding 1952 (the year of the great London smog), falling to a fifth higher. In women, Greater London had a third higher COPD rates than

non-conurbations in the early 1950s, *rising* to 40% higher in the early 1980s then falling to only 10% higher in the late 1990s partly due to a rise in rates in conurbations.

Lung cancer mortality

Lung cancer rates age-standardised to 1999 population (Figure 5-9, page 140) were lowest in non-conurbations throughout but were otherwise dissimilar to those for COPD in both time trends and relative differences between urban and less urban areas. In males, rates in all areas rose over time to peak in the mid 1970s then fell. In females, rates rose throughout in conurbations, but the rise levelled off in the late 1980s in Greater London and non-conurbations. No peaks in 1952 or 1953 could be detected in Greater London.

Unlike COPD mortality rates, highest rates were initially seen in Greater London in both sexes, with rates in conurbations overtaking Greater London in 1976 in men and 1986 in women. The differences between Greater London and non-conurbations lessened over time in males (Figure 5-9, page 140), but remained fairly similar between conurbations and non-conurbations (in fact, increasing slightly in the middle of the study period if looking at absolute differences (not shown), decreasing if looking at relative differences (Figure 5-9, page 140). In women, the differences between Greater London and non-conurbations also decreased over time if looking at relative differences, but rose slightly in absolute terms. In contrast with males, differences between conurbations and non-conurbations increased with time.

COPD : lung cancer ratios

Ratios of age-standardised COPD to lung cancer rates by area (Figure 5-10, page 141) showed a convergence from ratios of >5 in women and >1.5 in men initially to between 0.8 to 1.0 for women and 0.8 for men by the mid 1970s. Ratios were higher in women than in men and this was most noticeable in the earlier years of the study. Highest ratios in both sexes were seen in 1951 with ratios of 14.1 in conurbations, 9.1 in Greater London and 7.6 in non-conurbations for women; corresponding values for men were 2.9, 2.1 and 2.4 respectively. Ratios in women (but not men) in Greater London did not experience the marked drop in 1952 and 1953 seen in other areas. Conurbations generally experienced highest ratios until the mid 1970s for men and throughout for women. Ratios became almost the same for all areas for men between 1980-1992, with a small increase in ratio in Greater London from 1993-99.

Influenza mortality

Influenza mortality in England & Wales decreased over time to very low levels by the late 1990s, with some (mainly younger) age-groups experiencing no deaths in some years in the 1980s and 1990s (Figure 5-11, page 142). Twelve years where a peak in influenza mortality was superimposed on this overall decline were identified visually (Figure 5-8 on page 139 and detailed in Table 5-9 on pages 155-156 containing results from BAMP analyses) – only one of which occurred in the last 20 years (1989). These peak years clearly affected all age-groups (Figure 5-11, page 142) and therefore might be expected to produce a period effect.

Peaks in influenza mortality for England & Wales clearly coincided with peaks in COPD mortality in 1951 and 1953 (Figure 5-8, page 139) but did not appear to affect lung cancer mortality (Figure 5-9, page 140) and thus resulted in peaks in COPD:lung cancer mortality rate ratios (Figure 5-10, page 141). Visual inspection suggested some relationship between influenza peaks and COPD mortality more noticeably in women for some later years (1959, 1961, 1966, 1968, 1989) bearing in mind that influenza peaks in the winter months, which may straddle the calendar year period used in this analysis. However, some years with influenza peaks (1957, 1970, 1972, 1973, 1976) did not have an obvious related COPD mortality peak (except perhaps Greater London for 1976).

Figure 5-8 Directly age-standardised (to 1999 England & Wales population) mortality rates for COPD ages 15+ years in conurbations excluding Greater London, Greater London and non-conurbation areas and influenza in England & Wales for 1950-99

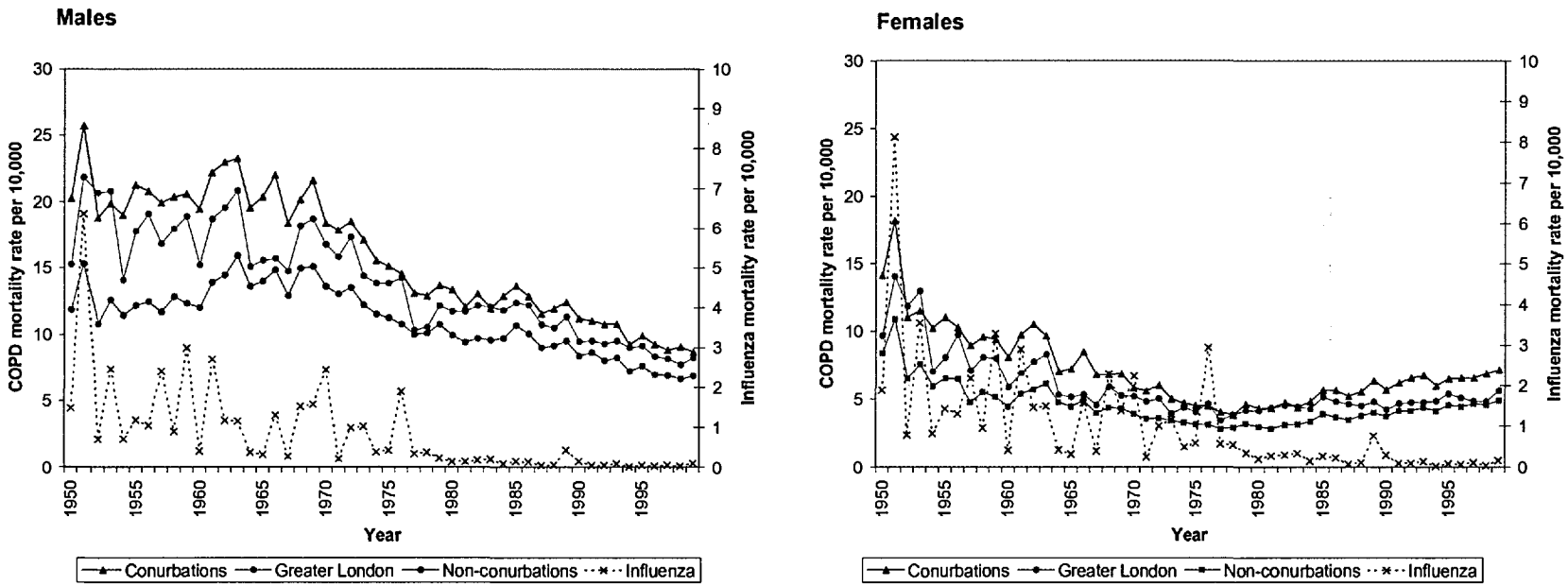


Figure 5-9 Directly age-standardised (to 1999 England & Wales population) mortality rates for lung cancer in ages 15+ in conurbations (excluding Greater London), Greater London and non-conurbation areas for 1950-99

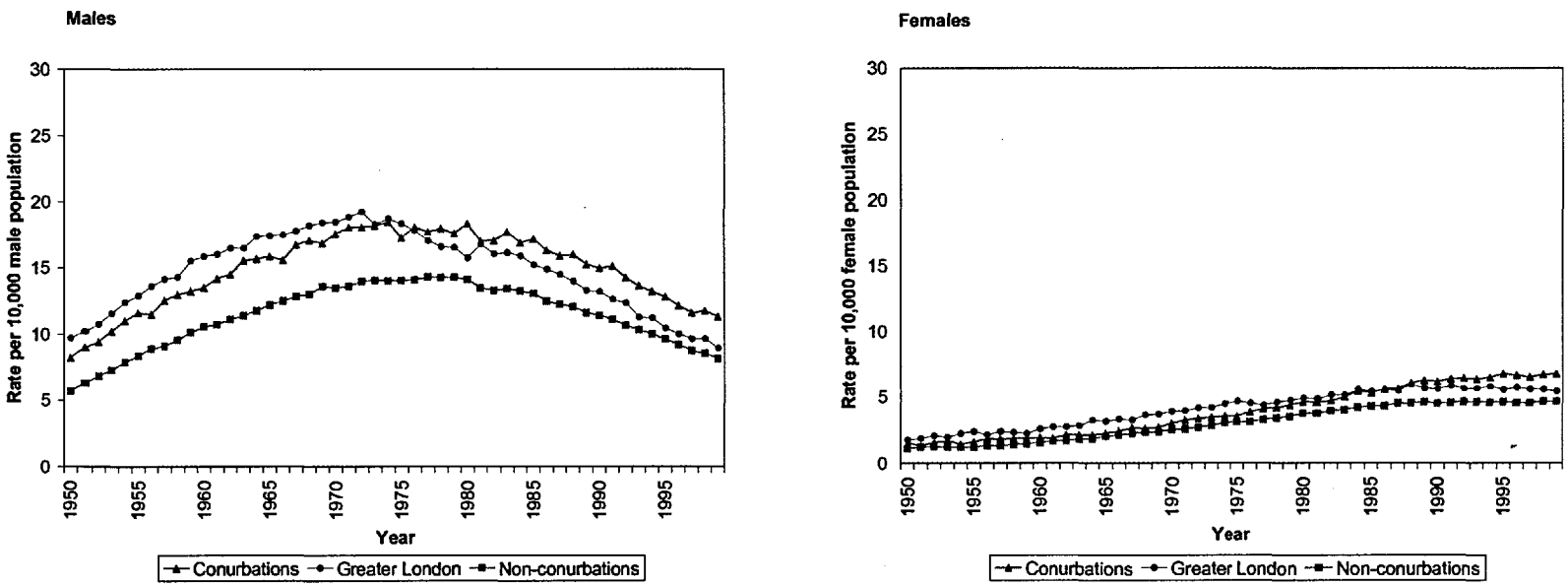


Figure 5-10 Ratio of COPD: lung cancer (age-standardised) mortality rates in conurbations (excluding Greater London), Greater London and non-conurbations areas for 1950-99

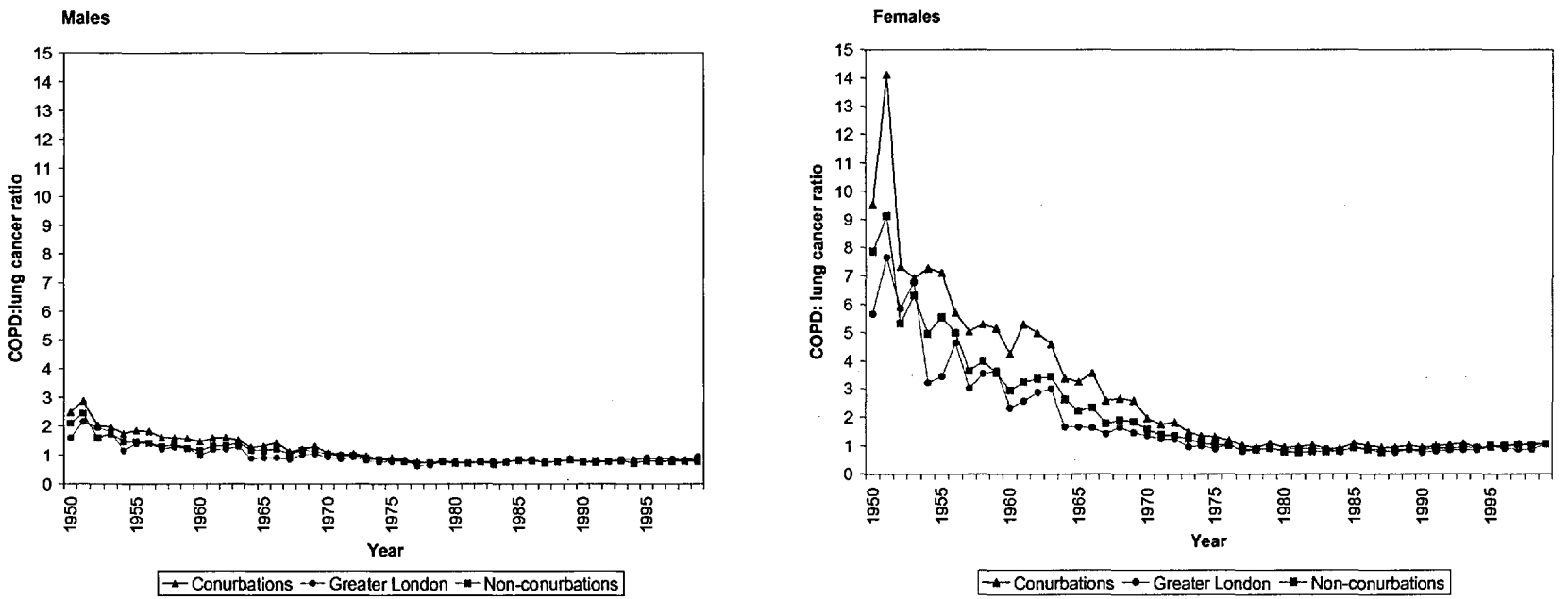
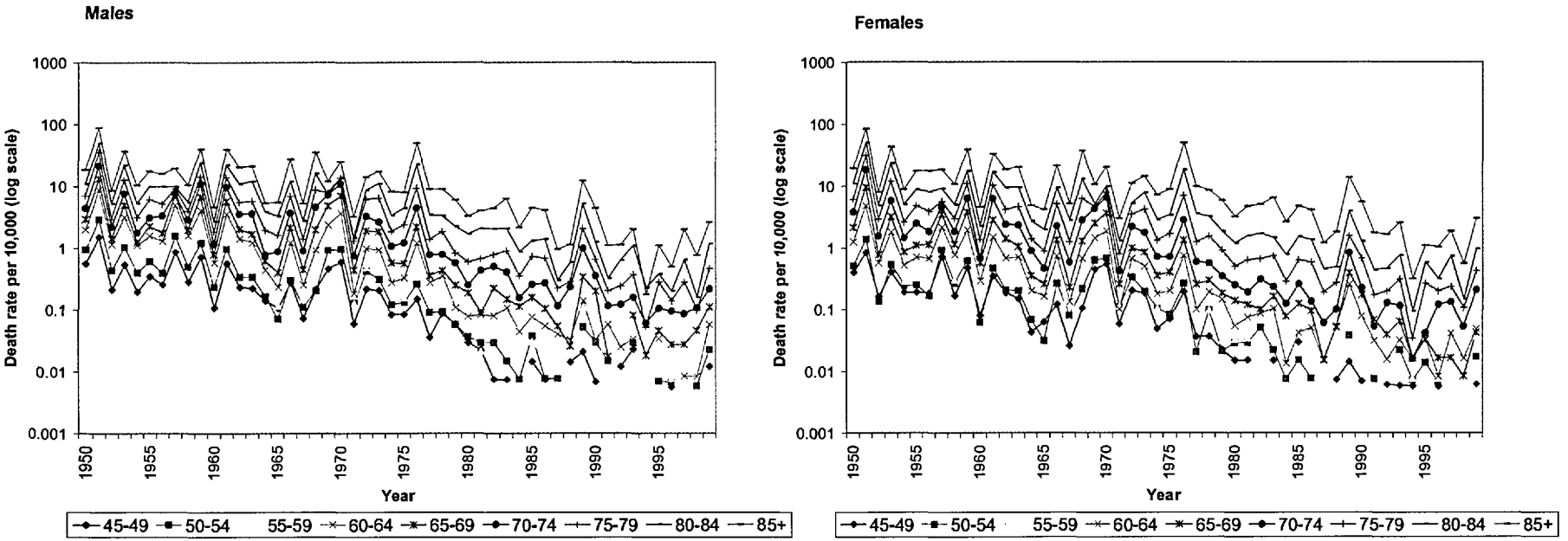


Figure 5-11 Mortality rates for influenza by age for ages 45+ years in England & Wales 1950-99
(NB gaps indicate zero deaths in that age-group in that year)



Bayesian age-period-cohort (BAMP) analyses by area

BAMP analyses were used to identify the timing of period and cohort effects using the RW2 smoothing constraint to identify specific years or cohorts, and further information obtained from graphs using the RW1 smoothing constraint. The y-axes on BAMP analysis graphs represent the difference from the internal average for the age, period or cohort parameter and the absolute values are therefore not comparable across different sex and area combinations. Due to the non-identifiability problem, as previously discussed, change points can be identified but general trends cannot. Change points can further be identified as maxima (a rise followed by a fall) and minima (a fall followed by a rise). Graphs of modelled rates by age-group and over-dispersion (not-shown), suggested a good fit to the data, but modelled rates for older age-groups (ages 55+ years) were much closer to observed rates than in younger age-groups with smaller numbers of deaths.

Statistical criteria based on credible intervals for second order differences for parameters excluding zero (Table 5-8, page 154) located many more period effects for COPD mortality than for lung cancer. Greater London had higher numbers of significant period parameters for COPD than other areas, which may have been partly related to smaller numbers. Few cohort effects were located unless using inclusive 25-75% credible intervals, where half the intervals located as significant may have arisen by chance. A large number of the significant cohort parameters were consecutive (with the same sign generally indicating maxima), suggesting they were part of the same cohort effect.

The numbers of parameters significant from RW2 analyses were fairly similar for period effects for both males and females in a particular area and for cohort effects for males or for females independent of area (Table 5-8, page 154), suggesting that period effects were more closely related to area, while cohort effects were more closely related to sex. Many more significant effects were located by RW1 than RW2 analyses (Table 5-8, page 154), presumably because RW2 smoothes trends more than RW1. More period effects were detected in London than in the other areas, which may be related to smaller numbers giving more rapid year to year fluctuations.

Period effects for COPD mortality in relation to hypotheses, artefacts and other influences

Period effects in relation to prior hypotheses

Timing of possible period effects in relation to prior hypotheses, data artefacts and other influences such as influenza were listed and evidence for period effects at similar times was sought. The main hypothesis was that COPD mortality might reflect trends in particulate and sulphur dioxide air pollution, with the greatest improvements occurring between the late 1950s and late 1970s. Consistent with this hypothesis, visual inspection of RW1 plots for COPD mortality for all areas and both sexes (Figure 5-12, pages 148-150) suggested the onset of a decline in period parameters starting around the late 1950s or early 1960s and tailing off in the late 1970s, with some superimposed year to year variations that were most marked in Greater London (with the smallest number of deaths). Statistical cut-off clearly located the tailing off of the decline to 1977-1978 in all areas and both sexes (Table 5-9, pages 155-156) with second order differences significant to 15-85% centiles (conurbations and non-conurbations) or 5%-95% centiles (Greater London). Since the air pollution analyses suggested that the decline in SO_2 may have levelled off later than the decline in black smoke, evidence was sought for a period effect in the early 1980s. However, there was only very weak evidence (significant to 25-75% centiles) for an effect in 1981 in three of six area-sex combinations (Table 5-9, pages 155-156).

It was more difficult to locate the exact timing of the onset of the decline than its levelling off, both visually and using statistical cut-offs. Statistical cut-offs using RW2 analyses located the onset of a decline in 1962-63 (Table 5-9, pages 155-156) but earlier maxima were located to 1956 and 1959 in Greater London (also air pollution episode years) and 1959 in non-conurbations (Table 5-9, pages 155-156) as well as 1951, which was also a year with high influenza activity (discussed below).

Period effects in relation to air pollution episodes

There was strong statistical support for change point maxima associated with air pollution episodes in Greater London in 1951 (also a year with high influenza mortality), 1956, 1957, 1959 for both sexes. Evidence was weaker for 1952 (males only, 25-75% centile) and 1962 (females only, 25-75% centile). No evidence was found

for Greater London air pollution episodes in 1975 and 1991 (with maximum air pollution levels an order of magnitude lower – Table 5.4, page 120).

Period effects in relation to possible data artefacts

ICD coding version changes occurred in 1958, 1968 and 1979. Any period effects in COPD mortality related to these might be expected in the years preceding the change, in all areas and both sexes. However, no consistent period effects were located (Table 5-9, pages 155-156) justifying not using bridge-coding factors to adjust the data. Also, no period effects were detected in relation to the Rule 3 interpretation changes in 1984 and 1992.

No areas had period effects located to 1974, when the major reorganisation of conurbation and non-conurbation boundaries took place.

Changes in published data leading to the exclusion of acute bronchitis codes in 1968 might have been expected to result in a small drop in the numerator, resulting in a maximum in 1967 or possibly minimum in 1968 (i.e. sudden change to a decline). 1968 was also a high influenza year, when maxima might be expected. Period effects were detected in Greater London in 1968, but these were maxima. Period effects in other areas were located to 1969, compatible with influenza effects on COPD mortality continuing into early 1969. This suggested that even if there was an effect of coding changes, this was outweighed by influenza effects.

Period effects in relation to influenza

Influenza mortality was particularly high in 1951 and there was evidence for a period effect in COPD mortality (Table 5-9, pages 155-156). Change points denoting maxima (rise followed by decline) in 1951 were located using statistical cut-offs in all areas except males in conurbations where, although not statistically significant, the effect was readily visible in plots of second order differences of period parameters (not shown). Statistical support for maxima was highest in Greater London pollution during that year. Other years with high influenza mortality in England & Wales were usually associated with change point maxima in Greater London, but rarely associated with them in conurbations or non-conurbations.

Period effects for lung cancer

Lung cancer period effects showed very different patterns from those for COPD. Effects were very gradual and most were identified by visual examination of RW1 plots (Table 5-10, page 156). No RW2 period effects showed higher statistical significance than 25% to 75% centiles. There were no period effects located to influenza years. Visual inspection of RW1 plots (Figure 5-13, pages 151-153) suggested a change from an increase to a decline in the early 1970s (Table 5-10, page 156) in males in all areas and some attenuation in the 1970s for females. All areas for females showed onset of a decline starting around 1988.

Cohort effects for COPD and lung cancer mortality

Main cohort effects were readily visible in RW1 plots (Figures 5-12 and 5-13, pages 148-153). Effects were different between men and women, but had similar timing for both lung cancer and COPD (Table 5-11, page 157). This was consistent with expectations that cohort effects would be similar for COPD and lung cancer reflecting lifetime smoking patterns.

The main cohort effect in men was a maximum occurring in those born around 1899-1903. This effect was diffuse and although readily apparent in RW1 plots (Figures 5-12 and 5-13, pages 148-153) and usually visible in plots of 2nd order differences from RW2 plots (not shown), did not show high levels of statistical significance (Table 5-11, page 157). Both lung cancer and COPD in males in non-conurbations showed an additional maximum for those born in cohorts centred on 1925-1926 (Table 5-11, page 157). In women, the main cohort effect was a maximum in those born in cohorts centred around 1925-29. This was readily visible in RW1 plots and statistically significant in all areas using 25-75% centiles in all areas and also in 15-85% centiles for COPD in Greater London and lung cancer in conurbations.

Original expectations also suggested there might be an additional COPD cohort effect in those born 1935-60, related to progressive reductions in air pollution exposure in childhood. Support for this from the data was weak and inconsistent. An additional later minimum or short-lived levelling off of the decline following by a second maximum was visible in some RW1 plots for females, COPD mortality in non-conurbations and all female lung cancer plots (Figures 5-12f on page 150, 5-13b on

page 151, 5-13d on page 152 and 5-13f on page 153). Statistical support for these second maxima using RW2 analyses was only detected for females in non-conurbations, timed to 1955-57, but was weak (significant only to 25-75% credible intervals). However, for other areas second maxima were visible although not statistically significant in plots of 2nd order differences for RW2 parameters (not shown) and were located to 1945 for female COPD in non-conurbations, 1949 for Greater London female lung cancer and 1958 for female lung cancer mortality in conurbations (Table 5-11, page 157).

Figure 5-12 BAMP COPD age -period-cohort RW1 analyses

Figure 5-12a COPD conurbations males

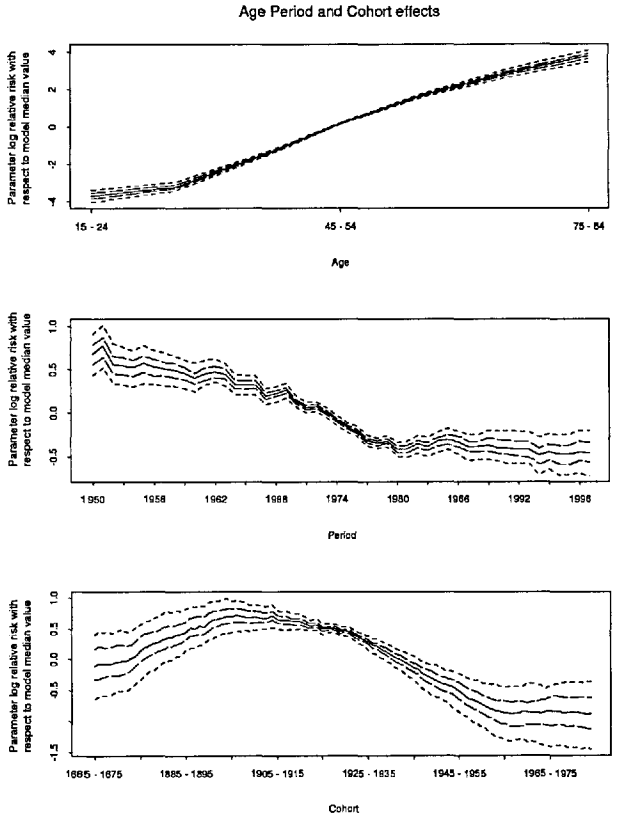


Figure 5-12b COPD conurbations females

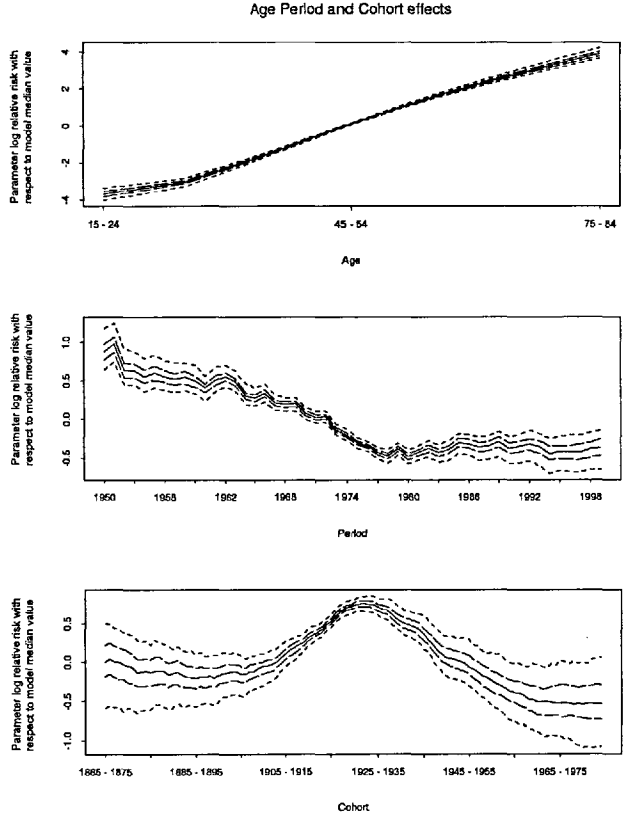


Figure 5-12c COPD Greater London males

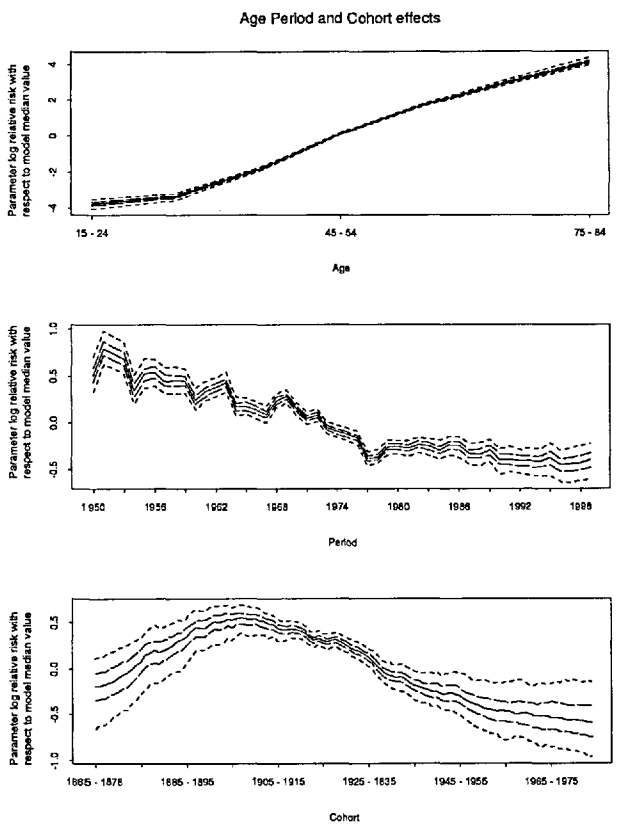


Figure 5-12d COPD Greater London females

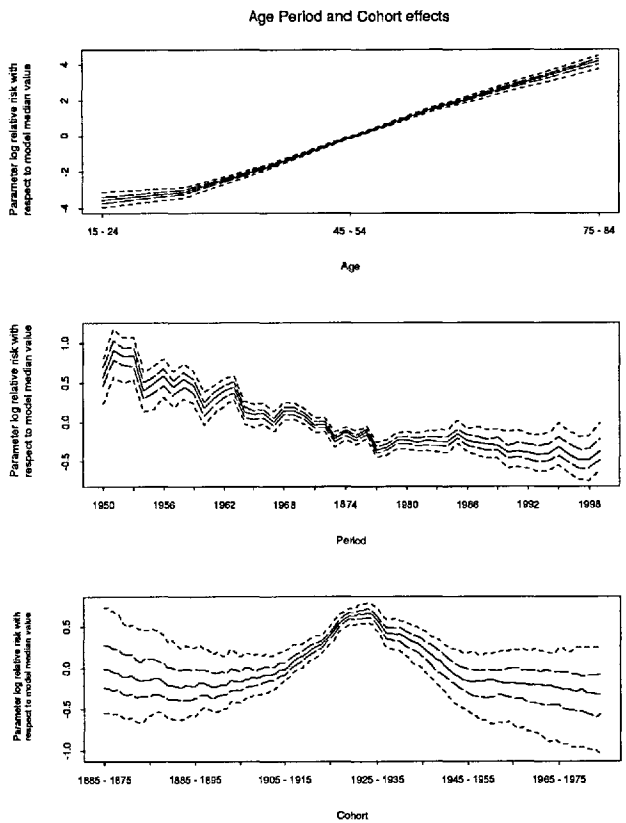


Figure 5-12e COPD non-conurbations male

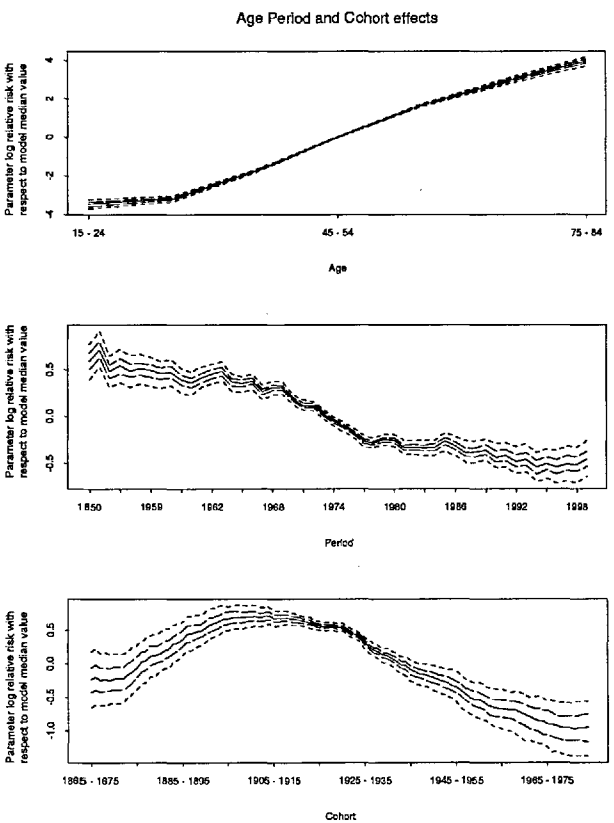


Figure 5-12f COPD non-conurbations female

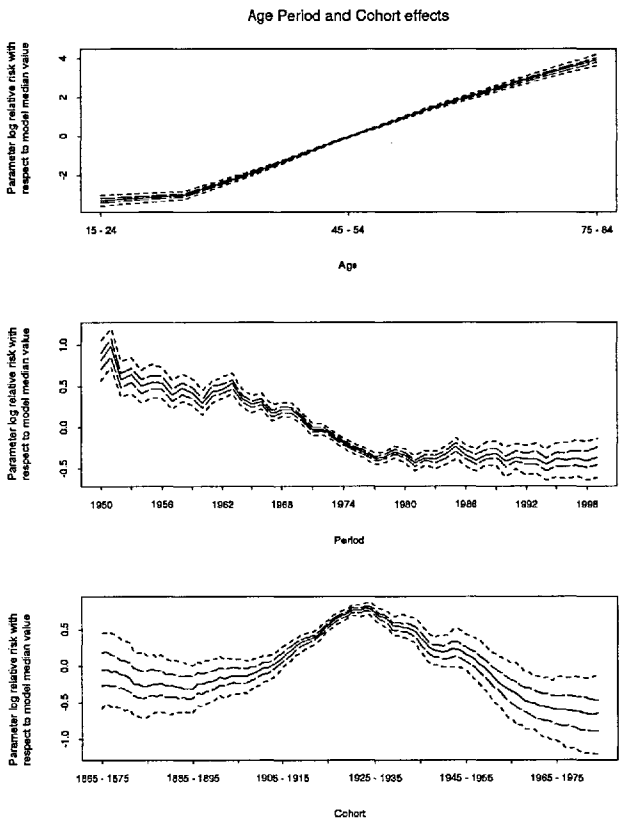


Figure 5-13 BAMP lung cancer age-period-cohort RW1 analyses

Figure 5-13a Lung cancer conurbations males

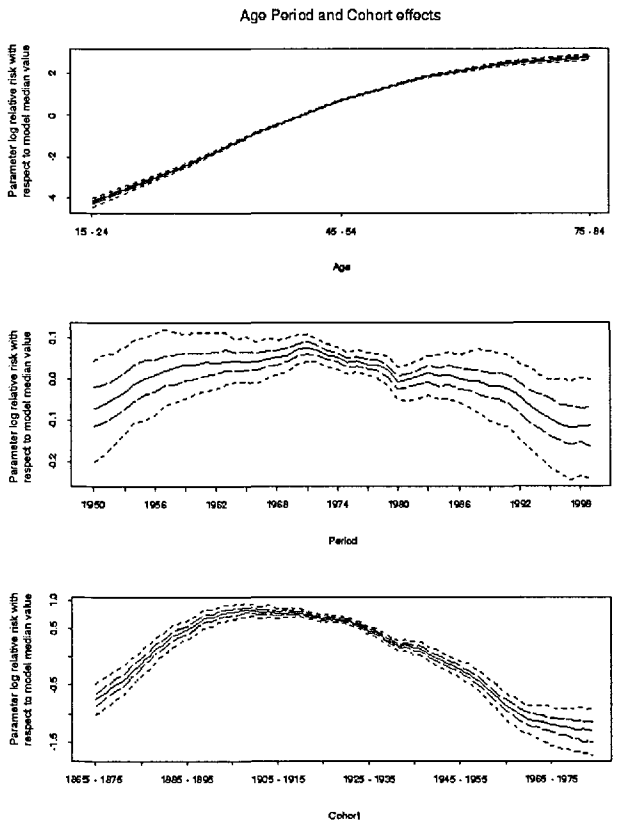


Figure 5-13b Lung cancer conurbations females

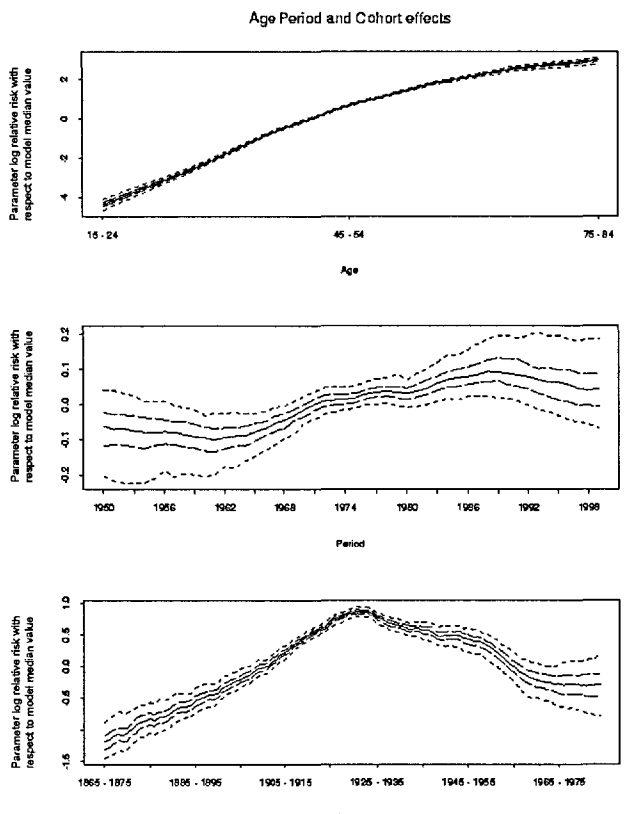


Figure 5-13c Lung cancer Greater London males

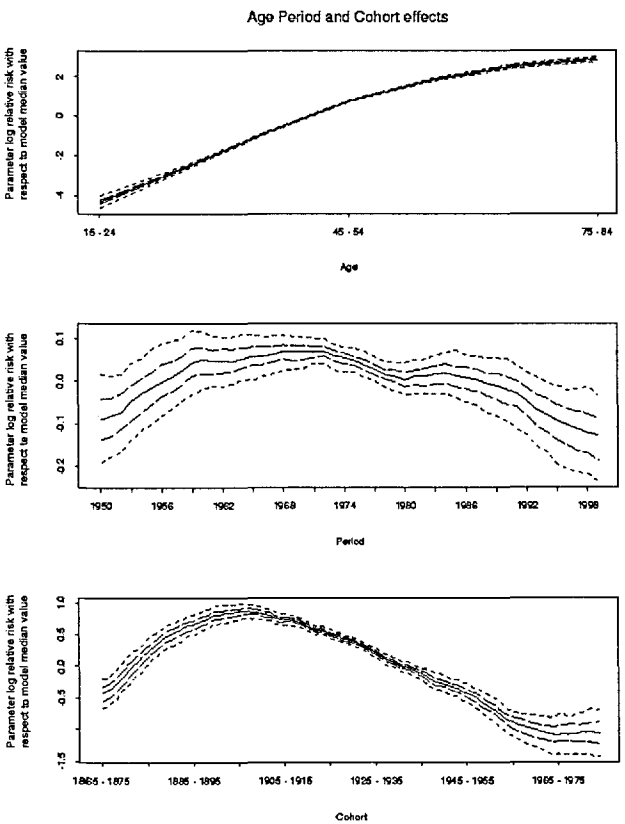


Figure 5-13d Lung cancer Greater London females

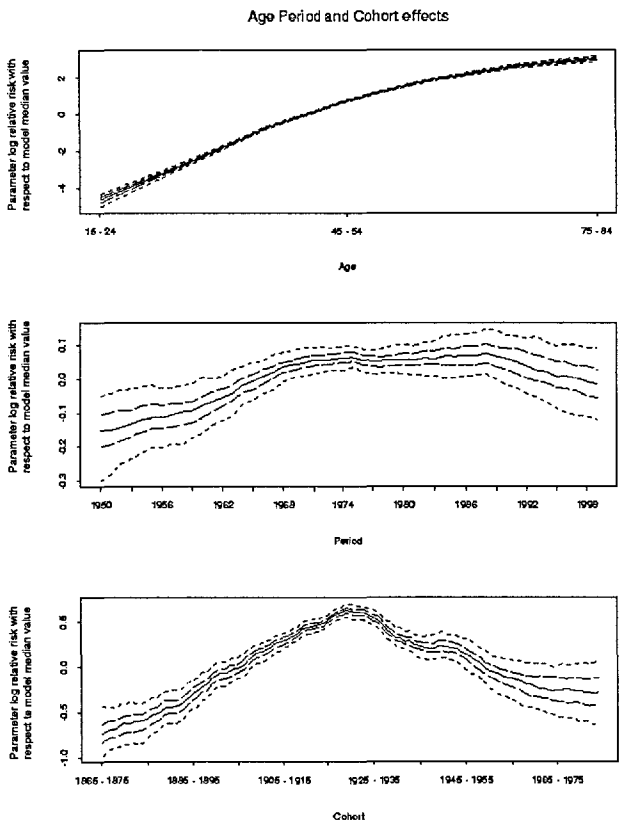


Figure 5-13e Lung cancer non-conurbations males

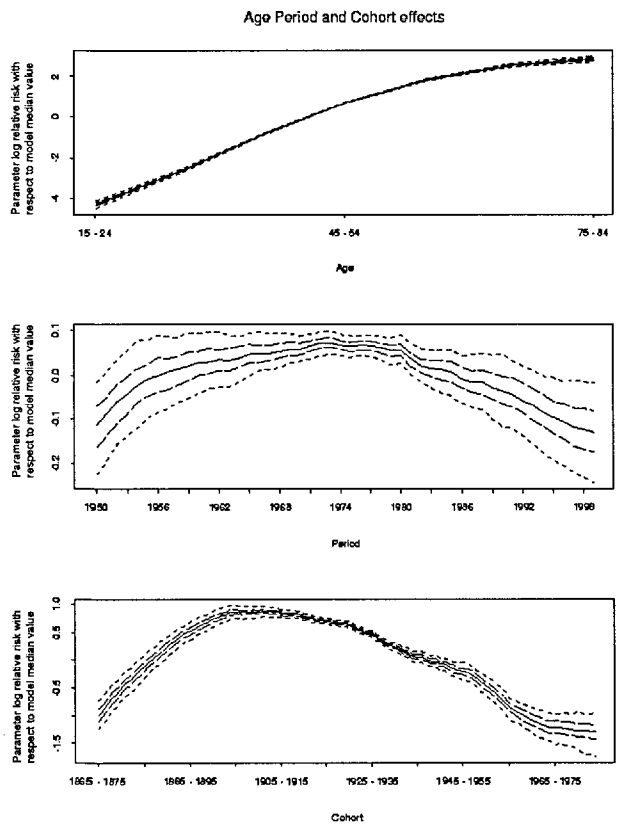


Figure 5-13f Lung cancer non-conurbations females

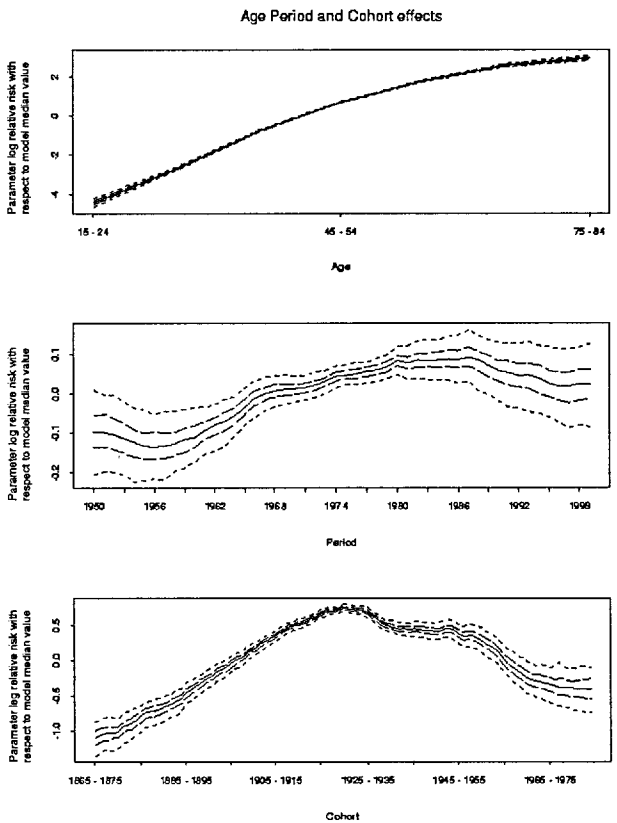


Table 5-8 BAMP analyses: number of second order differences for period parameters (n=48) and cohort parameters (n=108) where 5-95%, 10-90%, 15-85% or 25-75% credible intervals around parameters do not cross zero. RW2 analyses and RW1 analyses

(consec. = consecutive)

	Period					Cohort			
	5%-95% CI	10%-90% CI	15%-85% CI	25%-75% CI		5%-95% CI	10%-90% CI	15%-85% CI	25%-75% CI
COPD RW2									
Conurbations – male	0	0	2	8		0	0	0	0
Conurbations – female	0	0	3	10		0	0	0	6 (of which 5 consec.)
Greater London – male	14	19	21	33		0	0	0	0
Greater London – female	15	18	22	30		0	0	2	7 (all consec.)
Non-conurbations – male	0	1	2	13		0	0	0	2
Non-conurbations – female	1	2	7	13		0	0	0	6 (of which 5 consec.)
Lung cancer RW2									
Conurbations – male	0	0	0	1		0	0	0	2
Conurbations – female	0	0	0	0		0	0	1	6 (all consec.)
Greater London – male	0	0	0	0		0	0	0	4 (of which 2 consec.)
Greater London – female	0	0	0	0		0	0	0	3 (of which 2 consec.)
Non-conurbations – male	0	0	0	1		0	0	1	5 (of which 2 & 3 consec.)
Non-conurbations – female	0	0	0	2		0	0	0	7 (of which 4 consec.)
COPD RW1									
Conurbations – male	9			27		0			12 (three lots of 2 consec.)
Conurbations – female	8			28		0			3
Greater London – male	13			30		0			7 (of which 2 consec.)
Greater London – female	16			28		0			7 (of which 2 consec.)
Non-conurbations – male	11			34		0			11 (two lots of 2 consec.)
Non-conurbations – female	17			35		0			9 (three lots of 2 consec.)
Lung cancer RW1									
Conurbations – male	0			1		0			10 (three lots of 2 consec.)
Conurbations – female	0			0		1			10 (five lots of 2 consec.)
Greater London – male	0			0		0			3 (of which 2 consec.)
Greater London – female	0			1		0			4
Non-conurbations – male	0			1		0			13 (two of 3 consec. & 2 consec)
Non-conurbations – female	0			2		0			7 (of which 2 consec.)

Table 5-9 Expected timing of chronic air pollution effects, air pollution episodes, artefacts, and influenza peaks in England & Wales (flu) related to evidence for period effects for COPD mortality in those years using 2nd order differences from BAMP RW2 analyses.

Year	Comments	Conurbations – male	Conurbations – female	Greater London – male	Greater London – female	Non-conurbations – male	Non-conurbations – female
Hypothesis (Max or Min)							
1956-60 (Max)	Five years following 1956 clean air act	[1962-3 Max 25-75%]	[1962 Max 15-85%] [1963 Max 25-75%]	1956, 1959 Max 5-95% *1957 Min 5-95%* [1963 Max 5-95%]	1956, 1958-9 Max 5-95% *1957 Min 5-95%* [1962 Max 10-90%] [1963 Max 5-95%]	1959 Max 25-75% [1960 Max 25-75%] [1962 Max 25-75%] [1963 Max 15-85%]	*1957 Min 25-75%* [1960 Min 5-95%] [1963 Max 5-95%]
1975-80 (Min)	Black smoke decline levels off	1975-76 Min 25-75% 1977&78 Min 15-85%	1977 Min 25-75% 1978 Min 15-85% [1981 Min 25-75%]	*1976 Max 5-95%* 1977 Min 5-95%	*1976 Max 5-95%* 1977 Min 5-95%	1977 Min 15-85% 1978 Min 25-75% [1981 Min 25-75%]	1977 Min 15-85% *1979, Max 25-75%* [1981 Min 25-75%]
1980-85 (Min)	SO_2 decline levels off	1981 Min 25-75%	-	1981 Min 25-75% *1982 Max 25-75%* 1984 Min 25-75% *1985 Max 25-75%*	*1985 Max 10-90%*	1981 Min 25-75% *1985 Max 25-75%*	*1985 Max 15-85%*
Air pollution episodes, peak influenza years							
1951 (Max)	Flu, GL air pollution episode	-	Max 25-75%	Max 5-95%	Max 5-95%	Max 25-75%	Max 25-75%
1952 (Max)	GL air pollution episode	NA	NA	Max 15-85%	-	NA	NA
1953 (Max)	Flu	Min 25-75%	Min 25-75%	Max 5-95%	Max 5-95%	-	-
1956 (Max)	GL air pollution episode	NA	NA	Max 5-95%	Max 5-95%	NA	NA
1957 (Max)	Flu, GL air pollution episode	-	-	*Min 5-95%*	*Min 5-95%*	-	*Min 25-75%*
1959 (Max)	GL air pollution episode	NA	NA	Max 5-95%	Max 5-95%	NA	NA
1961 (Max)	Flu	-	-	-	-	-	-
1962 (Max)	GL air pollution episode	NA	NA	-	Max 10-90%	NA	NA
1966 (Max)	Flu	-	-	Max 25-75%	Max 15-85%	-	Max 25-75%
1968 (Max)	Flu	[1969 Max 25-75%]	-	Max 10-90% [1969 Max 5-95%]	Max 15-85% [1969 Max 25-75%]	[1969 Max 25-75%]	[1969 Max 25-75%]
1970 (Max)	Flu	-	-	*Min 25-75%*	-	-	-
1972 (Max)	Flu	-	-	Max 5-95%	Max 5-95%	-	-
1973 (Max)	Flu	-	-	*Min 25-75%*	*Min 5-95%*	-	-
1975 (Max)	GL air pollution episode	NA	NA	-	-	NA	NA
1976 (Max)	Flu	-	Min 25-75%	Max 5-95%	Max 5-95%	-	-
1989 (Max)	Flu	-	-	Max 10-90%	Max 25-75%	-	-
1991 (Max)	GL air pollution episode	NA	NA	[1990 Min 25-75%]	[1990 Min 15-85%]	NA	NA

Max 5-95% indicates that 5th-95th credible intervals (CIs) for 2nd order differences for period parameters were all negative (indicating a maximum: change from rise to a fall or levelling off). Min 5-95% indicates that 5th –95th CIs for 2nd order differences were all positive (a minimum: change from a fall to a rise or levelling off). 10-90% relates to 10th-90th CIs, 15-85% to 15th-85th CIs, 25-75% to 25th-75th CIs . See pages 103-4 and Figure 4.2, page 104 for further explanation of identification of change points.

NA = not applicable; – = not detected

Italics used to identified different sign of period effect from that predicted; [brackets] used to identify period effects in years close to but not identical to predicted

Air pollution episodes identified from Holland 1979[297] and Anderson 1995[298]

Table 5-9 (continued) Expected timing of chronic air pollution effects, air pollution episodes, artefacts, and influenza peaks in England & Wales (flu) related to evidence for period effects for COPD mortality in those years using `2nd order differences from BAMP RW2 analyses

Year	Comments	Conurbations – male	Conurbations – female	Greater London – male	Greater London – female	Non-conurbations – male	Non-conurbations – female
Artefact							
1957 (Either)	Last year using ICD6	-	-	Min 5-95%	Min 5-95%	-	Min 25-75%
1967 (Either)	Last year using ICD7	-	-	Min 5-95%	Min 5-95%	-	-
1968 (Min)	Acute bronchitis codes removed	-	-	*Max 10-90%*	*Max 15-85%*	-	-
1974 (Either)	Changes to area boundaries	-	-	-	-	-	-
1978 (Either)	Last year using ICD8	Min 15-85%	Min 15-85%	-	-	Min 25-75%	-
1983 (Min)	Year before Rule 3 coding change	-	-	-	-	-	-
1992 (Max)	Year after Rule 3 coding change	-	-	-	-	-	-

Max 5-95% indicates that 5th-95th credible intervals (CIs) for 2nd order differences for period parameters were all negative (indicating a maximum: change from rise to a fall or levelling off). Min 5-95% indicates that 5th–95th CIs for 2nd order differences were all positive (a minimum: change from a fall to a rise or levelling off). 10-90% relates to 10th-90th CIs, 15-85% to 15th-85th CIs, 25-75% to 25th-75th CIs . See pages 103-4 and Figure 4.2, page 104 for further explanation of identification of change points.
NA = not applicable; – = not detected
Italics used to identified different sign of period effect from that predicted; [brackets] used to identify period effects in years close to but not identical to predicted

Table 5-10 Timing of (all) period effects for lung cancer in conurbations (C), Greater London (GL) and non-conurbations (non-C) in males (m) and females(f) using BAMP analyses

Analysis	Conurbations – male	Conurbations – female	Greater London – male	Greater London – female	Non-conurbations – male	Non-conurbations – female
RW2						
	1954 Max 25-75%		-	-	1954 Max 25-75%	1955 & 1956 Min 25-75%
RW1 identified using statistical criteria or *in italics if could only be identified visually from RW1 graphs*						
	1971 Max (rise to decline) 1980 Min 25-75% *1985 Continuation of decline*	*1971 Levelling off of rise* *1978 Level starts rising* *1988 Levelling off of rise*	*1960 Levelling off of rise* *1972 Onset of decline* *1980 Levelling off of decline* *1984 Continuation of decline*	*1969 Levelling off of rise* 1988 Max 25-75% *1989 Onset of decline*	*1973 Max (rise to shallow decline)* 1980 Max 25-75%	*1952 Max (onset of decline)* *1968 Some attenuation of rise* *1978 Levelling off of rise* 1980 Max 25-75% 1987 Max 25-75% *1988 Onset of decline*

Max 25-75% indicates that 25th–75th credible intervals (CIs) for second order difference for period parameters were all negative (indicating a maximum: change from rise to a fall or levelling off). Min 25-75% indicates that 25th–75th CIs for second order difference for period parameters were all positive (indicating a minimum: change from a fall to a rise or levelling off). See pages 103-4 and Figure 4.2, page 104 for further explanation of identification of change points.

Table 5-11 Timing of (all) cohort effects for COPD and Lung Cancer using 2nd order differences from BAMP analyses

Italic font is used if the cohort effect could only be identified visually from RW1 graphs

	Cohort centre	Direction	Credible intervals (CI)	Comments
COPD				
Conurbations – male	-	-		Visual inspection of RW1 plots and of plots of RW2 posterior 25th,median and 75th centiles for 2nd order differences (copdcm2psi2b.gph) located maximum to the 1894-1904 & 1895-1905 cohorts (i.e. centred around 1899-1900)
Greater London – male	-	-	-	Visual inspection of RW1 plots suggested a maximum for the cohort centred around 1900. However, this was not seen on RW2 plots of posterior estimates for 2nd order differences
Non-conurbations – male	1878 1926	Minimum Maximum	25th-75th CI 25th-75th CI	Visual inspection of RW1 plots suggested a very gradual cohort effect with a maximum (end of a rise) for the cohort centred around 1900 (RW2 plots of posterior 25th,median and 75th centiles for 2nd order differences located this to the cohort 1896-1906, i.e. centred around 1901) followed by the 1926 maximum forming the start of a decline.
Conurbations – female	1926	Maximum	25th-75th CI	Cohorts 1918-1928 to 1924-1934 inclusive were significant for 25th-75th CIs
Greater London – female	1925 1929	Maximum Maximum	15th-85th CI, 15th-85th CI	Cohorts 1920-1930 to 1926-1936 inclusive were significant for 25th-75th CIs
Non-conurbations – female	1908 1928	Minimum Maximum	25th-75th CI 25th-75th CI	Cohorts 1921-1931 to 1925-1935 were significant for 25th-75th CIs Additional levelling off (minimum) visible in RW1 plot for cohort centred on 1945, also seen in RW2 plots of posterior 25th,median and 75th centiles for 2nd order differences
Lung cancer				
Conurbations – male	1888 1936	Maximum Minimum	25th-75th CI 25th-75th CI	Visual examination of RW1 plots suggests a gradual trend, peaking in cohorts born around 1900, with a possible location to cohort born 1898-1908 (i.e. centred around 1903) seen in RW2 plots of posterior 25th,median and 75th centiles for 2nd order differences.
Greater London – male	1885 1902-3 1905	Maximum Maximum Maximum	25th-75th CI 25th-75th CI 25th-75th CI	Cohorts 1897-1907 & 1898-1908 then 1900-1910 were significant for 25th-75th CIs
Non-conurbations – male	1900-1 1925	Maximum Maximum	25th-75th CI 15th-85th CI	Cohorts 1895-1905 & 1896-1906 were significant for 25th-75th CIs Cohorts 1920-1930 to 1922-1932 inclusive were significant for 25th-75th CIs
Conurbations – female	1929 *1958*	Maximum *Maximum*	15th-85th CI	Cohorts 1920-30 to 1925-35 inclusive were significant for 25th-75th CIs *RW1 plots suggested a later maximum, located to cohort born 1953-1963 using plots of 25-75% CIs for 2nd order differences for RW2*
Greater London – female	1925 1928-9 *1938* *1949*	Maximum Maximum *Minimum* *Maximum*	25th-75th CI 25th-75th CI	Cohorts 1920-1930, 1923-1933 and 1924-1934 were significant for 25th-75th CIs *RW1 plots suggested a levelling off (minimum), located to cohort born 1933-1943 using plots of 25-75% CIs for 2nd order differences for RW2* *RW1 plots suggested a later maximum located to cohort born 1944-1954 using plots of 25-75% CIs for 2nd order differences for RW2*
Non-conurbations – female	1925 1928/9 1955 1957	Maximum Minimum Maximum Maximum	25th-75th CI 25th-75th CI 25th-75th CI 25th-75th CI	Cohorts 1920-1930 then 1922-1932 to 1925-1935 inclusive were significant for 25th-75th CIs Cohorts 1950-1960 & 1952-1962 were significant for 25th-75th CIs

Maximum indicates a change from rise to a fall or levelling off, minimum indicates a change from a fall to a rise or levelling off. See pages 103-4 and Figure 4.2, page 104 for further explanation of identification of change points.

Classical age-period-cohort analyses

Classical analyses were used to provide some quantification of the changes in period and cohort effects, but as outlined in the methods, this was only possible by making comparisons of period effects in one area compared with another, since the absolute values of these effects are not identifiable.

Model fit

The results for the GLM Poisson regression assuming similar parameter effects for age, period and/or cohort in different areas (conurbations, Greater London and non-conurbations) can be found in Table 5-12 (page 159). Uppercase letters for age (A), period (P) or cohort (C) denote a model where the effect is assumed to be the same in all areas, while lowercase letters (a, p or c) denote a model where the effect is allowed to differ for each of the three areas. Using AIC and BIC more complex models (allowing two or all of age, period and cohort to vary by area) generally provided the best fit. The most dramatic improvement in fit as assessed by AIC and BIC was seen in the step from assuming age, period and cohort effects were similar in all areas to that where one parameter was allowed to vary by area.

When one parameter was allowed to vary, the best fit (Table 5-12, page 159) was achieved with different period effects in different areas for COPD in males and females and with different cohort effects in different areas for lung cancer in males and females (as assessed by AIC and BIC, except COPD females AIC only). When two parameters were allowed to vary by area, the best model fit was with different period and cohort effects but similar age effects in each area for COPD in males (supported by AIC), different age and period effects but similar cohort effects in different areas for COPD in females (AIC and BIC) and different age and cohort effects but similar period effects in different areas for lung cancer in males and females (AIC).

Table 5-12 Poisson regression model results from maximum likelihood analyses

Uppercase letters in model column denote shared age (A), period (P) or cohort (C) effects
Lowercase letters in model column denote age (a) period (p) or cohort (c) effects allowed to differ in different areas

Best model using AIC when varying one or two parameters is indicated in bold in Model column.
Best model using AIC/BIC for each disease/sex combination is indicated in bold in AIC/BIC columns.

Model	Residual deviance	Residual degrees of freedom	Overdispersion factor	AIC	BIC
COPD in males					
APC	28146	895	31.45	35156	35924
APc	4100	673	6.09	11518	13291
ApC	3921	795	4.93	11132	12396
aPC	4631	881	5.26	11669	12507
Apc	1239	593	2.09	8818	10986
aPc	1947	661	2.95	9389	11222
apC	1720	783	2.20	8954	**10278**
apc	1118	583	1.92	**8716**	10934
COPD in females					
APC	19512	891	21.90	25795	26552
APc	2351	669	3.51	9028	10752
ApC	2464	791	3.12	8947	10200
aPC	2697	877	3.07	9007	9834
Apc	985	589	1.67	7822	9940
aPc	1479	657	2.25	8179	9963
apC	1195	779	1.53	**7702**	**9014**
apc	888	579	1.53	7745	9912
Lung cancer in males					
APC	22662	891	25.43	30210	30967
APc	1201	675	1.78	9156	10920
ApC	3385	791	4.28	11133	12385
aPC	4209	877	4.80	11784	12611
Apc	844	595	1.42	8960	11118
aPc	927	663	1.40	**8906**	10729
apC	1338	779	1.72	9110	**10422**
apc	813	585	1.39	8948	11156
Lung cancer in females					
APC	9328	887	10.52	15775	16521
APc	881	672	1.31	**7721**	9436
ApC	1912	787	2.43	8558	9799
aPC	2655	873	3.04	9129	9945
Apc	747	592	1.26	7747	9856
aPc	860	660	1.30	7724	9498
apC	1166	775	1.50	7836	**9136**
apc	731	582	1.26	7751	9909

Source of data: Tables 6.1 to 6.3 in Breuninger, 2003[288]

Quantifying relative risks of period parameters relative to non-conurbations

In order to be able to quantify the period or cohort effects relative to non-conurbations, the best-fitting models allowing one parameter to vary was examined – results from models allowing more than one parameter to vary were not readily interpretable in terms of quantification. This was not the best fitting model overall except for female lung cancer, so the reductions in relative risk should be viewed as illustrative rather than definitive. Confidence intervals around the relative risks are shown in Figures 5-14 and 5-15 (both on page 162). Relative risks for period effects were almost always greater than 1.0 indicating statistically significant higher levels for period effects in conurbations and Greater London relative to non-conurbations. Relative risks for cohort effects were very tight round central values. To aid visual clarity, confidence intervals for relative risks are not shown in further graphs.

Variation of period effects in COPD mortality

Graphs of the relative risk of COPD period effects for urban areas compared with non-conurbations (the exponential of the difference in parameters) using models assuming similar age and cohort, but different period effects in different areas (ApC models) are shown in Figure 5-16 (page 163). These have not been smoothed, as year to year variations are of interest here e.g. relating to the 1952 smog in London. Patterns for males and females were broadly similar to each other, but there were more marked year to year fluctuations in females who experienced lower numbers of deaths. The graphs for RR of period effects were very similar to the relative differences between age-standardised rates in each area, presented for comparison in Figure 5-17 (page 164). This suggests that the factors causing the main differences between the more urban areas and the non-conurbations acted almost exclusively as period effects.

Linear regression of the relative risks against year to find average changes in the period effect relative risks (Table 5-13, page 166) suggested for COPD in conurbations an average absolute reduction in RR for period effects relative to non-conurbations of 38% in males (from an RR of 1.67 to 1.28) and 46% in females from 1957-1978 (being respectively the year following the Clean Air Act and the year when the decline levelled off according to BAMP analyses presented above). From 1979-1999 there was no significant change in relative risk in conurbations. For Greater London, the relative risk

for the period effect COPD in males in Greater London relative to non-conurbations showed a absolute average decline of 22% from 1957-1978 and no significant change from 1979 onwards. Although the RR of period effect in females also showed an overall decline of 21% over 1950-1999, most of this occurred 1979-1999 with a 38% reduction. The highest value for both RRs for period effects in Greater London relative to non-conurbations was seen in 1952 (1.9 in males and 1.8 in females, easily identifiable in Figure 5-17, page 164), which is almost certainly related to the London smog in that year.

Variation of cohort effects in lung cancer mortality

Graphs of the relative risk of lung cancer cohort effects for urban areas compared with non-conurbations using models assuming similar age and period, but different cohort effects in different areas (APc models) are shown in Figure 5-18, page 165. These have been smoothed using an arbitrary five cohort moving average. To calculate averages, the last 20 cohorts (born 1955-65 to 1974-84) and first five cohorts (born 1865-75 to 1869-79) were dropped from the linear regression to prevent unstable or missing estimates related to small numbers of deaths. Relative risks for cohort effects compared to non-conurbations differed between conurbations and Greater London. For cohorts born 1870-80 to 1869-79, relative risks had risen with successive cohorts in conurbations by around 30% on average (Table 5-14, page 166), but had fallen in Greater London from being twice the risk of that in non-conurbations for the oldest cohorts to very similar relative risks to those in non-conurbations for youngest cohorts (Figure 5-18, page 165).

Figure 5-14 Relative risk (95% confidence intervals) for period effects for COPD in conurbations and Greater London relative to non-conurbations, using a classical age-period-cohort quasi-likelihood model with interaction terms assuming age and cohort effects are shared across areas (males top two graphs, females below).

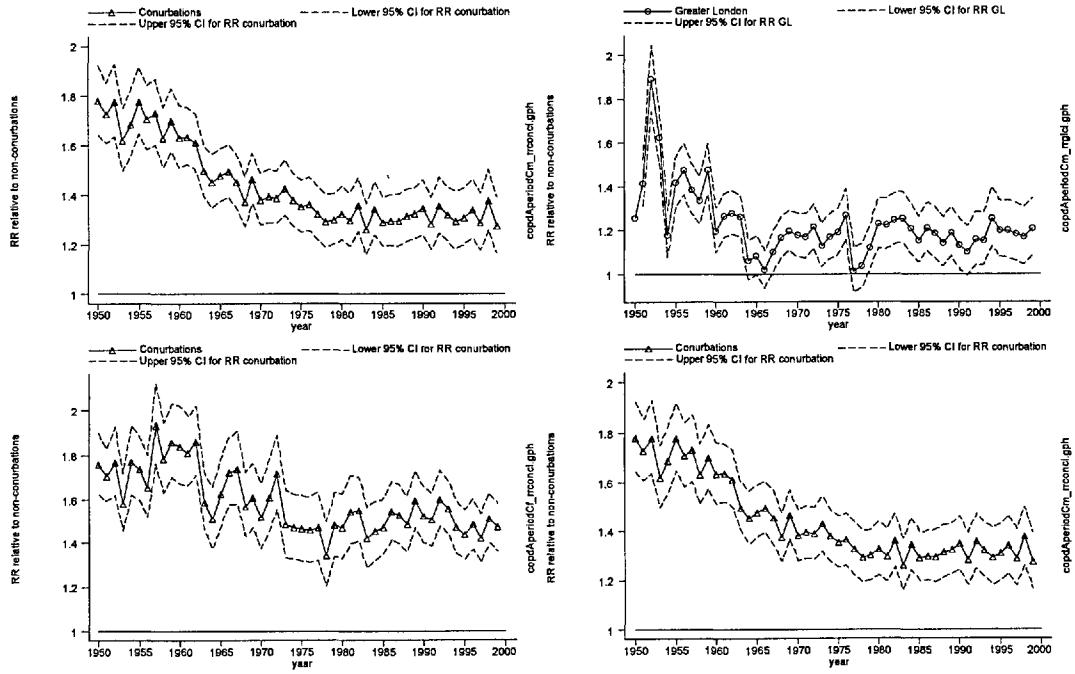


Figure 5-15 Relative risk (95% confidence intervals) for cohort effects for COPD relative to non-conurbations, using a classical age-period-cohort quasi-likelihood model with interaction terms assuming age and period effects are shared across areas (males top two graphs, females below, first 5 and last 20 cohorts omitted)

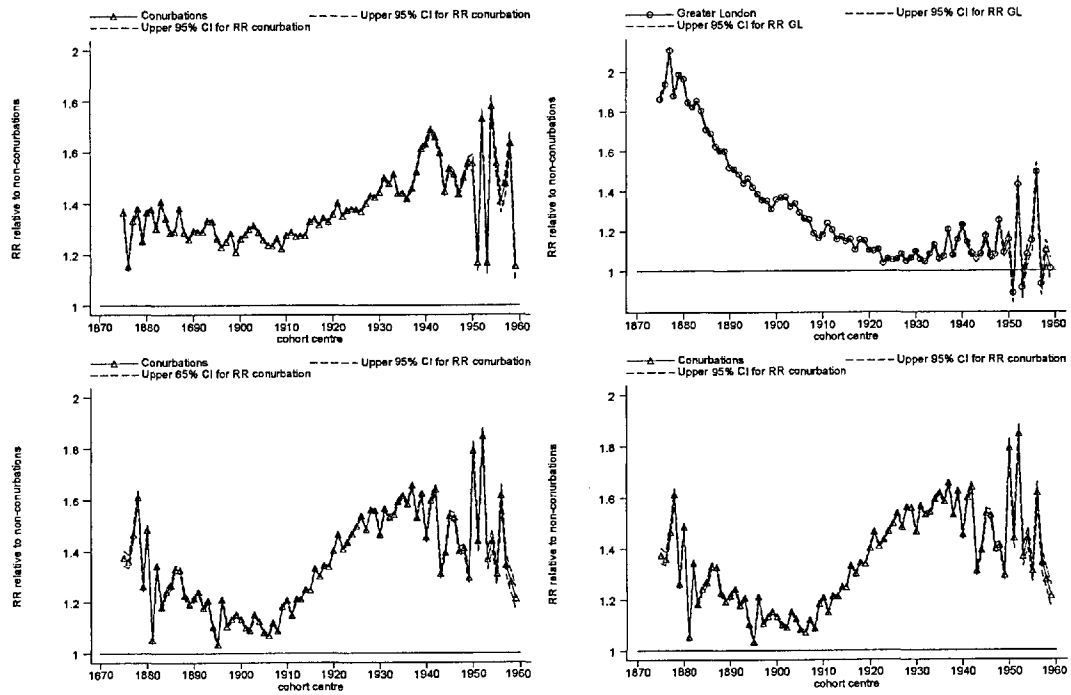
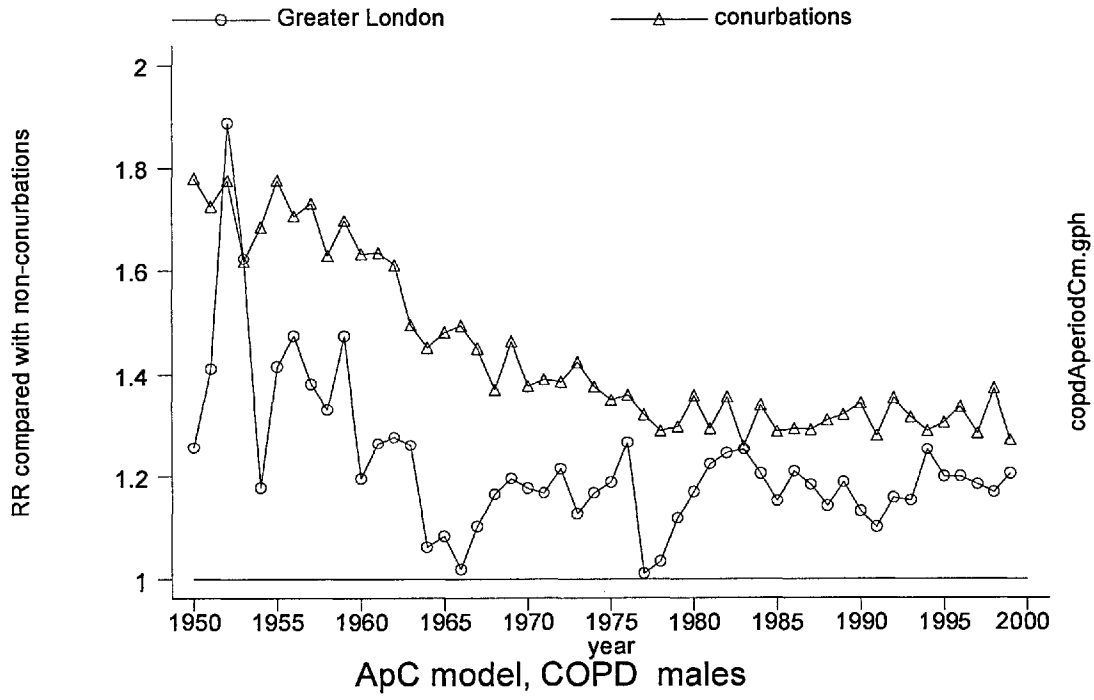


Figure 5-16 Relative risk for period effect for COPD for conurbations and Greater London compared with non-conurbations, using a classical age-period-cohort model with interaction terms assuming age and cohort effects are shared across areas. Upper graph males, lower graph females.



