Eleni Sofianopoulou

Spatio-temporal Analysis: Respiratory Prescribing in Relation to Air Pollution and Deprivation, in Primary Health Care

Thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Institute for Health and Society

Institute for Research on Environment and Sustainability

For *Nikolas, Christina, Roi, Giorgos*

Abstract

Asthma and Chronic Obstructive Pulmonary Disease (COPD) are two of the most common chronic respiratory diseases causing a major burden of ill health to populations across the world. Respiratory medication prescribing can be used as an indicator of air pollution effect on asthma and COPD, capturing patients with any severity of disease from mild to severe. In contrast, the traditional indicators of asthma exacerbation, such as hospital admissions and emergency room visits, only capture events of patients who suffer severe symptoms.

In this study, I aimed to develop statistical models for assessing the spatio-temporal patterns of salbutamol prescribing in relation to air quality, in a primary health care setting. Salbutamol represents 93% of short-acting β2-agonists, which are prescribed for quick-relief of symptoms and acute exacerbations to individuals that suffer from asthma or COPD. I analysed salbutamol medication (approximately 67 billion Average-Daily-Quantities) prescribed by 64 GP practices in Newcastle and North Tyneside Primary Care Trusts, Northeast England, in 2002-2006. I used a mixed-effects model suitable for data that are not independent in time or space.

My study found ambient Particulate Matter (PM_{10}) concentrations to have a significant relationship to salbutamol prescribing in primary care. An increase of 10μg/m^3 in ambient PM_{10} concentrations was associated with an increase of 1% in salbutamol prescribing. Income deprivation and average age of patients registered per practice also had a significant relationship with salbutamol prescribing. The findings showed that the variation of salbutamol prescribing was subject not only to health needs caused by deprivation and air quality, but also random effects that were practice specific, such as facilities within the practice or experience and prescribing pattern of practitioners. Overall, the findings demonstrated that respiratory prescribing in primary care can be used as indicator of air pollution effect on asthma and COPD, increasing the scope of its use for health surveillance in the future.

Acknowledgment

It is a pleasure to thank those who contributed to the completion of this PhD. I greatly thank Prof. Tanja Pless-Mulloli and Prof. Stephen Rushton for supervising this study, offering me their advice and guidance. Tanja and Stephen are great academics and interesting individuals who I had the pleasure to work with. I need to convey a special thank you to Tanja, for her encouragement and support from the preliminary to the final stages of my work. I am grateful to the Colt Foundation for awarding me a PhD fellowship, making this PhD work possible. The Colt Foundation Student days have been a source of good advice and feedback on my work. The Director of Colt Foundation, Ms Jackie Douglas, responded always kindly and efficiently to my research needs.

I am thankful to my PhD assessors, Dr Peter Lurz and Prof. Nigel Unwin for giving me helpful advice. I appreciate the support Dr Roy Sanderson offered me in software installation and Dr Mark Shirley's support on database management. Sincere thanks to my fellows, Dr Payam Dadvand and Dr Andrew Close for discussing issues related to statistical analysis with me as well as Ms Lindsay Bramwell for proof-reading my thesis. I really appreciate the friendship I have been offered by team members that made the PhD life easier. I want to thank Mrs Gillian Paczynski, manager of the Institute of Health & Society, who has kindly supported the important administration aspect of the PhD. I appreciate Mrs Vicky Ryan's assistance to report the final results of my statistical model and Mr Fraser Chalmers' support on computer software/hardware installation.

I would also like to thank Dr David Chappel from North East Public Health Observatory (NEPHO) for his role on giving me approval to access the health data, as well as, NEPHO's management team, Ms Louise Unsworth, Prof Gyles Glover and Prof John Wilkinson for supporting the writing up phase of this work. I would like to acknowledge Ms Sue Brent from RDTC for providing me with the prescribing data and Mr Ian Abernethy from Gateshead city council for providing the traffic data.

I am also thankful to Prof. Douglas Bates and Prof. Peter Diggle, whose remarkable work in Statistics - books, publications and development of statistical software - helped me to learn new aspects of statistical analysis. There are several people who I contacted in the duration of this work who are not mentioned above, however I appreciate these several contacts and think of them as part of my PhD journey.

I would also like to thank the two external examiners of my PhD, Prof. David Coggon and Prof. Peter Diggle, for their time to examine my work, and for

providing me with constructive criticism and advice that helped me to improve my work.

My loving thanks to Nikolas for his encouragement to start this PhD and his support for completing it. I want to thank my parents for supporting me and teaching me to enjoy working towards gaining some education and knowledge. Finally, I thank my sister and brother for their loving support and encouragment.

Statement of Contribution: I declare that the work contained in this thesis is my own and has not been produced in collaboration with, or with the assistance of, any other person except in an advisory capacity.

Preface

Asthma and Chronic Obstructive Pulmonary Disease (COPD) are two of the most common chronic respiratory diseases. According to the latest WHO estimates, approximately 300 million people have asthma and 210 million people have COPD. It is recognised that COPD is an under-diagnosed disease and is expecting to be the third biggest cause of death during the coming decades (2008). In Europe, it is estimated that 30 million people are asthmatics, six million suffer symptoms which are characterised as severe, and 1.5 million people live in fear of dying from an attack.

Early epidemiologic studies have used data from secondary and tertiary care, such as admissions to hospital or emergency room visits, to quantify air pollution effect on respiratory exacerbations. However, those indicators capture patients that suffer from relatively severe symptoms. Primary care data offer a great source of information as the vast majority of patients suffering from respiratory diseases are fully treated at primary care level. Prescribing of respiratory medication in primary care, as an indicator of air pollution effect on asthma and COPD can capture patients with all levels of severity.

The majority of primary care consultations in the UK are by patients with respiratory disease compared to any other type of illness (British Thoracic Society, 2006, Pinnock and Sheikh, 2009). This figure emphasizes the important role of primary care in managing respiratory diseases. The World Health Assembly set out an action plan in 2008, to prevent and control chronic non-communicable diseases, including asthma and COPD (World Health Organization, 2008a). Part of this action plan is to strengthen the management of the diseases at primary care level. The WHO report also suggested that accessibility to medication in primary care can be used as indicator to monitor progress (World Health Organization, 2008b). One of the main long-term priorities for National Health Service (NHS) reform in UK, is the shift from provision of hospital-based acute care to proactive care, delivered in primary care (Department of Health, 2006, Department of Health, 2005). In the case of asthma, 75% of hospital admissions are avoidable (Asthma UK, 2010), no equivalent information was available for COPD. The importance of exploiting data at primary care level and especially data of respiratory diseases is discussed below.

Aims and Objectives

Statistical models should account sufficiently for temporal and spatial variation of the variables that are not distributed independently in space and time. A key weakness of traditional statistical techniques is that they have to assume independence of health events. Complex spatio-temporal associations of health hazards, exposures and outcomes may therefore be misrepresented or result in associations being biased.

My study aimed to develop an adequate statistical model at the level of primary care to assess the relationship between asthma/COPD medication prescribing and air quality. I tested the hypothesis that exposure to air pollution increases the frequency and duration of asthma and COPD symptoms, generating a consequent increase in the use of salbutamol medication and consequently an increase of prescriptions. Salbutamol represents 93% of short acting β2-agonists (Prescription Pricing Authority, 2006), which are prescribed for quick-relief of symptoms and acute exacerbations to individuals that suffer from asthma or COPD. I evaluated possible time-lags in response to medication use and air quality. I also considered contextual factors of the local environment, such as income, employment and educational deprivation and demographic factors (age and gender of patients) as covariates.

To the best of my knowledge, this study is the first in the UK to assess the effect of air pollution on prescribing of respiratory medication, at GP practice level. In order to achieve the aim of this study, the following objectives had to be met:

- To collate data on potential explanatory variables of the association between air pollution exposure and respiratory prescribing.
- To explore the potential relevance of covariates, as well as understand the structure of the data, in order to build an adequate statistical model.
- To quantify and evaluate the relationship between salbutamol prescribing, air quality indicators and the other covariates.
- To examine delayed responses of salbutamol prescribing to the trigger factors of asthma/COPD exacerbations.

A Brief Outline of the Thesis

The thesis comprises five chapters. The **first chapter** outlines the scientific background and consists of five sections. I initially present an overview of the literature on asthma and COPD prevalence as well as their main risk factors. I then focus on the pharmacological interventions for asthma and COPD, and present a comprehensive review of studies that used respiratory prescribing as indicator of air pollution effect on respiratory diseases. The fourth section discusses the role of primary care on respiratory diseases and its application in environmental epidemiology. In the last section I present the literature on the main aspects of spatial and spatio-temporal analysis, as they apply to this study.

The **second chapter,** consisting of four sections, describes the data collection, collation and exploration process. In the first section I present the prescribing data and GP registered population that I used to create GP service areas. In the second section I give details on the indicators of air quality. The third part describes the deprivation data, and the final section reports on the demographic data.

Chapter three presents the statistical analysis and consists of four sections. The first section is dedicated to describing the methods of statistical analysis. The second section presents the preliminary stages of the statistical model development. The third section describes the modelling of salbutamol prescribing seasonal variation, while in the fourth section I present the final model and its findings. The main discussion points are presented in **chapter four** while the conclusions as well as main limitations and strengths of the thesis are presented in **chapter five**.

Contents

List of Figures

List of Tables

Abbreviations

ADMS	Atmospheric Dispersion Modelling System
AIRE	Asthma Insights & Reality in Europe
ACF	Autocorrelation Function
ATS	American Thoracic Society
BS	Black Smoke
BTS	British Thoracic Society
CDC	Centre of Disease Control and Prevention
CI	Confidence Intervals
CO	Carbon Monoxide
COPD	Chronic Obstructive Pulmonary Disease
DDD	Daily Defined Doses
DEFRA	Department for Environment Food and Rural Affairs
EC	European Community
ECRHS	European Community Respiratory Health Survey
EFA	European Federation of Allergy and Airways Diseases Patients Associations
ERS	European Respiratory Society
GAM	Generalized Additive Model
GARD	Global Alliance against Chronic Respiratory Diseases
GINA	Global Initiative for Asthma
GIS	Geographic Information System
GOLD	Global Initiative for Chronic Obstructive Lung Disease
GP	General Practitioner
IMD	Index of Multiple Deprivation
INSPIRE	Infrastructure for Spatial Information in Europe
ISAAC	The International Study of Asthma and Allergies in Childhood
LSOA	Lower Super Output Areas
MCMC	Markov Chain Monte Carlo
MCP	Minimum Convex Polygon
NEPHO	North East Public Health Observatory
NHS	National Health Service
NO_2	Nitrogen dioxide
O_3	Ozone
P-value	Probability value
PM_{10}	Particulate Matter with aerodynamic diameter less than 10μm
PCT	Primary Care Trust
RDTC	Regional Drugs and Therapeutics Centre
SABA	Short-acting β2 agonists
s.d.	Standard Deviation
SES	Socio-economic Status
SO_2	Sulphur Dioxide
SOA	Super Output Areas
TSP	Total Suspended Particles
US	United States
WHO	World Health Organisation

Chapter 1. Scientific Background

1.1 Chronic Respiratory Diseases – Asthma and Chronic Obstructive Pulmonary Disease (COPD)

Hundreds of millions of people suffer every day from chronic respiratory diseases, which are chronic diseases of the airways and other structures of the lung. The two most common respiratory diseases are asthma and Chronic Obstructive Pulmonary Disease (COPD), also known as obstructive lung diseases. Asthma is a disease in which inflammation of the airways causes airflow into and out of the lungs to be restricted (Figure 1-1). COPD is a lung ailment the main components of which are chronic bronchitis and emphysema (Global Initiative for Chronic Obstructive Lung Disease (GOLD), 2008). Chronic bronchitis and emphysema used to be considered separate conditions but the terms are now replaced by COPD (Global Initiative for Chronic Obstructive Lung Disease (GOLD), 2008). In people with chronic bronchitis, the airways are narrowed, tight, swollen and are often filled mucus, resulting in reduced airflow. In emphysema, air sacs (alveoli) in the lungs are damaged and overstretched, resulting in air being trapped in the lungs, limiting the space for air exchange (Figure 1-2).

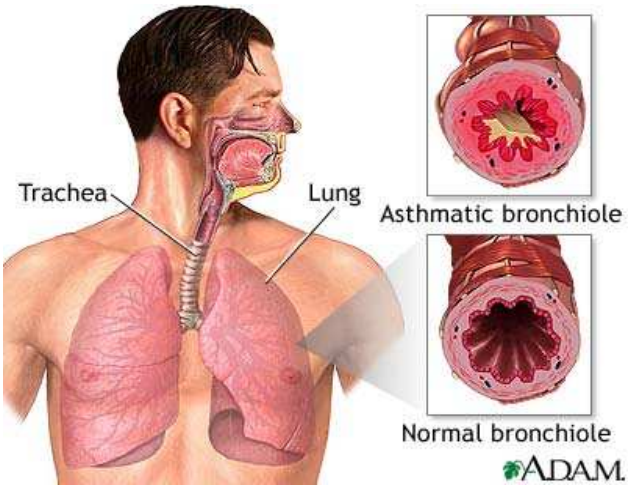


Figure 1-1. Airway (bronchiole) with inflammation (MedlinePlus - A.D.A.M. Inc, 2006).

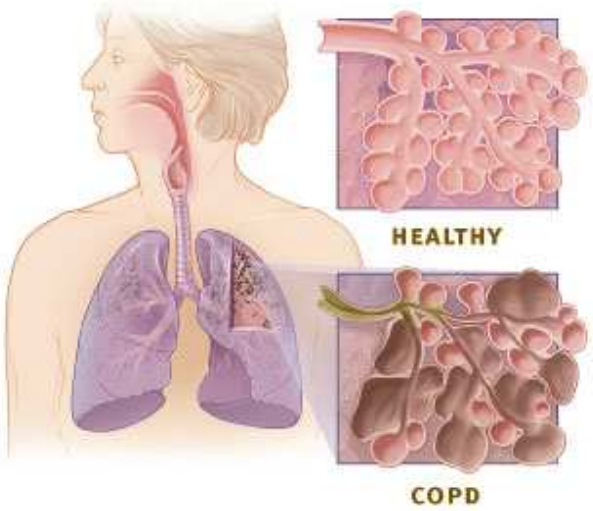


Figure 1-2. Alveoli damaged by COPD.

1.1.1 Asthma Prevalence

According to the latest WHO estimates, currently 300 million people have asthma (World Health Organization, 2009a). In 1992, the World Health organisation (WHO) formed the Global Initiative for Asthma (GINA) to promote the investigation of the disease. In Europe it is estimated that 30 million people are asthmatics, six million suffer symptoms which are characterised as severe, while 1.5 million people live in fear of dying from an attack (European Federation of Allergy and Airways Diseases Patients Associations, 2003). The United Kingdom (UK) is one of the countries with the highest asthma prevalence in the world (Woodruff and Fahy, 2001, Masoli M. et al., 2004c) along with USA, Canada, Australia, New Zealand, Brazil, Peru and South Africa (Figure 1-3).

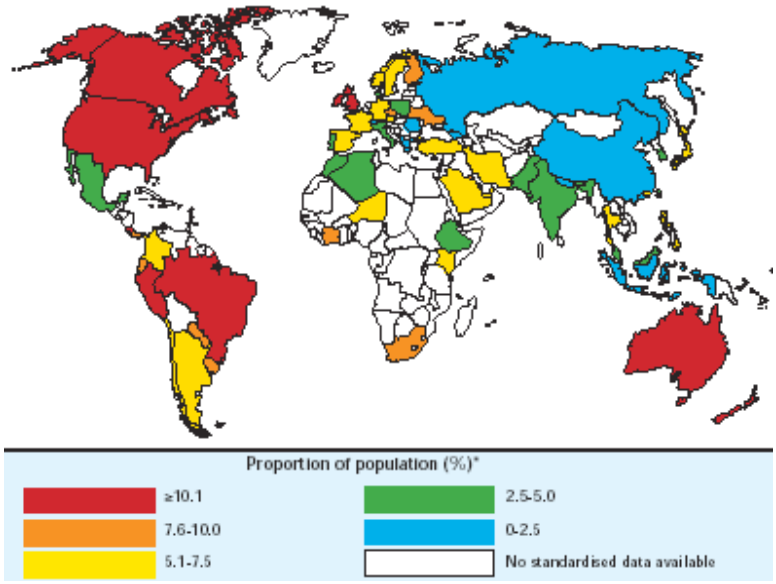


Figure 1-3. World map of The Prevalence of Asthma, Source: Global Initiative for Asthma (GINA) (2004) (Masoli M. et al., 2004b).

Figure 1-4 shows fatality rates of asthma; the highest mortality rates occurred in countries with some of the lowest prevalence of the disease, such as Russia, China and Mexico. In contrast, countries with a very high prevalence of asthma often have the lowest rates of fatality, including the UK, Canada, Finland, and Australia (Masoli M. et al., 2004c). Asthma fatality rate corresponds to the number of deaths caused by asthma to the number of diagnosed cases of asthma. The pattern observed indicate that the countries with low asthma fatality rate are better on containment of mild cases.

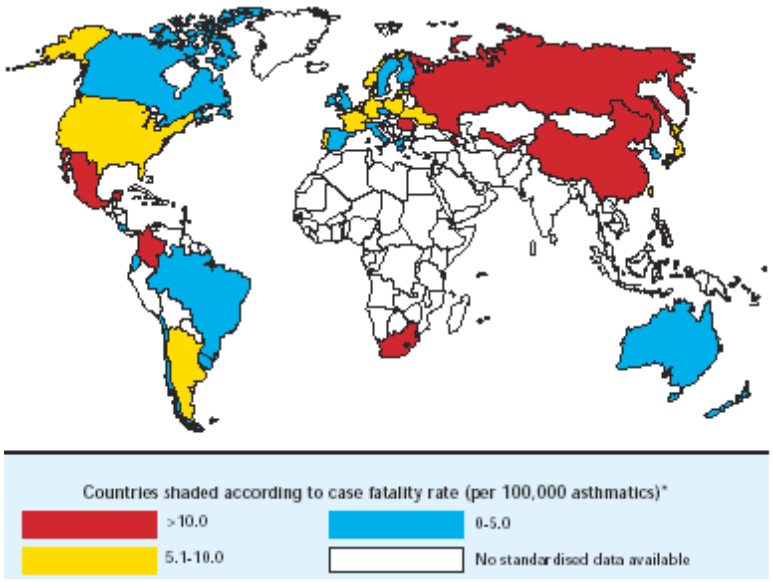


Figure 1-4. World map of the asthma case fatality rates, Source: Global Initiative for Asthma (GINA) (2004) (Masoli M. et al., 2004a).

For England, data from the Quality and Outcomes Framework (QOF) showed that asthma prevalence was 5.9% (NHS The Information Centre for Health and Social Care, 2010). Asthma UK estimated that there are approximately 5.4 million people in UK (8.9% of the total UK population) with asthma (Prescription Pricing Authority, 2006, Asthma UK, 2004); (Prescription Pricing Authority, 2007). Comparing asthma prevalence based on Asthma UK and QOF data, the latter appears to be an underestimate, however the figures are not directly comparable as QOF does not include the populations of Scotland, Wales and Northern Ireland.

I compared data from the 4th Morbidity Study and Asthma UK in order to review the trend of asthma prevalence in UK over the most recent decade. According to the 4th Morbidity Study (Calverley and Sondhi, 1998) the diagnosed prevalence of asthma in UK was 4% (approximately 2.5 million people) at the end of 1990s. Comparing this figure to current Asthma UK estimate of approx. 8.9%, would show an increase of 4.9%. I also compared the estimate of 5.4% of UK's population from Asthma UK in 1998 (called National Asthma Campaign) to the current Asthma UK estimate of 8.9% that showed a 3.5% increase in asthma prevalence, during the last decade. The two numbers (4.9% and 3.5%) differ by approximately 1.5%. These indicative numbers of asthma prevalence increase in England during the last decade, are lower compared to asthma prevalence increase rates observed in earlier decades.

Experts were finding it difficult to understand why rates, world-wide, were rising by 50% on average during the 1970s and 1980s (World Health Organization, 2000). Asthma prevalence was increasing worldwide at such a rate that asthma was classified as epidemic. Figure 1-5 shows the significant increase of asthma medication for England, between 1980 and 1993. However, for the last 10-15 years there are conflicting views on time trends of asthma, as recent trends show that the asthma epidemic has slowed down and prevalence may have declined. There are some studies indicating that asthma prevalence may be stabilising in high & middle income countries (Bollag et al., 2005, Fleming et al., 2000). Bollag *et al*., (2005) argued that asthma consultations in Switzerland have declined and suggest further research to be conducted in other cities and regions, in order to assess if the asthma epidemic has started to stabilise or even decline. In their recent work, Bollag *et al* (2009) also observed a fall in the proportion of asthma patients who had hay fever, and have argued that the asthma prevalence in Switzerland has been reduced because allergic induced asthma has declined.

Those results were consistent with findings of an English study in children (Anderson et al., 2004). Anderson et al (2004) also found no increase in prevalence of asthma symptoms in children 12-14 years old, in British Isles. The study on the British Isles reported decreases of eczema and hay fever as well. This would be similar with another study in Switzerland where prevalence rates of asthma and current asthmatic symptoms remained constant, and at the same time no further increase was observed for hay fever rates and allergic sensitisation rates (Braun-Fahrlander et al., 2004). However, a study in Australia found a fall in asthma prevalence but an increase in prevalence of eczema and hay fever (Robertson et al., 2004).

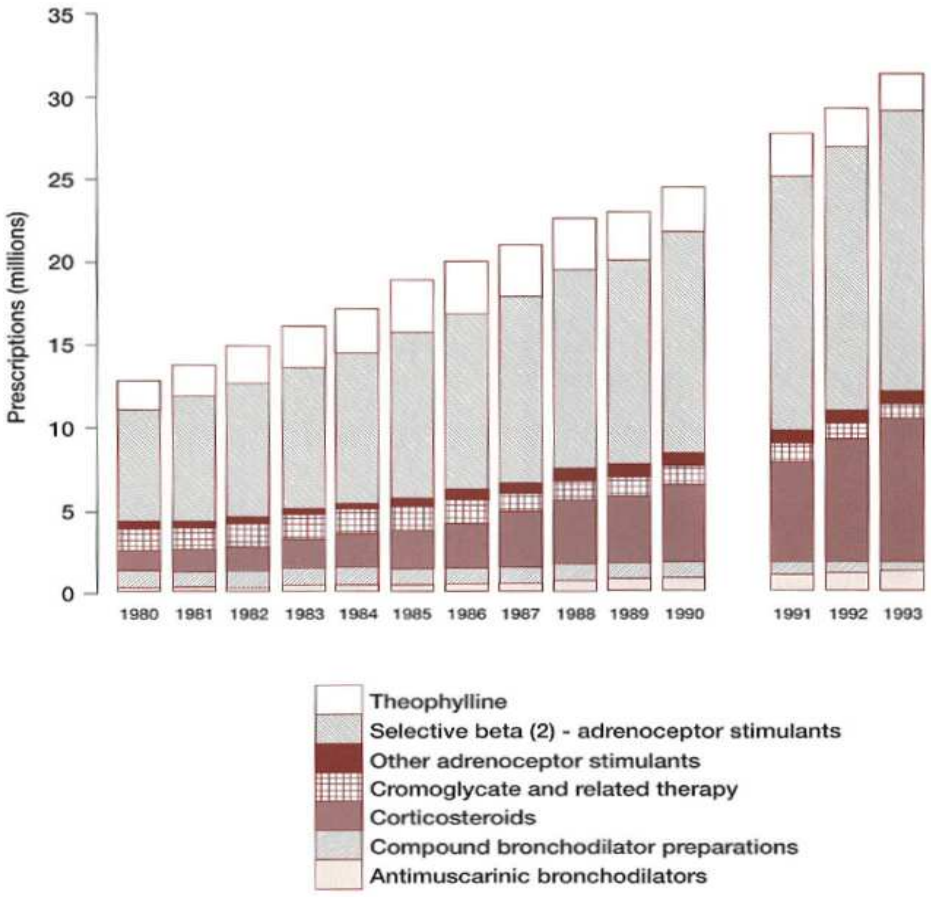


Figure 1-5. Number of prescriptions for asthma preparations in England between 1980-1993, Source: (Department of Health, 1995).

The reported findings on the stabilisation or even decrease of the asthma prevalence over the last decade are to some extent reassuring but "should not be taken to indicate that the global pandemic of asthma is easing and that the worst is over" (Pearce and Douwes, 2005) page 763. The International Study of Asthma and Allergy in Childhood (ISAAC) phase III showed that the increase of asthma prevalence in Latin American countries may lead to describing asthma as a "Spanish and Portuguese speaking rather than an English speaking disease" (Asher et al., 2006).

ISAAC is one of the two major multicentre studies that are able to provide evidence on the trends of asthma prevalence. The second one is the European Community Respiratory Health Survey (ECHRS) that focuses on centres of high income

countries, located mainly in Western Europe. I will now describe these two studies in more detail.

1.1.1.1 International Study of Asthma and Allergy in Childhood

ISAAC phase I has shown an approximately thirty-fold variation of asthma prevalence between countries, from data collected by June 1996 (ISAAC Steering Committee, 1998b). The survey was completed in 156 collaborating centres in 56 countries with a total of over 700,000 children, aged 6-7 and 13-14 years, participating. The highest 12-month prevalence for asthma symptoms were reported from centres in UK, New Zealand, Australia, and Republic of Ireland (Figure 1-6).

The time trend of the asthma prevalence was assessed by ISAAC survey phase III, carried out between 2001 and 2002 (Asher MI et al., 2006). This time 193,404 children aged 6-7 years in 37 countries and 304,679 children aged 13-14 years from 106 centres in 56 countries participated. For the UK an increase of the 12-month prevalence of asthma in children 6-7 years old was observed, while a decrease of asthma was reported for children aged 13-14 years (Asher MI et al., 2006). The results from the ISAAC III study showed no increase in the prevalence of asthma symptoms in centres with pre-existing high prevalence in the older age group (age 13-14), such as UK

(

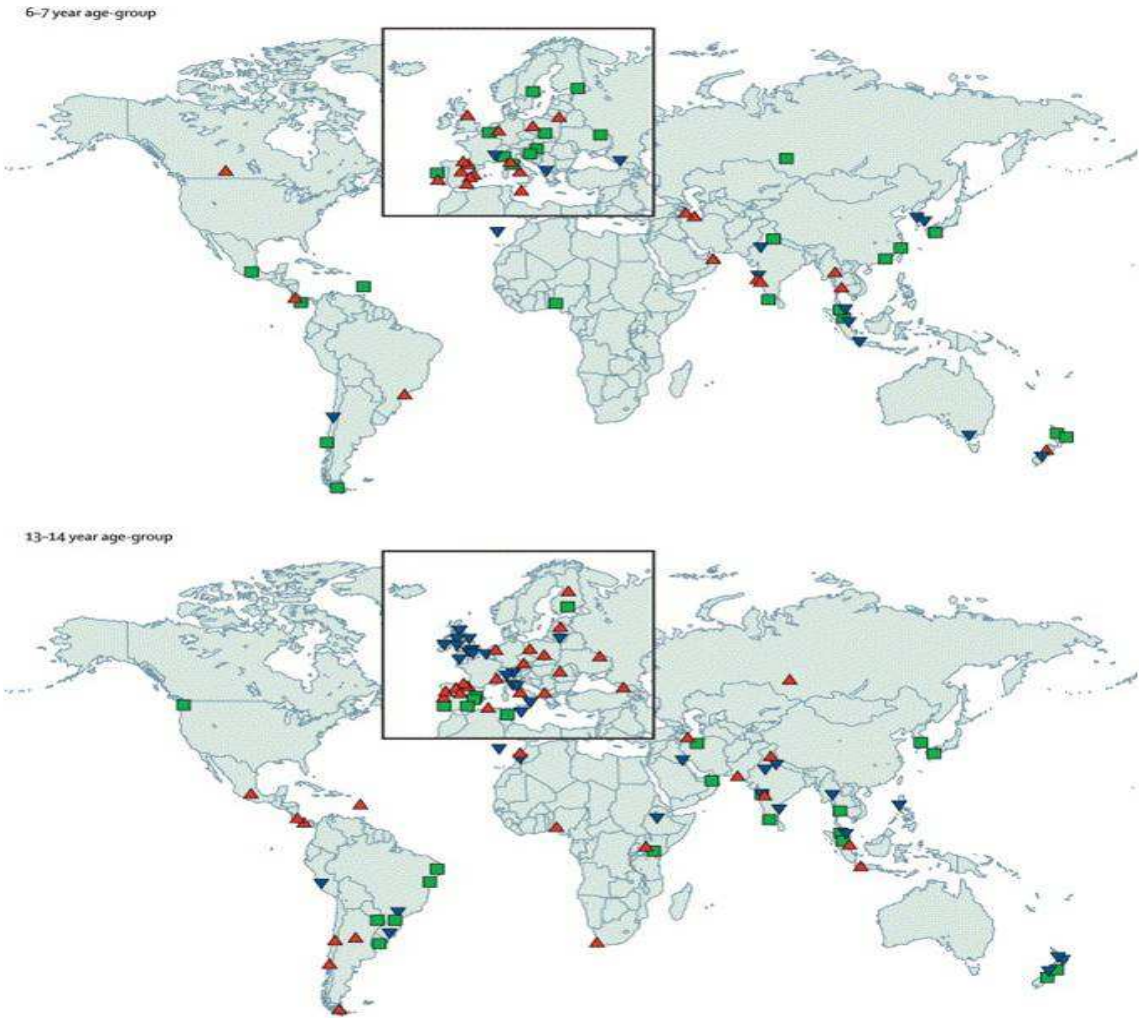


Figure 1-7). At the same time the results showed an increase in the prevalence of asthma in many other centres, such as Spain, Portugal and in countries of Latin America.

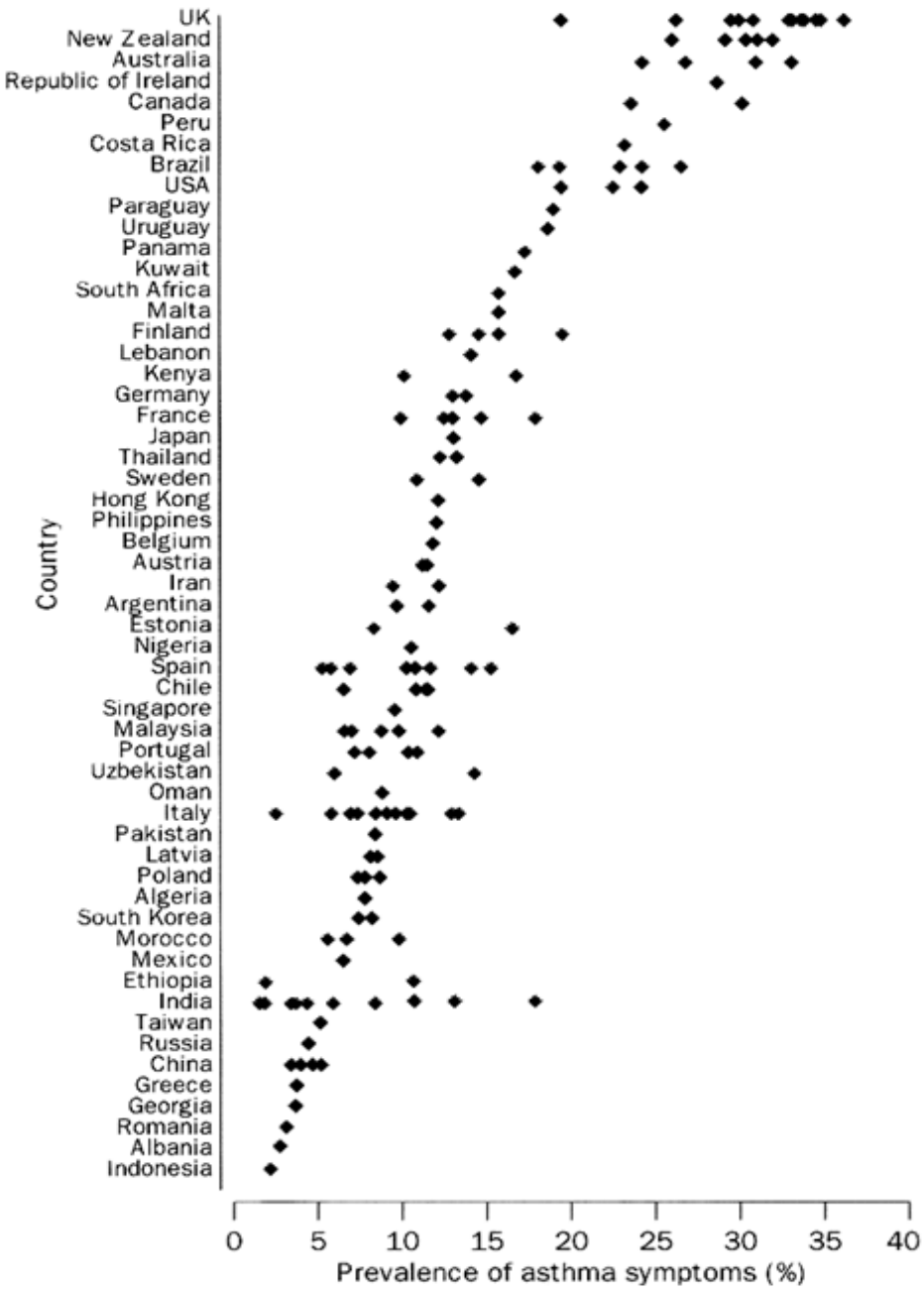


Figure 1-6. 12-month prevalence of self-reported asthma symptoms from written questionnaires, Source: ISAAC Steering Committee (1998) (ISAAC Steering Committee, 1998a).

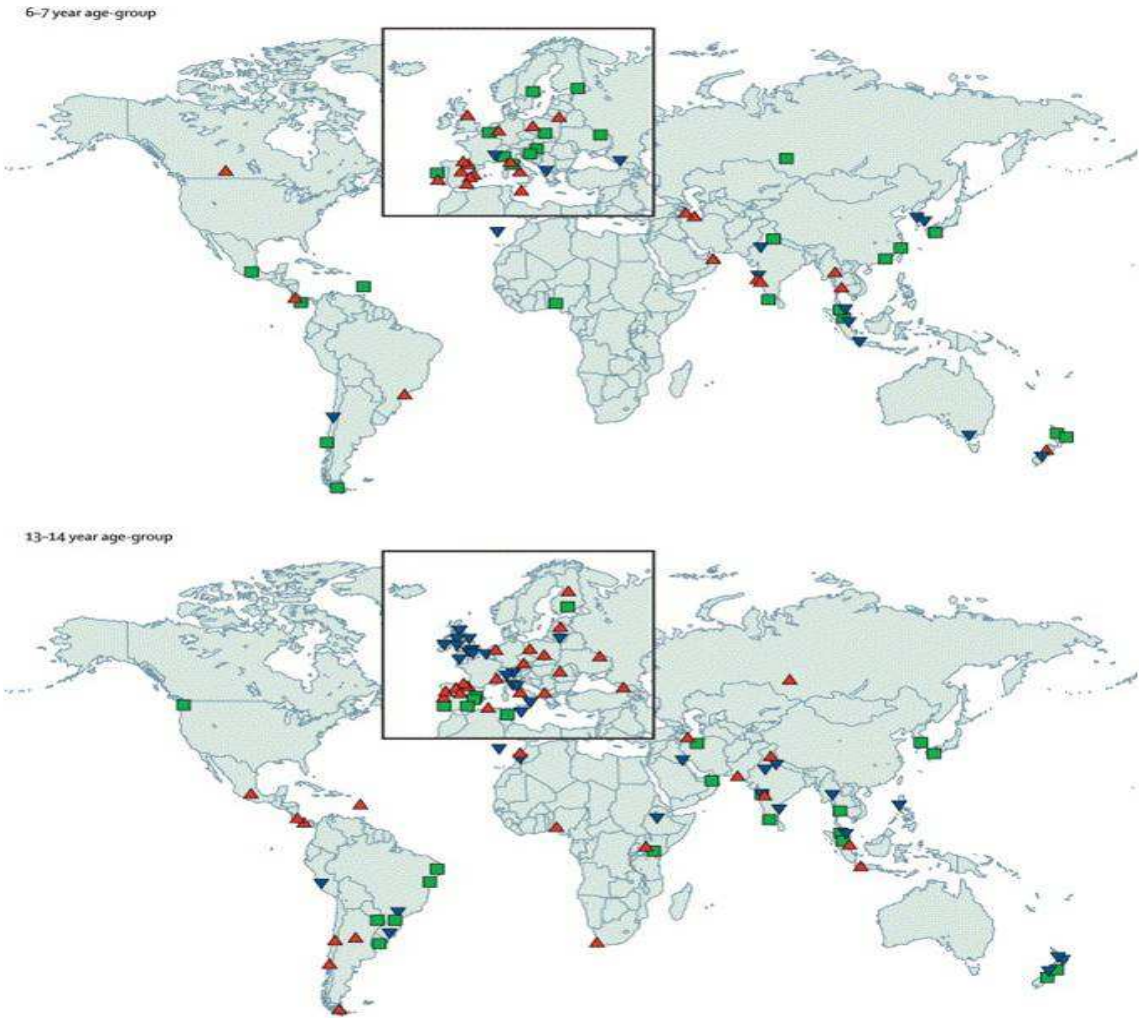


Figure 1-7. World map showing direction of change in prevalence of asthma symptoms for 6–7 year and 13–14 year age-group (Asher et al., 2006).

1.1.1.2 European Community Respiratory Health Survey

The European Community Respiratory Health Survey (ECRHS) started in 1993 aiming to estimate the geographical variation in asthma prevalence in young adults (20-44 years) (Janson et al., 1997). Data were collected on a total of 17,029 individuals from 34 centres in 14 countries. The highest prevalence of diagnosed asthma was found in New Zealand. In Europe, the highest asthma prevalence was found in all centres of the UK and in some of the centres in France (Figure 1-8).

A follow up of ECRHS was conducted in 14 countries 5-11 years later, collecting data between 1998 and 2003 from 11,168 subjects, 20-44 years, in 1991-3 (Chinn S et al., 2004, Janson et al., 2001).

Figure 1-9 shows the geographical variation of asthma as measured by people having an attack of asthma in the last 12 months and/or currently taking medicine for asthma. The ECHRHS found large geographical differences in the prevalence of asthma. Some of the UK fell within the category with the highest asthma prevalence >7% (Janson et al., 2001). The ECRHS follow-up found an increase in the number of people diagnosed with asthma but the same was not reported for the symptom prevalence.

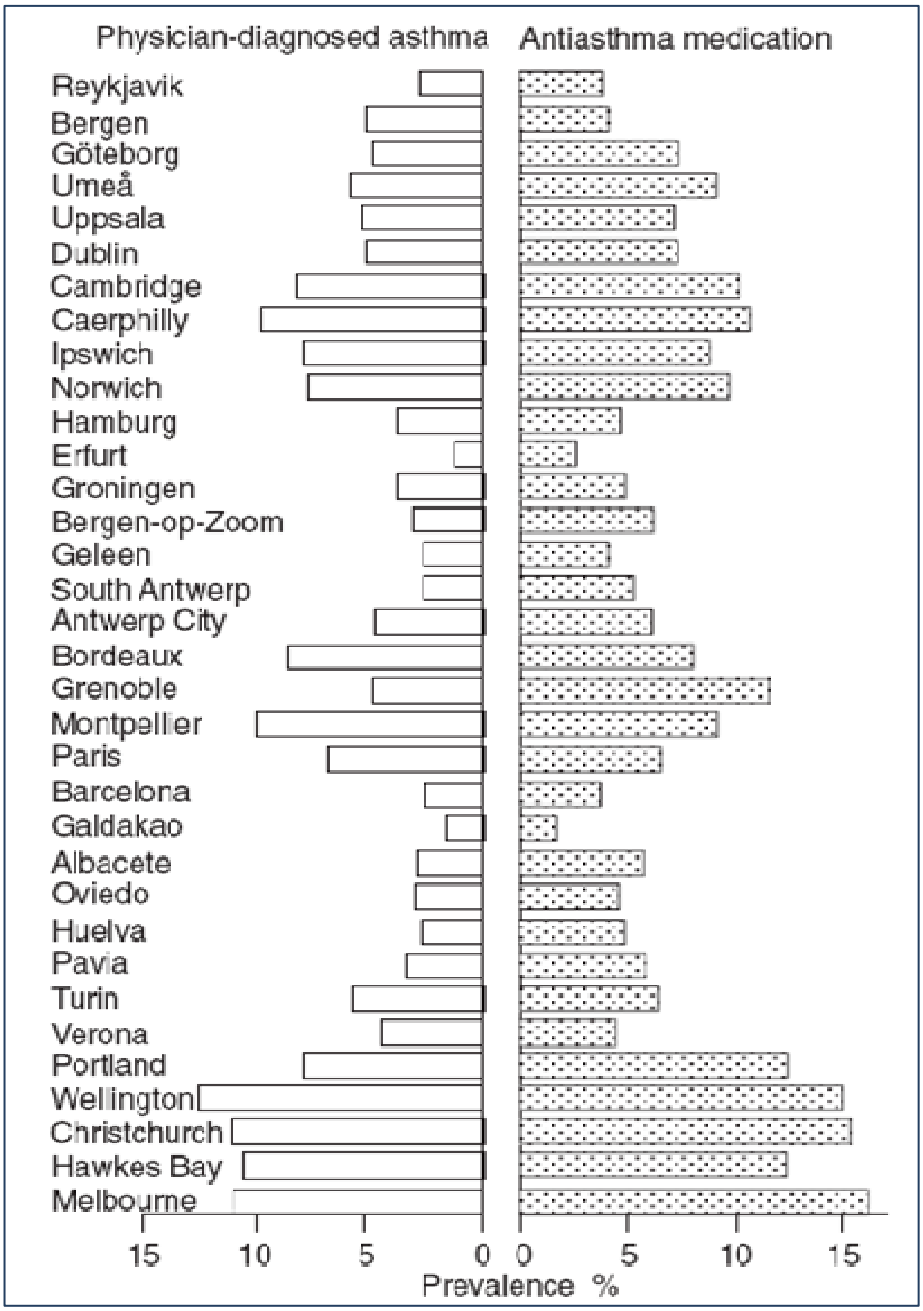


Figure 1-8. Age-sex standardised prevalence of physician-diagnosed asthma and antiasthma medication, adjusted for non-participation, Source: Janson et al (1997) http://www.ecrhs.org/.

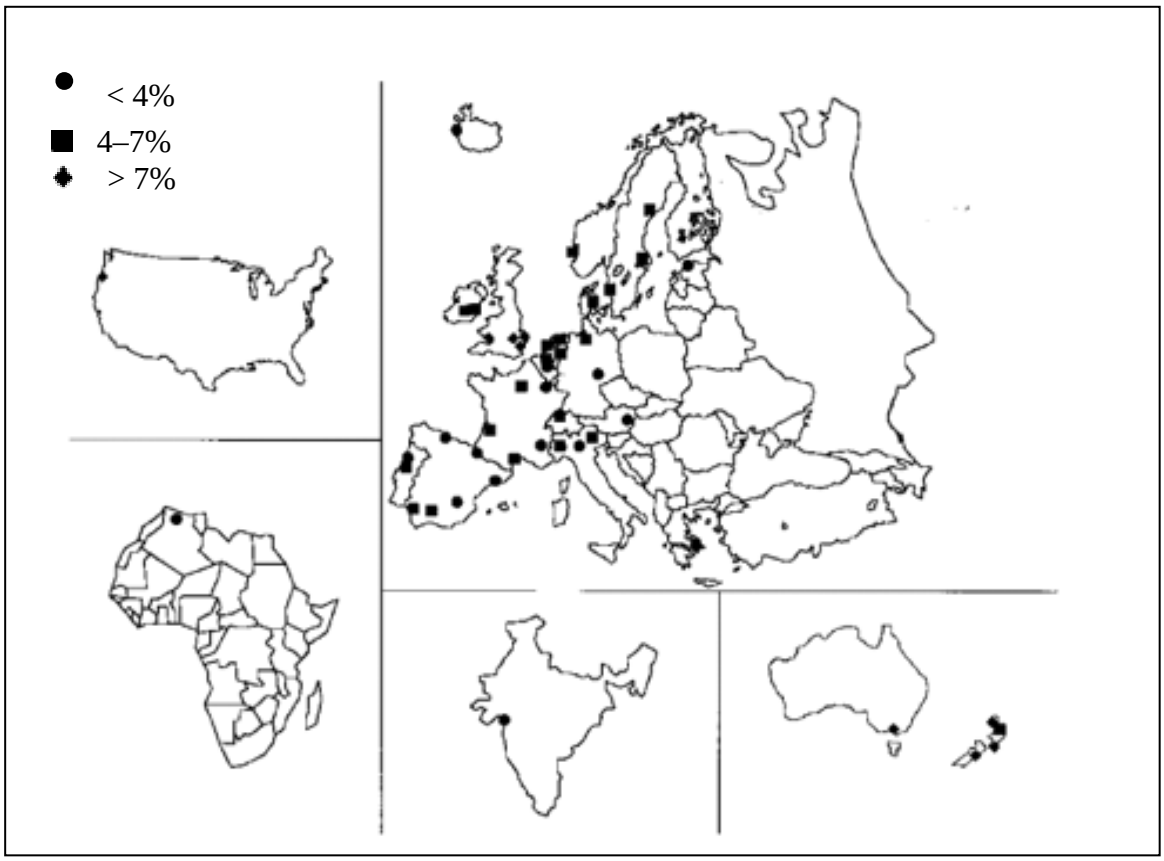


Figure 1-9. Geographical variation in prevalence of asthma (having an attack of asthma in the last 12 months and/or currently taking medicine for asthma), Source: Janson et.al. (2001)

1.1.2 COPD Prevalence

Similar maps to those produced for asthma prevalence by GINA do not exist for COPD. The Global Initiative for Chronic Obstructive Lung Disease (GOLD) was formed in 1998, in an effort to carry out similar work for COPD. According to the latest WHO estimates, currently 210 million people have COPD. However, it is recognised that COPD is an under-diagnosed disease (World Health Organization, 2009a). Estimates of the European Federation of Allergy and Airways Diseases Patients Associations (EFA) stated that "there are more than 600 million people worldwide with COPD", which is almost three times higher than the estimate provided by WHO (European Federation of Allergy and Airways Diseases Patients Associations, 2009a).

COPD is one of the leading causes of death worldwide and was predicted to become the third leading cause of death by 2030 (World Health Statistics, 2008). In 2005, 3

million people across the world died of COPD, which was considerably higher than the numbers who died of asthma (250,000). In 2006 though, WHO launched The Global Alliance against Chronic Respiratory Diseases (GARD) aiming to establish an integrated approach on monitoring (surveillance) of chronic respiratory diseases and their determinants as well as to strengthen national policies for their prevention and control. The WHO GARD Planning Committee also works with the EFA patient society as well as respiratory, allergological and general practitioners societies and government organisations (World Health Organization, 2009b, Viegi et al., 2007). EFA is a European network of allergy, asthma patient organizations that was founded in 1991, in Sweden and added COPD to its mandate, in 2002 (European Federation of Allergy and Airways Diseases Patients Associations, 2009b)

In Europe there are currently, approximately 44 million people suffering from COPD (European Federation of Allergy and Airways Diseases Patients Associations, 2009a). I reviewed data on COPD prevalence in UK. According to the 4th Morbidity Study (Calverley and Sondhi, 1998), the diagnosed prevalence of COPD in the UK was 1% (approximately 600,000 people). Data from the Quality and Outcomes Framework (QOF) showed that 816,341 (1.6% of England's population) people were registered with COPD between April 2009 and March 2010 (NHS The Information Centre for Health and Social Care, 2008), however the QOF database was considered to underestimate the prevalence of some disease domains (Martin and Wright, 2009, Prescription Pricing Authority, 2007). In addition, the QOF figure (1.4%) is not directly comparable to that published by the 4th Morbidity Study (1%) because the latter was based on the whole UK population, while the former was based on the population of England only.

COPD is often under-diagnosed (Coultas and Mapel, 2003, Vestbo, 2004) because repeated measurements are required to establish the diagnosis and differentiate it from other respiratory diseases, such as asthma (Vestbo and Lange, 2002). The availability of spirometry tests in primary care can also influence the degree of under-diagnosis and under-treatment (Cazzola et al., 2009, Walters et al., 2008). Because spirometry is underused for the assessment of COPD in primary care, the development of simple and standardised questionnaires has been suggested by Miravitlles et al (2009) for the assessment of COPD severity. A study in Northern Sweden by Lindberg *et al (2006a)* recently estimated the under-diagnosis of COPD by disease severity and reported that COPD under-diagnosis is related to disease-severity.

The Northern Ireland Cost and Epidemiology of COPD (NICECOPD) study, estimated COPD prevalence to be 6.3% in adults aged 44-69 years (Murtagh et al., 2005).

The project PLATINO (Proyecto Latinoamericano de Investigación en Obstrucción Pulmonar) in five Latin American cities on populations aged ≥40 years, showed prevalence estimates ranging from 11.9% in Mexico City (Mexico) to 19.4% in Montevideo (Uruguay) (Menezes et al., 2005, Menezes et al., 2008).

In Asian countries a high prevalence of COPD was reported as well. In Korea, the COPD prevalence was 17.2% among adults aged ≥45 years, while 7.8% among adults aged >18 years (Kim et al., 2005). Prevalence of airflow limitation was 10.9% in adults aged ≥ 40 years, in Japan (Fukuchi et al., 2004). The overall prevalence of COPD was 8.2%, in a Chinese study (Zhong et al., 2007).

In Scandinavia, surveys of population samples provided estimates of newly occurring cases (incidence rates). The 7-years cumulative incidence of COPD was 11.0% and 4.9% in middle-aged and elderly adults respectively, in Northern Sweden (Lindberg et al., 2006b). The 30-yr cumulative incidences of chronic bronchitis and COPD in middle-aged males from two rural Finish cohorts, were 42% and 32% for continuous smokers, respectively, compared to 22% and 12% for never smokers, respectively (Pelkonen et al., 2006).

Large variability in the prevalence and incidence of COPD may occur depending on the definition of airway obstruction (Viegi et al., 2000). Lindberg et al (2005) estimated prevalence of COPD using the guidelines of the British Thoracic Society (BTS), the European Respiratory Society (ERS), the Global Initiative for Chronic Obstructive Lung Disease (GOLD), and the American Thoracic Society (ATS). The ATS and ERS have issued guidelines for the diagnosis and treatment of patients with COPD (Celli et al., 2004), in order to make comparable data on COPD.

Halbert *et al.* (2006) conducted a systematic review and meta-analysis on global burden of COPD, the results of which approximated a COPD prevalence of 9-10% in adults aged ≥ 40 yrs. They suggested though that bodies such as the GOLD should help to standardise COPD prevalence measurement. A regional study in the North East of England conducted by Melville et al. (2010), in 6,000 males and 6,000 females aged 45–69 randomly selected from a primary care database, also showed a COPD prevalence of 10%.

The ECRHS project is the largest multicentre study providing information on COPD prevalence. More details on the results of the ECHRS are presented below.

1.1.2.1 European Community Respiratory Health Survey

The ECHRS assessed the prevalence of COPD in high income countries, mainly located in Western Europe, analysing data on more than 18,000 young adults (20-44 years). According to ECHRS results, in total 3.6 % had COPD and 11.8% had chronic respiratory symptoms, without airflow limitation (De Marco et al., 2004). The former is classed as COPD stage I+ and the latter COPD stage 0, according to WHO GOLD standards. Overall, ECHRS found COPD to be a considerable issue on young adults. This added to the evidence that COPD is not a burden only in elderly population. Studies that focused on middle age or elderly subjects limit the findings of real COPD burden and the view that COPD is a disease of those aged >50 yrs should be revised (Viegi et al., 2007, Vestbo, 2004).

ECHRS findings also reported that female gender was significantly associated with chronic cough and phlegm (Cerveri et al., 2003). This finding is contrary to the old belief that COPD is a disease of males (Siafakas et al., 1995). In 2000, the number of COPD deaths in USA was higher in females than males. A retrospective cohort of British patients with COPD, found that, from 1990 to 1997, COPD became more frequent in 20-44 year old females (Soriano et al., 2000). This is likely to be related to the increase in smoking prevalence in women.

1.2 Aetiology and Risk factors for Asthma and COPD

Asthma has a long history described first by Hippocrates and his school (460-360 B.C.) (Hippocrates, 1881), however it is difficult to determine whether in referring to "asthma," Hippocrates meant a distinct clinical entity or simply a symptom. Asthma in now well defined as a disease but its causes are not fully understood. Asthma aetiology has a genetic component and there is also evidence that environmental factors are linked to its development (Mutius E. et al., 1992, Cullinan P. and A., 2003). The development of COPD is also considered to be influenced by genetic susceptibility but its main cause is considered to be inhalation of tobacco smoke and other inhaled particles.

There is epidemiologic evidence that long-standing asthma can lead to COPD and patients can have both diseases (Bleecker, 2004, Global Initiative for Chronic Obstructive Lung Disease (GOLD), 2008, Viegi et al., 2007, Soriano et al., 2003, Viegi et al., 2004, Lange et al., 1998). A major difference between asthma and COPD is that the airways obstruction in asthma is reversible while that in COPD is not fully reversible and gets progressively worse over time. Asthma can disappear particularly in children, however in a substantial proportion of cases the disease may come back (Martinez et al., 1995, Strachan et al., 1996, Sears et al., 2003).

The incomplete understanding of causes of both asthma and COPD has motivated many studies investigating avoidable risk factors that exacerbate symptoms in those who already have the disease.

1.2.1 Exacerbation of Asthma and COPD

Studies have found that air pollution is a major risk factor for both asthma and COPD exacerbation (Romeo et al., 2006, Donaldson et al., 2000, Halbert et al., 2006, Anderson et al., 1997, Katsouyanni et al., 2001). Poor air quality caused by tobacco smoke or inhaled particles is one of the major triggers for worsening symptoms and exacerbations for COPD and asthma (Pope C.A. et al., 1995). Other environmental factors that can worsen symptoms are cold air and respiratory infections.

Asthma exacerbations are also linked to allergens like pollen and fungal spores that are seasonal. Few studies have focused on the association of allergens and asthma (Rossi et al., 1993, Atkinson et al., 2006, Glikson et al., 1995). The majority of studies for asthma and COPD have focused on analysing their association with air

pollution, as environmental exposure can potentially become an avoidable risk factor. Socioeconomic differences are associated with asthma and COPD like many other diseases (e.g ischaemic heart disease, many types of cancer, mortality related to alcohol and violence) and this has been reported in the literature since the 1970's.

1.2.2 Air Pollution as Risk Factor for Asthma and COPD Exacerbations

I present below major studies that have focused on analysing the association of asthma and COPD to air pollution. First, I describe the air pollutants analysed in those studies.

1.2.2.1 Air Pollutants

The concentration of air pollutants is often used to assess air quality and to define thresholds that can be harmful to human health and the environment. An air pollutant is a substance in the air that can have the form of a gas, solid particles and liquid droplets. Pollutants are classified as either primary or secondary based on their origin. The primary pollutants are emitted directly from sources such as vehicle exhausts, factories or volcanic eruptions. The major primary pollutants produced by human activities are:

- Sulfur dioxide (SO_2), which is produced in various industrial processes.
- Nitrogen dioxide (NO_2) that is emitted from high temperature combustion. Traffic emissions are a major source of NO_2.
- Carbon monoxide (CO) that is a product of incomplete combustion of fuels such as natural gas, coal or wood. Traffic emissions are a major source of carbon monoxide.
- Particulate matter (PM) consists of tiny particles of solid or liquid, suspended in a gas and is categorised with respect to size. The most widely used definition of PM size is the aerodynamic diameter, the ranges of which are presented in Table 1-1. Human activities, such as the burning of fossil fuels in vehicles, and power plants, and various industrial processes can generate significant amounts of PM.

Fraction	Size range
PM_{10}	≤10 μm
$PM_{2.5}$	≤ 2.5 μm
PM_1	≤ 1 μm
Ultrafine (UFP or UP)	≤ 0.1 μm

Table 1-1. Size of Particulate Matter Fractions.

- Volatile organic compounds (VOCs) are numerous and vary a lot as chemical compounds. They are often divided into the separate categories of methane (CH4) and non-methane (NMVOCs). Methane is a greenhouse gas. Within the NMVOCs, the aromatic compounds benzene, toluene and xylene are known or suspected carcinogens.

Secondary pollutants are formed when primary pollutants react or interact. The main secondary pollutants are:

- Particulate matter (PM) formed from gaseous primary pollutants and compounds in photochemical smog.
- Ground level ozone (O_3) formed by the reaction of sunlight on air containing NOx and VOCs.

There are many studies that have investigated the effect of a range of air pollutants on asthma and/or COPD, in an environmental epidemiology context. I present some of those studies in the next section.

1.2.2.2 Association of Air Pollution to Asthma and COPD

The Air Pollution and Health - A European Approach (APHEA) study analysed air pollution in relation to respiratory diseases in 10 European countries. APHEA is the largest scale study in Europe in the field of respiratory diseases and air pollution and analysed the following main types of pollutants: PM, SO_2, NO_2 and O_3. APHEA looked at hospital admissions for asthma, asthma and COPD and all-respiratory disease admissions. The study confirmed that air pollution in European cities was positively associated with increased numbers of admissions for respiratory diseases (Katsouyanni et al., 2001, Atkinson et al., 2001b, Anderson et al., 1997, Sunyer et al., 1997, Katsouyanni et al., 1997). Variation in air pollutants effect estimates between cities could be explained by cities' characteristics.

Romeo *et a.l* (2006) conducted a review of the time series and panel studies on the short term effects of PM_{10} on increases of the illness in childhood. The results showed that exposure to PM_{10} was associated with an increase in hospitalizations for asthma. In addition, exposure to PM_{10} of asthmatic children was associated with the frequency of asthmatic symptoms (wheezing and cough), the use of anti-asthma medications (in addition to regular therapy) and a decrease in lung function.

A study in London investigated the relationship between daily GP consultations for asthma and other lower respiratory diseases (LRD) and air pollution. Associations were found between air pollution and daily consultations for asthma and other lower respiratory disease. The most significant associations were observed in children and the most important pollutants were NO_2, CO, and SO_2, while in adults the only consistent association was with PM_{10} (Hajat et al., 1999).

Some studies assessed the relationship between air pollution and respiratory diseases by using indicators measured directly in individuals with asthma. In Germany, the association between particulate air pollution and asthma medication use and symptoms was investigated, in a panel study of 53 adult asthmatics, during the winter 1996/1997. The study reported that asthma medication use and symptom increase was associated with particulate air pollution and gaseous pollutants, such as nitrogen dioxide (von Klot et al., 2002). A study in eight North American cities, investigated the relationship between ambient concentrations of five pollutants and asthma exacerbations (daily symptoms and use of rescue inhalers) among 990 children (November 1993-September 1995). PM_{10} and O_3 were found unrelated to exacerbations while strong association was found with CO_3 and NO_2 (Schildcrout et al., 2006).