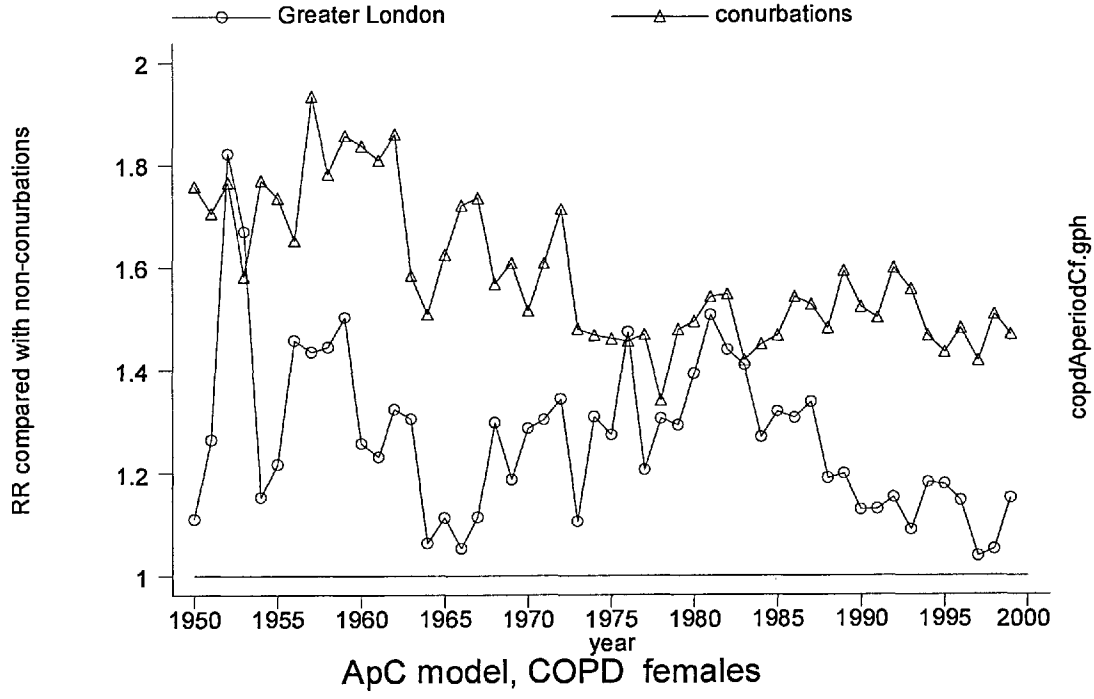
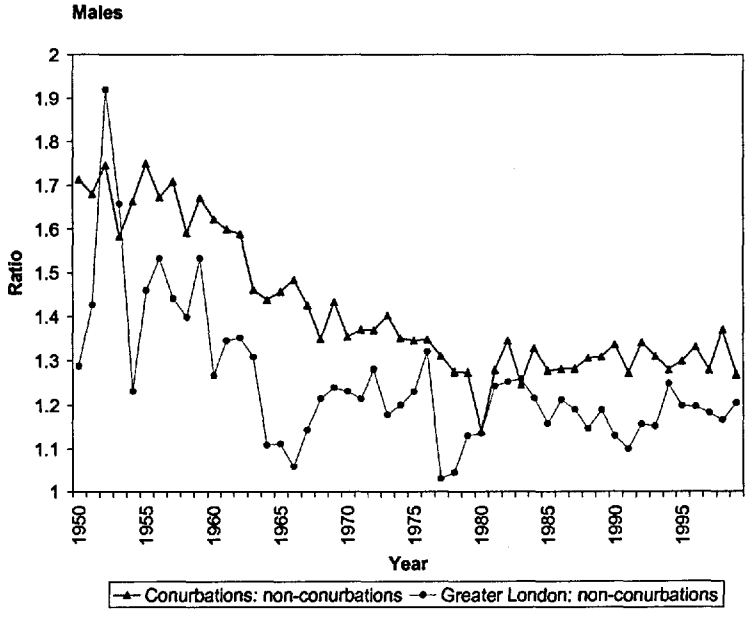


Figure 5-17 Relative differences between age-standardised rates for COPD mortality for conurbations (excluding Greater London) and Greater London compared with non-conurbations, 1950-1999. Upper graph males, lower graph females.



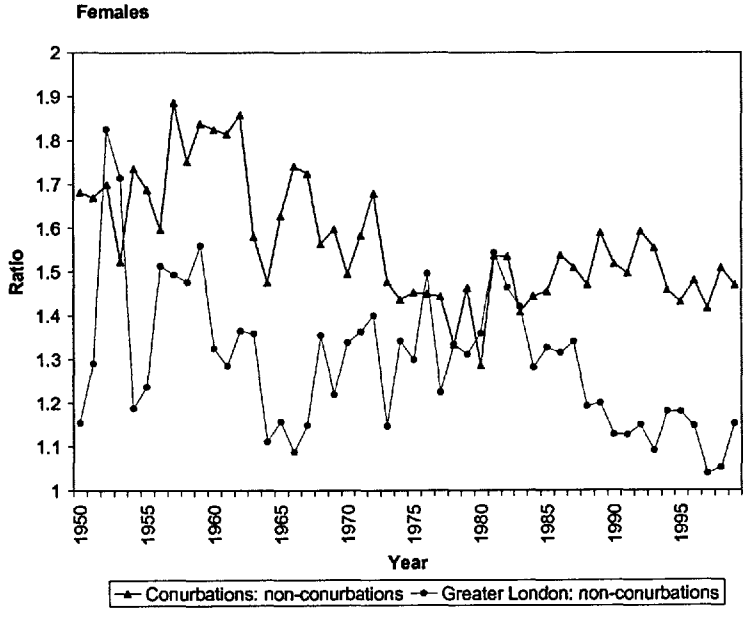
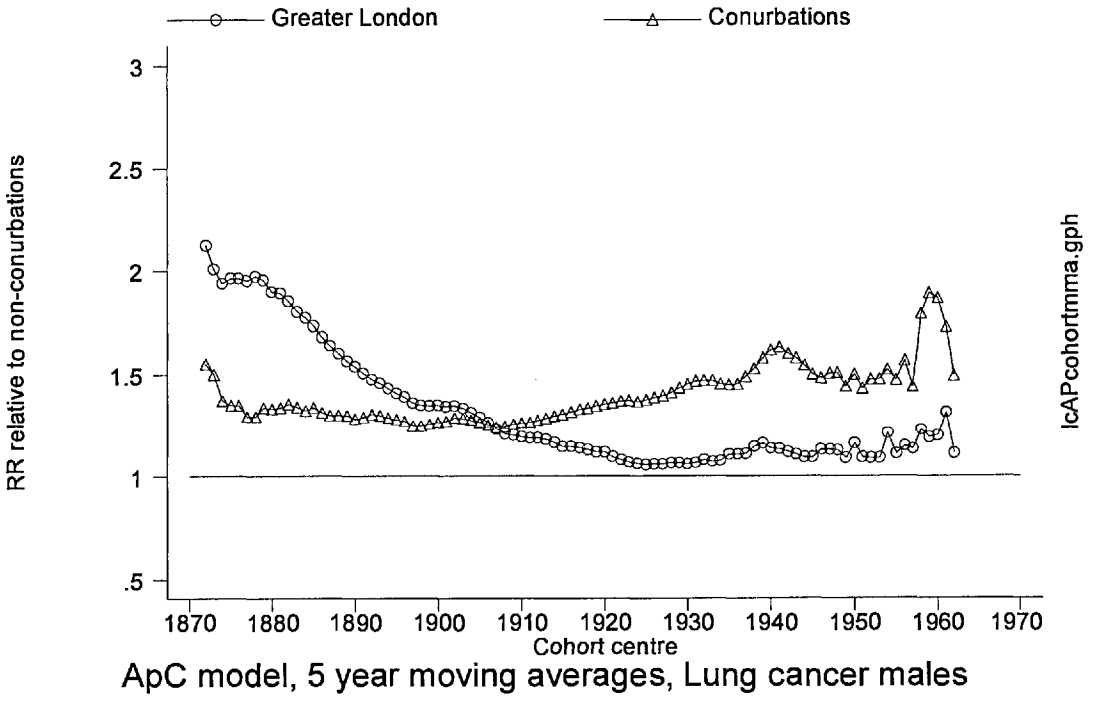
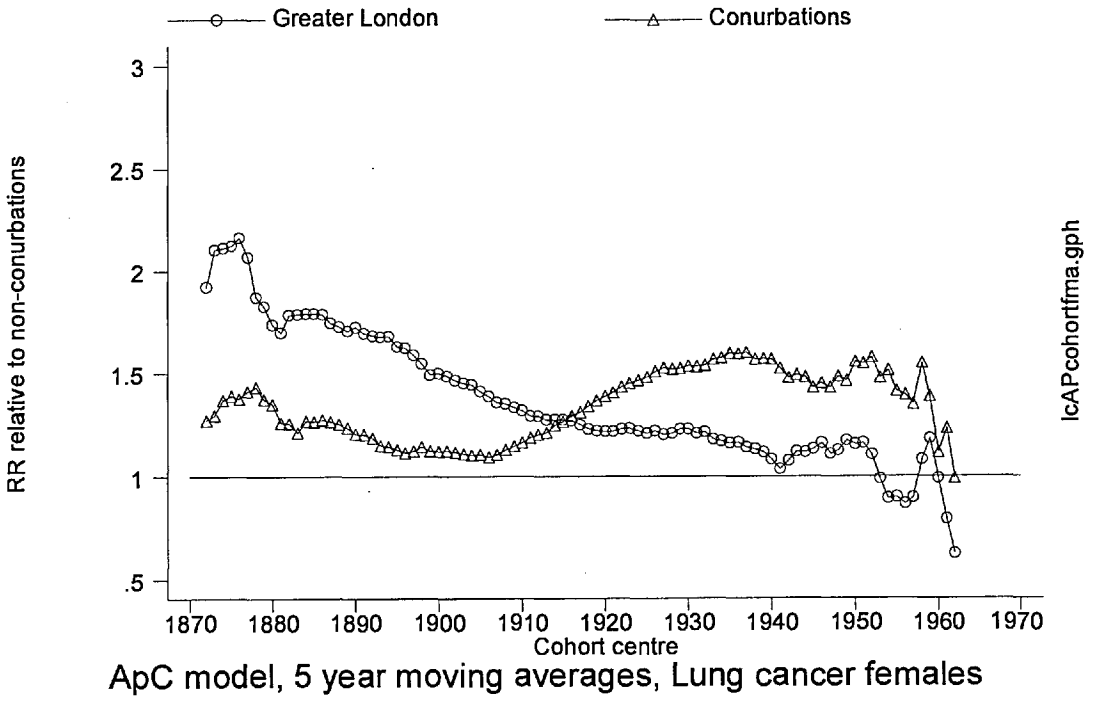


Figure 5-18 Relative risk for cohort effects for lung cancer for conurbations vs. non-conurbations and Greater London vs. non-conurbations, using a classical age-period-cohort model with interaction terms assuming age and period effects are shared across areas. Graphs have been smoothed using a five cohort moving average. Upper graph males, lower graph females.



ApC model, 5 year moving averages, Lung cancer males



ApC model, 5 year moving averages, Lung cancer females

Table 5-13 Absolute changes in average relative risk of COPD period effects for conurbations vs. non-conurbations and Greater London vs. non-conurbations from classical age-period-cohort models assuming shared age and cohort but varying period effects in the different areas (ApC model), assuming a linear trend

	Year	Predicted absolute % change in RR vs. non-conurbations	p-value for coefficient in linear regression of RR
COPD in males			
Conurbations	1950-1999	–48.4%	p<0.001
Greater London	1950-1999	–24.7%	p<0.001
Conurbations	1957-1978	–38.4%	p<0.001
Greater London	1957-1978	–22.1%	p=0.005
Conurbations	1979 onwards	+0.8%	p=0.715
Greater London	1979 onwards	–1.9%	p=0.559
COPD in females			
Conurbations	1950-1999	–34.2%	p<0.001
Greater London	1950-1999	–21.0%	p=0.006
Conurbations	1957-1978	–45.8%	p<0.001
Greater London	1957-1978	–6.4%	p=0.496
Conurbations	1979 onwards	–1.3%	p=0.732
Greater London	1979 onwards	–37.6%	p<0.001

% figures relate to absolute changes in predicted levels obtained from linear regression.

Table 5-14 Absolute changes in average relative risk of lung cancer cohort effects for conurbations vs. non-conurbations and Greater London vs. non-conurbations from classical age-period-cohort models assuming shared age and period but varying cohort effects (APc model)

	Absolute % change in predicted RR vs. non-conurbations averaged over cohorts born 1870-80 to 1954-64	p-value for coefficient in linear regression of RRs
Lung cancer in males		
Conurbations	+27.6%	p<0.001
Greater London	–78.5%	p<0.001
Lung cancer in females		
Conurbations	+34.1%	p<0.001
Greater London	–98.8%	p<0.001

The last 20 cohorts (born 1955-65 to 1974-84) and first five cohorts (born 1865-75 to 1869-79) were dropped to prevent unstable or missing estimates related to small numbers of deaths.
% figures relate to absolute changes in predicted levels obtained from linear regression.

Comparison of air pollution trends with mortality trends

Time trends in mortality and air pollution

Longer term time trends in COPD and lung cancer mortality were similar to those occurring at the same time for black smoke concentrations, but less so for sulphur dioxide. Age-standardised COPD and lung cancer mortality rates for both males and females showed a clear downward trend between 1950-1999, with rates in conurbations and Greater London converging towards those in non-conurbations up to the mid to late 1970s (Figure 5-8, page 139), with Greater London rates intermediate between conurbations and non-conurbations. Regional Warren Spring survey data[37] for black smoke between 1958 –1961 also showed declines over the same time period (Figure 5-3, page 118) as the mortality data with similar relative positions of conurbation regions, Greater London and non-conurbation regions. SO_2 data were similar for conurbations and non-conurbations, but the relative position of Greater London was different as Greater London had the third highest SO_2 levels initially (after Northwest and Yorkshire & Humberside regions) rising to highest levels in 1966-67 (Figure 5-3, page 118). Plots of annual averages from 13 long-running individual monitoring stations also showed a declining trend in black smoke and sulphur dioxide between the 1950s and 1990s. Again similarities between mortality and air pollution data were greater for black smoke, where the declines had levelled off by the 1980s (Figure 5-4, pages 123-125), than for sulphur dioxide, which continued to decline into the 1990s (Figure 5-5, pages 126-128).

Timing of period effects

As previously noted (pages 103-4), using the current formulation of BAMP it is not possible to distinguish between period effects relating to year-to-year fluctuations and those related to longer-term underlying trends. BAMP did identify a minimum period effect in COPD mortality in 1977-78 at similar timing to the longer term levelling off seen in black smoke concentrations as seen in data presented in this chapter (Figure 5-3, page 118 and Figure 5-4, pages 123-125) and consistent with reports on urban smoke concentrations.[244,245] It is possible that this period effect was related to the ICD coding change in 1979, but this seems unlikely as strongest statistical support was for a 1977 effect not 1978 (Table 5-9, pages 155-156) and artefactual changes in rates were not seen at the time in age-stratified rates (Figure 5-6, pages 132 & 133). The levelling off of a decline in SO_2 appears to have occurred later than that for black smoke, but it is

harder to confirm this as less has been published on SO_2 long-term trends. Readings from long-running individual monitoring sites in all areas (Figure 5-5, pages 126-128) and from a review of SO_2 levels in London[296] were suggestive that declines in SO_2 in all areas had converged to similar low levels by 1985. However, evidence for a period effect 1980-85 was weak and inconsistent, with a possible weak effect detected in 1981 in three of six area-sex combinations (Table 5-9, pages 155-156).

Period effects in COPD mortality seen as maxima, consistent with onset of a decline were located using BAMP to 1962-3 in all areas with less consistent maxima seen in the late 1950s. While this is reasonably consistent with long-term trends and the 1962-63 maxima coincide with peak levels for smoke and SO_2 air pollution in all regions according to annual averages from Warren Spring survey data[37] shown in Figure 5-3, page 118, the maximum in Greater London at least may correspond to an air pollution episode in December 1962 (Table 5-4, page 120) where recorded maximum 24-hour black smoke and SO_2 levels in Greater London were over 3000 $\mu g/m^3$.[297]

The Warren Spring air pollution data also showed a clear decline in air pollution levels in Greater London and non-conurbations in the four years 1958-9 to 1960-61 (Figure 5-3, page 118) consistent with the maxima in period effects detected in the late 1950s. However, air pollution episodes were documented[297] in London in December 1957 and January 1959 (Table 5-4, page 120), and it may be these rather than the decline in annual average levels which resulted in the observed period effects in these years. High influenza activity in 1957 (Table 5-9, pages 155-156) may also at least partially account for period effects in those years.

BAMP analyses did not locate any period effects in 1958-1961 in conurbations (Table 5-9, pages 155-156). This was consistent with concurrent Warren Spring annual black smoke and SO_2 concentrations for regions in conurbations, which were static or rose slightly between 1959-60 and 1962-63 (Figure 5-3, page 118).

Average falls in air pollution in conurbations and Greater London 1958/9 to 1970/1 relative to non-conurbations compared with drop in relative risk for period effects during the same period

There were similarities between averaged relative risks for period effects for COPD mortality for conurbations relative to non-conurbations from classical age-period-cohort

analyses and the relative differences between averaged black smoke concentrations (Table 5-15, page 169). Smoothed regression lines through regional air pollution averages using Warren Spring data generally fitted the data well as evidenced by high r^2 coefficients (Table 5-3, page 119), except for sulphur dioxide in non-conurbation regions, where a clear trend was not always seen (Figure 5-3, page 118). The black smoke levels relative to non-conurbations and RR for period effects were reasonably similar for Greater London but less so than with conurbations. The regression line for black smoke for Greater London had a high r^2 (=0.86), but the data suggested a curvilinear pattern (Figure 5-3, page 118) so actual black smoke levels were also presented in Table 5-15 below.

Sulphur dioxide only showed some similarities between concentrations in conurbations relative to non-conurbations and RR for period effects for COPD mortality for 1958/9 but not 1970/1. No similarities were seen for Greater London relative to non-conurbations in either year (Table 5-15). Using actual rather than smoothed air pollution averages for non-conurbations did not alter this finding (not shown).

Table 5-15 Comparison of relative risk for period effects for COPD mortality for conurbations and Greater London for 1958 and 1970 relative to non-conurbations with black smoke and SO_2 concentrations

	Predicted RR period effect relative to non-conurbations from regression of 1957-78 data from classical age-period-cohort analyses		Black smoke concentrations relative to non-conurbations taken from Table 5-3, page 119		SO_2 concentrations relative to non-conurbations taken from Table 5-3, page 119
	Males	Females	Using trend line averages	Using actual concentrations	Using trend line averages
Conurbations					
1958	1.65	1.84	1.73	-	1.67
1970	1.43	1.58	1.47	-	1.08
Greater London					
1958	1.48	1.64	1.35	1.70	2.07
1970	1.31	1.48	0.72	1.17	1.80

Period effects taken from classical age-period-cohort models with shared age and cohort effects, but differing period effects (ApC models).
Air pollution levels concentrations relate to 1958/9 and 1970/1 financial years. Data for conurbations were averaged for North, North West, Yorkshire & Humberside and West Midlands regions. Data for non-conurbations were averaged for East Anglia, East Midlands, South East, South West regions and Wales

Section III

The role of non-smoking factors in spatial variations in COPD mortality in Great Britain 1981-1999

Chapter 6 Methods

Chapter 7 Results

Chapter 6. The role of non-smoking factors in spatial variations in COPD mortality in Great Britain 1981-1999: Methods

Overview of Chapter 6

This part of the PhD concerned an investigation of national spatial variations in COPD mortality (such as the higher rates in more northern versus more southern areas of England[21]) with particular emphasis on the role of non-smoking related factors relating to hypothesis 5 (Section I, Chapter 1). The analysis was confined to 1981-1999 because (i) the temporal analyses suggested that the impact of declines in air pollution had levelled off by the end of the 1970s, suggesting the potential for increases in relative importance of other factors in explaining variations in mortality after this time and (ii) for pragmatic reasons as outcome and risk factor datasets became available at suitable geographical resolution and ICD9 coding was used for mortality throughout this time period.

Key points

The analyses were conducted at district level. This was a compromise attempting to give the highest levels of resolution possible to minimise the ecological fallacy, give a breadth of risk factor datasets to maximise information from the analyses (the coarser the spatial resolution, the more datasets were available), while providing sufficient counts in outcome and risk factor data to give adequate precision in risk factor estimates.

Creation of datasets

- National mortality data for COPD, lung cancer, pneumoconioses and asbestos related diseases for 1981-99 were used. SMRs were calculated, using expected numbers calculated for each year individually and summed over time.
- A district-level deprivation index was created based on the percentage of population living in Carstairs index quintile 5. The association of quintiles of

this district-level index with COPD mortality was compared with that of ward level Carstairs index and was considered likely to under-estimate the increased risk associated with higher levels of deprivation on COPD and lung cancer mortality as well as over-estimate the protective effect of lower levels of deprivation.

- Interpolated black smoke and SO_2 air pollution data were only available for 1996 and were obtained from NETCEN. To see what relation 1996 data might have with other years, readings from approximately 60 individual monitoring stations in operation over 1981-99 were used. Correlations of around 0.4 to 0.5 between annual means in 1981 and 1996 and of 0.7 between 1996 and 1999, suggesting that the data for 1996 would have at least a moderate relation with relative ambient exposures in the 1980s.
- Interpolated daily average minimum temperature and rainfall were obtained from the European Union "Monitoring Agriculture with Remote Sensing Unit (MARS) project and were based on approximately 30 weather stations.
- Average district fruit and vegetable purchase weights were obtained from the TNS super panel dataset, the validation of which was described in Section I, Chapter 3. The dataset used was restricted to households with data for at least four of every five weeks, representing 20,000 households taking part in the survey for a median of 63 weeks. To avoid unrepresentativeness, the contributions of households taking part in the survey for less than 12 weeks were weighted by the proportion of number of weeks in the survey divided by 12.

Statistical analyses

- Descriptive analyses were conducted for all datasets.
- A Bayesian shared component analysis was conducted to separate shared and independent lung cancer and COPD risks for districts. Lung cancer was assumed to be a good marker of cumulative smoking and removing risks shared between COPD and lung cancer would remove smoking-related risks.
- Linear regression was conducted using the log independent COPD risk estimated from the shared component analysis as the outcome and district-level deprivation, air pollution, occupational mortality (pneumoconioses and asbestos-related diseases), minimum temperature, rainfall and fruit & vegetable purchases as outcomes to see if these non-smoking risk factors could explain spatial patterns where COPD and lung cancer mortality distributions were discordant.

Data

District-level datasets for mortality from COPD and lung cancer 1981-1999 and population 1981-1999 were created. Using information from the literature search, datasets on a number of non-smoking related factors were compiled including

- area-level deprivation (based on the 1991 Census)
- ambient air pollution concentrations (interpolated ambient black smoke and sulphur dioxide levels for 1996)
- area-level fruit & vegetable intake (approximated by fruit and vegetable purchases for 1991-2000)
- meteorological factors (interpolated meteorological data on average daily minimum temperature and average daily rainfall 1981-1999)
- occupational exposures (mortality from pneumoconioses and asbestos-related diseases 1981-1999)

Full details about the creation of these datasets and checks on the validity of these exposure variables are given below.

Geographical resolution

Mortality were extracted at ward-level defined using 1991 boundaries (ward91), from the Small Areas Health Statistics Unit (SAHSU) databases. Analyses were conducted at district level where district was derived from ward by using the first four units of this six-unit code. Although other geographical ward boundaries were available (e.g. ward81 and ward96), all SAHSU data had already been mapped to ward91 boundaries and this was consistently available over all years 1981-1999. Mortality data were then aggregated to district after cleaning (see below). Fruit and vegetable purchases were available by ward91 and aggregated to district level. Air pollution interpolations were available on a 1km grid and meteorological factors were available on a 50km grid. GIS (geographic information systems) software was used to drop these data onto district-counties, using area-based weighted means where the grid covered more than one district. The creation of a district-level deprivation score based on ward-level deprivation index is described below.

Mortality datasets 1981-1999

Mortality datasets were extracted from SAHSU databases, which originated from the Office for National Statistics (ONS). COPD was defined as ICD9 codes 490-496 (i.e. including asthma), lung cancer as ICD9 162 and analyses were restricted to those aged 45+ years. Cleaning involved the exclusion of 1,042 (0.18% of 573,931 total) COPD deaths and 1,643 (0.23% of 706,145 total) lung cancer deaths mainly due to missing ward and/or year. Additional small numbers of deaths were excluded because wards had no population in the age-sex group in which the death occurred (see population paragraphs below), wards had male but no female population (25 wards) or wards had a missing deprivation score (10 wards). The final datasets were based on 10,510 (96%) wards out of a total possible of 10,933.

Pneumoconiosis deaths were defined as ICD9 codes 500 (coal workers' pneumoconiosis) and 502-505 (respectively pneumoconiosis due to other silica or silicates, pneumoconiosis due to other inorganic dust, pneumopathy due to inhalation of other dust, pneumoconiosis unspecified). Asbestos related deaths were defined as ICD9 codes 501 (asbestosis), 163 (malignant neoplasm of the pleura) and 158.8 and 158.9 (malignant neoplasm of the peritoneum), following that used by Coggon et al[299] to define asbestos-related deaths in the ONS Decennial supplements. This identifies about 60% of mesothelioma deaths as captured by the national (GB) mesothelioma register.[300]

SMRs for 1981-99 were calculated as the sum of observed counts divided by sum of expected numbers, where expected numbers were calculated for each year separately applying national rates by age and sex to each ward. Exact 99% confidence intervals were calculated assuming that the observed number of deaths had a Poisson distribution with expected values the calculated expected counts. The level of 99% was chosen to partly adjust for multiple comparisons.

Population datasets 1981-1999

ONS population data were obtained from SAHSU databases. SAHSU population figures from 1982 to 1990 had been estimated using a straight line interpolation between the 1981 and 1991 census. The population figures from 1992 to 1999 were derived from the ONS mid-year population estimates. These population estimates were based upon the 1991 Census with an allowance for under-enumeration in the Census

and incorporate information on births, deaths and net migration. Slight variations with ONS data occurred as ONS population estimates were released at district level and then distributed to ED91 geography using 1991 census demography by SAHSU and then summed to provide the current estimates, resulting in a small number of wards with deaths but no population in a particular age-sex group as described above.

Deprivation index for 1991

The deprivation index used was based on Carstairs quintile, being the quintile of Carstairs scores for wards as derived from the 1991 Census, in the middle of the time period of analysis. A district-level deprivation index covering the analysis period was in existence: the Index of Local Conditions 1991.[301] However, this was developed for use in funding allocation to local government and has rarely been used in epidemiological studies.

The deprivation index chosen in this analysis was the percentage of the population in the district living in the most deprived Carstairs quintile (quintile 5, where quintiles were based on scores from all 10,933 wards prior to cleaning exclusions). This index had the advantage of being based on Carstairs score rather than creating a new and untested variable. Also it was similar to one of six of the Index of Multiple Deprivation 2000 (IMD2000) indices making up the district-level IMD2000 scores (extent of deprivation – the proportion of a district's population living in wards that rank within the most deprived 10% of wards in England)[302]

Quintiles were then created of this district-level deprivation index. However, as 28% (i.e. more than 20%) of districts had no population living in wards assigned to Carstairs quintile 5, the 'quintiles' of the district deprivation index were chosen to be 0% for the first division with remaining percentage of population in districts living in wards in Carstairs quintile 5 divided into quartiles. For convenience, this index was been referred to as a quintile throughout.

To investigate how the district deprivation index performed with respect to COPD and lung cancer deaths, a comparison was made between the distribution of statistically significant SMRs (using 99% confidence intervals) by Carstairs quintiles for wards and by quintiles of the deprivation index for districts.

The expected association with ward-level deprivation was seen, with high COPD and lung cancer SMRs clustering in the most deprived wards (Figure 6-1, page 177) and low SMRs more common in the less deprived wards (Figure 6-3, page 178). The created district level deprivation index showed a reasonably similar pattern for high SMRs (Figure 6-2, page 177), with the bulk of high SMRS in quintiles 4 and 5. However, there were about 15% fewer of the high SMRs in the most deprived quintile than when using a ward-based measures, most of which appeared to be captured by quintile 4. With low SMRs (Figure 6-4, page 178) the district-level variable put more of these in quintile 1 (0% of population living in Carstairs quintile 5) – about 10% more in males and 20% more in females, with similar percentages in quintile 2 and lower percentages in quintiles 3,4,5 than with the ward-level deprivation index (Figure 6-3, page 178).

It was concluded that the bias using the created deprivation index was likely to under-estimate the increased risk associated with higher levels of deprivation on COPD and lung cancer mortality but also to over-estimate the protective effect of lower levels of deprivation.

Figure 6-1 Wards with COPD SMRs statistically significantly higher than 100 by Carstairs quintile (5 = most deprived)

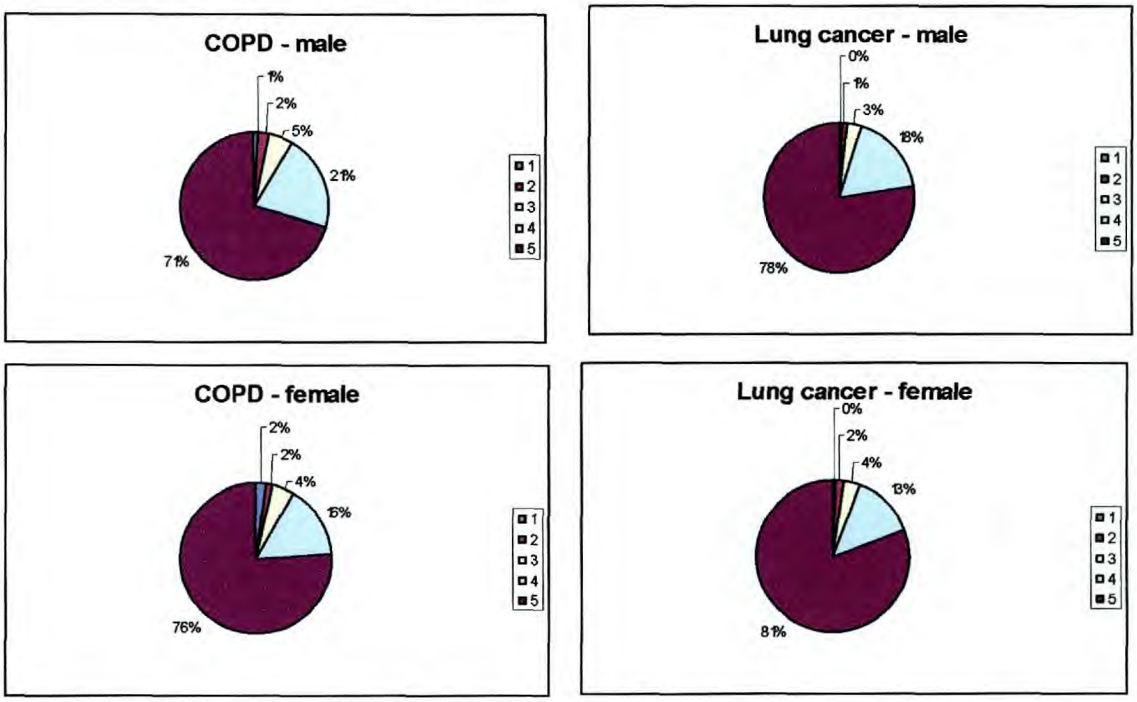


Figure 6-2 Districts with COPD SMRs statistically significantly higher than 100 by deprivation index quintile (5 = most population living in deprived areas)

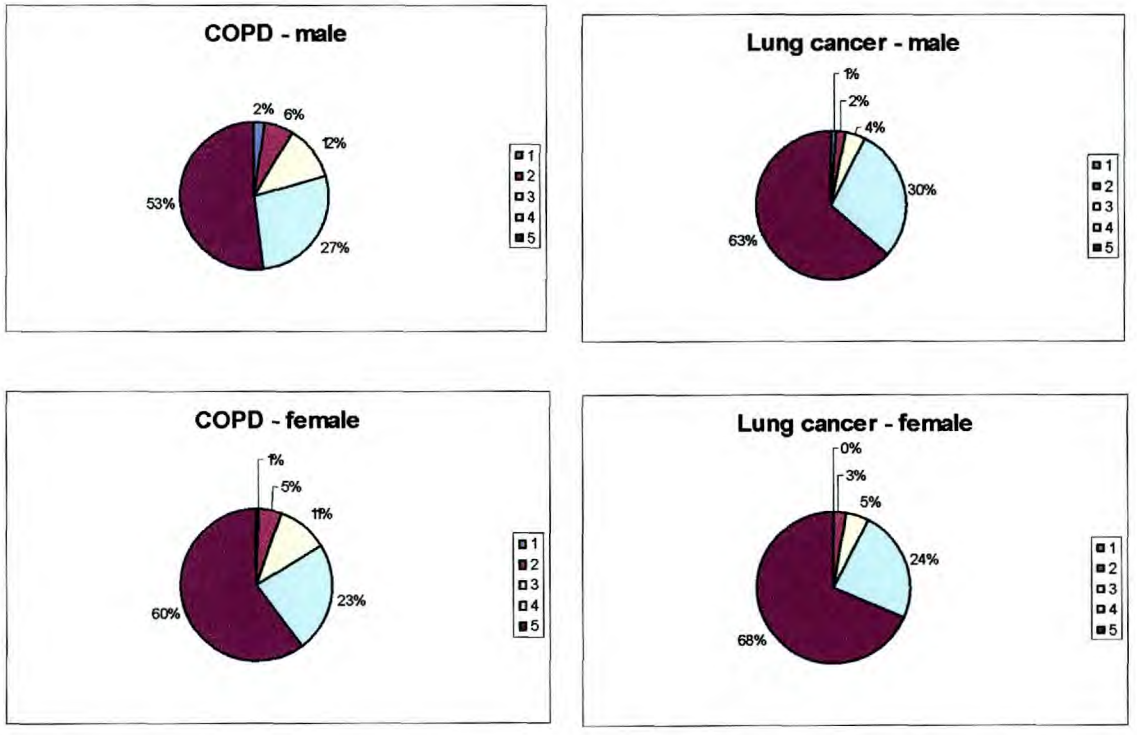


Figure 6-3 Wards with COPD SMRs statistically significantly lower than 100 by Carstairs quintile (5 = most deprived)

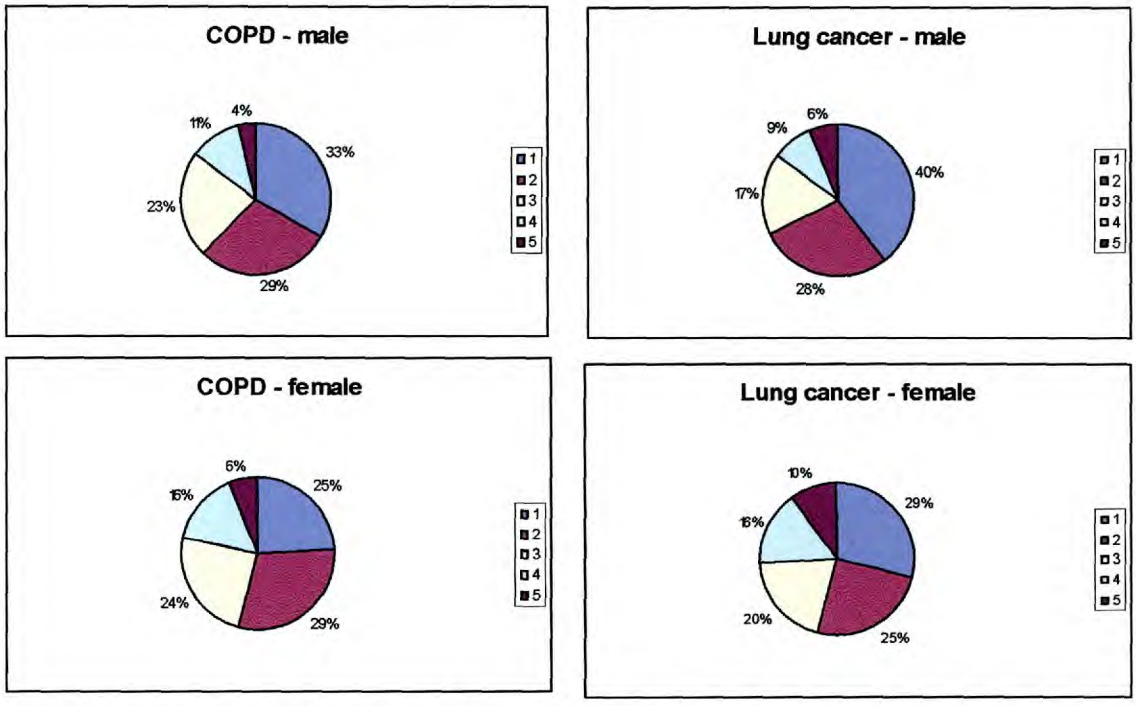
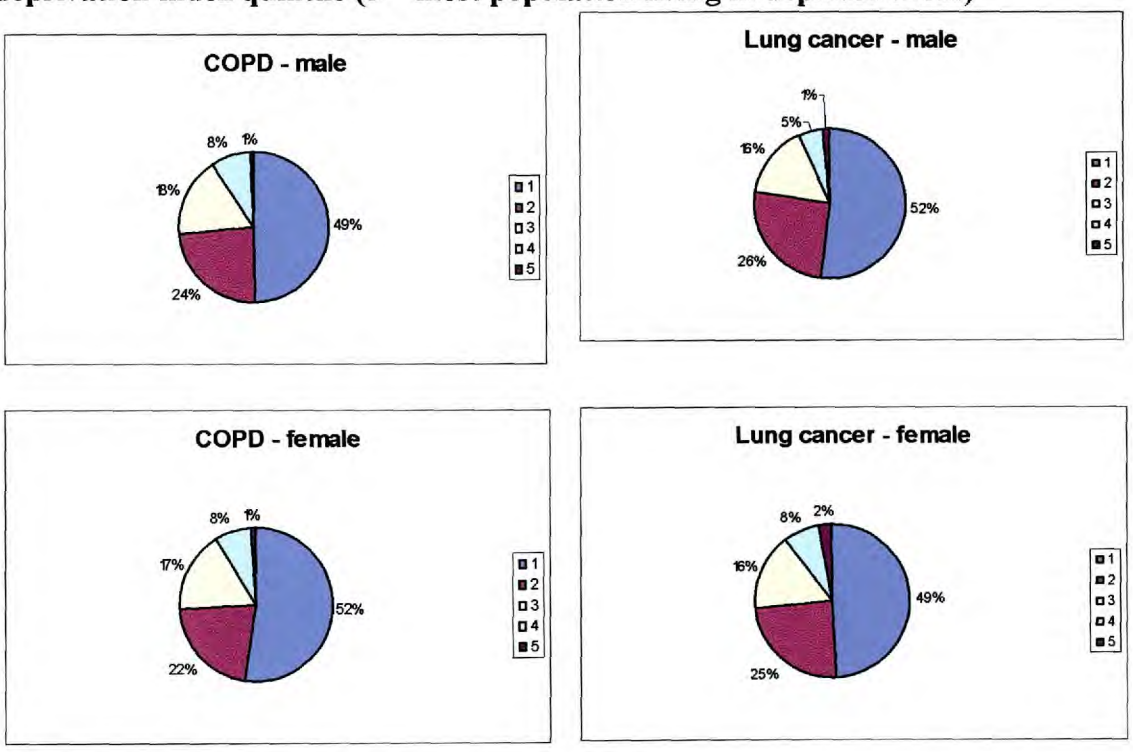


Figure 6-4 Districts with COPD SMRs statistically significantly lower than 100 by deprivation index quintile (5 = most population living in deprived areas)



Air pollution datasets: PM_{10} and SO_2 concentrations 1996

Interpolated background PM_{10} and SO_2 concentrations were only available for 1996 in the time period 1981-1999. Data for 1996 covering the whole of the UK were obtained from the Data Quality Archive, maintained by the AEA Technology's National Environmental Technology Centre (NETCEN).[303] These data were produced using a combination of emissions and monitoring data. Validation of the data by NETCEN[303] suggested that estimates of SO_2 were 'reasonably good' in city centre or rural locations where the majority of the automatic monitoring stations are located, but the map may have underestimated concentrations in many smaller urban or suburban areas. Validation of PM_{10} values by comparisons with vehicle emissions in 26 sites were stated to be 'reasonably consistent'[303] except in two central London sites.

Examining the validity of using 1996 data to apply to 1981-1999

To help interpret the changes in air pollution occurring over 1981-1999, data for annual average black smoke and SO_2 concentrations from 2,302 and 2,253 individual monitoring stations respectively by year for 1981-1999 were also obtained from Netcen. The correlations between air pollution levels by year for these stations were examined to see if 1996 area levels would be meaningful about exposures about other years in the 1980s or 1990s.

Using all data, but excluding missing values, correlation coefficients of annual mean values showed that as expected correlations fell off over time from about $r = 0.8$ to $r = 0.9$ for correlations between one year and the next year to around $r = 0.3$ for smoke and $r = 0.5$ for SO_2 between 1981 and 1999. Since stations went in and out of operation during this time period and some had missing values for one or more years, the factors causing this may have introduced bias affecting comparability over time, so further analyses were conducted on stations with readings available for each year.

Sixty-five stations located in 46 districts had black smoke readings for each year 1981-1999. All regions had more than one of these stations, except for East Anglia, but sitings were non-random – nine districts had more than one station, two districts – Barnsley and Wakefield in Yorkshire & Humberside region – had five of these stations each.

Sixty stations had annual mean SO_2 concentrations available for each year over this time period. All regions were represented except South West England and East Anglia again had only one station (Norwich). Eight districts had more than one station, one district had five – Barnsley in Yorkshire & Humberside region.

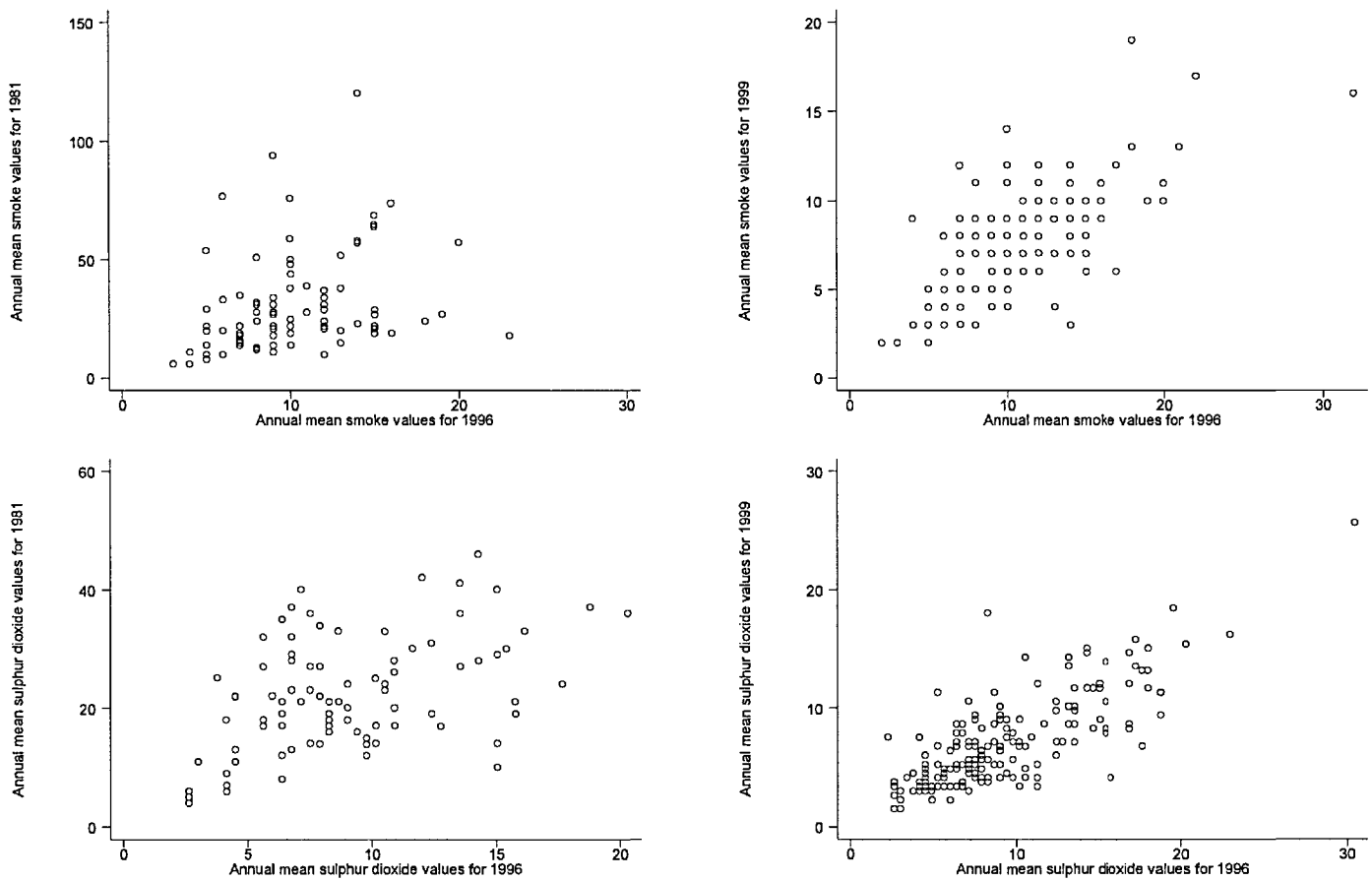
Correlations were moderate (around 0.5) between 1981 and 1996 and, not unexpectedly, higher (around 0.7) between 1996 and 1999 (Table 6-1). Scatterplots of the underlying data can be seen in Figure 6-5 (page 181). Spearman rank correlations were also performed and gave similar values to product moment correlation coefficients (Table 6-1 below) suggesting that the relative levels of stations had changed together with the values. Similar values for the correlations were seen (not shown) if using the using the full data set of all monitoring stations (with missing readings in some years).

Table 6-1 (Product moment) correlation and Spearman rank correlation coefficients for black smoke and SO2 data between 1981 & 1996 readings and 1996 & 1999 readings (n = number of stations)

	Correlation 1981 & 1996 (r)	Spearman rank correlation 1981 & 1996 (ρ)	Correlation 1996 & 1999 (r)	Spearman rank correlation 1996 & 1999 (ρ)
Black smoke (n=65)	0.53 **	0.58 **	0.74 **	0.73 **
SO_2 (n=60)	0.48 **	0.42 **	0.63 **	0.66 **

**p<0.001

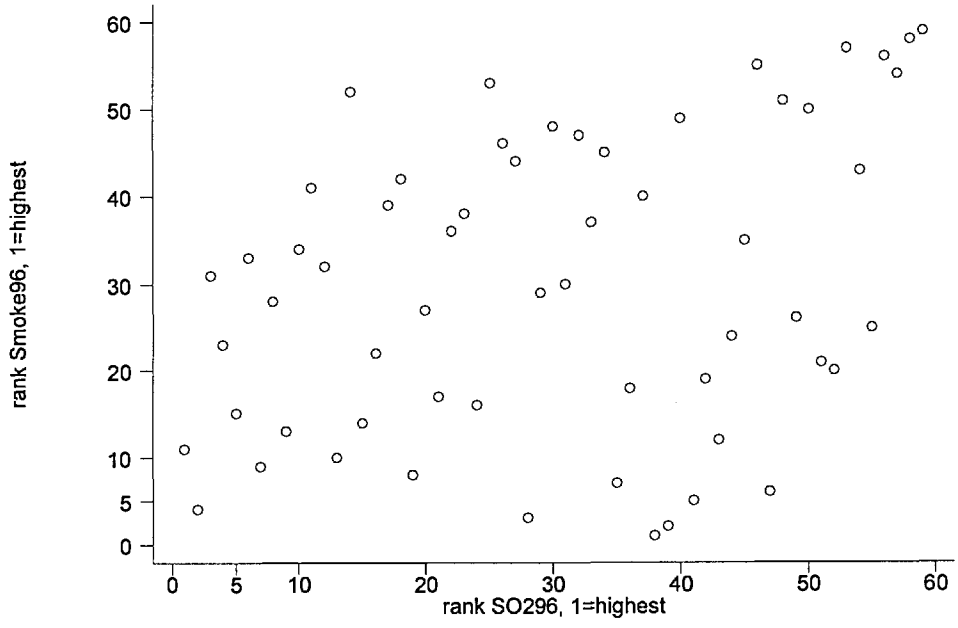
Figure 6-5 Scatterplots of annual mean black smoke in μg/m^3 and annual mean SO_2 in ppb from continuous-running monitoring stations for 1981 & 1996 and 1996 & 1999



Using 59 stations with information on both black smoke and SO_2 each year for 1981-99, SO_2 rankings in 1996 were only weakly correlated with black smoke rankings in the same year of 1996 (Figure 6-6 below) or with black smoke rankings in 1981 or 1999, giving spearman correlation coefficients (ρ) of around 0.3:

- Correlations between black smoke in 1996 and SO_2 in 1996 $\rho = 0.35$, $p = 0.006$;
- Correlations between black smoke in 1981 and SO_2 in 1996 $\rho = 0.35$, $p = 0.007$;
- Correlations between black smoke in 1999 and SO_2 in 1996 $\rho = 0.33$, $p = 0.01$;
- All comparisons were made using 59 monitoring stations with annual mean values for each year 1981-99, but similar values were obtained using the full data set of all monitoring stations with 88 to 234 stations involved in the correlations.

Figure 6-6 Scatterplot of rankings for black smoke and for SO_2 in 1996 for 59 stations with no missing values 1981-1999



Meteorological data: average minimum temperature and rainfall for 1981-1999

Daily temperature and rainfall data for 1981-1999 on a $50km^2$ grid were obtained from the European Union "Monitoring Agriculture with Remote Sensing Unit (MARS)".[254] This consisted of daily meteorological data from 30 weather stations in Great Britain on which quality checks had been carried out. Rainfall data were not interpolated but taken

from the weather station most similar to the grid centre using a score derived from the distance between the weather station and the grid centre, the altitude differences, the distance from the coast and presence or absence of a climate barrier. Temperature data *were* interpolated using data from up to four stations, with the number of stations and choice of stations dependent on a scoring system based on factors as described for rainfall.[304]

A number of possible climate variables were considered for these analyses, including winter temperatures (as highest COPD mortality is seen in the winter), the variability of temperatures and the number of dry days, but the final choice was for mean minimum temperature and mean rainfall as these were straightforward measures and the epidemiological evidence did not strongly suggest other measures.

Fruit and vegetable purchases using TNS 1991-2000

Creating the district-level dataset

A full description of the original dataset and the creation of a fruit and vegetable purchase variable were given in chapter 3 on Data Quality (page 72). Data were available by household for 1991-2000 and to maximise the available information, data for 2000 were left in. The assumption was made that spatial variations in average district patterns of fruit and vegetable purchases over the 10 year period 1991-2000 would also be representative of variations in the 1980s.

There were 33,177 households represented in the dataset for 1991-2000. Because of concerns about bias, households where more than one in five weeks had zero purchases were recorded were removed (13,171 or 39.7% of households). This left 20,006 households in the study. The median number of weeks in the survey for these households was 63 weeks (range 1 to 522 weeks).

Purchases by each household were summed over district level and two variables representing average fruit and vegetable purchases were created:

- Quintiles of total fruit & vegetable purchases in g per person per day
- Portions of purchased fruit & vegetable per person per day (taking a portion as 80g, except for fruit juice, where a portion was 125g (taken as approximating to 125ml – also note that no district had >= 1 portion fruit juice per day)

Approximately a fifth (4,417 of the 20,006) households were in the survey for less than 12 weeks. Because of concerns about representativeness of purchases it was decided to weight the contribution that these households made to the district-level averages by the number of weeks in the study divided by 12. The number of households per district was also weighted with any households contributing <12 weeks treated as a proportion of a household corresponding to [*number of weeks in survey/12*].

Descriptive analyses

Descriptive analyses were carried out for each dataset and the district-level SMRs for mortality variables or district-level average air pollution, minimum temperature, rainfall, or fruit & vegetable purchases were mapped using Arc View (a Geographical Information Systems (GIS) package).

Shared component analyses

The spatial analyses were concerned with areas where spatial patterns of lung cancer and COPD mortality were discordant, as set out in the fifth hypothesis (chapter 1). To identify these areas, a Bayesian shared component model was used, adapted from a model developed by Knorr-Held and Best.[215] This followed the usual steps involved in Bayesian analyses as described in Gelman et al[279] and discussed in Chapter 3 (page 83):

i. Setting up a full probability model – a joint probability distribution for all observable and unobservable quantities incorporating prior beliefs (or no prior beliefs) about the relationships under study
ii. Conditioning on observed data – calculating the conditional probability distribution of the unobserved quantities of interest, given the observed data
iii. Evaluating the fit of the model and implications of the resulting posterior distribution.

Setting up the full probability model

As set out in Knorr-Held & Best,[215] COPD and lung cancer are treated as two competing diseases, where each individual can fall into one of three categories – COPD death, lung cancer death or other (survived), which can be expressed using a

multinomial formulation. Rare disease assumptions for each condition allow replacement of the multinomial model by two independent binomial models for each disease and the replacement of these by Poisson models. Although COPD currently constitutes 5% of deaths[17] and lung cancer a further 5%,[14] the risk of death for the (live) population is rare. The Poisson rate can then be expressed as the product of the relative risk times the expected number of counts.[305]

The overall relative risk is assumed to be the same for both diseases. The SMRs are internally standardised so that the overall SMR for the study region is 100. The relative risk for district i is modelled as the product of a shared component θ_i and a disease-specific component ϕ_{di}, where the disease is lung cancer or COPD (in notation d=1,2). Disease counts y for each disease (1 or 2) and each district (i) are assumed to be conditionally independent Poisson random variables. In statistical notation this is written as

Equation 6-1

$$y_{1i} \sim Poisson(e_{1i}\theta_i^{\delta}\phi_{1i})$$

Equation 6-2

$$y_{2i} \sim Poisson(e_{2i}\theta_i^{1/\delta}\phi_{2i})$$

where

y = disease count for disease 1 or 2 in area i

e = expected count for disease 1 or 2 in area i

θ = shared (risk) component for disease 1 and 2 in district i

ϕ = disease-specific (risk) component for disease 1 or 2

δ = a scaling parameter

and where θ_i, ϕ_{1i} and ϕ_{2i} are assumed to be independent.

The scaling parameter δ weights the contribution of the shared component to the overall risk to allow a different risk gradient to be associated with this component . This is required to allow the shared component of risk to have a stronger association with one disease than the other (the shared component θ_i should be thought of as representing unmeasured risk factors – as with measured risk factors the regression coefficient(s) associated with these factors may be larger for one disease than the others.)

Another prior assumption is that the shared and disease-specific components have a strong spatial structure. In the Knorr-Held & Best model[215] this was modelled using spatial cluster models. Here, the formulation proposed by Besag, York and Mollié[306] was used as this can be fitted within readily available Bayesian statistical software. The Besag-York-Mollié formulation assumes that the risks are composed of spatially structured and unstructured elements – this is sometimes referred to as a convolution prior. This assumption was applied to both the disease-specific and shared components of the model. The unstructured elements of the log relative risk parameters (for shared and disease-specific components) were assumed to be independent of each other and drawn from a normal distribution with unknown mean μ and variance σ^2. The structured elements of shared and disease-specific components were assumed to follow a conditional auto-regressive distribution, sometimes referred to as a CAR prior. A CAR prior assumes that the parameter in each area is smoothed towards the mean risk of the neighbouring areas. The variance is inversely proportion to the number of neighbours. Also, the smaller the number of observations for each area, the greater the extent of smoothing towards the values of the neighbours.

The statistical notation for the CAR prior can be written as follows in Equation 6-33, using the example of the shared parameter θ in area i with neighbouring districts j :

Equation 6-3

$\theta_i|\theta_j, j \neq i \sim$ Normal (m_i, v_i)

The mean for area i denoted as $m_i = \frac{1}{n_i} \sum_{j \in \partial_i} \theta_j$

where

∂_i =set of areas adjacent to area i

n_i = number of neighbours.

The variance for area i denoted as $v_i = \frac{v^*}{n_i}$

where v^* is the overall scale (variance) parameter

As set out in Knorr-Held and Best,[215] in the limiting case that the shared component has no spatial structure θ_i will be constant for all i (all areas). The disease counts for each disease, y_{1i} and y_{2i} will be independent Poisson with spatial variation in relative risk

determined by the disease specific components ϕ_1 and ϕ_2. In the limiting case that the specific components are the same, i.e. $\phi_1 = \phi_2$, both diseases will have a common relative risk pattern determined through the shared component θ. The model framework allows the whole range between these limiting cases, with the data allowed to indicate the relative strengths of the shared and disease-specific components, through the choice of diffuse priors.

Choice of diffuse priors

A flat prior (prior distribution that is uniform) was chosen for mean values μ of the spatially unstructured elements, with mean relative risk parameters of both the disease-specific and shared components set at 0. Inverse gamma priors (prior values with an inverse gamma distribution) were assumed for the variances of both spatially structured and unstructured elements of the shared and disease-specific log relative risks.

The log of the scaling parameter δ was assigned a normal prior with mean 0, therefore any value δ_0 is as equally likely as the reciprocal value $1/\delta_0$. The relevance of this choice means that it does not matter which disease is labelled 1 or 2 in Equation 6-1 and Equation 6-2 on page 185, which contain δ and $1/\delta$ parameters respectively. Switching the label will result in the posterior value for δ changing to the reciprocal distribution, but the same posterior distributions for joint and specific components.

Conditioning on observed data

The model was implemented in WinBUGS version 1.4.[280] BUGS stands for **B**ayesian inference **U**sing **G**ibbs **S**ampling – this is widely used software for Bayesian statistical analysis.

District neighbours were assigned using the adjacency tool within WinBUGS. The spatial boundaries were then mapped and inspected to ensure that each district had at least one neighbour. Particular attention was paid to islands, which had to have neighbours manually assigned – generally, districts on the mainland and adjacent islands as appropriate. Particular care was needed with the Isle of Wight, which contained two districts neighbouring each other. These also had to be assigned neighbours on the mainland for the model to run successfully.

Initial sampling using two chains was conducted until convergence of both chains was reached (5,000 to 10,000 iterations). Convergence was checked using trace plots of parameters of sample values against iterations to examine for the point where chains were overlapping and not drifting apart. Realistic starting values for sampling based on past experience of the model were assigned, so that time to convergence was minimised.

Samples obtained during initial runs to reach convergence of the model were discarded. Quoted results were based on a further 50,000 iterations using 2 chains (100,000 samples in total).

Parameters calculated included the posterior means of the shared component risk θ, the disease-specific risk component ϕ for COPD and lung cancer and the associated posterior variances. The percentage of the variance of the total risk that was shared or disease-specific was also calculated. The combination of the shared and disease-specific components were also combined to produce a smoothed posterior relative risk for each area.

Evaluating the fit of the model

Assumptions of both a strong spatially structured and unstructured component to the relative risks were made in the model. Examination of posterior distributions found approximately 40% of the variance of the relative risk parameters were unstructured, so both structured and unstructured terms were retained in the model.

The variance of the spatial and unstructured components was compared by calculating the percentage the spatial variance made to the total variance (the sum of the spatially structured and unstructured variance.) As the variance of the spatially structured components calculated as part of the model is a conditional variance dependent on the neighbours as set out in Equation 6-3 on page 186, an empirical variance calculation for each area i was included in the model code to use for this calculation.

Linear regression

The final part of the analysis looked at the relationship between the exposure variables and the independent COPD risk from the shared component analysis – i.e. the risk not shared with lung cancer as identified by the model – using the outcome variable ϕ_1, the

(natural) log of the independent COPD risk factor (from here on referred to as COPDspec).

Grouping of exposure variables

In the regression, exposures of interest were included as categories e.g. quintiles of SO_2, because analyses were being made on the assumption that the relative differences between districts had remained relatively constant over time while absolute values may have changed. Not using absolute values prevented over-interpreting the data – e.g. the statement that a 1 unit rise was associated with a certain percentage rise in COPDspec. The temporal association or lag between exposure and outcome could not be explored in this analysis and any association may have related to the association of the exposure variable with past levels.

Exposure variables were chiefly grouped as quintiles as described in Table 6-2, except for SMRs for pneumoconioses and asbestos-related disease, which were grouped as to whether they were statistically significantly different from 100. Variables grouped as quintiles were also regrouped as tertiles and binary variables (greater or less than the median) to allow sensitivity analyses to the choice of categories. To avoid the influence of extreme SMRs on the later regression analyses, pneumoconiosis and asbestos SMRs were grouped into three, corresponding to statistically significantly less than 100 (using exact 99% confidence intervals), not significantly different from 100 and statistically significantly greater than 100.

Table 6-2 Groupings of exposure variables used in linear regression analyses

Exposure variable	Grouping	Description
Deprivation (% of district population living in wards in the most deprived Carstairs quintile, quintile 5)	Modified quintiles	Quintile 1=0% of ward populations living in Carstairs quintile 5. Quintile 2-5 = quartiles of all %s >0
PM_{10} in mcg/m^3	Quintiles	
SO_2 in ppb	Quintiles	
Temperature	Quintiles	
Rainfall	Quintiles	
Pneumoconiosis SMRs	3 categories	1 = not stat. sig. different from 100 2 = stat. sig. <100 (using 99% CIs) 3 = stat. sig. >100 (using 99% CIs)
Asbestos related SMRs	3 categories	1 = not stat. sig. different from 100 2 = stat. sig. <100 (using 99% CIs) 3 = stat. sig. >100 (using 99% CIs)
Fruit & vegetable purchases in g per week	Quintiles	

CI = confidence interval

Descriptive associations of exposure variables with COPDspec

Before proceeding to regression analyses, distributions of variables and graphical displays using box and whisker plots of the relationship between COPDspec and the exposure variables were examined. The degree of association between exposure variables was examined using Spearman rank correlations.

Linear regression model

Linear regression was conducted using COPDspec as the outcome variable. This corresponded to the model

$$\phi_1 = \alpha + \beta_1 x_1 + \beta_{2a} x_{2a} + \beta_{2b} x_{2b} + \beta_{2c} x_{2c} + \beta_{2d} x_{2d} + \beta_{2e} x_{2e} + \beta_3 x_3 + \ldots + \varepsilon$$

where ϕ_1 = ln (independent COPD risk factor)

α = a constant

β_1 = coefficient for linear variable x_1

β_2 = coefficient for categorical variable x_2 with 5 categories a to e

β_3 = coefficient for linear variable x_3

etc.

ε = residual error (assumed to be normally distributed)

Here, the relative change in the independent COPD risk (exp (COPDspec)) is represented by the exponential of coefficient β_x where a positive coefficient represents an increase in independent COPD risk with increase in exposure to factor x and a negative coefficient represents a protective effect related to increasing exposure to factor x.

The linear regression was weighted using analytic weightings equal to the precision of COPDspec (calculated as the 1/(square of the posterior standard deviation)).

Univariate analyses

Firstly, univariate analyses were conducted between COPDspec and each exposure variable. The grouped exposure variables were analysed first as both categorical and then as linear (interval scale) variables using likelihood ratio testing to indicate whether the variable could be treated as categorical in multiple regression.

Multivariate analyses

Next all variables were entered into the linear regression model as there were strong *a priori* reasons that all of these variables might be related to the outcome. Variables with $p > 0.05$ from the Wald test were dropped one at a time and the contribution of that variable to the model investigated using likelihood ratio testing. If two or more variables were dropped, the first and subsequent variables were added back in again to see if had now become significant (using Wald and likelihood ratio tests). The use of tertiles or binary variables rather than quintiles was explored – as models involved differences in variables, comparisons were made using Akaike's Information Criteria (AIC). Where variables were treated as categorical in the model, and not all of the categories were significantly associated with the outcome using Wald tests, the joint significance of the categories was explored using an F-test.

Regression diagnostics such as plots of residuals against fitted values were examined for the final main model. As a check that the shared component analysis had fully separated the independent COPD risk from risks shared with lung cancer, the (natural) log of the lung cancer SMR was added back into the final model. No association between the independent COPD risk factor and lung cancer SMRs was expected but if log lung cancer SMRs were significant (using Wald and likelihood ratio testing), it would have indicated some residual confounding.

Multiple regression analyses were conducted on 451 of 459 districts in Great Britain. Eight districts were excluded mainly because data were missing from the TNS dataset (Table 6-3).

Table 6-3 Excluded districts from linear regression and reason for exclusion

District code	District name	Reason for exclusion
01AA	City of London	Large numbers of missing weeks in TNS dataset
16FD	Isles of Scilly	Missing from TNS dataset, temperature and rainfall dataset
6121	Caithness	Large numbers of missing weeks in TNS dataset
6126	Skye and Lochalsh	Missing from TNS dataset
6127	Sutherland	Missing from TNS dataset
6554	Orkney	Missing from TNS dataset
6655	Shetland	Missing from TNS dataset, temperature and rainfall dataset
6756	Western Isles	Missing from TNS dataset

Secondary analyses

Two types of secondary analyses were conducted. Firstly, a sensitivity analysis to see if restricting analysis to districts with at least 10 households participating per district affected the significance or the regression coefficient for the fruit & vegetable variable. Secondly, an examination of interactions with deprivation was conducted. *A priori* expectations were that deprivation could potentially interact with all of the exposure variables of interest and interactions between deprivation and variables significant in the final model were investigated using Wald test and likelihood ratio testing. Similar interaction analyses were conducted to look at interactions between temperature and rainfall.

Chapter 7. The role of non-smoking factors in spatial variations in COPD mortality in Great Britain 1981-1999: Results

Overview of Chapter 7

This chapter contains the results from the spatial analyses section of the PhD.

Key findings

Descriptive analyses

- COPD and lung cancer SMRs were highly correlated and showed similar geographical distributions in conurbation areas: central and south-west Scotland, Tyneside, north-west and central England (Merseyside, Manchester, Leeds, Sheffield, Birmingham), south Wales, Greater London, the Thames estuary. These were also areas with higher deprivation and ambient SO_2 levels.
- PM_{10} data for 1996 were not considered to be representative of other years 1981-99 so PM_{10} data were not used in subsequent analyses.
- Maps suggested a south-west to north-east gradient in minimum temperature, with the warmest areas in the south-west. Rainfall showed a west-east gradient, with the wettest areas in south Wales and west and central Scotland.
- For men, districts with pneumoconiosis SMRs greater than 100 were concentrated in coal-mining areas (some Welsh valleys had SMRs >1000) and districts with high asbestos-related SMRs were in ship-building areas and central London. There were few districts with SMRS statistically significantly greater than 100 for pneumoconioses or asbestos-related diseases for women.
- District-level fruit and vegetable purchases showed variability across Great Britain, but the general impression was of highest purchases in southern England and lowest purchases in the north-west of England and Scotland.

Shared component analyses

- The percentage of the variance accounted for by the shared component was 64% for COPD and 70% for lung cancer smoothed relative risks in males and females i.e. approximately a third of overall risks could be thought of as disease-specific.
- Comparing both log COPD SMRs and log independent COPD risks from the shared component analyses with log lung cancer SMRs suggested that the shared component analysis had removed most of the risk shared between COPD and lung cancer, but that a small residual association remained.
- Higher COPD risks were seen in Tyneside, the north-central conurbation ring (the Liverpool, Manchester, Leeds, Sheffield and Birmingham conurbation areas), central London and south Wales but not central Scotland (which includes Glasgow and Edinburgh). Lower risks were located in East Anglia, the south coast of England, Devon, Edinburgh (but not Glasgow) and north-west Scotland.
- Areas with higher shared risks were the same as those with higher COPD risks plus the addition of central and southern Scotland.
- Districts with higher independent lung cancer risks were generally, but not exclusively different from those for the independent COPD risks, being located in central Scotland, Tyneside, Merseyside and (especially females) central London. Lower lung cancer risks were chiefly located in central England, non-conurbation districts in northern England and south and mid Wales.