COPD emergency department visits were associated with the daily ambient concentrations of PM_{10}, SO_2, NO_2, CO and O_3 in the City of São Paulo, Brazil (Arbex M A et al., 2009). PM_{10} and SO_2 readings showed both acute and lagged effects on COPD emergency department visits. Increases in CO concentration showed impacts in the female and elderly groups. NO_2 and O_3 presented mild effects on the elderly and in women, respectively (Arbex M A et al., 2009). A study in Barcelona, found significant associations between the number of emergency room admissions for COPD and SO_2 and Black Smoke (BS) (Sunyer et al., 1993).

Medina-Ramon *et al.* (2006) analysed hospital admissions of COPD and pneumonia in relation to O_3 and PM_{10} in 36 US cities. The study confirmed an increased risk of COPD and pneumonia admissions associated with ambient concentrations of PM_{10} and O_3. This study included a large number of cities and analysed more years of

follow-up than previous multicity studies on respiratory effects of PM_{10} and O_3 (Medina-Ramon et al., 2006).

Pope A.C. *et al* (2004) observed, unlike previous studies, no association of elevated exposures to PM with prevalence of COPD symptoms. A possible reason for this is that the study relied on cause-of-death coding that has a potential for estimation bias for specific causes of death, such as COPD. COPD patients are likely to die from pneumonia or cardiovascular disease (Speizer et al., 1989, Pope Iii et al., 2004). A study focusing on COPD patients living in rural areas of England, between 2006 and 2007, did not observe any positive association between PM_{10}, O_3 and COPD, while positive relationships were observed between CO, NO_2 and COPD admissions (Sauerzapf et al., 2009).

Overall, the existing evidence suggests that air pollution contributes to the exacerbation of asthma and COPD, while the type of pollutants as well as the air pollution levels is modified by city characteristics, such as widely varying climates and geomorphology.

1.2.3 Socioeconomic Status as Risk Factor for Asthma and COPD

Socio-economic factors are linked to asthma and COPD, just like so many other diseases. Socio-economic characteristics of an area expressed as indicators of poverty or housing are linked to susceptibility (Jerrett and Finkelstein, 2005, Yen and Syme, 1999, Pickett and Pearl, 2001). Using local areas' contextual data is an ecological approach accounting for population susceptibility. Even studies that control covariates at the individual level, such as cohort studies, can potentially misinterpret the impact of individuals susceptibility, when not taking into account local area effects (Jerrett and Finkelstein, 2005, Yen and Syme, 1999). According to Pickett and Pearl (2001) this is because "population inequalities in disease are not accounted for, by any known combination of individual genetic and environmental risk and must therefore be attributable to other unmeasured factors, some of which may operate at an aggregate level". The community-level variables are discussed in detail in relevant literature (Wilkinson, 1996, Robert, 1998, Kennedy et al., 1996, Diez-Roux et al., 1997, Diez-Roux, 1998).

1.2.3.1 Association of Socioeconomic Status to Asthma and COPD

A relationship between social class and respiratory symptoms in adults was observed in an early UK study (Speizer and Tager, 1979). Many late studies

confirmed this (Cohen et al., 1977, Marmot et al., 1984, Higgins et al., 1977, Burr and Holliday, 1987, Demissie et al., 1996, Lebowitz et al., 1990).

Two recent studies have shown significant association between greater neighborhood income inequality and higher childhood asthma hospitalization rates (Watson et al., 1996, Cagney and Browning, 2004). In New Zealand, Salmond et al. (1999) found a linear increase in a 12-month period prevalence of asthma with increasing area deprivation. In Vancouver, Canada, Lin et al. (2004) found that exposures to nitrogen dioxide were associated with asthma hospitalization for males in the low socioeconomic group but not in the high socioeconomic group. Nauenberg et al (1999) showed that low family income could predict better asthma exacerbations associated with air pollution than lack of insurance coverage, while Neidell (2004) reported that the effect of pollution was greater for children of lower socio-economic status (SES).

1.3 Medication for Asthma and COPD

The majority of epidemiological studies that have investigated risk factors for asthma and COPD have used hospitalization and visits to emergency departments, to assess the relationship between respiratory conditions and ambient air pollution. In recent years, prescribing of respiratory medication has been used increasingly as a marker of respiratory diseases' morbidity. I used prescribing of respiratory medication to assess asthma and COPD exacerbation in relation to air pollution. Next, I will describe the pharmacological interventions for asthma and COPD and review the studies that have adopted prescribing of respiratory medication as an indicator of air pollution effects on asthma and COPD.

1.3.1 Pharmacological Interventions in Asthma and COPD

The medication for asthma and COPD is classed into two main types: 1) β2-agonists that dilate the airways (bronchi and bronchioles) facilitating airflow and 2) corticosteroids that are used for long-term management of asthma and COPD (British National Formulary, 2009). There are two types of β2-agonists, the short-acting and long-acting β2-agonists. Short-acting β2-agonists (SABA) - often called short-acting bronchodilators or "quick relief" medication – help reduce asthma and COPD symptoms or stop the symptoms of an attack in progress. In cases where the airways obstruction is more severe the short-acting β2-agonists are prescribed regularly. In patients who remain symptomatic or have two or more exacerbations in a year, long-acting β2-agonists are prescribed (British National Formulary, 2009).

Mucolytic medication is related to COPD management, as it hydrolyzes mucus reducing its viscosity. This type of medication was used in many countries, such as Germany but currently its use is controversial. Studies on their efficacy at preventing exacerbations have shown that they are effective in preventing acute exacerbations (Allegra et al., 1996, Petty, 1990) but also that they are ineffective at preventing exacerbations or deterioration of lung function (Decramer et al., 2005). Long-term studies with mucolytics have shown little effect on lung function or symptoms, although some have shown a reduction in exacerbations. The GOLD guidelines state that the widespread use of mucolytics cannot be recommended for COPD management, given the current evidence (Global Initiative for Chronic Obstructive Lung Disease (GOLD), 2008).

1.3.1.1 Prescribing of short-acting β2- agonists

Respiratory prescriptions account for 7% of all prescribed drugs, in the UK (British Thoracic Society, 2006). In England in 2004 about 51 million prescriptions were dispensed in the community for the prevention and treatment of respiratory disease. Almost half of these (24,785,000 or 49%) were for inhaled β2-agonists (bronchodilators) used in the treatment of asthma (Table 1-2) (British Thoracic Society, 2006).

	Prescriptions (thousands)
General respiratory drugs	
Inhaled bronchodilators	24,785
Inhaled corticosteroids	13,480
Cromoglycate and realted therapy	74
Leukotriene receptor antagonists	597
Antihistamines, hyposensitisation and allergic emergencies	8,741
Oxygen	704
Mucolytics	97
Aromatic inhalations	22
Cough preparations	1,672
Systemic nasal decongestants	624
Antituberculous Drugs	**71**
Cystic fibrosis drugs (pancreatin)	**153**
Total	**51,020**
this table does not include drugs dispensed in hospitals	

Table 1-2. Prescriptions used in the prevention and treatment of respiratory disease. England 2004 (Department of Health, 2004).

Figure 1-10 shows the quantities of different types of medication prescribed for asthma and COPD in general practices in England. One of the most notable observations is that the short-acting β2-agonists were prescribed most often to treat asthma and COPD in England compared to other medication.

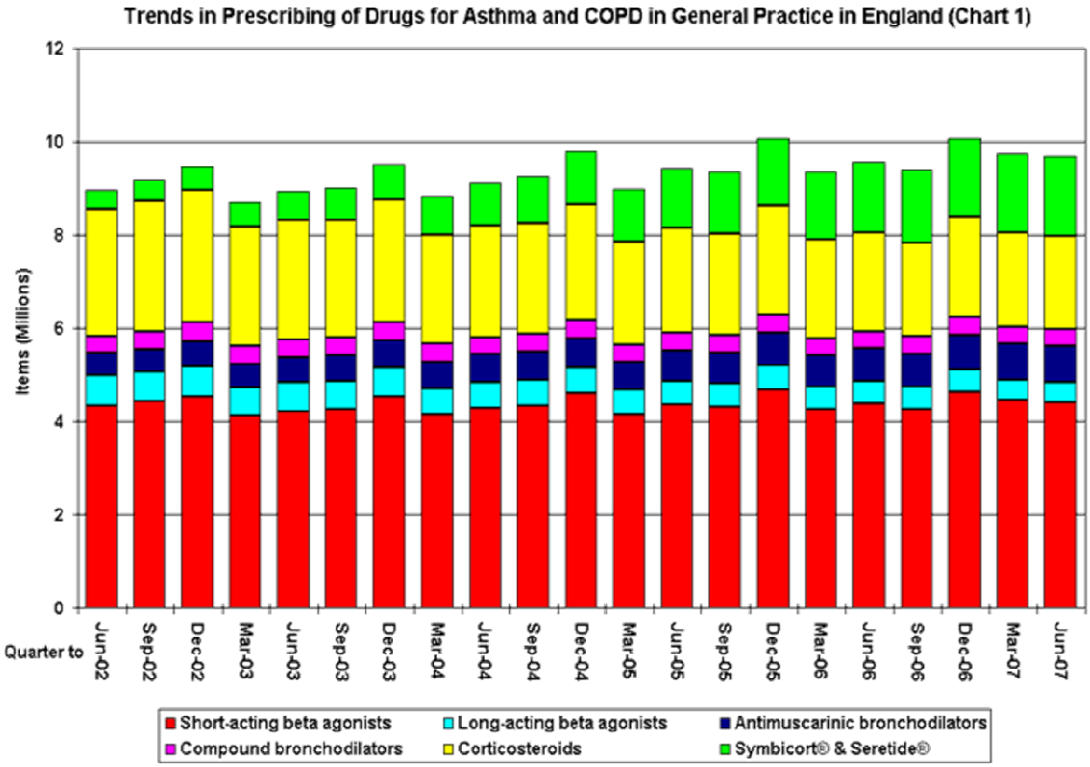


Figure 1-10. Prescribing of drugs for Asthma and COPD in General Practices in England (Prescription Pricing Authority, 2007).

A second observation is that prescribing of medication increased over the last five years and this is mainly due to increased prescribing of Symbicort and Seretider (Figure 1-10). These are combination products containing a long-acting β2-agonist plus corticosteroid. However, concerns have been expressed recently regarding the side effects of the components of such combination products and the effectiveness of therapy and discontinuing of drugs that do not produce measurable improvements is reviewed (Prescription Pricing Authority, 2007). The prescribing of short-acting β2-agonists has not had any noticeable changes during the period of 2002–2007, other than increasing in the quarter to December each year (Figure 1-10). The drug salbutamol represents 93% of all prescribing for short-acting β2-agonists and 89% of the cost (Prescription Pricing Authority, 2006).

1.3.1.2 Respiratory Prescribing as Indicator for Respiratory Diseases

The majority of studies that have analysed the association between air pollution and asthma and COPD used hospital admissions and attendance & emergency visits as indicators (Walters et al., 1994, Schwartz, 1994, Atkinson et al., 2001a, Tolbert

et al., 2000, Arbex M A et al., 2009, Sunyer et al., 1993, Sunyer et al., 1997, Medina-Ramon et al., 2006). These indicators capture patients with severe asthma or COPD symptoms.

According to the National Asthma Campaign only 20% of people with asthma are described as having a 'severe or very severe' asthma (National Asthma Campaign). There is lack of epidemiological data on COPD severity but a few studies that have investigated COPD severity reported that the majority of subjects analysed had mild COPD (Lindberg et al., 2006a). COPD prevalence was studied in Japan, and 56% of cases were found to be mild, 38% moderate, 5% severe, and 1% very severe (Fukuchi et al., 2004), while a study in Korea reported that the majority of these cases were found to be mild in degree (Kim et al., 2005).

The main advantage of prescribing in relation to traditional indicators for asthma and COPD is therefore the fact that prescribing is considered to capture patients from any severity class. The Asthma Insights and Reality in Europe (AIRE) Survey is a multi-national survey for the severity, control and management of asthma in children and adults (Vermeire et al., 2002). Medication usage was assessed among many other aspects of asthma control and management. Figure 1-11 shows that quick relief medication was prescribed to patients with any severity type. More details on AIRE study are presented in Appendix A.

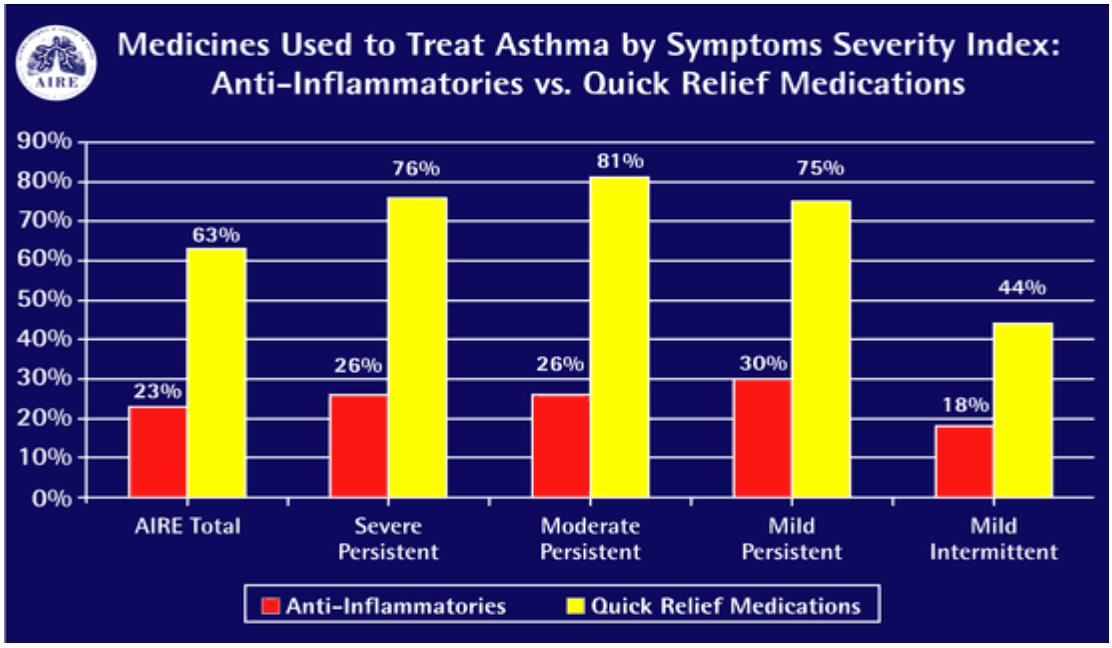


Figure 1-11. Asthma medication use per symptoms' severity, Source: *Asthma Insights and Reality in Europe (AIRE) Survey.*

No similar data exist on use of medication per COPD severity class. As mentioned earlier, COPD is under-treated, which means a lack of appropriate medication use, especially corticosteroids that are suitable for long-term management of COPD. The lack of appropriate prescribing for long term management of COPD leads to patients having more frequent exacerbation of symptoms, making them rely more on the use of β2-agonists (quick relief medication). Therefore, I would expect that β2-agonists would also be used often by COPD patients of any severity.

Naureckas et. al. (2005) evaluated the use of short acting β2-agonist prescriptions as a surrogate indicator for asthma exacerbation with respect to traditional indicators (hospital admissions & emergency-department visits) during 1996-1998 in Illinois, Chicago. They concluded that a very strong and significant association was observed between those two indicators, which suggested that β2-agonists prescriptions can be used as a marker for asthma morbidity (Naureckas et al., 2005).

1.3.2 Respiratory Prescribing as Indicator of Air Pollution Health Effects – Literature Review

I found only a few studies that used respiratory prescribing as an indicator of the air pollution effect on respiratory exacerbations. I used the databases Web of Science, Medline and Scopus for my search. The searching keywords were 1) Topic=(respiratory prescribing) AND Topic=(air pollution), 2) Topic=(respiratory drug sales) AND Topic=(air pollution) and 3) Topic=(respiratory drug sales) AND Topic=(air pollution).

My searches in May 2010 returned four studies that analysed the association of respiratory prescribing in relation to air pollution (Laurent et al., 2009, Pitard et al., 2004, Zeghnoun et al., 1999, Vegni et al., 2005). Three of the four studies were from France, while the fourth was conducted in Italy. To conduct such studies requires the availability of prescribing data at population level. Such availability is determined by the structure of the health care system of each country. I also located an unpublished study, currently in development, in California, US (Griffiths et al., 2009, Griffiths et al., 2003). I present below the studies and then discuss the main aspects on them.

1.3.2.1 Literature Review Outcome

Zeghnoun et al.'s (1999) study had a temporal ecological design, choosing respiratory medication sales data as an indicator for the short-term effects of

ambient air pollution in the city of Le Havre (200,000 inhabitants), in France, from June 1993 to December 1996. Mucolytic medication, prescribed for the management of COPD, as well as anti-cough medication for children and adults, were assessed in this study. The total number of daily respiratory drug sales data was related to daily ambient air concentrations of SO_2, NO_2, and black smoke (BS). The air pollutants concentration data were recorded by a number of stations and exposure was calculated as the average of values from the selected stations, defining a spatially homogenous air pollution exposure. They employed an autoregressive Poisson regression model adjusting for time trends, seasonal variations, influenza epidemics, and weather. Respiratory drug sale was associated with BS-24h, NO_2-24h and SO_2-1h, with reported lag 0 to 3 days. The strongest association was observed for BS-24h, where an increase of two standard deviations above the mean (13.2 to 36μg/m^3) was linked to a 3.7% increase in respiratory drug sales, for 1 day lag. For NO_2 and SO_2, an increase of two standard deviations above the mean of NO_2 h (33.7 to 64 μg/m^3, 1 day lag) and SO_2 h (97.8 to 355 μg/m^3, 3 days lag) was associated respectively with an increase of 3.3% and 2.7% of respiratory drug sales. Respiratory drug sales were associated with most pollutants studied with lags varying from 1 to 9 days (Zeghnoun et al., 1999).

Pitard et al. (2004) also used a temporal ecological design to assess the effect of air pollution on bronchodilators (β2-agonists), and cough and cold preparation sales in the city of Rouen (106,592 inhabitants), in France, from July 1998 to June 2000. Asthma and COPD medication was used (β2-agonists and corticosteroids) that represented 33% of the total respiratory drug sales in the city as well as cough and cold medication that represented 67% of the total respiratory drug sales. The air pollution exposure was calculated as the average of daily concentrations of NO_2, SO_2 and BS recorded by two air pollution stations, defining a spatially homogenous air pollution exposure. Lags of up to 10 days were included in the analysis for each pollutant. A generalized additive model (GAM) was employed, adjusting for day of the week, seasonal variation, influenza epidemics, weather and bank holidays. A 10- μg/m^3 BS increase was significantly associated with a 6.2% (95% CI, 2.4-10.1%) increase in the sales of anti-asthmatic and COPD medication, and with a 9.2% (95% CI, 5.9-12.6%) increase in the sales of cough and cold preparation for children aged under 15 years. For the anti-asthma and COPD medication, associations were found for lags ranging from 5 to 7 days with NO_2, and for lags ranging from 1 to 7 days with BS. There was no association between SO_2, and anti-asthmatic and COPD drug sales. The results of this study suggested that an increase in anti-asthmatic and COPD drug sales was directly associated with BS and NO_2 concentrations but not with SO_2 (Pitard et al., 2004).

Vegni et al. (2005) also used a temporal ecological design. They evaluated respiratory drug dispensing data as a health indicator in relation to air pollution in the city of Como (84,713 inhabitants), in Italy, for 1995-1997. For this study, respiratory drugs were defined as the anti-asthma therapeutic group including both β2-agonists as well as corticosteroids prescribed by GPs. They used two health indicators: 1) count of individual patients with respiratory drug dispensed (Cases) and 2) weekly dispensed Daily Defined Doses (DDD) of drugs. Weekly air mean concentrations of total suspended particles (TSP) were recorded by one available station, defining a spatially homogenous exposure. The health indicators were modelled using a random effects Poisson regression model adjusted for long-term trends, seasonal variations, calendar variations due to holidays, and weather. The analysis showed a strong association of health indicators with pollution, with an increase of 8.2% (95% CI 0.2%-16.9%) of cases and 13.7% (95% CI 4.4%-23.8%) for DDD from the 10th to 90th percentile of TSP (29-92 $\mu g/m^3$). Overall, these results were interpreted as an approximate increase of 13% of Cases and 22% of DDD for a 100 $\mu g/m^3$ increase of TSP (Vegni et al., 2005).

Laurent et al.'s (2009) spatio-temporal ecological study evaluated the short-term relations between ambient air pollution and sales of short-acting β2-agonists (SABA), in Strasbourg (450,000 inhabitants), France in 2004. The date, age group, sex and census block of residence were also extracted for each SABA sale. The Atmospheric Dispersion Modelling System (ADMS) urban Gaussian dispersion model was used to model hourly mean concentration of PM_{10}, NO_2 and O_3 at small-area level (census block). Socioeconomic status, meteorological data, pollen counts as well as counts of influenza cases were included in the conditional logistic regression model. Increases of 10$\mu g/m^3$ in ambient PM_{10}, NO_2, and O_3 concentrations were associated, respectively, with increases of 7.5% (95% confidence interval [CI], 4 to 11.2%), 8.4% (95% CI, 5 to 11.9%), and 1% (95% CI,- 0.3 to 2.2%) in SABA sales. Deprivation had no influence on these relations.

The study by Griffiths et al (2009) is work in progress and I have not located relevant publications in peer reviewed journals as yet. Griffiths et al.'s (2009) spatio-temporal ecological study examined the effects of chronic exposure to air pollution on asthma exacerbation through asthma prescriptions for quick-relief medications (β2-agonists) in California, US, for each quarter from 1998 to 2001. They aggregated information on the use of maintenance therapies by each patient at the 5 digit zip code level in California. Using individual data they were able to stratify prescribing data by asthma severity as well as by age. The spatial variation of exposure to air pollutants PM_{10} and O_3 was predicted for each zip code in California using ordinary kriging, capturing a spatial prediction of the ambient

pollution and weather conditions for each zip code, defining a heterogenous exposure (Griffiths et al., 2009). A negative binomial model was employed to approximate the expected number of prescriptions for each zip code with additional covariates, such as the population demographics (e.g., median household income, percent urban population, race/ethnicity). In general, they reported a positive relationship between asthma and both PM_{10} and O_3 levels (Griffiths et al., 2009).

1.3.2.2 Discussion of Literature Review

The points of discussion regarding these studies lie in the following areas: 1) air pollutants, 2) exposure assessment and 3) medication type. I will compare and discuss those areas below, in order to highlight the variation and challenges on this area of research.

Respiratory medication

The five studies presented above use asthma and/or COPD medication to investigate the association between air pollution and asthma and/or COPD. Two earlier studies (Pitard et al., 2004, Vegni et al., 2005) used β2-agonists and corticosteroids as indicators of respiratory symptoms (asthma and COPD), while the two most recent studies (Griffiths et al., 2009, Laurent et al., 2009) focused on short-acting β2-agonists as indicator of asthma exacerbation. The earliest study in this field, conducted by Zeghnoun et al. (1999), used mucolytic medication and anti-cough medication.

Zeghnoun et al.'s (1999) study initiated the use of respiratory prescribing as a health indicator of air pollution effects in respiratory diseases, however the specific medication they used as marker of COPD exacerbation should not be used anymore, given contradictory evidence on its efficacy. Choosing a medication as an indicator of a medical condition is not straightforward and a very good knowledge on medication and the disease under study is required.

Air pollutants

The Zeghnoun et al. (1999) and Pitard et al. (2004) studies from France used ambient concentrations of SO_2, NO_2, and BS as indicators of air quality. Vegni et al. (2005) in Italy used total suspended particles (TSP). The most recent published study by Laurent et al. (2009) used concentrations of PM_{10}, NO_2 and O_3, and the unpublished study by Griffiths et al. (2009) also used concentrations of PM_{10} and O_3 as indicators of air quality.

Black smoke consists of fine solid particles suspended in air, which mainly arise from the incomplete burning of fossil fuels, in the domestic, industrial or transport sectors. Total suspended particulate matter (TSP) is a collective name for airborne particles or aerosols that are less than 100 micrometers. Most modern air pollution stations have stopped measuring BS or TSP and instead record particulate matter of specific sizes (Table 1-1 page 19), such as PM_{10}. The BS and TSP were reported by the earliest studies (Zeghnoun et al., 1999, Vegni et al., 2005, Pitard et al., 2004), while the recent studies (Laurent et al., 2009, Griffiths et al., 2009) have used PM_{10}. I have also used PM_{10} as air quality indicator, therefore my study is most similar to the two most recent studies.

Pollutants recorded in the same station are correlated most of the time. Zeghnoun et al. (1999), Pitard et al. (2004), Laurent et al. (2009) and Griffiths et al. (2009) also reported that the mean levels of air pollutant concentrations measured at stations appear to be highly correlated in time. Due to the correlation of air pollutants replication of information can occur when using all in the same model, as indicators of air quality.

Air pollution exposure

The previous studies on air pollution and respiratory prescribing assumed homogenous air quality exposure except the study of Laurent et al. (2009) where heterogeneous exposure was estimated at census block level. The study under development has also used heterogeneous exposure (Griffiths et al., 2009). Preliminary work for this study in 2003, had assumed homogenous exposure and the authors had reported that their initial model suffered from the lack of heterogeneity (Griffiths et al., 2003).

The assumption of homogenous air pollution exposure on the other studies may be related to the absence of an air pollution monitoring network. Many epidemiological studies in the past have used homogenous exposure (Katsouyanni et al., 1996). Vegni et all (2005) explicitly stated the presence of only one monitoring station in the area. The most recent published study by Laurent et al. (2009) reported the use of the existed urban atmospheric dispersion model ADMS to model air exposure in time (hourly) and at small-area level (census block).

Covariates

All previous studies used air pollutants, which are the primary factors of concern but also included additional covariates that are relevant to worsening of respiratory symptoms. The three earlier studies (Zeghnoun et al., 1999, Vegni et al., 2005, Pitard et al., 2004) controlled for seasonal factors, influenza, pollen levels, and weather (temperature and humidity). In addition, the studies by Zeghnoun et al. (1999) and Pitard et al (2004) took into consideration the day of the week and bank holidays. Both studies analysed daily data and observed zero pharmacy sales of respiratory medication (prescribed by GPs) on specific days (i.e. Sundays). Vegni et al. (2005), on the other hand, stated that they aggregated the data on a weekly basis avoiding assumptions on the closure days of different pharmacies and complicating the model.

Vegni et al.'s (2005) model investigated only the effect of air pollution on a weekly basis. No attempt was made to assess lag times on cumulative periods of exposure and respiratory medication sales. In contrast, other studies tested whether there may be a lag period between air pollution exposure and sales/prescribing of respiratory medication. Zeghnoun et al. (1999) investigated lags of 1 to 14 days and found strong associations for lags 0 to 3 days as well as 6 to 9 days. Pitard et al. (2004) included lags up to 10 days. The strongest associations were observed for 6 to 7 days lag with NO_2, and 2 to 5 days lag for BS. The most recent published study (Laurent et al., 2009) observed associations involving latency periods of 4 to 10 days. Finally, no analysis of lag days was reported in the ongoing study of Griffiths et al (2009), even though they discussed that increase in asthma medication use due to air pollution exposure eventually (perhaps with a lag) leads to the filling of a prescription.

The recent study in France (Laurent et al., 2009) accounted for socioeconomic characteristics in addition to the other covariates. They used an index of socioeconomic status built from 52 socioeconomic variables (income, educational level, employment, housing characteristics). Overall, socioeconomic status was not found to be a significant covariate. The authors discussed that similar studies in other settings should confirm whether the lack of interaction with deprivation was due to specific local conditions. The ongoing study by Griffiths (2009) includes population characteristics, such as median household income, percent urban population and race/ethnicity as well as seasonal or quarterly factors. In this study household income was negative statistically significant, suggesting that households with lower incomes obtained more asthma prescriptions.

1.4 Primary Care and Chronic Diseases

In 2007, the World Health Assembly set out actions to be implemented over the next six years, in order to prevent and control chronic non-communicable diseases, such as asthma and COPD (World Health Organization, 2008a). Part of the plan to achieve this was to strengthen the management of the diseases in primary health care and use accessibility of medication in primary care as an indicator to monitor progress (World Health Organization, 2008b). In order to use primary care based prescribing data as an indicator, baselines of medication use have to be available, which will have to be established by collecting relevant data (World Health Organization, 2008b). In the UK, a consistent priority for NHS reform is the shift from provision of hospital-based acute care to proactive care delivered in primary care (Department of Health, 2006, Department of Health, 2005, Department of Health, 2010). The importance of exploiting data at primary care level and especially data of respiratory diseases is discussed below.

1.4.1 Primary Care and Chronic Respiratory Diseases

Most epidemiological studies use data from secondary and tertiary care such as hospitals and emergency care, even though the vast majority of patients are fully dealt with at the primary care level. The important role of primary care for health services has been well recognised (World Health Organization, 1978, World Health Organization, 2003) and the UK is one of the countries that has developed it comprehensively (Starfield B et al., 2005, Starfield B., 1994). The majority of primary care consultations in the UK are for patients with respiratory disease (Pinnock and Sheikh, 2009, British Thoracic Society, 2006), emphasizing the significant role of primary care in managing respiratory diseases.