Linear regression

- The final multivariate model showed significant associations between the log independent COPD risk and deprivation, ambient SO_2, temperature and rainfall in both males and females. Additionally significant associations were seen between pneumoconioses and asbestos-related disease SMRs in males and fruit & vegetable purchases in women. Final models explained 48% of the variance of the log independent COPD risk for males and 38% for females.
- Independent COPD risks were:
    - higher by 8.5% (6.1% to 11.0%) in males and 13.1% (10.1% to 16.3%) in females in most deprived compared with least deprived districts
    - higher by 9.3% (6.6% to 12.0%) in males and 9.8% (6.4% to 13.4%) in females in districts in highest compared with lowest quintile of SO_2

- lower by 5.0% (2.3% to 7.7%) for females in districts in highest quintile of purchases of fruit & vegetables compared with lowest
- higher by 7.4% (5.3% to 9.6%) in males where districts had pneumoconiosis SMRs >100 and lower by 3.5% (2.2% to 4.8%) in males where district pneumoconiosis SMRs were <100
- *lower* by 2.5% (4.5% to 0.5%) in males where asbestos-related diseases SMRs were higher than 100 (but univariate models had suggested *higher* risks for these districts in both males and females)
- statistically significant for minimum temperature and rainfall in the final model, but associations were complex and difficult to interpret.

- The addition of the (natural) log of the lung cancer SMR in to the final model was not significant on likelihood ratio testing.
- Significant interactions with deprivation were seen for SO_2, temperature and asbestos-related SMRs in males and with SO_2 for females. The effect of SO_2 appeared to be highest in the least deprived districts. The interaction effects of deprivation with temperature for males were difficult to interpret. Lower independent COPD risks in districts with high asbestos related disease SMRs in males were confined to the most and least deprived districts.
- Significant interactions were seen between temperature and rainfall in both sexes, with increasing independent COPD risks seen with increasing rainfall in the coldest districts (those in the lowest minimum temperature quintile).

Descriptive analyses

This section gives descriptive statistics for the variables used in the shared component and regression analyses. COPD and lung cancer mortality are examined first – these will be used in the shared component analyses. Next the distribution and geographical distributions of the district-level explanatory variables for the regression analyses are considered – pneumoconiosis and asbestos-related disease mortality, deprivation, air pollution, meteorological variables and fruit & vegetable purchases. District-level maps for all these variables can be found in the Appendix.

COPD and lung cancer mortality

Counts

There was a wide range in counts for COPD and lung cancer mortality in ages 45+ years (Table 7-1) with the smallest number of deaths and population in the Isles of Scilly and the City of London. All counts were skewed to the right (mean>median). The largest districts in terms of population were Birmingham with >6m population years represented and Leeds and Glasgow with >4m population years (aged >45 years). Not surprisingly, Leeds and Glasgow also had the highest numbers of deaths – Leeds the highest number of male COPD deaths (6,773) and Glasgow the highest numbers of female COPD deaths (4,395) and lung cancer deaths (10,357 males and 5,961 females).

Table 7-1 Summed counts for COPD and lung cancer deaths and population for ages 45+ years by district for Great Britain 1981-1999

	Total	Minimum	Maximum	Mean	Median	Standard Deviation
Males						
COPD	361,194	5	6773	786.9	577	717.6
Lung cancer	483,637	12	10,357	1,053.7	748	1011.0
Population	183,049,913	8140	3,090,359	398,802	332,344	292,087
Females						
COPD	211,695	4	4,395	461.2	322	461.2
Lung cancer	220,865	6	5961	481.2	341	505.3
Population	219,346,937	8759	3,662,107	477,880	390,493	360,432

SMRs

There was an approximately seven-fold range in COPD SMRs (Table 7-2, page 197). However, most of the extreme SMRs were based on small numbers – the two SMRs less than 33.3 occurred in the Isles of Scilly with 5 COPD male deaths and City of London with 4 COPD female deaths. There was only a 2.5 fold difference between COPD SMRs in the 5th and 95th percentiles (Table 7-2, page 197). There was a smaller

variation in lung cancer SMRs with a 3-fold (males) to 4-fold difference in range, and 2-fold (males) to 3-fold (females) difference between 5th and 95th percentiles.

Table 7-2 Descriptive statistics for COPD and lung cancer SMRs

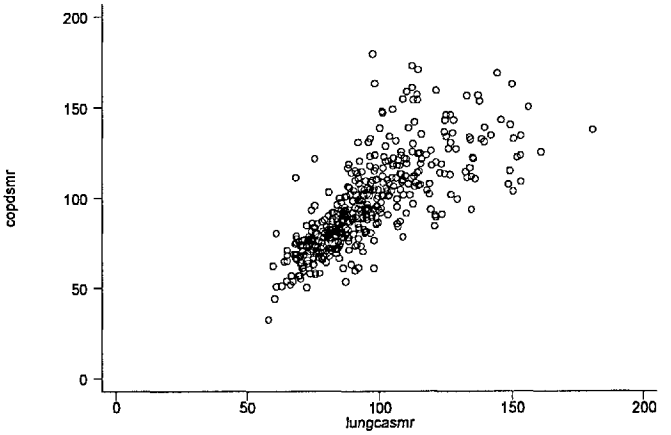
	Minimum	Maximum	Mean	Median	SD	5th percentile	95th percentile
Males							
COPD SMRs	32.6	179.6	95.3	90.9	25.4	61.0	143.3
COPD SMRs where counts>10	44.0	179.6	95.5	91.0	25.3	61.0	143.3
Lung cancer SMRS	58.0	180.7	94.8	91.6	20.5	68.9	135.5
Lung cancer SMRs where counts>10	58.0	180.7	94.8	91.6	20.5	68.9	135.5
Females							
COPD SMRs	30.9	184.9	94.2	88.9	28.1	59.1	148.9
COPD SMRs where counts>10	36.3	184.9	94.3	88.9	28.0	59.2	148.9
Lung cancer SMRS	53.4	202.7	94.0	89.2	23.9	64.5	144.6
Lung cancer SMRs where counts>10	53.4	202.7	94.1	89.4	23.9	64.5	144.6

Relationship between COPD and lung cancer SMRs

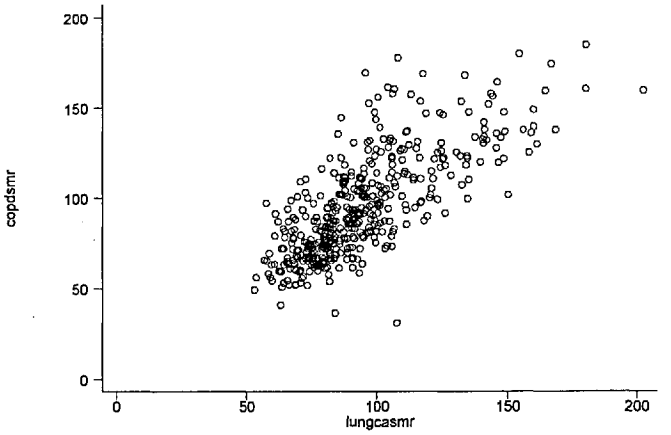
COPD and lung cancer SMRs were highly correlated ($r = 0.76$ for males, $r = 0.75$ for females, both p-values <0.0001), as seen in scatterplots (Figure 7-1, page 198).

Figure 7-1 Scatterplots for COPD and lung cancer SMRs for districts in Great Britain 1981-1999

Males



Females

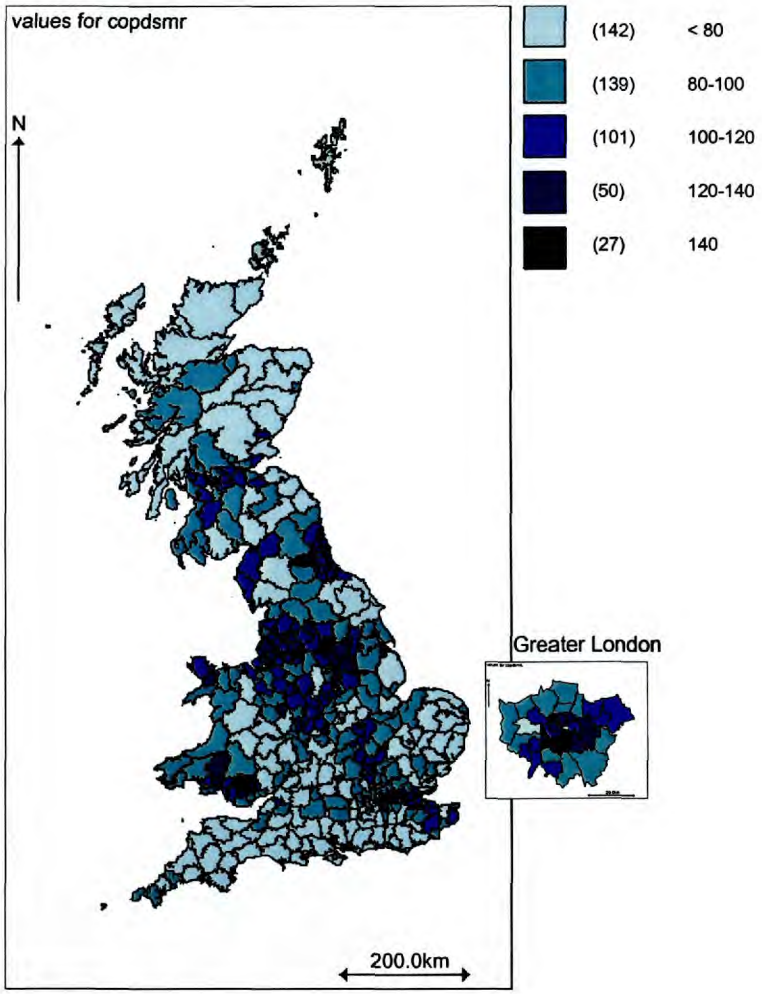


Geographical distribution of COPD and lung cancer SMRs

COPD (Figure 7-2, page 199) and lung cancer (Figure 7-3, page 200) SMRs showed similar patterns for males and females), with higher SMRs seen in conurbation areas of central and south-western Scotland (covering Edinburgh and Glasgow and the Clyde), north-eastern England (including Tyneside), north-western and north-central England (covering the Manchester, Merseyside and south Yorkshire conurbations), west Midlands (Birmingham and surrounds), south Wales, London (central, south-east and north-east) and the Thames estuary. There were slightly more districts with high SMRs for females in Scotland and northern England, while males had more in the west Midlands area of England.

Figure 7-2 Map of COPD SMRs for UK districts 1981-1999

Males



Females

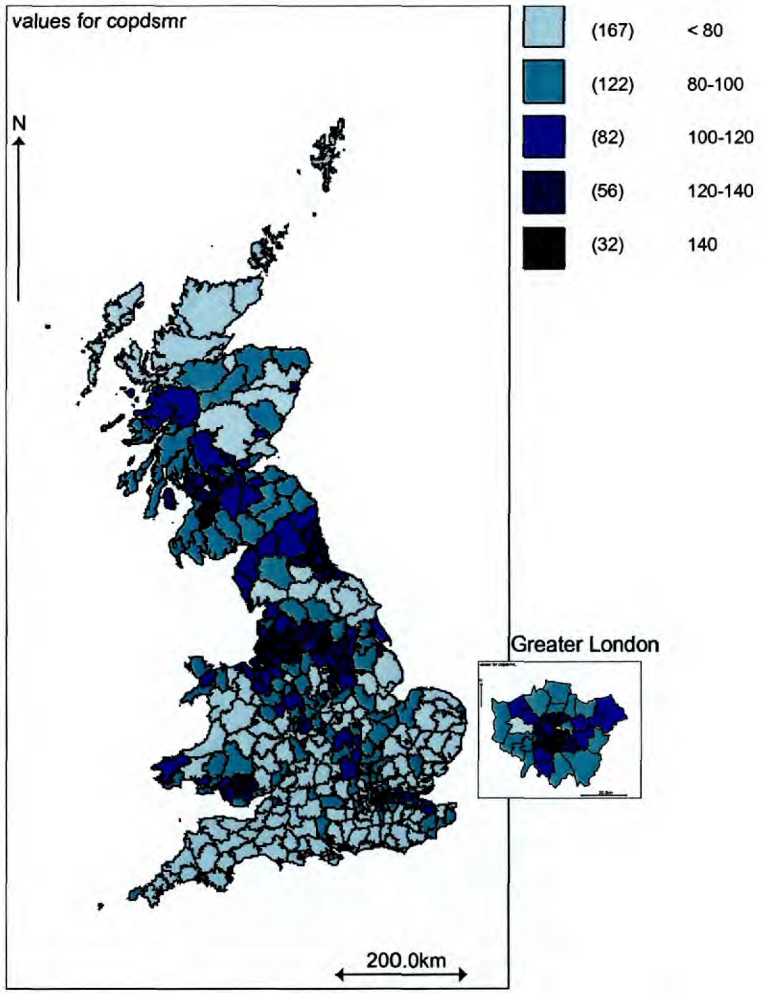
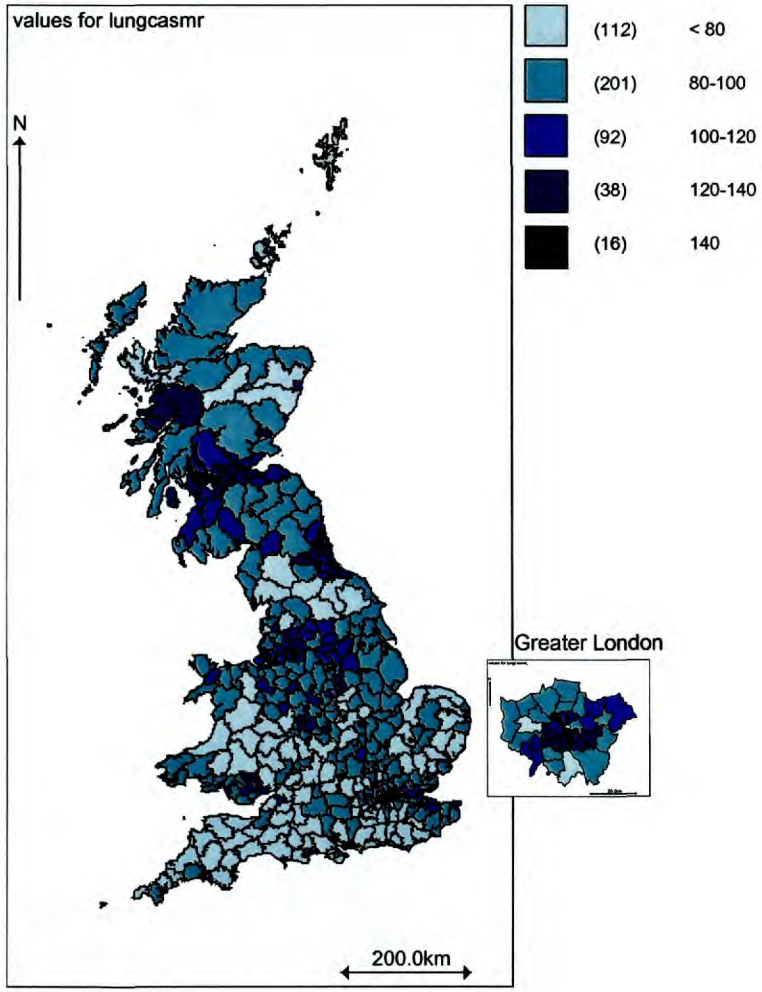
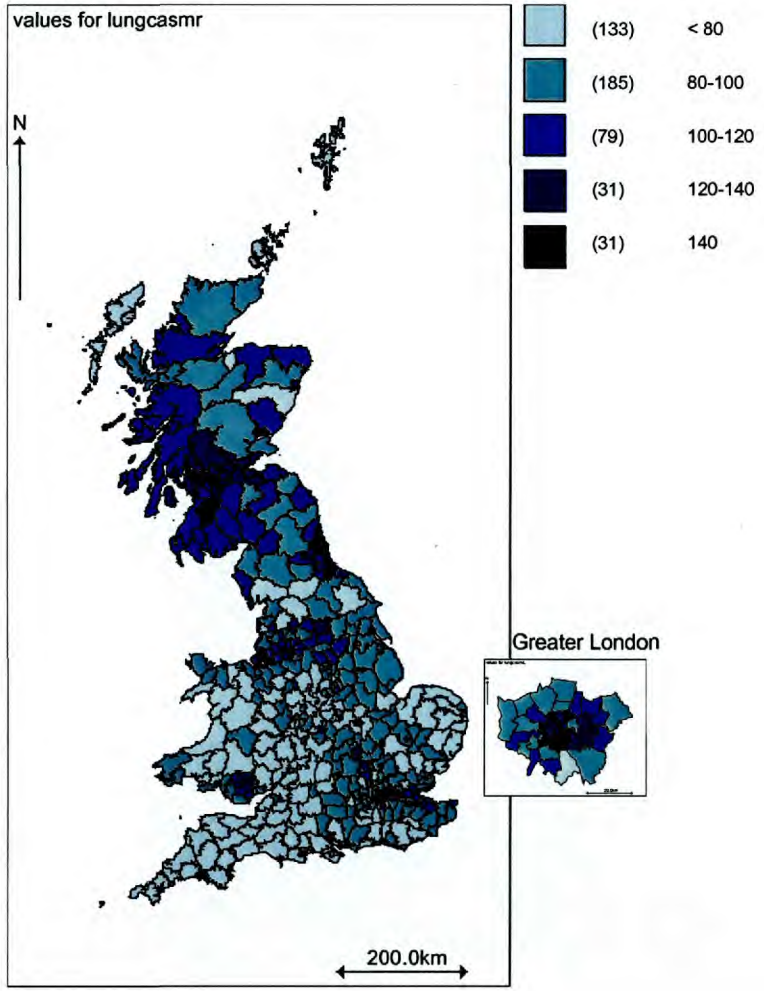


Figure 7-3 Map of Lung cancer SMRs for UK districts 1981-1999

Males



Females



Deprivation

Geographical location

Districts with higher levels of deprivation (the highest percentage of the population living in the most deprived wards – Carstairs quintile 5) were generally similar identified as having high COPD, lung cancer, pneumoconiosis and/or asbestos-related disease SMRs. These districts were found in Greater London, South Wales, West Midlands, Merseyside, Tyneside, central Scotland (including Edinburgh, Glasgow and Clydeside) and in some parts of northern Scotland (see map in Appendix, Figure A-6, page 358).

Relationship of deprivation with COPD SMRs

As expected, there was a significant trend for increasing COPD SMR with increasing deprivation in both males and females, shown in Table 7-3.

Table 7-3 Relationship of district level deprivation with COPD SMRs for 1981-1999 in Great Britain

Deprivation variable grouping	Observed	Expected	SMR
Males			
1 (least deprived)	58138	76463	76.0
2	49552	57430	86.3
3	53622	57926	92.6
4	89131	82556	108.0
5 (most deprived)	110751	86816	127.6
			χ^2 for trend 9953.72 (1 df, p=0.000)
Females			
1 (least deprived)	31257	42856	72.9
2	25696	31288	82.1
3	31931	34913	91.5
4	52931	49259	107.5
5 (most deprived)	69880	53381	130.9
			χ^2 for trend 7480.12 (1 df, p=0.000)

Deprivation variable key:
1 = 0% of population in district living in wards in Carstairs quintile 5 (most deprived quintile)
2, 3, 4, 5 = quartiles of percentage of population in districts living in wards in Carstairs quintile 5, where percentage is >0%

Ambient air pollution

Ranges

District annual averages of PM10 ranged from 12.72 $\mu g/m^3$ to 28.27 $\mu g/m^3$, while SO_2 ranged from 0 to 11.4 ppb. For the purposes of the later regression analyses, quintiles of these variables were created. Associated ranges are shown in Table 7-4, page 202.

Table 7-4 PM_{10} and SO_2 ranges per quintile

Pollutant	Quintile				
	1	2	3	4	5
PM_{10} ($\mu g/m^3$)	12.72 – 17.14	17.15 - 19	19.004 – 20.46	20.49 – 22.53	22.54 – 28.27
SO_2 (ppb)	0 – 1.30	1.37 – 2.12	2.13 – 3.001028	3.001095 – 3.94	3.94 – 11.38

District PM_{10} levels

Interpolated annual PM_{10} levels in Great Britain for 1996 showed a pronounced east-west gradient, with highest levels in the south-east of England (see map in Appendix, Figure A-7, page 359). PM_{10} levels are composed of the combination of local sources (generally highest in conurbation areas) and fine particulates coming in to the country across the Channel from Europe. Consultation with AEAT, who conducted the interpolation, suggested that 1996 was an unusual year in that annual levels were dominated by a small number of exceptionally high influxes of particulates from the continent and could not be held to be representative of years either side (personal communication, Professor David Briggs, Professor of Environmental Health Sciences, Imperial College). It was therefore decided to omit the 1996 district-level PM_{10} data from further analyses. Unfortunately, there were no other years of interpolated data for Great Britain within the 1981-99 analysis period.

District SO_2 levels

Interpolated annual SO_2 levels in Great Britain for 1996 showed the expected patterns (Appendix, Figure A-8, page 360) with highest levels in conurbation areas such as Greater London, south Wales, west Midlands, Merseyside, Manchester, south and west Yorkshire, Tyneside and central Scotland. Lowest levels were seen in south-west England, central and northern Wales, the Borders and the Highlands and islands of Scotland.

Meteorological variables – minimum temperature and rain

Descriptive statistics

Meteorological data were available for a total of 457 districts (out of a possible 459). The missing districts were the Isles of Scilly (16FD) and Shetland (6655). The range in average daily minimum temperature for districts was 3.5 to 8.8 °C and the range for average daily rainfall was from 1.6 to 3.7 mm (Table 7-5, page 203).

Table 7-5 Descriptive statistics for average daily minimum district temperature and district rainfall 1981-1999

Variable	Minimum	Maximum	Mean	Median	Standard Deviation
Minimum temperature (°C)	3.5	8.8	6.5	6.4	0.91
Daily rainfall (mm)	1.6	3.7	2.3	2.2	0.42

There were 22 districts where the average minimum temperature was above the 95th centile (8.0°C) of which 11 were in Wales, nine were in the south-west region and two were in West Midlands region. There were 22 districts where the average minimum temperature was below the 5th centile (4.9°C) of which five were in Northern region and the remainder were in Scotland.

There were 22 districts where the average minimum rainfall was above the 95th centile (3.5mm) of which 21 were in Scotland and one was in Wales (Llanelli – the area with the highest average rainfall in the UK). There were 22 districts where the average minimum rainfall was below the 5th centile (1.75mm) and these were in six of the 10 regions: Yorkshire & Humberside (one district), East Midlands (five districts), East Anglia (three districts), South-east (one district), South-West (five districts), Wales (seven districts).

Geographical patterns for temperature and rainfall

Maps suggested a south-west to north-east gradient in temperature, with the warmest areas for minimum temperature being in the south-western areas of Great Britain (Appendix, Figure A-9, page 361). Rainfall showed a west-east gradient, with the wettest areas in south Wales and west and central Scotland (Appendix, Figure A-10, page 362).

Pneumoconiosis and asbestos-related disease mortality

Counts

There were many more deaths in males than females, particularly for pneumoconioses (Table 7-6, page 204). There were only a total of 232 pneumoconiosis deaths in ages 45+ years in females over 1981-99. The small numbers particularly in younger age-groups, may have made SMRs unreliable in these age-groups due to uncertainty in the calculated expected counts.

Table 7-6 Counts for pneumoconiosis and asbestos-related disease deaths and population for ages 45+ years by district for Great Britain 1981-1999

	Total	Minimum district	Maximum district	Mean	Median	Standard Deviation
Males						
Asbestos-related deaths	9,492	0	259	20.7	12	28.9
Pneumoconioses	5,660	0	391	12.3	2	32.2
Population	183,049,913	8140	3,090,359	398,802	332,344	292,087
Females						
Asbestos-related deaths	2,099	0	99	4.6	3	7.2
Pneumoconioses	232	0	140	0.5	0	6.6
Population	219,346,937	8759	3,662,107	477,880	390,493	360,432

SMRs

For males, 60 districts had pneumoconiosis SMRs statistically significantly greater than 100. These were located in Lothian (around Edinburgh), north-east England (around Tyneside), north-west England (around Merseyside), central England (South and West Yorkshire, Derbyshire), South Wales and north-west Wales (Appendix, Figure A-11, page 363). These districts were similar to those identified by Coggon [248] as areas with concentrations of deaths in miners from coal-worker's pneumoconioses in 1979-80 and 1982-90. There were a number of extreme SMRs (SMR>1000) for pneumoconioses in males, particularly in Wales (Table 7-7, page 205). For females, only five districts had pneumoconiosis SMRs statistically significantly greater than 100 (using 99% confidence intervals) and these were in the north-west and west Midlands area (Appendix, Figure A-12, page 364). Two of these districts were different from those for males. The most noticeable single outlier was 140 pneumoconiosis deaths in women aged 45+ in 1981-1999 in Oldham giving a massive SMR of 16,123. Deaths were spread across wards – 18 of 21 wards in Oldham had one or more of such deaths – with most of the deaths occurring in the 1980s.

For asbestos-related diseases, there were 48 districts with SMRs statistically significantly greater than 100 for males and 13 such districts for females, five of which were different from those identified for males. There were some differences in geographical distribution from those for pneumoconioses, with the high SMRs tending to be in areas previously associated with ship-building areas – including Fife, Clydeside, Tyneside, Sunderland, Barrow-in-Furness, Merseyside, the Thames, Plymouth and Portsmouth (Appendix, Figure A-13, page 365). Districts in central London were also identified. The highest asbestos-related mortality SMR of 1030

(Table 7-7 below) was seen in males in Barrow-in-Furness. For females (Appendix, Figure A-14, page 366), higher SMRs were seen in central London, Clydeside, Tyneside, and north-central England (South Yorkshire, Lancashire, Nottinghamshire).

Table 7-7 Extreme (>1000) district-level SMRs for pneumoconioses and asbestos-related deaths

Region name	Region number	District code	District-name	Actual Number of deaths	Expected number of deaths	SMR
Pneumoconioses – males						
North	1	21GS	Derwentside	92	8.9	1033.5
Yorkshire & Humberside	2	05CC	Barnsley	391	22.0	1780.9
East Midlands	3	18FM	Bolsover	97	7.8	1236.4
Wales	9	49SN	Dinefwr	86	4.9	1762.9
Wales	9	51SZ	Arfon	116	5.2	2231.6
Wales	9	52TD	Cynon Valley	181	6.6	2729.9
Wales	9	52TE	Merthyr Tydfil	99	5.5	1788.8
Wales	9	52TF	Ogwr	140	13.5	1039.0
Wales	9	52TG	Rhondda	199	8.2	2421.1
Wales	9	52TH	Rhymney Valley	157	8.4	1870.4
Wales	9	52TJ	Taff-Ely	88	7.4	1187.7
Pneumoconioses – females						
West Midlands	7	42QP	Stoke-on-Trent	14	1.0	1374.5
North-west	8	03BP	Oldham	140	0.9	16123.3
North-west	8	03BQ	Rochdale	20	0.8	2581.5
Asbestos-related deaths – males						
North	1	17FF	Barrow-in-Furness	131	12.7	1030.2
Asbestos-related deaths – females						
-	-	-	-	0	-	-

SMR groupings used in regression analyses by district

For the purposes of the regression analyses, SMRs for pneumoconiosis and asbestos-related diseases were grouped into the following groupings: not significantly different from 100 using exact 99% confidence intervals (the reference group), statistically significantly less than 100, and statistically significantly greater than 100. The number of districts in each of the three groups showed that there were good numbers in the reference group (Table 7-8, page 206), but for females there were very few districts that had SMRs different from baseline.

Table 7-8 Number of districts in SMR groupings for pneumoconioses and asbestos-related disease mortality 1981-99 in ages 45+ years

	Coding	Key	Number of districts	% of total (n=459)
Males				
Pneumoconioses	1	Non-significantly different from 100	164	35.7
Pneumoconioses	2	Statistically significantly <100	235	51.2
Pneumoconioses	3	Statistically significantly >100	60	13.1
Asbestos-related deaths	1	Non-significantly different from 100	327	71.2
Asbestos-related deaths	2	Statistically significantly <100	84	18.3
Asbestos-related deaths	3	Statistically significantly >100	48	10.5
Females				
Pneumoconioses	1	Non-significantly different from 100	454	98.9
Pneumoconioses	2	Statistically significantly <100	0	0
Pneumoconioses	3	Statistically significantly >100	5	1.1
Asbestos-related deaths	1	Non-significantly different from 100	444	96.7
Asbestos-related deaths	2	Statistically significantly <100	2	0.4
Asbestos-related deaths	3	Statistically significantly >100	13	2.8

Average district fruit & vegetable purchases from the TNS Super Panel 1991-2000

Descriptive statistics

Data related to 453 districts out of a possible 459 as described in the methods. The median number of person weeks of purchases per district in the dataset used for these analyses was 14,218 (Table 7-9, page 207). The median of the weighted number of households per district was 29 (range 1 to 298) and 83 (18.4%) of 451 districts had less than 10 (weighted) households per district. Four districts had only one household in the study – three of these were in Scotland. Fruit and vegetable purchases were normally distributed by district (not shown), with a median of 2.4 portions per person per day (range 0.7 to 5.9). There were two districts with >5 portions per day and three districts with <1 portion per day, but all of these outlying districts had small numbers of households taking part.

Table 7-9 Descriptive statistics for district-level TNS dataset variables

	Mean	Median	Min	Max	10th percentile	25th percentile	90th percentile
Number of weeks in study	146	63	1	522	4	15	479
Weighted no. households per district	38.8	30.1	1	298	7.5	13.6	77
Average no. person-weeks per district	18,821	14,218	48	145,718	1734	3662	39,355
Average district purchases of fruit/vegetable (inc. fruit juice) in g per person per week	1460	1440	475	3637	1087	1257	1820
Average district purchases of portions fruit/vegetables per day (portion = 80g fruit/vegetables or 125g fruit juice)	2.5	2.4	0.7	5.9	1.8	2.1	3.1

Geographical distribution

Fruit and vegetable purchases showed some variability across Great Britain, but the general impression was of highest purchases in districts in the southern part of England and lowest purchase in districts in the north-west of England and in Scotland (Appendix, Figure A-15, page 367).

Shared-component analyses for COPD and lung cancer mortality

This section contains the results for the Bayesian model that partitions the district-level COPD and lung cancer risks into a shared component and two disease-specific components. As both diseases are closely associated with cumulative smoking, the shared component is assumed to capture most of the smoking attributable risk.

First the proportion of variances explained by shared and disease-specific components, then the proportion of variances explained by spatially structured and non-spatially structured elements is given. To check whether the shared component has indeed removed most of the association with lung cancer, the relationship between COPD SMRs and the independent COPD risk with lung cancer SMRs is explored graphically and through correlation coefficients.

Next, the maps are used to demonstrate the geographical variation of the independent COPD risk factor for males and females. As it is difficult to indicate statistical probability on these maps (the extra information hinders clarity), they are shown with maps showing the statistical probability that the independent COPD risk factor is greater or less than one (since it is measured on a relative risk scale). Finally, the shared

(COPD and lung cancer) and independent lung cancer risk maps are examined and compared with those for the independent COPD risks.

Proportion of variances that are explained by shared and disease-specific components and their spatial structuring

The shared component analyses suggested that 64% of the variance in smoothed COPD relative risks and 70% of variance in smoothed lung cancer relative risks was accounted for by the shared component in both males and females (Table 7-10). In fact, there was a striking similarity in these proportions between males and females. This corresponds to 36% of COPD relative risk being disease-specific and 30% of the lung cancer relative risk being disease-specific.

The proportion of the variance of both the shared and disease-specific components that were spatially structured was approximately 60% for males (Table 7-10). For females, approximately 70% of the variance of the shared component was spatially structured compared with approximately 60% of the disease-specific components (Table 7-10).

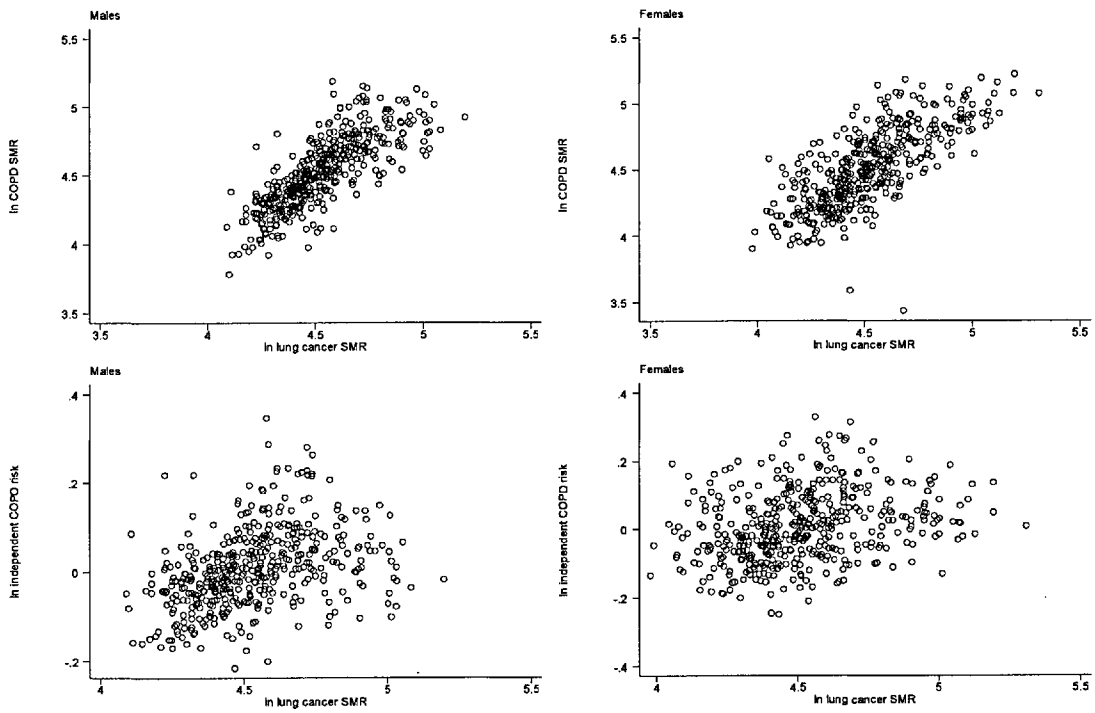
Table 7-10 Proportion of variance explained by shared component or spatially structured shared or disease-specific component: mean and 95% credible intervals from 100,000 samples

	Proportion of variance of:	that is:	mean proportion	sd	2.5%	97.5%
Male						
	COPD	explained by shared component	0.642	0.0551	0.534	0.726
	lung cancer	explained by shared component	0.702	0.0622	0.575	0.800
	shared component (θ)	spatially structured	0.634	0.0599	0.511	0.738
	disease-specific component (ϕ) for COPD	spatially structured	0.620	0.0712	0.489	0.752
	disease-specific component (ϕ) for lung cancer	spatially structured	0.609	0.0853	0.456	0.773
Female						
	COPD	explained by shared component	0.643	0.0594	0.523	0.745
	lung cancer	explained by shared component	0.702	0.0569	0.589	0.797
	shared component (θ)	spatially structured	0.710	0.0543	0.598	0.808
	disease-specific component (ϕ) for COPD	spatially structured	0.633	0.0708	0.498	0.764
	disease-specific component (ϕ) for lung cancer	spatially structured	0.615	0.0700	0.479	0.739

Residual association of log independent COPD risk (COPDspec) from shared component analyses with lung cancer SMRs

Plots comparing log COPD SMRs and log independent COPD risks from the shared component analyses against log lung cancer SMRs (Figure 7-4) suggested that the shared component analysis had removed much of the association between COPD and lung cancer, but that some residual association remained. The (product moment) correlations between log COPD SMRs and log lung cancer SMRs were 0.79 (p=0.0000) for males and 0.74 (p=0.0000) for females. By comparison, correlations between COPDspec and log lung cancer SMRs were 0.38 (p=0.0000) for males and 0.27 (p=0.0000) for females.

Figure 7-4 Plots of log COPD SMRs against log lung cancer SMRs (above) and of COPDspec, the log independent COPD risk, against log lung cancer SMRs (below – note different x and y scales) for males and females



Spatial pattern of smoothed relative risks for COPD and lung cancer mortality

As part of the process of modelling shared and independent risk factors, the model produces a combined (shared plus independent) relative risk for COPD and lung cancer

– relative being relative to the national average – smoothed towards that of the nearest neighbours. The geographical distribution of smoothed relative risks were generally similar to those for SMRs (shown in Appendix, Figures A-2 to A-5, pages 354-357). Smoothing tended to produce more higher relative risks (RRs) and fewer lower RRs (Table 7-11) except for male lung cancer in Scotland where relative risks were reduced. This extended outwards the higher risks seen in the conurbation areas described in the section on COPD and lung cancer SMRs. The biggest changes between SMR and relative risk maps were apparent in Scotland, but this was partly a visual impression because of larger district areas (related to sparser population).

Table 7-11 Descriptive statistics for SMRs, smoothed relative risks and independent and shared risks from shared component analyses for COPD and lung cancer mortality

	Mean	Median	Min	Max	10th centile	90th centile
Males						
COPD SMRs	95.6	91.1	44.0	179.6	67.0	131.1
COPD smoothed relative risks	103.3	98.5	55.4	191.7	72.8	141.2
Independent COPD risk	101.1	100.1	80.5	141.4	90.4	117.5
Lung cancer SMRs	94.9	91.6	59.8	180.7	72.4	124.4
Lung cancer smoothed relative risks	102.2	98.6	65.3	194.4	78.4	133.7
Independent lung cancer risk	100.8	98.7	79.6	137.9	91.8	113.6
Shared COPD and lung cancer risks	101.9	100.0	68.5	148.2	81.8	126.8
Females						
COPD SMRs	94.3	88.9	30.9	184.9	63.2	133.5
COPD smoothed relative risks	104.2	98.1	48.6	204.0	70.8	146.9
Independent COPD risk	101.3	100.3	77.9	138.6	88.8	115.3
Lung cancer SMRs	94.1	89.6	53.4	202.7	68.59	126.4
Lung cancer smoothed relative risks	102.9	97.6	62.0	221.1	76.2	138.0
Independent lung cancer risk	100.0	100.0	76.3	138.7	89.7	115.0
Shared COPD and lung cancer risks	102.4	98.9	66.7	171.1	80.7	171.1

Note that all risks in this table were multiplied by 100 to enable comparisons with SMRs

Independent COPD risks from shared component analyses

Posterior means for the independent COPD risks (based on 100,000 samples after convergence) showed less variability than for COPD SMRs and smoothed relative risks, with a less than two-fold variations between lowest and highest district risks compared with a four-six fold variation in COPD SMRs and smoothed relative risks (Table 7-11 above).

Independent COPD risks for males

Map of independent COPD risks

A map of independent COPD risks (Figure 7-5, page 213), not considering statistical significance, suggested higher risks in an enlargement of the areas in England & Wales highlighted in SMRs (Figure 7-2, page 199). This included many districts in northern and central England and a north-west to south-east axis down to Kent including Greater London roughly corresponding to the areas between and adjacent to the M40-M6 and M1 motorways. Much of South Wales and the north-west tip of north Wales were also included. However, only a small number of districts in Scotland were included, chiefly located to the south of Glasgow. Lower independent COPD risks were chiefly seen in East Anglia, south-western England and the highlands of Scotland.

Statistical significance of district independent COPD risks

Considering the statistical significance of independent COPD risks dramatically curtailed the extent of high and low COPD independent risks (Figure 7-5, page 213). A posterior probability of at least 0.8 was chosen, representing 80% confidence that the 'true' independent COPD risk was greater than 1, a common cut-off for Bayesian spatial analyses and found to give reasonable sensitivity and specificity in simulation studies.[307] Using this cut-off, statistically significant high risks (coloured yellow or red on the map) were seen in Tyneside, the north-central conurbation ring (containing the Liverpool, Manchester, Leeds, Sheffield and Birmingham conurbation areas), central London and south Wales.

Again using a probability >0.8 cut-off, statistically significant low risks were confirmed in East Anglia, the south coast of England, Devon. They were also seen in several districts in central Scotland including Edinburgh city (but not Glasgow) and in north-west Scotland.

Independent COPD risks for females

Map of independent COPD risks

The map of independent COPD risks not considering statistical significance (the map on the left in Figure 7-6, page 214), showed a very similar pattern of risks to that seen in males: i.e. many districts in northern and central England and a north-west to south-east axis down to Kent, much of South Wales, the north-west tip of north Wales and a small

number of districts in south-central Scotland. Again, lower independent COPD risks were chiefly seen in East Anglia, south-western England and the highlands of Scotland.

Statistical significance of district independent COPD risks

Considering the statistical significance of independent COPD again greatly curtailed the extent of high and low COPD independent risks. As with males (Figure 7-5, page 213), statistically significant high risks (coloured yellow or red on the maps in the centre and right of Figure 7-6, page 214) were seen in Tyneside, the north-central area (containing the Liverpool, Manchester, Leeds, Sheffield and Birmingham conurbation areas), and south Wales, but the pattern of districts affected in central London was different from that in males.

Females had higher numbers of districts with statistically significant low independent COPD risks than males (67 vs. 49). As with males (Figure 7-5, page 213), these were seen in East Anglia, the south coast of England and Devon (Figure 7-6, page 214). They were also seen in several districts near Glasgow (but not Glasgow or Edinburgh city) and in northern Scotland.

Figure 7-5 Map of posterior mean district independent COPD risks for males (left), map of the probability that posterior mean risk is greater than 1 (centre), map of the probability that posterior mean risk is less than 1 (right)

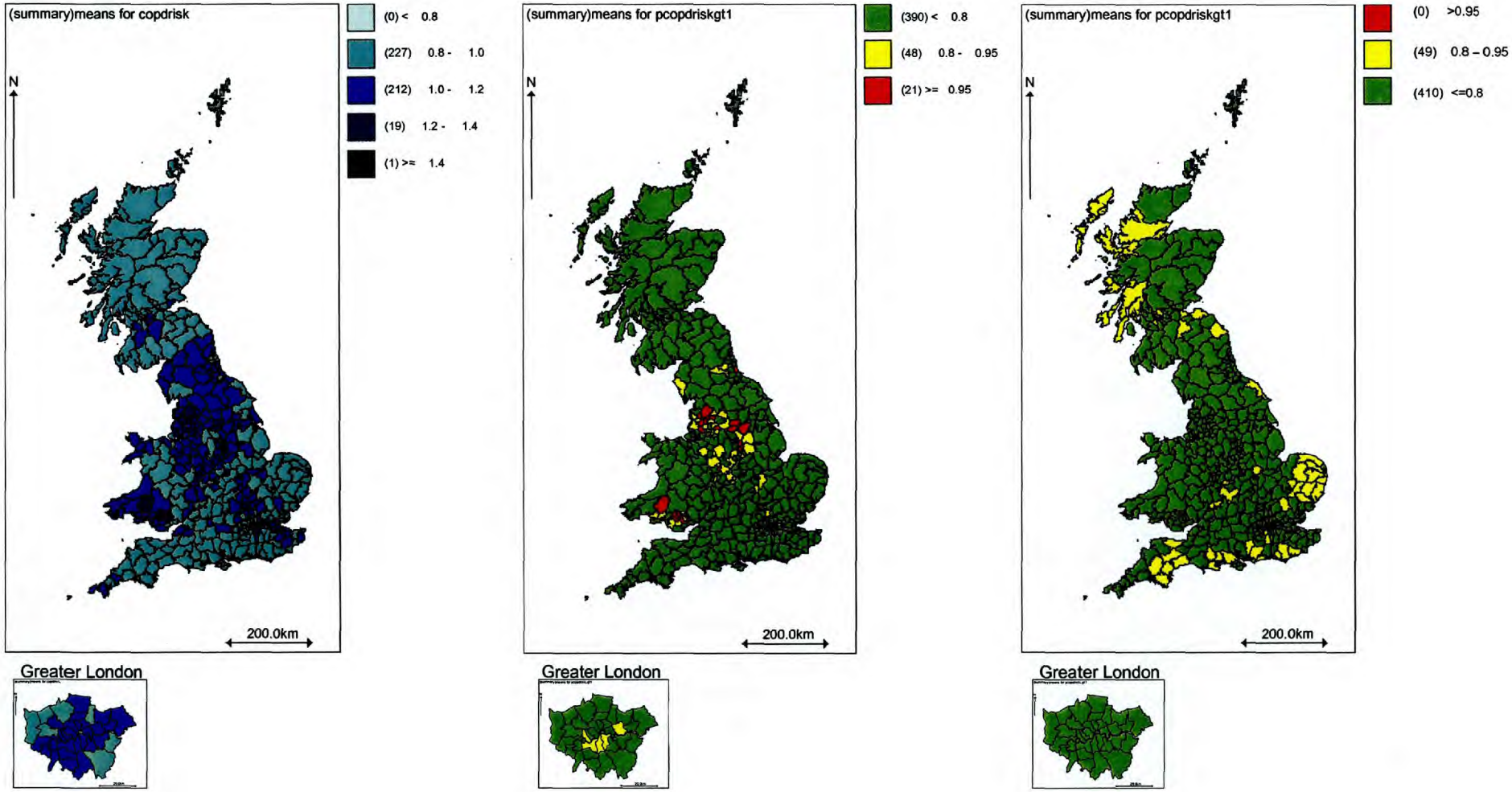
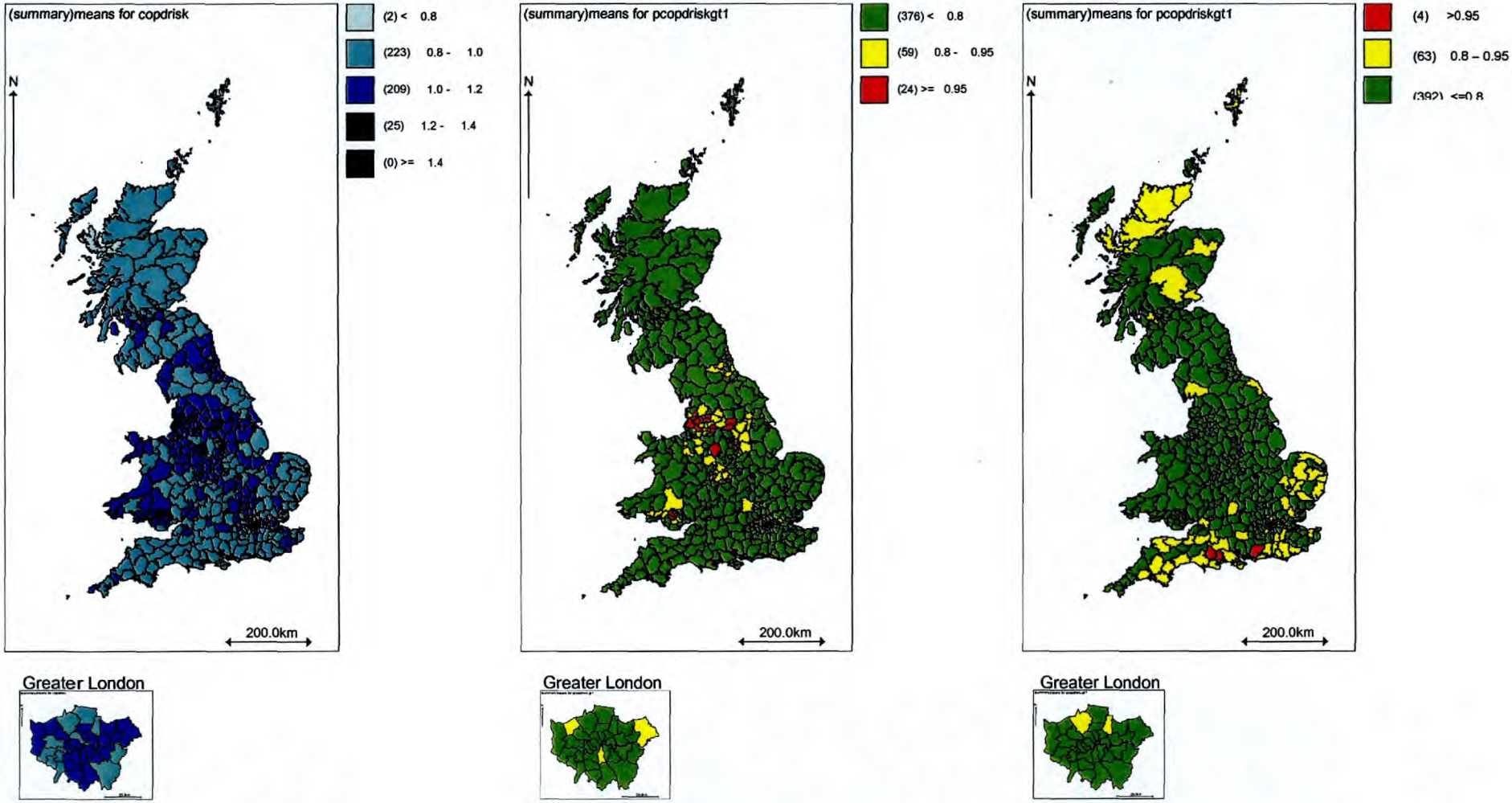


Figure 7-6 Map of posterior mean district independent COPD risks for females (left), map of the probability that posterior mean risk is greater than 1 (centre), map of the probability that posterior mean risk is less than 1 (right)



Shared (COPD and lung cancer) risks from shared component analyses

The geographical distribution of posterior means for the shared COPD and lung cancer risks (maps on the left in Figure 7-7, page 216 and Figure 7-8, page 217) showed some similarities with independent COPD risks (Figure 7-5, page 213 and Figure 7-6, page 214). Increased shared risks were seen in both males and females in Tyneside, the north-central conurbation ring (Liverpool, Manchester, Leeds, Sheffield and Birmingham conurbation areas), south Wales and central London. However, unlike the independent COPD risks, they were also increased in central Scotland and in females in southern Scotland. Risks also tended to be higher (shaded darker in the maps) for shared than independent risks. More of the districts with high (>1) posterior mean shared risks were had high probability that these risks were greater than one than seen for the independent COPD risks.

Maps of probabilities that posterior means were less than one (maps in the centre and right of Figure 7-7, page 216 and Figure 7-8, page 217) showed that many English districts in southern England and Wales (excluding Greater London and south Wales around Cardiff) had high probabilities (>=0.95) that shared risks were less than one. For males, a number of northern and southern Scotland districts had statistically significant low risks (i.e. risks <1.0), but this was not the case for females.

Figure 7-7 Map of posterior mean district shared COPD and lung cancer risks for males (left), map of the probability that posterior mean risk is greater than 1 (centre), map of the probability that posterior mean risk is less than 1 (right)

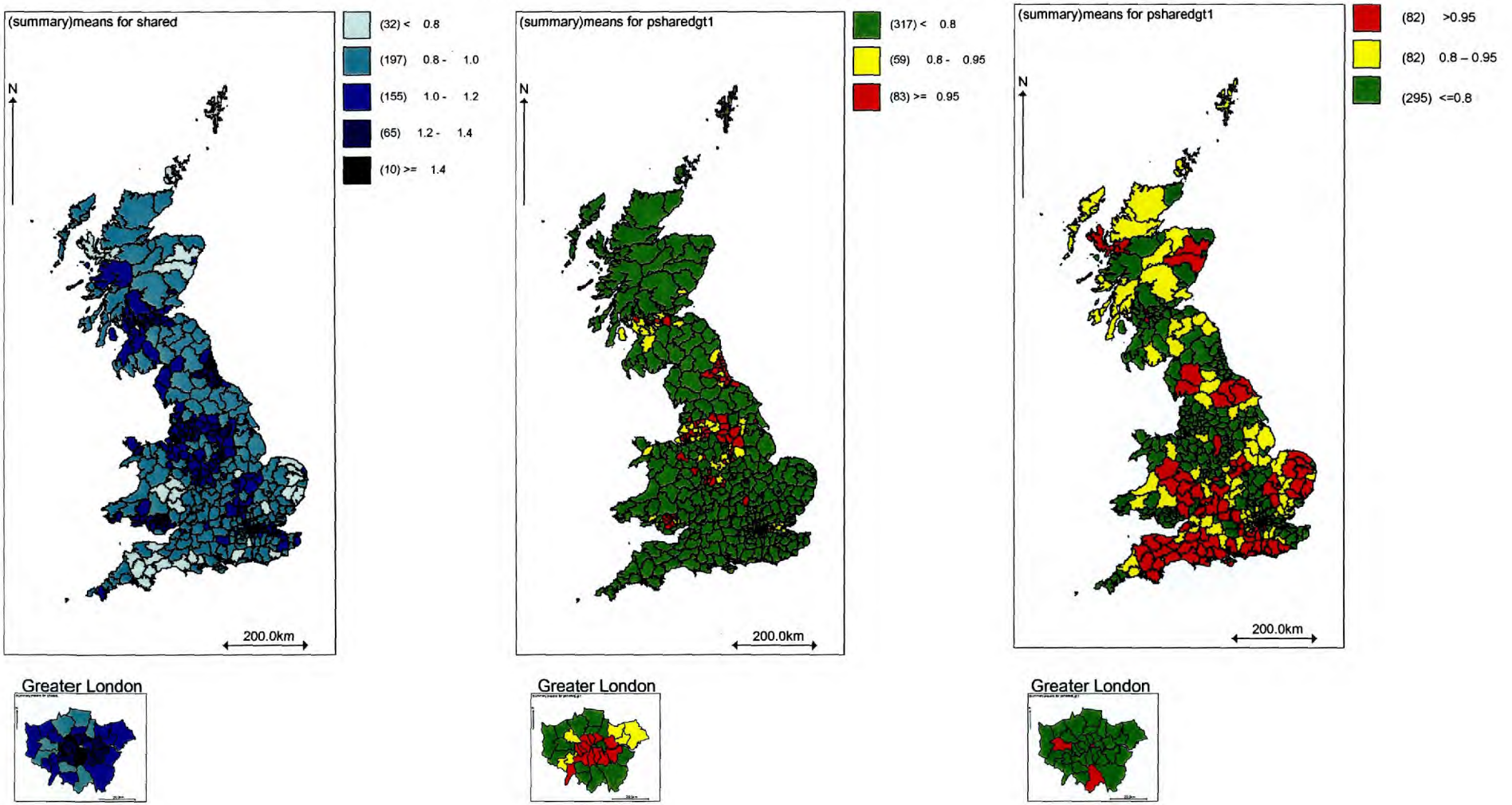
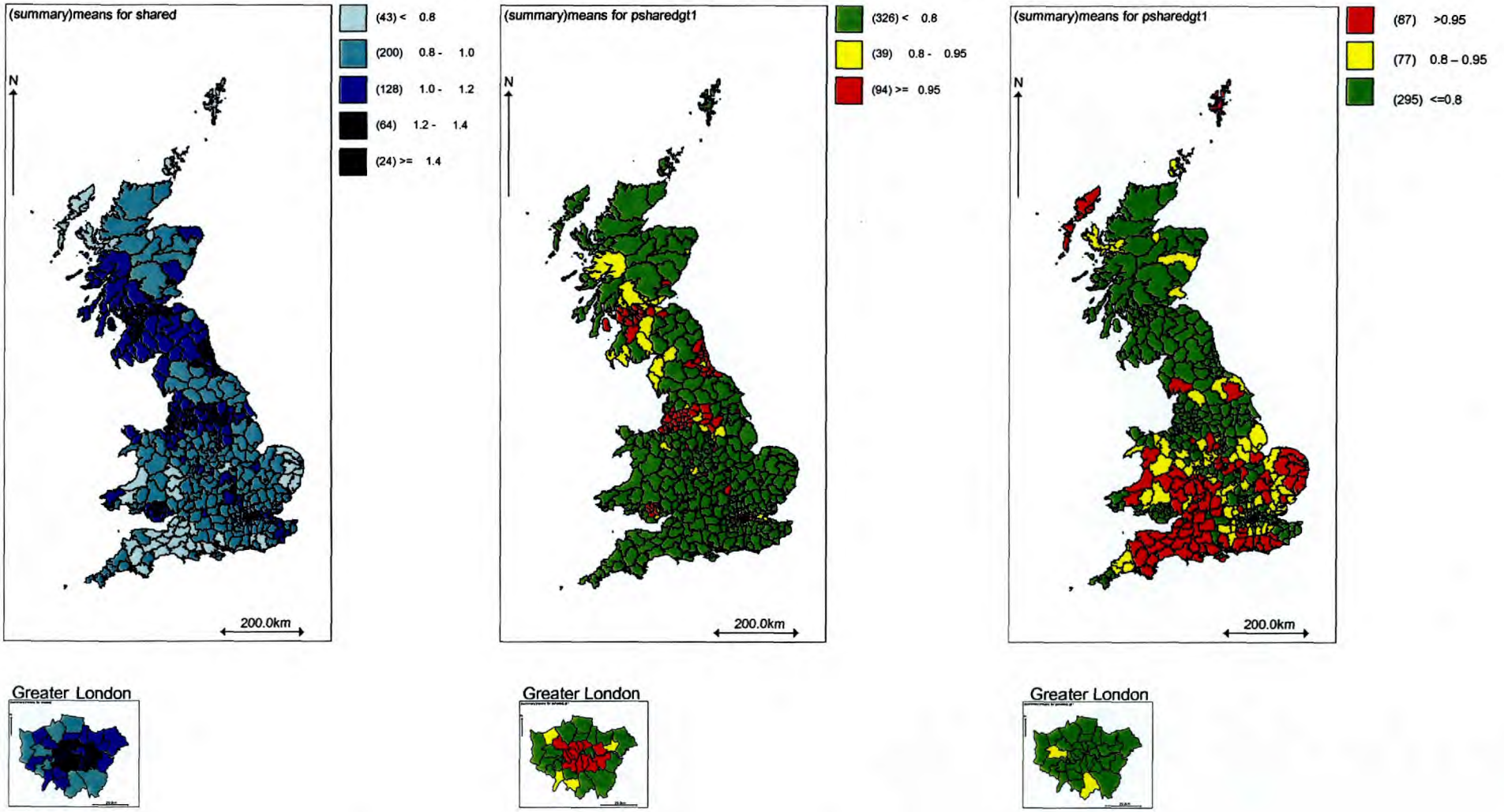


Figure 7-8 Map of posterior mean district shared COPD and lung cancer risks for females (left), map of the probability that posterior mean risk is greater than 1 (centre), map of the probability that posterior mean risk is less than 1 (right)



Independent lung cancer risks from shared component analyses

The geographical distribution of posterior means for the independent lung cancer risks (Figure 7-9, page 219 and Figure 7-10, page 220) were generally, but not exclusively (e.g. Greater London, Merseyside, Tyneside) different from those for the independent COPD risks (seen in Figure 7-5, page 213 and Figure 7-6, page 214). Increased independent risks were seen in both males and females in northern England and most of Scotland and also in the south-east corner of England (especially in females).

Many fewer districts had statistically significant posterior means for independent lung cancer risks that were greater than one (maps in the centre of Figure 7-9, page 219 and Figure 7-10, page 220). These were chiefly located in central Scotland, Tyneside, Merseyside and (especially females) central London. Statistically significant posterior means that were less than 1 (maps on the right of Figure 7-9, page 219 and Figure 7-10, page 220) were chiefly located in central England, non-conurbation districts in northern England and in south and mid Wales.

Figure 7-9 Map of posterior mean district independent lung cancer risks for males (left), map of the probability that posterior mean risk is greater than 1 (centre), map of the probability that posterior mean risk is less than 1 (right)

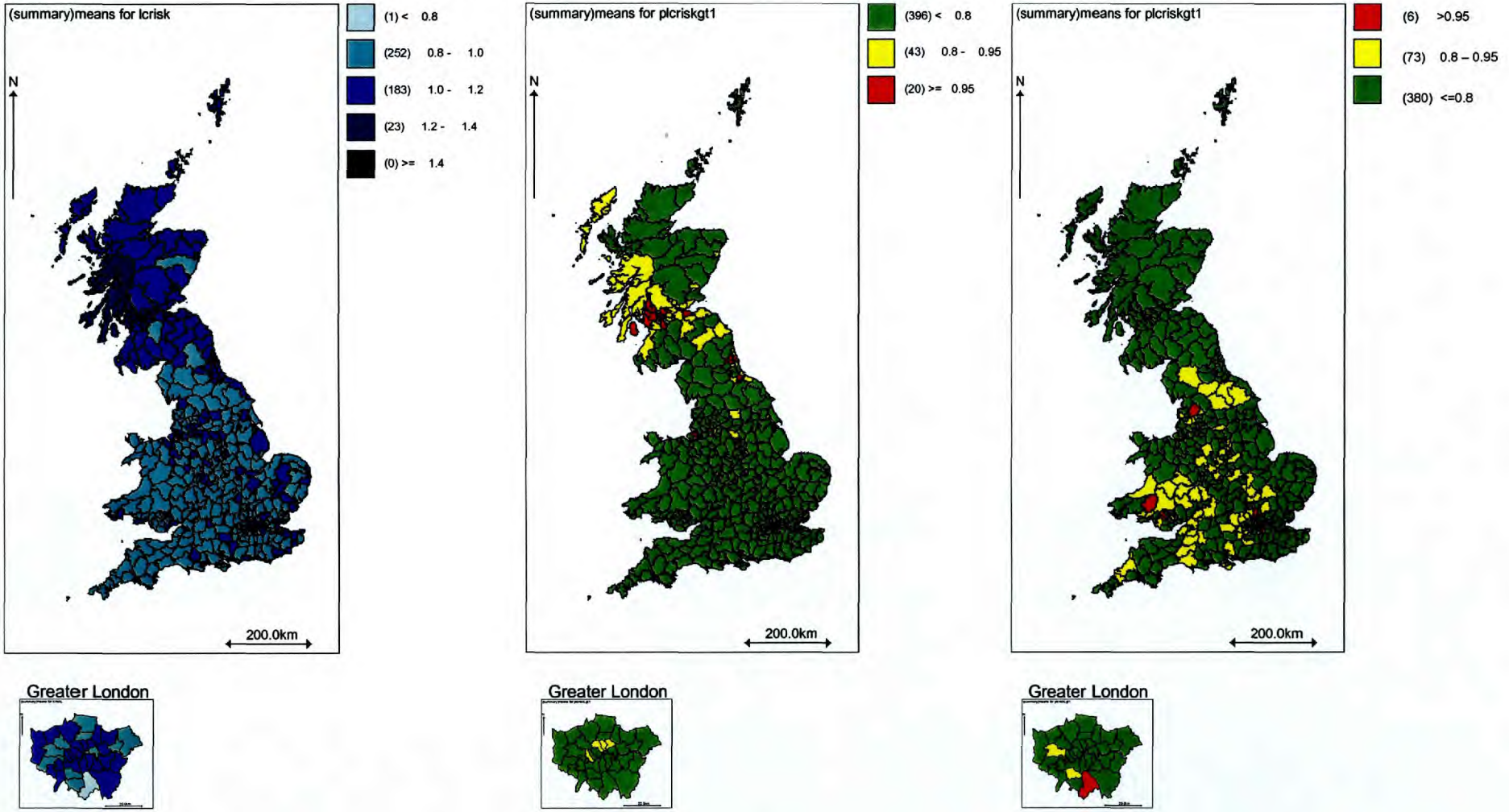
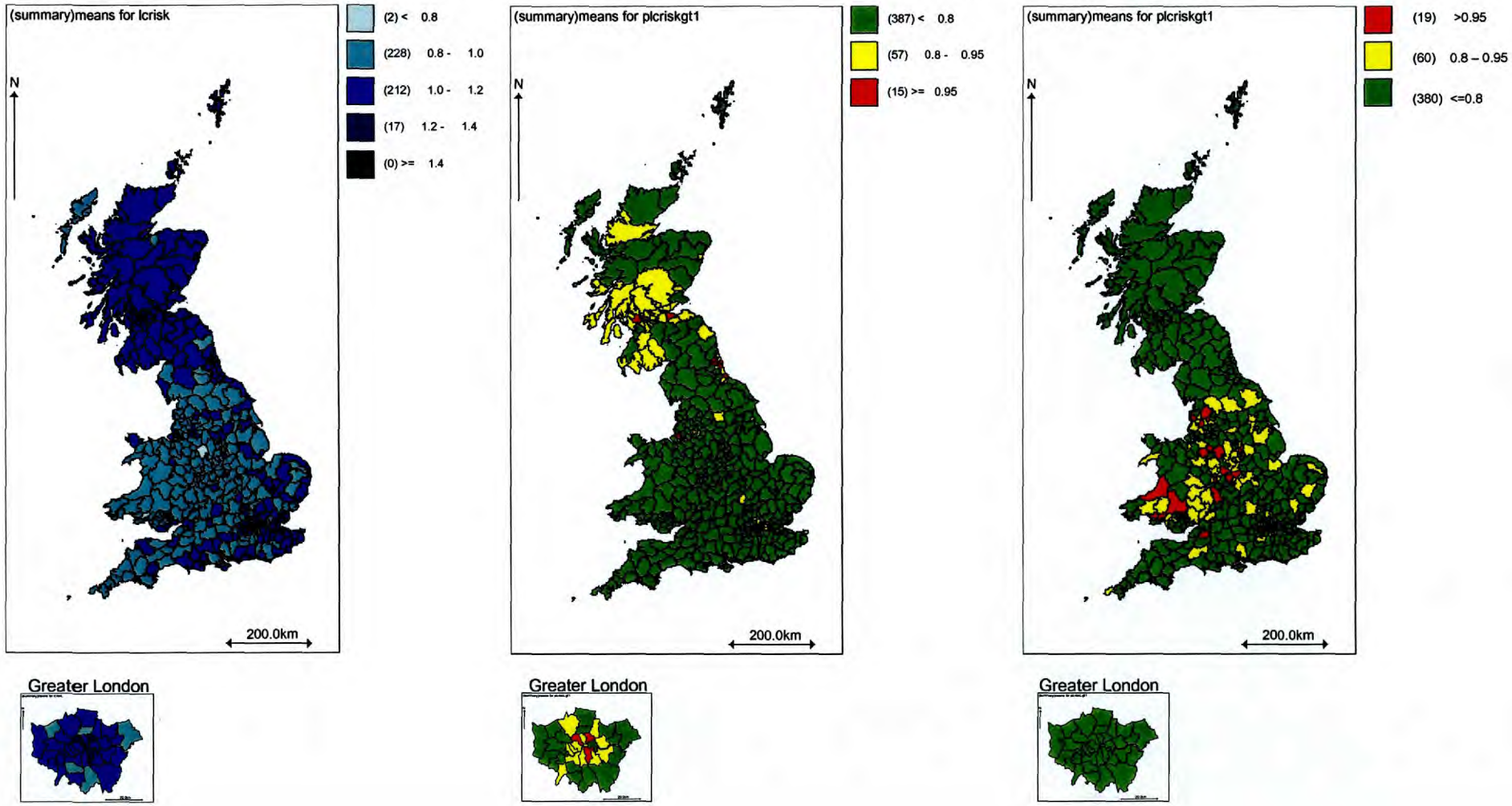


Figure 7-10 Map of posterior mean district independent lung cancer risks for females (left), map of the probability that posterior mean risk is greater than 1 (centre), map of the probability that posterior mean risk is less than 1 (right)



Linear regression

This section contains the results for the linear regression of the natural log disease-specific district-level COPD risk ϕ_i (COPDspec) obtained from the shared component analyses – against the seven potential explanatory variables. It was assumed that the partitioning process had removed shared risks with lung cancer of which the major component would be cigarette smoking and significant associations would reflect associations of district-level non-smoking risk factors for COPD mortality.

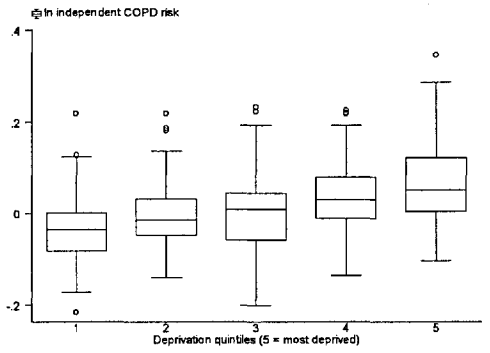
Firstly, the associations of COPDspec with the explanatory variables are explored through graphical displays and univariate linear regression. Next the results from multivariate linear regression are described. Finally, secondary analyses including sensitivity analyses and an examination for interaction with deprivation of the significant variables from the final multivariate model are given.

Descriptive plots of the association of explanatory variables with COPDspec

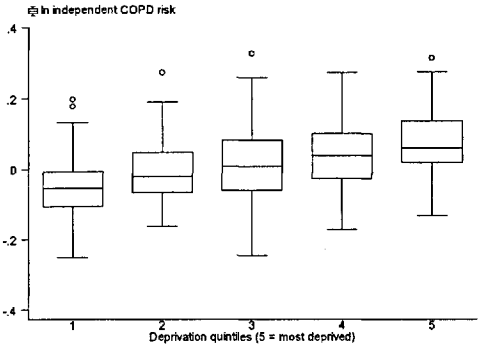
Descriptive box and whisker plots are displayed in Figure 7-11, pages 222-223. These suggested increases of COPDspec with increasing deprivation and increasing SO_2 levels in both males and females. Higher levels of COPDspec were also suggested with pneumoconiosis SMRs statistically significantly greater than 100 in both males and females and with asbestos-related disease SMRs statistically significantly greater than 100 in females (but not males). A possible decrease in COPDspec was suggested with increasing fruit and vegetable purchases in both sexes. There was no obvious relationship with average minimum temperature. There was some suggestion of an inverted U-shape relationship with average rainfall with higher levels in the middle quintile.

Figure 7-11 Box (showing interquartile range) and whisker plots* of the association of explanatory variables with the ln independent COPD risk

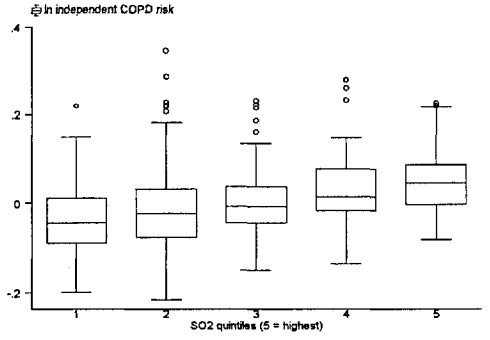
Deprivation – males



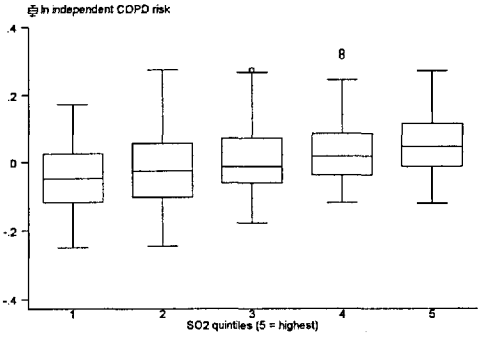
Deprivation – females



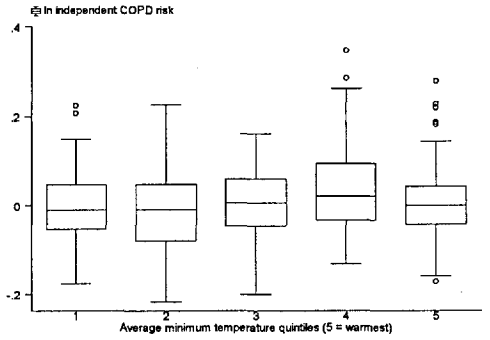
Sulphur dioxide – males



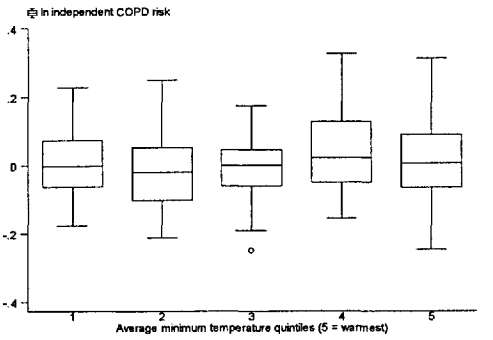
Sulphur dioxide – females



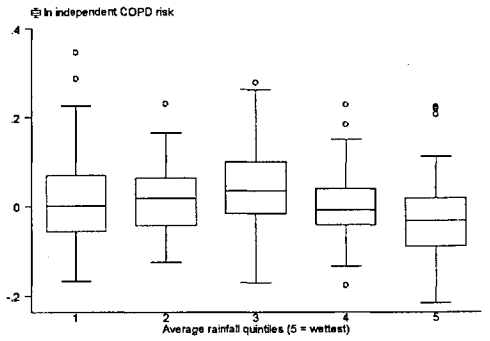
Minimum temperature – males



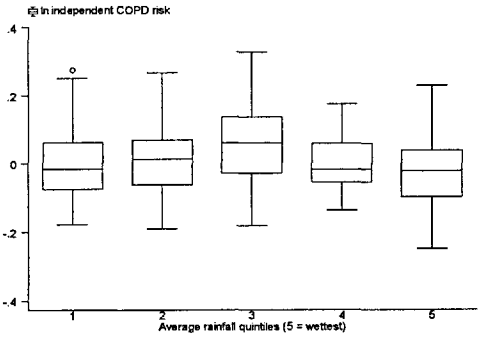
Minimum temperature – females



Rainfall – males



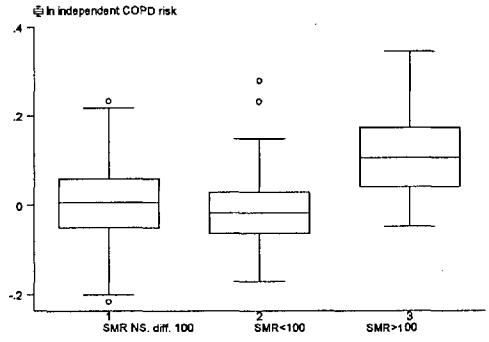
Rainfall - females



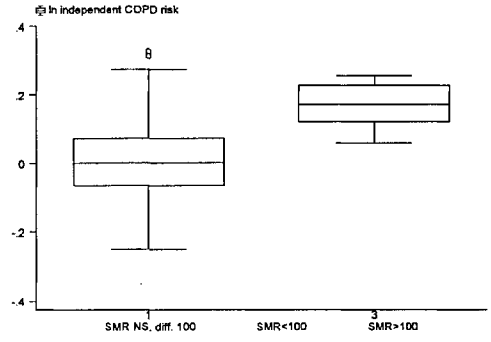
** The box extends from the 25th percentile to the 75th percentile of the data (the inter-quartile range or IQR), with the line in the middle of the box representing the median. The lines emerging from the box (whiskers) extend to the next data point beyond 1.5 x IQR. Any more extreme data points are individually plotted.*

Figure 7-11 continued

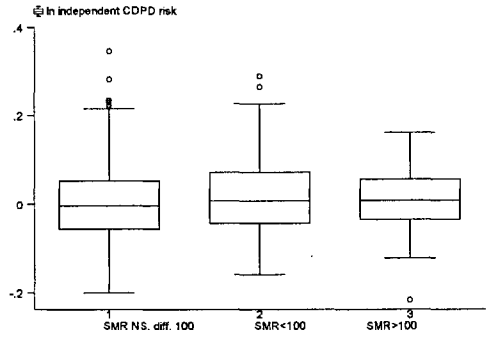
Pneumoconiosis SMRs – males



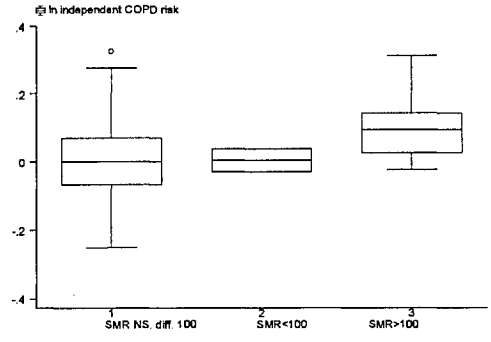
Pneumoconiosis SMRs – females



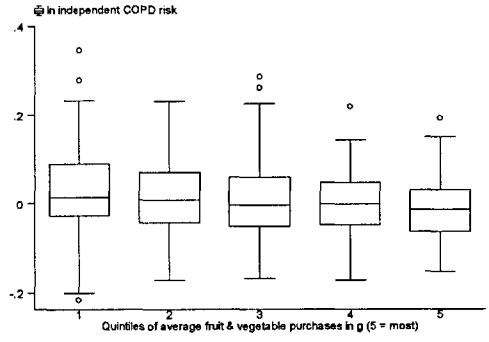
Asbestos related SMRs– males



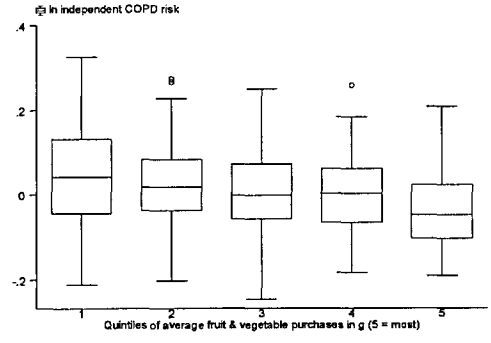
Asbestos related SMRs - females



Fruit and vegetables – males



Fruit and vegetables - females



** The box extends from the 25th percentile to the 75th percentile of the data (the inter-quartile range or IQR), with the line in the middle of the box representing the median. The lines emerging from the box (whiskers) extend to the next data point beyond 1.5 x IQR. Any more extreme data points are individually plotted.*

Spearman rank correlations of explanatory variables

Spearman rank correlations (denoted by ρ) between explanatory variables used in the final model (Appendix Table A-4, page 351) were all generally low. The highest correlation was seen between rain and SO_2 quintiles (exposure variables the same in both male female datasets) with a value of $\rho = -0.42$. All other correlations had values of $\rho<0.2$ except for deprivation and SO_2 quintiles ($\rho = 0.23$ in males, $\rho = 0.24$ in females), deprivation and asbestos-related disease SMRs in males only ($\rho = -0.25$) deprivation and fruit & vegetable purchase quintiles ($\rho = -0.26$ in males, $\rho = -0.25$ in females). This suggests that collinearity of explanatory variables is unlikely to have been a problem for these analyses.

Univariate linear regression

There was no significant difference in univariate models treating quintiles of deprivation, ambient SO_2 or TNS fruit & vegetable purchases as linear or categorical variables, so these were treated as linear variables in the model for simplicity. All other variables were treated as categorical.

All 459 districts were included in the variables for deprivation, SO_2, pneumoconiosis and asbestos-related disease mortality. The temperature and rainfall variables covered 457 of 459 districts and TNS variables covered 451 districts.

Univariate regression analyses found that deprivation, ambient SO_2, rainfall, pneumoconiosis SMRs and fruit & vegetable purchases were significantly associated with independent COPD risks in both males and females (Table 7-12, page 226). Additionally, temperature and asbestos-related disease SMRs were statistically significant in females. The size of associated risks was generally modest and similar to those typically encountered in environmental epidemiology analyses. Results for exposure groupings treated as linear variables could be more forcefully expressed as the difference between quintiles 1 and 5:

- districts in deprivation quintile 5, with the highest percentage of population living in wards in Carstairs quintile 5, had a 13.3% (males) and 16.9% (females)

higher independent COPD risk than those in quintile 1 (with no population in wards in Carstairs quintile 5)

- districts with the highest exposures to SO_2 (average annual exposures >3.9ppm) had a 11.7% (males) and 13.2% (females) higher independent COPD risk compared with the lowest quintile (average annual exposures <1.5 ppm)
- districts with the highest levels of fruit & vegetable purchases (> approx. 2.7 portions per day) had a 4.6% lower (males) and 8.6% lower (females) independent COPD risk compared with the lowest quintile of purchases (< approx. 2 portions per day)

Coefficients for rainfall supported a U-shaped relationship suggested in the descriptive plots with highest risk seen in the middle quintile.

Table 7-12 Results of single variable linear regression of the natural log independent COPD risk (COPDspec) against explanatory variables

Variable	Treatment of variable in model - interpretation of coefficient	Percentage change in independent COPD risk* (95% CI)	p-value
MALES			
Deprivation quintile	Linear – one category increase in quintile	**+2.53 (+2.05 to +3.01)**	**0.000**
SO_2 quintile	Linear – one category increase in quintile	**+2.24 (+1.71 to +2.77)**	**0.000**
Minimum temperature_quintile 2	Categorical - compared with quintile 1	–1.48 (–3.84 to +0.93)	0.225
Minimum temperature_quintile 3	Categorical - compared with quintile 1	+0.23 (–2.20 to +2.72)	0.854
Minimum temperature_quintile 4	Categorical - compared with quintile 1	+1.93 (-0.49 to +4.41)	0.119
Minimum temperature_quintile 5	Categorical - compared with quintile 1	–0.32 (–2.72 to +2.14)	0.796
F-test for joint distribution of temperature			0.069
Average rainfall_quintile 2	Categorical - compared with quintile 1	+0.99 (–1.33 to +3.35)	0.405
Average rainfall_quintile 3	Categorical - compared with quintile 1	**+3.71 (+1.37 to +6.10)**	**0.002**
Average rainfall_quintile 4	Categorical - compared with quintile 1	–0.27 (–2.45 to +1.97)	0.812
Average rainfall_quintile 5	Categorical - compared with quintile 1	**–2.75 (–5.05 to –0.39)**	**0.023**
F-test for joint distribution of rainfall			**0.000**
Pneumoconiosis SMR_group 2	Categorical - compared with group 1	**–2.68 (–4.10 to –1.24)**	**0.000**
Pneumoconiosis SMR_group 3	Categorical - compared with group 1	**+10.16 (+7.63 to +12.75)**	**0.000**
Asbestos-related disease SMR_group 2	Categorical - compared with group 1	+1.41 (–0.52 to +3.37)	0.153
Asbestos-related disease SMR_group 3	Categorical - compared with group 1	+0.46 (–2.15 to +3.14)	0.732
F-test for joint distribution of asbestos-related disease SMRs			0.358
Fruit & vegetable purchase quintile	Linear – one category increase in quintile	**-0.93 (–1.47 to –0.40)**	**0.001**
FEMALES			
Deprivation quintile	Linear – one category increase in quintile	**+3.18 (+2.62 to +3.74)**	**0.000**
SO_2 quintile	Linear – one category increase in quintile	**+2.51 (+1.88 to +3.15)**	**0.000**
Minimum temperature_quintile 2	Categorical - compared with quintile 1	**–2.81 (–5.55 to –0 .004)**	**0.050**
Minimum temperature_quintile 3	Categorical - compared with quintile 1	–1.65 (–4.45 to +1.24)	0.260
Minimum temperature_quintile 4	Categorical - compared with quintile 1	+1.55 (–1.30 to +4.49)	0.289
Minimum temperature_quintile 5	Categorical - compared with quintile 1	–0.90 (–3.74 to +2.03)	0.544
F-test for joint distribution of temperature			**0.023**
Average rainfall_quintile 2	Categorical - compared with quintile 1	+1.11 (–1.59 to +3.88)	0.423
Average rainfall_quintile 3	Categorical - compared with quintile 1	**+7.09 (+4.23 to +10.03)**	**0.000**
Average rainfall_quintile 4	Categorical - compared with quintile 1	+0.85 (–1.73 to +3.50)	0.520
Average rainfall_quintile 5	Categorical - compared with quintile 1	–1.71 (–4.40 to +1.05)	0.222
F-test for joint distribution of rainfall			**0.000**
Pneumoconiosis SMR_group 2	Categorical - compared with group 1	-	-
Pneumoconiosis SMR_group 3	Categorical - compared with group 1	**+17.27 (+7.42 to +28.02)**	**0.000**
Asbestos-related disease SMR_group 2	Categorical - compared with group 1	+0.73 (–10.52 to +13.39)	0.904
Asbestos-related disease SMR_group 3	Categorical - compared with group 1	**+10.62 (+4.70 to +16.87)**	**0.000**
F-test for joint distribution of asbestos-related disease SMRs			**0.002**
Fruit & vegetable purchase quintile	Linear – one category increase in quintile	**–1.71 (–2.32 to –1.09)**	**0.000**

*Percentage change in independent COPD risk calculated as = 100 x (exp (β coefficient) – 1)

Multiple linear regression

Results for the final models for multivariate regression of COPDspec, the log independent COPD risk against the district-level exposure variables are shown in Table 7-13, page 228. The effect of all coefficients was reduced in the multivariate model as compared with the univariate analyses, except for asbestos-related disease SMRs in males, where the effect of SMRs greater than 100 became significant. The final model explained 48% of the variance in COPDspec for males ($r^2 = 0.4827$) and 38% of the variance for females ($r^2 = 0.3839$).

In both males and females, deprivation, ambient SO_2, temperature and rainfall were significant in the final models. Penumoconioses and asbestos-related disease SMR groupings were also significant for males, but not for females despite significant associations in univariate analyses. However, there were very small numbers of districts where females had high SMRs for these conditions (five with pneumoconiosis SMRs >100, 13 with asbestos-related disease SMRs >100), which may be one reason why these were not included in the final model. It should also be noted that asbestos-related disease SMRs groupings were associated with increases in independent COPD risks in univariate analyses, but reductions in risk for males in the final multivariate model (the only risk factor whose effect changed sign between univariate and multivariate analyses). District fruit and vegetable purchases were significant in both sexes in univariate analyses (although smaller effects were seen for males), but only remained significant for females in the final model.

Results for exposure groupings treated as linear variables expressed as the difference between quintiles 1 and 5 suggested a 8.5% (95% confidence intervals 6.1% to 11.0%) in males and 13.1% (10.1% to 16.3%) in females higher independent COPD risk in districts with more deprivation, a 9.3% (6.6% to 12.0%) in males and 9.8% (6.4% to 13.4%) in females higher independent COPD risk in districts with highest levels of SO_2 and a 5.0% (2.3% to 7.7%) lower independent COPD risk for females in districts with highest purchases of fruit & vegetables.

Table 7-13 Results of final model for multiple linear regression of ln independent COPD risk (COPDspec) against explanatory variables

Variable	Treatment of variable in model - interpretation of coefficient	Percentage change in independent COPD risk* (95% CI)	p-value†
MALES			
Deprivation quintile	Linear – one category increase in quintile	+1.65 (+1.20 to +2.11)	0.000
	Most compared with least deprived quintile	+8.54 (+6.12 to +11.02)	
SO_2 quintile	Linear – one category increase in quintile	+1.79 (+1.29 to +2.30)	0.000
	Highest compared with lowest quintile	+9.28 (+6.61 to +12.03)	
Minimum temperature_quintile 2	Categorical - compared with quintile 1	–0.67 (–2.52 to +1.22)	0.486
Minimum temperature_quintile 3	Categorical - compared with quintile 1	+0.48 (–1.52 to +2.52)	0.643
Minimum temperature_quintile 4	Categorical - compared with quintile 1	+3.73 (+1.74 to +5.75)	0.000
Minimum temperature_quintile 5	Categorical - compared with quintile 1	+1.49 (–0.52 to +3.54)	0.148
F-test for joint distribution of temperature			0.000
Average rainfall_quintile 2	Categorical - compared with quintile 1	+0.61 (–1.21 to +2.46)	0.512
Average rainfall_quintile 3	Categorical - compared with quintile 1	+0.39 (–1.58 to +2.40)	0.700
Average rainfall_quintile 4	Categorical - compared with quintile 1	–0.08 (–2.03 to +1.91)	0.936
Average rainfall_quintile 5	Categorical - compared with quintile 1	–3.20 (–5.22 to –1.15)	0.002
F-test for joint distribution of rainfall			0.002
Pneumoconiosis SMR_group 2	Categorical - compared with group 1	–3.50 (–4.77 to –2.21)	0.000
Pneumoconiosis SMR_group 3	Categorical - compared with group 1	+7.41 (+5.31 to +9.56)	0.000
Asbestos-related disease SMR_group 2	Categorical - compared with group 1	–0.29 (–1.74 to +1.17)	0.692
Asbestos-related disease SMR_group 3	Categorical - compared with group 1	–2.50 (–4.48 to –0.48)	0.015
F-test for joint distribution of asbestos-related disease SMRs			0.053
FEMALES			
Deprivation quintile	Linear – one category increase in quintile	+2.50 (+1.94 to +3.06)	0.000
	Most compared with least deprived quintile	+13.13 (+10.07 to +16.28)	
SO_2 quintile	Linear – one category increase in quintile	+1.89 (+1.24 to +2.54)	0.000
	Highest compared with lowest quintile	+9.79	
Minimum temperature_quintile 2	Categorical - compared with quintile 1	–2.43 (–4.76 to –0.06)	0.045
Minimum temperature_quintile 3	Categorical - compared with quintile 1	–3.04 (–5.53 to –0.49)	0.020
Minimum temperature_quintile 4	Categorical - compared with quintile 1	+2.25 (–0.29 to +4.84)	0.082
Minimum temperature_quintile 5	Categorical - compared with quintile 1	0.08 (–2.45 to +2.69)	0.949
F-test for joint distribution of temperature			0.000
Average rainfall_quintile 2	Categorical - compared with quintile 1	+0.37 (–1.93 to +2.74)	0.753
Average rainfall_quintile 3	Categorical - compared with quintile 1	+3.99 (+1.41 to +6.63)	0.002
Average rainfall_quintile 4	Categorical - compared with quintile 1	+1.08 (–1.43 to +3.65)	0.401
Average rainfall_quintile 5	Categorical - compared with quintile 1	–2.09 (–4.66 to +0.54)	0.118
F-test for joint distribution of rainfall			0.000
Fruit & vegetable purchase quintile	Linear – one category increase in quintile	–1.03 (–1.58 to –0.47)	0.000
	Highest compared with lowest quintile	–5.02 (–2.33 to –7.65)	

*Percentage change in independent COPD risk calculated as = 100 x (exp (β coefficient) – 1)
† Wald test for single variables, F-test for joint distribution of categorical variables treated as categorical. Significance of variable within the model established using maximum likelihood

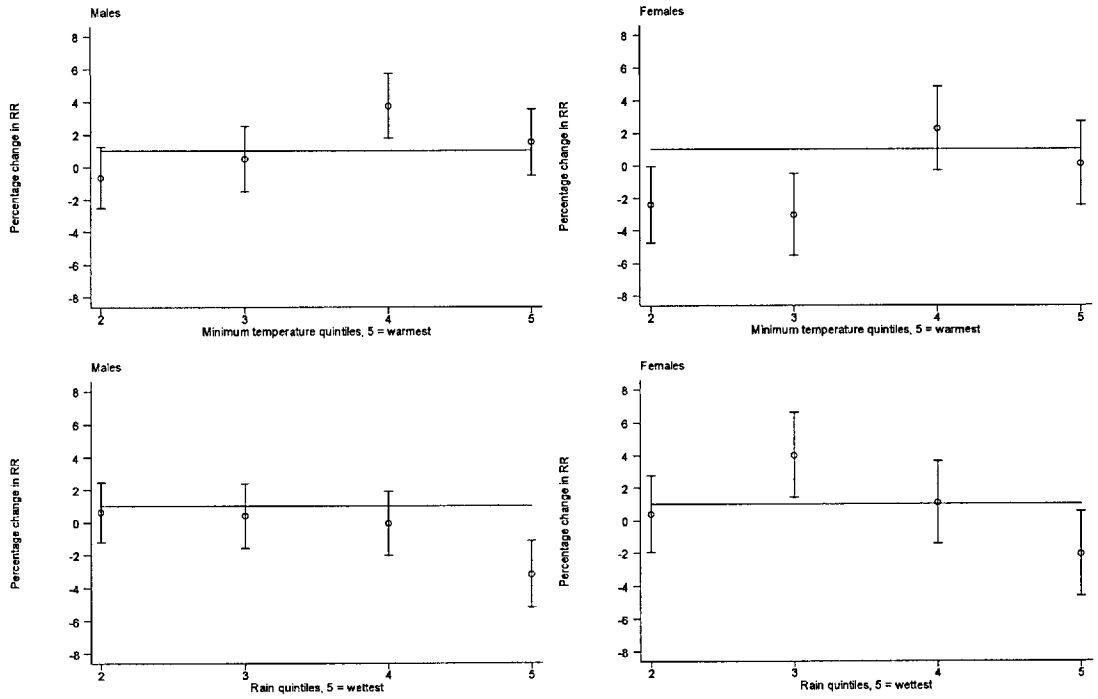
For males, there was a 7.4% (5.3% to 9.6%) increase in independent COPD risk where districts had pneumoconiosis SMRs >100 and a 3.5% (4.8% to 2.2%) reduction in independent COPD risk where pneumoconiosis SMRs were <100. For asbestos-related diseases in males there was a 2.5% (4.5% to 0.5%) *reduction* in independent COPD risk where SMRs were higher than 100.

The associations with minimum temperature and rainfall were complex (Table 7-13, page 228). Using AIC criteria quintiles provided a better fit than tertiles in the model. The percentage change in relative risks is illustrated graphically in Figure 7-12 below. The results do not support the prior hypotheses that higher risks would be seen in colder and wetter districts. An inverted U-shaped relationship with rain remains possible, but results could also suggest that wettest districts have lower independent COPD risks.

Testing the final model – adding log lung cancer SMRs

The addition of the (natural) log of the lung cancer SMR in to the final model was not significant on likelihood ratio testing, suggesting that the shared risks with lung cancer had successfully been removed during the shared component analysis.

Figure 7-12 Percentage change in relative risk for independent COPD risk factor for quintiles of minimum temperature and rainfall compared with first quintile (coldest and driest respectively)

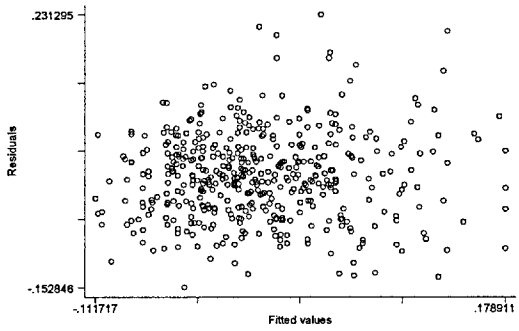


Testing the final model - model fit

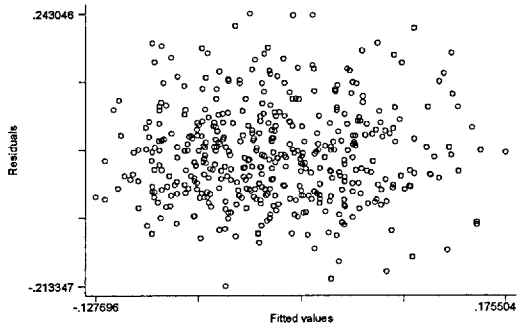
The final model for males had a significant Cook-Weisberg test for heteroskedasticity suggesting increases in variance with increases in fitted values. Examination of a plot of residuals against fitted values (on the left in Figure 7-13) suggested that this was minor. Initial actions to counter this would be to do a weighted analysis and to log the outcome variable – but these had already been done. This should not affect the coefficients, but may results in an underestimate of the confidence intervals. The Cook-Weisberg test for heteroskedasticity for the final model for females was not significant and the plot of residuals against fitted values confirmed this (on the right, Figure 7-13).

Figure 7-13 Plot of residuals against fitted values for the final linear regression models

Males



Females



Secondary analyses

Sensitivity

Comparisons with and without deprivation

Adjustment for deprivation may overadjust the model to some extent, as some of the effect of deprivation is likely to have been removed during the shared component analysis. Comparisons were made of changes in independent COPD risk derived from the final model and changes derived from the final model but excluding deprivation (Table 7-14, page 231).

Removing deprivation from the model generally increased the size of coefficients. For example, the effect of SO_2 on the independent COPD risk comparing highest with lowest quintile rose from +8.5% (6.1% to 11.0%) to +12.1% (9.3% to 14.9%) for males and from +9.8% (6.4% to 13.4%) to +14.9% (11.2% to 18.7%) for females. The effect

of district fruit and vegetable purchases comparing highest with lowest quintile in females rose from a 5.0% (2.3% to 7.7%) to 7.9% (5.1% to 10.5%) reduction in independent COPD risk. Asbestos-related SMRs became non-significant in this model for males. Temperature and rain remained difficult to interpret but temperature effects became more negative (or less positive), while rainfall effects remained consistent with an inverted U-shaped relationship.

Table 7-14 Percentage change in independent COPD risk in the final model and in the final model excluding deprivation

Variable	% change in independent COPD risk* (95% CI)	
	Final model with deprivation	**Final model excluding deprivation**
Males		
SO_2 quintile	+1.79 (+1.29 to +2.30)	+2.30 (+1.79 to +2.82)
Minimum temperature_quintile 2	–0.67 (–2.52 to +1.22)	–0.90 (–2.85 to +1.09)
Minimum temperature_quintile 3	+0.48 (–1.52 to +2.52)	+0.12 (–1.98 to +2.26)
Minimum temperature_quintile 4	+3.73 (+1.74 to +5.75)	+2.78 (+0.72 to +4.87)
Minimum temperature_quintile 5	+1.49 (–0.52 to +3.54)	+0.61 (–1.48 to +2.74)
Average rainfall_quintile 2	+0.61 (–1.21 to +2.46)	+1.35 (–0.58 to +3.31)
Average rainfall_quintile 3	+0.39 (–1.58 to +2.40)	+1.44 (–0.63 to +3.56)
Average rainfall_quintile 4	–0.08 (–2.03 to +1.91)	+0.72 (–1.34 to +2.83)
Average rainfall_quintile 5	–3.20 (–5.22 to –1.15)	–1.91 (–4.03 to +0.26)
Pneumoconiosis SMR_group 2	–3.50 (–4.77 to –2.21)	–4.13 (–5.46 to –2.79)
Pneumoconiosis SMR_group 3	+7.41 (+5.31 to +9.56)	+9.02 (+6.81 to +11.27)
Asbestos-related disease SMR_group 2	–0.29 (–1.74 to +1.17)	+0.42 (–1.11 to +1.96)
Asbestos-related disease SMR_group 3	–2.50 (–4.48 to –0.48)	–0.58 (–2.64 to +1.51)
Females		
SO_2 quintile	+1.89 (+1.24 to +2.54)	+2.82 (+2.15 to +3.49)
Minimum temperature_quintile 2	–2.43 (–4.76 to –0.06)	–2.96 (–5.46 to –0.40)
Minimum temperature_quintile 3	–3.04 (–5.53 to –0.49)	–3.23 (–5.91 to –0.46)
Minimum temperature_quintile 4	+2.25 (–0.29 to +4.84)	+0.93 (–1.76 to +3.70)
Minimum temperature_quintile 5	+0.08 (–2.45 to +2.69)	–1.34 (–4.02 to +1.43)
Average rainfall_quintile 2	+0.37 (–1.93 to +2.74)	+1.42 (–1.09 to +4.00)
Average rainfall_quintile 3	+3.99 (+1.41 to +6.63)	+5.99 (+3.18 to +8.87)
Average rainfall_quintile 4	+1.08 (–1.43 to +3.65)	+2.72 (–0.01 to +5.52)
Average rainfall_quintile 5	–2.09 (–4.66 to +0.54)	+0.71 (–2.06 to +3.56)
Fruit & vegetable purchase quintile	–1.03 (–1.58 to –0.47)	–1.62 (–2.20 to –1.04)

*Percentage change in independent COPD risk calculated as = 100 x (exp (β coefficient) – 1)

Restricting analyses to districts with at least 10 households in TNS survey

Restricting analyses to the 368 districts with at least 10 (weighted) households contributing to the TNS super-panel increased by a quarter the associated reduction in independent COPD risk in females comparing the highest and lowest quintile of fruit and vegetable purchases, with reduction in risk rising from 5.0% (2.3% to 7.7%) to 6.2% (3.1% to 9.2%). However, fruit and vegetable purchases remained non-significant for males.

Interactions

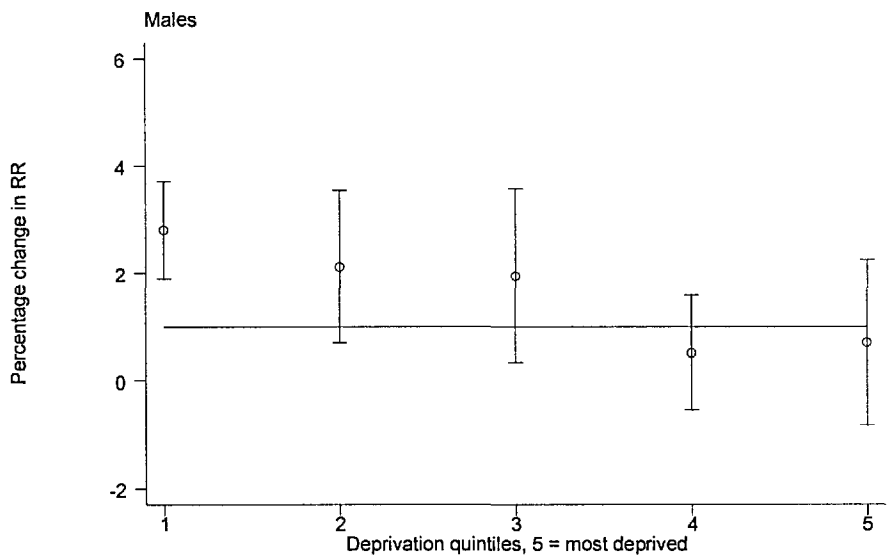
Interactions with deprivation

Significant interactions with deprivation quintiles were seen for quintiles of SO_2 (in fixed ranges as detailed in Table 7-4, page 202) and quintiles of temperature and groupings of asbestos-related SMRs in males and with fixed range quintiles of SO_2 for females. Contrary to expectations, the effect of SO_2 appeared to be highest in the least deprived districts a suggestion of decreasing effect with increasing deprivation (Figure 7-14, page 233). The interaction effects of deprivation with temperature for males were difficult to interpret (Figure 7-15, page 234), with most effects not significantly different from the baseline quintile (coldest). Temperature interactions were for females are not presented but were borderline significant (p=0.08). The interaction effects of deprivation with high or low asbestos related disease SMRs in males was also difficult to interpret (Figure 7-16, page 235), but was, perhaps surprisingly, suggestive of a lower independent COPD risk in districts with high asbestos related disease SMRs (seen in the least and most deprived quintiles).

Interactions between temperature and rain

There were statistically significant interactions between rain and temperature effects that were only borderline significant in males (p=0.049) but strongly significant in females (p=0.0006). These suggested an increase in independent COPD risk with increases in rainfall, seen only in the coldest districts (those in the lowest quintile for average minimum temperature) and were easier to see in graphs for females (Figure 7-17b, page 237) than for males (Figure 7-17a, page 236).

Figure 7-14 Percentage change in independent COPD risk per quintile increase in ambient SO_2 concentration by deprivation quintile



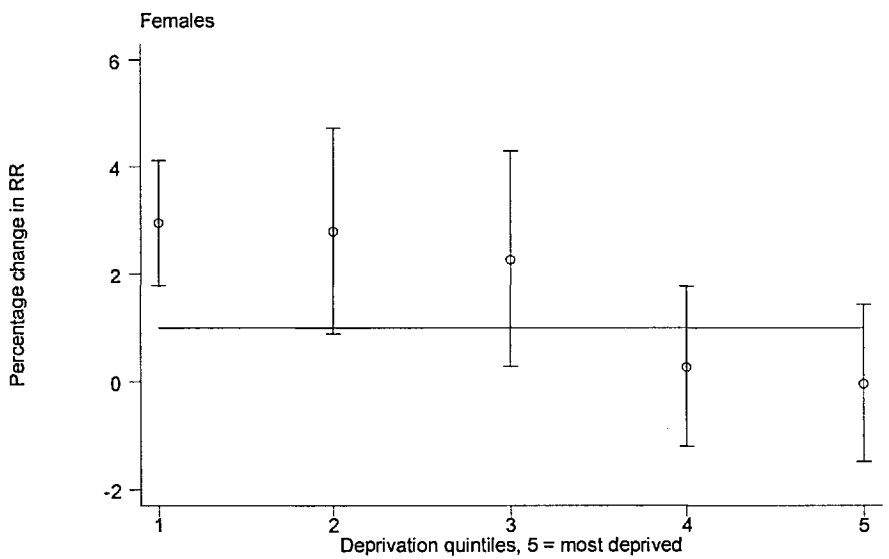


Figure 7-15 Percentage change in independent COPD risk with increasing average minimum temperature (quintiles compared with quintile 1 – lowest temperature) for males by deprivation quintile

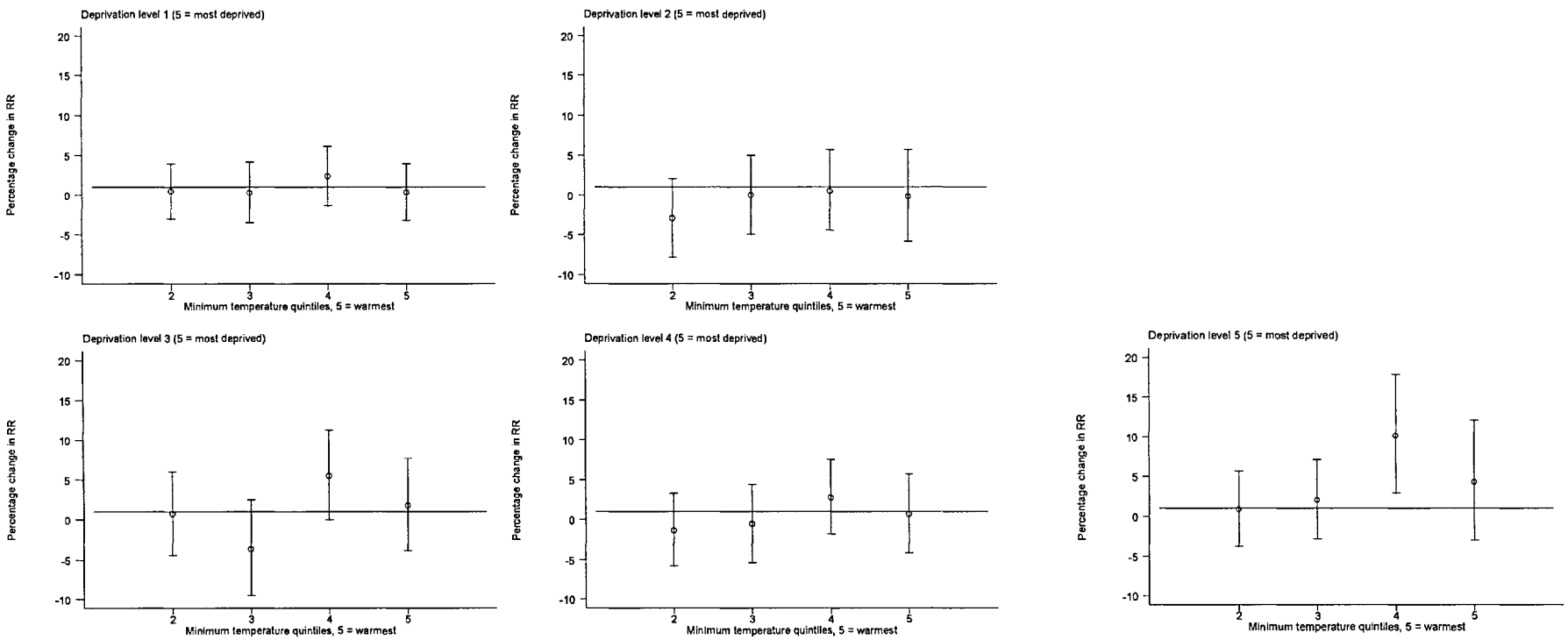


Figure 7-16 Percentage change in independent COPD risk with asbestos-related disease SMRs <100 or >100 (compared with SMRs non-significantly different from 100) for males by deprivation quintile

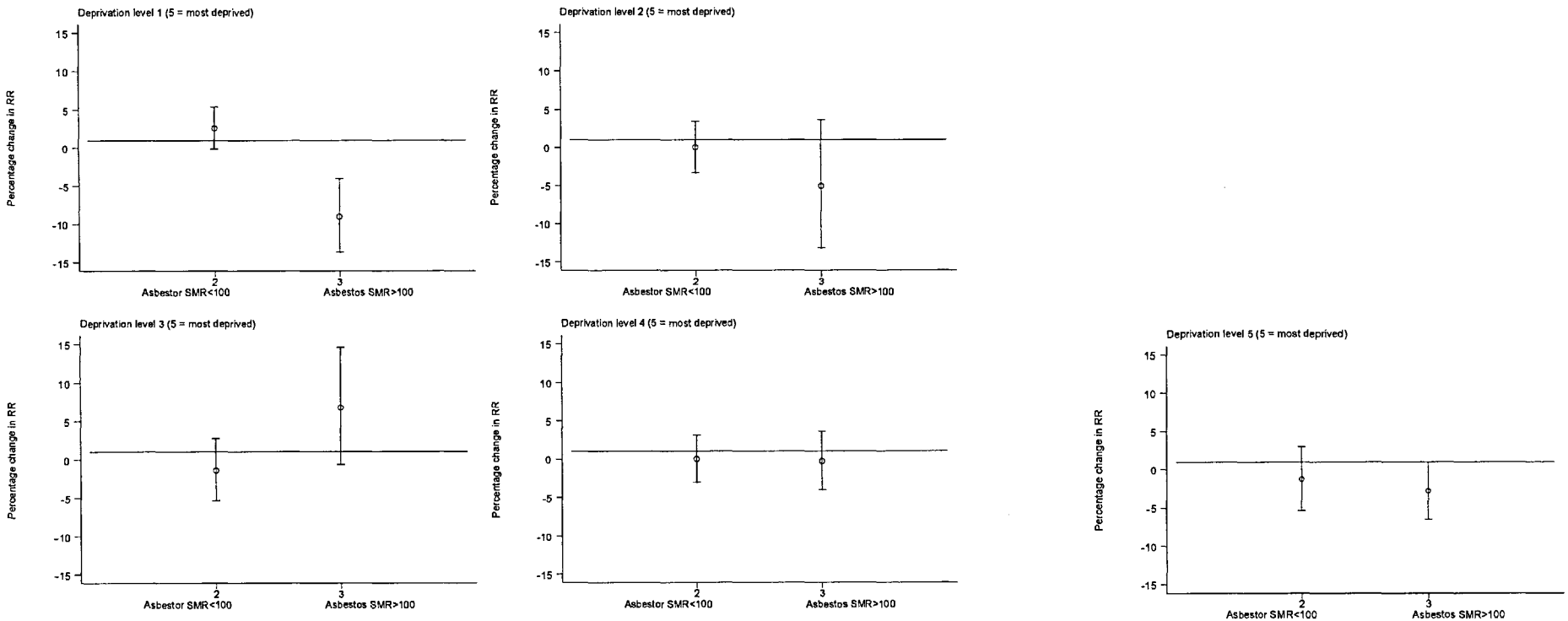


Figure 7-17a Percentage change in independent COPD risk with increasing rainfall by temperature quintile – males

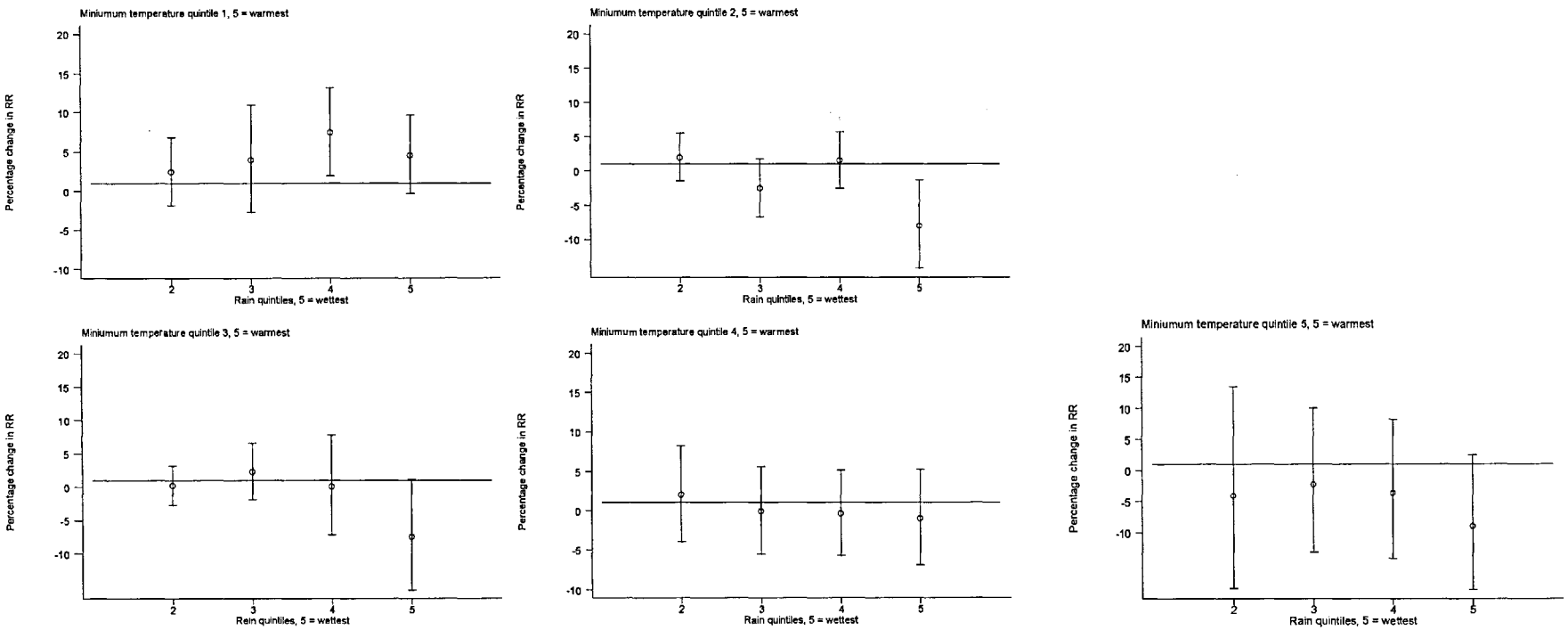
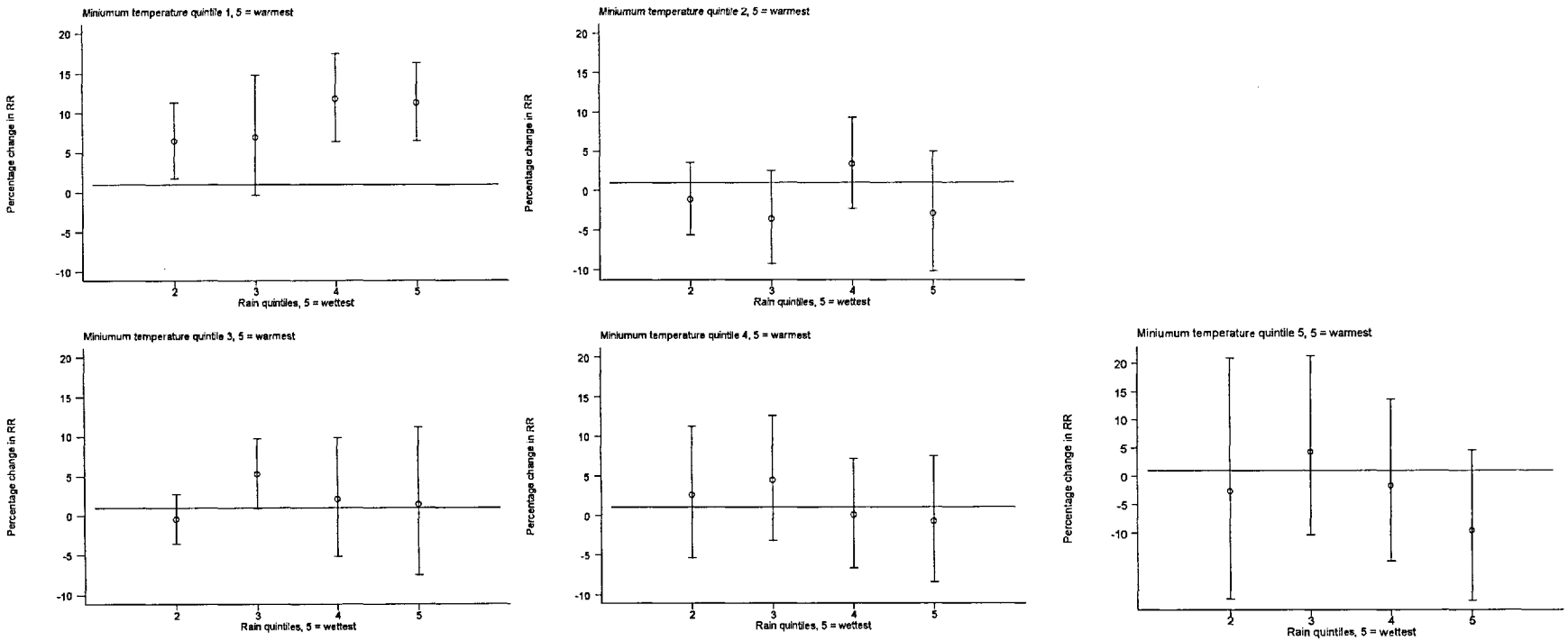


Figure 7-17b Percentage change in independent COPD risk with increasing rainfall by temperature quintile – females



Section IV

The epidemiology of chronic obstructive pulmonary disease in the UK: spatial and temporal variations

Discussion and Conclusions

Chapter 8. Discussion

Overview and summary of Chapter 8

This section discusses the results from the temporal and spatial analyses.

Discussion of methodological issues common to both analyses

- The main outcome used was COPD mortality. As well as being an important health endpoint in its own right, geographical variations in mortality are likely to be similar to those for COPD morbidity. While differences in diagnostic labelling and coding have occurred over time, these were not thought to have significantly affected the analyses.
- Lung cancer was used as a surrogate for cumulative smoking. This approach was used by Peto[46] and others[100] in the Global Burden of Disease studies. It may be a better measure than current smoking prevalence for chronic diseases such as COPD where cumulative exposures are important. Drawbacks of this approach include the association of lung cancer with other exposures of interest in these analyses such as air pollution and the possibility of different lags for smoking on COPD and lung cancer.
- The chief sources of bias in the analyses presented are likely to be the ecological fallacy, migration and misclassification of exposures.

Discussion of temporal analyses

- Results suggested that air pollution was an important influence on COPD mortality in the 1950s to 1970s, but after it declined to lower levels in the late 1970s smoking became the major determinant.
- Prior beliefs about changes in air pollution were consistent with actual data.
- COPD mortality trends (over time and the relative difference between areas with different levels of urbanisation) were found to be in agreement with *a priori* expectations if air pollution changes following the 1956 Clean Air Act had important effects on COPD mortality. A greater similarity was seen between trends for black smoke and COPD mortality than for SO_2.

- The interpretation that ambient black smoke and SO_2 exposures were important in explaining COPD mortality trends was consistent with population-level studies in Dublin[308] and Hong Kong[309] in the 1990s on the effects of falls in black smoke and/or SO_2 on respiratory mortality. The size of the effect of air pollution was possibly higher in the UK in the 1950s to 1970s, when exposures were much higher.
- The interpretation was also coherent with pathological studies, cross-sectional and cohort studies looking at lung function, studies of COPD incidence and morbidity, and time-series studies looking at acute effects on COPD mortality.
- Changes in cigarette composition, COPD treatments (especially influenza vaccination and antibiotic use) and diet may also have influenced COPD mortality but only air pollution changes could fully explain observed trends.
- It was not possible to determine the lag or cumulative impact of air pollution on COPD mortality, but maximum effect within a few years of exposure would be most consistent with results.
- The use of non-conurbations as a comparator may have provided a wandering baseline with differential changes in air pollution and other risk factors.
- Timings of cohort effects for males and females and period effects were consistent with other classical age-period-cohort analyses, but the BAMP method was able to pinpoint the timing more precisely to specific years.

Discussion of spatial analyses

- The spatial analyses suggested that factors other than smoking remained important in determining geographical patterns of COPD mortality in the 1980s and 1990s.
- Results were consistent with analyses by Law and Morris[21] suggesting that smoking may only explain half of the current geographical variability in COPD mortality.
- Quintiles or groupings rather than absolute values were used because of changes over time. Analyses relied on the assumption that relative differences between areas remained constant over time, for which justification was given.
- All datasets used had potential sources of bias. Minor differences between associations of COPD mortality and district-level versus ward-level deprivation were seen, but the underlying 'true' relationship with deprivation is unknown.
- Interpretations of specific associations were:

- Deprivation may have mainly captured residual smoking effects.
- The numerical size of the association between ambient SO_2 was similar to that seen in American analyses of chronic air pollution effects, but as spatial analyses used COPD specific risks after the removal of estimated effect of cumulative smoking and not total COPD mortality as an outcome, direct comparison with other studies were difficult.
- Significant associations with fruit and vegetable consumption were seen in multivariate analyses in females. This may be because women typically consume more of these foods. However, this variable may be solely acting as a marker for other lifestyle factors.
- The association with pneumoconioses may have been partly related to an increased willingness to diagnose COPD in mining areas as the disease is eligible for industrial compensation.
- Changes in coefficients for asbestos-related disease SMRs between univariate and multivariate analyses were strongly suggestive of confounding rather than a true effect.
- Temperature and rainfall showed complex effects and interactions with both each other and with deprivation, which could be linked through housing quality.

- The analyses conducted were designed to establish whether risk factors other than smoking were important in spatial variations in COPD mortality. Further analyses would be needed to assess the size of the association of exposures with COPD risk and thence attributable risks.

Methodological issues common to both temporal and spatial analyses

Relevance of COPD mortality as an outcome

These analyses considered COPD mortality rather than hospitalisations or prevalence, as mortality was the most readily available over the time spans used in the analyses. COPD mortality is an important health outcome, causing 5% of all deaths in recent years.[17] Additionally, completeness of capture of deaths is excellent[236] and a lot of information on quality issues such as coding is available.[258-261]

Results presented were based on the underlying cause of death, but COPD is a chronic condition often mentioned as a contributing cause in Part II of death certificates Cohort studies suggest that a mention of COPD (including as underlying cause of death) on the death certificate is more likely in individuals with severe COPD,[222] but as few as one in five of individuals with severe COPD in life may have COPD mentioned as the underlying cause of death.[263] A multiple-cause coding analysis for England & Wales 1993-1999[17] suggested that time trends in underlying cause of death and in any mention of COPD were similar in 1993-1999. However, the relative proportions for earlier years are unknown and changes in this proportion over time could be related to both fashions in labelling (inclusion of contributing causes of death) and to the underlying prevalence and severity of COPD in the population.

It is less clear whether results suggesting effects of air pollution and other factors on mortality can be used to infer effects on COPD prevalence, severity or health-care use. The relationship between morbidity and mortality in Britain has only been closely examined in recent years. As mentioned in Chapter 3, a study examining the relationship between levels of COPD mortality using underlying cause of death and hospital admissions, general practice consultations and symptoms in 1991-1995 in England[83] found a good spatial correlation at health authority level with Spearman rank correlations around 0.8 between mortality and morbidity rates. This implied that areas with high COPD mortality would also have high levels of hospital admissions and GP consultations. Results from spatial analyses in this PhD found that different distribution of specific COPD risk factors explain some of the variability in COPD mortality. It is

plausible, therefore, that the same risk factors will also explain some of the geographical variability in COPD morbidity and prevalence, although exact percentages may differ e.g. if those with the most severe COPD and the greatest risk of dying are the most affected by that risk factor. Also, asthma was included in the definition of COPD in the spatial analyses (and contributed approx 7% overall of obstructive lung disease deaths[17] during this time period).
Correlation analyses for asthma in the study mentioned above[83] found poor correlations between health authority levels of asthma mortality and morbidity.

Diagnostic labelling and coding variations in mortality

Differences in diagnostic labelling of COPD over time or in different geographical areas may have affected the findings in the analyses. Changes in coding using national data from 1950 actually suggested fairly constant proportions of deaths coded to asthma and chronic bronchitis (chronic bronchitis plus chronic airways obstruction in ICD9) over time (Appendix, Table A-5, page 352). However, death certificate[258-260] and health-care[49,50] analyses have suggested variation in the diagnostic labels used for obstructive lung disease in individual patients. To attempt to minimise this misclassification bias, as inclusive definition of COPD as possible was used in both temporal and spatial analyses.

While inclusive diagnoses may reduce such bias, this may be at the expense of increasing heterogeneity of the clinical condition identified and thus some reduction in power to detect associations. For example, Lee[27] suggested that emphysema was more closely associated with cigarette smoking than chronic bronchitis in analyses for 1941-85. Some specificity was introduced by the use of an age cut-off at 45 years for spatial analyses, where asthma was included in the definition of COPD. For temporal analyses, including younger individuals, asthma diagnoses were excluded. However, it was not possible to disaggregate emphysema in the data available for temporal analyses and there are concerns about the ability to distinguish between the two conditions using death certificate data.[50] Nationally, emphysema deaths formed 7% or less of COPD deaths (Appendix, Table A-5, page 352) so it is unlikely that emphysema deaths are responsible for the differences between conurbations and non-conurbations in temporal analyses.

The impact of changes in coding were investigated through investigation of conversion factors derived from bridge-coding analyses and dual coding in relation different interpretations of coding rule 3 1984-1992 and additionally through descriptive analyses looking for artefactual step changes in rates, as described in Chapter 4. No specific effects were identified either through visual inspection of rates or in a detailed examination of timing of period effects in age-period-cohort analyses. This may reflect on the small size of estimated changes. However, it may also indicate a lack of sensitivity of visual inspection and the age-period-cohort models used to detect small changes – say less than 10% change in rates (rule 3 changes between 1984-1992 in England & Wales were suggested to have increased the numbers of deaths in COPD by 7%, with up to 10% increases in those aged 75+ years[261]).

Spatial analyses were also potentially affected by small artefactual changes in mortality predicted for COPD but not lung cancer[261] as a result of different application of coding rule 3 in 1984-1992 in England & Wales (but not Scotland). Data were grouped for 1981-1999 and it is unlikely that these changes would have varied spatially in the same way as exposure data. However, it is possible – but unlikely – that this could have affected the shared component analyses based on comparisons between COPD and lung cancer, perhaps through some distortion in the expected number of cases for COPD (based on pooled numbers for England & Wales and Scotland for each year).

Diagnostic labelling and coding issues in relation to lung cancer

Different histological types of lung cancer have different relationships with smoking and other risk factors[310] and a general shift from a predominance of squamous cell to adenocarcinoma has been noted over the last 50 years.[310] A broad label of lung cancer was used as the different types of lung cancer cannot be readily identified in mortality data and it is likely that better histopathological information would be available for more recent years than for the 1950s and 1960s. While improvements in diagnostic techniques such as the fibreoptic bronchoscope, thin-needle aspiration and improved cytological staining occurred in the 1970s,[310] these are more likely to have affected histological than clinical diagnosis. X-rays and sputum cytology are arguably the mainstay of clinical diagnosis and these would have been available from the 1950s onwards. Diagnostic improvements between 1941-75 were not thought to have affected a previous age-period-cohort analysis of lung cancer mortality in England & Wales.[311]

Using lung cancer mortality as a surrogate for cumulative smoking

Both cancer of the lung and chronic obstructive lung disease are considered to be strongly and causally related to smoking.[28] In the absence of appropriate data on smoking to the geographical resolution required, lung cancer mortality was used as an indirect indicator of cumulative smoking in both the temporal and spatial analyses presented in this PhD. It is arguably a better measure of smoking for diseases such as COPD where a cumulative effect of smoking is seen[108] than a single measure of smoking prevalence derived from a survey.

The analyses presented were dependent on the assumption that smoking-attributable risks for lung cancer mortality were high, but were lower for COPD mortality. For spatial analyses in particular, small differences in smoking attributable risks between lung cancer and COPD could have limited the ability to detect impacts of non-smoking factors (although the size of the absolute differences between smoking-attributable risks for COPD and lung may not be so important if one remembers that attributable risks from different causes can add up to more than 100%[209]).

Magnitude of risks attributable to smoking for lung cancer and COPD mortality

Lung cancer mortality was first used as an indirect indicator of cumulative smoking by Peto et al[96] when estimating the global burden of mortality associated with tobacco and the method was further refined by Ezzati and Lopez.[100] The methods used by Peto[96] and Ezzati[100] calculated smoking rates as the excess rates for lung cancer in smokers, using data on lung cancer mortality rates for the general population and lung cancer mortality rates in non-smokers from the American Cancer Society cohort study. In contrast, methods used in this PhD assumed that lung cancer mortality rates captured almost all of the smoking effect. This assumption was also made in a British study by Strachan and Maheswaran[144] estimating the effect of air pollution exposure in childhood (derived from an index of coal consumption) and death 60 years later, using the Longitudinal Study cohort. It has been employed in a geographical analysis of COPD mortality rates in the USA.[166]

This approach is reasonable, given that over 90% of lung cancer was found to be attributable to smoking in 1950-2000 using the method described by Peto et al[96] (some 95% in men and 85% in women[28]). Peto also suggested reasonable differential between

smoking attributable mortality for lung cancer and COPD, with approximately 75% of COPD mortality in the UK attributable to smoking over 1950-2000[96] (varying from 53-86%[96,99] depending on time period studied, gender and age-group). This compares with the findings in the shared component analysis presented in this PhD that 64% of variance in overall COPD risk could be accounted for risks shared between COPD and lung cancer, which can be thought of as largely attributable to smoking, with the addition of other shared factors.

Different methodological approaches to estimate disease risks attributable to smoking give risk estimates compatible to those by Peto. A Danish study[312] using a risk factor model taking detailed account of the evolution of smoking (prevalence, quit rates, age at starting to smoke, inhalation) in the previous 20 years suggested attributable risks for lung cancer mortality of 83% for males and 82% for women – lower than the Peto method – but attributable risks for COPD mortality that were similar to those obtained using the Peto method. An age-cohort analysis method incorporating information on smoking[311] in British males 1941-75 was in close agreement with the Peto findings in that 88% of lung cancer mortality was attributable to smoking.

The methods used to determine risks attributable to smoking are all dependent on a number of assumptions. The Danish risk factor method[312] assumed that 20 years information on smoking habits is sufficient, but this may underestimate the impact of smoking on lung cancer[313] The age-cohort method[311] makes assumptions about the statistical relationship between smoking and lung cancer. Both the Peto[96] method and the Danish study method assume relative risks for cause-specific mortality from smoking derived from the American Cancer Prevention studies are directly transferable to other populations.[100] This is not always tenable e.g. general population lung cancer rates (including smokers and non-smokers) in males in the UK in 1950-1970 were found to be higher than rates for lifelong smokers in the American Cancer Prevention study[100] (although Peto considered these American rates were similar to those for professional men in the UK[313]), which might suggest a lower attributable risk for smoking in the UK at least in these years.

Relationship of air pollution and lung cancer

Both temporal and spatial analyses made the implicit assumption that any association of lung cancer with air pollution (and other risk factors) was small. This is consistent with

findings in the analyses presented here. In the temporal analyses, strong period effects for COPD were detected that were similar to trends in air pollution, but no consistent period effects were seen for lung cancer. Spatial analyses found an effect of SO_2 on COPD mortality in the 1980s and 1990s after removal of a component shared between lung cancer and COPD, suggesting either an excess effect of SO_2 on COPD mortality or a small effect of ambient SO_2 on lung cancer or both.

Long term exposures to high levels have been certainly associated with lung cancer mortality.[41,265,314] For example, the American Cancer Society Cancer Prevention II study showed statistically significant associations between lung cancer mortality in the 1980s and 1990s and particulates (as $PM_{2.5}$ but not as total suspended particulates), sulphate particles and SO_2[265] and the AHSMOG study of Seventh Day Adventists[315] also showed associations with PM_{10} and SO_2. Exposure to smoky coal fuel, the major component of urban air pollution in the 1950s and 1960s, has been associated with increases in lung cancer in China.[316] However, Doll and Peto estimated in 1981[42] that perhaps 1-2% of lung cancer cases in the United States were related to air pollution and this was considered to hold in recent reviews.[310,317] Higher estimates have been suggested in some studies with 10-12% of lung cancer in cities attributed to air pollution.[317] Air pollution (from combustion products of fossil fuels) was estimated to have caused some 5-10% of lung cancers in men in the UK in the 1950s-1970s[318] possibly in association with tobacco smoking and occupational exposures, with estimates of around 4% of cancers in non-smoking women.

Risk factors for lung cancer other than air pollution are considered later in the paragraphs relating to spatial analyses.

Lag period for effects of smoking for both COPD and lung cancer

Another implicit assumptions in using lung cancer as a marker of cumulative smoking is that lag periods for COPD and lung cancer are similar (where lag reflects the average time over the population of individuals between onset of smoking and disease or mortality). The lag period for both diseases is difficult to ascertain in the literature and it likely that the lag period between onset of smoking and disease is itself only a marker for accumulation of pack years of smoking[310,313] in combination with other factors (genetic susceptibility, protective factors such as diet, interactions or additive effects from occupational or other exposures).

Lung cancer mortality rates are often stated to lag smoking rates by about 20 year[310] and this relationship is suggested in graphs of tobacco consumption in Great Britain against lung cancer mortality (Figure 8-1, page 258). Lag effects of smoking on COPD mortality are more rarely explicitly stated, but one author[319] suggested 10-40 years. Graphs of tobacco consumption in Great Britain against COPD mortality (Figure 8-1, page 258) suggest a relationship between tobacco consumption and COPD mortality from the 1980s onwards – the relationship for females suggests a slightly longer lag of around 25 years but the lag is difficult to establish for males. This timing is consistent with observations in Japan, where an increase in COPD mortality started approximately 25 years after increases in cigarette consumption in Japan occurring in the mid 1950s.[320]

Cohort effects in temporal analyses in this PhD suggested similar timing for both lung cancer and COPD mortality cohort effects (Table 5-11, page 157), but the effects were diffuse and probably not sensitive enough to pinpoint a difference of five years.

Lag analyses for shared component analyses

Because of concerns about the differences in lags, the Bayesian shared component analyses were repeated using lung cancer mortality for 1981-90 and COPD mortality for 1991-1999. Analyses were conducted using the same statistical model as described in Chapter 6, the only difference being that expected numbers were based on averaged national rates for all years 1981-1999 combined rather than for each year singly. Results (not shown) in terms of both the proportions of variance explained by the shared component for COPD and lung cancer risks and the geographical variation in shared and independent risks were very similar to analyses using all years combined presented in Chapter 7. The similarity may reflect similar lag periods but could also reflect some constancy in the relative smoking levels – districts with higher smoking levels in the 1990s also probably had higher smoking levels in the 1980s.

Bias and confounding – general issues

The inferences made from this study should be regarded as hypothesis generating rather than definitive: they are based on descriptive analyses of group level data. Data were prone to a number of biases, and individual level information on major confounders such as cigarette smoking were not available. The issue of biases in the data and the problem of migration are discussed in more detail in the discussion of the temporal and

spatial analyses, but two issues will be briefly discussed here – the ecological fallacy and changes in the nature of air pollution data over time.

Ecological fallacy

Group-level analyses such as those presented here have risks of bias additional to those seen in estimation of individual level studies if there is heterogeneity of exposure levels within groups, if there are variations in the background rate of disease in the unexposed population in different groups, or if the rate difference for the exposure effect at the individual level varies across groups.[209] The latter two effects may occur if modifying risk factors or confounders are differently distributed across groups, if there are contextual effects of the ecologic exposure or if disease risk depends on the prevalence of that disease in other members of the groups (e.g. infectious respiratory illnesses increasing the risk of COPD mortality). As a result the risk estimates seen at the group level may not reflect risk estimates at the individual level. This is termed the ecological fallacy.

The size and effect of such biases can be very difficult to estimate. Adjustment for confounders such as smoking based on group-level data may inadequately control for confounding.[209] The relevance of this is that even if lung cancer is a good indicator of cumulative smoking, ecologic bias may have limited its ability to fully capture the impacts of smoking. An additional problem is that the direction of effect of misclassification bias cannot be predicted in group-level analyses.[209] As set out in detail in Rothman and Greenland's Modern Epidemiology,[209] in individual analyses, nondifferential misclassification of exposure nearly always biases the effect estimates towards the null, in groups the direction of the bias cannot be readily predicted. This is of particular concern for the spatial analyses here, which used a number of exposure variables with varying issues about the way in which the data were collected and compiled. Measurement of regression model fit is not helpful to determine if ecologic bias is present – the model may fit the data well, but still have considerable bias.[209]

A strategy for reducing ecologic bias is to use the smallest geographical units possible – this was attempted in the spatial analysis but there was not enough information on most exposures to permit analyses at area levels smaller than districts. Another strategy would be to increase the information available on exposures within groups in areas e.g. proportions of fruit & vegetable consumption by ethnic group (if available) i.e. the

within group variability. It is possible to incorporate such information in both classical[209,321] and Bayesian frameworks.[321] Generally information on most exposure variables used in the spatial analyses was too limited to permit this. Analyses were stratified by sex, but it had to be assumed that some exposure variables were similar for males and females (some reasonably so – temperature, rainfall, ambient air pollution – others less so – fruit & vegetable consumption). As an indirect check on the plausibility of results, the size and direction of risk estimates were compared with those from published analyses.

Air pollution

As noted in Chapter 3, changes in air pollution over time may have affected the validity of the measurement techniques (e.g. black smoke methods are less good at detecting particles lighter in colour and may underestimate particulate levels in more recent years). Also the chemical constituents of particulates and the component acid gases collectively attributed to SO_2 have changed over time. It is therefore possible that the risk relationships of the pollutants as measured with diseases such as COPD and lung cancer have changed over time. The nature of these measurement biases are unknown, but it seems unlikely that they can explain the large changes in period effects seen with COPD mortality although they may have had some effect on the size of the risk coefficients for the spatial analyses for 1981-99.

Temporal analyses

Consistency of results with other age-period-cohort analyses

The period and cohort effects detected in the temporal analyses were consistent with those found in previous age-period-cohort analyses for England & Wales using classical statistical techniques.

Stevens and Moolgavkar[311] analysed lung cancer mortality in males in England & Wales 1941-75. Lee *et al*[27] analysed lung cancer, chronic obstructive lung disease (COLD) and emphysema deaths for England & Wales 1941-85. Marks and Burney[26] analysed COPD (excluding asthma) mortality rates for 1901-1991. Both of the latter analyses[26,27] used bridge coding factors to adjust for change years. All three analyses found cohort effects in all conditions analysed for males born around 1900. Lee[27] found

cohort effects for females born around 1925 for all diseases analysed. Marks[26] commented that '*it may be too early to see the cohort effect on COPD mortality as a result of smoking in women*', but a cohort effect in women born around 1925 is visible in the graphs presented by him.

Period effects were not directly commented on by either Lee[27] or Marks[26] - the paragraph titled 'period effects' in the latter analysis[26] actually describes the age-sex specific mortality rates. However, graphs of both period and cohort effects presented in Lee *et al*[27] suggest possible maxima (rise followed by a fall) in 1950, 1965 and a minimum in 1975 in COLD mortality in both males and females – extremely similar to the timing of period effects in COPD mortality in 1962-3 and 1977-8 in the analyses for this PhD, bearing in mind that Lee's analyses used a five year grid. Lee's analyses[27] also suggest a period effect around 1950 or 1955 in lung cancer and emphysema (maxima for emphysema, a maximum for male lung cancer and a minimum for females lung cancer). This is also consistent with timing and direction of period effects for lung cancer detected in three of the six area-sex combinations in the analyses presented in this PhD (maxima period effects seen in 1954 for males in conurbations and non-conurbations and minimums in 1955-56 for females in non-conurbations (Table 5-10, page 156). Stevens[311] did not find period effects for lung cancer mortality in males 1941-75 in a basic age-period-cohort model, but found an effect in 1951-55 in a more complex model adjusting for the effect of smoking.