High quality research, focusing on primary care management, is fundamental to understanding and developing a comprehensive healthcare service for respiratory disease. Several studies have been conducted regarding management of respiratory diseases at a primary care level. However, I located only two studies that utilized GP level respiratory data from an environmental health perspective, in UK (Hajat S. et al., 1999, Dunn et al., 1995). Dunn et al. (1995) investigated airborne emissions from a factory (producing plastic coated wallpaper) and prevalence of asthma as identified from a computerised repeat prescribing system of three GP practices. Hajat et al. (1999) investigated the relationship between daily GP consultations for asthma and other lower respiratory diseases (LRD) and air pollution in London.

One of the main challenges in environmental epidemiology is linking health data to environmental and socioeconomic data that are available on spatially and temporally misaligned units.

1.4.2 Spatial Analysis of Primary Care Population Data

If patients want to access primary care services in UK, they register with a GP practice, and there are no geographic constraints on their choice. However, most people choose to register with a practice that is close to the place they live or work. Several factors such as area deprivation and transportation links or other factors may influence their choice. Consequently, there is no formal definition of boundaries of primary care service areas and the population of any given area can be affiliated to a number of practices. A traditional way to define GP practice catchment areas has been to align them to administrative boundaries (Congdon and Best, 2000). If we assume that 60% of a practice population A come from the administrative spatial unit of Lower Social Output Area (LSOA) 1 and 40% come from practice population B, then this LSOA would be part of practice A catchment area.

The use of administrative boundaries is a common method for the regionalisation of health care and there are certain advantages of such approach. Main advantages are that these spatial units are readily available and represent to some extent the area the practice population live. Main disadvantages are that they provide no information on the overlap between catchment areas, as they form adjacent spatial units, and they are rough approximations of the distribution of registered patients, so few will probably relate directly to the GP health service utilisation areas. In addition, because they are not accurate, any measure on catchment areas or health, well-being and determinants of ill-health within them can result in compromising the reliability of those measurements. In particular, the spatial distribution of environmental data does not follow the spatial distribution of administrative boundaries. Therefore, when using the traditional GP catchment areas to derive measures of environmental exposure of a practice population increases error in exposure misclassification.

1.5 Spatial Analysis

Spatial analysis is defined as the: "the quantitative study of phenomena that are located in space" (Bailey and Gatrell, 1995). Epidemiology has a geographical basis when it characterises disease occurrence by time and place. Consequently disease mapping has a long tradition (Timmreck, 2002), dating back to 1800s when maps of diseases were used to analyse causes and spread of disease outbreaks (Walter, 2000).

Over the last two decades, epidemiologists have applied statistical spatial analysis in geographical information systems (GIS). Studies have looked at cases of leukaemia, lymphoma, larynx and lung cancer in relation to industries (Gardner et al., 1990, Elliott et al., 1992, Elliott and Wartenberg, 2004). This type of spatial analysis has been previously used extensively in other scientific disciplines, such as ecology, economics and sociology (Diggle, 1983 , Bronars and Jansen, 1987, White et al., 1981, Rushton et al., 2006, Rushton et al., 1997, Lurz et al., 1997). Geographical Information Systems (GIS) and statistical spatial analyses are applications of what is known as spatial analysis. Spatial analysis is used to: 1) increase the basic understanding of spatial processes, 2) test hypothesis regarding the relationship of the variables under study and at best 3) predict the strength of the relationship of variables in the future.

1.5.1 Geographical Information Systems

GIS is a computer based technology that captures, stores, analyses and presents data which are referenced spatially on the earth's surface. Spatial data used in a GIS environment can be generated by satellites, digital aerial photographs and Global Positioning Systems (GPS) as well as by scanning and digitising paper maps and aerial photographs (Schuurman, 2004). Conversion of postcodes to grid references is also a method for acquiring spatial data, called geo-coding (Crampton, 2005). Core functionality of GIS is the creation and display of maps electronically. Two of the most useful GIS operations are overlay and proximity analysis (Heywood et al., 1998). With overlay analysis it is possible to integrate large amounts of data in map layers format (i.e. overlay asthma health outcomes on air pollution and socio-economic data). Proximity analysis is used to determine relationships between selected points or areas and their neighbours (i.e. identify the population present in a 2.5km radius around an industrial facility).

The true power of GIS, however, lies in its modelling and exploratory capability. Modelling is applied when we want to test a hypothesis or estimate the relationship between two or more variables with some precision. Exploratory analysis on the other hand may lead to insights into the data and formulation of a hypothesis (Fotheringham and Rogerson, 1994, Bailey and Gatrell, 1995). Modelling and exploratory analysis can not always be clearly distinguished. This is due to the level of complexity of various techniques. Examples of this are the widely used interpolation techniques. With such techniques it is possible to estimate values of variables in unsampled areas, based on data provided from sampled areas (i.e. estimate air pollution based on air pollution monitoring). Some interpolation techniques such as ordinary or universal kriging require statistical modelling and are described as models, while other simpler techniques such as Thiessen polygons fall into exploratory analysis (Bailey and Gatrell, 1995). Statistical analysis is not a primary operation in GIS software, therefore it is linked to statistical software when statistical modelling is involved.

A range of GIS operations and functionalities have been used for research on asthma and air pollution. Wilkinson (1999) assessed the hospital admission of children with asthma in relation to road traffic in London, using proximity analysis (Wilkinson et al., 1999). Lu et al. (2003) analysed ozone and asthma in children in the South Coast of California, using a kriging interpolation technique to estimate air pollution and a Poisson spatial regression model (Lu et al., 2003). Maantay (2007) investigated asthma and air pollution in the Bronx, New York, using proximity analysis as well as applying air dispersion modelling for estimating exposure.

GIS techniques have been widely used for public health research. WHO, European Community (EC) and Centre of Disease Control and Prevention (CDC) recommend such techniques for environmental health studies (World Health Organization, 1994, National Center for Health Statistics, 2005, European Commission, 2004). A variety of GIS software is now available, both commercial and open source (software developed and distributed free), that allows storing and manipulating of all different types of digital spatial data. The commercial software ArcGIS and the open source GRASS GIS (Neteler and Mitasova, 2004) are two of the most software packages (Kennedy, 2006).

1.5.2 Spatial Statistical Modelling

Spatial statistical modelling is determined by the theories of spatial statistics that give explicit consideration to the possible importance of the spatial arrangement of spatial data. Spatial statistics was developed during the late 40s and early 50s by

statisticians such as Moran, Geary and Whittle (Florax and van der Vlist, 2003). The development of spatial statistics was slow until the early 1970s when Cliff and Ord presented their influential work on spatial autocorrelation (Cliff and Ord, 1973). In the 1980s, Cliff and Ord added models and applications of spatial processes with an example from epidemiology. They analysed the spatial pattern of cholera in London (Cliff and Ord, 1973, Cliff and Ord, 1981). Since then the literature on spatial statistical analysis has been expanded as well as their application (Diggle and Ribeiro Jr, 2007, Lawson, 2001, Lawson, 2009, Elliott et al., 2000, Waller and Gotway, 2004b, Cressie, 1993).

Spatial statistics focus on two spatial effects: spatial heterogeneity and autocorrelation. Spatial heterogeneity refers to the variability between entities and processes in space (Fotheringham and Rogerson, 1994). Spatial autocorrelation refers to the level of interdependence between those processes as well as the nature and strength of their interdependence (Cliff and Ord, 1973). The idea of spatial interdependence can be to some extent expressed by Tobler's Law of Geography: *"everything is related to everything else, but near things are more related than distant things" (Tobler, 1970).* Spatial autocorrelation tests should be carried out for the residuals of a model, as the presence of such correlation amongst the residuals indicates inadequacy of the regression model (Maguire et al., 2005). In my analysis, I used spatial autocorrelation tests to examine the residuals of the final statistical model for spatial interdependence.

Accounting for spatial autocorrelation allows more realistic inferences to be made, although due to the wide range of spatial models the selection of an appropriate model is often not straightforward (Langford, 1994, Goldstein, 1995). Some spatial models assess the variation of spatial patterns over time and space. These are called spatio-temporal models and are considered an improvement of simple spatial models. Various techniques have been applied for spatio-temporal analysis of mortality rates, ranging from log-linear regression (Cogdon, 1994), to more complex techniques such as thin plate splines (van der Linde et al., 1995), and more recently to mixed effects or random effects models. Mixed effects models take into account fixed and random effects related to the phenomenon under study that in turn allows accounting for complex spatio-temporal structures (Gelman, 1995, Gelman et al., 2009, Pinheiro and Bates, 2000).

Mixed-effects models can be implemented either in a frequentist or Bayesian context (Xia and Carlin, 1998, Richardson et al., 2006). Markov Chain Monte Carlo (MCMC) algorithms have revolutionized both approaches to frequentist and Bayesian statistical inference but primarily Bayesian (Maritz and Lwin, 1989, Carlin

and Louis, 1996). Bayesian models have been used for assessing paediatric asthma hospitalizations in relation to traffic density in San Diego (Zhu et al., 2000) as well as ambient O_3 concentrations and paediatric asthma emergency room visits in Atlanta (Zhu et al., 2003).

1.5.3 Spatial Analysis of Health Data and Policy Making

In May 2007, a Directive establishing an Infrastructure for Spatial Information in Europe (INSPIRE) was published and came into effect on 31 December 2009. One of the themes of INSPIRE is Human Health. INSPIRE does not intend to initiate a programme of new spatial data collection but instead "it is designed to optimise the scope of exploiting the data that are already available" (Commission of the European Communities, 2004). One of the most important Directives that INSPIRE complements is the Directive on the Re-use of Public Sector Information. The Directive on Re-use of Public Sector Information applies amongst others to NHS and local authorities and was implemented in the UK on 1st July 2005 (Office of Public Sector Information, 2005).

The INSPIRE directive complements other important policies, such as the Global Monitoring for Environment and Security (GMES) Directive as well as the European Environment and Health Strategy. GMES directive will bring an understanding of environmental factors potentially having adverse health effects and identifies air pollution as a major environmental health problem (Commission of the European Communities, 2004). In order to increase knowledge on those issues it is recognised that links should be established between environmental, geographical and health data (Commission of the European Communities, 2004).

In addition, the European Environmental and Health Action Strategy proposes the development of tools that link spatial health and environmental data, picturing of the demography and exposure patterns contributing to adverse health effects. Moreover it identifies asthma as one of the six priority diseases for which further research is required. This policy, complemented by INSPIRE could, for example, improve the identification of those at risk of respiratory diseases and target measures to reduce those risks (INSIPRE Framework definition support working group, 2003).

A preliminary estimation of the benefits of INSPIRE shows that one of the greater benefits will be the development of policies that reduce the impact of environmental pollution on health (INSIPRE Framework definition support working group, 2003). A pan-government initiative to improve the sharing and re-use of public sector

location information and implementation of INSPIRE, called UK Location, was established in UK (DEFRA, 2010). This study utilises existing spatial data as well as re-using information from the NHS and Local Authorities, to provide results on respiratory health outcome in primary care.

Chapter 2. Data Collection and Exploration

This is a study of ecological design at primary health care level, accounting for both temporal and spatial variability. The study area comprised the area of Newcastle-upon-Tyne and North Tyneside Primary Care Trusts (PCTs), in the Northeast of England, for the period of January 2002 until July 2006 (Figure 2-1). PCTs are local organisations and their role is to commission health services. The recent reform of the NHS has set GP consortia the leading role for commissioning health services (Department of Health, 2010). Every UK resident has access to GP practice services regardless of income or insurance cover. The total population under study was approximately 450,000 according to the 2001 Census (Newcastle: 260,000, North Tyneside: 90,000).

All data had to be subject to exploratory analysis, in order to be converted to the appropriate temporal and spatial format that would allow them to be used as variables in a statistical model. In the first section, I present the respiratory prescribing data as well as the patients' data and how they were used to define GP service areas. In the second section I present the air quality indicators, while the socioeconomic and demographic (age and sex) characteristic of patients registered per GP practice are presented in the third and forth section, respectively.

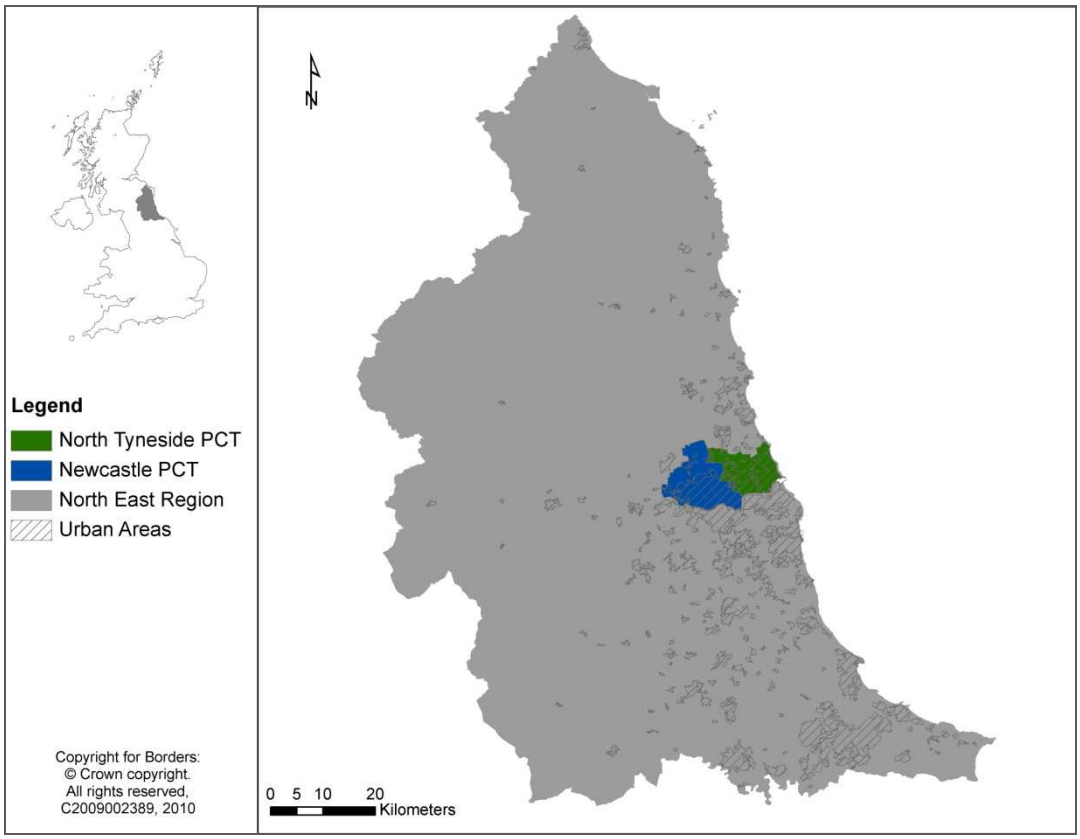


Figure 2-1. Study area

2.1 Health Data

I accessed respiratory prescribing data from the Regional Drugs and Therapeutics Centre (RDTC) and the patients' data per GP practice from the Exeter database, via the North East Public Health Observatory (NEPHO). In order to access both datasets, ethical approval was required by the PCTs' Caldicott Guardians. It took over a year to get approval for accessing the health data.

2.1.1 Respiratory Prescribing Data

The Newcastle RDTC holds prescribing data for England. They are based on prescriptions that are handed into pharmacies. These get forwarded to RDTC where they are added into a database. The data I accessed are not linked to individual patients but only to their prescribing GP practices. The prescribing data were available to me aggregated per month. Prescribing data were retained on the prescribing system for 60 months. The asthma prescribing data were acquired on 22/12/2006 for the period: 01/11/2001–31/10/2006. In order to link the prescribing data to the other datasets of interest, I removed two months from the

beginning and three months from the end of the dataset, analysing a total of 55 (01/01/2002 – 31/07/2006) months of salbutamol prescribing data. The prescribing data for this period covered approximately 67 billion Average Daily Quantities (ADQs) (Prescribing Support Unit, 2000). ADQ is a unit of measurement developed in order to calculate drug volume usage more accurately than the prescription item (Walley and Roberts, 2000).

Consulting a respiratory physician and a GP as well as studies on asthma and COPD pharmacological interventions (British National Formulary, 2009), I decided to access data for short-acting β2-agonists as an indicator of asthma and COPD exacerbations (British Thoracic Society, 2006). Short-acting β2-agonist inhalers are quick-relief medication, available as salbutamol, salmeterol, turbetaline and formoterol (British National Formulary, 2009). Salbutamol was employed in this study, as it is the drug most frequently prescribed, accounting for 93% of all short-acting β2 agonists prescriptions (Prescription Pricing Division, 2008).

2.1.1.1 Exploring Prescribing Data

The WHO produces the Defined Daily Dose (DDD) values for measuring prescribing volume, however DDD included data from many countries outside Europe, making these values less appropriate for measuring drug volume usage in a European country (Whiteside et al., 2001). The measurement unit ADQ was developed by an expert group in 1995 to be used for the analysis of English prescribing data (Prescribing Support Unit, 2005). ADQ is designed to be used as a numerator or denominator in studying variations in prescribing between general practices or primary care groups. All salbutamol data were standardised by total number of people registered with each GP before the analysis, therefore the unit of analysis was ADQs per 1,000 population.

The inhaler containing salbutamol medication is prescribed in a standard size, in England. Each inhaler contains 200 inhalations, also called doses or puffs, which is equivalent to 100mcg. The maximum suggested daily dose for adults is 8 puffs. This is important as all patients have the same amount of medication to consume, therefore the period between replenishing an inhaler is affected only by the frequency of medication use. I created a conceptual model on three simple scenarios of inhaler usage. The scenarios I have created are a coarse representations of reality as inhaler consumption is determined by several factors that often vary daily. Nonetheless, these scenarios can provide a broad estimate of the period in which patients would consume most of their inhaler and then seek to replenish it. The first scenario assumes that 8 doses (the maximum suggested

doses) are consumed daily, while 5 and 2 doses are assumed on the second and third scenarios, respectively (Table 2-1). According to the first scenario, a patient can consume most of the inhaler in 21 days (three weeks). A less frequent use represented by the second scenario showed that 30 days (four weeks) is the time that most of the medication would be consumed, while the last scenario of infrequent use showed that the medication could last for almost two months.

I wanted to check how frequently one can consume the daily maximum suggested 8-doses of salbutamol. After communication with a patient who suffered from severe chronic asthma, I found that salbutamol may also be prescribed "as needed", allowing consumption of more than 8 doses per day (Appendix B). In addition, this patient consumed double the maximum suggested daily dose for a few months (Personal communication with Alison Copeland, May 2010), which led to a much shorter period (7-14 days) before need for inhaler replenishment.

Scenario A : 8 doses - maximum suggested dose

Days of Inhaler Consumption	Dose	Total Dose	Proportion of Inhalers Consumed
58	8	464	2.3
51	8	408	2.0
44	8	352	1.8
37	8	296	1.5
30	8	240	1.2
21	8	168	**0.8**
14	8	112	0.6
7	8	56	0.3

Scenario B : 5 doses

Days of Inhaler Consumption	Dose	Total Dose	Proportion of Inhalers Consumed
58	5	290	1.5
51	5	255	1.3
44	5	220	1.1
37	5	185	0.9
30	5	150	**0.8**
21	5	105	0.5
14	5	70	0.4
7	5	35	0.2

Scenario C : 2 doses - minimum usual dose

Days of Inhaler Consumption	Dose	Total Dose	Proportion of Inhalers Consumed
58	3	174	0.9
51	3	153	**0.8**
44	3	132	0.7
37	3	111	0.6
30	3	90	0.5
21	3	63	0.3
14	3	42	0.2
7	3	21	0.1

Table 2-1. Scenarios on inhaler consumption

2.1.1.2 Discussion/Conclusions

Newcastle RDTC archives the prescribing data older than 5 years and it was not possible to easily access them after they had been archived. Accessing prescribing data for periods longer than 5 years where data exist, is important for

epidemiological studies which benefit from longer time trends. It is a great advantage to have prescribing data in electronic format for several years.

Not all countries can use prescribing data in epidemiological public health research because the population the prescriptions refer to is not always known. The availability of population based prescribing data depends on the health system of each country. For example, in United Kingdom and France it is possible to know the population that the prescribing refers to, while in other countries such as Germany this is not known and therefore it is not possible to use this as health indicator of a population.

The electronic prescribing data were well edited and organised, requiring little editing before analysing them. Finally, the RDTC database system did not allow a distinction between salbutamol prescribed to people suffering from asthma or COPD, however plans existed on how the database could also provide prescribing data per diagnosis, in the future.

The three salbutamol consumption scenarios I created are very basic, however they show that a patient would look to replenish his/her inhaler after a period of three weeks (21 days) to two months and probably more. In addition, after my communication with a patient who was prescribed salbutamol medication, I found evidence that more than 8-doses can be used daily, causing shorter replenishment periods (7-14 days).

2.1.2 Practice Population Data

Postcodes are an abbreviated form of address in the UK. On average there are 15 delivery points per postcode (Cabinet Office, 2009). The patients' postcodes per practice were provided for the 1st of April for each year of interest and assumed to remain the same for the rest of the year. The patients' data were unlinked to individuals and to prescribing data. I obtained over 2 million postcodes of patients for the period 2002-2006. I used an Access database and GIS to clean the datasets, explore and link them to prescribing data. The main issues I faced when editing the population data are presented below.

Outliers: I removed outliers from the patients' postcodes dataset. There were a few postcodes (approximately 10) that were not based within the Northeast Region (e.g. Wales). I considered it unlikely that those postcodes were correct patients' residence addresses. They were probably caused by data entry mistakes. I considered these as outliers and removed them from the dataset.

Datasets Linkage: I linked the GP population dataset to the prescribing dataset based on the matching GP practice codes. I was in contact with the North East Family Health Services Agency (2008), to check why some GP practices did not match between the two datasets. I went through the unmatched GP codes, in order to check that no one, who should have been included was missing. Some practice codes that appeared in the population data were irrelevant to asthma prescribing, as some practices specialized in specific health outcomes, such as sexual diseases. In another case a GP practice had ceased operation during the study period.

Missing Data: I discovered that some patients had no GP practice assigned to them. A cluster of missing data appeared in North Tyneside PCT. The North East Family Health Services Agency could not help me in tracing individuals' records; therefore I contacted the North Tyneside PCT directly. After providing information on the missing data and relevant maps, I found that the GP practice those patients were registered with, was the A86029 practice (Figure 2-2). This had been caused due to a database management error, during the data transfer between two government bodies.

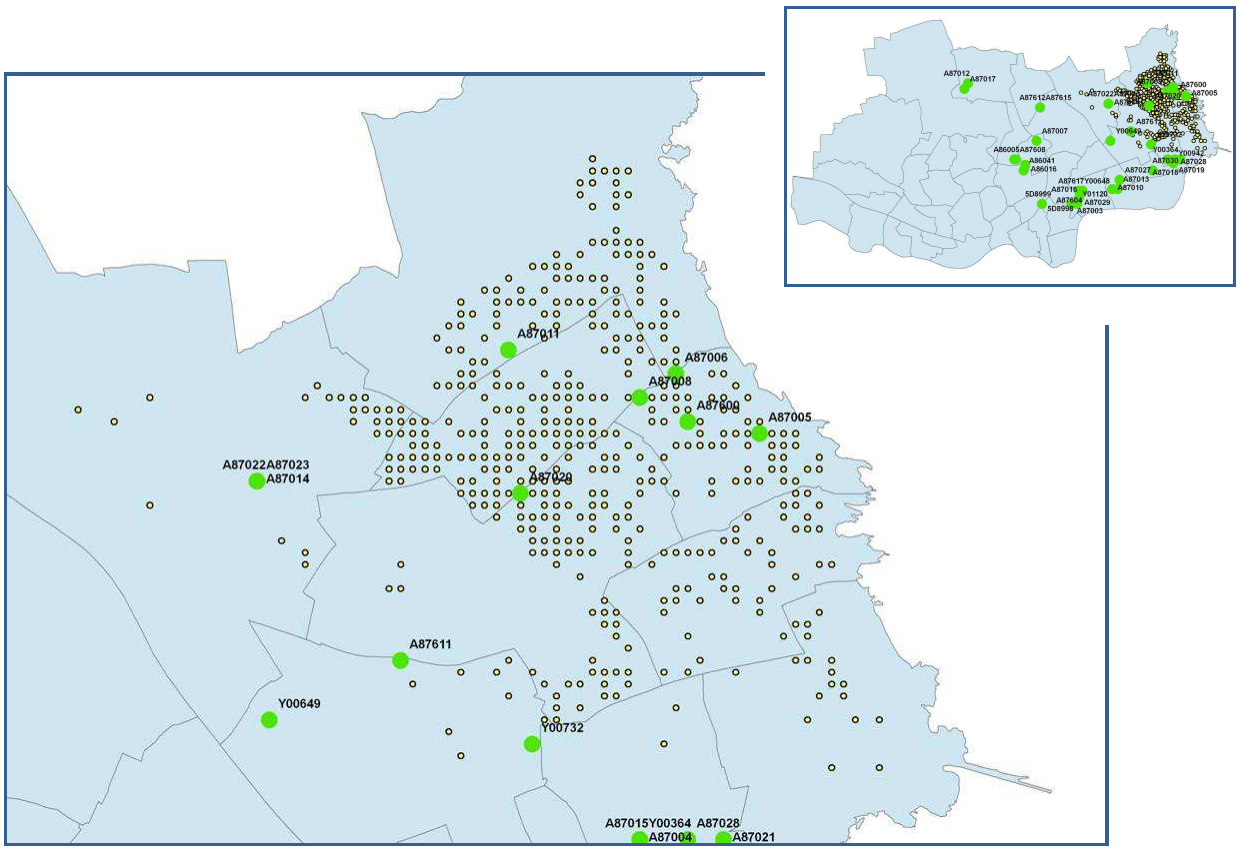


Figure 2-2. Postcodes of patients with no GP practice assigned to them

The annual numbers of patients per PCT are presented on Table 2-2, while a map of the GP practices is shown in Figure 2-2. A list of practices' names that correspond to each practice code is presented in Appendix C.

Year	**Newcastle PCT (35 GP practices)**	**North Tyneside PCT (29 GP practices)**
2002	230,464	173,123
2003	236,181	179,391
2004	244,023	185,314
2005	252,845	191,711
2006	264,614	199,545

Table 2-2. Number of patients per PCT per year.

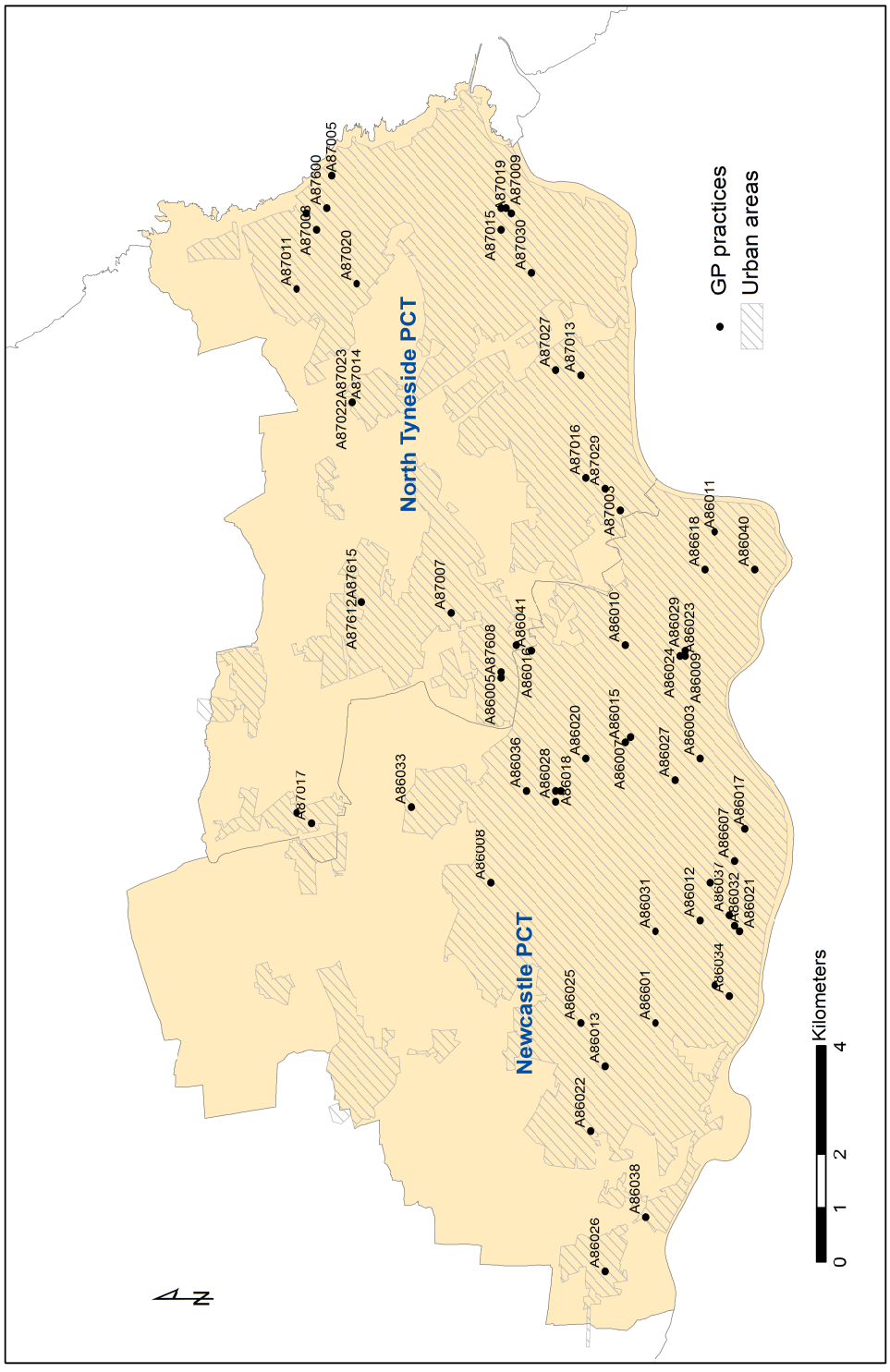


Figure 2-3. GP practices in Newcastle and North Tyneside PCTs.