Agreement of results with original hypotheses

The first four hypotheses investigated in this PhD were formulated *a priori* with regard to expected changes in COPD and lung cancer mortality if air pollution changes following the 1956 Clean Air Act had important effects on COPD:

> ***1. Overall rates of COPD mortality would be higher in more urban than more rural areas from 1950 onwards. Alternatively, COPD rates would be higher in more urban areas until the late 1950s after which they would converge towards those in more rural areas.***
>
> ***2. Declines in COPD mortality rates would been seen in both urban and rural areas starting from the late 1950s and this would be located as a period effect***

in age-period-cohort analyses. Lung cancer would not show the same period effects.

3. Age effects in an age-period-cohort analyses of rates 1950 onwards would be similar in all areas.

4. Cohort effects in age-period-cohort analyses would be similar in COPD and lung cancer reflecting lifetime smoking patterns. There might be an additional COPD cohort effect in those born 1935-60, related to progressive reductions in air pollution exposure in childhood.

In fact, both of the expected trends in rates were seen in relation to the first hypothesis, with both higher rates in the more urban areas and evidence of a convergence towards those in non-conurbations. Furthermore, convergence was seen for COPD but not lung cancer rates between the late 1950s and late 1970s, when the greatest national improvements in particulate and sulphate air pollution were seen. That the convergence was not seen in lung cancer argues against an effect due to smoking. Also, timing of the convergence was similar in males and females, although trends in smoking patterns are known to differ markedly by sex.

In line with the second hypothesis, the decline in COPD mortality was seen in all areas as might be expected from the national effects of the 1956 Clean Air Act. This decline was located as a period effect – i.e. affecting all age-groups and representing effects once the strong effects of age and cohort effects (probably representing generational smoking patterns and discussed later) were removed. The timing of onset identified by Bayesian age-period-cohort analyses was diffuse but had certainly occurred by 1962-63 and this could reflect the gradual impact of the legislation or a lagged effect on mortality. Both COPD mortality rates and period effects (from classical age-period-cohort analyses assuming shared age and cohort effects) suggested falls that were more marked in conurbations and Greater London than non-conurbations. Relative differences in COPD mortality period effects between areas (from classical age-period-cohort analyses) were comparable with relative differences in air pollution levels between these area, particularly for black smoke but less so for SO_2. Lung cancer mortality, used as a marker of cumulative smoking, was not found to show the same period effects as COPD mortality as predicted, suggesting that period effects in COPD mortality were unlikely to be due to smoking.

The third hypothesis was that age effects on COPD mortality would be similar in both urban and rural areas. However, this hypothesis could not be tested in the analyses as this assumption was used to partition mortality data for younger age groups published in 20 year agebands into 10 year agebands for 1950-62 (described in Chapter 4, page 92).

The fourth hypothesis, that similar cohort effects would be seen for COPD and lung cancer was met for each area and sex group and these were consistent with smoking patterns, with women lagging behind men by about 25 years.[28] A second possible cohort effect in COPD mortality those born 1935-60, related to progressive reductions in air pollution exposure in childhood, was not seen except possibly in one of the six area-sex groupings with weak evidence for a cohort effect around 1945 for females in non-conurbations (Table 5-11, page 157).

Can changes in air pollution explain observed patterns in COPD mortality trends and period effects?

Examination of actual trends in ambient black smoke and SO_2 confirmed expectations of changes resulting from the 1956 Clean Air Act. It is argued that these changes are the most likely causes of the observed patterns in COPD mortality. There are a number of arguments in favour of this interpretation:

- *A priori* hypotheses of the most likely patterns in COPD mortality if air pollution were an important factor were generally realised, as discussed in preceding paragraphs.
- The time period during which a decline was seen and the relative differences between conurbations and Greater London compared with non-conurbations in COPD mortality were similar to those for air pollution especially for black smoke.
- Quantitative age-period-cohort analyses suggested that the relative levels of the period effect in conurbations and Greater London relative to non-conurbations were similar to relative levels of air pollution in these areas, particularly for black smoke but less so for SO_2.

Consistency of findings of COPD mortality trends and period effects with research into health impacts of particulates and SO_2 reductions on respiratory mortality

No other analyses assessing health impacts of the 1956 Clean Air Act were found during the course of this PhD. The most relevant comparisons for this study are the natural experiments relating to a ban on bituminous coal in Dublin[308] and restrictions of sulphur content of fuels in Hong Kong,[309] which both have information on respiratory mortality effects.

Restrictions on coal sales in Dublin, 1990

The ban on the sale and distribution of bituminous coals within the city of Dublin in 1990 resulted in a fall in black smoke levels from averages of 50.2 $\mu g/m^3$ for 1984-1990 to 14.6 $\mu g/m^3$ for 1990-1996 and average SO2 levels from 33.4 $\mu g/m^3$ to 22.1 $\mu g/m^3$. In a Poisson regression, this was associated with a 16.1% fall in respiratory disease mortality, which changed little (15.5% reduction) after adjustment for temperature, relative humidity, day of week, respiratory epidemics and trends in respiratory mortality in the rest of Ireland. The authors[308] noted that effects on total mortality (an 8% reduction) in Dublin were very similar to those that might be predicted using a study of the effects of a fall in ambient particulate levels following a strike at a steel mill in Utah,[322] but were higher than those predicted from studies of acute effects of air pollution.[323]

These results confirm important chronic effects of particulate and SO_2 outdoor air pollution on respiratory mortality and it is reasonable to assume impacts on COPD mortality (comprising a third of respiratory mortality in England & Wales in 1991-95[15]). Using the available area-based data (Table 5-3, page 119) the average level of black smoke for conurbations in England & Wales in 1970/1 was 52.6 $\mu g/m^3$ – comparable to that in Dublin in 1984-1990 – while the average level in non-conurbations was 35.9 $\mu g/m^3$ – over twice as high as that for Dublin in 1990-1996. Simple comparison of respiratory death rates in Dublin 1984-1990 to 1990-1996 using data provided in the Dublin paper[308] gave a ratio of 1.19. In the temporal analyses presented in this PhD, the ratio of period effects in 1970/71 between conurbations and non-conurbations was 1.43 for males and 1.58 for females (Table 5-15, page 169) – i.e. much higher than in Dublin.

It is difficult to interpret direct quantitative comparisons between the Dublin study and the current study, because methods were very different – the temporal analyses here adjust for smoking (by removal of cohort effects), while the Dublin study adjusts for meteorological and temporal effects, comparisons were made between areas for the same time period in the temporal analyses but relate to the same area in different time periods in Dublin, the Dublin analyses relate to respiratory mortality but the temporal analyses relate to COPD mortality. Also, area-based air pollution levels are not currently available for the UK over the time period of the temporal analyses in this PhD (air pollution data in the Air Quality Archive relate to monitoring stations).

If direct comparisons between results from this PhD and the Dublin study are valid, the higher effects seen in this study could relate the presence of much higher levels of SO_2: levels of SO_2 in England & Wales were 79.6 $\mu g/m^3$ in conurbations and 73.8 $\mu g/m^3$ in non-conurbations in 1970/1, over twice as high as levels seen in Dublin before and after the ban of sale of coal. Another explanation for the higher effects in England & Wales include a non-linear association between particulate exposure and COPD. The latter would be consistent with findings in the AHSMOG study in non-smoking Californian Seventh Day Adventists[264] where there was a non-linear relationship between increasing total suspended particulate (TSP) exposure and COPD symptoms – increasing exposure from 1,000 hours/year at >100 $\mu g/m^3$ TSP to >200 $\mu g/m^3$ increased relative risks from 1.04 to 1.30. Other plausible possibilities are greater effects of air pollution on COPD than respiratory mortality, that adjusting for smoking increases the observed association between black smoke and mortality, or that air pollution only explains part of the observed COPD mortality trends in England & Wales in the 1950s to 1970s (i.e. that exposures to other risk factors or types of treatment available were different from those in Dublin in the 1990s).

Restrictions on sulphur content of fuel oil in Hong Kong, 1990

Restrictions were introduced in Hong Kong in July 1990 obliging power plants and road vehicles to use fuel oil with a sulphur content of less than 0.5% by weight. Average ambient SO2 levels fell from 44.2 $\mu g/m^3$ in the year preceding the restrictions to around half this level one year later, maintained up to five years afterwards.[309] No marked changes in levels of ambient PM_{10} or other pollutants were seen following the restrictions. Using Poisson regression adjusting for time trends, seasonality, temperature and relative humidity, the results were interpreted as a reduction of 2% in

annual respiratory mortality (in ages 15+ years) for every 10 $\mu g/m^3$ fall in SO_2 – effects double that seen in the same study for all-cause mortality.

These results suggest important effects of ambient SO_2 on respiratory mortality, independent of particulates and it is reasonable to assume that impacts on COPD mortality were at least similar. Again, there are differences in methodology between the Hong Kong study and analyses in this PhD making quantitative comparisons difficult. The SO_2 levels in Hong Kong were much lower than those seen in England & Wales in the 1950s to 1970s (Table 5-3, page 119). If the reduction of 2% in annual respiratory mortality (in ages 15+ years) for every 10 $\mu g/m^3$ fall in SO_2 calculated from the Hong Kong results[309] can be applied to COPD mortality in the UK and at higher ranges of ambient SO_2, then, for example, a 14% difference in mortality might be expected between conurbations and non-conurbations in 1958/9 when average SO_2 levels were 169.5 $\mu g/m^3$ in conurbations compared with 101.3 $\mu g/m^3$ in non-conurbations and the relative risks of period effects for COPD mortality between these two areas at this time were 1.65 for men and 1.84 for women (Table 5-15, page 169).

Similar explanations for the apparent higher impacts in the UK as for the Dublin comparison above can be employed: particulate levels in England and Wales were several times higher than those in Hong Kong, there may be a non-linear effects of SO_2 with greater effects at higher levels, air pollution may have greater impacts on COPD than respiratory mortality, adjusting for smoking (through removal of cohort effects) may increase the observed association between black smoke and mortality, or air pollution only partly explains the COPD mortality trends in England & Wales.

Consistency of findings with air pollution trends

The interpretation that changes in air pollution can explain observed trends and period effects in COPD mortality is consistent with other observations. Chronic bronchitis was known as the 'English disease' from the early years of the 20th century[29] as the disease had a high prevalence here, even before smoking became common[3] and at a time when air pollution levels in most urban areas in the UK were at their highest. Urban-rural differences in bronchitis deaths (in both men and women) were also apparent in the late 19th century prior to the start of the smoking epidemic.[29] Both air pollution (Figure 5.1, page 113) and COPD mortality[324] fell in the first half of the 20th century.[27]

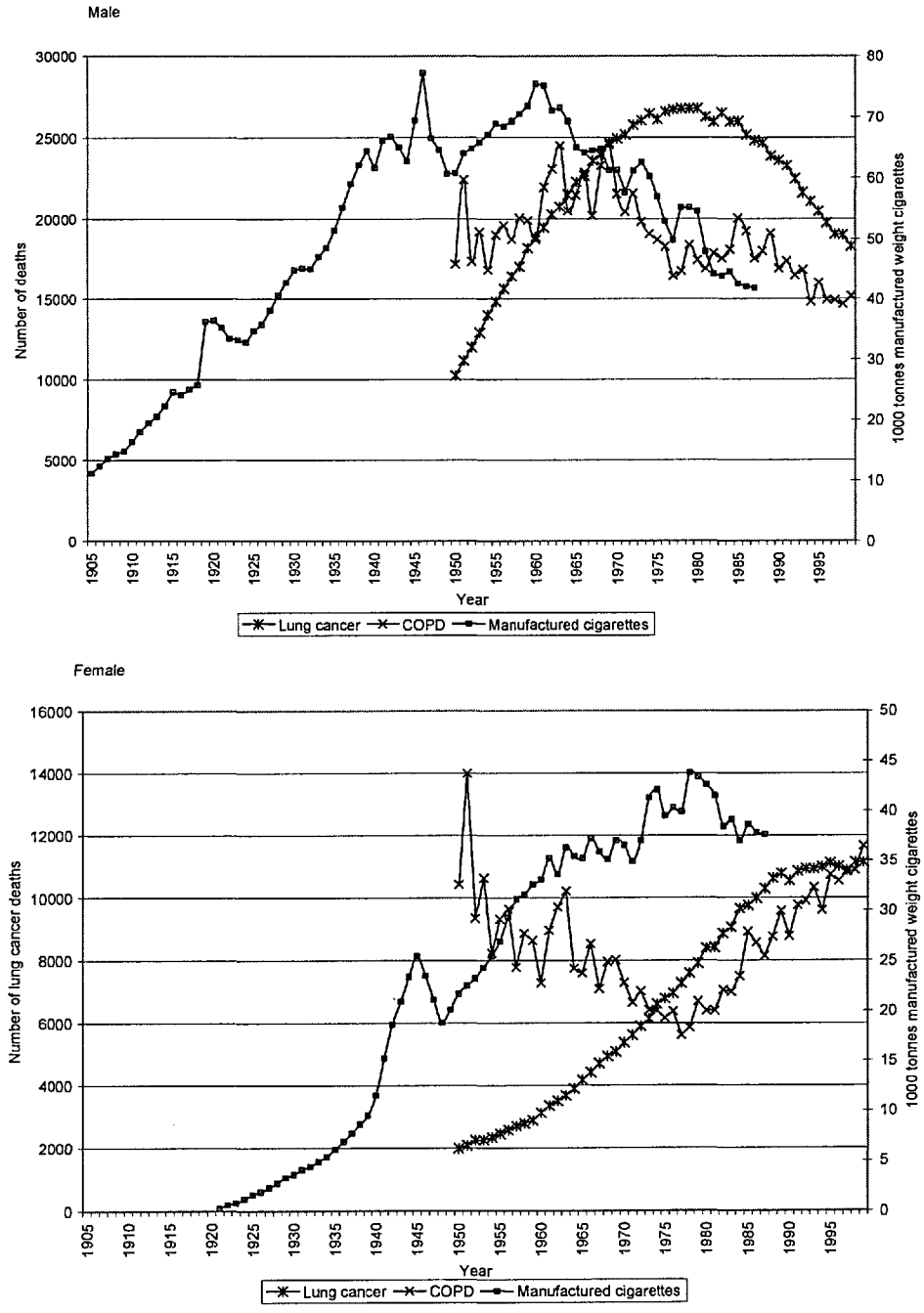
Holland remarked in 1979[297] that air pollution concentrations had reduced so significantly in the previous decade that they no longer had a measurable effect on health. This impression was also held by other authors at the time.[325] This perception is consistent with the temporal analyses presented in this PhD, which suggested that impacts of air pollution were readily visible in simple analyses of COPD mortality until the late 1970s. It is not wholly consistent with Holland's remarks, however, as he went on to suggest that health effects were not discernible at annual mean levels in excess of 140 $\mu g/m^3$ black smoke in the presence of SO_2 levels greater than 180 $\mu g/m^3$ – levels last seen in the UK in the late 1960s (Figure 5.3, page 118) rather than mid 1970s – these remarks have been disputed by other authors.[264] However, Holland's remarks[297] were made in the context of a review of studies using less sophisticated designs and analytical methods than those conducted in more recent years. Holland himself commented on the limitations of the studies reviewed in terms of accounting for other environmental, economic, social and meteorological data. It may be that once air pollution has fallen below a certain level, more sophisticated designs or analyses are necessary to detect air pollution effects, as the relative impacts of factors became more important. For example, in spatial analyses relating to 1980-99 in this PhD an effect of SO_2 air pollution was demonstrated in multiple regression analyses, but the size of the effect was lower than that seen with deprivation.

Holland et al[297] also argued that increases in deaths were only discernible with acute air pollution episodes where black smoke exceeded 500-800 $\mu g/m^3$ together with SO2 of more than 700-1,000 $\mu g/m^3$ (both 24 hour averages) – these would relate to those experienced up to the early 1960s in this country (Table 5-4, page 120). This is consistent with the findings of the BAMP analyses, where period effects detected in the same years as known air pollution episodes in Greater London were only detected up to 1962, although as discussed in the section on BAMP methods (page 285 below), a single air pollution episode may not necessarily have registered as a period effect.

Assigning an important role of air pollution in COPD mortality is also consistent with observed trends in COPD and lung cancer mortality in relation to trends in tobacco consumption (Figure 8-1, page 258). Trends in tobacco consumption in Great Britain in the 20th century were closely paralleled by lung cancer mortality trends around 20 years later, COPD mortality trends did not show this pattern until around 1975 (females) or 1980 (males). In fact, in females, numbers of COPD deaths fell between 1950-1975

when female tobacco consumption over the preceding 20 years was rising. Female COPD deaths then started to rise after 1975, and this did parallel tobacco consumption patterns 25 years previously (Figure 8-1).

Figure 8-1 Tobacco cigarette consumption (tonnes weight) in Great Britain 1905-1987 and number of deaths from lung cancer and COPD 1950-1999 in England & Wales for ages 15+ years



Manufactured cigarette consumption in males shows lower but parallel trends to all tobacco consumption. Female tobacco consumption has always been almost exclusively limited to manufactured cigarettes.[113]
Source: Tobacco data from UK Smoking Statistics[113]

A more detailed analysis by Lee et al[27] using a better measure of cumulative smoking found that lung cancer time trends in the 20th century in England & Wales closely paralleled cumulative constant tar consumption, as did emphysema deaths but not deaths from COPD excluding emphysema. One interpretation of this is that cigarette smoking lagged by about 20 years has always been the strongest determinant of lung cancer mortality, but it did not become so for COPD mortality until air pollution had fallen to lower levels. (The possible role of smoking will be discussed in more detail later).

Coherence of research presented with other studies assessing associations between COPD and ambient particulate and SO_2 levels

Few studies have specifically looked at COPD as an outcome in relation to air pollution – presumably because of the numbers involved analyses involving all respiratory diseases are more common, although lung cancer mortality is sometimes examined and the numbers of deaths from this are not much larger than for COPD. As a result, and perhaps surprisingly, the attributable risk of air pollution on COPD mortality has not been estimated, although it has for lung cancer mortality.[310,317] Most studies report on air pollution effects on respiratory mortality, half of which might be expected to be due to pneumonia and a third due to COPD if following similar patterns to those seen for England & Wales in 1991-95[15] (using underlying cause of death). Where studies have used COPD as an outcome, the effect estimates – at least for acute effects – are higher than for total mortality.[137]

There is evidence for effects of ambient air pollution on small airway remodelling, reduced lung function, COPD incidence, hospitalisations for COPD and acute effects of air pollution on COPD mortality with effects seen more consistently for particulates than for SO_2 and this is discussed in more detail next. This is coherent with the suggested role of air pollution changes in explaining the findings specific to COPD mortality from the temporal analyses.

Note that some of the studies quoted below have used differing measures for particulates. Dockery & Pope[38] suggested that black smoke could be considered as approximately equivalent to PM_{10}, and that PM_{10} could be considered approximately equal to a factor of 0.55 multiplied by TSP. An alternative interpretation is that black smoke is equal to PM_{10} as a lower bound and TSP as an upper bound.

Small airway remodelling

Autopsy samples from lungs of non-smoking women without a history of chronic lung disease, occupational dust or biomass fuel exposures but living in Mexico city, which has high levels of particulate air pollution, were found to have visible dust, fibrosis and excess muscle in small airway walls to a much greater extent than that seen from control subjects living in Vancouver, a city with low particulate levels[326] The changes seen in the lungs of the Mexican women were similar to those seen in chronic airflow obstruction and the authors considered that they were comparable to changes seen in cigarette smokers or workers with high level exposures to dust.[326] This provides some biological plausibility for a relationship between air pollution and COPD, although it should be noted that this study[326] did not have information on environmental tobacco smoke exposure.

Reduced lung function

Both cross-sectional and a small number of cohort studies have found associations between ambient air pollution and reduced lung function and increases in decline of lung function over time in children and adults.[139,140] The associations have usually[141,231,327,328] but not always[328] been seen with both particulates and with SO_2. Decrements seen in cross-sectional studies correspond to 3-5% drop in FEV_1 for a 50 $\mu g/m^3$ increase in particulates (expressed as TSP[328,329] or PM_{10}[327]) but smaller effects for SO_2.[327,329]

COPD incidence

A small case-control study in Athens, Greece[330] found that long-term exposure to black smoke was associated with higher odds of having COPD (using self-report, confirmed by spirometry) and that the relationship was stronger for exposures over the past 5 years than for exposures over the last 20 years. The AHSMOG study[264] suggested a strongly significant dose-response relationship between COPD symptoms (the definition included a doctor's diagnosis of COPD diagnosis or asthma) and cumulative TSP exposures over an 11 year time period. A relationship with SO_2 was only seen at higher levels of exposure (but exposure measures used were hourly levels that are not directly comparable with annual means used here).

COPD hospitalisation

The APHEA (Air Pollution and Health, a European Approach) time series analyses[138] found associations between hospitalisations for COPD in six European cities and ambient black smoke and SO_2 levels of similar magnitude. The relative risk of hospitalisation were 1.04 (95% CI 1.01 to 1.06) for black smoke and 1.02 (0.98 to 1.06) for SO_2 for 50 $\mu g/m^3$ increases lagged 1-3 days . The APHEA 2 study[331] did not find a statistically significant association with SO_2 for COPD hospitalisations independent of particulates, but noted that correlations with particulates were very high, which may account for the failure to detect an association.

Acute effects on COPD mortality

Acute effects (same and preceding days) of particulate and SO_2 air pollution on COPD mortality have been seen in several studies. Effects were seen for each pollutant when treated separately in a study in Philadelphia,[137]but SO_2 effects were not significant when TSP was included in the model, which included adjustments for meteorological and seasonal variations. Death (from all causes) was significantly associated with PM_{10} in the preceding 7-15 days in a cohort of COPD patients in Barcelona,[332] but SO_2 was not investigated in this study. An increase in obstructive lung disease death deaths and hospitalisations compared with the same week of preceding years was seen during an air pollution episode in London in December 1991[298] in which particulate and NO_2 levels were raised, and to a lesser extent SO_2.

Which ambient air pollutant is important in COPD mortality?

The temporal analyses found that patterns in black smoke more closely resembled patterns in COPD mortality than SO_2 in terms of timing and relative levels between different areas. In terms of timing, SO_2 levels fell later than COPD mortality or black smoke. The relative differences in age-standardised COPD mortality rates, with conurbations having highest levels, Greater London intermediate and non-conurbations (Figure 5.8, page 139) lowest, were more consistent with relative black smoke levels in these areas than SO_2 (Figure 5.3, page 118). Further, quantitative analyses of the relative risk of period effects in conurbations and Greater London with respect to non-conurbations compared better with relative ratios for black smoke in these areas than relative ratios for SO_2 (Table 5-15, page 169 – although, as discussed on page 284, these comparisons should be treated as illustrative rather than definitive). The conclusion that falls in particulates were more important in explaining changes in

COPD mortality will only hold if chief air pollution effects are in the short-term (acute) or short-medium term (a few years) as it seems likely from the limited evidence available that both pollutants had fallen over the first half of the 20th century (Figure 5.1, page 113), although this was less clear for SO_2.[296] The issue of lag in air pollution effects is discussed on following pages.

Unfortunately, it was not possible to analyse black smoke in the spatial analyses due to problems with the interpolated black smoke dataset, but a relationship was found between SO_2 and the independent COPD risk. It seems unlikely that SO_2 levels were solely acting as an indirect measure of black smoke as the association between the two pollutants from 59 long-running independent monitoring stations in the 1980s and 1990s was found to be relatively weak (Figure 6.6, page 182), with correlation coefficients around 0.3. However, the low correlations may be due to selection factors resulting in a station being long-running. Again, as relative differences between areas are likely to have shown some similarity over time, it is not possible to determine whether the observed relationship with SO_2 is a short-term relationship or reflects a lagged effect related to similar relative differences between areas.

It seems unlikely that pollutants not considered in these PhD analyses such as nitrogen oxides and ozone, were responsible for observed trends in mortality over 1950-99. These pollutants, chiefly related to emissions from combustion engines and have only been monitored from the 1970s onwards.[333] Ambient NO_2 levels rose slightly from the 1970s to peak around 1990[333] followed by a fall – very different trends to those seen for COPD mortality. Ozone showed almost no change or a small upward trend at most sites 1972-1995[334] with rural levels of ozone are generally higher than urban levels (as urban ozone is scavenged by nitrogen oxides, present in higher quantities in urban areas). Again, this is different from observed higher levels of COPD mortality in conurbations and Greater London. One of the few studies to consider the impacts of nitrogen oxides, ozone and carbon monoxide did not find a relationship between these pollutants and mortality in a cohort of patients known to have COPD.[332]

It seems likely that chronic exposures to both particulates and SO_2 have impacts on COPD mortality – although this may be mediated through sulphate particles and acid aerosols (of which SO_2 and black smoke are major determinants[335]). Black smoke and sulphur dioxide have been found to be the major determinants of acidic aerosol

composition in Greater London[335] and the similarities between air pollution and COPD mortality may be related to acidic aerosol formation, rather than particulates. However, the health effects of acidic aerosols are reported much less frequently than effects of particulates or SO_2. Also, information from published studies on health effects of mixtures of air pollution is limited,[336] as most studies estimate the effect of single pollutants, but the current limited evidence suggests that effects of different air pollutants are additive rather than synergistic.[336]

The relative importance of particulates and SO_2 cannot be answered by the analyses presented, nor is it possible to determine the nature of the combined effects. Inferring definitive quantitative risk relationships between COPD mortality and ambient particulate and SO_2 concentrations using the observational analyses presented to make comparisons will not be attempted as it would overinterpret the data in the absence of more detailed information on confounding factors. The literature is not particularly helpful as to the relative importance of particulates or SO_2 on COPD mortality as COPD is rarely studied as a separate outcome in studies of either acute or chronic health effects of air pollution, even though numbers involved are comparable to those for lung cancer mortality.

If air pollution is responsible is this an acute or lagged effect?

Assuming black smoke and SO2 have important effects on COPD mortality, the effects may be a summation of acute effects related to higher levels or to air pollution episodes, or could be due to longer term, chronic effects or both.

Acute effects of particulate and SO_2 air pollution on COPD mortality have been documented in a study in Philadelphia,[137] with the strongest associations with the mean pollution of the current and preceding days. However, studies on chronic effect of air pollution have generally found higher effects[308] – at least on total mortality – than might have been predicted from short-term associations of air pollution with mortality.[323]

Most time trends studies limit the investigation of lags to a few weeks, but effects of air pollution episodes may have impacts on COPD mortality lasting months, which could be termed 'delayed acute'. Bell and Davis[294,337] suggested that the 1952 London smog was responsible for increased mortality up to four months following the air pollution episode,[337] either through a direct effect or through increasing susceptibility to

infections such as influenza. This would suggest an excess of at least 12,000 deaths (from all causes) rather than the more customary 4,000 deaths attributed to the smog.[337] It is unclear how many of these excess deaths were due to COPD. About a quarter of the excess deaths in the first two weeks from the start of the smog were attributed to bronchitis,[293] but the published data do not specify whether this refers to any bronchitis deaths (codes for 'bronchitis' and 'chronic bronchitis' were available in ICD6, used to code deaths at that time). In contrast, an investigation by Schwartz[338] into harvesting effects using mortality data from Boston with up to four months lag argued against an effect of PM_{10} air pollution on COPD mortality beyond one month. The analysis by Schwartz[338] provided some evidence that short-term impacts of air pollution on COPD mortality relate to deaths brought forward, termed mortality displacement, rather than excess deaths. Schwartz's analyses relate to the USA in the 1980s, with lower air pollution than in the UK than in the 1950s-70s. If the results can be transferred to the UK, then any effects of particulates on COPD mortality would have to be due to more chronic effects.

An intermediate-term effect (lasting months to 1-2 years) of chronic air pollution on COPD, which could be termed 'maintenance of chronicity', is plausible but has not been directly investigated in published studies to date. Here, high background ambient levels of particulates and other pollutants may create a pro-inflammatory state in the lungs,[135] which would worsen pre-existing oxidant stress seen in the lungs of COPD patients.[154] This could result in increased sputum production,[136] associated with increased airway resistance and mucus plugging of smaller peripheral airways and would increase susceptibility to infections.[136]. Mucus hypersecretion is itself associated with increased COPD mortality[85,86] (although less strongly than airways obstruction[86]), while infections are associated with (and probably the cause of) at least 60% of exacerbations,[339] which are in turn related to high mortality rates.[340]

It also possible that air pollution has a long-term cumulative effect, through the mechanisms mentioned in the preceding paragraph and through deleterious effects on lung function. Many cross-sectional studies have found an association between air pollution and lower lung function as have the few published cohort studies to look at this.[139] The complex ecological analyses by Lipfert and Morris[341] looking at ambient air pollution in the USA between 1960-97 argued against a long-term effect on all-cause mortality, but whether this holds for COPD mortality is unclear. Also, air pollution data

were obtained from different sources and the attributable risks varied by data source. The declines in COPD mortality from the temporal analyses were located to period effects. This and the timings of the falls in air pollution argue against the predominant impact of air pollution on COPD mortality being through reduction of lung function in childhood,[139] resulting in increased COPD prevalence and thence COPD deaths.

The time period over which air pollution affects COPD mortality is important for the interpretation of temporal analyses in particular. Temporal analyses would certainly be consistent with predominant summed acute effects, 'delayed acute' effects or 'maintenance of chronicity' effects of black smoke acting over a few days to up as much as 5 years. They would also be consistent with predominant long-term cumulative effects that are lagged over a long time period – for example, the period effect showing a levelling off of the decline in COPD mortality in 1977-78 may relate to black smoke or SO_2 levels falling to some 'threshold' range in the 1960s, below which effects become undetectable in descriptive analyses. Spatial analyses are less affected as some constancy in relative differences between areas in ambient SO_2 has already been assumed (by using estimates of ambient levels of SO_2 for 1996 in analyses of mortality in the 1980s and 1990s) and this would hold across different time frames of action.

Were improvements in air pollution (and presumed associated falls in COPD mortality) a result of the Clean Air Acts?

The temporal analyses have been related to the 1956 Clean Air Act, assuming that the observed falls in ambient air pollution levels were related to this piece of legislation.

Clean air legislation

There have actually been three Clean Air Acts in Britain in 1956, 1968 and 1993. The 1968 Act extended the 1956 Act, in particularly introducing the use of tall chimneys for industries burning coal, liquid or gaseous fuels. The 1993 Act consolidated changes and amendments to the previous two acts, prohibiting the emission of dark smoke from chimneys including domestic premises. From the 1970s, when the UK joined the European Union, European Community directives has increasingly been introduced to control industrial and transport-related air pollution. In response to European legislation, the UK government passed the Environmental Protection Act (1990) and the Environment Act (1995). The latter resulted in the publication of the National Air

Quality Strategy in 1997, with objectives for seven pollutants set out in the Air Quality (England) Regulations 2000, the Air Quality (Wales) Regulations 2000 and the Air Quality (Scotland) Regulations 2000.[244]

Impact of the 1956 Clean Air Act

From a historical perspective, government action on air pollution has a somewhat chequered record. The first commission on air pollution set up to tackle air pollution in London,[35] which has suffered from bad air pollution for centuries, was actually in 1285 and the first recorded official action on air pollution was a proclamation banning sea-coal, issued in 1306 (but largely ignored).

It seems likely that the 1956 Act (described in Chapter 5, page 114) accelerated improvements in air pollution that had been occurring since the start of the 20th century, rather than being the key reason for the reduction in air pollution levels. There were certainly sharp falls in smoke emissions starting in 1955, when the 1956 Clean Air Act was under discussion and ending in the early 1960s, mainly due to reductions in industrial emissions through improving efficiency of coal combustion or switching to oil.[37] Improvements in domestic emissions of smoke occurred more slowly. However, air pollution had been improving over previous decades in most parts of the country and this can be dated to around 1900 in London (Figure 5-1, page 113) for which figures are available. The National Survey of Air Pollution 1961-71[37] suggested that the act could be credited with

> *"having greatly accelerated developments that would, however, of themselves, have taken place slowly but inevitably. In fact, like other successful legislation, the Clean Air Act was 'swimming with the tide' of industrial development."*

Auliciems and Burton[36] found no evidence to support the strong claims of effectiveness made on behalf of the 1956 Act using an analysis of data from Kew from 1922/3 to 1970/1 and for the 1960s from three cities without major smoke control programmes.

The Clean Air Acts applied to smoke but not sulphur dioxide. The only relevant legislation for sulphur dioxide at the time of the first two Clean Air Acts was the Alkali Acts, which applied to registered industrial plants. However, smoke control measures contained in the Clean Air Acts indirectly encouraged the reduction of sulphur dioxide levels by encouraging the use of oil, electricity and gas (which was by the 1950s was changing to sulphur-free natural gas, rather than gas made from coal and coke).[296]

Smokeless fuels only resulted in a minimal reduction in sulphur dioxide emissions. The first general legislation to limit emissions of sulphur dioxide was introduced in London in 1971 and brought into force in 1972 when the City of London (Various Powers) Act prohibited the use of fuel oil containing more than 1 per cent sulphur in the city.[37] Virtually all premises had converted to low sulphur fuels by 1983 and this was reflected in a decline in sulphur dioxide levels in the City.[296] Another factor reducing sulphur dioxide levels was the closure of electricity generating stations in cities such as London[296] (sometimes resited in rural areas).

Can other factors can explain the observed mortality trends and period effects in COPD mortality?

Any single factor explaining the observed mortality trends timing of period effects would need to obey all of the following four criteria

(i) decline over the late 1950s to late 1970s (or increase, if a protective factor) if effects are contemporary OR identifiably decline (or increase) over a twenty-year period in earlier years if effects are lagged (i.e. 1950-1970 if a 10 year lag, 1920-40 with a 40 year lag etc.) AND

(ii) to have had a greater effect in conurbations than non-conurbations and an intermediate effect in Greater London throughout the time period of action (either higher levels or more susceptible populations in the more urban areas) AND

(iii) to have had similar effects on males and females (either similar exposures in each or different exposures with different susceptibilities) AND

(iv) be localised to a period effect i.e. affecting all age-groups at a particular period in time

The risk factor could have either an effect on mortality or incidence of disease or both, but a factor affecting incidence alone would be expected to have a lagged impact on mortality. All these criteria were met by air pollution. These criteria will now be used to assess whether other factors could plausibly explain the differences in mortality trends and period effects in COPD mortality.

Could changes in smoking habits explain observed trends and period effects?

Period effects and trends in COPD mortality compared with lung cancer mortality

As discussed above (pages 247-248), the lag period for effects of smoking on lung cancer and COPD mortality are probably reasonably similar at around 20 and 25 years respectively. Both diseases have high attributable risks (>80%) for smoking – but that for lung cancer is higher than for COPD. It seems most plausible that, given the close relationship to cumulative smoking of both COPD and lung cancer that both diseases would have similar underlying period and cohort effects related to smoking, but that given the lower attributable risk for smoking for COPD, other factors would explain any discrepancies in period and/or cohort effects between the two diseases. In the analyses, presented here cohort effects were seen to be very similar, but period effects very different between the two diseases as set out next.

BAMP age-period-cohort analyses presented here found that cohort effects by sex were extremely similar in timing for COPD and lung cancer. These cohort effects were consistent with known smoking trends by sex.[26,113] In contrast, period effects for COPD mortality were different from period effects for lung cancer mortality. BAMP analyses detected marked period effects for COPD mortality between the late 1950s and late 1970s, but no consistent period effects for lung cancer (period effects for lung cancer were seen in the early 1950s in three of six area-sex combinations, two of these three being in maxima and the third a minimum, see Table 5-10, page 156).

Relative (quantitative) differences between areas in terms of rates and period effects also differed between COPD (Figure 5-8, page139 for rates, Figure 5-12, pages 148-150 for period effects) and lung cancer mortality (Figure 5-9, page 140 for rates, Figure 5-13 on pages 151-153 for period effects). Rates for COPD mortality (Figure 5-17, page 164) and period effects (Figure 5-16, page 163) suggested convergence towards non-conurbations. While rates of lung cancer mortality for males and females in Greater London relative to non-conurbations also showed a fall (Appendix, Figure A-1, page 353), those in conurbations relative to non-conurbations showed little change until 1980 onwards when there was a modest rise, particularly in females. Relative differences between rates of lung cancer mortality in different areas were similar to relative risks for period effects for lung cancer using a classical age-period-cohort model assuming similar age and cohort effects but different period effects in all areas (not shown).

However, just observing that there are differences in period effects between lung cancer and COPD mortality, despite similar cohort effects in age-period-cohort analyses is not conclusive proof that smoking trends cannot explain the observed mortality trends and period effects. A more detailed discussion of changes in smoking follows.

Period effects in COPD mortality in relation to observed trends in smoking prevalence

The main form of tobacco use in the UK over the time period of the temporal analyses were manufactured cigarettes. In the UK, women have almost exclusively smoked manufactured cigarettes since smoking first started to become popular among women in UK in the 1940s.[113] Men smoked pipes and cigars prior to the 1900s and first started smoking cigarettes at the end of the 19th century. Cigarettes rapidly became the most popular form of smoked tobacco and from the 1940s onwards, approximately 80% of men who smoked tobacco smoked manufactured cigarettes.[113] Hand-rolled cigarettes have accounted for only about 6% of total annual cigarette sales by weight from the 1960s onwards. Filter tip cigarettes became popular from the 1960s onwards,[113] but Lange[86] found no difference on FEV_1 decline between plain and filter tips over a five year follow-up study of over 4,000 Danish smokers in the 1970s to 1980s. The main impact of filters was to reduce tar yields, which are discussed below.

The prevalence of smoking fell continuously in the second half of the 20th century from just under 80% in males in the UK[113] to 32% in England in 1999[342] and 34% in Scotland in 1998.[234] In women, data on smoking prevalence are first available from 1956[113] when smoking prevalence was 42%. Prevalences in women remained just over 40% until 1979 then fell[113] to 27% in England in 1999[342] and 32% in Scotland in 1998.[234]

It seems unlikely that smoking prevalence alone can explain the trends in COPD mortality as it does not obey all four criteria set out above. In particular:

(i) Smoking prevalences for women do not show the required declines over the 1950s-1970s– even if considering lags in women as prevalences did not start to decline until the end of the 1970s AND

(iii) Similar effects of the factor on males and females would be needed to give rise to similar period effects but time trends in smoking prevalences were too dissimilar in males and females AND

(iv) similar period effects showing declines in prevalence of smoking in males and females would need to be demonstrated but this does not seem plausible for women in the 1940s to 1970s when prevalences were static.

In relation to criterion (ii) relating to a greater effect of the factor in conurbations and greater London, smoking prevalence is generally thought to have been higher in urban than non-urban areas,[343] but whether conurbations had higher smoking prevalences than Greater London is unknown as these historical data were not available.

While trends in smoking prevalence may not explain the observed period effects in COPD mortality, there were marked changes in cigarette composition over the time period of analyses that also need to be discussed as these could plausibly result in period effects.

Reductions in tar content

There has been a reduction in the average tar exposure of smokers over the 20th century from both reductions in tar in cigarette brands available and a switching by smokers to filter-tipped and lower tar cigarettes. Tar yield of all brands of manufactured cigarettes fell slowly between the 1930s and 1970 and more rapidly thereafter.[113] For plain (no filter) cigarettes, sales-weighted average tar yields fell from 33mg/cigarette in 1934-40 to 30mg/cigarette in 1960 and 17mg/cigarette in 1988. For unventilated filter cigarettes sales-weighted average tar yields fell from 24mg/cigarette in the 1960s to 14 mg/cigarette in 1988. Even lower tar levels were seen with filters with air ventilation holes[344] first introduced in the 1960s and expanded tobacco in the 1970s – 'low tar' cigarettes with 8-14mg tar per cigarette and 'very low tar' cigarettes' (<7mg per cigarette). These falls in tar yields would have affected both men and women.

While declines in tar yields did occur over the time period required (criterion (i)), could be expected to have had similar effects on males and females (criterion (iii)) and also could be localised to a period effect (criterion (iv)), it seems unlikely that they would show greater effects in conurbations, intermediate effects in Greater London and lowest effects in non-conurbations (criterion (ii) above) as it seems unlikely that brands smoked would vary to such an extent between conurbations and non-conurbations to explain the relative differences between areas.

Additionally, the decline in tar yields starts earlier but also continues later than the period effects for COPD mortality seen in the temporal analyses and is more gradual. It seems unlikely therefore that a contemporary action could explain the observed period effects in COPD mortality as they would have been expected to start earlier unless tar has a 10-20 year lagged effect on COPD mortality. In the case of a lag, the gradual reductions in tar yields and switching to lower tar brands would certainly be consistent with the diffuse onset of the long-term COPD period effect in the late 1950s and early 1960s. Studies such as those by Lee[117], using data on COPD from the American Cancer Society Study following over 1 million individuals in the USA from 1959-60 over 12 years and from a retrospective case-control study 1966-72 in north-east England, do suggest that smokers of higher tar plain cigarettes have higher COPD mortality than smokers of lower tar filter cigarettes. However, one would have to hypothesise that there is a threshold below which tar has much less effect on COPD mortality and that this was responsible for the period effect in 1977-78. It seems unlikely that this hypothesised threshold cut-off would have been reached in all age-groups and both sexes to result in the sharp levelling off of the decline in period effects observed in all areas and both sexes in 1977-78.

Also, if changes in tar yields had occurred non-gradually enough to give rise to the COPD mortality period effects, one might also have expected changes in period effects in lung cancer mortality. Harris,[344] using the Cancer Prevention Study II in the US which followed approximately one million individuals from 1982 for six years, found higher risks of lung cancer in those smoking high tar non-filter brands at baseline than those smoking medium tar filter cigarettes (but no difference in risk between smokers of low or very low tar and medium tar filter cigarettes). Lee[117] using the American Cancer Society Study and data from north-east England as above, also found that smokers of higher tar plain cigarettes have higher lung cancer mortality than smokers of lower tar filter cigarettes. Changes in tar yield over the 1960s to 1980s might therefore have been expected to produce observable changes in period effects for lung cancer during the period of analysis – possibly during the 1980s to 1990s if a twenty year lag is seen – but the only period effects observed for lung cancer mortality were in the early 1950s and these were inconsistent.

Other changes to cigarettes

Many changes have occurred over the last century in the composition of cigarettes in terms of additives, tobacco curing and delivery mechanisms such as composition of cigarette papers. Tar is a rather imprecise term referring to '*that portion of condensate from tobacco smoke trapped on a Cambridge filter from which nicotine and moisture have been subtracted*'[345] and many changes in delivered chemical constituents have been noted including declines in known carcinogens such as polynuclear aromatic hydrocarbons and increases in others such as nitrosamines (related to changes in tobacco curing[344]). These changes may have affected differences between lung cancer and COPD.

The arguments for these changes are likely to be similar to those for changes in tar and it seems likely that they would also fail criterion (ii) as it seems unlikely that brands smoked would vary to such an extent between conurbations and non-conurbations to explain the relative differences between areas and convergence over time in COPD mortality.

Social patterns in smoking

Period effects in smoking might also be possible in relation to social patterns in smoking common to all age groups such as the gradual reduction of smoking at work and the increasing proportion of smokers in more deprived groups – as individuals from manual backgrounds were more likely to smoke and less likely to quit than those form non-manual groups.[197] However, it is difficult to see how these changes would have had greatest effects in conurbations and intermediate effects in Greater London (criterion (ii)).

Could changes in infections plus improvements in treatment explain observed trends and period effects?

Changes in treatments for COPD in the 20th Century

Many new treatments became available for COPD over the last 50 years, with potential impacts on mortality.

The first antibiotics including penicillin G and streptomycin became available in the 1940s, with chloramphenicol introduced in the late 1940s, erythomycin in the early 1950s, ampicillin and trimethoprim in the early 1960s and cephalosporins in the

1970s.[346] The usefulness of antiobiotics in treating chronic bronchitis was recognised in the 1950s[347,348] and reviews confirm a reduction of mortality from the use of antibiotic treatment in exacerbations.[54]

Long-term antibiotic prophylaxis with a tetracyline (or trimethoprim or sulphonamides) in winter months was used from the mid to late 1950s[349,350] and early 1960s.[351] This dropped out of general use – probably in the 1970s as very few trials of its use were published after this time.[352] A recent Cochrane systematic review[352] found a small but significant effect of long-term antibiotic prophylaxis in reducing days of illness. The authors also suggested[352] that reducing bacterial colonisation might reduce the severity of exacerbations, but did not assess mortality as an outcome. The recent NICE guidelines on COPD[54] were more cautious about the evidence presented in the Cochrane review stating

> '*there is insufficient evidence to recommend prophylactic antibiotic therapy in the management of stable COPD*'.

The first selective β-agonist bronchodilator treatment, isoprenaline, became available as a sublingual preparation – initially for use in asthma – in 1948[353] and as a metered dose inhaler in the 1960s.[354] The impact of short-acting β-agonists on COPD mortality has not been studied.[355] Long-acting β_2-agonists became available in the 1990s[356] – long after the period effects observed for COPD mortality. Short-acting anticholinergic bronchodilator inhalers first began to be used in chronic bronchitis in the mid 1970s[357] – also too late to explain the observed period effects in COPD mortality.

Steroid treatment also became available in the 1950s,[350] but there is no proven benefit on COPD mortality from the use of systemic steroids in exacerbations on mortality,[358] nor from inhaled[54] or oral[54] corticosteroid use in stable COPD.

Methylxanthines such as theophylline have been used in COPD for many years, with the earliest article relating to its use in a Cochrane systematic review of theophylline[359] dating to 1967[360] suggesting introduction in the 1960s and early 1970s. Cochrane systematic reviews have suggested theophyllines are not effective in acute exacerbations of COPD[361] but do have some beneficial effects on morbidity in terms of lung function and exercise tolerance[359] – neither review was able to look at the impact on mortality due to lack of available studies.

Vaccination against influenza has been in use in COPD patients since the 1950s, with the first trial published in 1959.[362] Vaccination against influenza is associated with a 70% reduction in (all cause) death rates in patients with chronic lung disease and in the elderly in influenza seasons.[363] The first randomised controlled trial to investigate the use vaccination against pneumococcal disease in COPD patients was in 1987.[364] Pneumococcal vaccine is associated with an approximately 30% reduction in (all cause) mortality over influenza seasons.[54] Vaccination of COPD patients against both influenza and pneumoccocal disease is associated with approximately 80% reduction in mortality over influenza seasons.[54] The use of influenza vaccine is consistent with the observed drops in influenza mortality rates over 1950-99, especially the fall in size and frequency of peaks in deaths from influenza, and could also help to explain the reduction in year to year variability in COPD mortality (Figure 5.8, page 139).

Long-term oxygen therapy was first introduced in the 1980s[54] and non-invasive ventilation in the 1990s – both of these treatments have been proven to reduce mortality.[54]

Could changes in treatments explain COPD mortality trends and period effects?
Treatment effects on COPD mortality would be expected to occur acutely or in the short-term, to time from the introduction of the drug or vaccine into clinical practice and show increasing impacts as use became more common. In terms of timing, the most likely treatment factors that could explain the onset of a period effect in the late 1950s and early 1960s would be antibiotic use (either acutely or prophylactically), β-agonist inhalers or influenza vaccination (criterion (i)). Additional treatments could explain the continuation of the decline with no additional benefit on population mortality from the anything introduced in the 1970s – although it is less clear why a levelling off of effect should be located so precisely to 1977-78 in all areas and both sexes. It is interesting that no consistent period effect was located that could be associated with treatments with proven effect on mortality[54] such as long-term oxygen therapy (introduced in the early 1980s) and non-invasive ventilation (from the 1990s).

Treatments would be expected to have had similar effects on males and females (criterion (iii)) and to be localised to a period effect i.e. affecting all age-groups at a particular period in time (criterion (iv)).

However, it is very difficult to conceive that *use* of these treatments was greatest in conurbations than non-conurbations, intermediate in Greater London and lowest in non-conurbations throughout the time period of analysis (criterion (ii)). If changes in treatments are to explain the observed trends and period effects in COPD mortality, one would have to hypothesise that *effectiveness* of new treatments was greatest in conurbations, intermediate in Greater London and lowest in non-conurbations due to some susceptibility or interacting risk factor common to the populations in these areas. High ambient air pollution exposure comes again to mind. Leaving this aside, an alternative theory might be increased prevalence and/or severity of respiratory infections including influenza in urban than less urban areas, possibly related to over-crowding, with the difference between areas decreasing over time. In favour of this theory is the observation that year to year variations in COPD mortality, presumably related to respiratory tract infections, were much more marked in the first 20 years of analysis (Figure 5.8, page 139), even when examining rates on a log scale (not shown). It seems unlikely that this was due to influenza alone (unless influenza vaccination was increasingly systematically used in COPD patients) as peaks in COPD mortality coincided with those for influenza mortality in 1951 and 1953, but the relationship became less clear for later years, particularly in men (Figure 5.8, page 139). However, it is less clear why conurbations should be worse affected with infections than Greater London.

In conclusion then, new treatments for COPD cannot wholly explain observed trends and period effect in mortality – their use would meet criteria (i), (iii) and (iv), but require a combination of new treatments and another factor to meet criterion (ii).

Are there other factors that could explain observed trends and period effects?

Other important risk factors for COPD that have changed over the 20th century and could potentially explain the observed trends in COPD mortality are occupational exposures to dusts especially in coal miners[43] and diet (especially fruit or fish consumption[162]), both considered next. While respiratory infections in early childhood are considered a risk factor for later COPD,[177] and might be expected to affect men and women equally unless other factors were in operation, they would be expected to show a cohort rather than a period effect.

Changes in occupational exposures

Occupational exposures to dusts and minerals, which might be expected to be higher in conurbation areas (criterion (ii) above), have improved greatly over the last century (criterion (i) above) and would affect all age-groups exposed (criterion (iv)). Also, crop farming and animal husbandry have been associated with approximately two-fold increases in the risks of reporting chronic bronchitis symptoms[365] and there has been a large decrease in the numbers of people involved in farming in the UK since the 1950s. However, occupational exposures would be expected to have impacts predominantly on men, thereby failing criterion (iii) .

Diet

The National Food Surveys, conducted in the UK since 1940, suggest a sustained increase in fruit consumption over this period,[272,273] an increase in vegetable consumption (excluding potatoes) from the mid 1950s that levelled off in the late 1980s (Appendix, Figure A-15, page 367)[272,273] and an increase in consumption of fruit juice.[366] This would be consistent with criterion (i) – a protective factor rising over time, but one would have to hypothesise that there was an upper threshold beyond which increased fruit & vegetable consumption did not offer additional benefits on COPD or that some other changes in fruit & vegetables (e.g. type, preparation methods etc.) had changed either in the 1970s or earlier if the effect were lagged. Changes in diet would probably have occurred in both males and females (consistent with criterion (iii)) and their effects could plausibly be localised to a period effect (consistent with criterion (iv)). However, it would be harder to explain the relative differences between areas (criteria (i) and (ii)) as the National Food Surveys suggest that households in the London region have consistently had the highest consumption of fruit of all regions – at least since 1955[273] (vegetable consumption is more difficult to interpret as published data include potatoes).

Are combinations of factors needed to explain observed trends and period effects?

Trends in tobacco smoking in terms of prevalence, cigarette consumption or social changes were found to be unlikely alternative explanations for the observed trends and period effects in COPD mortality. Changes in treatment alone or in combination with infections were also considered unlikely to explain observed effects as were changes in occupational exposures or diet.

It is certainly possible that some combination of the factors discussed – air pollution, changes in smoking, treatments, respiratory infections, occupational exposures and diet – are responsible for the observed trends in COPD mortality. Attempting to explain effects on mortality by a single factor (such as air pollution) may well be too simplistic. On the other hand, the simplest explanation – changes in air pollution – would be preferred if following the principle of Occam's razor (*'Pluralitas non est ponenda sine neccesitate'* or 'Plurality should not be posited without necessity').

Other results from temporal analyses – cohort effects for COPD and lung cancer, period effects for lung cancer

Period effects for lung cancer

BAMP analyses found period effects in the early 1950s in three of six areas in RW2, consistent with analyses by Lee.[27] The RW1 analyses were less consistent in terms of direction or timing of period effects detected (Table 5-10, page 156) and due to the smoothing assumptions, probably less reliable for this purpose than RW2 analyses.

Using data for the general population may have obscured period effects differences between smokers and non-smokers

Stevens and Moolgavkar's analysis,[311] partitioning lung cancer deaths into those among smokers and non-smokers, suggested that mortality rates among smokers increased over 1941-75, but that the rate among non-smokers rose from 1941 to 1956-60, then declined until 1971-75 – i.e. the decline in lung cancer rates in non-smokers was coincident with the implementation of the Clean Air Act of 1956. Lee *et al*[27] in an age-period-cohort analysis of lung cancer mortality 1941-85 commented that lung cancer rates in younger age groups from 1950 were not fully explained by trends in cigarette consumption and this was particularly clear in women. They also suggested a decline in the background (non-smoking related) rate of lung cancer, through declines in occupational hazards, air pollution and also possibly through declines in bronchitis, which might itself predispose to lung cancer. The age-period-cohort analyses presented in this PhD relate to a general population and observed effects will be a combination of effects in smokers and non-smokers. It is possible that period effects for lung cancer were inconsistent across areas and by sex because of the combination of smokers and non-smokers with greater effects of the 1956 Clean Air Act on lung cancer in non-smokers.

Cohort effects for COPD and lung cancer mortality

As discussed previously, maxima for COPD and lung cancer mortality in male cohorts in the temporal analyses for this PhD centred on 1899-1903 (taking up smoking in and just after the First World War) and those in female cohorts centred on 1925-9 (taking up smoking in and just after the Second World War) were consistent with cohort trends in smoking with peak lifetime exposures occurring in these cohorts.[26,27] However, the analyses presented in this PhD described an additional maximum centred on 1925-26 for both COPD and lung cancer male cohorts in non-conurbations, which may suggest that the increase of smoking in women also had an influence on smoking in men in these areas. This has not been described previously.

Significance of using the cohort effect to capture most of the influence of smoking

The cohort effect was assumed to capture all or most of the smoking effects in these analyses. This assumption has been made by other authors.[311] The significance of this assumption for results from the temporal analyses in this PhD is that period effects represent effects at least partially adjusted for lifetime smoking patterns (although there was potential for period effects related to smoking as discussed previously).

Relative differences in cohort effects between conurbations, Greater London and non-conurbations in relation to smoking

Age-standardised rates showed highest lung cancer mortality in Greater London than in conurbations until the mid 1970s in males and mid 1980s in females (Figure 5.9, page 140) and lowest rates throughout in non-conurbations. Relative risks for conurbations and Greater London (relative to non-conurbations) were greater than 1 for both cohort (Figure 5-18, page 165) and period (not shown) effects for lung cancer in the classical age-period-cohort analyses, consistent with the lower age-standardised rates of lung cancer (Figure 5-9, page 140). As period effects in lung cancer were few and were not detected in the 1970s and 1980s (as discussed above) it is likely that the cross-over between age-standardised rates for lung cancer in conurbations and Greater London represented differences in cohort smoking patterns. It is unlikely to be due to factors such as air pollution as the timing of the changeover was different in males and females.

Evidence for different cohort effects in different areas was found. The age-stratified analyses suggested differences in rates between areas in a cohort pattern (Figure 5-7,

pages 134-135). The timing of cohort effects in different areas using BAMP analyses appeared relatively similar and therefore unlikely to explain the differences between Greater London and conurbations. However, the cohort effects detected in BAMP analyses were diffuse (Table 5-11, page 157) and could be considered to be consistent with the age-stratified analyses for lung cancer mortality, which suggested a cohort pattern with falls in rates occurring 5-10 years earlier in Greater London than in non-conurbations in males (Figure 5-7a, page 134), but similar timing of cohort effects in females (Figure 5-7b, page 135).

The classical age-period-cohort analyses also supported the hypothesis of different cohort effects in Greater London and conurbations. Firstly, the best-fitting model for lung cancer when only one parameter was allowed to vary was allowing different cohort effects in different areas (Table 5-12, page 159). This may have been related to differences in magnitude of cohort effect between areas and/or in males slightly different timing, (but it may also have reflected a very similar (or absent) period effect in all areas, such that small variations in cohort effects become more important in driving variation in the model). Secondly, quantification of the relative risk of the cohort effect for lung cancer for males and females (Figure 5-18, page 165) suggested that there were differences in the magnitude of the cohort effect between areas. There was a marked fall in relative risk with progressively younger cohorts for Greater London compared with non-conurbations. In comparison conurbations relative to non-conurbations showed a slight decline in relative risk of cohort followed by a marked rise with progressively younger cohorts.

As timings of the cohort effects were reasonably similar (or at most 10 years different from each other), the most likely explanation is that different areas had different average cumulative tobacco exposures. Smoking-related explanations include:

(i) Uptake of smoking and therefore smoking prevalences were different in different areas – there is some evidence for this as cities were thought to have higher smoking rates than country areas[343]) – and these differences have reduced over time.

(ii) Smokers in London started quitting earlier than in conurbations and smokers in non-conurbations quit more than in conurbations (thus increasing the differential between the two) – while analyses suggest that the age at which smokers quit has been falling in successive cohorts[367] no geographical

analyses have been located. If this interpretation is correct, age-stratified rates can be interpreted as suggesting that male cohorts born around 1905-1910 and female cohorts born around 1915 in Greater London were the first cohorts to have lower cumulative smoking exposures than similar cohorts in conurbations.

(iii) There have been changes in the numbers of cigarettes smoked in different geographical areas e.g. smokers in conurbations have cut down less than smokers in other areas or there were changes in the cigarette brands smoked by area – average cigarette consumption was higher in conurbations than rural districts in 1970[343] but it is unclear if this is due to differences in prevalence or differences in numbers of cigarettes smoked by smokers.

However, the results may also reflect biases such as migration affecting London more than conurbations (see page 289). It was not possible to make comparisons with actual smoking prevalence or smoking cessation data as these were not available for either the time period or the geographical areas used in the analyses. National data only are available prior to 1974.[113] Regional data became available from 1974 from the General Household Survey, but these areas did not correspond to the areas used in this analysis.

Temporal analyses did not support cohort effects of air pollution exposures

Another factor with the potential to cause a cohort effect in both COPD and lung cancer mortality is air pollution, with both highest cumulative lifetime levels and childhood exposures likely to be experienced in those born in the late 1890s and early 1900s (the oldest cohort in this study was born 1865-75). An analysis of the 1946 birth cohort[176] suggested that childhood exposures could be important for later respiratory disease prevalence. At age 36 years, atmospheric pollution was one of the best predictors of adult lower respiratory tract problems, in combination with poor home environment, parental bronchitis, childhood lower respiratory illness and later smoking.[176] This analysis found evidence for a cohort effect around the 1900s in males (that was attributed to smoking) but no evidence for a cohort effect at this time in females, suggesting that if present, this is quantitatively a small effect.

Original hypotheses (set out in Chapter 1) also suggested there might be an additional COPD cohort effect in those born 1935-60, related to progressive reductions in air pollution exposure in childhood. Statistical support for this from the data was weak

with inconsistent timing (Table 5-11, page 157) and only seen in females. Any effect from air pollution may have been diffuse and therefore compatible with weak and inconsistent support, and relate to younger age-groups (<65 years by the end of the analysis period) with smaller numbers of deaths, both of which may make a true pattern harder to detect. However, the fact that effects were only seen in females and mainly for lung cancer suggests that the second maxima observed in the data was more likely to be related to other factors such as changes in female cohort smoking patterns rather than an effect of air pollution.

The inferred absence of an effect of air pollution in childhood with COPD mortality concurs with findings from an unpublished case-control study by Strachan and Maheswaran.[144] Using the Longitudinal Study, the study found that childhood exposure to smoke and sulphur dioxide air pollution in the 1930s, assessed using a four-point index of air pollution derived from coal consumption, was not associated with deaths from chronic obstructive pulmonary disease (COPD) or lung cancer deaths sixty years later. Confounders adjusted for included social class, place of residence in childhood and the 1990s at individual level and infant mortality at district level. However, the measure of exposure used in this study was crude.

Methodological issues in temporal analyses

Data issues

Data entry

Great care was taken to minimise data entry errors for manually entered data including double entry and checking of crude rates, but the possibility of data entry errors cannot be excluded – they can also not be excluded from the published source tables (errata source tables in the year following publication in Registrar General volumes were also checked and no amendments to COPD or lung cancer mortality were reported).

Age in age-period-cohort analyses

The similarity of the age effect by area in age-period-cohort analyses may partly rely on the use of the assumption of similar age effects in similar areas to partition mortality data by area for younger age groups published in 20 year agebands into 10 year agebands for 1950-62 using all England & Wales data. However, it seems unlikely that this can fully explain the similarities seen. Younger age-bands comprised small

numbers of deaths and this related to 12 out of the 50 years of analysis. Also, the classical age-period-cohort analyses detected some variability in age effects – when allowing two of three of age, period and cohort to vary suggested better fit with age and period varying for COPD in females and age and cohort to vary for lung cancer in males and females (Table 5-12, page 159). The extent of variability in age in different areas is difficult to estimate and may have been related to minor changes in either timing or size of effects or even less variability in cohort effects in female COPD mortality or period effects in lung cancer mortality.

Air pollution levels

Although data from individual monitoring stations are available, there are no area-based air pollution datasets covering the UK over the last 50 years and the earliest interpolated area-based concentration data come from 1996. Air pollution monitoring sites are generally chosen because of high levels of air pollution, which are near to populated areas. Over the time period of analysis, as stations were concentrated in areas of high population (in 1966, stations were sited in 95% of towns with >100,000 population but only 24% of towns with 5,000-20,000[37]) and were more representative of urban than country areas (in 1968 there were 192 monitoring stations sites in the country and 1122 in town sites[37]). As sites came in and out of use over time, there were not enough long-running stations to restrict analyses to these sites – only three of over 1000 monitoring stations had been in operation during the entire time period 1950-1996. Sites may not therefore be representative of air pollution levels experienced across geographical areas, although it is likely that trends presented will have at least some relationship with the trends in average exposures experienced by individuals in that area. It also implies that the air pollution monitoring data for 1958 to 1971 presented (Figure 5-3, page 118, Table 5-3, page 119 and Table 5-15, page 169) will have more accurately reflected ambient concentrations in conurbations and Greater London, while average ambient levels experienced by individuals in the non-conurbation regions may have been lower than those presented.

Because of concerns about the potential biases, it was decided not to produce simple averages of air pollution levels across all available stations for all 50 years of analysis for use in this PhD. However, averaging across available stations was the basis of the approach by Warren Spring (formerly holders of the Data Air Quality Archive) to provide regional trend data in annual average air pollution for the period 1958-1971 that

were presented graphically in Chapter 5 (Figure 5-3, page 118). This represented a 13 year period and it was considered that the siting of stations would have changed considerably more over the whole 50 year period and therefore have limited the usefulness of the comparisons. These regional data were then averaged across conurbations, Greater London and non-conurbations (presented in Table 5-3, page 119 and Table 5-15, page 169), which, as argued above, may have just combined biases within the data and given a misleading picture, so conclusions – particularly about the size of the relative differences between areas – should be interpreted accordingly. However, in the absence of other data, averaging across stations was the only way in which comparisons could be made between areas used in these PhD analyses to look at the years of greatest change in ambient air pollution levels.

Non-conurbations

The terms 'conurbations' or 'metropolitan counties' implies land use that comprises highly urbanised and industrialised landscape, while 'non-conurbations' implies small towns or villages and rural areas. However, conurbation areas and Greater London also comprise areas of countryside, while non-conurbations will have contained industrial areas (in smaller cities and towns and industries purposely sited in rural areas) as well as the South Yorkshire area, first labelled as a conurbation area during the 1974 local government reorganisation, but included in the non-conurbation areas for these analyses. However, even though the categories of urbanisation were relatively crude, real differences between areas were detected in terms of COPD and lung cancer mortality (Figure 5-8, page 139 and Figure 5-9, page 140) and in air pollution levels (Figure 5-3, page 118) that persisted over time.

The choice of non-conurbations as a baseline was used because these areas were thought likely to have had the lowest exposures to factors likely to influence COPD mortality. Land use has changed over time, with increasing urbanisation of the entire country, which would tend to diminish the differences between conurbation and non-conurbation areas. A rising baseline may account for some of the convergence between conurbations and non-conurbations. It seems unlikely that this fully explains the differences in trends between areas as improvements in lifestyle factors including occupation and environmental exposures would be expected in all areas – this was seen with air pollution, which fell in both urban and non-conurbation areas (Figure 5-3, page 118) while remaining lowest in non-conurbation areas.

Implications for comparisons of classical age-period-cohort area comparisons with air pollution data

Caution must be used in interpreting the comparisons of air pollution data by regions with falls in period effects (Table 5-15, page 169), as discussed above. Firstly, non-conurbations may have provided a non-static baseline in the age-period-cohort analyses. Secondly, the air pollution data chiefly related to urban areas within conurbations, Greater London and non-conurbations (although these urban areas will contain the majority of the population in these areas) and may have underestimated the differences in average exposures between urban and non-conurbation areas. Thirdly, the sites were not chosen at random as outlined above and while giving some information about trends, they are not the same as using interpolated data.

The comparisons presented should be considered as illustrative given the biases inherent in the methods used but it is reasonable to conclude that falls in ratios between areas (conurbations:non-conurbations; Greater London:non-conurbations) over 1958-1971 were seen for both air pollution and period effects in mortality. These conclusions are consistent with that expected, if air pollution declines over time are considered to provide the best single explanation of the observed mortality trends.

Analyses using age-standardised rates

A non-log scale was chosen for the COPD and lung cancer age-standardised rates (Figures 5-8 and 5-9, pages 139-140) as this showed the differences between the three areas most clearly. However, it also had the effect of visually exaggerating the year to year peaks in earlier years when rates were higher and the convergence. However, convergence was not just an artefact of the presentation, as graphs of relative and absolute differences in rates between different areas (not shown) suggested similar conclusions.

Age-stratified results were examined as well as age-standardised rates. This was because age-stratified mortality rates suggested that trends were not parallel across age-groups, particularly for COPD mortality in females (Figure 5-6b, page 133) and lung cancer mortality in both sexes (Figures 5-7a and 5-7b, pages 134-135) and therefore, the use of age-standardised rates was not strictly completely justified. As most of the deaths were seen in older age-groups, the age-standardised results will mainly reflect

those older age-groups. The other advantage of examining age-stratified rates was that additional inferences about cohort effects in different areas was possible.

Statistical techniques

These analyses used two new analytical techniques, both based on methods developed for use in a classical framework by Clayton and Schifflers.[286,368] These methods were implemented in a Bayesian framework using specially written BAMP sofware[281] and this was used to locate the timing of period or cohort effects, but was not able to assess the magnitude of differences between areas. The second new technique employed the usual classical age-period-cohort model, but with age and period on different grids and choosing one or more of age, period and cohort effects to be the same in all areas. Results for different areas were then subtracted from each other to draw inferences about period or cohort effects in one area relative to another. Differences between areas in the classical model may have resulted from differences in timing or differences in magnitude of the effect, which is where combination with information on timing from BAMP analyses was helpful.

Some assessment of the validity of these new techniques was made by comparing results with those from other published analyses and by comparisons with descriptive analyses of age-standardised and age-stratified rates presented within this PhD.

Use of BAMP

The BAMP programme has mainly previously been used to make mortality predictions.[211] The timings of period and cohort effects in BAMP analyses were completely consistent with those located in previous classical age-period-cohort analyses for COPD and lung cancer[26,27,311] and with patterns visible in age-stratified and age-standardised rates. One of the advantages of this Bayesian model over classical analyses on the same grid is that period and cohort effects were readily identifiable on a finer scale than that seen in previous age-period-cohort analyses.[26,27,311]

Use of BAMP to detect change-points

Both visual means and statistical cut-offs (credible intervals at specified cut-offs that did not cross zero) were used to identify change points. Very gradual changes were not easily identified as statistical significant as the rate of change for any particular year or cohort was too gradual. Visual inspection of RW1 plots was particularly useful

identifying these very gradual changes, such as those seen in cohort parameters for males. For example, a cohort effect for those cohorts born around 1900 was visually identifiable from RW1 plots for males dying of COPD in conurbations (Figure 5-12a, page 148) and Greater London (Figure 5-12c, page 149) but this was not picked up by statistical criteria (Table 5-11, page 157). Similarly, it was difficult to time the onset of the decline in the period effect for COPD mortality. There were marked year to year fluctuations (Figures 5-12b, 5-12c, 5-12d, 5-12e, pages 148-150) in the late 1950s and early 1960s, some of which were years with known air pollution episodes and/or high influenza activity (Table 5-9, pages 155-156). Within the current formulation of BAMP, it is not possible to distinguish between maxima reflecting acute peaks (such as those related to air pollution episodes or annual influenza activity) and maxima relating to the onset of a longer term decline. RW1 plots (Figure 5-12, pages 148-150) give the visual impression of a declining trend in period effects in the 1960s and 1970s. The year 1963 is a potential candidate for the onset of this longer term trend as statistically significant maxima were detected in all three types of areas (conurbations, Greater London and non-conurbations) for both males and females (Table 5-9, pages 155-156) using RW2 analyses. However, RW1 plots (Figure 5-12, pages 148-150) also suggest acute maxima for period effects in this year superimposed the previously mentioned longer term decline. Other sources did not identify 1963 as a year with high influenza levels or an air pollution episode (Table 5-9, pages 155-156) – factors that may have given rise to acute maxima – but the winter of 1962-3 was exceptionally cold, which may have affected COPD mortality rates in all ages resulting in the observed acute period effects. A new formulation of BAMP, with the ability to distinguish between acute and longer term trends is currently under development.

When using statistical criteria, RW1 analyses (using first order differences) were particularly sensitive to short-term (year-to-year) fluctuations, but RW2 analyses (using second order differences) were more robust to these short-term fluctuations. For example, using statistical criteria for RW2 analyses to detect cohort effects for COPD mortality in females in conurbations identified the strong cohort effect centred on 1926 (readily seen in Figure 5-12b, page 148). Using statistical criteria on RW1 analyses, the 1926 centred cohort effect was ignored, but visually minor effects were identified for single cohorts born 1867-1877 and 1882-1892 (maxima) and 1880-1890 (minima).

Sensitivity of BAMP to year to year changes in mortality – how large do changes need to be to be detected?

Results suggested that BAMP will only identify period effects in mortality if there is a change of at least 10% with the preceding year. Use of BAMP did not detect a period effect related to the rule 3 changes between 1984-1992 in England & Wales, which dual coding by ONS suggested had increased the numbers of deaths in COPD by 7%.[261] 1968 was a year with high influenza activity and mortality peaks were detected in this year in Greater London (Table 5-9, pages 155-156) – here the excess numbers of deaths in Greater London compared with the preceding year was around 10%. Very prominent peaks in period effects were seen in Greater London in 1951 (a year of high influenza activity), 1952 (the year of the great fog) and 1953 (another year with high influenza activity). Comparing observed numbers of deaths in these years with neighbouring years 1950 and 1954, suggested a 25-30% increase in numbers of male deaths and a 20-33% increase for females (this would be expected to be located to a period effect as it is unlikely that changes in cohort effects would have had a large impact over this short time period). BAMP analyses were found to readily identify the strong period effect in pneumonia mortality between 1984-1992 related to rule 3 interpretation changes (not shown), which resulted in a halving of the number of deaths attributed to pneumonia.[237]

Interpretation of period effects

The detection of a period effect will depend not only on the change in mortality in the year in question, but also what is happening to mortality in adjacent years. For example, it may seem surprising that the infamous 1952 London smog was not identified as a strong period effect in both sexes in Greater London (Table 5-9, pages 155-156), especially as it was visible in descriptive analyses of rates (Figure 5-8, page 139), but surrounding years also had high mortality – the smog followed a year of very high influenza activity, while there was another air pollution episode in 1953 – so the effect of the 1952 smog relative to the surrounding years (on which the second order differences are based) was less marked.

Another issue is that period effects detected could result from a single factor (e.g. an influenza activity) or a combination of factors (e.g. influenza plus an air pollution episode). Inspecting the mortality figures suggests that only part of the period effect in 1952 was due to the December air pollution episode. Contemporary reports suggested that approximately 1,150 of the excess deaths in the weeks during and after the fog in

December and up to the end of 1952 were due to bronchitis,[293] i.e. approximately 12% increase to the year's total mortality for Greater London (using totals in column 4 in Table 5-4, page 120). However, comparing the observed annual COPD mortality numbers (in ages 15+ years) for 1952 with 1950, COPD mortality for the whole year was higher by 16% in females and 24% in males. It is therefore possible that the period effect in 1952 relates to both the December smog and to increased deaths related to the known high influenza activity in the preceding winter.

Similarly a combination of factors with opposite effects on mortality (e.g. a mild winter plus an air pollution episode) may have resulted in a period effect not being detected, or have weakened its statistical significance e.g. 1968 was a year with high influenza activity, which might be expected to increase deaths, plus the removal of acute bronchitis codes from mortality data used, which might have resulted in an artefactual drop in mortality. In the event, maxima (with credible intervals for 10^{th} to 90^{th} centiles for males and 15^{th} to 85^{th} centiles for females not crossing zero) were detected in Greater London only (Table 5-9, pages 155-156).

Because of the non-identifiability problem, period parameters could not be used to examine the extent of year to year variation, only to identify change years. Strictly speaking, it was possible to interpret trends in the RW1 analyses (i.e. to identify a decline over time), but this was only because an arbitrarily chosen set of constraints was used – using a different set of constraints should not have made a difference to change points, but could have changed the interpretation of trends.

Using classical analyses allowing one or more of age, period or cohort parameters to vary across areas and different grids for age-groups and periods

Use of different grids

Using different grids for age-groups and periods resulted in marker artefactual periodicity in the data, such that plots of results were not interpretable if examined by area separately (e.g. conurbations alone). However, subtracting parameters of non-conurbations from parameters for conurbations or Greater London was interpretable as the periodicity artefacts were similar in all areas and therefore removed. Resulting parameters were meaningful using a two-dimensional graph if only one of age-period-cohort were examined, but if two parameters were examined they were not (as two factors were varying at the same time) and it was difficult to interpret resulting trends.

Overlapping cohorts

This analysis suffers from the fact that, because of the choice of differing grids, there is even more overlap between the cohort parameters than that normally found in age-period-cohort analyses. This increased overlap could be taken into account by the choice of an appropriate smoothing prior in a Bayesian framework, but this was not possible within a classical framework. Undertaking the development of the statistical methodology needed was beyond the scope of this PhD, but this is being taken forward separately by Professor Held. The increased overlap will have affected the assumption of independence of parameters and the statistical significance presented (for example, confidence intervals in Figures 5-14 and 5-15 on page 162). This provides another justification for not presenting results with confidence intervals, but concentrating on the patterns suggested between areas.

Consistency with other analyses

If allowing one of age, period and cohort to vary, allowing period to vary across area was selected as the best fitting models for both males and females for COPD mortality. This was consistent with the age-stratified analyses suggesting different period effects in different areas for COPD mortality. The results were also consistent with air pollution in that falls in ratios between areas were seen over time (as discussed above).

Bias in temporal analyses

Migration

The most important bias to consider affecting these temporal analyses is migration. The interpretation of the results has assumed that those dying in particular areas also experienced life-time air pollution levels in those areas. This may not be justified as people move. Bias will be introduced if migrants are healthier than non-migrants or if those who consider that their lungs have been affected by living or working in conurbations to move to non-conurbation areas.

Internal migration

According to the National Statistics website (http://www.statistics.gov.uk/), during much of the 20th century there was a movement of population from the old coal, shipbuilding and steel industries in the north of England, Scotland and Wales to the

light industries and services of the south of England and the Midlands. The regional data on internal migration available for 1976 onwards[369] suggest differences in the scale of effects. London and the South-East experienced the largest inward and outward migrations, with numbers for both at least twice as high as in the North-East, Yorkshire Humberside and Midlands government regions, which had the lowest inward and outward migration. Both in and out-migration showed small gradual increases over time.

Analyses for stomach cancer and stroke mortality[370] suggested that internal migrants in the UK (here defined as those dying in the 1990s in a different county from the one they were a child in the 1930s) had lower mortality risks than non-migrants.

It is therefore possible that the trends in COPD mortality rates and period effects for the three geographical areas analysed (Figure 5-8, page 139) may be partly explained by internal migration particularly in terms of the trend towards convergence over time, but if so, it seems strange that similar effects were not also observed for lung cancer mortality (Figure 5-9, page 140).

Ethnic changes

There have been changes in the ethnic composition of areas over time. The 1950s and 1960s saw an influx of Caribbean migrants chiefly into cities such as London and Birmingham, while the 1970s and 1980s saw an influx of Asians from East Africa and then from South Asia again concentrated in London and major cities. These groups, at least initially, are likely to have different smoking and dietary habits from the original population that may have affected their COPD risk but very little has been published on the differences in COPD risk between ethnic groups. ONS analyses for 1991-93[371] suggested that men (aged 20-64 years) born in the Caribbean and the Indian sub-continent had lower SMRs for lung cancer and all respiratory diseases, but those born in East Africa had higher SMRs for all respiratory disease. If these results also apply to COPD mortality, it suggests that immigration could have tended to reduce the differences between non-conurbations and the other two areas.

Conclusions from temporal analyses

In conclusion, air pollution seems the most likely single factor that can explain the observed the convergence of COPD mortality rates in conurbations and Greater London

towards non-conurbations and the changes in period effects, with changes certainly observed within five years of the 1956 Clean Air Act and possibly occurring earlier given the limitations of the methods used. Other possible candidates cannot fully explain the changes in COPD mortality on their own – these factors include trends in smoking, introduction of new treatments (such as influenza vaccination, antibiotics and β-agonist inhalers), changes in diet and changes in occupational exposures. It is not possible to exclude combinations of a number of factors as the explanation for the observed trends and the results will also have been influenced by biases from internal and external migration, which are likely to have reduced differences between areas over time.

Spatial analyses

Findings in relation to original hypothesis

The *a priori* hypothesis with respect to spatial variations in COPD, hypothesis 5 stated:

The spatial distribution of areas where patterns of lung cancer and COPD mortality are discordant reflects the distribution of risk factors other than smoking that influence COPD development and/or mortality.

In effect, this hypothesis was suggesting that the spatial variations in SMRs for COPD seen across the country were not solely due to differences in smoking. A number of exposures associated with COPD development and/or mortality from individual-level and ecological studies were selected on the basis of the literature search and these were all found to be associated with COPD mortality risks at district-level. This is perhaps not surprising in retrospect, but it was unclear before conducting the analyses whether the influence of smoking and deprivation would make contributions of other factors negligible in explaining spatial variations in COPD mortality. It also has public health relevance in suggesting that smoking is not the only modifiable risk factor with potential impacts on COPD mortality.

Analyses were carried out in a two-stage process because of the lack of information at sub-regional level on smoking (for either prevalence or cumulative exposure), using a relatively new Bayesian statistical technique[215] to remove smoking effects, which were

identified as the component of district-level COPD mortality risk that is shared with lung cancer. This was followed by a frequentist weighted logistic regression. The implications of this choice of analytical methods will be discussed first, followed by an interpretation of the findings.

Methodological issues

Choice of analysis years

The temporal analyses had suggested that, while air pollution was a key determinant of differences in COPD mortality between geographical areas in the 1960s and 1970s, its effect lessened after this and the relative contribution of other factors became more important. Data to investigate spatial patterns risk factors also first became available for the 1980s and 1990s. While district-level meteorological variables and occupational mortality data were available for the 1980s and 1990s, dietary data were only available for the 1990s, ambient air pollution data only available for 1996 and the deprivation variable chosen was based on information from the 1991 Census. It was decided conduct the analysis using the full span of available and compatible mortality data (19 years), rather than restricting the analysis to the 1990s, to maximise the numbers involved and therefore the power to detect associations, but this may have been at the expense of introducing some misclassification bias. Investigation of changes over time during this 19 year period for datasets only available in the 1990s was conducted where possible and suggested some consistency in the relative ranking of areas for these risk factors over time.

Choice of shared component followed by linear regression analyses

The use of a two-stage process made explicit the removal of smoking associated risks, through the use of lung cancer mortality as a proxy and followed closely the way in which the spatial hypothesis was formulated.

One of the few other spatial analyses to look at COPD mortality in the UK was by Law and Morris,[21] which also had to use an indirect method of looking at smoking. They first conducted a district-level Poisson analysis examining the association of cause-specific mortality by age-sex group with deprivation, latitude and urbanisation, adjusted for percentage of the population belonging to an ethnic minority group. Law and Morris[21] then examined the relationship between deprivation and latitude with age-

adjusted smoking data from the 1972 General Household Survey, which is available by socio-economic group and by region (but not at higher spatial resolution, either in the published volume[372] or in the dataset available from the UK Data Archive). Finally they used smoking-related relative risks derived from the British doctors cohort[108] to calculate expected cause-specific mortality by deprivation and latitude and compared these with the actual results obtained.

The approach used in this study allowed examination of a wider range of factors than those possible using the Law and Morris method e.g. using temperature directly rather than using latitude as a proxy,[21] using information on diet and occupational mortality and adjusting for smoking at district rather than regional level. Also, as discussed previously, there were also some theoretical advantages in using lung cancer mortality instead of smoking prevalence data as it may better capture information on cumulative smoking.

Alternative methodological approaches

An alternative methodological approach would have been to conduct a Poisson regression (with appropriate adjustment for the overdispersion of the data) including the log lung cancer SMR as a covariate and proxy for smoking. This is not as explicit as the method used here. Another problem is that several of the other variables (asbestos-related SMRs, air pollution, fruit & vegetables, deprivation) are also risk factors for lung cancer. Strongest correlations were seen with deprivation quintiles with coefficients around 0.8 in Spearman rank correlations, with the next strongest seen with coefficients around 0.3 for SO_2 quintiles and asbestos-related disease groupings. Correlation of covariates would be expected to reduce the precision of the estimates. Finally, a theoretical issue is that inclusion of lung cancer in the model implies it is a risk factor for COPD, which is not believed (although it is possible that the reverse is true as chronic inflammation seen in some COPD patients may predispose to lung cancer[310]).

It would also have been theoretically possible to have fitted a hierarchical linear regression model for the second stage of the analysis making use of ward-level mortality data. For this the shared-component model would have needed to have been run at ward level. The outcome variable COPDspec would then have been at ward level, while covariates could be incorporated at the highest spatial resolution available –

e.g. asbestos-related disease and pneumoconiosis mortality at ward-level, dietary data at district-level as it was too sparse use at small area level. The associations could have been estimated using a random effects model, with unexplained variation partitioned into spatially structured (representing unmeasured ward-level covariates) and random variation.

The advantage of this approach would be reduction of ecological bias by using small area level data where possible. Theoretical drawbacks of this approach are that the fine scale of results may have made recognition of patterns (such as higher risks in conurbation areas) more difficult to detect and that a finer scale may have made data sparse and therefore more prone to chance fluctuations. However, the chief limitation was practical – running the shared-component model for 459 districts took approximately 4-5 hours of dedicated computer time after convergence plus one hour burn-in, but running the model for 10,000 small areas would have taken several weeks. Also, geographical boundary information for winBUGS are not available at ward-level and their creation at district-level for the analyses presented in this PhD required some complex text-editing.

Shared component methodology

One of the strengths of the Bayesian shared component approach was the shared component has the potential to remove nuisance effects common to both diseases (e.g. related to errors in population estimates, geographical boundaries, etc.) in addition to the effects of shared unmeasured risk factors.

Maps of the results from the shared component analyses were displayed with posterior mean values as well as the probabilities that risks were greater than or less than 1 as the maps of values alone can be misleading. Use of probabilities can both minimise the false positive results and maximise the sensitivity.[307] A large simulation study suggested that Bayesian disease mapping models (such as the shared component model) are conservative with high specificity,[307] but that looking at probabilities that the relative risk is above 1 (or other values) with a cut-off of 80% is a way of increasing the sensitivity.

Linear regression

Smoothing in shared component analyses

The risks produced by shared component analyses represent risks smoothed towards those of surrounding districts, with the degree of smoothing dependent on the number of deaths in each district. This will have reduced the effect of chance fluctuations. However, there are concerns that smoothing may obscure the underlying pattern of risk.[307] Smoothing is unlikely to have had a large effect on the results in this particular case as (i) the numbers (of deaths) per district were not sparse and (ii) maps of COPD SMRs showed reasonably similar patterns to the smoothed posterior relative risk estimates (Figure 7-2, page 199, Appendix, Figure A-2 and A-3, pages 354-355).

Weighted linear regression analyses

Weighting in the linear regression took account of precision with which the outcome variable was estimated, by using analytic weightings equal to the precision of COPDspec (calculated as the 1/(square of the posterior standard deviation)).

Measurement error in covariates

Information on the measurement error of covariates was not included in this model. This could be included in both a regression or Poisson model in a classical framework, but is arguably easier to incorporate in a Bayesian framework. For this to be worth doing, the information on measurement error variance should be good. However, for most of the exposure variables used in the analysis, the measurement error information was incomplete. Taking fruit and vegetable consumption as an example, the number of households used to estimate average district consumption was available and measures of the variability of consumption by week or by year could also have been calculated. But this would not have given information on the accuracy of the TNS data on reflecting the average dietary consumption in those districts. For previously interpolated data such as temperature, rainfall and SO_2, no information on the errors associated with interpolation were given. Although it would have been possible to use the standard errors of the mean values calculated for the meteorological data, these may potentially have been smaller than the errors introduced by interpolation.

The issue of measurement error could not be ignored and will be discussed in relation to individual covariates in the interpretation of findings below. However, the use of

quintiles rather than absolute values will have helped to reduce the impact of random measurement errors. Also, weighting of the TNS households in the survey for less than 12 weeks many have reduced bias of unrepresentative purchases. Sensitivity analyses were conducted for the fruit and vegetable variable by restricting analyses to districts with at least 10 households surveyed. Finally, the plausibility of the results were considered by comparison with other analyses.

Examining for interactions

The power of linear regression analyses to examine for interactions between explanatory variables is limited, and this was further limited by the use of quintiles particularly if used as categorical variables. For example, a statistically significant interaction ($p=0.04$) was detected in females if using binary deprivation and binary fruit and vegetable variables, but not when using quintiles. The interaction detected suggested a four-fold higher protective effect of fruit and vegetable consumption on independent COPD risk in the more deprived compared with the less deprived districts. However, quintiles were preferred to binary variables if assessing model fit using AIC criteria and so this result should be treated cautiously.

Migration

The mortality analyses related to a 19 year period. Exposures were measured in different time-scales, but it was assumed that relative differences between areas were similar over time – but this relies on the assumption that people have exposures related to the areas where they die. Migration will affect these assumptions, but as this was an ecological study, the direction of the bias cannot be predicted in advance.[209]

Findings of shared component analyses

What does the COPD and lung cancer shared component represent?

Removing the shared component between lung cancer and COPD in spatial analyses will have removed risks common to both diseases in terms of both nature and size of risk. Smoking is common to both diseases but the differences in smoking-related attributable risks for COPD and lung cancer suggest that there will be an additional disease-specific smoking-related risk for lung cancer. Smoking is not the only risk factor common to both diseases and the shared component is likely to include other risk factors common to both diseases.

What are the risk factors for lung cancer other than smoking?

Established non-smoking causes of lung cancer include[28,310]:

- Radon
- Air pollution
- Asbestos
- Occupational exposures (nickel, hexavalent chromium, bischloromethyl ether, arsenic, polycyclic aromatic hydrocarbons such as benzo[373]pyrene found in tar and soot, mustard gas)
- Silica dust (classified as a carcinogen by IARC,[374] but this is controversial[310])

A protective effect of fruit and vegetable consumption on lung cancer has also been seen in some but not all studies, and the effect of vegetables may be stronger than that for fruit.[310]

Risk factors shared between COPD and lung cancer other than smoking

Several of the known and potential risk factors for lung cancer, including air pollution, some occupational exposures and fruit & vegetable consumption are also known or potential risk factors for COPD and therefore likely to contribute to the shared COPD and lung cancer mortality risk. The estimates from linear regression analysis using independent COPD risk as an outcome may differ from risk estimates using overall COPD mortality risks depending on the contribution to the shared risk.

Another consideration is whether a risk factor acts additively or synergistically with smoking. Smoking is known or is likely to act synergistically with most of the other causes of lung cancer.[310] Radiological doses are generally considered to be additive to other risks,[318] but radon and smoking influence lung cancer risk in a manner that is supra-additive, but submultiplicative.[310] It is not clear whether risk and protective factors for COPD act synergistically or additively with smoking, except for coal mining dust exposure, which is generally considered to act additively.[43] This is important for the shared component as factors acting synergistically with smoking for one disease but additively for the other will contribute to both the shared component and the independent risk for one disease.

Doll[28] considered that only radon, asbestos exposure and ambient air pollution may have increased lung cancer mortality risks at population level and of these only ambient air pollution can be definitely considered a shared risk with COPD.

Asbestos in conjunction with cigarette smoking is credited with contributing up to 5% of lung cancer in the 1970s.[28] Peak asbestos exposures in the UK are likely to parallel the peak in asbestos imports in 1955-1980[375] and therefore the contribution of asbestos to lung cancer mortality may have been higher than 5% in the 1980s and 1990s (the time period of analysis) – but most published considerations of the health impacts of asbestos exposure are confined to mesothelioma.[375] The contribution of asbestos dust exposure to COPD is not clear[131] but the spatial pattern of asbestos-related disease deaths may serve as a marker for industrial exposures to other dusts that may increase the risks of COPD.

In the analyses presented in this PhD, asbestos-related disease SMRs were found to be higher in central London and areas associated with ship-building such as Clydeside, Tyneside, Plymouth and Portsmouth. These areas would also be expected to show higher lung cancer risks if considerable numbers were exposed. The maps for asbestos-related disease SMRs (Appendix, Figure A-13, page 365 and Figure A-14, page 366) showed more similarity with maps for the shared COPD and lung cancer risks (Figure 7-7, page 216 and Figure 7-8, page 217) than for independent lung cancer risks (Figure 7-9, page 219 and 7-10, page 220), suggesting that some part of the shared risk could be associated with the presence of types of industries where asbestos use was more common.

The spatial pattern of independent lung cancer risks for males (Figure 7-9, page 219) showed more consistency with those for females (Figure 7-10, page 220), who had small numbers of asbestos-related disease deaths, than with SMRs for asbestos-related disease (Appendix, Figure A-13, page 365), particularly in more southern districts, suggesting that spatial patterns for the independent lung cancer risks reflect factors other than asbestos exposure – possibly smoking risks in excess of those shared with COPD.

The role of air pollution in lung cancer was considered on pages 246-247 with estimates ranging from 1-3% overall in the USA[42] to 10-12% in polluted urban areas.[317,318] The spatial distribution of interpolated ambient SO_2 for 1996 (Appendix, Figure A-8, page

360) showed most similarity with the spatial patterns of the shared COPD and lung cancer risks (Figures 7-7 and 7-8, pages 216 & 217), but there was also some similarity with COPD independent risks (Figures 7-5 and 7-6, pages 213-214), suggesting effects of SO_2 on both COPD and lung cancer mortality, plus additional effects on COPD.

Linear regression associations of exposure variables with independent COPD risks

General issues

Linear regression was carried out on the independent COPD risk (but not shared or independent lung cancer risks) to investigate the hypothesis that areas where spatial patterns of lung cancer and COPD mortality are discordant would contain different distributions of risk factors thought to influence COPD development and/or mortality. Specific issues to do with the data and the findings will be discussed in turn. However, all the results may be prone to the ecological fallacy, as discussed on page 249 above.

Some of the exposure data were only available for single years. This may not prejudice the analyses, as it seems likely that the relative differences between areas persist over time or change slowly. For example, Dorling[376] found that many of the areas with high mortality in the 1990s had also experienced high mortality rates in the 1950s. This suggests that there were similarities in individuals living in an area and/or that clustering of risk factors associated with an area persisted over time. This premise is consistent with findings from the shared component analysis where lagging COPD mortality 10 years (using lung cancer mortality for the 1980s and COPD mortality for the 1990s) produced similar spatial patterns of risks (not shown).

The associations with each variable in turn will be discussed in light of issues with the data and interpretation of the associations detected. This will be followed by a discussion of interactions.

Association with deprivation

Issues related to data

There was no 'off-the-shelf' deprivation measure available for use at the geographical resolution of this study. The measure was based on Carstairs score for 1991 (an area-based measure), whose use in epidemiological studies is well-established. Carstairs

index shows good correlation with other deprivation indices such as Jarman scores and the Townsend index.[255] The deprivation variable used in these analyses also had similarities to the criteria for one of the IMD2000 indices.[302] However, it was concluded in Chapter 4 that in comparison with the use of Carstairs score at ward-level, the created deprivation index was likely to under-estimate the increased risk associated with higher levels of deprivation on COPD and lung cancer mortality but also to over-estimate the protective effect of lower levels of deprivation. It is difficult to know the significance of this for the validity of the results presented. Deprivation is likely to be a proxy for some composite of a range of individual-level factors[277] accumulated over the life-course[176,377,378] and area-level risk factors (such as environmental pollution, health and community services, land use and social networks)[277] and it is difficult if not impossible to know whether existing deprivation measures adequately capture the association with disease. Moreover, while associations with most diseases are strong,[277] the dose-response relationship and indeed the mathematical form of the association with specific diseases is unclear.

The variable used for deprivation related to a single year 1991, but there may have been changes in areas over time that were not captured. This is, however, likely to be more of an issue for small-area analyses, as factors such as new development are more likely to profoundly affect the social make-up of a ward than an entire district. Also, the measure relates to the middle of the analysis period, which is more likely than a measure at the beginning or end of the period to capture average deprivation levels over time.

Interpretation

An association with deprivation is seen with most diseases[277] including lung cancer and COPD. In the datasets used, there were correlations of around 0.7 between the deprivation variable and log COPD SMRs and around 0.7 between the deprivation variable for log lung cancer SMRs in females and 0.8 in males. It is perhaps surprising then, that deprivation continued to show a significant association with independent COPD risks after removal of the shared component. This raises again the issue that it is unclear what deprivation actually represents. There is a very strong relationship between area-level deprivation and smoking prevalence[150] and Jarvis[379] had suggested that much of the socio-economic variation in mortality in adults is related to variations in smoking. It may be that in this situation the deprivation variable is mainly picking up

residual associations with smoking, not fully removed through the shared component analysis. This would be consistent with the finding of a residual association between the independent COPD risk (COPDspec) and log lung cancer SMRs (Figure 7-4, page 209). The correlations of 0.38 for males and 0.27 for females (between COPDspec and log lung cancer SMRs) might have suggested a stronger association with deprivation for males than for females but the reverse was found in both multivariate and univariate analyses. On the other hand, it has been suggested that females may be more sensitive to the adverse effects of smoking [208]

When a number of variables are included in the model as here, inclusion of social class of the area may result in overadjustment for deprivation, resulting in an observed relationship between exposure and outcome that gives an over-conservative estimate of the 'true' association. For example, air pollution levels are generally higher in deprived areas and adjusting for deprivation in these analyses resulted in a 20-30% reduction in the observed change in relative risk per quintile of ambient SO_2 (Table 7-14, page 231), which could be an underestimate of the true air pollution risk. This is a difficult problem affecting many environmental studies. Because of this, the change in relative risk estimates before and after adjustment for various confounders were presented in Table 7-14 (page 231). Adjusting for deprivation was found to reduce the size of the coefficients for SO_2, pneumoconioses and fruit & vegetable consumption. The coefficient for asbestos related SMRs became larger (but there are some problems with interpretation of this factor as discussed below), while changes in coefficients for temperature and rainfall were difficult to interpret.

Association with SO_2 air pollution

Air pollution data

The measure of particulate air pollution used was PM_{10}. Unfortunately, 1996 was an unusual year and average PM_{10} levels in that year could not be held to be representative of years either side.

Interpolated ambient SO_2 concentrations were produced from a combination of emissions data 'calibrated' using concentrations from automatic monitoring sites. In the Netcen report relating to these interpolations,[303] interpolated values were held to be '*a reasonably good estimate of concentrations*' in city centre or urban locations but levels in suburban or smaller urban areas may have been underestimated. However, the

meaning of *'reasonably good'* was clarified later in the same paragraph when comparisons of interpolated data with Basic Urban Network measurements suggested actual measured values were 1.6 to 2.3 times those of interpolated values. Reasons given included the influence of very local sources on monitoring sites, inaccuracies in the way that emission sources were spatially distributed within the emissions inventory and inaccuracies in measurements.[303] It is likely that this under-estimate of actual concentrations in the interpolated values was minimised by the use of quintiles rather than absolute values, but some distortion of the relative levels of districts remains possible if there was differential under-estimation.

Interpolated air pollution data were only available for 1996, so relative differences between areas had to be assumed to remain relatively constant over time. Consistency in rankings over time using concentrations from individual stations suggested this was reasonable (assuming that concentrations from individual station bear some relationship to the interpolated data), but was less valid for the early 1980s. Spearman rank correlation coefficients for annual average concentrations of SO_2 in 1996 and concentrations for other years were only around 0.4 between 1981 and 1996, but rose steadily each year to 0.6 in the late 1980s and 0.7 in the early 1990s, were 0.9 in 1995 and 1997 and fell back to 0.7 by 1999. Using quintiles of SO_2 concentrations rather than absolute levels or rankings would also be expected to have improved the validity of relative positions of districts over time.

Interpretation

The analyses suggested an excess COPD specific risk of approximately 10% in districts in highest compared with lowest quintile of ambient SO_2, where there were differences of 4.3 ppb between mean annual average levels in the lowest and highest quintile of SO_2 (the range for all districts was 0 to 11 ppb). Coefficients were very similar in males and females.

These numerical size of the association (assuming linear extrapolations can be made) was higher than that found in another British study investigating the association between ward-level mortality data with air pollution data from individual monitoring stations 1981-1998.[149] This ward-level study[149] found an excess risk for respiratory mortality of 10.6% (95% confidence intervals 6.0% to 15.4%) for a 10 ppb rise in SO_2 with four year lag and an excess of 2.2% (1.4% to 2.9%) with a 10 year lag after

adjustments for deprivation using Carstairs decile. However, the COPD specific risk represents approximately a third of the variance in overall COPD mortality relative risk and the component of the excess SO_2 risk associated with the component of the risk shared with lung cancer may have a lower or absent association with ambient SO_2, which would make the two studies more comparable. Also, those dying from COPD may be a subgroup of those dying from respiratory disease who are more sensitive to the effects of pollution.

The numerical size of the association between SO_2 and the COPD specific risk was more comparable with that seen in the Harvard Six Cities study,[41] which found a 37% (11% to 68%) excess risk for cardiopulmonary mortality in the 1970s to 1991 in the most compared with the least polluted city, representing approximately 22 ppb difference in mean ambient SO_2. It is also comparable with the association between cardio-respiratory mortality and concurrently measured ambient SO_2 air pollution in the Health Effects Institute (HEI) reanalysis[380] of the American Cancer Society study, where a 48% excess risk (33% to 64%) in summer and 29% excess (20% to 38%) in winter, corresponding to approximately 23 ppb difference in SO_2 concentration. Again, it is difficult to make exact comparisons because the COPD specific risk associated with air pollution may be different from that for COPD risk as a whole, and those dying from COPD are a subgroup of those dying from cardio-pulmonary disease.

Although exact numerical comparisons of the preceding two paragraphs should not be over-interpreted, the size of the effect of SO_2 air pollution is certainly plausible in comparison with the British ward-level, the Harvard Six Cities and the HEI reanalysis of the American Cancer Society studies into chronic health effects of air pollution.

It seems unlikely the observed association with SO_2 is a proxy for particulate exposure. As discussed in the section relating to the temporal analyses (page 254-263), the literature suggests that SO_2 air pollution has effects on respiratory disease independent of particulates (although this may be mediated through sulphate particles). It was not possible to test this directly as the interpolated PM_{10} data could not be used in the analyses. However, poor correlations between black smoke and SO_2 of individual monitoring stations (Figure 6-6, page 182) suggested that SO_2 levels were a poor measure of PM_{10} concentrations, thus suggesting an independent effect of SO_2. In fact, Spearman rank correlation coefficients between black smoke and SO_2 readings at the

same stations in the same year fell over time from $\rho \approx 0.6$ in 1981, to $\rho \approx 0.3$ from 1996 onwards, which is consistent with changes in pollution sources over time.[266]

Association with fruit & vegetables

Issues related with data

The TNS dataset was perhaps the dataset of most concern of those used in the spatial analyses. Use of these data in epidemiological studies has been limited. Only one validation has previously been published[253] and this looked at macronutrients. Comparisons of fruit and vegetable levels with the usual national nutrition reference studies, the National Food Surveys and the National Diet and Nutrition Surveys, detailed in Chapter 3 showed consistently lower values in the TNS. Use of TNS data in this PhD was acceptable only if assumptions were made that the differences between TNS and other surveys were constant across geographical areas. However, these comparisons were problematic in that they also highlighted problems with the national reference studies in terms of small numbers and methodology.

Although households in the TNS are chosen to be broadly representative of the UK (with some over-sampling of large households[267]), the potential for misclassification bias was high. The largest number of households per district was 298 and this compares with district populations ranging in size from around 20,000 to over 3 million. Some districts had very small numbers of participating households (the 10th percentile was 7). It is also likely that participating households are not fully representative of the general population, by virtue of the fact that they are prepared to participate in this type of ongoing commitment.

The analysis presented here used quintiles of weight of fruit and vegetables plus fruit juice purchased as this could be directly derived from the data. The Food Standards Agency[268] follows the WHO in suggesting an intake of five portions of fruit and vegetable per day, defining one portion of fruit or vegetable as 80g or 125 ml fruit juice. Using portions as the fruit and vegetable variable would have down-weighted the effect of fruit juice compared with the analyses presented. It is unclear whether this would have been appropriate as the relative importance of fruit or fruit juice with lung function is unclear. Strachan et al[159] found an association between frequency of drinking fruit juice or eating fresh fruit in the winter and lung function (FEV_1). The difference

between the lowest versus the highest quintile of weight purchased was equivalent to 1.5 portions of fruit and vegetables per person per day.

Observed association in females

It is possible that the observed associations are solely the result of biases in the data. It was difficult to predict the direction of the biases and as stated previously, misclassification bias in an ecological study may *increase* as well as decrease the size of any apparent association.[209] However, a number of points suggest that this could be a real effect, worthy of further investigation using individual level analysis.

(i) The size of effect is in line with those seen with the other factors in the linear regression analyses presented.

(ii) The size of the association (in females) increased when the analysis was limited to districts with larger numbers of households involved i.e. when the precision of the exposure measure improved.

(iii) The association was seen in both males and females in univariate analyses, but was stronger in females and only persisted in multivariate analyses in females. Women in general eat more fruit and vegetables and drink more fruit juice than men,[366] suggesting that they may well have consumed a higher percentage of these foods purchased per household.

(iv) The association is consistent in that other ecological analyses[162] have found associations between fruit and vegetable consumption and COPD.

(v) The association is coherent with different types of studies. A cross-sectional study of a Dutch cohort[381] found higher lung function (FEV_1) and lower reports of COPD symptoms in subjects with higher intake of fruit. Longitudinal analyses of smokers in the EPIC-Norfolk cohort found that those in the lowest quartile of plasma vitamin C had a greater decline in FEV_1 than those in the upper three quartiles,[158] while longitudinal analyses of a cohort in Nottingham[382] found greater declines in FEV_1 in adults with lower dietary reported vitamin C intake.

(vi) There are plausible biological mechanisms through which fruit & vegetable consumption may protect against development of COPD.[153]

Size of association

The size of the association in an ecological analysis of data from the Seven Countries Study[162] was much larger than in the current analysis. Tabak *et al*[162] compared average

consumption of fruit and vegetables compiled from dietary reports in 16 cohorts (from seven developed countries) with COPD mortality 25 years later. They found a mortality rate ratio of 0.52 (95% confidence intervals 0.38 to 0.73) associated with an increase in mean solid fruit (apples and pears) consumption at baseline of 13.1g (a 10% increase). In the current analysis, restricted to districts with at least 10 households participating in the TNS survey, the rate ratio for COPD mortality in women was 0.94 (0.91 to 0.97) for an increase of 113g fruit and vegetables and fruit juice per day (corresponding to 44 g total fresh fruit). It is of note that restricting the analysis to districts with more participating households increased the size of the coefficient while reducing the difference in average consumption between the highest and lowest quintile. Another analysis involving Tabak,[383] this time a Cox proportional hazards analysis using data for middle-age men from three of the cohorts used in the Seven Countries Study, suggested that a 100 g increase in fruit intake at baseline was associated with a 24% lower COPD mortality risk after 20 years of follow-up. Again no association was seen with vegetable consumption.

Both the Seven Countries analysis and the analyses presented here in Section III suggest that a modest increase in fruit and vegetable consumption could give appreciable effects on mortality at a population level. The apparent lower effect size seen in the analyses presented here should not be over-interpreted. The Seven Countries analyses[162,383] used very different methodology from this study. That ecological study[162] was also subject to marked ecological bias as no association was seen at group level between baseline prevalence of cigarette smoking and COPD mortality, although this could be detected in individual-level analyses conducted for some cohorts. These comparisons raise the possibility that the effect of fruit and vegetables may be much larger with a more precise measure of intake and/or accumulate over time. The data used here were concurrent with the second half of the mortality period (although the relative differences between areas were presumed to remain relatively constant over time).

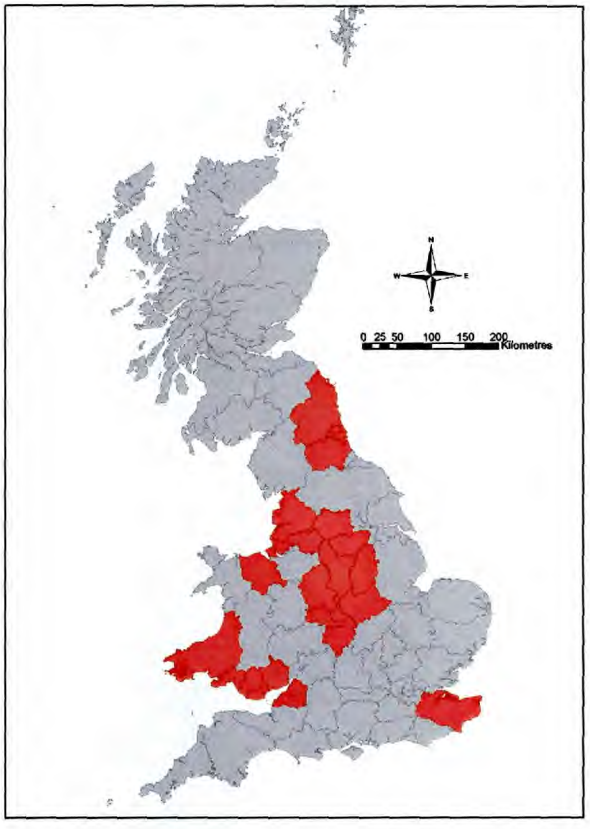
Confounding

Even if the association detected in this study was real, fruit & vegetable consumption in observational studies may be acting as a marker for other lifestyle factors such as smoking, physical activity and socio-economic status[384] that were incompletely adjusted for in the analyses.

Pneumoconioses

There was a marked similarity between the spatial pattern of the independent COPD risk (Figure 7-5, page 213 and Figure 7-6, page 214) and mining areas,[248] shown in Figure 8-2 below. Part of this similarity may be artefactual as, from 1992 onwards, Industrial Injuries Disablement Benefit was first made available for miners developing COPD.[123] This may have increased the awareness of or willingness to diagnose COPD. Higher levels of COPD mortality were also seen in these areas for women who are not generally employed in areas exposed to high levels of coal dust, which supports this interpretation. However, it may also reflect a true increased risk related to other factors. Higher rates of bronchitis mortality and hospital admissions in mining areas were observed in the 1950s and 1960s in Scotland[385] in both males and females, suggesting this is not just a diagnostic artefact related to benefit changes in the 1990s and the factor(s) responsible have persisted over time.

Figure 8-2 Map highlighting 22 counties in England & Wales where most (96%) of deaths in miners occurred 1979-80 and 1982-90



Note: There are an average of approximately seven districts per county.
Source: Coggon 1995[248]

In the spatial analyses presented in Section III, a semi-qualitative indicator was used to indicate areas with high pneumoconiosis levels – those where the lower limit of the 99% confidence interval for the SMR was than 100. While this is likely to be a reasonable identifier of exposed versus less exposed areas, it may underestimate the association between coal mining and COPD. However, it may be a better choice than using pneumoconiosis SMRs. Firstly, there were a number of outlying districts with extremely high SMRs of >1000 (Table 7-7, page 205), which may have had undue influence on the data. Secondly, an analysis by Coggon et al[248] of pneumoconiosis and COPD proportional mortality ratios (PMRs) in miners in England and Wales in 1979-1990, found that 21of 22 counties with high pneumoconiosis PMRs had high COPD PMRs, but correlations between these PMR values were poor (correlation coefficient of 0.22). This was interpreted as variations in risk for each disease related to differences in coal dusts by area, suggesting that absolute pneumoconiosis risk would not be a good predictor of COPD risk.

Only five districts had pneumoconiosis SMRs statistically significantly greater than 100 in women and the small number of districts is probably why this variable was not significant in multivariate analyses for women, despite the obvious association with log COPD specific risk evident in box plots (Figure 7-11, pages 222-223) and univariate analyses (Table 7-12, page 226).

There are very few studies with which to compare the size of the excess risk in areas with high pneumoconiosis SMRs in males from the spatial analysis, which found an excess in COPD specific risk of 7.4% (5.3% to 9.6%) where districts had pneumoconiosis SMRs >100 and reduction in risk of 3.5% (2.2% to 4.8%) in males where district pneumoconiosis SMRs were <100. Some of this excess may be artefactual related to disease labelling as discussed above. While a number of studies describe impacts of coal-mining on COPD development,[124] there are none with information on COPD mortality in coal miners. Oxman[124] used individual-level data from a cross-sectional analysis[126] to suggest that following a lifetime of moderate exposure to coal dust, 0.45% of non-smokers and 0.74% of smokers would be expected to develop chronic bronchitis symptoms attributable to dust and 0.8% of non-smokers and 0.66% of smokers would develop clinically important (defined as >20%) drop in FEV_1 directly attributable to dust. To be able to relate this to COPD mortality (about 5% of all deaths[17]), one would need to know the percentage of the population exposed

at this level and the risk that those developing chronic bronchitis symptoms or having low FEV_1 would actually be recorded as dying from COPD and this has not been established for UK populations. Also, the results presented may be an under-estimate as it is based on coal-face workers who had all contributed to two previous analyses,[126] leaving plenty of scope for differential loss to follow-up and those most affected by the dust most likely to have dropped out of the study.

Asbestos-related disease

The mortality data used in the analyses utilised ICD9 coding, which did not have a specific mesothelioma code, unlike ICD10.[386] The ICD9 codes used have been found to identify about 60% of mesothelioma deaths as captured by the national (GB) mesothelioma register.[300] While this would not be expected to have differed across districts, it may have resulted in fewer districts with SMRs statistically significantly different from 100 and thereby reduced the power of the analysis to detect an association.

In females, univariate analyses (Table 7-12, page 226) and box plots (Figure 7-11, pages 222-223) both suggested an increase in COPD specific risk for districts with SMRs for asbestos-related disease. The association became non-significant in multivariate analyses (Table 7-13, page 228), perhaps because of the small number (13) of such districts (Table 7-8, page 206). Box plots for males (Figure 7-11, pages 222-223) did not suggest an association of asbestos-related disease SMRs with COPD specific risk. A small positive excess risk for high asbestos-related disease SMRs was seen in univariate analyses (Table 7-12, page 226) but this was not statistically significant. However, in multivariate analyses (Table 7-13, page 228) the association with asbestos-related disease SMRs changed sign and became borderline significant in the model on likelihood ratio testing (p value = 0.0479). This is strongly suggestive of confounding by other variables, possibly deprivation, the variable it was most correlated with (Appendix, Table A-4, page 351) although the correlation coefficient (ρ) was low: $\rho = 0.26$. Confounding would also be the simplest explanation for the observed interaction with deprivation, which suggested a statistically significant lower independent COPD risk in districts with high asbestos related disease SMRs, but only in the least and most deprived quintiles (Figure 7-16, page 235). This explanation is also consistent with the generally accepted belief that asbestos exposure is not associated with COPD mortality.

Association with climate

Temperature and rainfall

While the meteorological data were provided at an area-based level, they were only based on 30 weather stations across the UK. This could have resulted in some bias in the results. However, the spatial distribution of these averages seen in maps of average values (Appendix, Figures A-9 and A-10, pages 361 & 362) were certainly plausible.

International comparisons in the 1950s came to the conclusion that COPD was a disease of temperate, damp, low-lying industrial areas.[387] A number of different meteorological variables were considered for use in the analyses and arguably future analyses should include a variable representing a composite of temperature and rainfall, particularly in light of the interactions between temperature and rainfall (discussed below).

While temperature and rainfall were significant in the final model for both males and females, it was difficult to interpret the observed associations. A better fit to the data was found when using quintiles than more simple divisions and it may be that relationships with meteorological variables are non-linear as suggested in the box–plots (Figure 7-11, pages 222-223). Another possibility is that associations were difficult to interpret because of the nature of the interactions between temperature, rainfall and deprivation (discussed below).

Variables not used

Urbanisation

Urban areas of England and Wales have been noted to have higher mortality from COPD than rural areas.[20] A variable related to urbanisation was not used, but other work in the department (personal communication, Danielle Vienneau) using land use and satellite data on light emissions as a marker for urban areas gave very similar spatial locations to those seen with high independent COPD risks and shared COPD risks. However, independent COPD risks also showed similar spatial distributions to those of deprivation and ambient SO_2 levels, while analyses by Barker and Osmond[177] suggested higher infant mortality and bronchitis in these areas in the first half of the 20th century (and presumed higher rates of respiratory tract infection in childhood that may increase risk of later COPD development), so this may just represent a clustering of a number of risk factors in similar areas. With respect to COPD mortality, the main risk factors

expected to be higher in urban areas would be occupational exposures, air pollution and deprivation, all of which were included in the current analyses and it is unclear whether an urbanisation variable would have brought extra information to this analysis.

Latitude

Latitude was used by in the analysis by Law and Morris[21] but was interpreted as a proxy for temperature. The analyses here used temperature directly, which showed a west-east as well as north-south gradient.

Early life exposures

Information on early life exposures (with presumed impact on COPD but not lung cancer[167]) could have been captured in the use of infant mortality in say the 1920s or 1930s. However, while the mortality data are available there would have been difficulties of incompatibility of geographical boundaries and problems with migration over and above those related to the present analyses.

Interactions

Investigation of interactions was intentionally restricted to those with the most plausibility, although most of the variables considered could plausibly interact with each other – for example, diet may interact with air pollution effects.[388] Interpretation of interactions became difficult when considering multiple possibilities. Also, three-way interactions could not be readily computed and power to consider these type of interactions was limited. For example, there were interactions between deprivation and temperature and between temperature and rainfall.

Interactions with deprivation

Significant interactions between deprivation quintiles and fixed range SO_2 quintiles were seen in males and females with effects of SO_2 highest in the least deprived districts and suggestive of decreasing impact with increasing deprivation (Figure 7-14, page 233). This was not expected and has not been reported previously. It may well suggest that the influence of other risk factors in deprived areas have larger effects on the COPD specific risk than ambient SO_2 thereby obscuring its impact on respiratory disease or reflect an adaptation response in people living in more deprived areas who might be expected to be exposed to higher levels of air pollution.

The interactions of deprivation with temperature were complex to interpret – these were just statistically significant in males (likelihood ratio testing gave a p value = 0.04) but just non-significant in females (p value = 0.08), which is probably related to the low power to detect an association when using quintiles. No pattern or linear trend was detected in interactions for either males (Figure 7-15, page 234) or females (not shown).

While no statistically significant interaction was seen with fruit and vegetable consumption and deprivation if using quintiles, using binary deprivation and binary fruit and vegetable variables revealed a statistically significant interaction (p=0.04). Using these variables, the protective effect of higher compared with lower fruit and vegetable consumption (an average 0.8 portions per day difference) on independent COPD risk was four-fold higher in the more deprived compared with the less deprived districts, showing a change in relative risk of –4.2% (95% confidence intervals –6.3% to 2.0%) in the more deprived districts compared with –1.0% (–3.2% to +1.1%) in the least deprived. For comparison, the highest vs. the lowest fruit and vegetable purchase quintiles related to a difference of on average 1.5 portions per day and the final multiple regression model suggested an overall change in independent COPD risk of –5% (–2.3% to –7.7%). No other study was located during the course of this PhD or in a specific literature search reporting on this particular interaction.

The interaction between deprivation and asbestos-related SMRs was difficult to interpret and probably related to confounding, although the risk factor(s) causing the confounding is unclear.

Interactions between temperature and rainfall

There were significant interactions of rain and temperature in both sexes. Again these did not show a clear picture, but did suggest increases in COPD risk with increasing rainfall, but only in the coldest districts.

Climate and deprivation effects – or housing?

It is possible that the interactions between rain and temperature and rain and deprivation were complex because of a three-way interaction between these variables and this may be mediated through poor quality housing. Blane *et al*[47] found that most counties of Britain with a poor climate (addition of Z-scores of average rainfall and temperature 1961-1990) were those with poor quality housing including Scotland, most of Wales,

most of north-east England and the areas around Leeds, Manchester and Birmingham – all areas identified as having high independent COPD risks. Cumbria and areas to the south-west had poorer climate but good-quality housing. South-east England except London had good climate and good quality housing. This information came from the nationally representative sample of 6,700 people surveyed in the Health and Lifestyle in 1984-85 and housing quality was assessed as the composite Z-scores of area of residence, housing tenure, crowding, outdoor toilet, sole use of domestic facilities, living room temperature and indoor carbon monoxide levels. While poor climate and poor climate were associated, Blane also suggested[47] that socio-economic class was likely to be a proxy for poor quality housing. The spatial analyses presented in this PhD submission adjusted for area-level deprivation, which is likely to at least partly adjust for poor quality housing and this may explain some of the observed association.

Implications

Shared component risk compared with smoking attributable risks

The shared component analysis in this PhD suggested that 64% of the variance in smoothed COPD relative risks for districts in 1981-1999 in both males and females was accounted for by a component shared with lung cancer, which is likely to be largely related to smoking but may also include other shared risk factors. This figure was 5-10% lower in females and 12-18% lower in males than attributable risks for smoking calculated using the Peto method[96] utilising comparisons with lung cancer mortality (presented in Table 2-4, page 36) or the Danish method[312] using information on smoking habits. Given the methodological issues with the Peto and Danish method discussed earlier in this chapter, the results from this study are certainly comparable.

Possible explanations for a lower percentage are that the shared component model did not fully identify all smoking related risks, which may be related to the strong spatial assumptions made in the model. Some evidence that residual smoking effects remain in the model for this was given by the residual association between the log independent COPD risk and lung cancer (Figure 7-4, page 209), with correlations between the log independent COPD risk and log lung cancer SMRs of 0.38 (p=0.0000) for males and 0.27 (p=0.0000) for females. Alternatively, the lower explanatory power of (presumed largely) smoking related factors found in this study may be a true finding and relate to different attributable risks for smoking for COPD (and lung cancer) in the UK over this

time period from those derived from the American Cancer Prevention Study,[100] used to calculate attributable risk in the Peto[96] and Danish[312] studies.

Geographical variations

This one of very few analyses to look at the risk factors associated with uneven distributions of COPD mortality in the UK. The patterns in COPD mortality observed – in particular higher COPD mortality in northern areas of England, southern Scotland and southern Wales – have been consistently documented over the past 40 years,[18,19,22] and it is perhaps surprising that so little research into the relative effects of modifiable risk factors has been conducted. COPD mortality is also an important contribution to the excess mortality seen in more deprived areas, estimated in one analysis[21] to contribute 13% of the excess risk (the third highest contribution after 33% from ischaemic heart disease and 22% from lung cancer).

A spatial analysis of COPD mortality in the USA,[166] using areas with approximately half the average population of districts used in these analysis, was limited by the available datasets – for example, there were no suitable air pollution data. In the American analysis, a Poisson regression of smoothed COPD mortality rates in white males aged 65+ years found higher rates where lung cancer rates (used as a marker of smoking) were high and in less densely populated areas (which may represent access to healthcare), but districts with highest COPD mortality rates were concentrated in major metropolitan areas – consistent with the spatial analyses presented in this PhD. Higher rates were also seen[166] where elevation was higher and in areas of low rainfall, but this association was dominated by higher rates in the western desert areas with greater distances to access to medical care and it is difficult to know if this is directly comparable with the UK.

Role of smoking in geographical variations in COPD mortality

The methodology of one of the few previous geographical analyses conducted by Law and Morris[21] was described previously on page 292-293. They concluded that differences in prevalence of smoking accounted for much of the variation in (all-cause) mortality between districts in England and Wales in 1992, with results presented in tables[21] suggesting this might account for approximately 50% of the differences in COPD mortality by latitude. This is consistent with the figure obtained in the analyses presented in Section III, where 64% of the variance in smoothed COPD relative risks

for districts in 1981-1999 was accounted for by a component shared with lung cancer, which is likely to be largely related to smoking but may also include other shared risk factors.

Role of air pollution in geographical variations in COPD mortality

The temporal analyses suggested that ambient air pollution was an important determinant of geographical variations in COPD mortality (i.e. between conurbations, Greater London and non-conurbations), particularly over the 1950s to the 1970s. The spatial analyses suggested that air pollution continues to be an important factor in COPD mortality, with effects of ambient SO_2 comparable in size and additional to those of deprivation. It may seem surprising that the current low air pollution levels continue to have presumably causal associations with mortality, but it should be remembered that ambient concentrations are not the same as dose – the average adult inhales about 10,000 litres of air in the course of a typical day.[310]

Role of other factors in geographical variations in COPD mortality

The findings presented here suggesting that factors other than smoking explain appreciable amounts of the remaining variation in COPD mortality risk is also consistent with Law and Morris's analysis (where only half of the excess risk was explained by smoking).[21] However, their results suggested that up to a third of the excess risk was related to temperature, based on the association of temperature with latitude. Their analyses were restricted in terms of exploration of other risk factors and the analyses presented here using temperature directly suggest a more complex relationship with COPD mortality, with interactions with deprivation and rainfall.

Urbanisation

The areas with highest risks shared between COPD and lung cancer and highest independent COPD risks were the conurbation areas of the UK. Such areas generally have poor health indices such as high infant mortality or premature mortality[378] as well as clustering of risk factors potentially affecting many diseases – including smoking, deprivation, air pollution, heavy industry, mining. They may also have environmental contamination relating to industrial processes over the last 150 years. As these are areas of high population, there is also greater potential for a role of infectious diseases in initiating or promoting chronic disease. This makes it difficult to disentangle specific risk factors. Correlations between variables actually used in the analyses presented

were measured and found to be moderate or low (Appendix, Table A-4, page 351), which argues against this. It remains possible that the significant variables were merely acting as proxy measures for some other factor associated with urbanisation or heavy industry. However, the variables chosen all had some *a priori* evidence of association with COPD from other studies for inclusion in the analysis, which strengthens the plausibility of the findings.

Further analyses

All of the risk factors investigated in the spatial analyses were selected on the basis that they had been associated with COPD in other studies. Confirmatory analyses with individual-level adjustment for smoking to determine the relative importance of different factors in COPD mortality in the same population would be useful. However, there are few UK cohorts that are both old enough to look at COPD mortality and that have information on both smoking and the range of risk factors used in this analysis. It might be possible to consider using COPD prevalence defined from lung function data, but the relationship between COPD prevalence, morbidity and mortality has been little investigated, especially in UK cohorts. The spatial analyses also suggested that the role of the meteorological factors in COPD mortality, more specifically cold and damp potentially modified by housing and deprivation, should be reassessed.

The analyses conducted were designed to establish whether or not known risk factors other than smoking and deprivation were important in explaining spatial variations. Further analyses would need to be conducted to assess the contributions that these risk factors make to the spatial patterns of the shared lung cancer and COPD risk component before being able to assess the size of the association of these risk factors with overall COPD mortality risk. Such analyses should also permit calculation of attributable risks.

Conclusions from spatial analyses

The analyses presented were designed to investigate the hypothesis that factors other than smoking could explain some of the spatial patterns of COPD mortality in the UK. The findings suggested that deprivation, temperature, rainfall, occupational dust exposures, fruit and vegetable consumption and air pollution all have independent and probably causal associations with spatial variations in COPD mortality after adjustment for smoking. Further work is need to assess the relative contribution of these risk factors to spatial variations in overall COPD mortality risk.

Chapter 9. Conclusions

Overview of Chapter 9

This chapter presents the overall conclusions from this PhD. Issues raised relate to those arising from planning and conducting the PhD analyses as well as the interpretation of the results obtained relating to changes in COPD mortality over time and the spatial variations in COPD mortality. Finally, the extra insights obtained through considering both temporal as well as spatial variations in the epidemiology of this significant respiratory disease are highlighted.

Conducting population level temporal and spatial epidemiological studies

A common theme from both temporal and spatial studies was that population level studies with direct relevance to public health such as those presented here are limited by the availability of appropriate environmental and lifestyle data.

While the UK has a unique long-running set of air pollution data from the 1950s onwards, using similar methodological techniques across the whole country (unlike data available for the USA[341]), these UK data are not available at area-level prior to 1996 limiting their use in epidemiological studies.

Smoking is a very important confounder in ecological studies,[21] but smoking data are not available at sub-regional level in historical or in current surveys such as the General Household Surveys or the Health Survey for England, probably because the numbers surveyed were thought too small to release at such level. More recent statistical developments in smoothing techniques could potentially allow the use of these data at county or at district-level. An alternative potential source of smoking data is from primary care data. However, investigation of primary care databases such as the General Practice Research Database suggest that such data tend to under-record past smoking.[389] Another issue is that smoking data from surveys commonly relate to smoking prevalence and numbers of cigarettes smoked, but for chronic diseases such as COPD and lung cancer, estimated pack years would be a more useful measure.

Dietary data are potentially a very important covariate to consider in many diseases. However, national dietary surveys such as the National Food Survey, the National Diet and Nutrition Survey and the Scottish and Welsh health surveys are not large enough to permit geographical analyses even at district-level, let alone small-area level. The analyses used here produced credible results from market research dietary data and more validation work is warranted to see if this data source could be used more widely for spatial analyses.

Mortality and (for more recent years) hospital data are readily available and of reasonable to good quality for use as outcome data in ecological or small-area analyses. While such studies can provide information on risks in a general population that is directly relevant to public health, they are also prone to a number of biases including ecological bias. Unfortunately, the potential for confirmatory individual-level UK analyses on the role of environmental factors in COPD and other multi-factorial chronic diseases is currently limited, as many British cohorts are currently too young to fully assess the impact of cumulative life-time exposures or have limited information on environmental exposures.

Conclusions from temporal analyses

The question of whether the 1956 Clean Air Act was associated with reductions in COPD mortality is particularly relevant to many developing and middle-income countries experiencing severe smogs and high outdoor air pollution in cities, but high levels similar to those seen in English cities in the 1950s and 1960s occasionally also affect more developed cities (e.g. both Moscow and Sydney were affected by smog from forest fires in 2002[390,391]). There are more general implications about the public health importance of policy decisions on air pollution and air pollution sources such as transport in the UK and other industrialised countries.

The temporal results plausibly suggest that strict and enforced regulation of air pollution can have an observable impact on respiratory health at a population level. Results presented in this PhD were consistent with findings that regulation of sulphur content of fuel oil in Hong Kong[309] and bituminous coal sales in Ireland[308] had important impacts on respiratory mortality that were higher than suggested from time-series analyses of acute effects. This current study suggested a more dramatic effect observable in

mortality rates at the much higher levels of air pollution experienced in the UK in the 1950s and 1970s. However, it has to be acknowledged that many other changes in factors affecting COPD mortality also occurred over the last 50 years that could contribute to the observed higher effects. While air pollution was found to be the single factor most able to explain the observed trends, changes in other factors such as tar content of tobacco, treatments for COPD demonstrated to reduce mortality (e.g. influenza vaccination, use of antibiotics) and increases in fruit and vegetable consumption may also have had an impact. Additionally, differential internal migration and immigration of ethnic groups with lower rates of mortality from respiratory disease[371] may also have affected COPD mortality rates.

The hypothesis that reductions in air pollution were the most important factor in UK trends in COPD mortality in the 1950s to 1970s could be tested by repeating the analyses in other populations that have experienced dramatic changes in air pollution, such as the reductions from very high levels in much of Eastern Europe or increases in cities in China over the last 20 years. More detailed analyses at individual level could also be attempted in the UK, if work was carried out to interpolate the comprehensive air pollution monitoring data from individual monitoring stations available for the last 50 years in the UK. Area-based scores could then be applied to individuals in long-running British cohort studies such as the Longitudinal Study or the 1946 birth cohort. Such studies would be able to account for individual confounders and potentially examine for interactions – for example, that seen with deprivation in the spatial analyses.

Conclusions from spatial analyses

Spatial variations suggested that smoking cannot fully explain spatial variations in COPD mortality in the UK and was consistent with an analysis for 1992[21] suggesting that smoking only explained about half of the variation. Exposures to factors such as fruit and vegetable consumption, occupational exposures and air pollution are all potentially modifiable, so the results of this analysis have important public health implications. COPD is the fourth leading cause of mortality worldwide and in the UK accounts for as many as one in eight hospital admissions.[16] Spatial patterns in COPD mortality correlate well with morbidity measures such as primary care consultations and hospital admissions, so it is likely that the significant risk factors identified in the spatial

analyses also have impacts on morbidity. More research into the time course over which these factors act would help to refine advice on policy. A potential role for cold and damp, perhaps interacting with deprivation through housing quality, was also suggested, consistent with clinical observations from the 1950s. Research revisiting this area is warranted.

Further analyses of data and results presented in Section III could be conducted to assess the attributable risk of factors other than smoking to spatial variations in COPD mortality risk. It would also be possible to extend the current ecological analyses to UK hospital admission data available (at acceptable quality) from the mid 1990s onwards. However, the scope for conducting analyses using individual-level data and/or individual-level adjustment for the major confounding factor, cumulative smoking, is limited.

Importance of considering both spatial and temporal variations in COPD mortality

The temporal analyses demonstrated that while COPD mortality fell over the last 50 years and the differences in mortality rates between areas (conurbations, Greater London and non-conurbations) became less over time, the relative position of areas with respect to COPD mortality was maintained to the end of the 1990s. This was relevant for the spatial analyses, where risk factors over the last 20 years such as air pollution have changed over time.

The temporal analyses concentrated on the impact of changes in air pollution on COPD mortality. The hypothesis carried forward into the spatial analyses was that air pollution had fallen to lower levels by the end of the 1970s, such that the relative importance of other factors in explaining variations in mortality in the 1980s and 1990s had increased. This arose from consideration of tempero-spatial variations relating to relatively broad areas (e.g. conurbations, non-conurbations), but it was reasonable to assume that this would extend to variations at finer spatial resolution such as districts. Fortuitously, it first became possible to investigate this hypothesis in data from the 1980s on, as outcome mortality data were available electronically at small area (therefore flexible) geographical resolution and important risk factor data became

available for the first time e.g. ambient interpolated air pollution, or for the first time at an appropriate spatial resolution e.g. dietary data.

The striking theme from both temporal and spatial analyses was the continued importance of ambient air pollution in COPD mortality. Results in Section II suggested that air pollution was probably an important factor in spatial and temporal variations in COPD mortality in the 1950s to 1970s, while results in Section III suggested it remained a significant factor in spatial variations in COPD mortality in the 1980s and 1990s, despite much lower levels of exposure. The relative importance of black smoke or SO_2 could not be established with the analyses presented in this PhD – the spatial analyses suggested an independent effect of SO_2 on COPD mortality, while the temporal analyses were consistent with a greater impact of particulates. Findings from the temporal analyses would be more consistent with a relatively short-term impact persisting over a few years, but the time period over which chronic exposure to air pollution has maximum effect (considering lag or cumulative exposures) has yet to be established. The UK must be one of the best countries to consider exploring this relationship as it has long had high levels of COPD and also has a unique data archive of air pollution exposure data for black smoke and SO_2 that have used the same measurement techniques for the past 50 years.

Finally, the analyses argue against the current trend in the medical literature towards considering COPD as a single-factor disease. While smoking is undeniably the most important risk factor, it is not the only risk factor involved in COPD deaths. A range of other factors may influence COPD mortality either directly or through modification of the effects of smoking and these other exposures are themselves potentially amenable to improvement through health promotion or government policy .

New Castle Coale, as an expert Physician affirms, causeth Consumptions, Phthisicks, and the Indisposition of the Lungs, not only by the suffocating aboundance of Smoake; but also by its Virulency... Therefore those Diseases (saith this Doctor) most afflict about London, where the very Iron is sooner consum'd by the Smoake thereof, than where this Fire is not used... The Consequences then of all this is, that (as was said) almost one half of them who perish in London, dye of Phthisical and Pulmonic distempers; That the Inhabitants are never free from Coughs and importunate Rheumatisms, spitting of Impostumated and corrupt matter.

Evelyn J. Fumifugium, or the Inconvenience of the Aer and the Smoak of London Dissipated. London 1661

Patients with COPD: do we fail them from beginning to end? Respiratory physicians should take some responsibility for what could be regarded as the neglect of patients with COPD and need to raise the profile of the disease with governments and funding bodies. The aim is to prevent its cause, modify its natural history, focus research and ensure the implementation of all measures that may reduce the suffering.

Partridge MR. Editorial in Thorax 2003;58:373-4

Reference List

(1) World Health Organization. The World Health Report 2003. Shaping the future. Geneva, Switzerland: World Health Organization, 2003.

(2) Murray CJL, Lopez AD. Alternative projections of mortality and disability by cause 1990-2020: Global Burden of Disease Study. Lancet 1997; 349(May 24):1498-1504.

(3) Strachan DP. Epidemiology: a British perspective. In: Calverley PMA, Pride NB, editors. Chronic Obstructive Pulmonary Disease. London: Chapman & Hal, 1995: 47-67.

(4) Pride NB, Soriano JB. Chronic obstructive pulmonary disease in the United Kingdom: trends in mortality, morbidity, and smoking. Curr Opin Pulm Med 2002; 8:95-101.

(5) Reid DD, Fletcher CM. International studies in chronic respiratory disease. Br Med Bull 1971; 27:59-64.

(6) Thom TJ. International comparisons in COPD mortality. Am Rev Respir Dis 1989; 140:S27-S34.

(7) Anderson HR, Esmail A, Hollowell J, Littlejohns P, Strachan D. Lower Respiratory Disease. In: Stevens A, Raftery J, editors. Health Care Needs Assessment. Oxford: Radcliffe Medical Press Ltd, 1994: 256-340.

(8) de Marco R, Accordini S, Cerveri I, Corsico A, Sunyer J, Neukirch F et al. An international survey of chronic obstructive pulmonary disease in young adults according to GOLD stages. Thorax 2004; 59(2):120-125.

(9) Halbert RJ, Isonaka S, George D, Iqbal A. Interpreting COPD prevalence estimates: what is the true burden of disease? Chest 2003; 123(5):1684-1692.

(10) Littlejohns P, Ebrahim S, Anderson R. Treatment of adult asthma: is the diagnosis relevant? Thorax 1989; 44:797-802.

(11) Littlejohns P, Ebrahim S, Anderson R. Prevalence and diagnosis of chronic respiratory symptoms in adults. BMJ 1989; 298:1556-1560.

(12) Dickinson JA, Meaker M, Searle M, Ratcliffe G. Screening older patients for obstructive airways disease in a semi-rural practice. Thorax 1999; 54:501-505.

(13) Soriano JB, Maier WC, Egger P, Visick G, Thakrar B, Sykes J et al. Recent trends in physician diagnosed COPD in women and men in the UK. Thorax 2000; 55:780-794.

(14) British Thoracic Society. The Burden of Lung Disease. A statistics report from the British Thoracic Society. London: British Thoracic Society, 2001.

(15) Strachan D, Hansell A, Hollowell J, McNiece R, Nichols T, Anderson HR. Collation and comparison of data on respiratory disease. Report to the Department of Health, August 1999. http://www.sghms.ac.uk/depts/laia/COLLATE/respdata.htm. London: Department of Health, 1999.

(16) Price D, Duerden M. Chronic Obstructive Pulmonary Disease. BMJ 2003; 326(7398):1046-1047.

(17) Hansell AL, Walk JA, Soriano JB. What do COPD patients die from? A multiple cause coding analysis for England & Wales. Eur Respir J 2003; 22(5):809-814.

(18) Howe GM. National atlas of disease mortality in the United Kingdom. London: Thomas Nelson & Sons Ltd, 1970.

(19) Gardner MJ, Winter PD, Barker DJP. Atlas of mortality from selected diseases. Chichester: John Wiley & Sons, 1983.

(20) Britton M. Mortality and geography. A review in the mid-1980s. DS9. [9]. 1990. London, HMSO. DS series. The Registrar General's decennial supplement for England and Wales. OPCS.

(21) Law MR, Morris JK. Why is mortality higher in poorer areas and in more northern areas of England and Wales? J Epidemiol Community Health 1998; 52:344-352.

(22) LAIA. The geography of COPD mortality in the elderly. 1996. LAIA Factsheet. http://www.sghms.ac.uk/depts/laia/

(23) Office for National Statistics. Mortality statistics, cause. Review of the Registrar General on deaths by cause, sex and age, in England and Wales, 1999. 26 ed. London: The Stationery Office, 2000.

(24) Calverley PMA. Chronic obstructive pulmonary disease: the key facts. London: British Lung Foundation, 1998.

(25) NHS Executive. Burdens of disease: a discussion document. Wetherby: Department of Health, 1996.

(26) Marks G, Burney P. Diseases of the respiratory system. In: Charlton J, Murphy M, editors. The health of adult Britain 1841-1994. London: The Stationery Office, 1997: 93-113.

(27) Lee PN, Fry JS, Forey BA. Trends in lung cancer, chronic obstructive lung disease and emphysema death rates for England and Wales 1941-85 and their relation to trends in cigarette smoking. Thorax 1990; 45:657-665.

(28) Doll R, Darby S, Whitley E. Trends in mortality from smoking-related diseases. In: Charlton J, Murphy M, editors. The Health of Adult Britain 1841-1994. London: The Stationery Office, 1997: 128-155.

(29) Collis EL. The general and occupational prevalence of bronchitis and its relation to other respiratory diseases. Journal of Industrial Hygiene 1923; 5:264-276.

(30) Logan WPD. Fog and mortality. Lancet 1949; i:78.

(31) Reid DD, Fairbairn AS. The natural history of chronic bronchitis. Lancet 1958; i:1147-1152.

(32) Fairbairn AS, Reid DD. Air pollution and other local factors in respiratory disease. Br J Prev Soc Med 1958; 12:94-103.

(33) Higgins ITT. Respiratory symptoms, bronchitis and ventilatory capacity in random sample of an agricultural population. BMJ 1957; ii:1198-1203.

(34) Burr ML, Holliday RM. Why is chest disease so common in South Wales? Smoking, social class and lung function: a survey of elderly men in two areas. J Epidemiol Community Health 1987; 41:140-144.

(35) Brimblecombe P. The Big Smoke. A history of air pollution in London since medieval times. London: Methuen, 1987.

(36) Auliciems A. Trends in smoke concentrations before and after the Clean Air Act of 1956. Atmos Environ 1973; 7:1063-1070.

(37) Warren Spring Laboratory. National Survey of Air Pollution 1961-71. London: HMSO, 1972.

(38) Dockery DW, Pope CA3. Acute respiratory effects of particulate air pollution. Annu Rev Public Health 1994; 15:107-132.

(39) Viegi G, Baldacci S. Epidemiological studies of chronic respiratory conditions in relation to urban air pollution in adults. In: D'Amato G, Holgate ST, editors. The impact of air pollution on respiratory health. Leeds: European Respiratory Society Journals Ltd, 2002: 1-16.

(40) Zemp E, Elasser S, Schindler C, Künzli N, Perruchoud AP, Domenighetti G et al. Long-term ambient air pollution and respiratory symptoms in adults (SAPALDIA study). Am J Respir Crit Care Med 1999; 159:1257-1266.