2.1.2.1 Visual Exploration of Patients Data

Figure 2-4 shows the distribution of patients' postcodes that were mainly distributed within the study area. I plotted the same data also in grid format, in order to show the counts of postcodes in the study area.

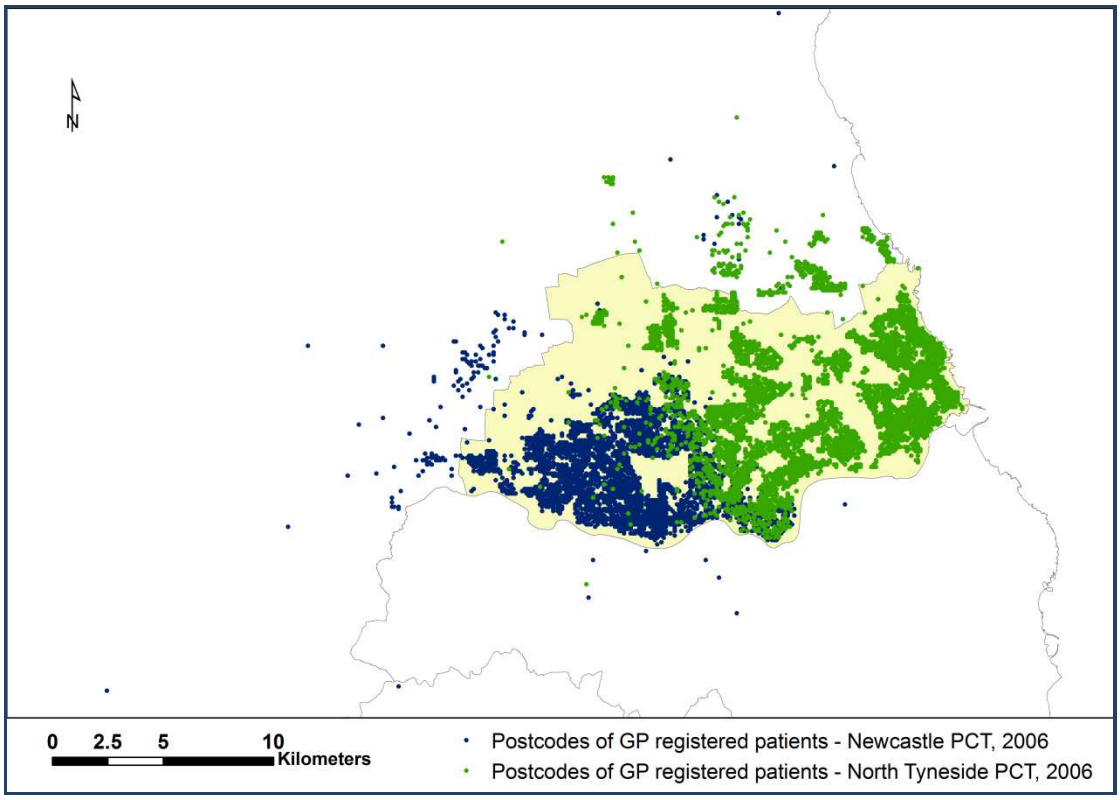


Figure 2-4. Postcodes of GP registered patients in the study area.

Figure 2-5 shows the number of postcodes for sub-regions of the study area, in grid format. We noted that some grids had thousands of postcodes, others had hundreds or tens, while a few postcodes were distributed sparsely in the periphery. Comparing the output of the grid plot to local knowledge of the area, revealed that the high numbers of postcodes coincided with the two main residential areas; Newcastle upon Tyne city centre and the area of Tynemouth, located at the eastern border of North Tyneside.

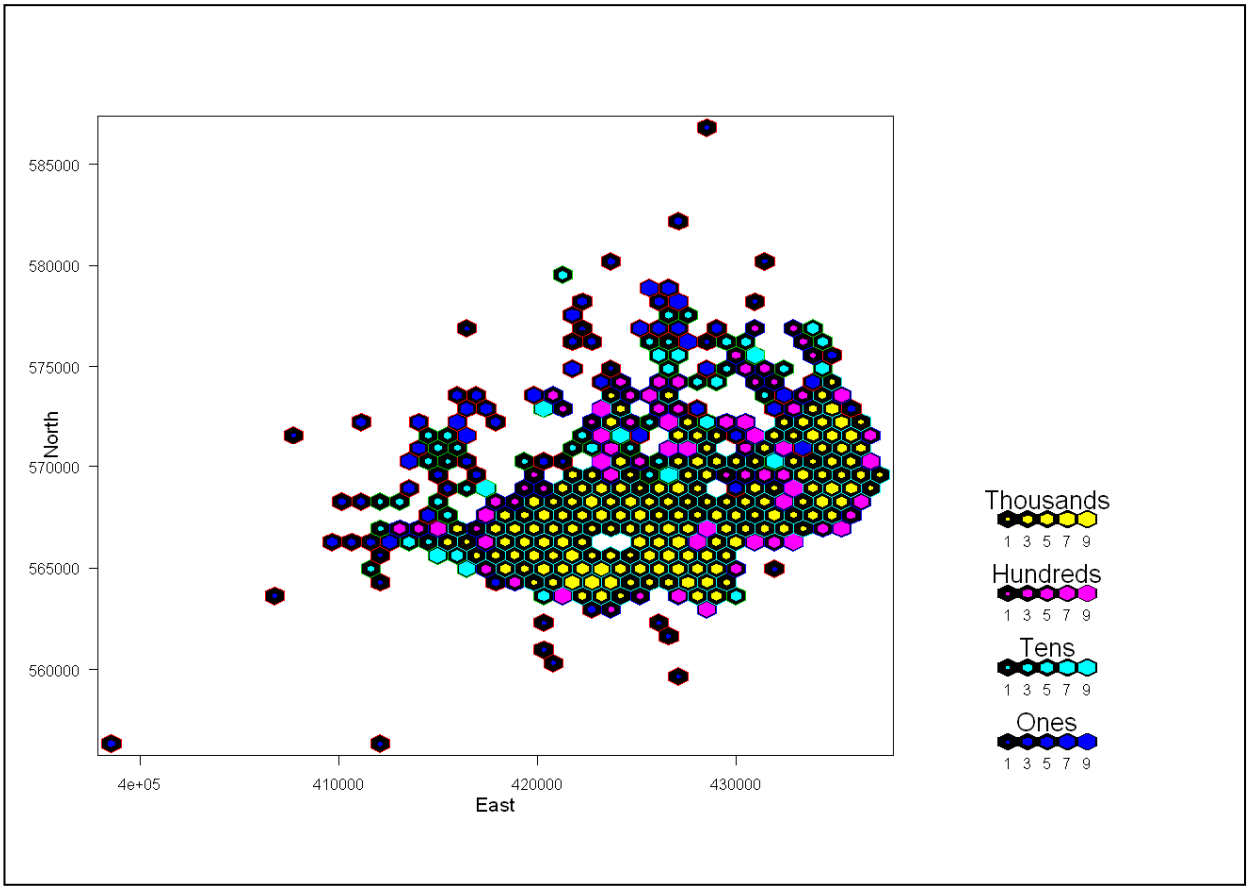


Figure 2-5. Grid format: Counts of patients' postcodes in sub-regions of the study area.

2.1.2.2 Spatial Exploration of Patients' Data

I started the analysis by separating the patients' postcodes by practice and drawing a Minimum Convex Polygon (MCP) for each one. Figure 2-6 depicts the service areas by GP practice for Newcastle and North Tyneside PCTs. It appeared that there was a tremendous amount of overlap between GP service areas within and between the two PCTs. Many GP practices appeared to serve very similar areas that covered the majority of the study area. Therefore, this method was not useful for capturing the GP service areas for the purpose of my further analysis. Some refinement was required to estimate service areas. As a next step, I used kernel analysis to define the area that each practice served.

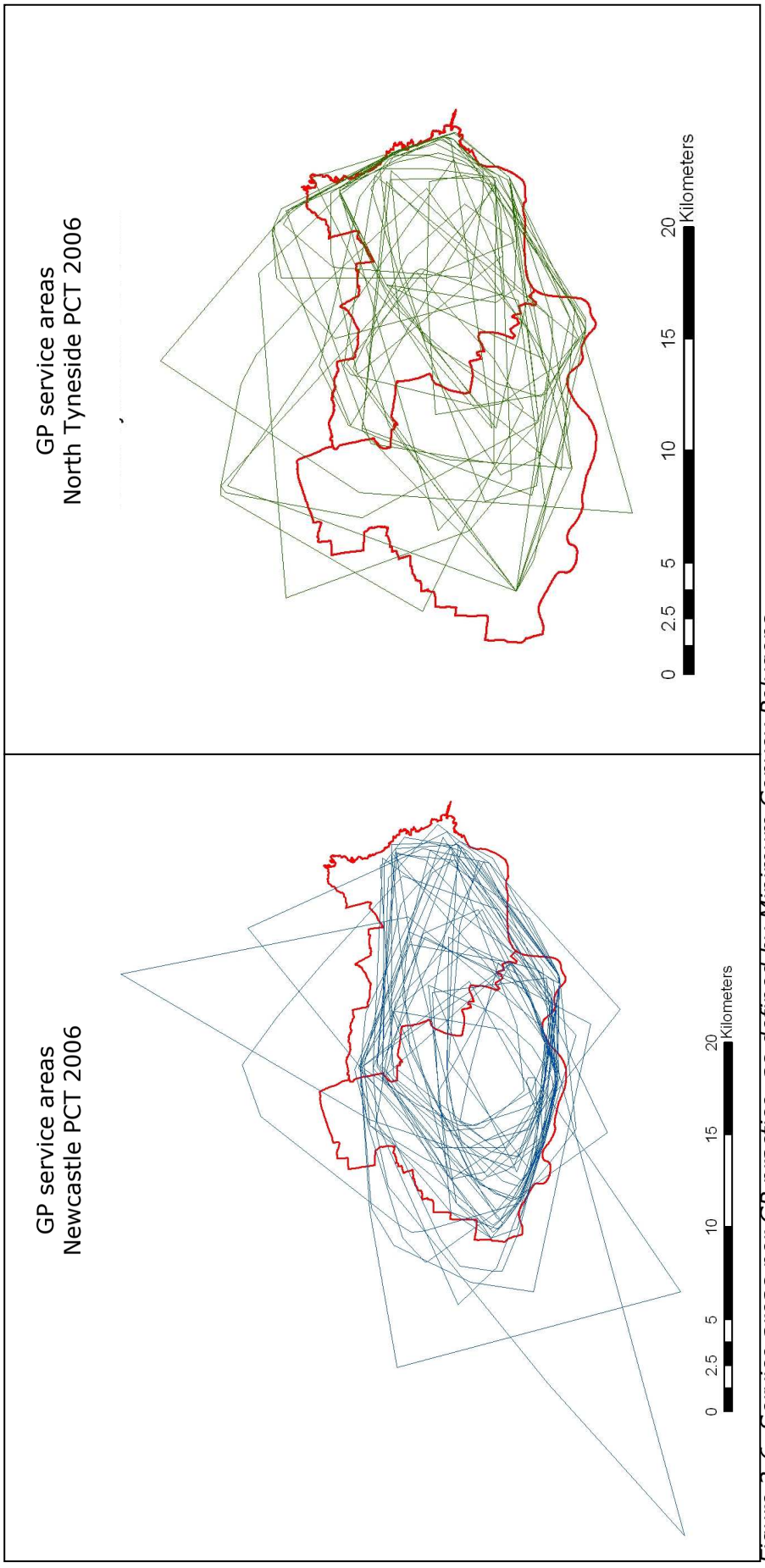


Figure 2-6. Service areas per GP practice, as defined by Minimum Convex Polygons.

Kernel estimation was developed to obtain a smooth estimate of a probability density from an observed sample of observations. The density of points is calculated using a bandwidth specified to a circle of a given radius centred at each point of interest. I used kernel analysis to estimate the patients density per GP and define a spatial unit that could depict the area where the majority of patients were expected to reside. There are various types of kernels, but the quartic kernel is the most widely used (Bailey and Gatrell, 1995). The formula for this is given as:

$$\lambda_{\tau}(s) = \sum_{h_i \leq \tau} \frac{3}{\pi}(1 - h_i^{\,2} / \tau^2) \qquad (1)$$

where, $\lambda\tau(s)$ is the intensity at the point of estimate s, h_i is the distance between the point s and the observed event location s_i and τ is the bandwidth which is sampled around point s. The region of influence within which observed events contribute to $\lambda\tau(s)$ is a circle of radius τ centred on s. The selection of the bandwidth is critical for kernel estimation, as it can cause over or under smoothing. Methods to estimate bandwidths are discussed in the relevant literature (Waller and Gotway, 2004a, Silverman, 1986, Scott, 1992). I used Equation 2 for calculating the bandwidth, as suggested by Bailey and Gatrell (1995):

$$\tau = 0{,}68n^{-0.2} \qquad (2)$$

where n is the number of spatial observations (postcodes of registered patients) per GP. The output of this equation was scaled up to fit the particular dimensions of the areas that the total n per GP was distributed in. Those areas were estimated by calculating the surface of each minimum convex polygon. The output values of kernel analysis produced continuous maps in raster format with a cell size of 25m.

I then used contour lines to define spatial units that represented GP service areas. Three contour lines were created per GP practice that contained 95%, 98% and 99% of the practice population, respectively. I conducted the analysis for these three cut off points to check how sensitive the analysis was to this factor. The contour lines cut-off points are produced automatically by most software, according to the values of the raster map, and it is not possible for the user to define cut-off points. For this analysis, I required consistent cut-off points to define the contour lines, so that service areas would be comparable per year and per GP practice. By adding an algorithm in ArcGIS I managed to define the cut-off for the contours' creation. The analysis of GP practice population data was conducted using ArcGIS

software (ESRI) with the Spatial Analyst and Hawth's Tool extensions (Beyer, 2004).

2.1.2.3 Results – GP Practice Service Areas

Kernel analysis took into account the density of patients' postcodes. Service areas were estimated for each of the 64 GPs for five years, creating in total 320 service areas. Because of the large number of GP practices I used seven GP practices to demonstrate the results of kernel analysis.

Figure 2-7 presents an example of a few raster maps produced by kernel analysis, showing the density of patients for seven GP practices in the area. In some cases a lot of patients were concentrated in a small area creating high values of intensity, while in other cases the intensity was less as the postcodes were distributed in a wider area. As a final step of the analysis, the borders of each GP practice area were calculated based on the values of kernel probability density distribution. Figure 2-8 shows the area that 98% of registered patients are expected to live in per GP practice. In Figure 2-9, I also present the area that all registered patients live per practice, in order to allow comparison between the areas drawn using the minimum convex polygons and kernel analysis.

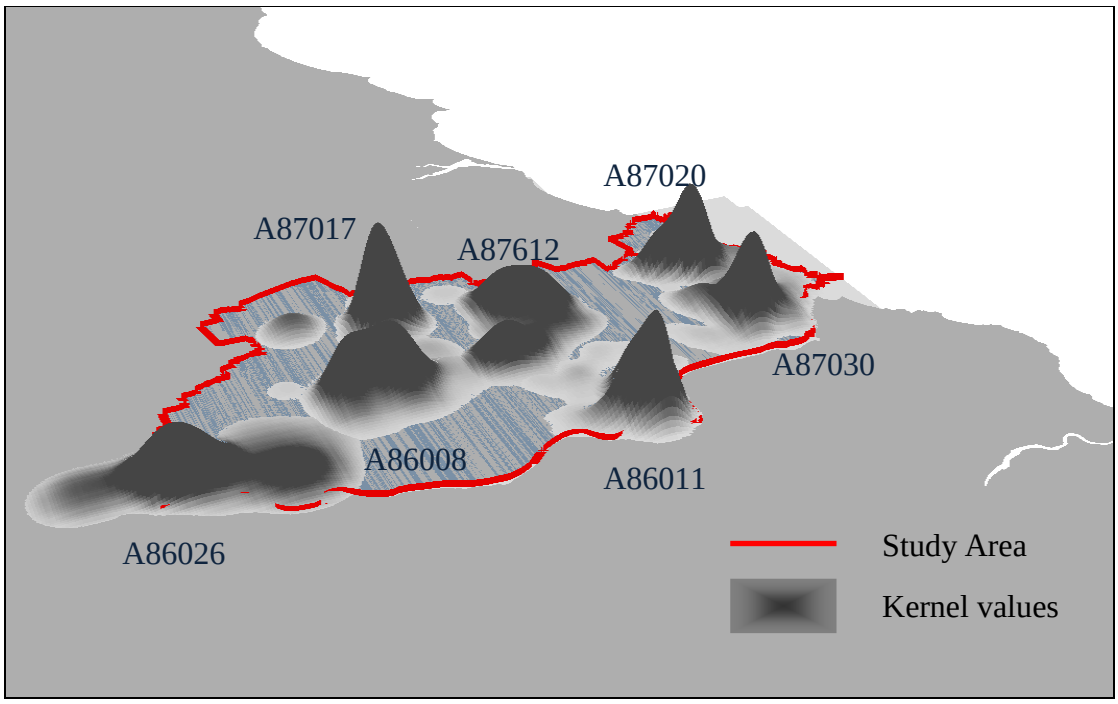


Figure 2-7. Raster maps produced by kernel analysis showing the density of patients per GP practice.

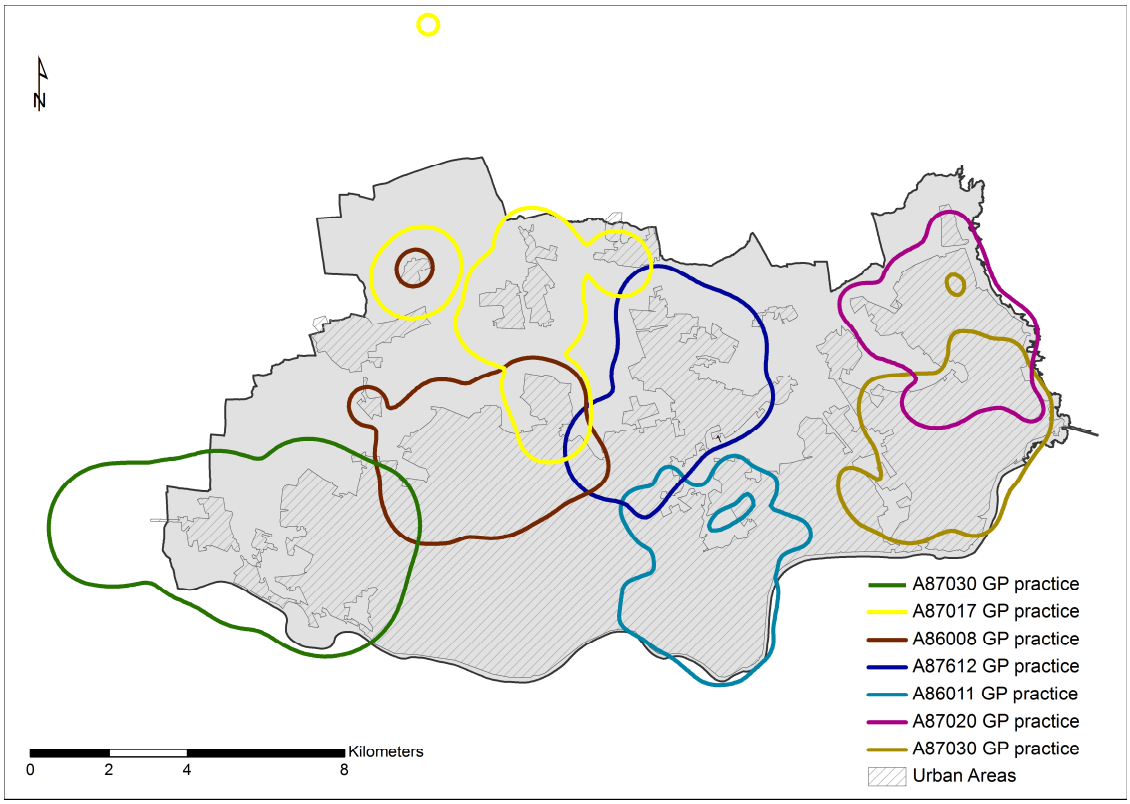


Figure 2-8 Contour lines representing the GP practice service areas.

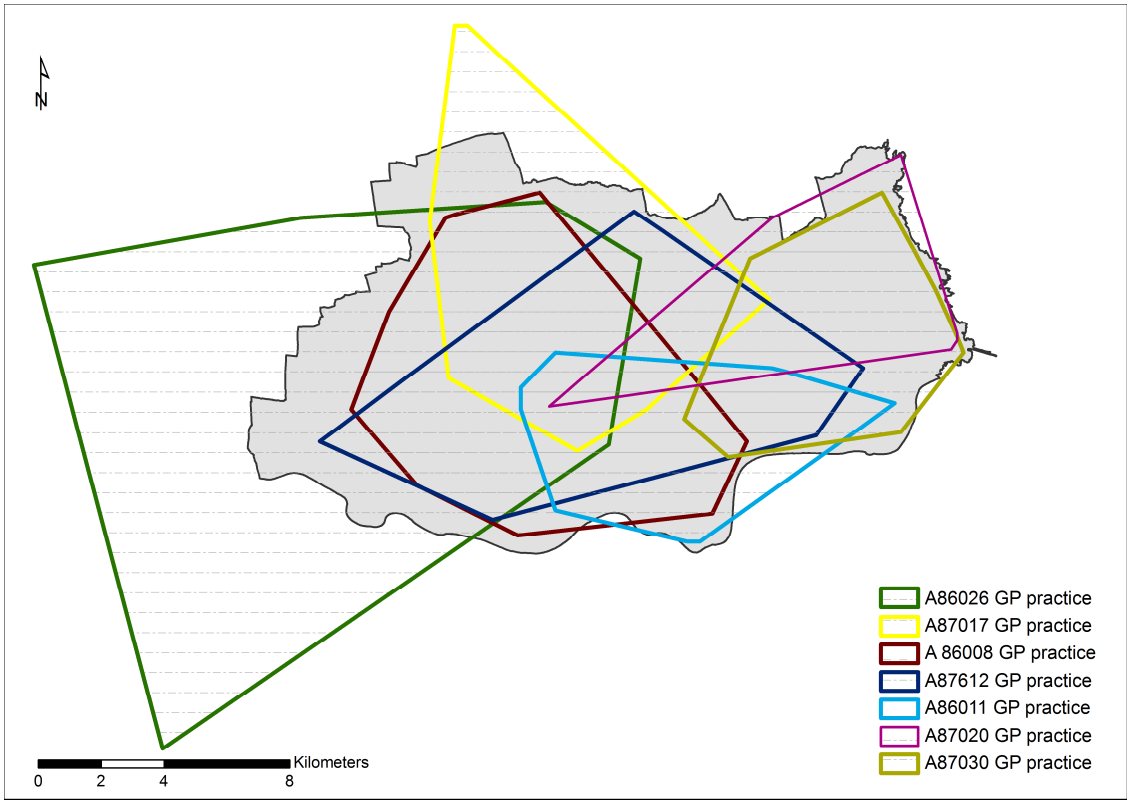


Figure 2-9 Minimum convex polygons showing the area that all registered patients live, per GP practice.

As mentioned earlier, I created three contour lines per GP practice that contained the 95%, 98% and 99% of the practice population, in order to check the sensitivity of the GP service area to different cut off points. Figure 2-10, Figure 2-11, Figure 2-12 and Figure 2-13 show examples of the contour lines that were produced as GP service area, for the 95%, 98% and 99% of kernel probability density distribution, respectively. Little variation was observed between the three cut off points. I decided to use the middle contour line that is the area that 98% of the patients are expected to live in, to define the GP service areas. I chose such a high value as the cut off point, as I wanted to remove areas that did not represent the exposure conditions of the majority of patients but at the same time I wanted to include the area in which the vast majority of registered patients lived.

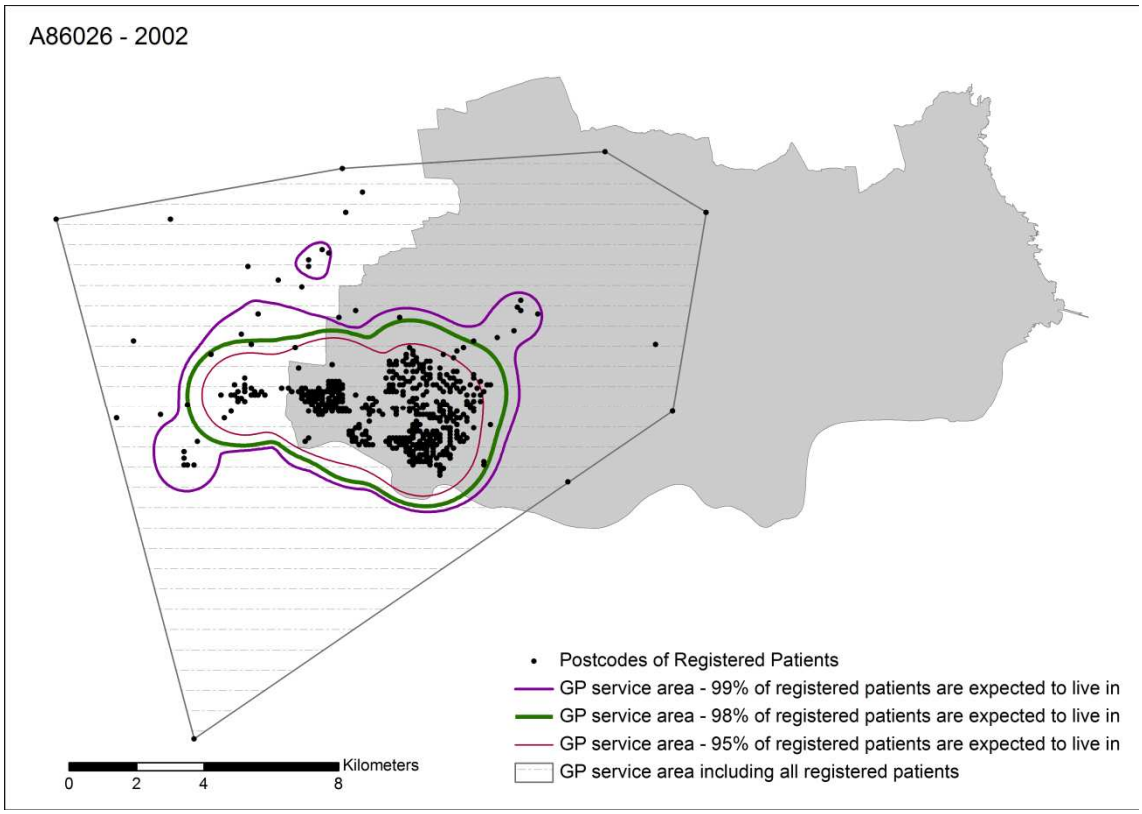


Figure 2-10 Examples of GP practice service areas.

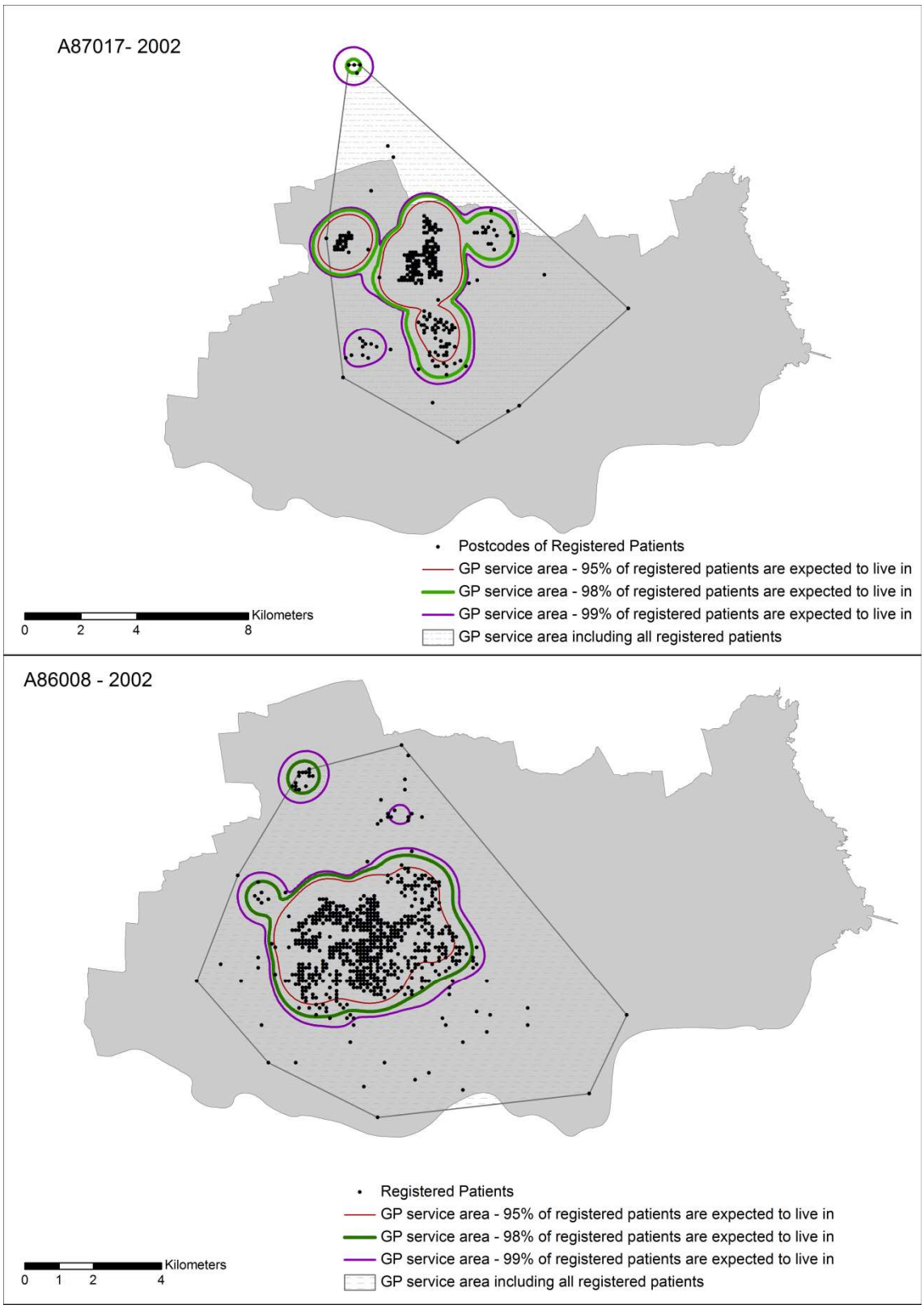


Figure 2-11 Examples of GP practice service areas

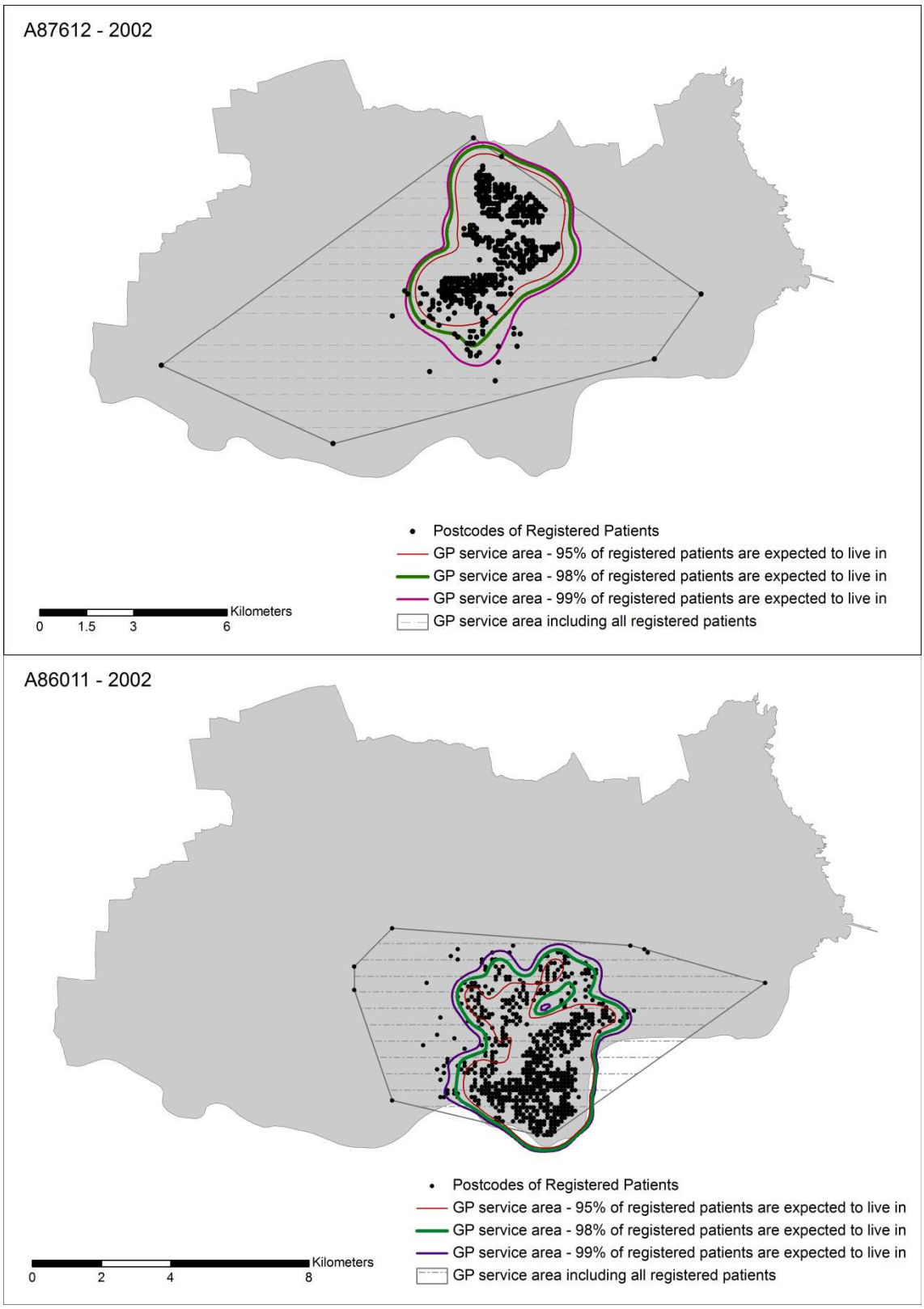


Figure 2-12 Examples of GP practice service areas

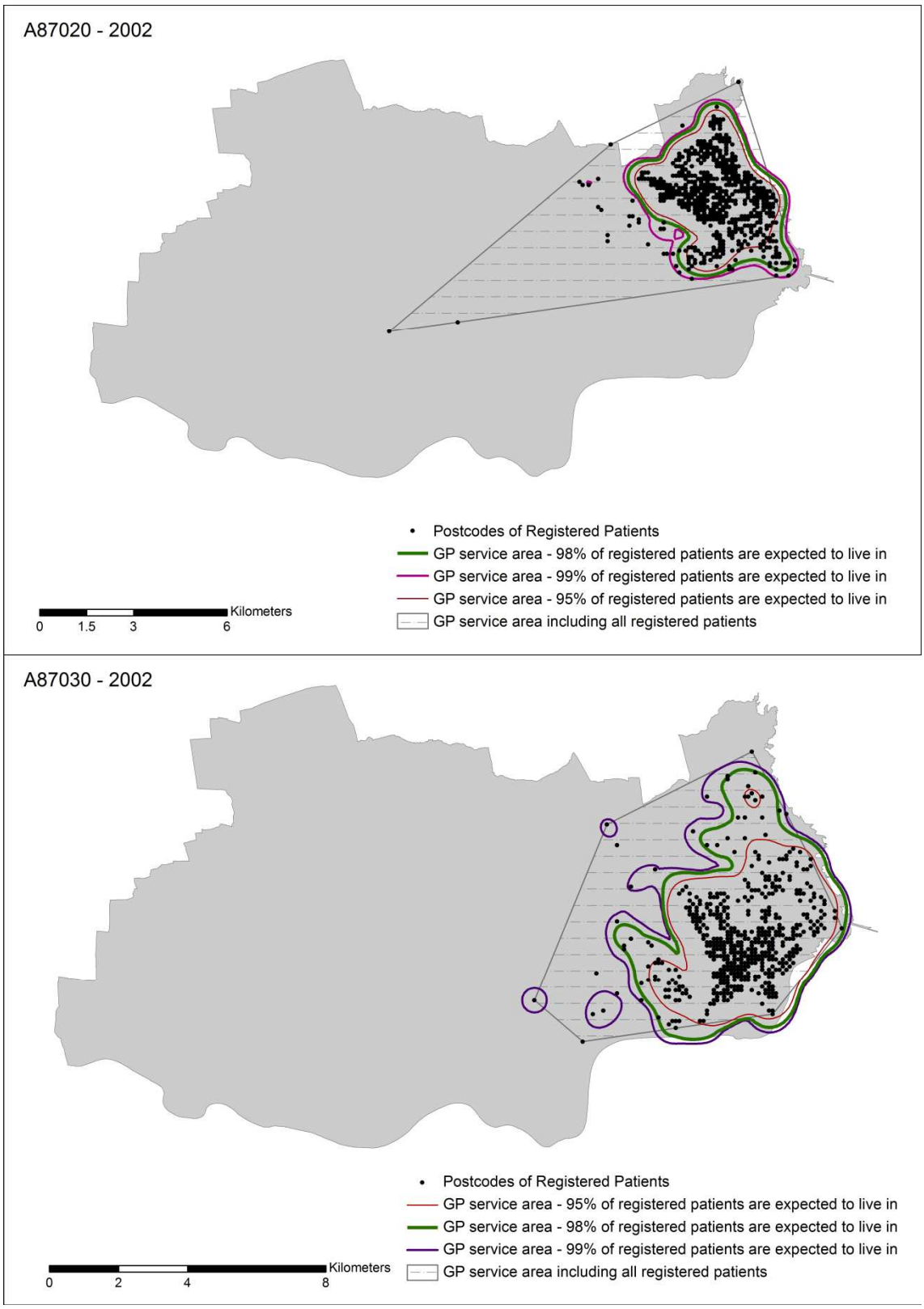


Figure 2-13 Examples of GP practice service areas

The service areas for all 64 GP practices, as summarised by kernel analysis, are presented in Appendix D. In Appendix E, I present the service areas for each of the five years of the study to demonstrate that very little variation existed in service areas through the years of the study.

2.1.2.4 Discussion/Conclusions

Results

Kernels have often been used in ecology to assess habitat selection and population dynamics and have been recently used in health research (Petersen J. et al., 2008, Lurz et al., 1997). Gatrell AC et al. (1996) have discussed the use of kernel density in geographical epidemiology and its application has been demonstrated in public health research recently (Shi X, 2009, Carlos et al., Petersen et al., 2009). Caution is required to define the bandwidth when applying kernel analysis. There is no unique equation to determine the selection of bandwidth, however there are a few proposed, like the one I employed. The responsibility lies on the researcher to examine whether the kernel output is a meaningful and representative summary of point observations. I considered the contour lines created based on kernel analysis to be representative summaries of the areas that registered patients were expected to live in.

Potential for Future Applications

Patients' postcodes are considered as sensitive data by the NHS and confidentiality rules apply to their storage and analysis. Kernel analysis allowed me to completely anonymise them by converting point data to area data and therefore removing consideration of individuals' identification. Ecologial studies that use aggregated rather than point data are useful in epidemiology, provided that their limitations are understood and their results are interpreted carefully.

The fact that GP catchment areas formed overlapping rather than adjacent polygons limited their spatial analysis. Kernel analysis reduced the degree of overlap allowing them to be used meaningfully for my study and as tools for surveillance. A variety of data that do not conform with administrative boundaries can be visualised within the spatial boundaries produced by this analysis, giving great scope to their future applicability. In the case of GP practices, using kernel estimates to define their service areas can be a practical tool to assist their new role as health services commissioners that was assigned to them under the recent fundamental NHS reform (Department of Health, 2010).

2.2 Air Quality Indicators

The association between salbutamol prescribing rate and air pollution was assessed in the regression model I developed. Ideally, individuals would be monitored on a regular basis using portable exposure equipment to derive estimates of their ambient or total exposure to environmental hazards, as those can provide greater and more significant effect estimates (Wilson, 2004, Ebelt et al., 2004). In reality, this is complex due to cost and practical difficulties, therefore proxies of human exposure to air pollution are often used in research (Jerrett and Finkelstein, 2005).

2.2.1 Ambient Concentration of Air Pollutants

In epidemiological studies, the ambient concentration of air pollutants is often utilized as a proxy for ambient or total exposure of populations. I used the ambient air pollution concentration as a surrogate to estimate the inhaled volume of air pollutants likely to occur.

I firstly identified monitors recording ambient air quality in my study area. There are over 1500 sites across the UK which monitor air quality and these are classed into two major categories: automatic and non-automatic networks. Automatic networks produce hourly pollutant concentrations from individual sites. Non-automatic Networks measure less frequently - either daily, weekly or monthly - and samples are collected by some physical means (such as diffusion tube or filter). These samples are then subjected to chemical analysis, and final pollutant concentrations are calculated from these results. The non-automatic monitors can be moved to more than one location.

In my study area there was only one automatic monitor that measured the five major air pollutants - PM_{10}, SO_2, NO_2, CO and O_3. This monitor was located in Newcastle city centre (Figure 2-14). I accessed the data recorded by the automatic monitor via the UK National Air Quality Archive website (UK National Air Quality Archive, 2010). I examined the availability of air pollution data recorded by the non-automatic network in the area. I also accessed these data from the Tyne and Wear Air Quality Information website (Tyne and Wear Air Quality Information, 2010) maintained by Sunderland University, however this source had only recent data and did not cover the initial years of my study period. Aiming to find previous records of these data I contacted the relevant Local Authorities but only few records were held for the early years of my study period. Unfortunately, the datasets from the non-automatic network were incomplete and I could not use them.

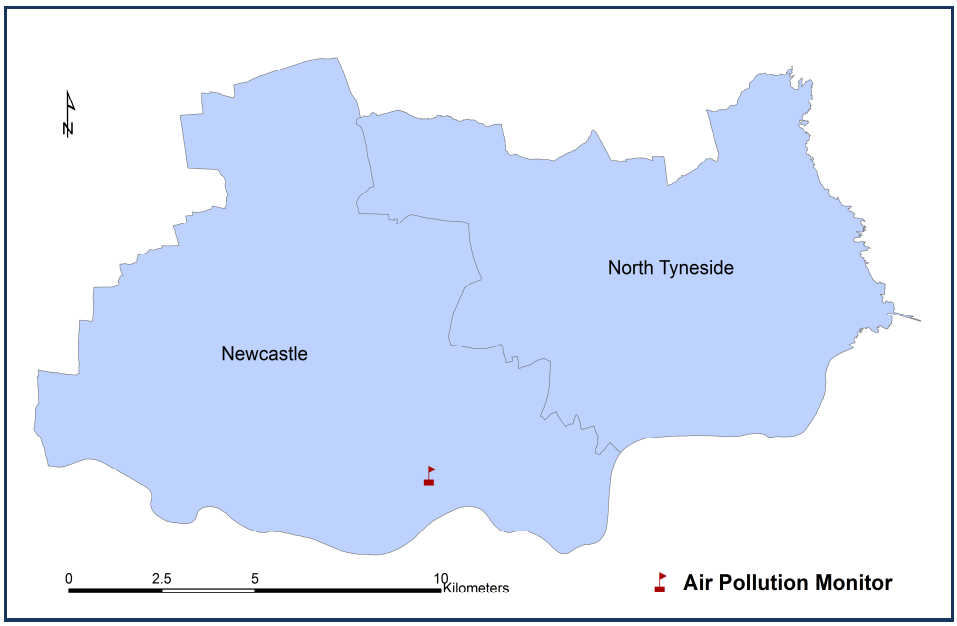


Figure 2-14. Air pollution monitor

I focused on deciding which and how many of the pollutants recorded by the automatic monitor in Newcastle, I should use. Levels of pollutant concentrations on the same monitor correlate, therefore it may be sufficient to use only one pollutant in regression for assessing the relationship of interest. I was primarily interested in PM_{10} and O_3 concentrations as indicators of ambient air pollution exposure. O_3 is long-range transport pollutant that does not only capture local air pollution emissions and is more likely to have similar concentrations through out my small study area. Primary particulate matter of small size also remains in suspension for some hours or days and travel considerable distances from the source. PM_{10} pollution episodes in UK have been attributed in UK sources but in other cases from sources outside the UK (Malcolm et al., 2000).

PM_{10} concentration levels are attributed in both long-range transport and in local traffic sources (ApSimon et al., 2000).The highest concentrations of PM_{10} are likely to occur in urban and industrialised environment and attention has been drawn to formation of secondary PM_{10}, particularly sulphate, and nitrate components resulting from oxidation of SO_2 and NOx emissions (CORINAIR - Core Inventory of Air Emissions Methodology, 1996). Secondary PM_{10} exhibits greater spatial variability than primary $PM_{10,}$ as it is mainly attributable to traffic. Air pollution concentrations linked to vehicular traffic pollution, are reduced significantly with distance from the road network. There are detailed emission inventories for the secondary particulate matter (CORINAIR - Core Inventory of Air Emissions Methodology, 1996) and local authorities have been responsible for controlling

particulate matter episodes. In late 1990s a few studies provided evidence on the contrary to the belief that the elevated levels of particulate matter in urban areas were largely related to traffic emissions, showing that rural measurements of PM_{10} become elevated at the same time as urban measurements during pollution episodes in the UK (King and Dorling, 1997, Stedman, 1997). This indicated that elevated levels of PM_{10} in urban centres were also attributable to long-range transport and originated from outside the local area.

Overall, given that only one monitoring site was available in the study area, PM_{10} and O_3 concentrations would provide a better exposure indicator, due to their partly long-range transport characteristics described on the previous paragraphs. The PM_{10} data had very few missing records while in the case of O_3 many months of data were missing. Based on data quality issues I used only the recordings of PM_{10}. The main concern associated with particulates is their potential effect on human health, notably the respiratory system, as particles of small size can be inhaled into and deposited in the respiratory system and remain there for long periods of time. Epidemiological studies have shown an impact of particles below 10μm (PM_{10}) on health. Future legislation in Europe and US is also focusing on monitoring of fine particles (i.e. $PM_{2.5,}$ PM_1) in addition to PM_{10}. However, colocated parallel measurements of PM_{10} and finer particles have revealed strong associations in their concentrations, and therefore the efficiency of their colocated monitoring is under question (Gehrig and Buchmann, 2003).

The PM_{10} data were accessed as 24h mean daily data. Figure 2-15 shows the daily PM_{10} concentrations recorded by the monitor located in Newcastle. I averaged the daily means to monthly values in order to assess its relationship to monthly salbutamol prescribing. I evaluated possible time lags between salbutamol prescribing and PM_{10} concentrations, therefore monthly averages of PM_{10} recorded 7, 14, 21 and 30 days preceding the month of prescribing were calculated.

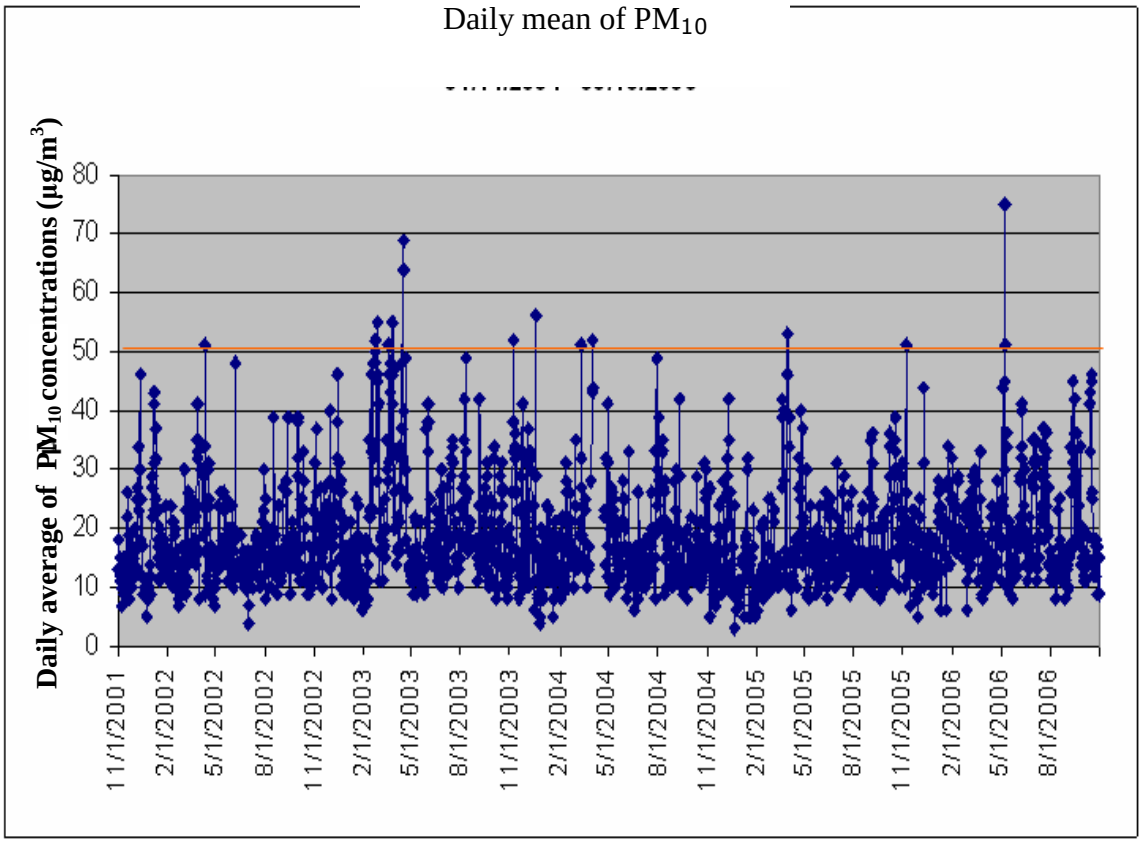


Figure 2-15. 24h-mean daily PM_{10} ambient concentrations.

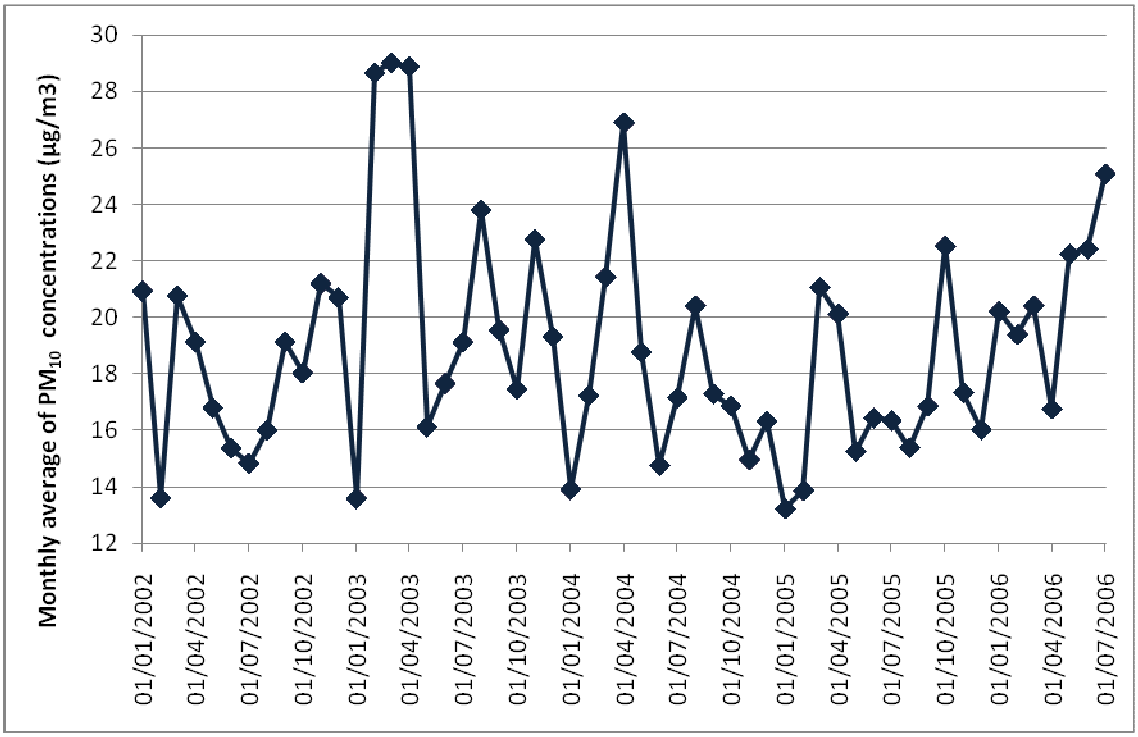


Figure 2-16 Monthly averaged PM_{10} ambient concentrations.

2.2.1.1 Discussion/Conclusions

The PM_{10} data were of good quality provided in fine temporal resolution but monitored only in one site. In an attempt to increase the spatial information on ambient PM_{10} concentrations, I also accessed PM_{10} data recorded by two non-automatic monitors; however those covered only the last two years of my 5 year study period. Therefore, I only used ambient air pollution data from one monitoring station for my analysis and I had to assume homogenous PM_{10} concentrations across my study area, like many previous similar studies (Pitard et al., 2004, Vegni et al., 2005, Zeghnoun et al., 1999, Katsouyanni et al., 2001). However, I knew that in addition to background air pollution there were local sources of pollution, possibly creating within city-variation of air quality which I would have liked to capture. Given that no ambient concentration data were available on a local level, I looked for data on sources of pollution. The study area is predominantly an urban environment, with traffic the main mobile pollution source. I present below how I used traffic data to develop an air quality indicator with a temporal and spatial domain.

2.2.2 Traffic Data

Road type has been used in several epidemiological studies as an indicator for exposure. "A" roads are usually related to higher levels of air pollution exposure while "B" roads and minor roads to lower levels. The road network was accessed via the Digimap website of the EDINA database. I downloaded A and B roads by Ordnance Survey Meridian 2 map, at 1:50,000 scale.

Traffic flow monitors recording the numbers of vehicles passing by individual sites 24 hours per day were also available for the study area. I accessed the traffic flow data from the Tyne and Wear Traffic and Accident Data Unit Transport Centre, based at Gateshead City Council. Monthly traffic flows were estimated, based on 7-day or 5-day averages. The latter excluded weekends and therefore provided average traffic flows only during working days. This measuring unit can better differentiate sites with increased/congested traffic at peak hours. Sites with congested traffic lead to higher exposures to traffic related air pollution, therefore I used the monthly flows based on 5-day averages.

2.2.2.1 Visual Exploration of Traffic Data

I accessed monthly traffic flow data for 50 monitoring sites which had complete data for my study period, located on A-roads and on a few B-roads (Figure 2-17). Details of the traffic monitoring sites are presented in Appendix F.

Figure 2-17 shows that the number of vehicles using some A and B roads could be similar and in some cases B roads had more traffic than A roads. In addition, great variability of traffic flows among A roads was observed. Due to the great variability of traffic flows, I considered that the traffic flow data could create a better proxy of traffic conditions than the length or type of road.

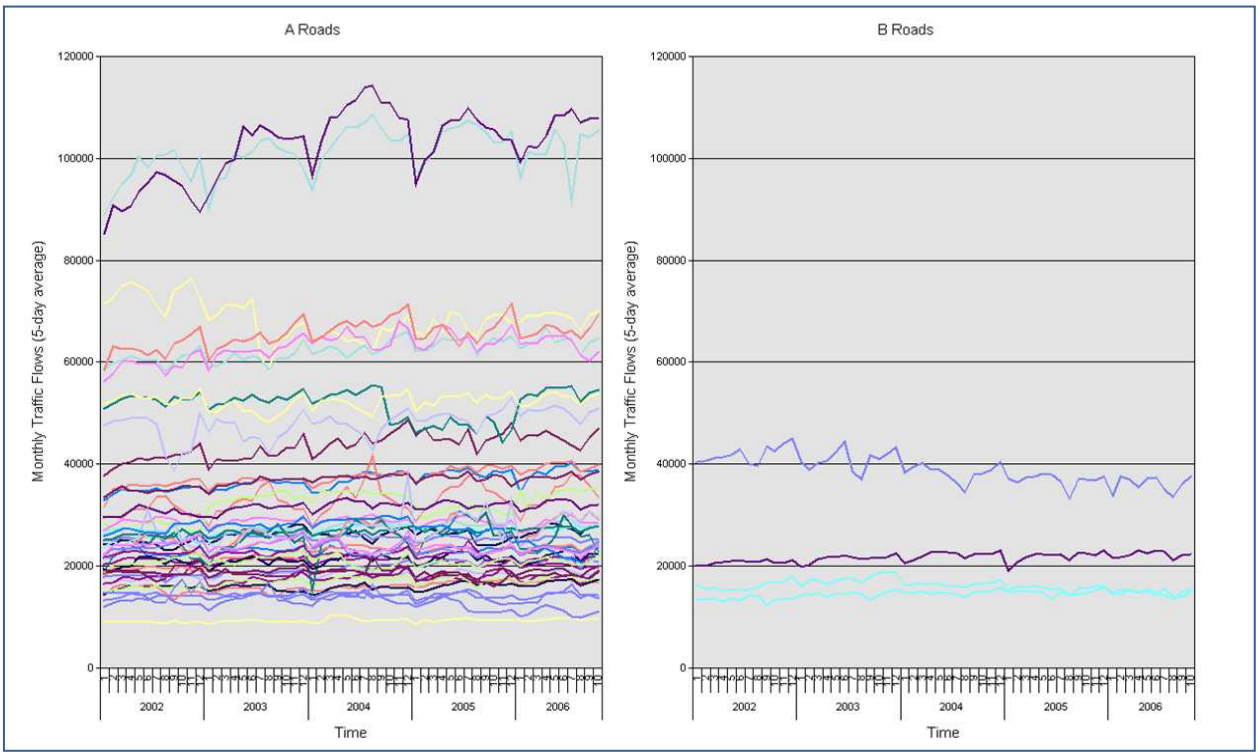


Figure 2-17. Traffic flows per monitoring site and type of road.