(41) Dockery DW, Arden Pope CI, Xu X, Spengler JD, Ware JH, Fay ME et al. An association between air pollution and mortality in six US cities. New England Journal of Medicine 1993; 329(24):1753-1759.

(42) Doll R, Peto R. The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. J Natl Cancer Inst 1981; 66:1191-1308.

(43) Coggon D, Newman Taylor A. Coal mining and chronic obstructive pulmonary disease: a review of the evidence. Thorax 1998; 53(May):398-407.

(44) Fletcher C, Peto R, Tinker CM, Speizer FE. The natural history of chronic bronchitis and emphysema. Oxford: Oxford University Press, 1976.

(45) Dockery DW, Speizer FE, Ferris BGJ, Ware JH, Louis TA, Spiro AI. Cumulative and reversible effects of lifetime smoking on simple tests of lung function in adults. Am Rev Respir Dis 1988; 137:286-292.

(46) Peto R, Lopez AD, Boreham J, Thun M, Heath CJr. Mortality from tobacco in developed countries: indirect estimation from national vital statistics. Lancet 1992; 339:1268-1278.

(47) Blane D, Mitchell R, Bartley M. The "inverse housing law" and respiratory health. J Epidemiol Community Health 2000; 54:745-749.

(48) The COPD Guidelines Group of the Standards of Care Committee of the BTS. BTS guidelines for the management of chronic obstructive pulmonary disease. Thorax 1997; 52(Supplement 5):S1-S28.

(49) Soriano JB, Davis KJ, Coleman B, Visick G, Mannino D, Pride NB. The proportional Venn diagram of obstructive lung disease. Chest 2003; 124:474-481.

(50) Sunyer J, Antó JM, McFarlane D, Domingo A, Tobías A, Barceló M-A et al. Sex Differences in Mortality of People Who Visited Emergency Rooms for Asthma and Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med 1998; 158:851-856.

(51) Siafakas NM, Vermeire P, Pride NB, Paoletti P, Gibson J, Howard P et al. Optimal Assessment and Management of Chronic Obstructive Pulmonary Disease (COPD). Eur Respir J 1995; 8:1398-1420.

(52) American Thoracic Society. Standards for the diagnosis and care of patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med 1995; 152:S77-S120.

(53) Pauwels RA, Buist AS, Calverley PMA, Jenkins CR, Hurd SS. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Diseases. NHLBI/WHO Global Initiative for Chronic Obstructive Lung Disease (GOLD) Workshop Summary. Am J Respir Crit Care Med 2001; 163:1256-1276.

(54) National Collaborating Centre for Chronic Conditions. Chronic Obstructive Pulmonary Disease. National clinical guideline on management of chronic obstructive pulmonary disease in adults in primary and secondary care. NICE guideline No. 12. Thorax 59[Supplement 1]. 2004.

(55) Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease: NHLBI/WHO Workshop. http://www.goldcopd.com/workshop/. 2001.

(56) Fabbri L, Caramori G, Beghe B, Papi A, Ciaccia A. Chronic obstructive pulmonary disease international guidelines. Curr Opin Pulm Med 1998; 4:76-84.

(57) Medical Research Council. Definition and classification of chronic bronchitis. Lancet 1965; ii:775-779.

(58) Antó JM, Vermeire P, Vestbo J, Sunyer J. Epidemiology of chronic obstructive pulmonary disease. Eur Respir J 2001; 17:982-984.

(59) Celli BR, Halbert RJ, Isonaka S, Schau B. Population impact of different definitions of airway obstruction. Eur Respir J 2003; 22:268-273.

(60) Viegi G, Pedreschi M, Pistelli F, Di Pede F, Baldacci S, Carrozzi L et al. Prevalence of airways obstruction in a general population. European Respiratory Society vs American Thoracic Society Definition. Chest 2000; 117:339S-345S.

(61) Tsoumakidou M, Txanakis T, Voulgaraki O, Mitrouska I, Chrysofakis G, Samiou M et al. Is there any correlation between the ATS, BTS, ERS and GOLD COPD's severity scales and the frequency of hospital admissions? Respir Med 2004; 98:178-183.

(62) Global Initiative for Chronic Obstructive Lung Disease. Pocket guide to COPD diagnosis, management and prevention. A guide for health care professionals. http://www.goldcopd.com/revised_gp.pdf . 2003.

(63) American Thoracic Society. Lung function testing: selection of reference values and interpretative strategies (statement). Am Rev Respir Dis 1991; 144:1201-1218.

(64) Hospers JJ, Schouten JP, Weiss ST, Rijcken B, Postma DS. Asthma attacks with eosinophilia predict mortality from chronic obstructive pulmonary disease in a general population sample. Am J Respir Crit Care Med 1999; 160(6):1869-1874.

(65) Thomason MJ, Strachan DP. Which spirometric indices best predict subsequent death from chronic obstructive pulmonary disease? Thorax 2000; 55:785-788.

(66) Jeffery PK. Pathology of Asthma and COPD: a Synopsis. European Respiratory Review 1997; 7(43):111-118.

(67) Magnussen H, Richter K, Taube C. Are chronic obstructive pulmonary disease (COPD) and asthma different diseases? Clin Exp Allergy 1998; 28(Supplement 5):187-194.

(68) Hunt LW, Silverstein MD, Reed CE, O'Connell EJ, O'Fallon WM, Yuginger JW. Accuracy of the death certificate in a population-based study of asthmatic patients. J Am Med Assoc 1993; 269(15):1947-1952.

(69) Dodge R, Cline MG, Burrows B. Comparisons of asthma, emphysema and chronic bronchitis diagnoses in a general population sample. Am Rev Respir Dis 1986; 133:981-986.

(70) Campbell DA, Mclennan G, Coates JR, Frith PA, Gluyan PA, Latimer et al. Accuracy of asthma statistics from death certificates in South Australia. Medical Journal of Australia 1992; 156:860-863.

(71) Barnes PJ. Molecular genetics of chronic obstructive pulmonary disease. Thorax 1999; 54:245-252.

(72) Samet JM, Lange P. Longitudinal studies of active and passive smoking. Am J Respir Crit Care Med 1996; 154(Suppl):S257-S265.

(73) Kerstjens HAM, Rijcken B, Schouten JP, Postma DS. Decline of FEV1 by age and smoking status: facts, figures, and fallacies. Thorax 1997; 52(9):820-827.

(74) Sherrill DL, Lebowitz MD, Knudson RJ, Burrows B. Smoking and symptom effects on the curves of lung function growth and decline. Am Rev Respir Dis 1991; 144:17-21.

(75) Tager IB, Segal MR, Speizer FE, Weiss ST. The natural history of forced expiratory volumes. Effect of cigarette smoking and respiratory symptoms. Am Rev Respir Dis 1988; 138:837-849.

(76) Rijcken B, Weiss ST. Longitudinal analyses of airway responsiveness and pulmonary function decline. Am J Respir Crit Care Med 1996; 154(6 SUPPL.):S246-S249.

(77) Glindmeyer HW, Dien JE, Jones RN, Weill H. Noncomparability of longitudinally and cross-sectionally determined annual change in spirometry. Am Rev Respir Dis 1982; 125:544-548.

(78) Xu X, Laird N, Dockery DW, Schouten JP, Rijcken B, Weiss ST. Age, period, and cohort effects on pulmonary function in a 24-year longitudinal study. Am J Epidemiol 1995; 141(6):554-566.

(79) Ware JH, Dockery DW, Louis TA, Xu XP, Ferris BGJ, Speizer FE. Longitudinal and cross-sectional estimates of pulmonary function decline in never-smoking adults. Am J Epidemiol 1990; 132:685-700.

(80) Hole DJ, Watt GCM, Davey Smith G, Hart CL, Gillis CR, Hawthorne VM. Impaired lung function and mortality risk in men and women: findings from the Renfrew and Paisley prospective population study. BMJ 1996; 313(21 September):711-715.

(81) Mahler DA, Wells CK. Evaluation of clinical methods for rating dyspnea. Chest 1988; 93(3):580-586.

(82) Hajiro T, Nishimura K, Tsukino M, Ikeda A, Koyama H, Izumi T. Analysis of clinical methods used to evaluate dyspnea in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med 1998; 158(4):1185-1189.

(83) Hansell AL, Hollowell J, McNiece R, Nichols T, Strachan D. Validity and interpretation of mortality, health service and survey data on COPD and asthma in England. Eur Respir J 2003; 21:279-286.

(84) Pride NB, Burrows B. Development of impaired lung function: natural history and risk factors. In: Calverley PMA, Pride NB, editors. Chronic Obstructive Pulmonary Disease. London: Chapman & Hall, 1995: 69-91.

(85) Annesi I, Kauffmann F. Is Respiratory Mucus Hypersecretion Really an Innocent Disorder? Am Rev Respir Dis 1986; 134:688-693.

(86) Lange P, Nyboe J, Appleyard M, Jensen G, Schnohr P. Relation of ventilatory impairment and of chronic mucus hypersecretion to mortality from obstructive lung disease and from all causes. Thorax 1990; 45:579-585.

(87) Vestbo J, Prescott E, Lange P. Association of chronic mucus hypersecretion with FEV1 decline and chronic obstructive pulmonary disease morbidity. Copenhagen City Heart Study Group. Am J Respir Crit Care Med 1996; 153:1530-1535.

(88) Peto R, Speizer FE, Cochrane AL, Moore F, Fletcher CM, Tinker CM et al. The Relevance in Adults of Air-flow Obstruction, but not of Mucus Hypersecretion, to Mortality from Chronic Lung Disease. Am Rev Respir Dis 1983; 128:491-500.

(89) McCloskey SC, Patel BD, Hinchliffe SJ, Reid ED, Wareham NJ. Siblings of Patients With Severe Chronic Obstructive Pulmonary Disease Have a Significant Risk of Airflow Obstruction. Am J Respir Crit Care Med 2001; 164:1419-1424.

(90) Hoidal JR. Genetics of COPD: present and future. Eur Respir J 2001; 18:741-743.

(91) Chen Y. Genetic epidemiology of pulmonary function. Thorax 1999; 54:818-824.

(92) Budhi A, Hiyama K, Isobe T, Oshima Y, Hara H, Maeda H et al. Genetic susceptibility for emphysematous changes of the lung in Japanese. International Journal of Molecular Medicine 2003; 11(3):321-329.

(93) He JQ, Ruan J, Connett JE, Anthonisen NR, Paré PD, Sandford AJ. Antioxidant gene polymorphisms and susceptibility to a rapid decline in lung function smokers. Am J Respir Crit Care Med 2002; 166:323-328.

(94) Sandford AJ, Chagani T, Weir TD, Connett JE, Anthonisen NR, Paré PD. Susceptibility Genes for Rapid Decline of Lung Function in the Lung Health Study. Am J Respir Crit Care Med 2001; 163:469-473.

(95) Sandford AJ, Silverman EK. Chronic obstructive pulmonary disease - 1: Susceptibility factors for COPD the genotype-environment interaction. Thorax 2002; 57:736-741.

(96) Peto R, Lopez AD, Boreham J, Thun M, Heath CJr. Mortality from smoking in developed countries 1950-2000. Indirect estimates from national vital statistics. Oxford: Oxford University Press, 1994.

(97) Health Education Authority. The smoking epidemic: a manifesto for action in England. London: Health Education Authority, 1992.

(98) Coultas DB. Passive smoking and risk of adult asthma and COPD: an update. Thorax 1998; 53:381-387.

(99) Tobacco Advisory Group of the Royal College of Physicians. Nicotine addiction in Britain. London: Royal College of Physicians, 2000.

(100) Ezzati M, Lopez AD. Estimates of global mortality attributable to smoking in 2000. Lancet 2003; 362:847-852.

(101) Bohadana A, Teculescu D, Martinet Y. Mechanisms of chronic airway obstruction in smokers. Respir Med 2004; 98:139-151.

(102) Cook DG, Strachan DP, Carey IM. Parental smoking and spirometric indices in children. Thorax 1998; 53:884-893.

(103) Savanes C, Omenaas E, Jarvis D, Gulsvik A, Burney P. Parental smoking in childhood and adult obstructive lung disease: results from the European Community Respiratory Health Survey. Thorax 2004; 59:295-302.

(104) Upton MN, Davey Smith G, McConnachie A, Hart CL, Watt GCM. Maternal and Personal Cigarette Smoking Synergize to Increase Airflow Limitation in Adults. Am J Respir Crit Care Med 2004; 169:479-487.

(105) Tager IB, Munoz A, Rosner B, Weiss S, Carey V. Effect of cigarette smoking on the pulmonary function of children and adolescents. Am Rev Respir Dis 1985; 131:752-759.

(106) Gold DR, Wang X, Wypij D, Speizer FE, Ware JH, Dockery DW. Effects of cigarette smoking on lung function in adolescent boys and girls. New England Journal of Medicine 1996; 335(13):931-937.

(107) Burge PS, Calverley PMA, Jones PW, Spencer S, Anderson JA, Maslen TK. Randomised, double blind, placebo controlled study of fluticasone propionate in patients with moderate to severe chronic obstructive pulmonary disease: the ISOLDE trial. BMJ 2000; 320(13 May):1297-1303.

(108) Doll R, Peto R, Wheatley K, Gray R, Sutherland I. Mortality in relation to smoking: 40 years observations on male British doctors. BMJ 1994; 309:901-911.

(109) Lange P, Groth S, Nyboe J, Mortensen J, Appleyard M, Jensen G et al. Effects of smoking and changes in smoking habits on the decline of FEV. Eur Respir J 1989; 2:811-816.

(110) Pride NB, Scanion PD, Connett JE, Waller LA, Altose MD, Bailey WC et al. Smoking cessation: effects on symptoms, spirometry and future trends in COPD. Thorax 2001; 56(2S):ii7-ii10.

(111) Pelkonen M, Tukiainen H, Tervahauta M, Notkola I, Kivela S, Salorinne Y et al. Pulmonary function, smoking cessation and 30 year mortality in middle aged Finnish men. Thorax 2000; 55:746-750.

(112) Xu X, Weiss ST, Rijcken B, Schouten JP. Smoking, changes in smoking habits, and rate of decline in FEV1: New insight into gender differences. Eur Respir J 1994; 7(6):1056-1061.

(113) Wald N, Nicolaides-Bouman A. UK Smoking Statistics. 2nd ed. Oxford: Oxford University Press, 1991.

(114) Peach H. Smoking and respiratory disease excluding lung cancer. In: Zaridze D, Peto R, editors. Tobacco, a major international health hazard. Lyon: International Agency for Research on Cancer, 1986: 61-72.

(115) Buist AS, Vollmer WM, Wu Y, Tsai R, Johnson LR, Hurd S et al. Effects of cigarette smoking on lung function in four population samples in the People's Republic of China. Am J Respir Crit Care Med 19995; 151:1393-1400.

(116) Withey CH, Papacosta AO, Swan AV, Fitzsimons BA, Ellard GA, Burney PGJ et al. Respiratory effects of lowering tar and nicotine levels of cigarettes smoked by young male middle tar smokers. II . Results of a randomised controlled trial. J Epidemiol Community Health 1992; 46:281-285.

(117) Lee PN, Garfinkel L. Mortality and type of cigarette smoked. J Epidemiol Community Health 1981; 35:16-21.

(118) Strachan DP. Causes and control of chronic respiratory disease: looking beyond the smokescreen. J Epidemiol Community Health 1992; 46(3):177-179.

(119) Lundbäck B, Lindberg A, Lindström M, Rönmark E, Jonsson AC, Jönsson E 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.

(120) Whittemore AS, Perlin SA, DiCiccio Y. Chronic Obstructive Pulmonary Disease in lifelong nonsmokers: results from NHANES. Am J Public Health 1995; 85(5):702-706.

(121) Brown CA, Crombie IK, Tunstall-Pedoe H. Failure of cigarette smoking to explain international differences in mortality from chronic obstructive pulmonary disease. J Epidemiol Community Health 1994; 48:134-139.

(122) Becklake MR. Occupational exposures: evidence for a causal association with chronic obstructive pulmonary disease. Am Rev Respir Dis 1989; 140:S85-S91.

(123) Rudd R. Coal miner's respiratory disease litigation. Thorax 1998; 53:337-340.

(124) Oxman AD, Muir DCF, Shannon HS, Stock SR, Hnizdo E, Lange HJ. Occupational dust exposure and chronic obstructive pulmonary disease. Am Rev Respir Dis 1993; 148:38-48.

(125) Love RG, Miller BG. Longitudinal study of lung function in coal-miners. Thorax 1982; 37:193-197.

(126) Marine WM, Gurr D, Jacobsen M. Clinically important respiratory effects of dust exposure and smoking in British coal miners. Am Rev Respir Dis 1988; 137:106-112.

(127) Cochrane AL, Haley TJL, Moore F, Hole D. The mortality of men in the Rhondda Fach, 1950-1970. Br J Ind Med 1979; 36(1):15-22.

(128) Humerfelt S, Eide GE, Gulsvik A. Association of years of occupational quartz exposure with spirometric airflow limitation in Norwegian men aged 30-46 years. Thorax 1998; 53(8):649-655.

(129) Ulvestad B, Bakke B, Eduard W, Kongerud J, Lund MB. Cumulative exposure to dust causes accelerated decline in lung function in tunnel workers. Occupational and Environmental Medicine 2001; 56:663-669.

(130) Bakke B, Ulvestad B, Stewart P, Eduard W. Cumulative exposure to dust and gases as determinants of lung function decline in tunnel construction workers. Occupational and Environmental Medicine 2004; 61:262-269.

(131) Christiani DC. Organic dust exposure and chronic airways disease. Am J Respir Crit Care Med 1996; 154:833-834.

(132) Becklake MR. Asbestos-related diseases of the lungs and pleura. Am Rev Respir Dis 1982; 126(2):187-194.

(133) Hendrick DJ. Occupation and chronic obstructive pulmonary disease (COPD). Thorax 1996; 51:947-955.

(134) Omland Ø. Exposure and respiratory health in farming in temperate zones - a review of the literature. Annals of Agriculture and Environmental Medicine 2002; 9:119-136.

(135) MacNee W, Donaldson K. Mechanism of lung injury caused by PM_{10} and ultrafine particles with special reference to COPD. Eur Respir J 2003; 21(Suppl. 40):47s-51s.

(136) MacNee W, Donaldson K. Exacerbations of COPD. Chest 2000; 117(3908):3978.

(137) Schwartz J, Dockery DW. Increased mortality in Philadelphia associated with daily air pollution concentrations. Am Rev Respir Dis 1992; 145(3):600-604.

(138) Anderson HR, Spix C, Medina S, Schouten JP, Castellsague J, Rossi G et al. Air pollution and daily admissions for chronic obstructive pulmonary disease in 6 European cities: results from the APHEA project. Eur Respir J 1997; 10(5):1064-1071.

(139) Sunyer J. Urban air pollution and chronic obstructive pulmonary disease: a review. Eur Respir J 2001; 17:1024-1033.

(140) Ackermann-Liebrich U. Outdoor air pollution. In: Annesi-Maesano I, Gulsvik A, Viegi G, editors. Respiratory epidemiology in Europe. Sheffield: European Respiratory Society, 2000: 400-411.

(141) Dockery DW, Speizer FE, Stram DO, Ware JH, Spengler JD, Ferris BG. Effects of inhalable particles on respiratory health of children. Am Rev Respir Dis 1989; 139:587-594.

(142) Detels R, Tashkin DP, Sayre JW, Rokaw SN, Massey FJJr, Coulson AH et al. The UCLA population studies of CORD: X. A cohort study of changes in respiratory function associated with chronic exposure to SOx, NOx, and hydrocarbons. Am J Public Health 1991; 81(3):350-359.

(143) Hazenkamp-von Arx ME, Gotschi Fellmann T, Oglesby L, Ackermann-Liebrich U, Gislason T, Heinrich J et al. PM2.5 assessment in 21 European study centers of ECRHS II: Method and first winter results. Journal of the Air and Waste Management Association 2003; 53(5):617-628.

(144) Strachan D, Maheswaran R. Air pollution exposure in childhood and fatal chronic respiratory disease: a nationwide case-control study. (Personal communication) Summary at http://www.doh.gov.uk/comeap/statementsreports/comeapjan0101.pdf page 13. 2000.

(145) Daley C. Air pollution and causes of death. Br J Prev Soc Med 2003; 13:14-27.

(146) Stocks P. Cancer and bronchitis mortality in relation to atmospheric deposit and smoke. Br Med J 1959; i:74-79.

(147) Gardner M, Crawford M, Morris J. Patterns of mortality in middle age and early old age in the county boroughs of England and Wales. Br J Prev Soc Med 1969; 23:133-140.

(148) Chinn S, Baldwin I, Gorgol M. The relation of mortality in England and Wales 1969-73 to measurements of air pollution. J Epidemiol Community Health 1981; 35:174-179.

(149) Shaddick G, Wakefield J, Briggs D, Elliott P. Modelling the association between mortality and long term exposure to air pollution (Abstract 10447). Epidemiology 2002; 13(Suppl 4).

(150) Kleinschmidt I, Hills M, Elliott P. Smoking behaviour can be predicted by neighbourhood deprivation measures. J Epidemiol Community Health 1995; 49 (Suppl 2):S72-S77.

(151) Strachan DP. Respiratory and allergic diseases. In: Kuh D, Ben-Shlomo Y, editors. A life course approach to chronic disease epidemiology. Oxford: Oxford University Press, 1997: 101-120.

(152) Greenwell LL, Moreno T, Jones TP, Richards RJ. Particle-induced oxidative damage is ameliorated by pulmonary antioxidants. Free Radical Biology & Medicine 2002; 32(9):898-905.

(153) MacNee W. Oxidants / antioxidants and COPD. Chest 2000; 117(5 Suppl 1):303S-317S.

(154) Rahman I, Swarska E, Henry M, Stolk J, MacNee W. Is there any relationship between plasma antioxidant capacity and lung function in smokers and in patients with chronic obstructive pulmonary disease? Thorax 2000; 55(3):189-193.

(155) Heffner JE, Repine JE. Pulmonary strategies of antioxidant defense. Am Rev Respir Dis 1989; 140:531-554.

(156) Sargeant LA, Jaeckel A, Wareham NJ. Interaction of vitamin C with the relation between smoking and obstructive airways disease in EPIC Norfolk. Eur Respir J 2000; 16(3):397-403.

(157) Smit HA, Grievink L, Tabak C. Dietary influences on chronic obstructive lung disease and asthma: a review of the epidemiological evidence. Proc Nutr Soc 1999; 58:309-319.

(158) Patel BD, Jakes RW, Welch AA, Binghma SA, Luben RN, Day NE et al. Are smokers with low plasma vitamin C levels at particular risk of COPD? (Abstract S28). Thorax 2003; 58(Suppl III):iii10.

(159) Strachan DP, Cox BD, Erzinclioglu SW, Walters DE, Whichelow MJ. Ventilatory function and winter fresh fruit consumption in a random sample of British adults. Thorax 1991; 46:624-629.

(160) Britton J. Dietary fish oil and airways obstruction. Thorax 1995; 50(Suppl 1):S11-S15.

(161) Schwartz J, Weiss ST. The relationship of dietary fish intake to level of pulmonary function in the first National Health and Nutrition Survey (NHANES I). Eur Respir J 1994; 7(10):1821-1824.

(162) Tabak C, Feskens EJ, Heederik D, Kromhout D, Menotti A, Blackburn HW. Fruit and fish consumption: a possible explanation for population differences in COPD mortality (The Seven Countries Study). European Journal of Clinical Nutrition 1998; 52(11):819-825.

(163) Burney PGJ. Asthma mortality in England and Wales: Evidence for a further increase, 1974-84. Lancet 1986;(9 August):323-326.

(164) Schwartz J, Weiss ST. Dietary factors and their relation to respiratory symptoms. Am J Epidemiol 1990; 132(1):67-76.

(165) Devereux G, Beach JR, Bromly C, Avery AJ, Ayatollahi SM, Williams SM et al. Effect of dietary sodium on airways responsiveness and its importance in the epidemiology of asthma: an evaluation in three areas of northern England. Thorax 1995; 50(9):941-947.

(166) Nandram B, Sedransk J, Pickle LW. Bayesian analysis and mapping of mortality rates for chronic obstructive pulmonary disease. J Am Stat Assoc 2000; 95(452):1110-1118.

(167) Barker DJP. Mothers, Babies and Health in Later Life. 2nd ed. Edinburgh: Churchill Livingstone, 1998.

(168) 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; 303(21 September):671-675.

(169) Rona RJ, Gulliford MC, Chinn S. Effects of prematurity and intrauterine growth on respiratory health and lung function in childhood. BMJ 1993; 306(817):820.

(170) Stein CE, Kumaran K, Fall CHD, Shaheen SO, Osmond C, Barker DJP. Relation of fetal growth to adult lung function in South India. Thorax 1997; 52:895-899.

(171) Shaheen SO, Sterne JAC, Tucker JS, Florey CdV. Birth weight, childhood lower respiratory tract infection, and adult lung function. Thorax 1998; 53:549-553.

(172) Gold DR, Tager IB, Weiss ST, Tosteson TD, Speizer FE. Acute lower respiratory illness in childhood as a predictor of lung function and chronic respiratory symptoms. Am Rev Respir Dis 1989; 140:877-884.

(173) Britten N, Wadsworth J. Long term respiratory sequelae of whooping cough in a nationally representative sample. BMJ 1986; 292:441-444.

(174) Johnston IDA, Strachan DP, Anderson HR. Effect of pneumonia and whooping cough in childhood on adult lung function. New England Journal of Medicine 1998; 338(9):581-587.

(175) Shaheen SO, Barker DJP, Shiell AW, Crocker FJ, Wield GW, Holgate ST. The relationship between pneumonia in early life and impaired lung function in late adult life. Am J Respir Crit Care Med 1994; 149:616-619.

(176) Mann SL, Wadsworth MEJ, Colley JRT. Accumulation of factors influencing respiratory illness in members of a national birth cohort and their offspring. J Epidemiol Community Health 1992; 46:286-292.

(177) Barker DJP, Osmond C. Childhood respiratory infection and adult chronic bronchitis in England and Wales. BMJ 1986; 293:1271-1275.

(178) Strachan DP, Seagroatt V, Cook DG. Chest illness in infancy and chronic respiratory disease in later life: an analysis by month of birth. Int J Epidemiol 1994; 23(5):1060-1068.

(179) Britton J, Martinez FD. The relationship of childhood respiratory infection to growth and decline in lung function. Am J Respir Crit Care Med 1996; 154:S240-S245.

(180) Brown PJ, Greville HW, Finucane KE. Asthma and irreversible airflow obstruction. Thorax 1984; 39:131-136.

(181) Peat JK, Woolcock AJ, Cullen K. Rate of decline of lung function in subjects with asthma. European Journal of Respiratory Disease 1987; 70:171-179.

(182) Lange P, Ulrik CS, Vestbo J, The Copenhagen City Heart Study Group. Mortality in adults with self-reported asthma. Lancet 1996; 347(May 11):1285-1289.

(183) Djukanovic R. AHR and collagen deposition in recently diagnosed versus long-standing mild asthma. European Respiratory Journal Topic 2001; 7:32.

(184) Lewis S, Butland B, Strachan D, Bynner J, Richards D, Butler N et al. Study of the aetiology of wheezing illness at age 16 in two national British birth cohorts. Thorax 1996; 51(7):670-676.

(185) Upton MR, McConnachie A, McSharry C, Hart CL, Davey Smith G, Gillis CR et al. Intergenerational 20 year trends in the prevalences of asthma and hay fever in adults: the Midspan family study surveys of parents and offspring. BMJ 2000; 321:88-92.

(186) Keuhni CE, Davis A, Brooke AM. Are all wheezing disorders in very young (preschool) children increasing in prevalence? Lancet 2001; 357:1821-1825.

(187) Speight ANP, Lee DA, Hey EN. Underdiagnosis and undertreatment of asthma in childhood. BMJ 1983; 286:1253-1256.

(188) Kaur B, Anderson HR, Austin J, Burr M, Harkins LS, Strachan DP et al. Prevalence of asthma symptoms, diagnosis and treatment in 12-14 children across Great Britain (international study of asthma and allergies in childhood, ISAAC UK). BMJ 1998; 316:118-124.

(189) Kwong GNM, Proctor A, Duggan R, Das C, Whyte MKB, Powell CV et al. Increasing prevalence of asthma diagnosis and symptoms in children is confined to mild symptoms. Thorax 2001; 56:312-314.

(190) O'Connor GT, Sparrow D, Weiss ST. The role of allergy and nonspecific airway hyperresponsiveness in the pathogenesis of chronic obstructive pulmonary disease. Am Rev Respir Dis 1989; 140:225-252.

(191) Hospers JJ, Schouten JP, Weiss ST, Postma DS, Rijcken B. Eosinophilia is associated with increased all-cause mortality after a follow-up of 30 years in a general population sample. Epidemiology 2000; 11(3):261-268.

(192) Burr ML, Butland BK, King S, Vaughan Williams E. Changes in asthma prevalence: two surveys 15 years apart. Arch Dis Child 1989; 64:1452-1456.

(193) Powell CV, White RD, Primhak RA. Longitudinal study of free running exercise challenge: reproducibility. Arch Dis Child 1996; 74:108-114.

(194) Sunyer J, Soriano J, Antó JM, Burgos F, Pereira A, Payo F et al. Sensitization to individual allergens as risk factors for lower FEV1 in young adults. Int J Epidemiol 2000; 29:125-130.

(195) Prescott E, Vestbo J. Socioeconomic status and chronic obstructive pulmonary disease. Thorax 1999; 54:737-741.

(196) Eames M, Ben-Shlomo Y, Marmot MG. Social deprivation and premature mortality: regional comparison across England. BMJ 1993; 307:1097-1102.

(197) Jefferis BJMH, Power C, Graham H, Manor O. Changing social gradients in cigarette smoking and cessation over two decades of adult follow-up in a British birth cohort. Journal of Public Health 2004; 26(1):13-18.

(198) Marmot MG, Shipley MJ, Rose G. Inequalities in death: specific explanations of a general pattern? Lancet 1984; ii:274-276.

(199) Prescott E, Lange P, Vestbo J. Socioeconomic status, lung function and admission to hospital for COPD: results from the Copenhagen City Heart Study. Eur Respir J 1999; 13(5):1109-1114.

(200) Krzyzanowski M, Kauffmann F. The relation of respiratory symptoms and ventilatory function to moderate occupational exposure in a general population. Int J Epidemiol 1988; 17(2):397-406.

(201) Prynne CJ, Paul AA, Mishra GD, Hardy RJ, Bolton-Smith C, Wadsworth MEJ. Sociodemographic inequalities in the diet of young children in the 1946 British birth cohort. Public Health Nutrition 2002; 5(6):733-745.

(202) Wadsworth MEJ. Health inequalities in the life course perspective. Social Science & Medicine 1997; 44(6):859-869.

(203) Holland P, Berney LR, Blane D, Davey Smith G, Gunnell D, Montgomery SM. Life course accumulation of disadvantage: childhood health and hazard exposure during adulthood. Social Science & Medicine 2000; 50:1285-1295.

(204) Kauffmann F, Becklake MR. Sex and gender. In: Annesi-Maesano I, Gulsvik A, Viegi G, editors. Respiratory epidemiology in Europe. Sheffield, UK: European Respiratory Society, 2000: 288-304.

(205) Prescott E. Tobacco-related diseases: the role of gender. An epidemiologic study based on data from the Copenhagen Centre for Prospective Population Studies. Dan Med Bull 2000; 47(2):115-131.

(206) Langhammer A, Johnsen R, Gulsvik A, Holmen TL, Bjermer L. Sex differences in lung vulnerability to tobacco smoking. Eur Respir J 2003; 21:1017-1023.

(207) Leynaert B, Bosquet J, Henry C, Liard R, Neukirch F. Is bronchial hyperresponsiveness more frequent in women than in men? Am J Respir Crit Care Med 1997; 156:1413-1420.

(208) Prescott E, Bjerg AM, Andersen PK, Lange P, Vestbo J. Gender difference in smoking effects on lung function and risk of hospitalization for COPD: results from a Danish longitudinal population study. Eur Respir J 1997; 10:822-827.

(209) Rothman KJ, Greenland S. Modern Epidemiology. Philadelphia, PA: Lippincott-Raven Publishers, 1998.

(210) Elliott P, Wakefield J, Best N, Briggs D. Spatial Epidemiology. Methods and Applications. Paperback ed. Oxford: Oxford University Press, 2001.

(211) Knorr-Held L, Rainer E. Prognosis of lung cancer mortality in West Germany: a case study in Bayesian prediction. Biostatistics 2001; 2(1):109-129.

(212) Bray I, Brennan P. Recent trends and future directions for lung cancer mortality. Br J Cancer 2002; 87:43-48.

(213) Breslow N. Biostatistics and Bayes. Stat Sci 1990; 5(3):269-298.

(214) Pascutto C, Wakefield JC, Best NG, Richardson S, Bernardinelli L, Staines A et al. Statistical issues in the analysis of disease mapping data. Stat Med 2000; 19:2493-2519.

(215) Knorr-Held L, Best NG. A shared component model for detecting joint and selective clustering of two diseases. J R Stat Soc Ser A Stat Soc 2001; 164(1):73-86.

(216) Wadsworth MEJ. Changing social factors and their long-term implications for health. Br Med Bull 1997; 53(1):198-209.

(217) Blane D, Berney LR, Davey Smith G, Gunnell D, Holland P. Reconstructing the life course: health during early old age in a follow-up study based on the Boyd Orr cohort. Public Health 1999; 113:117-124.

(218) Davey Smith G, Hart C, Watt G, Hole D, Hawthorne V. Individual social class, area-based deprivation, cardiovascular disease risk factors and mortality: the Renfrew and Paisley study. J Epidemiol Community Health 1998; 52:399-405.

(219) Moser KA, Goldblatt PO. Occupational mortality of women aged 15-59 years at death in England and Wales. J Epidemiol Community Health 1991; 45(2):117-124.

(220) Ebi-Kryston K. Predicting 15 year chronic bronchitis mortality in the Whitehall study. J Epidemiol Community Health 1989; 43:168-172.

(221) Truelsen P, Prescott E, Lange P, Schnohr P, Boysen G. Lung function and risk of fatal and non-fatal stroke. The Copenhagen city heart study. Int J Epidemiol 2001; 30(1):145-151.

(222) Camilli AE, Robbins DR, Lebowitz MD. Death certificate reporting of confirmed airways obstructive disease. Am J Epidemiol 1991; 133(8):795-800.

(223) Tang J-L, Morris JK, Wald NJ, Hole D, Shipley M, Tunstall-Pedoe H. Mortality in relation to tar yield of cigarettes: a prospective study of four cohorts. BMJ 1995; 311(9 December):1530-1533.

(224) Davey Smith G, Hart C, Blane D, Hole D. Adverse socioeconomic conditions in childhood and cause specific adult mortality: prospective observational study. BMJ 1998; 316(30 May 1998):1631-1635.

(225) Marmot M, Shipley M, Brunner E, Hemingway H. Relative contribution of early life and adult socioeconomic factors to adult morbidity in the Whitehall II study. J Epidemiol Community Health 2001; 55:301-307.

(226) Kalandidi A, Trichopoulos D, Hatzakis A, Tzannes S, Saracci R. Passive smoking and chronic obstructive lung disease. Lancet 1987; 2(8571):1325-1326.

(227) Dennis RJ, Maldonado D, Norman S, Baena E, Martinez G. Woodsmoke exposure and risk for obstructive airways disease among women. Chest 1996; 109(1):115-119.

(228) Morris SE, Sale RC, Wakefield JC, Falconer S, Elliott P, Boucher BJ. Hospital admissions for asthma and chronic obstructive airways disease in east London hospitals and proximity of residence to main roads. J Epidemiol Community Health 2000; 54(1):75-76.

(229) Zuskin E, Mustajbegovic J, Schacter EN, Kanceljak B, Godnic-Cvar J, Sitar-Srebocan V. Respiratory symptoms and lung function in wool textile workers. Am J Ind Med 1995; 27(6):845-857.

(230) Bodner C, Godden D, Brown K, Little J, Ross S, Seaton A. Antioxidant intake and adult-onset wheeze: a case-control study. Eur Respir J 1999; 13(1):22-30.

(231) Dockery DW, Brunekreef B. Longitudinal studies of air pollution effects on lung function. Am J Respir Crit Care Med 1996; 154:S250-S256.

(232) Prescott-Clarke P, Primatesta P. Health Survey for England 1996. Volume 1: Findings. 1 ed. London: The Stationery Office, 1998.

(233) Scottish Office Dept of Health. Scottish Health Survey 1995. London: JOINT HEALTH SURVEYS UNIT, at The National Centre for Social Research and Department of Epidemiology & Public Health, UCL, 1997.

(234) Shaw A, McMunn A, Field J. The Scottish Health Survey 1998. Volume 1. London: Joint Health Surveys Unit, at The National Centre for Social Research and Department of Epidemiology & Public Health, UCL, 2000.

(235) Cerveri I, Accordini S, Verlato G, Corsico A, Zoia MC, Cesali L et al. Variations in the prevalence across countries of chronic bronchitis and smoking habits in young adults. Eur Respir J 2001; 18(1):85-92.

(236) Devis T, Rooney C. Death certification and the epidemiologist. Health Statistics Quarterly 1999; 01(Spring):25-33.

(237) Rooney C, Devis T. Mortality trends by cause of death in England and Wales 1980-94: the impact of introducing automated cause coding and related changes in 1993. Popul Trends 1996; 86(Winter):29-35.

(238) Hansell AL, Bottle RA, Shurlock L, Aylin P. Accessing and using hospital activity data. J Public Health Med 2001; 23(1):51-56.

(239) Goldacre MJ, Duncan M, Cook-Mozaffari P, Griffith M. Mortality rates for common respiratory disease in an English population 1979-1998: artefact and substantive trends. Journal of Public Health 2004; 26(1):8-12.

(240) van Grunsven PM, van Schayck CP, Derenne JP, Kerstjens HAM, Renkema TEJ, Postma DS et al. Long term effects of inhaled corticosteroids in chronic obstructive pulmonary disease: a meta-analysis. Thorax 1999; 54:7-14.

(241) Thun MJ, Heath CWJr. Changes in Mortality from Smoking in Two American Cancer Society Prospective Studies since 1959. Preventive Medicine 1997; 26:422-426.

(242) Wald N, Kiryluk S, Darby S, Doll R, Pike M, Peto R. UK Smoking Statistics. Oxford: Oxford University Press, 1988.

(243) Forey BA, Lee PN, Fry JS. Updating UK estimates of age, sex and period specific cumulative constant tar cigarette consumption per adult. Thorax 1998; 53:875-878.

(244) Harrison RM. Pollution: Causes, effects & control. 2nd Edition ed. Cambridge: Royal Society of Chemistry, 1990.

(245) Quality of Urban Air Review Group. Airborne Particulate Matter in the United Kingdom. Birmingham: Institute of Public and Environmental Health, School of Chemistry, University of Birmingham, 1996.

(246) Ayres JG. Trends in air quality in the UK. Allergy 1997; 52(Suppl 38):7-13.

(247) Douglas JWB, Waller RE. Air pollution and respiratory infection in children. Br J Prev Soc Med 1996; 20:1-8.

(248) Coggon D, Inskip H, Winter P, Pannett B. Contrasting geographical distribution of mortality from pneumoconiosis and chronic bronchitis and emphysema in British coal miners. Occupational & Environmental Medicine 1995; 52(8):554-555.

(249) Gardner MJ, Winter PD, Taylor CP, Acheson ED. Atlas of cancer mortality in England & Wales 1968-78. Chichester: John Wiley & Sons, 1983.

(250) Ministry of Agriculture Fisheries and Food. National Food Survey 1997. London: The Stationery Office, 1998.

(251) Billson H, Pryer JA, Nichols R. Variation in fruit and vegetable consumption among adults in Britain. An analysis from the dietary and nutritional survey of British adults. European Journal of Clinical Nutrition 1999; 53:946-952.

(252) Henderson L, Gregory J, Swan G. The National Diet and Nutrition Survey: adults aged 19 to 64. London: HMSO, 2002.

(253) Robertson C, Best N, Diamond J, Elliott P. Tracing ingestion of 'novel' foods in UK diets for possible health surveillance - a feasibility study. Public Health Nutrition 2004; 7(2):345-352.

(254) European Commission Joint Research Centre. Monitoring Agriculture with Remote Sensing Unit. http://mars.jrc.it/. 2004.

(255) Morris R, Carstairs V. Which deprivation? A comparison of selected deprivation indexes. J Public Health Med 1991; 13(4):318-326.

(256) Noble B, Drever F, Shah N. Source and methods. In: Drever F, editor. Occupational health: decennial supplement. The Registrar General's decennial supplement for England and Wales. London: OPCS, 1995: 5-15.

(257) General Register Office for Scotland. GROS 1996 Annual Report. http://www.gro-scotland.gov.uk/grosweb/grosweb.nsf/pages/96_appl . 2004

(258) Guite HF, Burney PGJ. Accuracy of recording of deaths from asthma in the UK: the false negative rate. Thorax 1996; 51:924-928.

(259) Smyth ET, Wright SC, Evans AE, Sinnamon DG, MacMahon J. Death from airways obstruction: accuracy of certification in Northern Ireland. Thorax 1996; 51:293-297.

(260) Jones K, Berrill WT, Bromly CL, Hendrick DJ. A confidential enquiry into certified asthma deaths in the North of England, 1994-96: influence of co-morbidity and diagnostic inaccuracy. Respir Med 1999; 93(12):923-927.

(261) Office for National Statistics. Mortality statistics. Cause. 1993 (revised) and 1994. Series DH2, no 21. London: HMSO, 1996.

(262) Charlton J, Murphy M. Monitoring health - data sources and methods. In: Charlton J, Murphy M, editors. The Health of Adult Britain 1841-1994. London: The Stationery Office, 1997: 2-16.

(263) Mannino DM, Buist AS, Petty TL, Enright PL, Redd SC. Lung function and mortality in the United States: data from the First National Health and Nutrition Examination Survey follow up study. Thorax 2003; 58:388-393.

(264) Euler GL, Abbey DE, Magie AR, Hodgkin JE. Chronic obstructive pulmonary disease symptom effects of long-term cumulative exposure to ambient levels of total suspended particulates and sulfur dioxide in California Seventh-Day Adventist residents. Arch Environ Health 1987; 42(4):213-222.

(265) Pope CA3, Burnett RT, Thun MJ, Calle EE, Krewski D, Ito K et al. Lung cancer, cardiopulmonary mortality and long-term exposure to fine particulate air pollution. J Am Med Assoc 2002; 287(9):1132-1141.

(266) Quality of Urban Air Review Group. Urban Air Quality in the United Kingdom. First Report of the Quality of Urban Air Review Group. Bradford, West Yorkshire: Dept of the Environment, 1993.

(267) Taylor Nelson Sofres. Superpanel Home Page. https://www.tnsinfo.com/tnsinfo/Rooms/DisplayPages/LayoutInitial?Container=com.webridge.entity.Entity%5BOID%5BD8C4419C35C873498FAF113EBC0DE9D9%5D%5D. 2004

(268) Food Standards Agency. Five a day http://www.foodstandards.gov.uk/healthiereating/asktheexpert/fruitandveg/fivebackground. 2004.

(269) Carey IM, Strachan DP, Cook DG. Effects of changes in fresh fruit consumption on ventilatory function in healthy British adults. Am J Respir Crit Care Med 1998; 158:728-733.

(270) Cook DG, Carey IM, Whincup PH, Papacosta O, Chirico S, Bruckdorfer KR et al. Effect of fresh fruit consumption on lung function and wheeze in children. Thorax 1997; 52(628):633.

(271) Wenlock RW, Buss DH, Derry BJ. Household food wastage in Britain. British Journal of Nutrition 1980; 43(1):53-70.

(272) Fifty years of the National Food Survey 1940-1990. London: HMSO, 1991.

(273) Ministry of Agriculture Fisheries and Food. National Food Survey 2000. London: The Stationery Office, 2001.

(274) Ministry of Agriculture Fisheries and Food. National Food Survey 1992. London: The Stationery Office, 1993.

(275) Carstairs V. Deprivation indices: their interpretation and use in relation to health. J Epidemiol Community Health 1995; 49(Suppl 2):S3-S8.

(276) Sloggett A, Joshi H. Higher mortality in deprived areas: community or personal disadvantage. BMJ 1994; 309:1470-1474.

(277) Macintyre S. Geographical inequalities in mortality, morbidity and health-related behaviour in England. In: Gordon D, Shaw M, Dorling D, Davey Smith G, editors. Inequalities in health. The evidence presented to the Independent Inquiry into Inequalities in Health, chaired by Sir Donald Acheson. Bristol: The Policy Press, 1999: 148-154.

(278) Bland M. An introduction to medical statistics. Oxford: Oxford University Press, 1995.

(279) Gelman AB, Carlin JS, Stern HS, Rubin DB. Bayesian data analysis. London: Chapman & Hall/CRC, 1995.

(280) winBUGS version 1.4 (http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml). UK: MRC Biostastistics Unit, Cambridge and Imperial College London, 2004.

(281) Bayesian Age-Period-Cohort Modelling and Prediction (BAMP) Ludwig-Maximilians-Universität, München, Deutschland: http://www.stat.uni-muenchen.de/~schmidt/bamp/ . 2004.

(282) Clifford RE, Smith JWG, Tillett HE, Wherry PJ. Excess Mortality Associated with Influenza in England and Wales. Int J Epidemiol 1977; 6(2):115-128.

(283) Tillett HE, Smith JWG, Gooch CD. Excess Deaths Attributable to Influenza in England and Wales: Age at Death and Certified Cause. Int J Epidemiol 1983; 12(3):344-352.

(284) Coggon D, Inskip H, Winter P, Pannett B. Occupational mortality of men. In: Drever F, editor. Occupational health: decennial supplement. The Registrar General's decennial supplement for England and Wales. London: OPCS, 1995: 23-42.

(285) Department of Scientific and Industrial Research. The Investigation of Atmospheric Pollution. London: HMSO, 1960.

(286) Clayton D, Schifflers E. Models for temporal variation in cancer rates. II: Age-period-cohort models. Stat Med 1987; 6:469-481.

(287) Bray I, Brennan P, Boffetta P. Projections of alcohol- and tobacco-related cancer mortality in central Europe. Int J Cancer 2000; 87:122-128.

(288) Breuninger R. Modellierung von heterogenen Zeittrends in Alters-Perioden-Kohorten-Modellen. München: Institut für Statistik, Ludwig-Maximilians-Universität, 2003.

(289) Tango T, Kurashina S. Age, period and cohort analysis of trends in mortality from major diseases in Japan, 1955 to 1979: peculiarity of the cohort born in the early Showa era. Stat Med 1987; 6:709-726.

(290) Atmospheric pollution. Nature 1914; 94(2355 (December 17)):433-434.

(291) Shaw N. The treatment of smoke: a sanitary parallel (quoted in Brimblecombe 1987). Nature 1891; 66:667-670.

(292) Harrison RM. Pollution: Causes, Effects and Control. 4th Edition ed. Cambridge: Royal Society of Chemistry, 2001.

(293) Logan WPD. Mortality in the London fog incident, 1952. Lancet 1953; i:336-338.

(294) Bell ML, Davis DL. Reassessment of the lethal London fog of 1952: novel indicators of acute and chronic consequences of acute exposure to air pollution. Environ Health Perspect 2001; 109(3):389-394.

(295) Longhurst JWS, Mann C. Changing Air Quality in Greater Manchester. Manchester: ARIC, 1994.

(296) Laxen DPH, Thompson MA. Sulphur Dioxide in Greater London, 1931-1985. Environmental Pollution 1987; 43:103-114.

(297) Holland WW, Bennett AE, Cameron IR, Florey CdV, Leeder SR, Schilling RSF et al. Health effects of particulate air pollution: reappraising the evidence. Am J Epidemiol 1979; 110:525-639.

(298) Anderson HR, Limb ES, Bland JM, de Leon AP, Strachan DP, Bower JS. Health effects of an air pollution episode in London, December 1991. Thorax 1995; 50(11):1188-1193.

(299) Coggon D, Inskip H, Winter P, Pannett B. Occupational mortality by cause of death. In: Drever F, editor. Occupational health: decennial supplement. The Registrar General's decennial supplement for England and Wales. London: OPCS, 1995: 62-76.

(300) Hutchings S, Jones J, Hodgson J. Asbestos-related disease. In: Drever F, editor. Occupational Health. Decennial Supplement. DS no.10. London: HMSO, 1995: 127-147.

(301) 1998 Index of Local Deprivation. London, Office of the Deputy Prime Minister, Department of the Environment, Transport and the Regions, 1998.

(302) Index of Multiple Deprivation 2000. London, Department of the Environment Transport and the Regions, 2000.

(303) Stedman JR. Revised High Resolution Maps of Background Air Pollutant Concentrations in the UK: 1996. 20008001/006. AEAT - 3133. 1998. Abingdon, Oxfordshire UK, AEA Technology, National Environmental Technology Centre.

(304) van der Groot E. Technical description of interpolation and processing of meteorological data in CGMS. European Commission Joint Research Centre: Monitoring Agriculture with Remote Sensing (MARS) unit, 1997.

(305) Wakefield JC, Best NG, Waller LA. Bayesian approaches to disease mapping. In: Elliott P, Wakefield JC, Best NG, Briggs DJ, editors. Spatial Epidemiology: Methods and Applications. Oxford: Oxford University Press, 2000.

(306) Besag J, York J, Mollié A. Bayesian image restoration with two applications in spatial statistics. Ann Inst Statist Math 1991; 43(1):1-59.

(307) Richardson S, Thomson A, Best N, Elliott P. Interpreting Posterior Relative Risk Estimates in Disease-Mapping Studies. Environ Health Perspect 2004; 112(9):1016-1025.

(308) Clancy L, Goodman P, Sinclair H, Dockery DW. Effect of air-pollution control on death rates in Dublin, Ireland: an intervention study. Lancet 2002; 360:1210-1214.

(309) Hedley AJ, Wong CM, Thach TQ, Ma S, Lam TH, Anderson HR. Cardiorespiratory and all-cause mortality after restrictions on sulphur content of fuel in Hong Kong: an intervention study. Lancet 2002; 360:1646-1652.

(310) Alberg AJ, Samet JM. Epidemiology of lung cancer. Chest 2003; 123(Suppl 1):21S-49S.

(311) Stevens RG, Moolgavkar SH. A cohort analysis of lung cancer and smoking in British males. Am J Epidemiol 1984; 119:624-641.

(312) Brønnum-Hansen H, Knud J. Estimating Mortality Due to Cigarette Smoking: Two Methods, Same Result. Epidemiology 2000; 11(4):422-426.

(313) Peto R, Darby S, Deo H, Silcocks P, Whitley E, Doll R. Smoking, smoking cessation, and lung cancer in the UK since 1950: combination of national statistics with two case-control studies. BMJ 2000; 321(5 August):323-329.

(314) Pope CA3, Thun MJ, Namboodiri MM, Dockery DW, Evans JS, Speizer FE et al. Particulate air pollution as a predictor of mortality in a prospective study of U.S. adults. Am J Respir Crit Care Med 1995.

(315) Beeson WL, Abbey DE, Knutsen SF. Long-term concentrations of ambient air pollutants and incident lung cancer in California adults: results from the AHSMOG study. Environ Health Perspect 1998; 106(12):813-823.

(316) He X, Chen W, Liu Z, Chapman RS. An epidemiological study of lung cancer in Xuan Wei County, China: current progress. Case-control study on lung cancer and cooking fuel [conference proceeding]. Environ Health Perspect 1991; 94(August):9-13.

(317) Cohen AJ. Outdoor air pollution and lung cancer. Environ Health Perspect 2000; 108(Suppl 4):743-750.

(318) Cederlöf R, Doll R, Fowler B, Freiberg L, Nelson N, Vouk V. Air pollution and cancer: risk assessment methodology and epidemiological evidence. Environ Health Perspect 1978; 22:1-12.

(319) Davis RM, Novotny TE. The epidemiology of cigarette smoking and its impact on chronic obstructive pulmonary disease. Am Rev Respir Dis 1989; 140:S82-S84.

(320) Izumi T. Chronic obstructive pulmonary disease in Japan. Curr Opin Pulm Med 2002; 8:102-105.

(321) Richardson S, Monfort C. Ecological correlation studies. In: Elliott P, Wakefield JC, Best NG, Briggs DJ, editors. Spatial Epidemiology. Methods and Applications. Oxford: Oxford University Press, 2000: 205-220.

(322) Pope CA3, Schwartz J, Ransom MR. Daily mortality and PM_{10} pollution in Utah Valley. Arch Environ Health 1992; 47:211-217.

(323) Katsouyanni K, Touloumi G, Spix C, Schwartz J, Baldacci S, Medina S et al. Short-term effects of ambient sulphur dioxide and particulate matter on mortality in 12 European cities: results from time series data from the APHEA project. Air Pollution and Health: a European Approach. BMJ 1997; 314:1658-1663.

(324) Logan WPD, Benjamin B. Loss of Expected years of life - A perspective view of changes between 1848-72 and 1952. Monthly Bulletin of the Ministry of Health and the Public Health Laboratory Service 1953; 12:244-252.

(325) Macfarlane A. Daily mortality and environment in English conurbations. 1: Air pollution, low temperature, and influenza in Greater London. Br J Prev Soc Med 1977; 31:54-61.

(326) Churg A, Brauer M, del Carmen Avila-Casado M, Fortoul TI, Wright JL. Chronic exposure to high levels of particulate air pollution and small airway remodeling. Environ Health Perspect 2003; 111:714-718.

(327) Ackermann-Liebrich U, Leunberger P, Schwartz J, Schindler C, Monn C, Bolognini G et al. Lung function and long term exposure to air pollutants in Switzerland. Am J Respir Crit Care Med 1997; 155:122-129.

(328) Schwartz J. Lung function and chronic exposure to air pollution. A cross-sectional analysis of NHANES II. Environmental research 1989; 50:309-321.

(329) Frye C, Hoelscher B, Cyrys J, Wijst M, Wichmann HE, Heinrich J. Association of lung function with declining ambient air pollution. Environ Health Perspect 2003; 111:383-387.

(330) Karakatsani A, Andreadaki S, Katsouyanni K, Dimitroulis I, Trichopoulos D, Benetou V et al. Air pollution in relation to manifestations of chronic pulmonary disease: A nested case-control study in Athens, Greece. European Journal of Epidemiology 2003; 18:45-53.

(331) Sunyer J, Atkinson R, Ballester F, Le Tertre A, Ayres JG, Forastiere F et al. Respiratory effects of sulphur dioxide: a hierarchical multicity analyses in the APHEA 2 study (electronic paper). Occupational and Environmental Medicine 2003; 60(8):e2.

(332) Sunyer J, Basagaña X. Particles, and not gases, are associated with the risk of death in patients with chronic obstructive pulmonary disease. Int J Epidemiol 2001; 30:1138-1140.

(333) Air Quality Expert Group. Nitrogen dioxide in the United Kingdom. London: Department for Environment, Food and Rural Affairs, 2004.

(334) Photochemical Review Group. Ozone in the United Kingdom. Fourth Report of the Photochemical Review Group, 1997. Edinburgh: Dept of the Environment, Transport and the Regions; Centre for Ecology and Hydrology, 1997.

(335) Ito K, Thurston GD. Characterization and Reconstruction of Historical London, England, Acidic Aerosol Concentrations. Environ Health Perspect 1989; 79:35-42.

(336) Advisory Group on the Medical Aspects of Air Pollution Episodes. Health effects of exposures to mixtures of pollutants. London: HMSO, 1995.

(337) Bell ML, Davis DL, Fletcher T. A Retrospective Assessment of Mortality from the London Smog Episode of 1952: The Role of Influenza and Pollution. Environ Health Perspect 2004; 112(1):6-8.

(338) Schwartz J. Harvesting and long term exposure effects in the relation between air pollution and mortality. Am J Epidemiol 2000; 151(5):440-448.

(339) Seemungal T, Harper-Owen R, Bhowmik A, Moric I, Sanderson G, Message S et al. Respiratory viruses, symptoms, and inflammatory markers in acute exacerbations and stable chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2001; 164(9):1618-1623.

(340) Groenewegen KH, Schols AMWJ, Wouters EFM. Mortality and Mortality-Related Factors After Hospitalization for Acute Exacerbation of COPD. Chest 2003; 124(2):459-467.

(341) Lipfert FW, Morris SC. Temporal and spatial relations between age specific mortality and ambient air quality in the United States: regression results for counties, 1960-97. Occupational and Environmental Medicine 2002; 59:156-174.

(342) Erens B, Primatesta P, Prior G. Health Survey for England - The Health of Minority Ethnic Groups '99. London: The Stationery Office, 2001.

(343) Doll R. Atmospheric Pollution and Lung Cancer. Environ Health Perspect 1978; 22:23-31.

(344) Harris JE, Thun MJ, Mondul AM, Calle EE. Cigarette tar yields in relation to mortality from lung cancer in the cancer prevention study II prospective cohort 1982-8. BMJ 2004; 328(10 January):72-76.

(345) Stellman SD, Muscat JE, Hoffmann D, Wynder EL. Impact of Filter Cigarette Smoking on Lung Cancer Histology. Preventive Medicine 1997; 26:451-456.

(346) Timeline of antibiotics. The Free Dictionary.com . http://encyclopedia.thefreedictionary.com/Timeline%20of%20antibiotics. 2004.

(347) McVay L. Antibiotic prophylaxis in chronic respiratory diseases. Arch Intern Med 1953; 92:833-846.

(348) Edwards G, Buckley AR, Fear EC, Williamson GM, Zinnemann K. Adult Chronic Bronchitis - the Infective Factor and its Treatment. Br Med J 1957; ii(Aug 3):259-264.

(349) Buchanan J. Long-term prophylactic administration of tetracycline to chronic bronchitics. Lancet 1958; ii(October):719-723.

(350) Moyes EN, Kershaw RA. Long-continued treatment with tetracycline and prednisolone in chronic bronchitis. Lancet 2004; ii(Dec. 14):1187-1191.

(351) Pridie R. A trial of continuous winter chemotherapy in chronic bronchitis. Lancet 1960; ii(October 1):723-727.

(352) Staykova T, Black P, Chacko E, Ram FSF, Poole P. Prophylactic antibiotic therapy for chronic bronchitis (Cochrane Review). The Cochrane Library, Issue 2. Chichester, UK: John Wiley & Sons, Ltd, 2004.

(353) Herxheimer H. Aleudrine and Anthisan in bronchial spasm. Lancet 1948; 1:667-671.

(354) Beasley R, Smith K, Pearce N, Crane J, Burgess C, Culling C. Trends in asthma mortality in New Zealand, 1908-1986. Medical Journal of Australia 1990; 152:570-573.

(355) Sestini P, Renzoni E, Robinson S, Poole P, Ram FSF. Short-acting beta 2 agonists for stable chronic obstructive pulmonary disease (Cochrane Review). The Cochrane Library, Issue 2. Chichester, UK: John Wiley & Sons, Ltd, 2004.

(356) Appleton S, Poole P, Smith B, Veale A, Bara A. Long-acting beta2-agonists for chronic obstructive pulmonary disease patients with poorly reversible airflow limitation (Cochrane Review). The Cochrane Library, Issue 2. Chichester, UK: John Wiley & Sons, Ltd, 2004.

(357) Baigelman W, Chodosh S. Bronchodilator action of the anticholinergic drug, ipatropium bromide (Sch 1000), as an aerosol in chronic bronchitis and asthma. Chest 1977; 71(3):324-328.

(358) Wood-Baker R, Walters EH, Gibson P. Oral corticosteroids for acute exacerbations of chronic obstructive pulmonary disease (Cochrane Review). The Cochrane Library, Issue 2. Chichester, UK: John Wiley & Sons, Ltd, 2004.

(359) Ram FSF, Jones PW, Castro AA, de Brito Jardim JR, Atallah AN, Lacasse Y et al. Oral theophylline for chronic obstructive pulmonary disease (Cochrane Review). The Cochrane Library, Issue 2. Chichester, UK: John Wiley & Sons, Ltd, 2004.

(360) Atuk NO, Blaydes MC, Westervelt FB, Wood JEJr. Effect of aminophylline on urinary excretion of epinephrine and norepinephrine in man. Circulation 1967; 35(4):745-753.

(361) Barr RG, Rowe BH, Camargo CAJr. Methylxanthines for exacerbations of chronic obstructive pulmonary disease (Cochrane Review). The Cochrane Library, Issue 2. Chichester, UK: John Wiley & Sons, Ltd, 2004.

(362) Joint Working Party of the Medical Research Council's Committee on Influenza and Other Respiratory Virus Vaccines, Research Committee of the British Tuberculosis Association. Field Trial of Influenza Virus Vaccine in Patients with Chronic Bronchitis during the Winter 1957-8. Br Med J 1959; ii(Nov 7):905-908.

(363) Poole PJ, Chacko E, Wood-Baker RWB, Cates CJ. Influenza vaccine for patients with chronic obstructive pulmonary disease (Cochrane Review). The Cochrane Library, Issue 2. Chichester, UK: John Wiley & Sons, Ltd, 2004.

(364) Leech JA, Gervais A, Ruben FL. Efficacy of pneumococcal vaccine in severe chronic obstructive pulmonary disease. Can Med Assoc J 1987; 136:361-365.

(365) Danuser B, Weber C, Künzli N, Schindler C, Nowak D. Respiratory symptoms in Swiss farmers: An epidemiological study of risk factors. American Journal of Industrial Medicine 39[4], 410-418. 2001.

(366) Charlton J, Quaife K. Trends in diet 1841-1994. In: Charlton J, Murphy M, editors. The health of adult Britain 1841-1994. London: The Stationery Office, 1997: 93-113.

(367) Kemm JR. A birth cohort analysis of smoking by adults in Great Britain 1974-1998. J Public Health Med 2001; 23(4):306-311.

(368) Clayton D, Schifflers E. Models for temporal variation in cancer rates. I: Age-period and age-cohort models. Stat Med 1987; 6:449-467.

(369) Office for National Statistics. Internal migration, (constituent countries of the United Kingdom and Government Office Regions of England): Population Trends 116. Dataset Name: PT11681. Office for National Statistics, editor. 2004. London, Office for National Statistics.

(370) Maheswaran R, Strachan DP, Dodgeon B, Best NG. A population-based case-control study for examining early life influences on geographical variation in adult mortality in England & Wales using stomach cancer and stroke as examples. Int J Epidemiol 2002; 31:375-382.

(371) Harding S, Maxwell R. Differences in mortality of migrants. In: Drever F, Whitehead M, editors. Health Inequalities. Decennial Supplement. DS No. 15. London: The Stationery Office, 1997: 108-121.

(372) Office of Population Censuses and Surveys. The General Household Survey. London: HMSO, 1972.

(373) Abbey DE, Mills PK, Petersen FF, Beeson WL. Long-term ambient concentrations of total suspended particulates and oxidants as related to incidence of chronic disease in California Seventh-Day Adventists. Environ Health Perspect 1991; 94:43-50.

(374) International Agency for Research on Cancer (IARC). IARC monographs on the evaluation of carcinogenic risks to humans: silica, some silicated, coal dust and para-aramid fibrils. 68. 1997. Lyon, France, World Health Organization, IARC.

(375) Treasure T, Waller D, Swift S, Peto J. Radical surgery for mesothelioma. BMJ 2004; 328:237-238.

(376) Dorling D. Death in Britain. How local mortality rates have changed: 1950s-1990s. York: Joseph Rowntree Foundation, 1997.

(377) Power C, Hertzman C. Social and biological pathways linking early life and adult disease. Br Med Bull 1997; 53(1):210-221.

(378) Shaw M, Dorling D, Gordon D, Davey Smith G. The widening gap. Bristol: The Policy Press, 1999.

(379) Jarvis MJ, Wardle J. Social patterning of individual health behaviours: the case of cigarette smoking. In: Marmot M, Wilkinson RG, editors. Social determinants of health. Oxford: Oxford University Press, 1999: 240-255.

(380) Krewski D, Burnett RT, Goldberg MS, Hoover K, Siemiatycki J, Jerrett M et al. Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality. Boston MA: Health Effects Institute 2000 (errata 2001 at http://www.healtheffects.org/pubs-special.htm), 2000.

(381) Tabak C, Smit HA, Heederik D, Ocke MC, Kromhout D. Diet and chronic disease obstructive pulmonary disease: independent beneficial effects of fruits, whole grains and alcohol (the MORGEN study). Clin Exp Allergy 2001; 31(5):747-755.

(382) McKeever TM, Schrivener S, Broadfield E, Jones Z, Britton J, Lewis SA. Prospective study of diet and decline in lung function in a general population. Am J Respir Crit Care Med 2002; 165:1299-1303.

(383) Walda IC, Tabak C, Smit HA, Räsänen L, Fidanza F, Menotti A et al. Diet and 20-year chronic obstructive pulmonary disease mortality in middle-aged men from three European countries. European Journal of Clinical Nutrition 2002; 56(July):638-643.

(384) Lawlor DA, Davey Smith G, Bruckdorfer KR, Kundu D, Ebrahim S. Those confounded vitamins: what can we learn from the differences between observational versus randomised trial evidence? Lancet 2004; 363:1724-1727.

(385) Crofton EC. A study of lung cancer and bronchitis mortality in relation to coal-mining in Scotland. Br J Prev Soc Med 1969; 23:141-144.

(386) World Health Organization. International Statistical Classification of Diseases and Related Health Problems. Tenth Revision. Geneva: World Health Organization, 1992.

(387) Chronic Bronchitis. Br Med J 1959; i(Jan 17):157-158.

(388) Kelly FJ, Sandström T. Air pollution, oxidative stress, and allergic response. Lancet 2004; 363(January 10):95-96.

(389) Lewis JD, Brensinger C. Agreement between GPRD smoking data: a survey of general practitioners and a population-based survey. Pharmacoepidemiology and drug safety. In press.