2.2.2.2 Spatial Exploration of Traffic Data

The traffic flow data were point data recorded in individual sites of the road network for each month of my study period (55 months). I interpolated them within the main road network, in order to estimate traffic flows in between the monitoring points. I employed a model called spline with barriers that produced a minimum curvature surface at the desired row and column space. The spline interpolation estimated values using "a series of simple functions, such as polynomials, which are fitted to successive groups of data points and constrained to give some degree of continuity at their joins" (Bailey and Gatrell, 1995). Splines are analogous to flexible rulers that pass through the points while minimising the total curvature of the surface, producing a smooth surface. The software ArcGIS with the Spatial Analyst extension was used to conduct the analysis of filling new grid nodes on a finer grid from a coarser grid. It is beyond the scope of this work to explain the algorithm but one can review the literature of the mathematical approach applied in

Refinement Griding (Terzopoulos, 1988, Zoraster, 2003, Smith W.H.F. and Wessel P., 1990).

I carried out the spline interpolation for each set of monthly traffic flows. To evaluate the output of the interpolation, I re-ran the spline model after removing one of the 50 monitoring sites and then comparing the predicted value to the original data on that site. This validation process was conducted for 10 randomly selected monitoring sites, which was 20% of the total traffic monitors. The spline output was also assessed in time by repeating the validation process for those 10 monitoring sites at 6 randomly selected months out of the 55 months, which was almost 10% of the total number of months.

As a final step, I linked the traffic flow maps to GP service areas (2.1.2.3) and summed the estimated traffic flows per service area, in order to create an index of traffic conditions per GP practice. I built an automated process within ArcGIS for executing the linkage of traffic flow maps to the GP practice service areas for each year, producing 3,520 traffic indices.

2.2.2.3 Results – Traffic Indices per GP Practice Service Area

Continuous maps of traffic flows within the main road network were created, for each one of the 55 months of the study period. Figure 2-18 presents a map of the interpolated traffic flows within the road network as well as the raw traffic flows recorded on the monitoring sites. The darker colour in the road network symbolizes areas with more traffic flows, while the lower traffic flows are represented by lighter shades. The monitors located in the south west of the road network are located on England's main motorway (A1) and recorded the highest number of traffic flows in the study area (around 100,000 vehicles monthly averages). I had concerns that the interpolation model may have over-predicted the traffic flows on the south-western part of the network. The prediction on this part of the network was influenced by the two sites with the highest recorded traffic flows. I had no monitoring data for the roads on the south-west to check the degree of overprediction, and since the predicted values fell within the range of plausible traffic flows, I accepted the model's output.

.

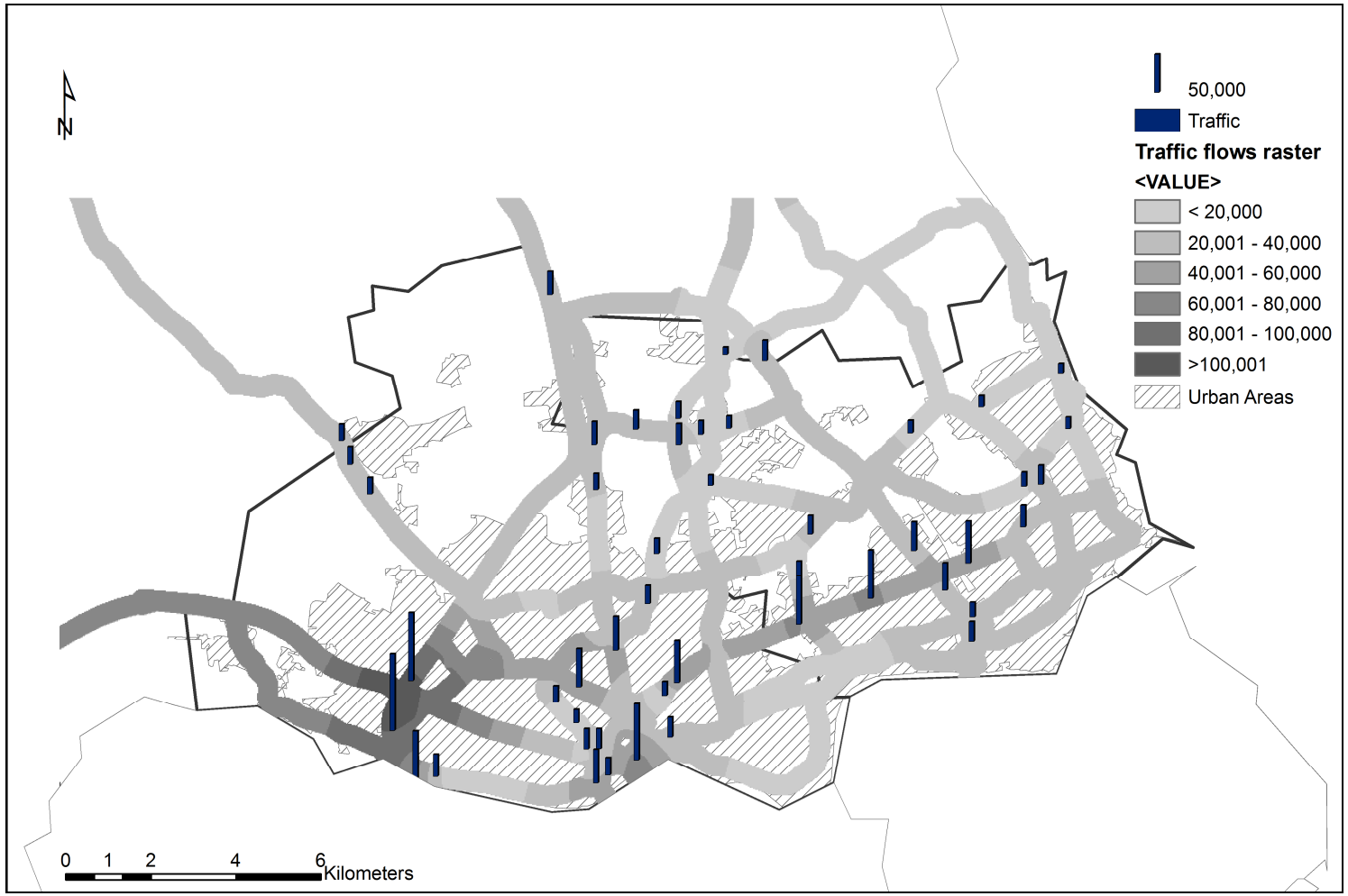


Figure 2-18 Traffic monitoring sites and estimated traffic flows within the main road network.

I validated the interpolation results by removing a monitoring site and after re-running the splines model, I compared the interpolated value to the observed. I used 10 randomly selected sites and repeated the process for six randomly selected months per site. The observed and predicted values on those ten sites at six randomly selected months correlated highly, with R^2 values ranging from 0.77 to 0.85 (Table 2-3).

Month	R^2
Mar-02	0.77
Sep-02	0.85
Oct-03	0.83
Dec-04	0.79
Jul-05	0.83
May-06	0.81

Table 2-3 R-squared values of observed traffic flows against estimated values.

I then summed the estimated traffic flows per service area, in order to create an index of traffic related pollution per GP practice. A summary of descriptive statistics is presented in Table 2-4. The traffic index values were divided to 1,000,000 to enhance practicality during the analysis. I present histogram of the traffic indices per GP service area in Figure 2-19. In Appendix G present the traffic index per month during the five years for each GP practice.

	Min.	1st Quartile	Median	Mean	3rd Quartile	Max.
Traffic Index	191,903,920	495,411,112	781,850,528	903,878,380	1,267,545,088	3,235,041,024

Table 2-4 Descriptive statistics for estimated traffic index

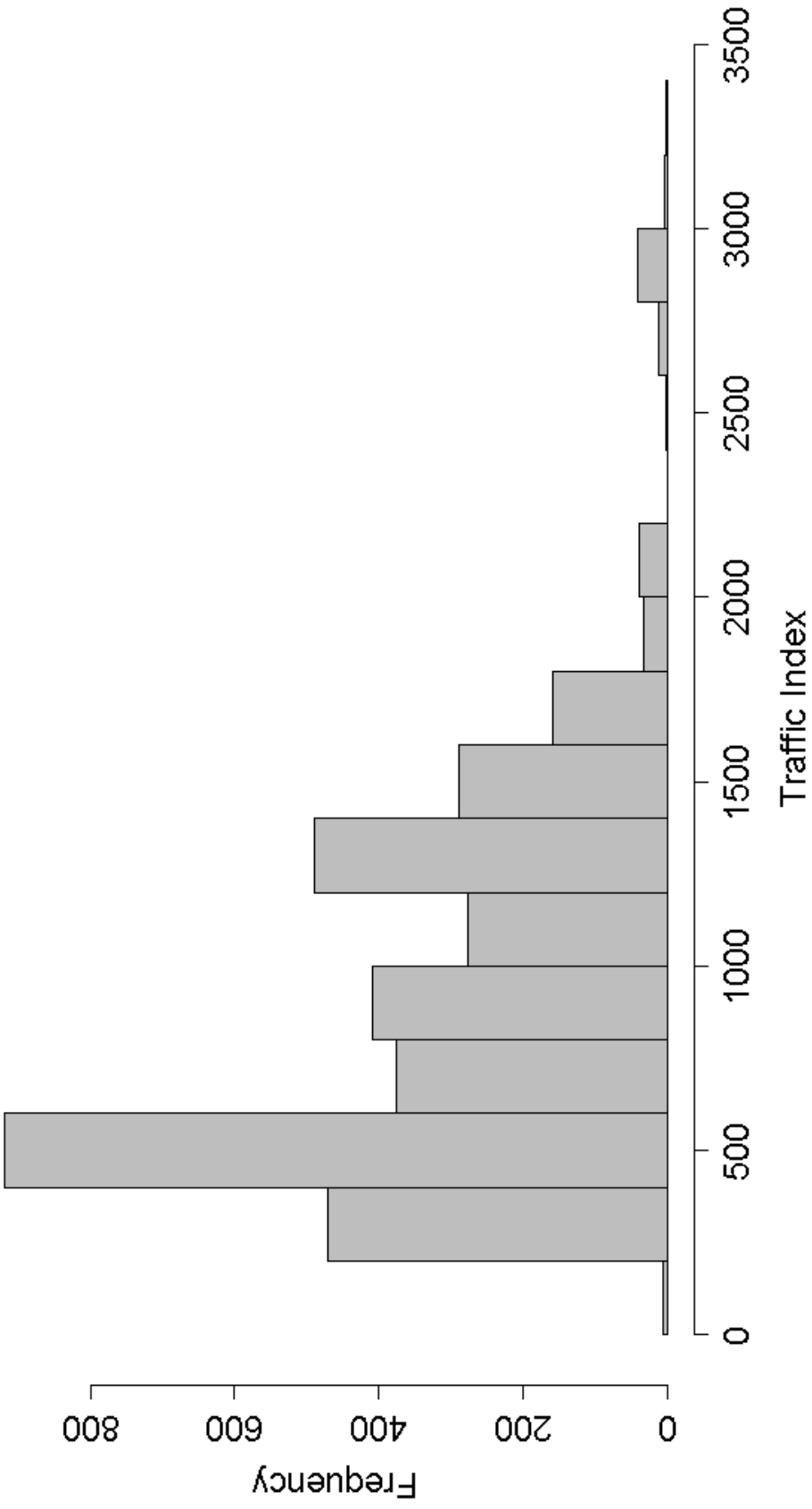


Figure 2-19 Histogram of estimated traffic index per GP service area

2.2.2.4 Discussion/Conclusions

Traffic Flow Interpolation

The splines with barriers technique has been previously applied in geology and geosciences (Zoraster, 2003) and I used it in a new context by applying it to traffic data. The interpolation of flows was used in order to create continuous maps of traffic flows that would in turn allow assessment of traffic within a variety of spatial units. The spline interpolation with barriers may have over-predicted the traffic flows in the south-western part of the road network. I could have checked that and perhaps alleviated the potential over-prediction, if I had accessed traffic data for the wider area surrounding my study area. This problem is also related to the fact that information about the direction of the traffic flows could not be included in the interpolation model. For instance, heavy traffic observed in monitors that are located on the motorway (at the south-western part of the road network), would not influence the traffic flows on A-roads, should the model allowed to control the direction of traffic flow along the motorway. To the best of my knowledge there is limited availability of interpolation techniques within lines that allow accounting for flow direction (Cressie et al., 2006).

Overall, I had no monitoring data for the roads on the south-west to check the degree of over-prediction, and since the predicted values fell within the range of plausible traffic flows, I accepted the model's output. The model performed well in predicting traffic flows based on the results of the validation process. The validation process showed that the observed and predicted values on ten randomly selected sites, at six randomly selected months, correlated highly with R^2 values ranging from 0.77 to 0.85.

Estimation of Traffic Index

In the absence of a network of air pollution monitors in the study area I created a traffic index, by summing the interpolated traffic flow per GP service area. I added the traffic flows values depicted at each 25m cells of the raster maps, thus the index does not represent actual numbers of cars that pass through the road network within each GP cathment area. The traffic index values are analogous to the number of cars as well as the length of road network within each GP catchment area. The sum of the interpolated traffic flows can cause bias in the index calculation. This is because a large GP service area with long road network but low volume of traffic, situated probably in the periphery of the study area, may has an index value close to that of a small GP service area with high traffic flows, situated

probably in the urban centre. I have not been able to measure the number of catchment areas that may have been affected by this measurement bias. However, I understand that its effect on the analysis is to smooth the differences between traffic flow variability among catchment areas.

In addition, the traffic index I constructed does not take into account the spatial distribution of patients within a GP service area. The continuous raster traffic maps could have been linked to raster maps created by kernel analysis or actual postcodes of registered patients, to create exposure metrics of traffic related air pollution. Such approaches would have the advantage of accounting for the non-homogenous distribution of patients within a GP service area or any other population under study. Continuous traffic flow maps allow assessment of traffic conditions in a variety of spatial units, such as Local Authorities and Lower Super Output Areas.

Overall, given the absence of spatial air pollution data, I constructed an index of traffic conditions per GP practice, providing a proxy measure of traffic conditions' spatial variability within the study area. However, analyses that use data such as number of cars or number of cigarettes smoked, to derive proxies of exposure, give rise to misclassification of exposure. I consider such measures of exposure to be further from the true exposure concentrations, when compared to indicators that use ambient air pollutant concentrations. Therefore, the main exposure metric in this study is the homogenous monthly ambient PM_{10} air pollution concentrations, while the traffic index was constructed in an attempt to account at some extent for the spatial variability of air pollution within the study area. Future research would be to use mobile PM_{10} monitors to investigate the association between monitored PM_{10} and traffic flows.

2.3 Socioeconomic Data

The association between salbutamol prescribing rate and socio-economic conditions of the group of patients registered per GP practice was also assessed in the regression model. Socio-economic characteristics of an area expressed as indicators of deprivation are linked to susceptibility (Jerrett and Finkelstein, 2005, Yen and Syme, 1999, Pickett and Pearl, 2001). Using local areas' contextual data is an ecological approach to account for population susceptibility. It is expected that sub-populations living in areas characterised by high deprivation, will be more susceptible to respiratory disorders.

2.3.1 Index of Multiple Deprivation

I accessed socioeconomic data by the Index of Multiple Deprivation (IMD). IMD was published in 2000 for England and was provided by the Office for National Statistics. In 2004, an improved version was released, called IMD 2004 (Communities and Neighbourhoods, 2004). IMD 2004 was mainly based on data collected between 2002 and 2004. One of the key features of the updated version was that census data (Department of the Environment Transport and the Regions, 2000) were not used in the derivation of the index, which made it possible to update it more frequently than a decade. IMD 2007 was published in 2008 (Communities and Neighbourhoods, 2008), while IMD 2010 is expected to be released soon.

One key characteristic of IMD 2004 is that it is available for smaller spatial units, compared to IMD 2000. Those spatial units are called Super Output Areas (SOAs) and consist of similar numbers of residents. Parliamentary wards have been divided into Super Output Areas (SOA) by the Office for National Statistics, creating three layers, defined by spatial scale: lower, middle and upper.

2.3.1.1 Exploration of Index of Multiple Deprivation

I used IMD 2004 as it is relevant to the time period of my study. IMD 2004 consists of seven domains: 1) income, 2) employment, 3) education, skills & training, 4) health, 5) living environment, 6) crime, and 7) barriers to housing and service. I extracted data for the three main aspects of socio-economic deprivation; income, employment and educational deprivation. The index of income deprivation captures people living in families dependent on income benefits. The employment deprivation index measures the working age population characterized as

involuntarily excluded from work, while the index of educational deprivation captures the proportion of working age adults with no or low qualifications as well as pupils' attainment. The indices are measured using different scores (Government Office for the North East, 2007) and each was weighted differently in the construction of the IMD 2004. Table 2-5 presents the scales used for depicting the deprivation scores of interest.

Deprivation Indices (IMD 2004)	**Range of Deprivation Scores in England**	
	least deprived	most deprived
Income	0.00	0.96
Employment	0.00	0.69
Education, Skills & Training	0.03	99.22

Table 2-5 : Scale of measurement of Deprivation Scores (2004)

The deprivation data were accessed in the lower layer of SOA, with a mean population of 1,500. The borders of LSOAs are available in the Edina database. I created maps of income, employment and educational deprivation in the study area (Figure 2-20, Figure 2-21, Figure 2-22). The levels of deprivation have been categorised in quartiles (4-quantiles).

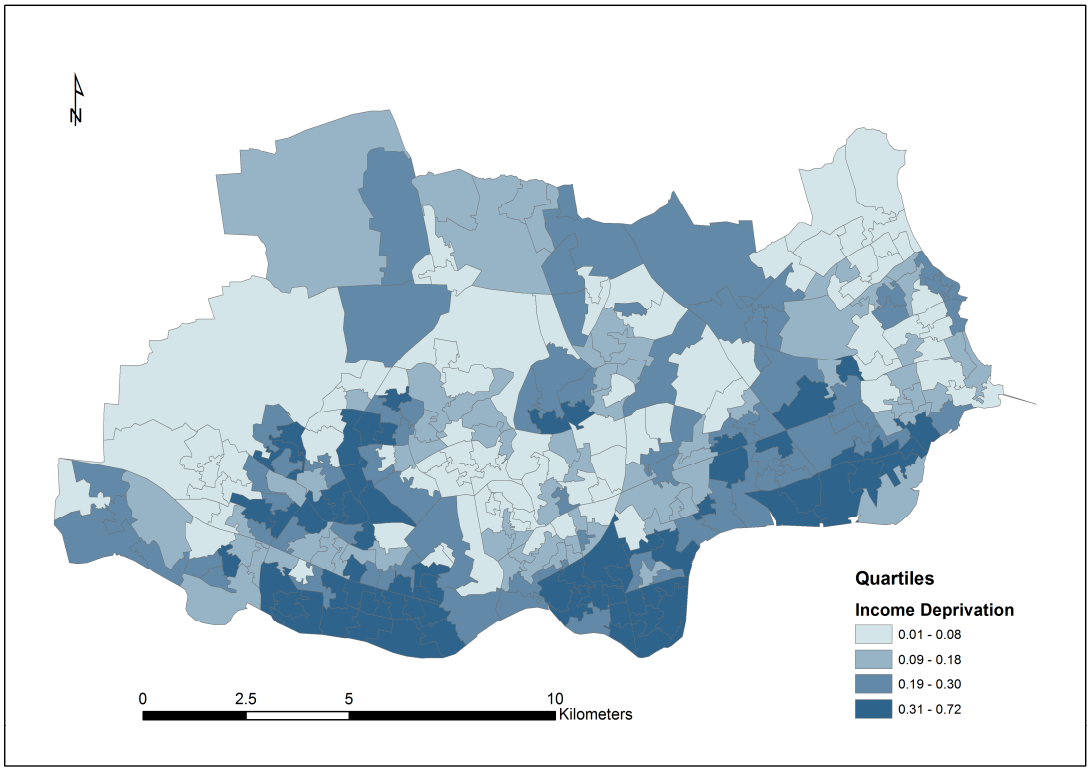


Figure 2-20. Income Deprivation

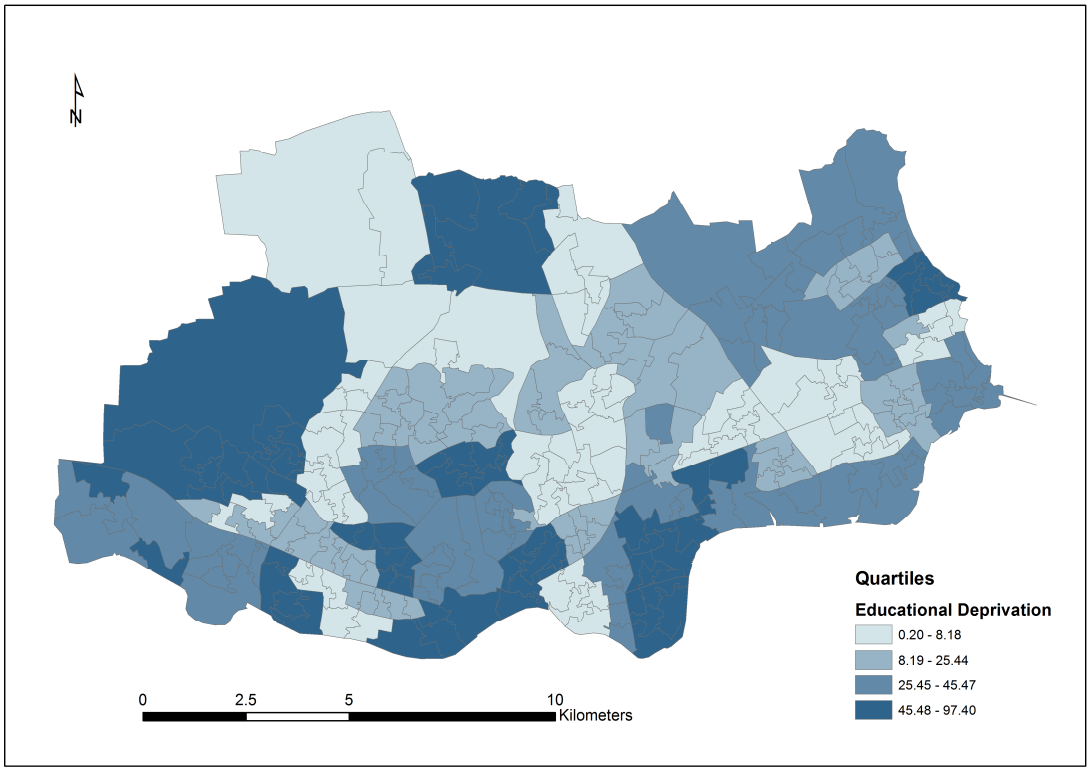


Figure 2-21. Educational Deprivation

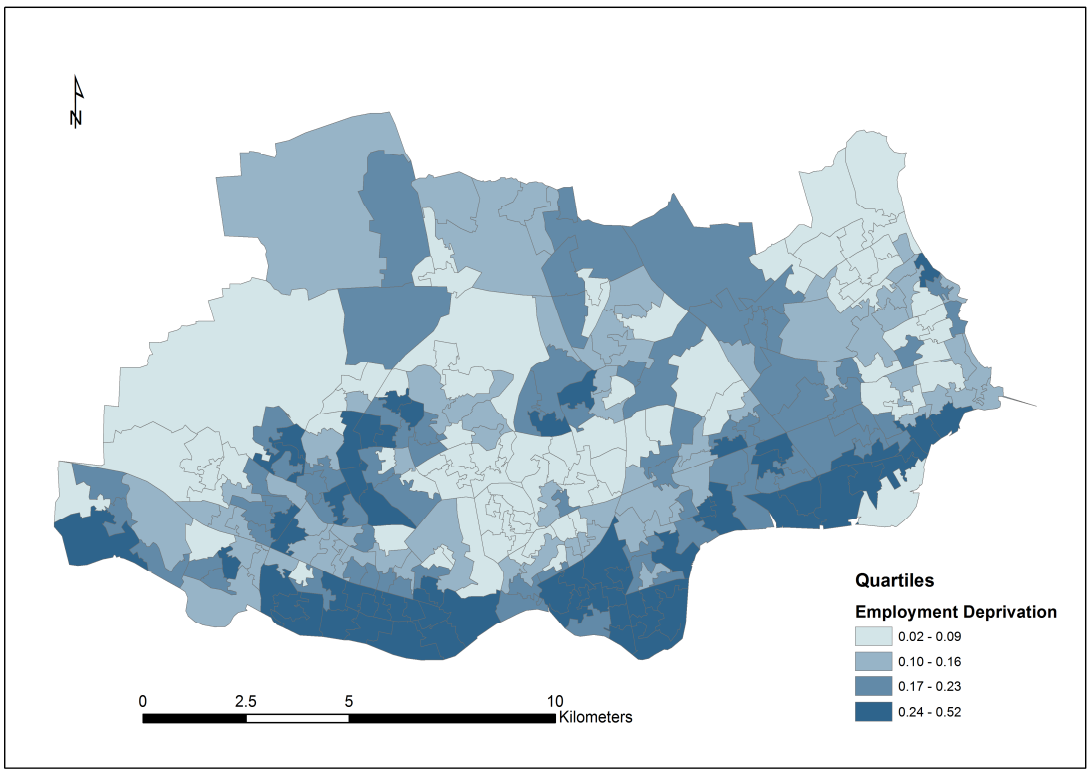


Figure 2-22. Employment Deprivation

2.3.1.2 Results - Practice Level Deprivation

I estimated a practice level deprivation score for each practice. This was derived by assigning each postcodes with the deprivation scores of the LSOA it fell within. Then, I averaged the deprivation scores over all postcodes affiliated to each practice. Three deprivation scores were estimated for each practice: 1) income, 2) employment and 3) education. Figure 2-23, Figure 2-24 and Figure 2-25 show histograms of the deprivation indices.