(390) Moscow 'suffocated' by smog. BBC News . 2002. http://news.bbc.co.uk/2/hi/europe/2238080.stm

(391) Bushfire smog shrouds Sydney. BBC News 2002 Nov 8. http://news.bbc.co.uk/2/hi/asia-pacific/2419525.stm

Appendix

List of Tables

Table A-1 ICD1 To ICD9 Codes For Chronic Obstructive Pulmonary Disease And Allied Conditions (Including Asthma) 348

Table A-2 Bridge-Coding Conversion Factors ICD6 To ICD7 And ICD7 To ICD8 For COPD Excluding Asthma And Lung Cancer.....349

Table A-3 ICD5 To ICD9 Codes Used For Mortality In Spatial And Time Trend Analyses.....350

Table A-4 Spearman Rank Correlation Coefficients Between Explanatory Variables Used In The Linear Regression Analyses In Chapter 7 351

Table A-5 Numbers Of COPD Deaths In England & Wales 1950-1996 For All Ages By ICD Code.....352

List of Figures

Figure A-1 Relative differences between age-standardised rates for lung cancer mortality for conurbations (excluding Greater London) and Greater London compared with non-conurbations, 1950-1999 353

Figure A-2 Map of smoothed COPD relative risks (combined spatially structured and unstructured parameters from shared component analyses) for UK districts 1981-1999 – males 354

Figure A-3 Map of smoothed COPD relative risks (combined spatially structured and unstructured parameters from shared component analyses) for UK districts 1981-1999 – females 355

Figure A-4 Map of smoothed lung cancer relative risks (combined spatially structured and unstructured parameters from shared component analyses) for UK districts 1981-1999 – males 356

Figure A-5 Map of smoothed lung cancer relative risks (combined spatially structured and unstructured parameters from shared component analyses) for UK districts 1981-1999 – females 357

Figure A-6 Map of deprivation – percentage of district population in 1991 living in wards in the most deprived Carstairs quintile 358

Figure A-7 Map of interpolated annual PM_{10} levels in UK districts in 1996 359

Figure A-8 Map of interpolated annual SO_2 levels in UK districts in 1996 360

Figure A-9 Map of average daily minimum temperature (°C) in the UK 1981-1999 361

Figure A-10 Map of average daily rainfall (mm) in the UK 1981-1999 362

Figure A-11 Map of pneumoconiosis SMRs in UK districts 1981-1999 – males .. 363

Figure A-12 Map of pneumoconiosis SMRs in UK districts 1981-1999 – female .. 364

Figure A-13 Map of asbestos-related disease SMRs in UK districts 1981-1999 – males 365

Figure A-14 Map of asbestos-related disease SMRs in UK districts 1981-1999 – female 366

Figure A-15 Map of average fruit and vegetable purchases for UK districts 1991-2000 367

Figure A-16 Household consumption of fruit & vegetables (weight per week in g) from the National Food Surveys 1942-2000 368

Table A-1 ICD1 to ICD9 codes for chronic obstructive pulmonary disease and allied conditions (including asthma)

Code (years used in UK mortality coding)	Description
ICD9 (1979-2000)	
490	Bronchitis, unspecified
491	Chronic bronchitis
492	Emphysema
493	Asthma
494	Bronchiectasis
495	Extrinsic allergic alveolitis
496	Chronic airways obstruction, not elsewhere classified
ICD8 (1968-78)	
493	Asthma
490	Bronchitis, unqualified
491	Chronic bronchitis
492	Emphysema
518	Bronchiectasis
516.1	Other pneumoconioses and related diseases due to inhalation of other dust
502.0 & 502.1	Chronic bronchitis
526.0	Bronchiectasis
519.8	Chronic airways obstruction (introduced in 1978 – ICD first published in 1967)
ICD7 (1958-67)	
241	Asthma
501	Bronchitis, unqualified
502.0 & 502.1	Chronic bronchitis
526	Bronchiectasis
527.1	Other diseases of lung and pleural cavity: emphysema without mention of bronchitis
ICD6 (1950-57)	
241.0	Asthma
501.0	Bronchitis, unqualified
502.0 & 502.1	Chronic bronchitis
526.0	Bronchiectasis
ICD5 (1940-49)	
106b	Chronic bronchitis
106c	Bronchitis not distinguished as acute or chronic
112 (6)	Asthma without any of the complications specified in (1-5) i.e. influenza, chronic endocarditis, myocardial disease, arteriosclerosis, chronic nephritis
113	Pulmonary emphysema
ICD4 (1931-39)	
106b	Chronic bronchitis
106c	Bronchitis not distinguished as acute or chronic
112	Asthma
113	Pulmonary emphysema
ICD3 (1921-1930)	
99b,c,d	Chronic bronchitis and bronchitis not distinguished as acute or chronic
105	Asthma
106	Pulmonary emphysema
ICD2 (1911-20)	
89,90	Bronchiectasis, bronchial catarrh, other bronchitis
96	Asthma
97	Pulmonary emphysema
ICD1 (1901-10)	
119	Emphysema, asthma

Table A-2 Bridge-coding conversion factors ICD6 to ICD7 and ICD7 to ICD8 for COPD excluding asthma and lung cancer (for males and females)

	COPD excluding asthma				Lung cancer			
	Male		Female		Male		Female	
Age (years)	ICD6to7	ICD7to8	ICD6to7	ICD7to8	ICD6to7	ICD7to8	ICD6to7	ICD7to8
0-4	1.13	1.06	1.09	1.01	0.97	1.00	1.00	1.00
5-9	1.13	1.06	1.09	1.01	0.97	1.00	1.00	1.00
10-14	1.13	1.06	1.09	1.01	0.97	1.00	1.00	1.00
15-19	1.13	1.06	1.09	1.01	0.97	1.00	1.00	1.00
20-24	1.13	1.06	1.09	1.01	0.97	1.00	1.00	1.00
25-29	1.13	1.06	1.09	1.01	0.97	1.00	1.00	1.00
30-34	1.13	1.06	1.09	1.01	0.97	1.00	1.00	1.00
35-39	1.08	1.04	1.25	1.07	1.00	0.98	1.00	0.98
40-44	1.10	1.05	1.20	0.98	0.99	1.00	1.00	1.01
45-49	1.15	1.03	1.19	1.03	1.00	1.00	1.01	0.98
50-54	1.08	1.01	1.07	1.02	1.00	1.00	0.98	1.00
55-59	1.06	1.01	1.09	1.02	1.00	1.00	0.99	0.99
60-64	1.07	1.01	1.09	1.02	1.00	1.00	1.00	1.00
65-69	1.07	1.01	1.09	1.00	1.00	1.00	1.01	1.00
70-74	1.07	1.00	1.06	1.00	1.00	1.00	1.00	0.99
75-79	1.05	1.00	1.04	1.01	1.00	1.00	1.00	1.00
80-84	1.05	1.01	1.03	1.01	1.00	1.00	1.00	0.99
85+	1.06	1.00	1.02	1.00	1.00	1.00	1.00	0.99

Conversion factors for ICD6 to ICD7 obtained from Registrar General's publications in 1957[RGI957III]
Conversion factors for ICD7 to ICD8 obtained from Registrar General publications in 1967[RG1967I, RGI967III]

Table A-3 ICD5 to ICD9 codes used for mortality in spatial and time trend analyses

Description	ICD codes
ICD9 1979-2000	
COPD	7
Bronchitis, unspecified	490
Chronic bronchitis	491
Emphysema	492
Asthma	493
Bronchiectasis	494
Extrinsic allergic alveolitis	495
Chronic airways obstruction, not elsewhere classified	496
Lung cancer	162
Pneumoconioses	
Coalworkers' Pneumoconiosis	500
Pneumoconiosis Due To Other Silica Or Silicates	502
Pneumoconiosis Due To Other Inorganic Dust	503
Pneumopathy Due To Inhalation Of Other Dust	504
Pneumoconiosis, Unspecified	505
Asbestos-related diseases	
Specified Parts Of Peritoneum	1588
Peritoneum, Unspecified	1589
Malignant Neoplasm Of Pleura	163
Asbestosis	501
ICD8 1968-78	
COPD	
Bronchitis, unqualified	490
Chronic bronchitis	491
Emphysema	492
Asthma	493
Bronchiectasis	518
Chronic airways obstruction	519.8
ICD7 1958-67	
COPD	
Asthma	241
Bronchitis, unqualified	501
Chronic bronchitis	502.0 & 502.1
Bronchiectasis	526
Emphysema without mention of bronchitis	527.1
ICD6 1950-57	
COPD	
Asthma	241
Bronchitis, unqualified	501
Chronic bronchitis	502.0 & 502.1
Bronchiectasis	526
ICD5 1940-49	
COPD	
Chronic bronchitis	106b
Bronchitis not distinguished as acute or chronic	106c
Asthma with influenza as contributory or secondary cause	112(1)
Asthma without complications specified in (1-5)*	112 (6)
Pulmonary emphysema	113

**influenza, chronic endocarditis, myocardial disease, arteriosclerosis, chronic nephritis*

Table A-4 Spearman rank correlation coefficients between explanatory variables used in the linear regression analyses in Chapter 7

	Deprivation quintile	SO_2 quintile	Minimum temperature quintile	Rain quintile	Pneumoconiosis SMR grouping	Asbestos-related disease SMR grouping
MALES						
Deprivation quintile	1.0000					
SO_2 quintile	0.2230	1.0000				
Minimum temperature quintile	–0.1446	–0.0663	1.0000			
Rain quintile	0.0713	–0.4220	0.1128	1.0000		
Pneumoconiosis SMR grouping	0.1067	0.1407	0.0562	–0.1782	1.0000	
Asbestos-related disease SMR grouping	0.2599	0.1601	0.0393	0.0195	0.1087	1.0000
Weight of fruit & vegetable purchases quintiles	–0.2531	0.0668	0.0563	–0.0987	0.0496	0.0436
FEMALES						
Deprivation quintile	1.0000					
SO_2 quintile	0.2388	1.0000				
Minimum temperature quintile	–0.1405	–0.0652	1.0000			
Rain quintile	0.0699	–0.4231	0.1095	1.0000		
Pneumoconiosis SMR grouping	0.1102	0.1345	0.0901	0.0008	1.0000	
Asbestos-related disease SMR grouping	0.1866	0.1483	0.0100	–0.0072	0.1647	1.0000
Weight of fruit & vegetable purchases quintiles	–0.2484	0.0668	0.0630	–0.1052	–0.1185	0.0133

Table A-5 Numbers of COPD deaths in England & Wales 1950-1996 for all ages by ICD code

Female COPD codes

ICD6 1950-57			ICD7 1958-67			ICD8 1968-78			ICD9 1979-1996		
codes	ndths	%	codes	ndths	%	codes	ndths	%	codes	ndths	%
241 (asthma)	10,554	13%	241 (asthma)	8,938	10%	490 (bronchitis)	6,070	7%	490 (bronchitis)	3,833	2%
501 (bronchitis)	9,260	12%	501 (bronchitis)	8,555	9%	491 (chronic bronchitis)	64,545	74%	491 (chronic bronchitis)	46,372	27%
502 (chronic bronchitis)	54,373	69%	502 (chronic bronchitis)	64,797	72%	492 (emphysema)	3,431	4%	492 (emphysema)	9,039	5%
526 (bronchiectasis)	4,229	5%	526 (bronchiectasis)	5,498	6%	493 (asthma)	7,770	9%	493 (asthma)	18,216	11%
			527 (emphysema)	2,315	3%	518 (bronchiectasis)	3,597	4%	494 (bronchiectasis)	6,487	4%
						519 (chronic airways obstruction)	1,291	1%	495 (extrinsic allergic alveolitis)	99	0%
									496 (chronic airways obstruction)	89,358	52%

Male COPD codes

ICD6 1950-57			ICD7 1958-67			ICD8 1968-78			ICD9 1979-1996		
codes	ndths	%	codes	ndths	%	codes	ndths	%	codes	ndths	%
241 (asthma)	10,874	7%	241 (asthma)	6,869	3%	490 (bronchitis)	5,865	2%	490 (bronchitis)	2,663	1%
501 (bronchitis)	9,484	6%	501 (bronchitis)	8,384	4%	491 (chronic bronchitis)	203,243	87%	491 (chronic bronchitis)	112,635	34%
502 (chronic bronchitis)	128,152	82%	502 (chronic bronchitis)	194,881	85%	492 (emphysema)	11,710	5%	492 (emphysema)	24,413	7%
526 (bronchiectasis)	8,715	6%	526 (bronchiectasis)	9,861	4%	493 (asthma)	5,477	2%	493 (asthma)	12,711	4%
			527 (emphysema)	9,586	4%	518 (bronchiectasis)	4,853	2%	494 (bronchiectasis)	6,442	2%
						519 (chronic airways obstruction)	3,660	2%	495 (extrinsic allergic alveolitis)	236	0%
									496 (chronic airways obstruction)	167,951	51%

Source: Analysis uses ONS data for England & Wales, 1950-1996

Figure A-1 Relative differences between age-standardised rates for lung cancer mortality for conurbations (excluding Greater London) and Greater London compared with non-conurbations, 1950-1999

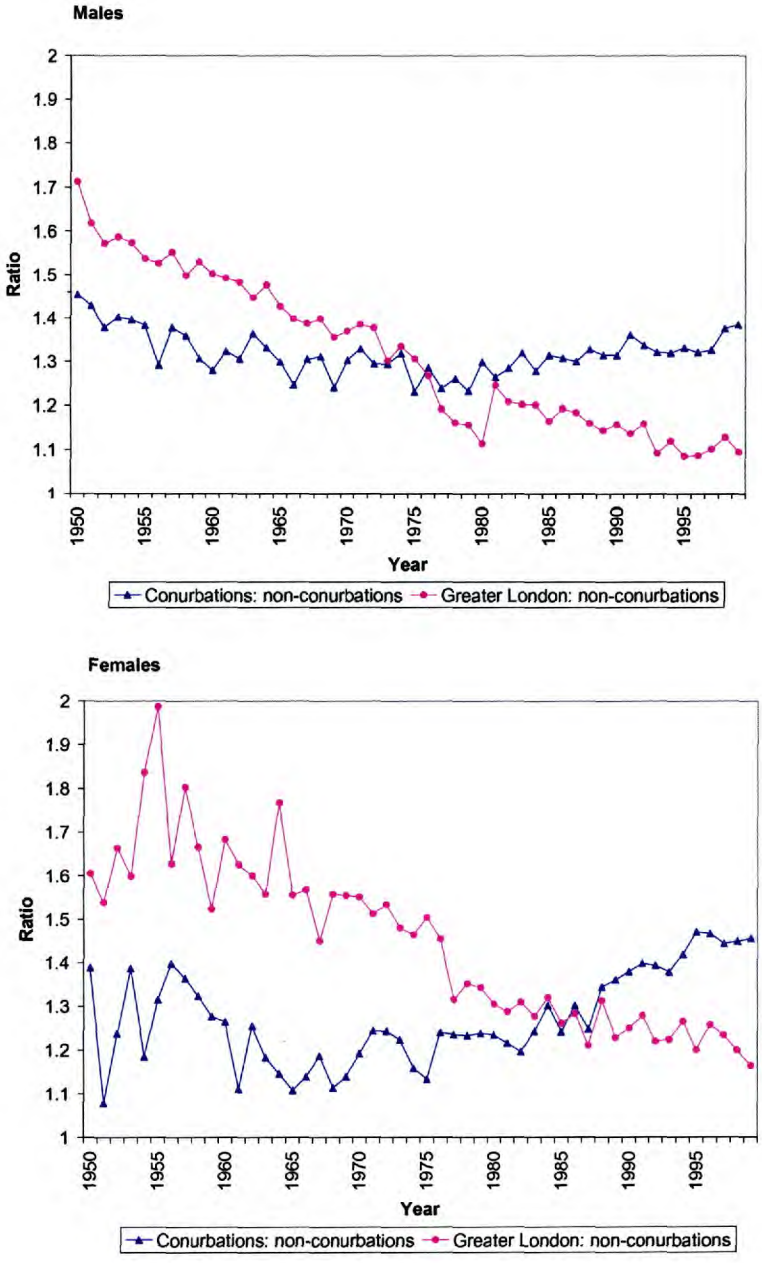


Figure A-2 Map of smoothed COPD relative risks (combined spatially structured and unstructured parameters from shared component analyses) for UK districts 1981-1999 – males

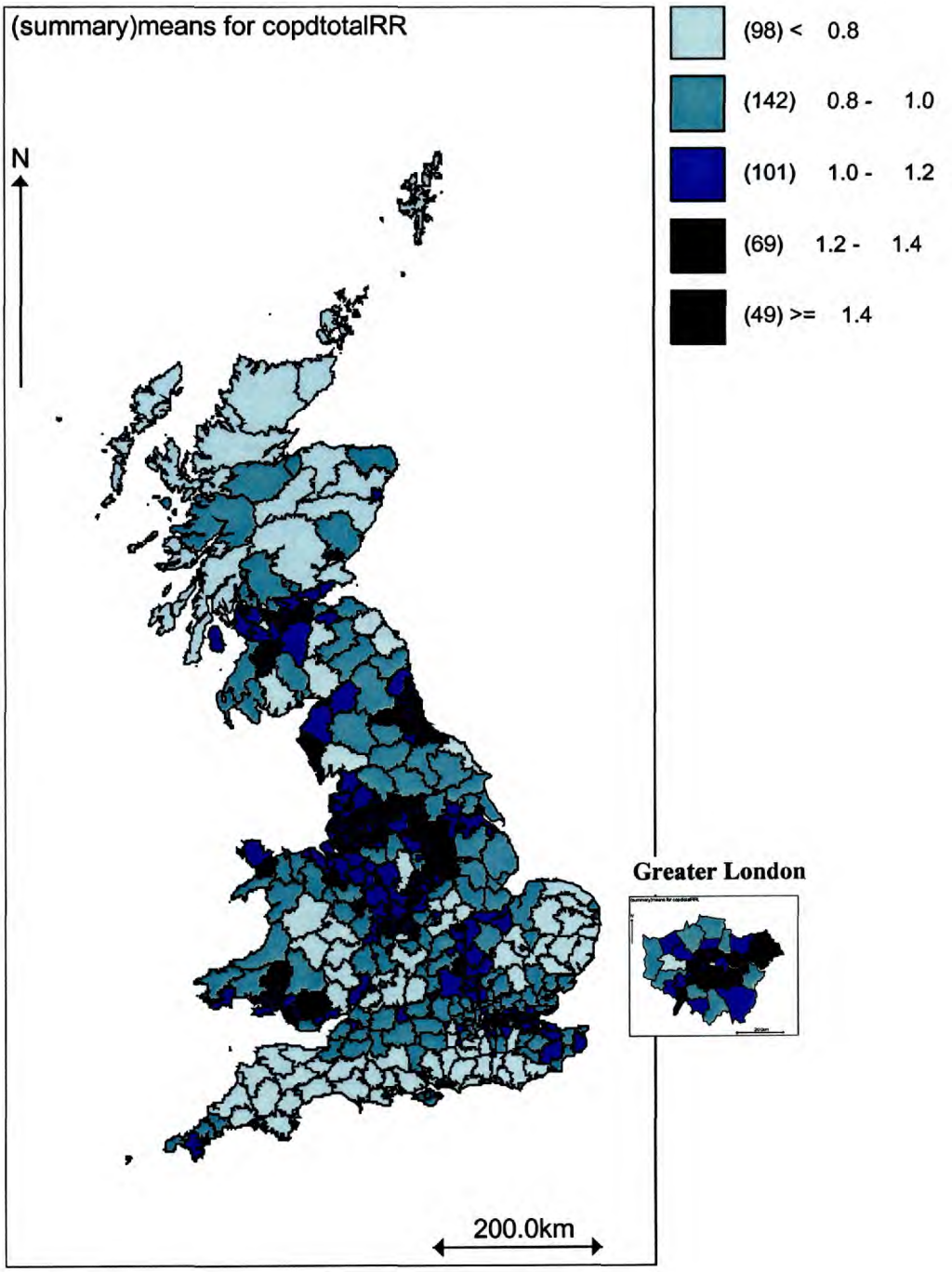


Figure A-3 Map of smoothed COPD relative risks (combined spatially structured and unstructured parameters from shared component analyses) for UK districts 1981-1999 – females

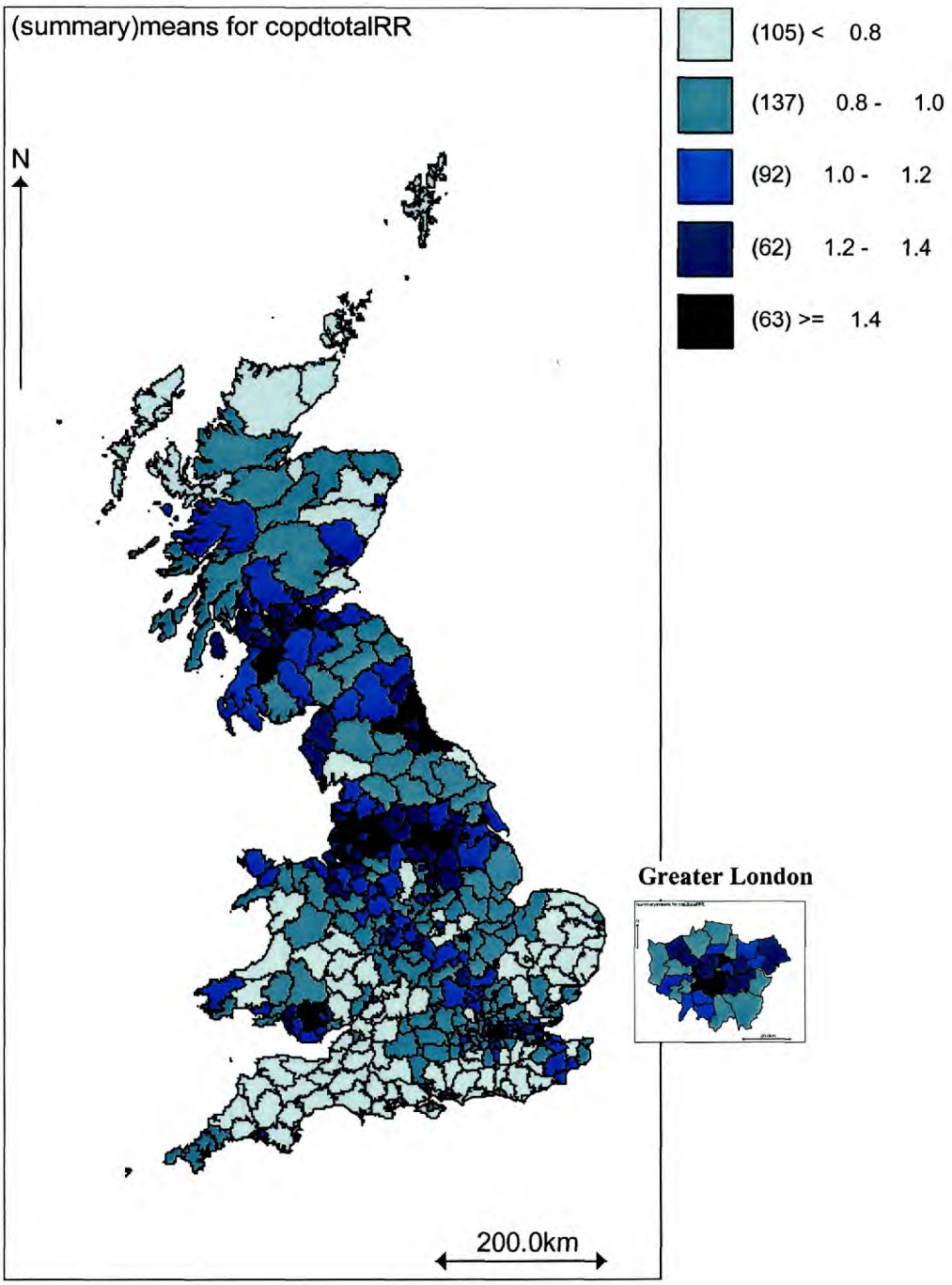


Figure A-4 Map of smoothed lung cancer relative risks (combined spatially structured and unstructured parameters from shared component analyses) for UK districts 1981-1999 – males

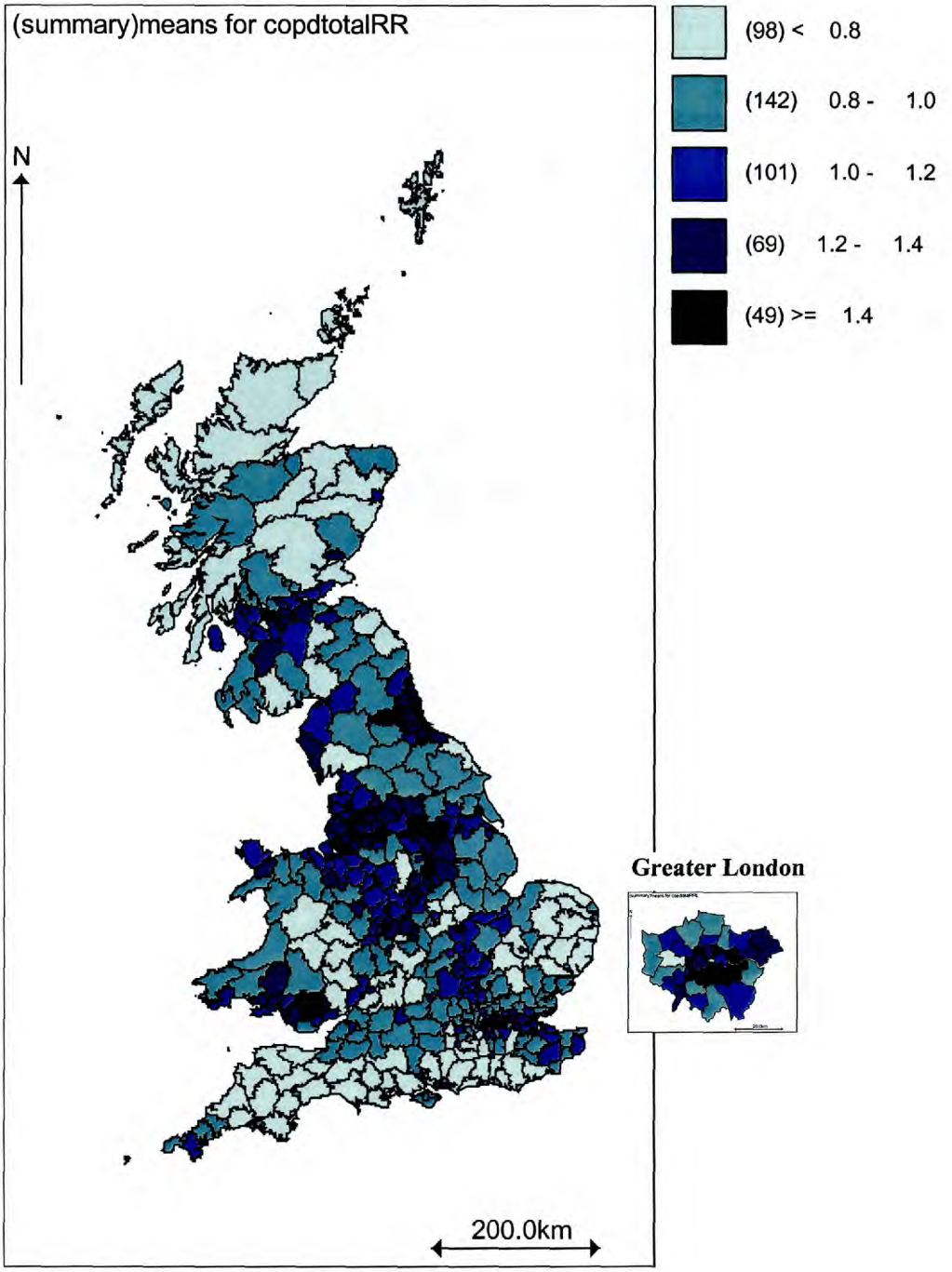


Figure A-5 Map of smoothed lung cancer relative risks (combined spatially structured and unstructured parameters from shared component analyses) for UK districts 1981-1999 – females

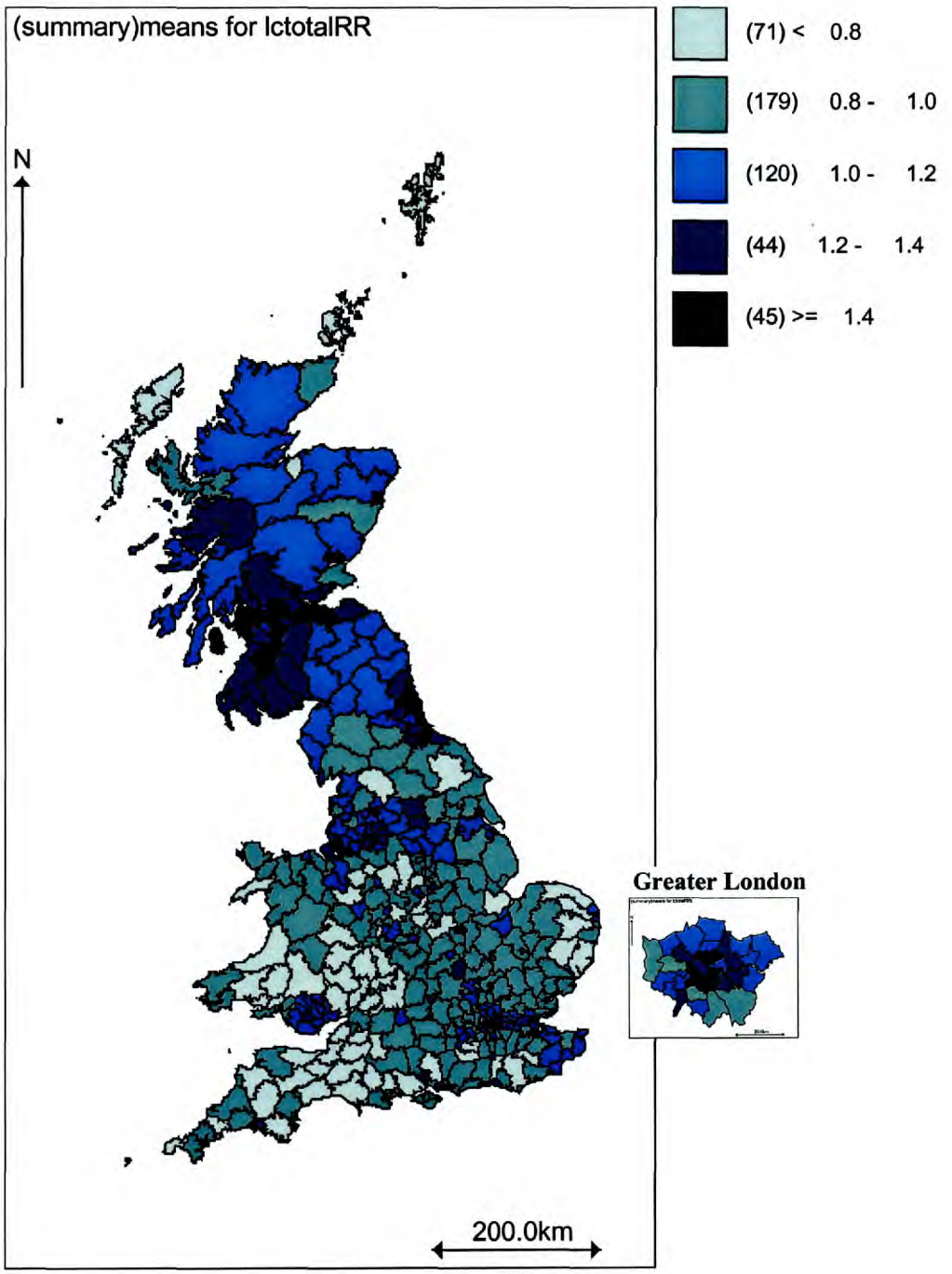


Figure A-6 Map of deprivation – percentage of district population in 1991 living in wards in the most deprived Carstairs quintile

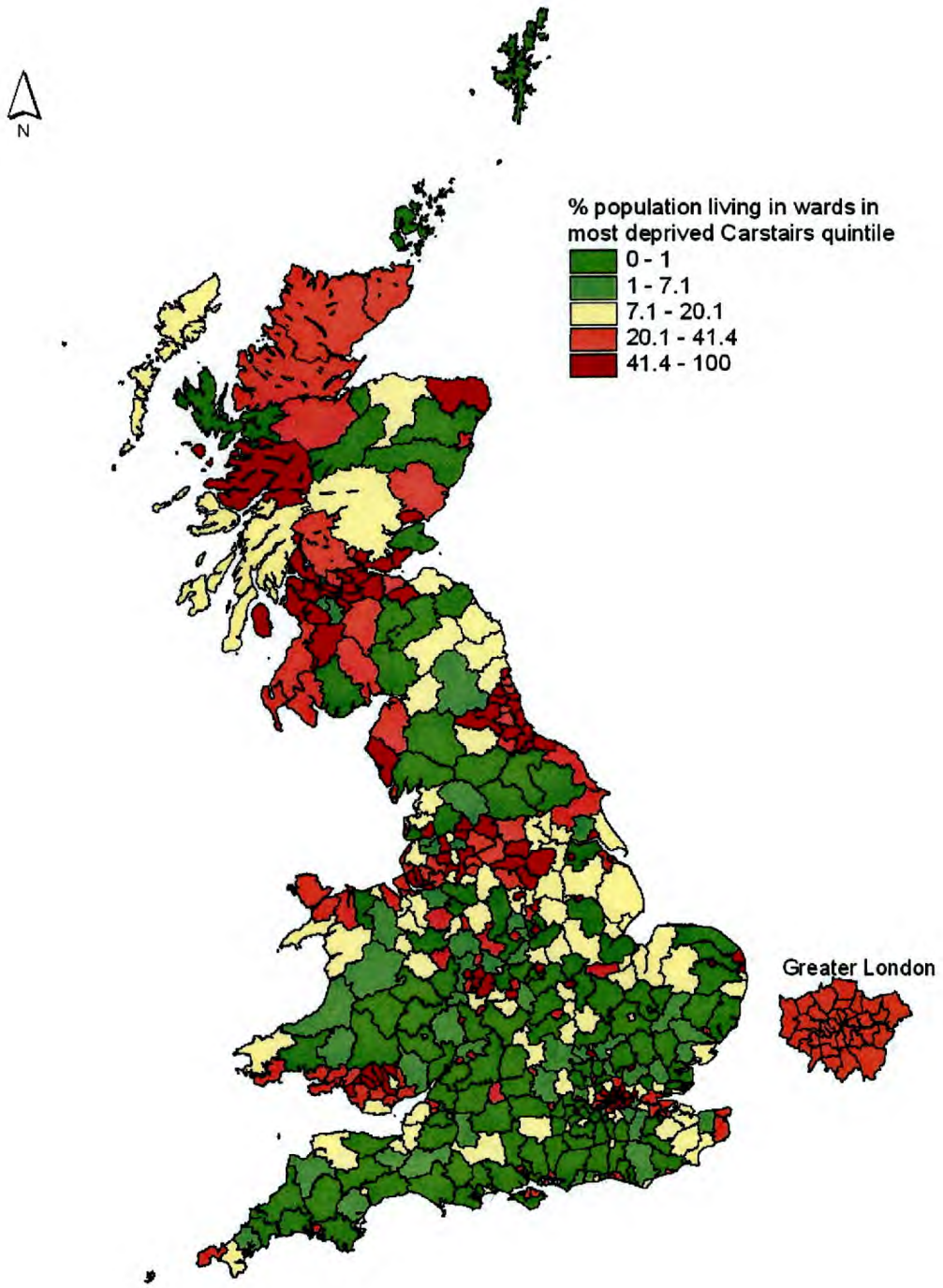


Figure A-7 Map of interpolated annual PM_{10} levels in UK districts in 1996

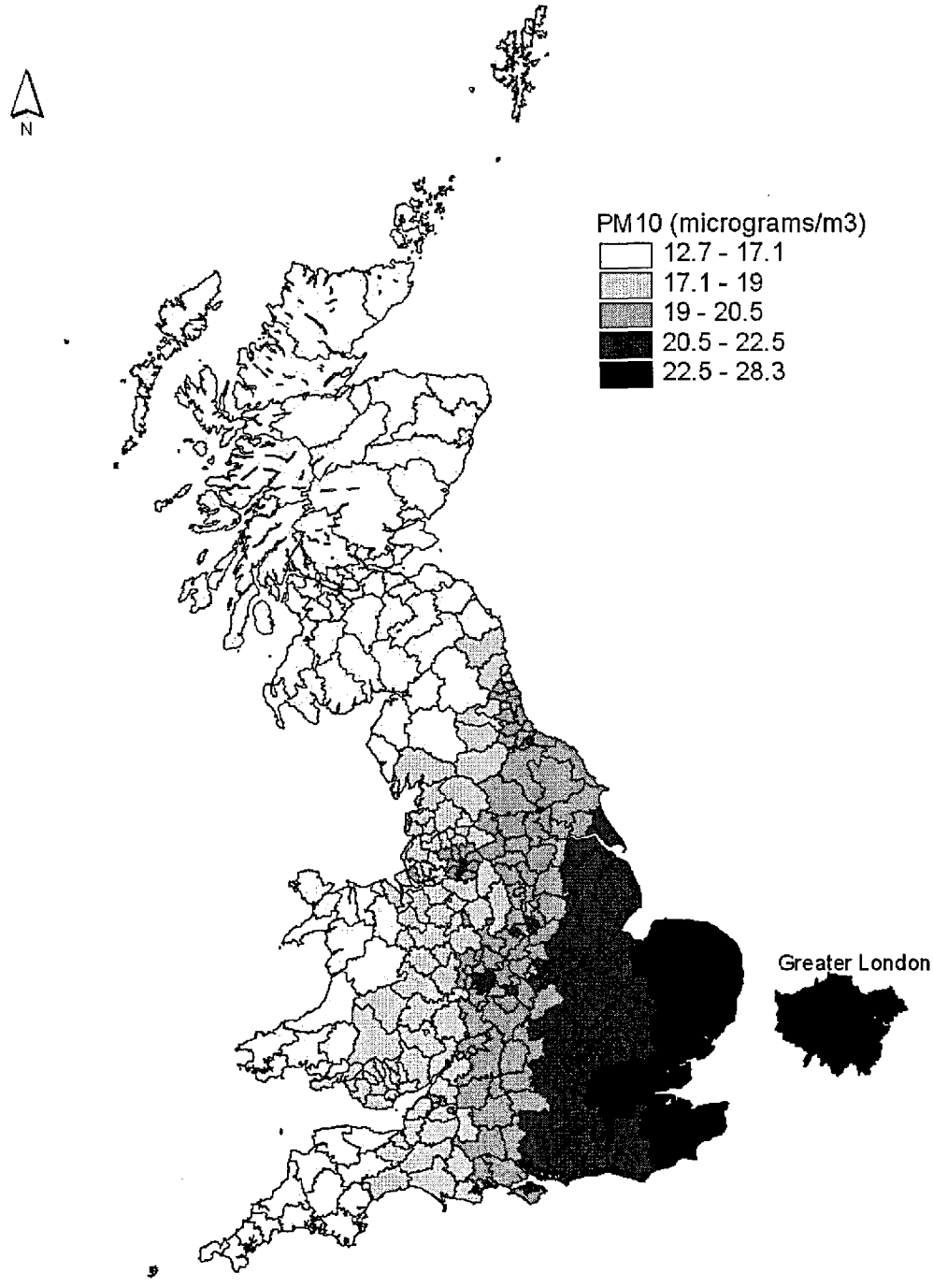


Figure A-8 Map of interpolated annual SO_2 levels in UK districts in 1996

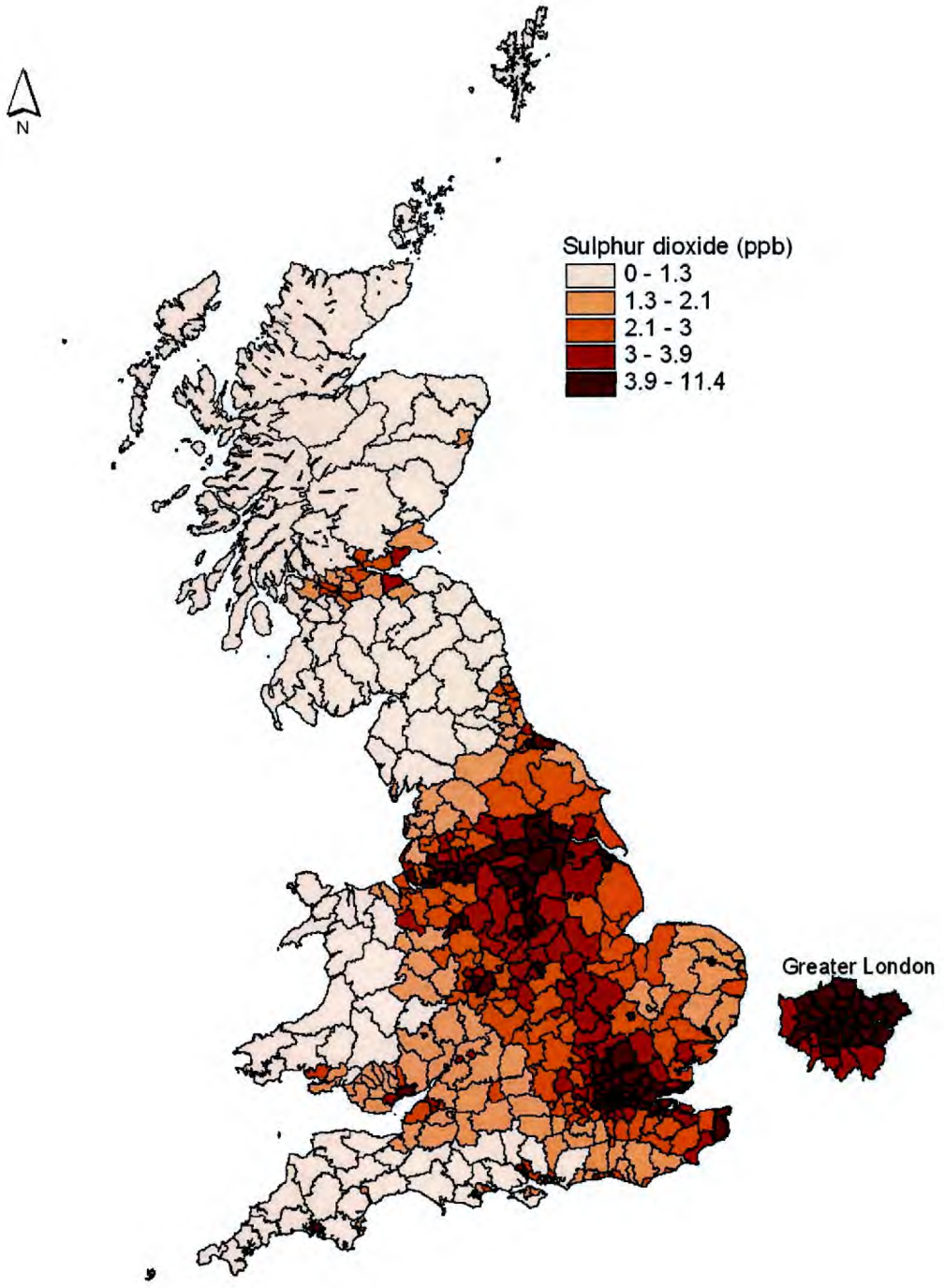


Figure A-9 Map of average daily minimum temperature (°C) in the UK 1981-1999

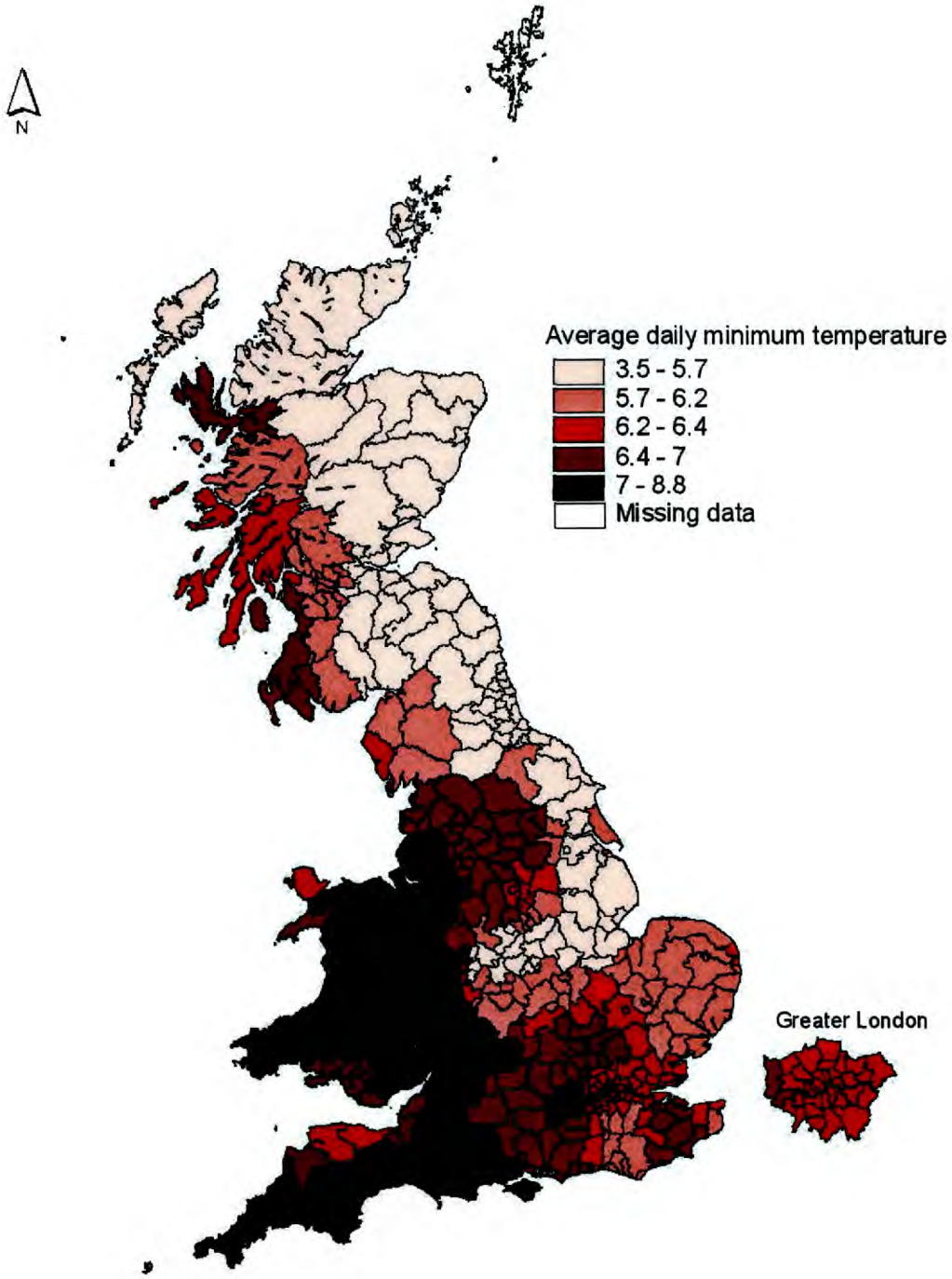


Figure A-10 Map of average daily rainfall (mm) in the UK 1981-1999

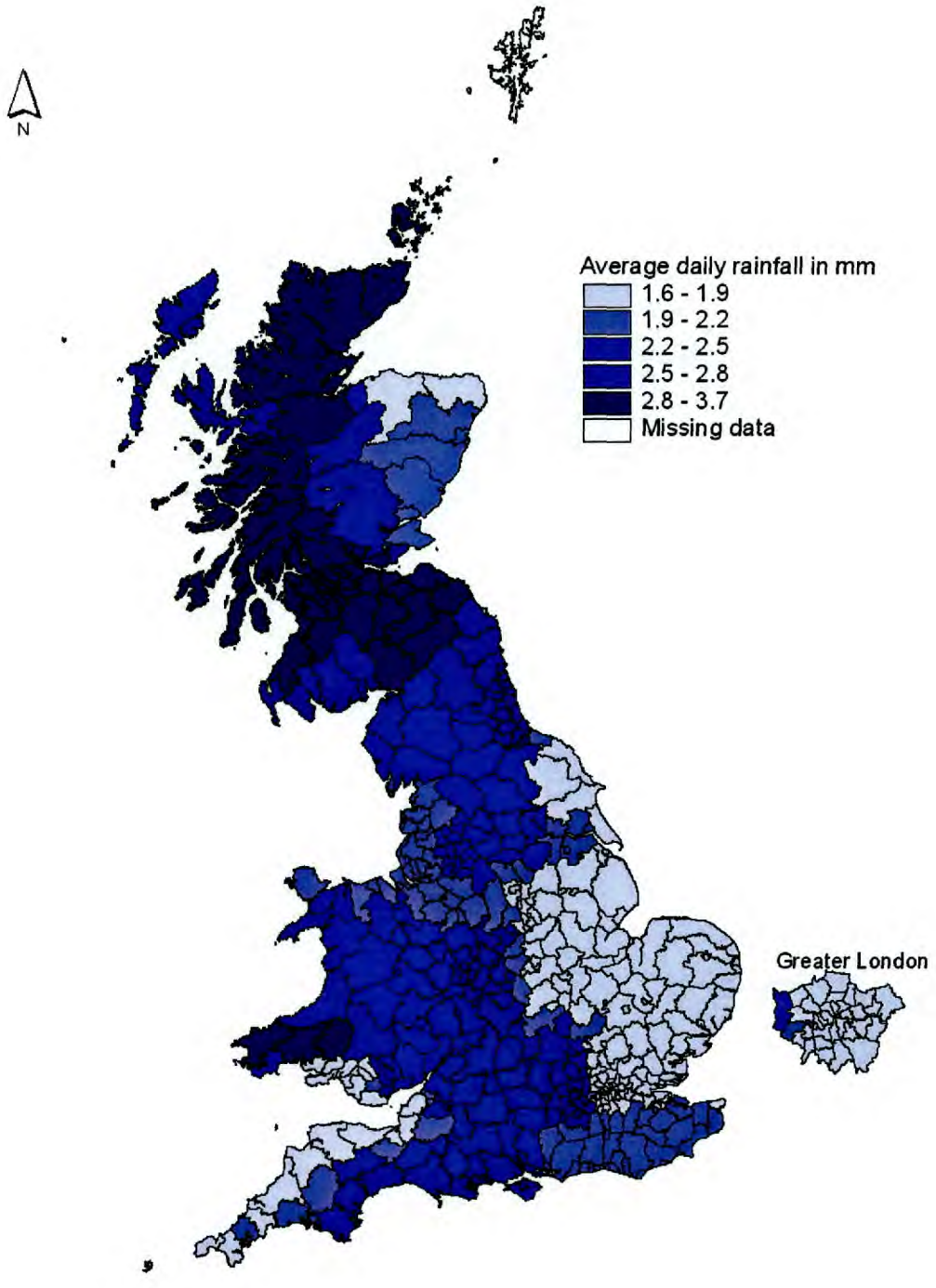


Figure A-11 Map of pneumoconiosis SMRs in UK districts 1981-1999 – males

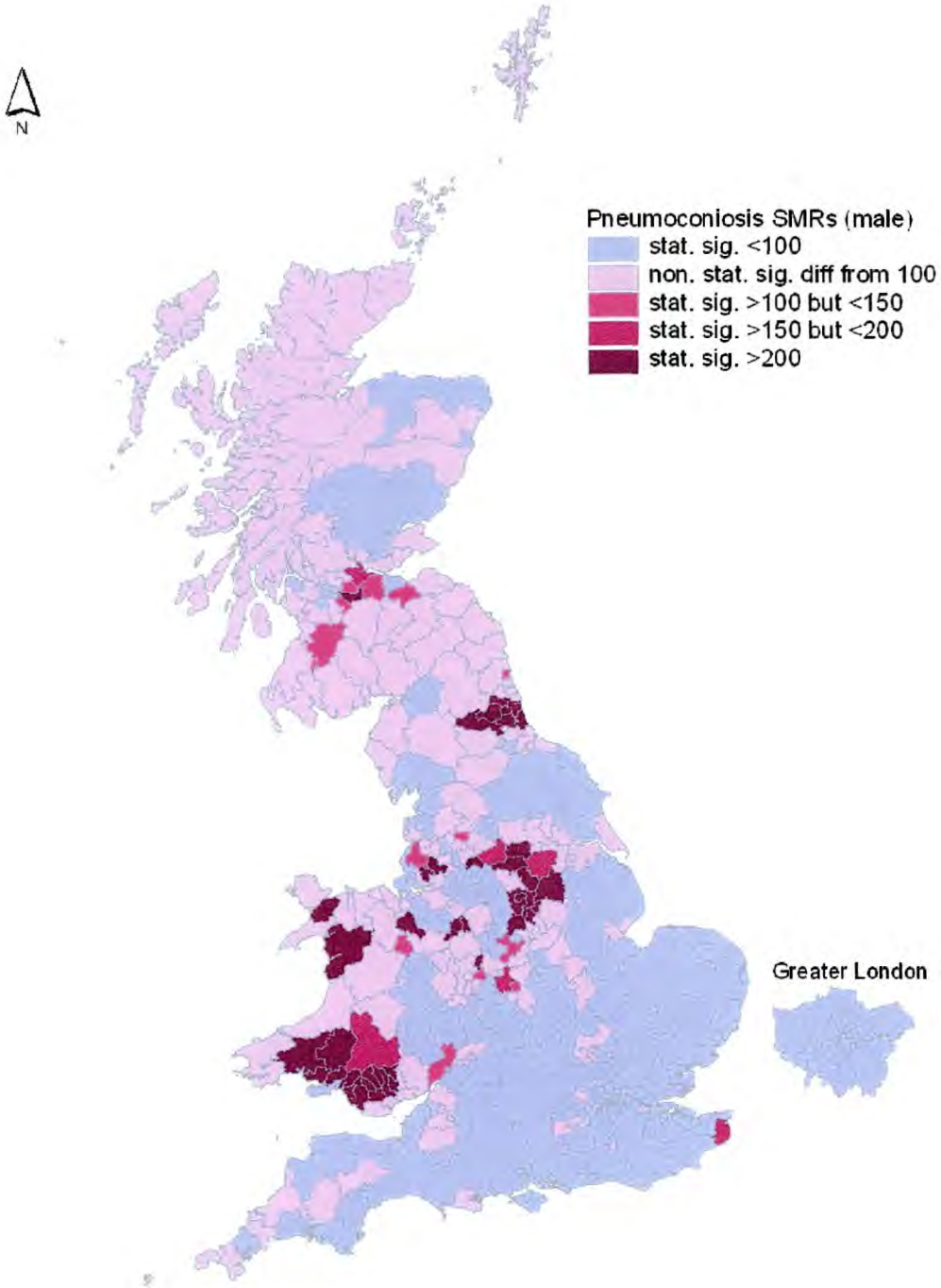


Figure A-12 Map of pneumoconiosis SMRs in UK districts 1981-1999 – female

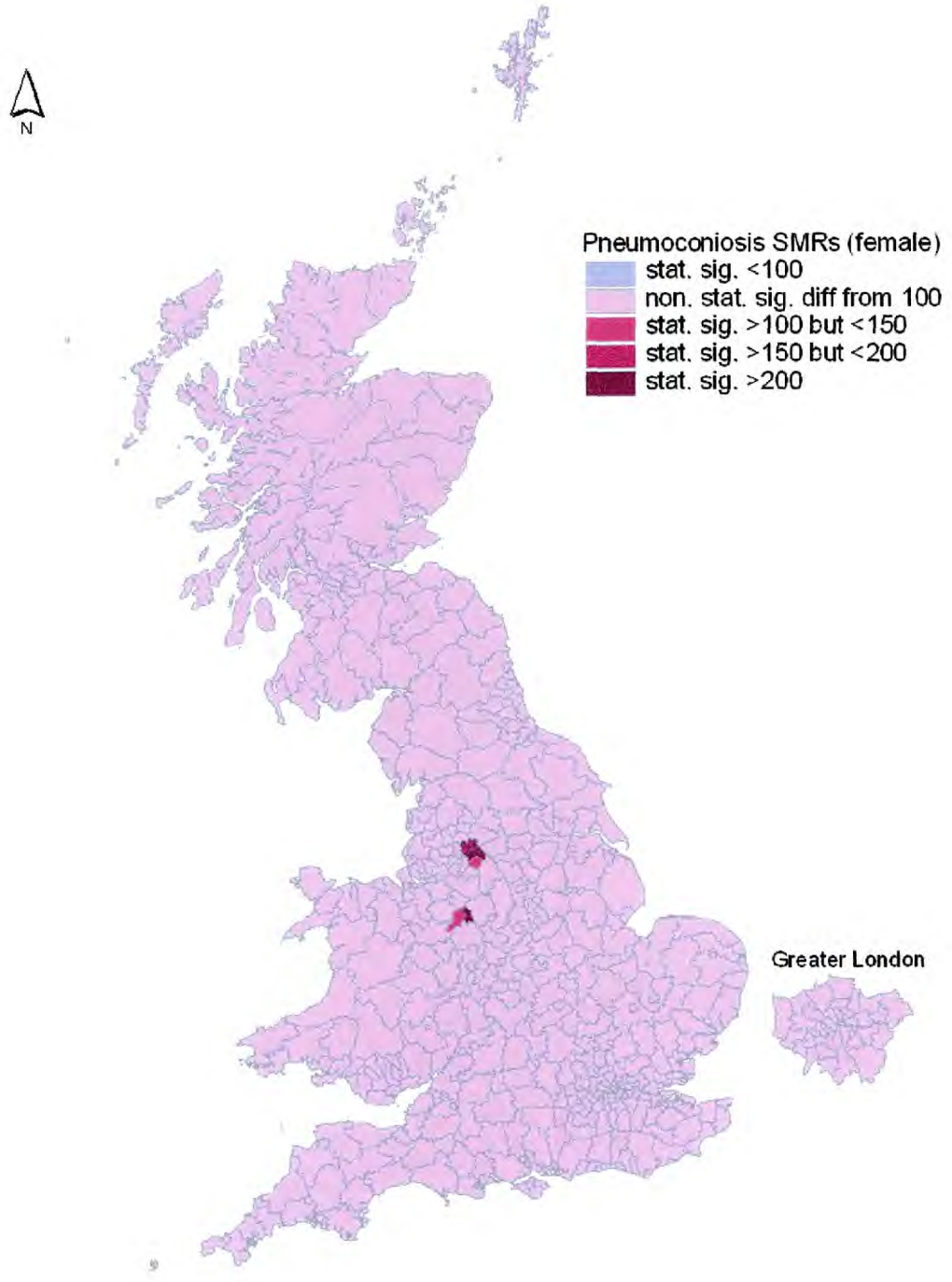


Figure A-13 Map of asbestos-related disease SMRs in UK districts 1981-1999 – males

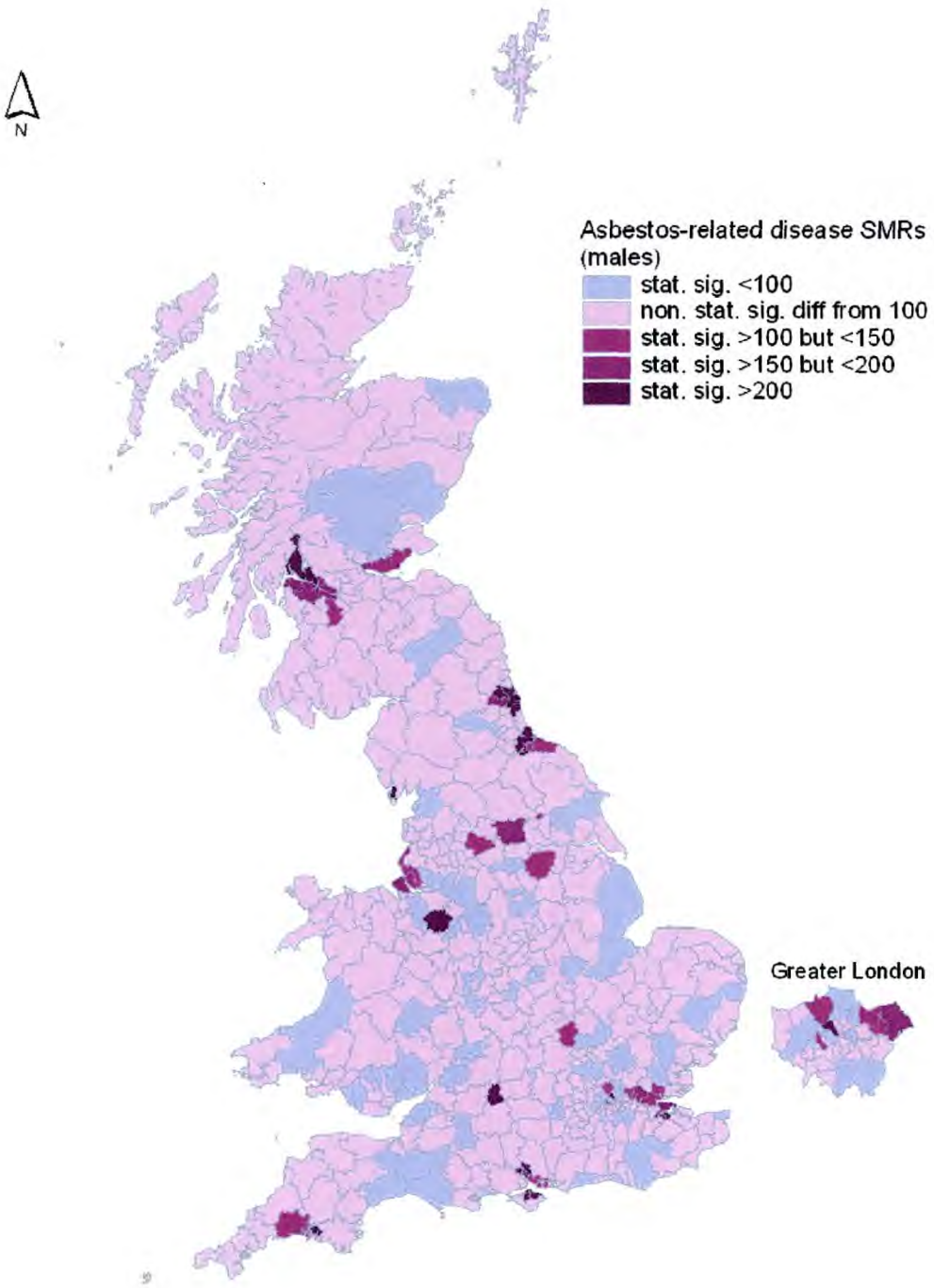


Figure A-14 Map of asbestos-related disease SMRs in UK districts 1981-1999 – female

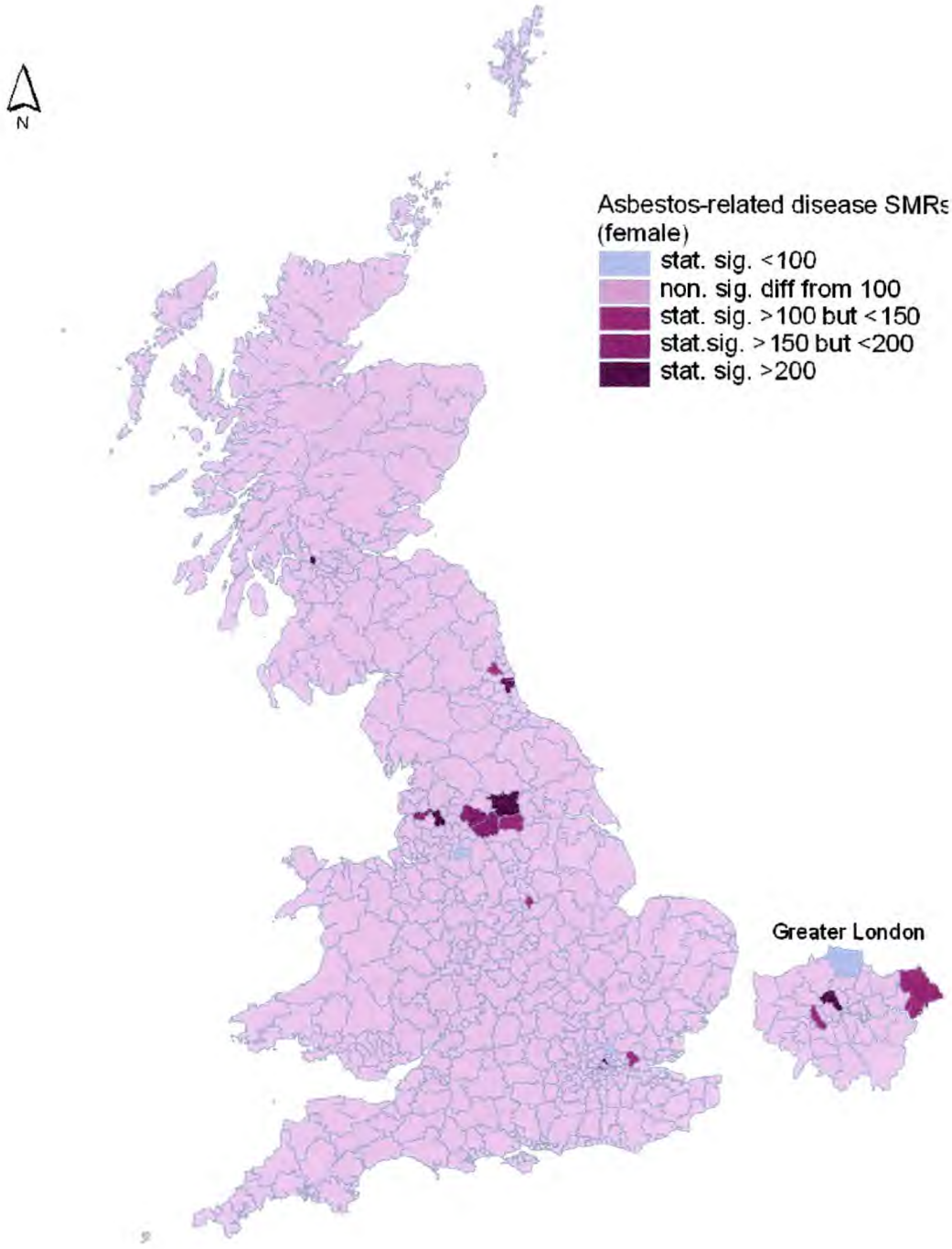


Figure A-15 Map of average fruit and vegetable purchases for UK districts 1991-2000

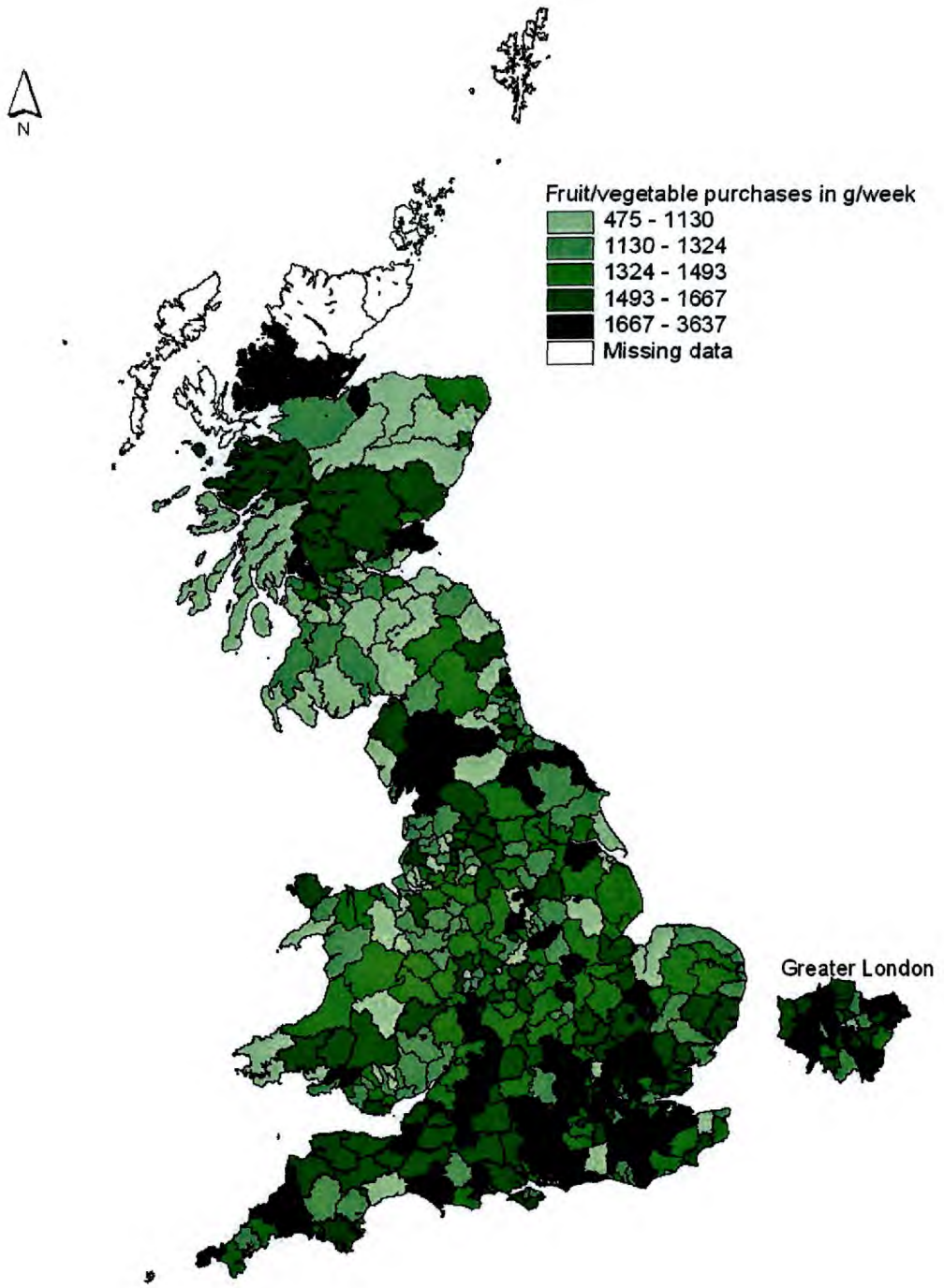
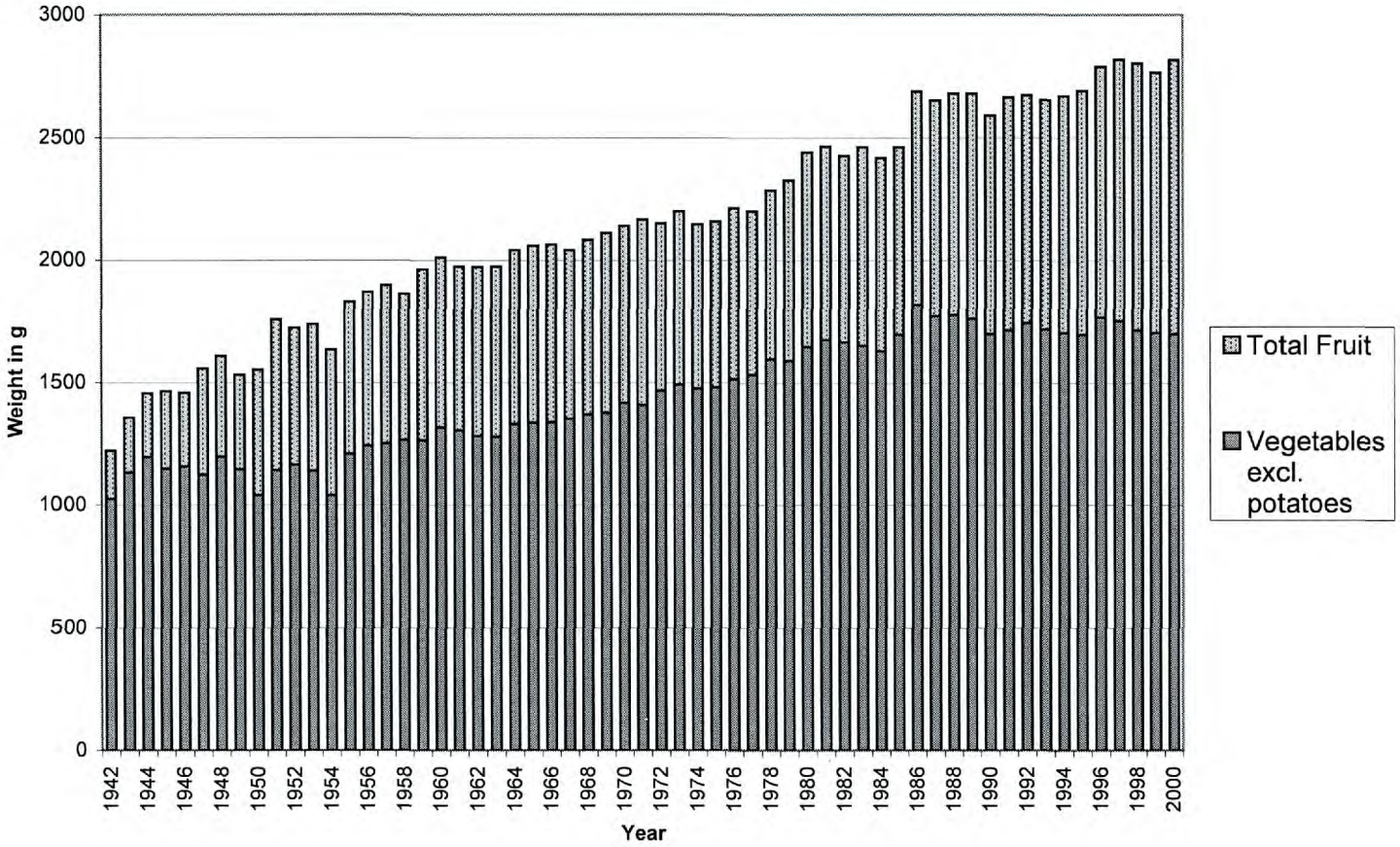


Figure A-16 Household consumption of fruit & vegetables (weight per week in g) from the National Food Surveys 1942-2000



Source of data: http://statistics.defra.gov.uk/esg/publications/nfs/default.asp
Note that definitions of fruit & vegetables differ from those used in comparisons of TNS and NFS data in Section I of this PhD, hence higher value