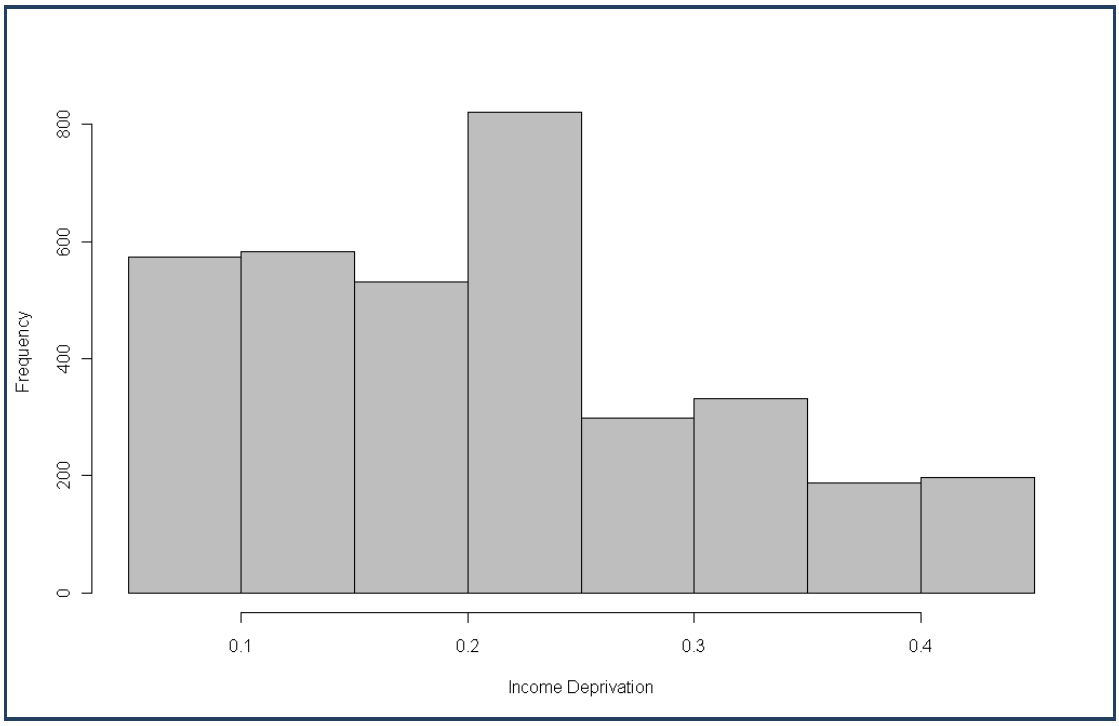


Figure 2-23. Histogram of Income Deprivation

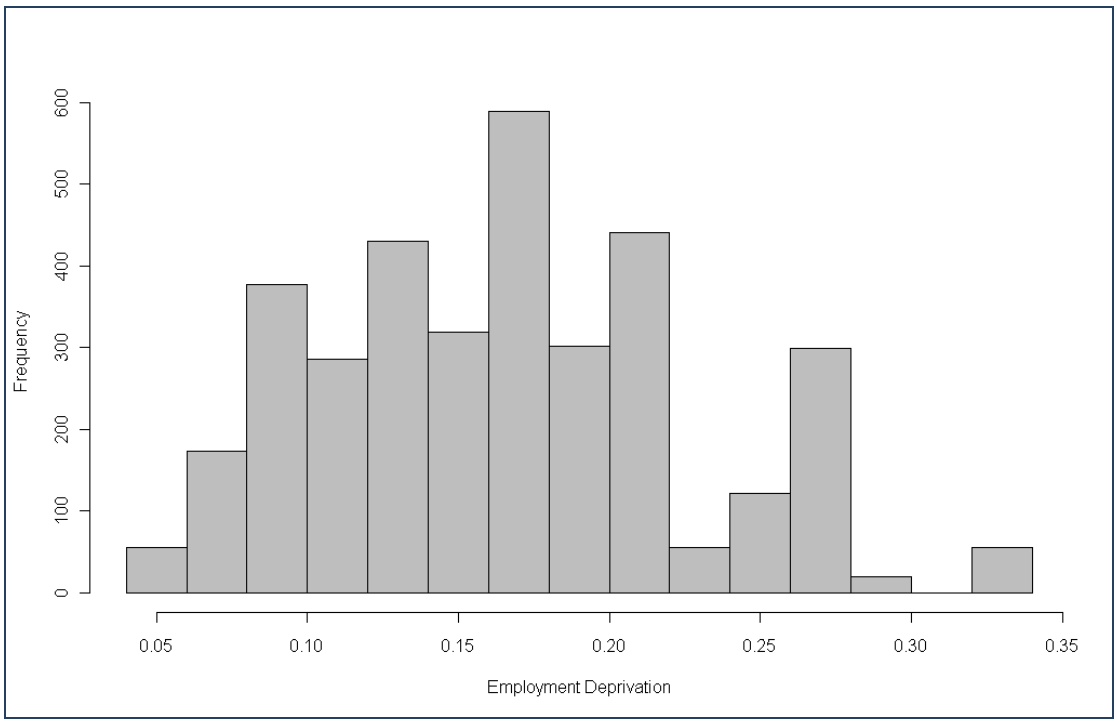


Figure 2-24. Histogram of Employment Deprivation

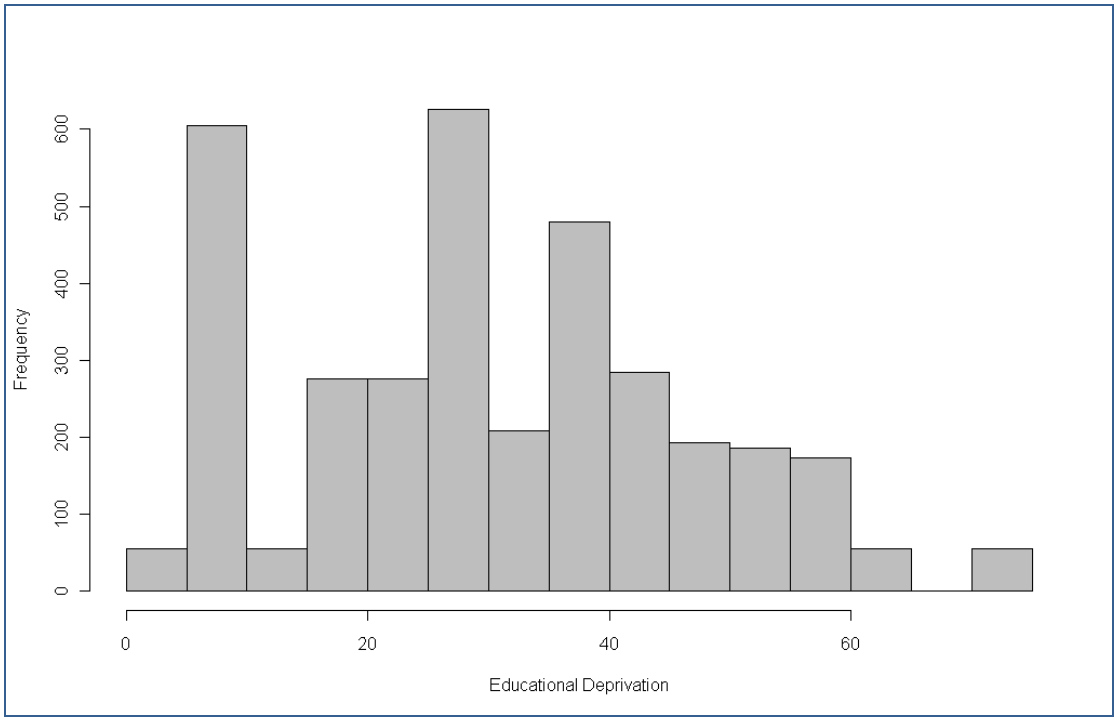


Figure 2-25. Histogram of Educational Deprivation

2.3.1.3 Discussion/Conclusions

An alternative way of estimating the practice deprivation would have been by assigning the deprivation scores to the postcode of practices' premises. However, such practice deprivation scores would not be representative of the contextual conditions in the areas that each practice served. To be representative, the majority of registered patients would have to live within the same LSOA where the practice premises were based or in LSOAs with similar deprivation. This could be true for some practices with small service areas, but would not apply to practices with large service areas. Overall, I considered that the estimated practice level deprivation scores were representative of the social composition of the patients registered to each practice.

.

2.4 Demographic Data

Demographic data are important for understanding disease prevalence and incidence in epidemiological studies. The two main demographic characteristics are age and sex of the subjects or population under study. The age and sex characteristics of practice populations were assessed in relation to salbutamol prescribing rate in the regression model I developed.

2.4.1 Age and Gender of Patients Registered per Practice

The age and sex data of registered patients were accessed from the Exeter database via NEPHO, after receiving ethical approval by the PCTs' Caldicott Guardians. I received data for each of the five years of study, extracted for the 1st of April for each year.

2.4.1.1 Exploration of Age and Gender Data

Figure 2-26 shows the age profile of populations registered per practice. Most of the practices had a similar age profile, while two of the practices had very high number of young population registered with them. These practices were close to Newcastle University and had a high number of students registered with them.

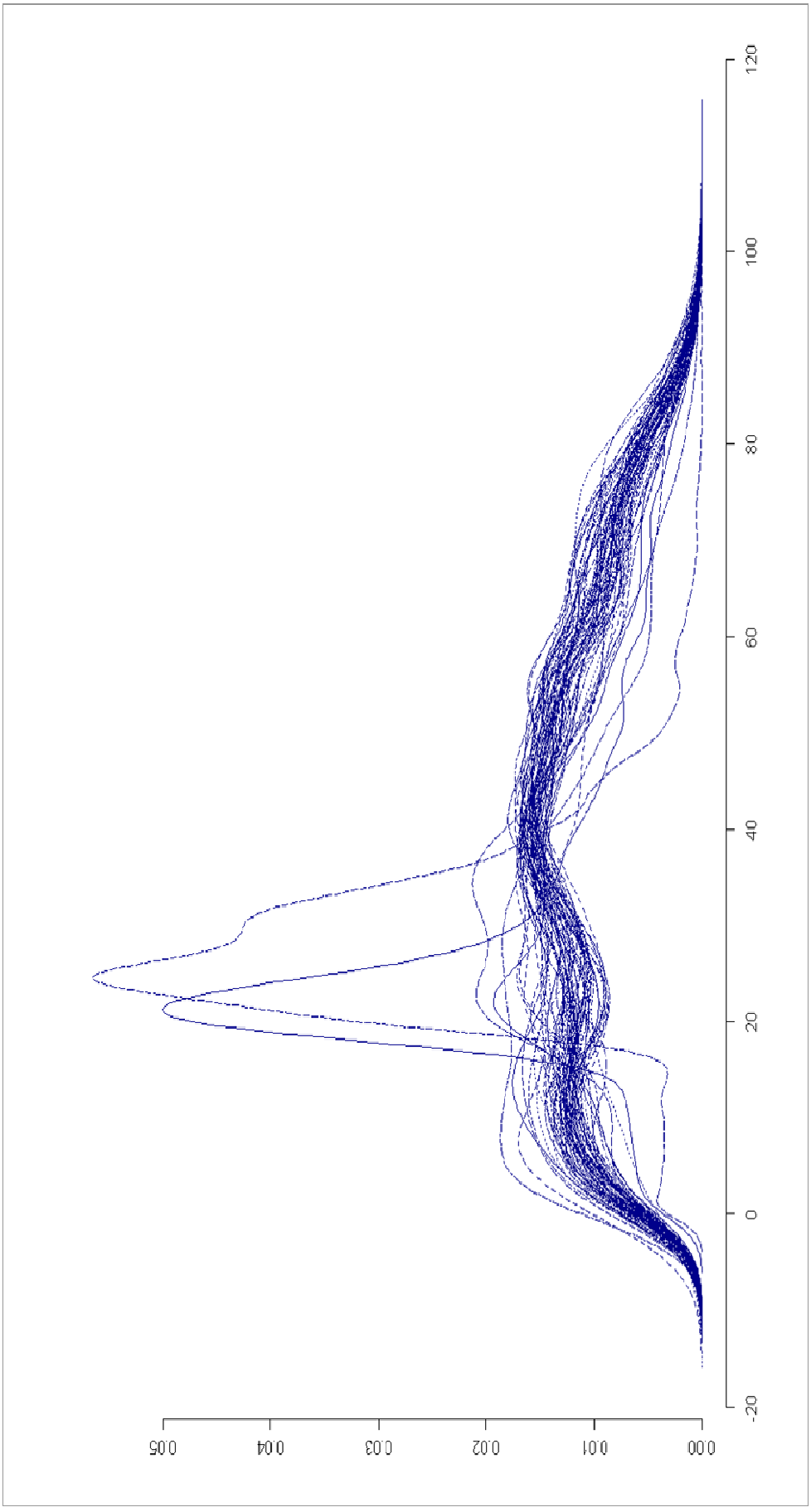


Figure 2-26 Age profiles of practice populations

The proportions of males and females in the total population registered per practice were estimated. These proportions are presented in Figure 2-27 using violinplots, which are similar to boxplots but depict density as well. The violinplots for males and females are mirror images because if 55% of a practice population consists of males the remaining 45% will have to be females and vice versa.

The most notable observation appeared in the violinplots of Newcastle PCT where a practice's population consisted of 65% males and 35% females. In order to investigate it further, I created violinplots per GP practice in Newcastle and North Tyneside PCT. In Figure 2-28, I present only the proportion of males per practice as the respective graphs for females were mirror images, as explained above. The practice with the high proportion of males was the A86027 practice, which is located within the campus of Newcastle University. I considered that the high proportion of registered males reflected the high number of males working for the university.

Two practices in North Tyneside PCT appeared to have a higher proportion of male patients than the other practices. However, the proportion was approximately 54%, which was much lower that the 65% observed in practice A86027. The vast majority of GP practices in North Tyneside appeared to have more female registered patients than males.

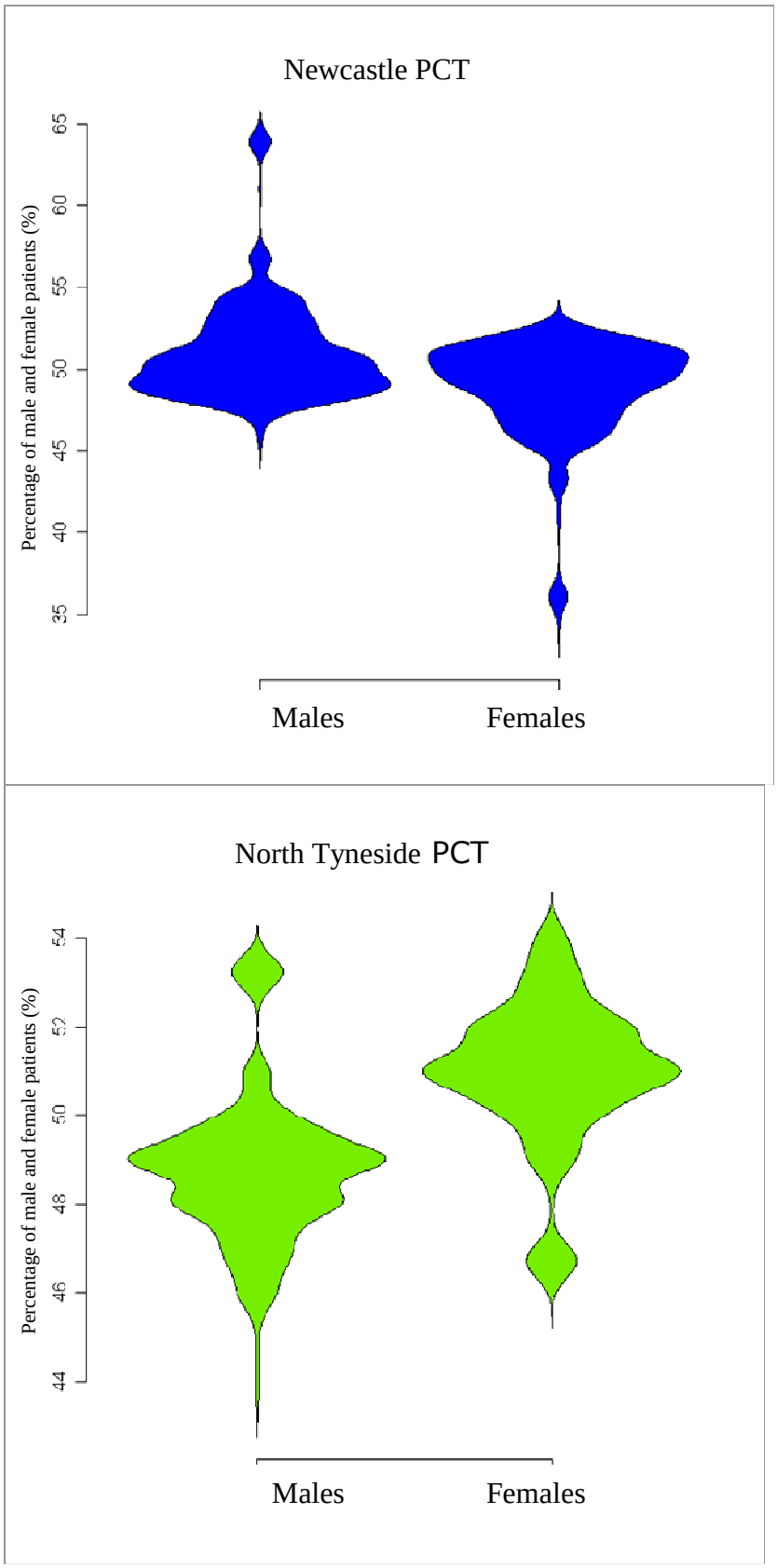


Figure 2-27 Percentage of males and females registered to GP practices, per PCT.

Figure 2-28 Percentage of males registered to GP practices, per practice

2.4.1.2 Results

I averaged the age of population per practice in order to create an indicator of the practice population age. I calculated the ratio of males and females registered per practice, with higher values when a practice population consists of more males than females and with lower values in the other way around.

Table 2-6 shows the descriptive statistics of the two variables to be included in my regression model. The histograms of the two variables are presented in Figure 2-29 and Figure 2-30.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Average age	29.0	38.0	39.0	39.0	41.0	46.0
Sex ratio	0.8	0.9	1.0	1.0	1.0	1.8

Table 2-6. Descriptive statistics of covariates

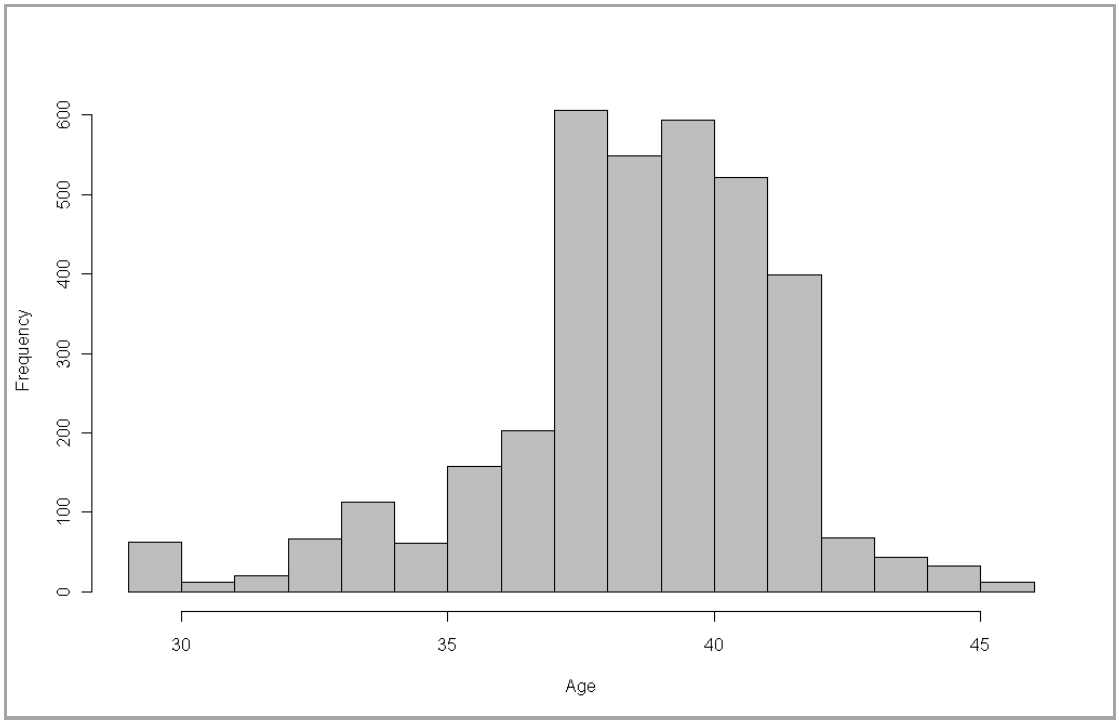


Figure 2-29 Histogram of average age of registered patients

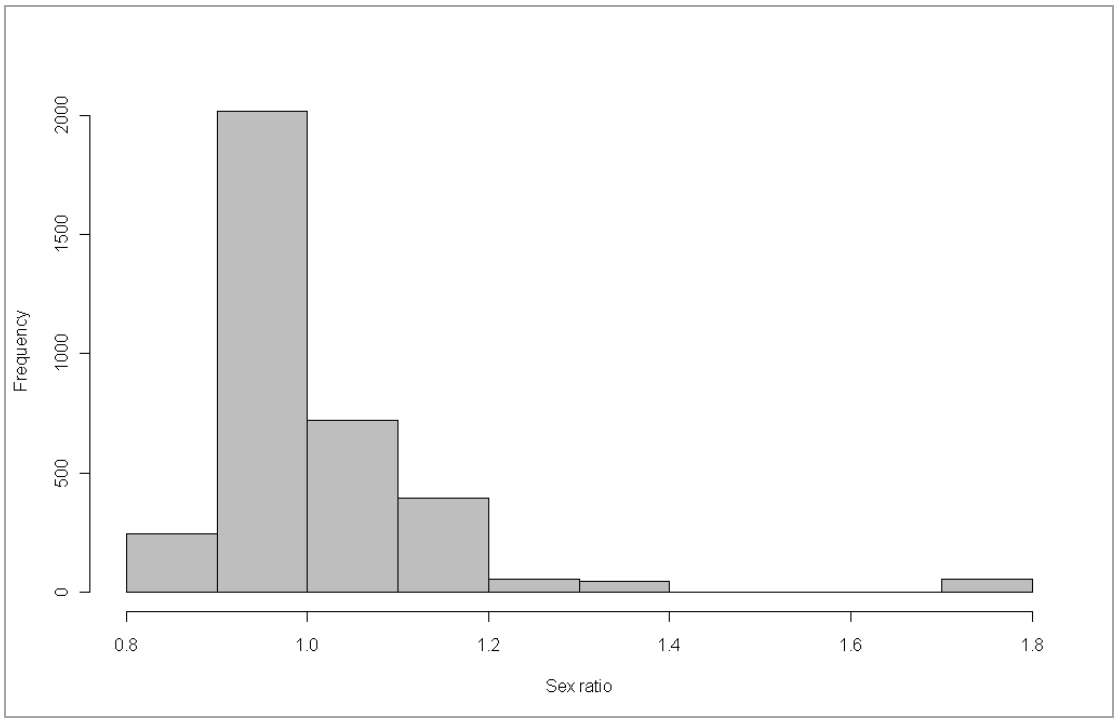


Figure 2-30 Histogram of sex ratio of registered patients

2.4.1.3 Discussion/Conclusions

The average age of registered patients per practice and the ratio of males and females gave a profile of the population served by each practice. Older people tend to suffer more from COPD, whilst asthma is more common in children. Gender appears not to be related to asthma or COPD. COPD was once considered a disease of males, probably due to the historically high smoking prevalence among males but it is not anymore. If the prescribing data were available per diagnosis, it would be possible to test per diagnosis how the age and gender profile of patients relates to respiratory prescribing, in a primary health setting.

Chapter 3. Statistical Modelling

3.1 Statistical Method

The statistical method to be employed was considered in parallel with the exploratory analysis. It is useful to have some consideration on the type of statistical model that in principle would be appropriate for the data and questions of interest before the model building process. A linear regression model was the starting point of my model building strategy. Careful consideration was required as to whether the data under analysis could satisfy the assumptions of the model to be used. There are several approaches that can be adopted to address the linear regression assumptions, depending on which assumption is not met. The grouped nature of prescribing data, with repeated measurements per GP practice over time, violated the basic assumption of independence that underlies the statistical methods used in linear regression (2000). Mixed models or mixed-effects models are a form of generalized linear model that has been extended to dependent data (Breslow and Clayton, 1993, Diggle and Ribeiro Jr, 2007, J.C. Pinheiro and Bates, 2000). I present below an overview of the theory behind mixed-effects model, in order to clarify the reason for considering this statistical method to be appropriate for my study.

3.1.1 Mixed-effects Model

Mixed-effects models provide a powerful tool for analysing grouped data. Data can be grouped in time and/or space so mixed models could be used to analyse repeated measurements over time (e.g. response of a patient to treatment as a series of observations over time) and/or data recorded at several centres (e.g. hospital, clinics, general practices). The term mixed-effects is used to describe the

fact that the models account for both fixed and random effects. Fixed effects refer to the parameters of the model that have a constant value and are estimated by traditional regression models. Random effects assume that some parameters take values that have arisen from a distribution. They are selected at random from the population of interest and are called effects because they represent a deviation from the overall mean (Pinheiro and Bates, 2000). Those effects assist in depicting more accurately the phenomenon under study by accounting for the random variation associated with given parameters.

Fixed effects models in classical statistics only have random variation in the error term, while mixed effects models introduce additional random variation in the parameters of the model. Any number of random effects can be specified in a model. For example, when assessing the effect of two treatments that are provided to several individuals by several centres, it is possible for the model to capture the by-treatment random effect as well as to include the centre providing the treatment as a random effect. This type of design is called hierarchical or multilevel as individuals are grouped per treatment and centre. Mixed-effects models are often called hierarchical (Raudenbush and Bryk, 2002, Lindley and Smith, 1972) or multi-level models (Goldstein, 2003). However, not all mixed-effects model are hierarchical (Pinheiro and Bates, 2000). Careful consideration is required when assigning random effects to variables as they have to be representative of the real case. Cases where an identical treatment is given to the same individual, no random effect by treatment is necessary, and only centre would be included as random effect.

More precise estimates of the treatments' effects can be estimated by assigning random variation to both centre and centre/treatment effects. In such hierarchical designs, the treatment effects are allowed to vary randomly across centres and the treatment standard error increases to allow for this. The inference of this model can be applied to the whole population of centres unlike conventional models where centre and centre/treatment effects are fixed and inferences on treatment effects are specific to the centres observed (Brown and Prescott, 2006). This important aspect of mixed-effects models is that the question of interest is for the whole population and not only for the population being sampled (Pinheiro and Bates, 2000). Especially when analysing health data that need ethical and confidentiality considerations to be assesed, using the fewest patients' data as possible to make inferences with same level of accuracy is desirable.

3.1.1.1 Application of Mixed-effect Models

Overall there are many reasons for using such model and its benefits can be applied to a range of data (Demidenko, 2004). Mixed effects models have started becoming established as the appropriate tool for modelling data with complex structure. Mixed effects models have been frequently applied in agriculture to analyse crop variety trials (Talbot 1984) and have been used extensively in animal breeding to predict heritability and predict genetic gain from breeding programmes (Smith et al., 2001, Johnson and Thompson, 1995). They are also applied in other fields such as engineering, ecology, medical science and in social sciences like education, linguistics, economics (Elmhagen and Rushton, 2007, Zhou et al., 2003, Revelt and Train, 1998, Clayton et al., 1996, Baayen et al., 2008).

The adoption of mixed effects models has been increasingly accepted in medical research, especially after a review of clinical trials by Brown and Kempton (1994). Mixed-effects models have been used in different aspects of asthma and COPD research, such as the response of portable peak expiratory flow meters to changes in true peak expiratory flow in 12 children with asthma (Burton et al., 1998), as well as to evaluate the effect of inhaled budesonide in subjects with mild COPD who continued smoking (Pauwels et al., 1999). The last two decades have seen several examples of using mixed models in relation to health (Pitt et al., 2000, Drake et al., 1999, Clark et al., 1996, Unutzer et al., 2002), however their use is still in the process of becoming regular practice. Mixed-effects models are also accepted by regulatory authorities. The US Food and Drug Administration website recommends such models for assessing the degree to which two drugs are the same in terms of efficacy and safety (Centre for Drug Evaluation and Research, 2001).

The application of mixed effects models is very much linked to the development of computers. Mixed effects models are fitted using the method of maximum likelihood or the residual maximum likelihood (RELM) method, proposed by Patterson and Thompson (1971). Likelihood-based methods are computationally intensive and demanding in terms of computational power, which has been a factor in restricting the application of such models.

A variety of commercial software, including SAS, SPSS, MLwin, HLM, Stata and S-plus is available for fitting mixed-effects. They differ mainly in whether they include programmes that allow fitting mixed-effects to normal, non-normal or categorical data. The commercial software Egret specialises in the application of mixed-effects models for epidemiological and biomedical studies. Software like WinBUGS and BayesX apply a Bayesian approach to such models. To the best of my knowledge WinBUGS is the most widely applied software for Bayesian statistics and can be

used to fit models of all types of data. It is open source software (distributed freely in the public domain) developed by the Medical Research Council Biostatistics Unit in Cambridge. Both WinBUGS and BayesX are integrated within R and can be used via this software.

R is open source statistical software that includes algorithms for fitting mixed-effects models. R allows one to fit mixed models to larger datasets compared with other software and is one of the most versatile software packages for application of mixed-effects models, using "nlme" and "lme4" packages. The latter package has been launched in recent years and is currently under continuous development to improve computation of mixed models and include new functions that deal with more complex datasets. For instance, lme4 allows the use of an offset in a mixed-effects model, which is not available in the nlme package. An offset specifies a priori known components to be included in the linear predictor during fitting. R also contains special functions to improve the modelling of the aforementioned types of data, like a function for fitting longitudinal models with missing data that is based on work in this field by Diggle and Kenward (1994) and Diggle and Farewell (2007). A recent addition in R is the package called "geoR" and its extension "geoRglm" that allows the application of mixed-effects models to geostatistical data (Diggle and Ribeiro Jr, 2007).

3.1.1.2 Mixed-effects model in perspective

Mixed effects models have random variation in the error term but also introduce additional random variation for which assumptions have to be made about the type of distribution they follow. The two sources of random variation are the main characteristic of mixed effects model. This is a potential disadvantage of mixed models as they rely on more distributional assumptions. Often a class of mixed-effects models is used that assume that both the random effects and errors follow Gaussian distributions. This assumption always has to be checked, by examining the linearity of residuals and random effects with normality plots. Non-linear mixed models are also applied (Pinheiro and Bates, 2000).

As mentioned above, in contrast to classical statistics, the mixed-effects model assumes two sources of variation, within groups and between groups. The within groups variation has the same meaning as in traditional statistics and is depicted by the model's error terms (residual variance). The between groups variation is also a random source of variation like the error terms, even though may behave like parameters in a model. As such we do not "estimate" them but form predictions of

the values of those random parameters, given the data observed (Pinheiro and Bates, 2000).

The fact that a mixed-effects model assumes that the parameters are random makes it similar to Bayesian statistical models. In a Bayesian context where all unknown parameters are considered as "random", the mixed-effect model is known as a random-effects model (Gelman, 1995). As in the Bayesian approach, in a mixed-effects model one has to make a decision in defining the prior distribution of random parameters. In the frequentist approach, the unknown parameters of the distribution are estimated from the data. In a mixed effects model, the fixed effects can also be estimated based on specified repeated-sampling properties that are equivalent to the Bayesian posterior inference under a non-informative (uniform) prior distribution (Gelman, 1995). In summary, a mixed model combines major features of the frequentist and Bayesian approaches.

3.1.1.3 Example of Mixed-effects Model

To illustrate the importance of accounting for the grouping factor of this study's prescribing data by means of mixed-effects model, I used a basic fixed effects regression model initially. For simplification I considered the relationship of particulate matter as an explanatory variable x_k with salbutamol prescribing in several GP practices as response y_k, where K=*1,…,k* in our study K=55 months. The model took the form:

$$y_k = \mu + \alpha x_k + \varepsilon_k \qquad 3\text{-}1$$

Where, y_k is the salbutamol prescribing rate, μ *is* the mean salbutamol prescribing across the population of GP practices being sampled, and the ε_{ij} are independent and identically distributed random variables with zero mean and constant variance, $N(0,\sigma^2)$. The parameter (coefficient) of the model is *a* with respect to the particulate matter covariate x_K. In this model it is assumed that the data were collected from similar, homogenous practices. However the practices are not homogenous and prescribing varied a lot between them. Figure 3-1 illustrates that some practices tended to have systematically higher or lower prescribing than others and it would be better to account for that.

The grouping factor of prescribing is categorical data. There is no inherent ordering of GP practices so I reordered the data to make the plot of prescribing between groups more informative. Figure 3-2 shows a dotplot of salbutamol prescribing per GP practice that reinforced the impression of considerable variability between GP

practices. The line observed in Figure 3-2 joined the mean salbutamol prescribing of the 64 general practices, which had been reordered accounted to increasing mean salbutamol prescribing.

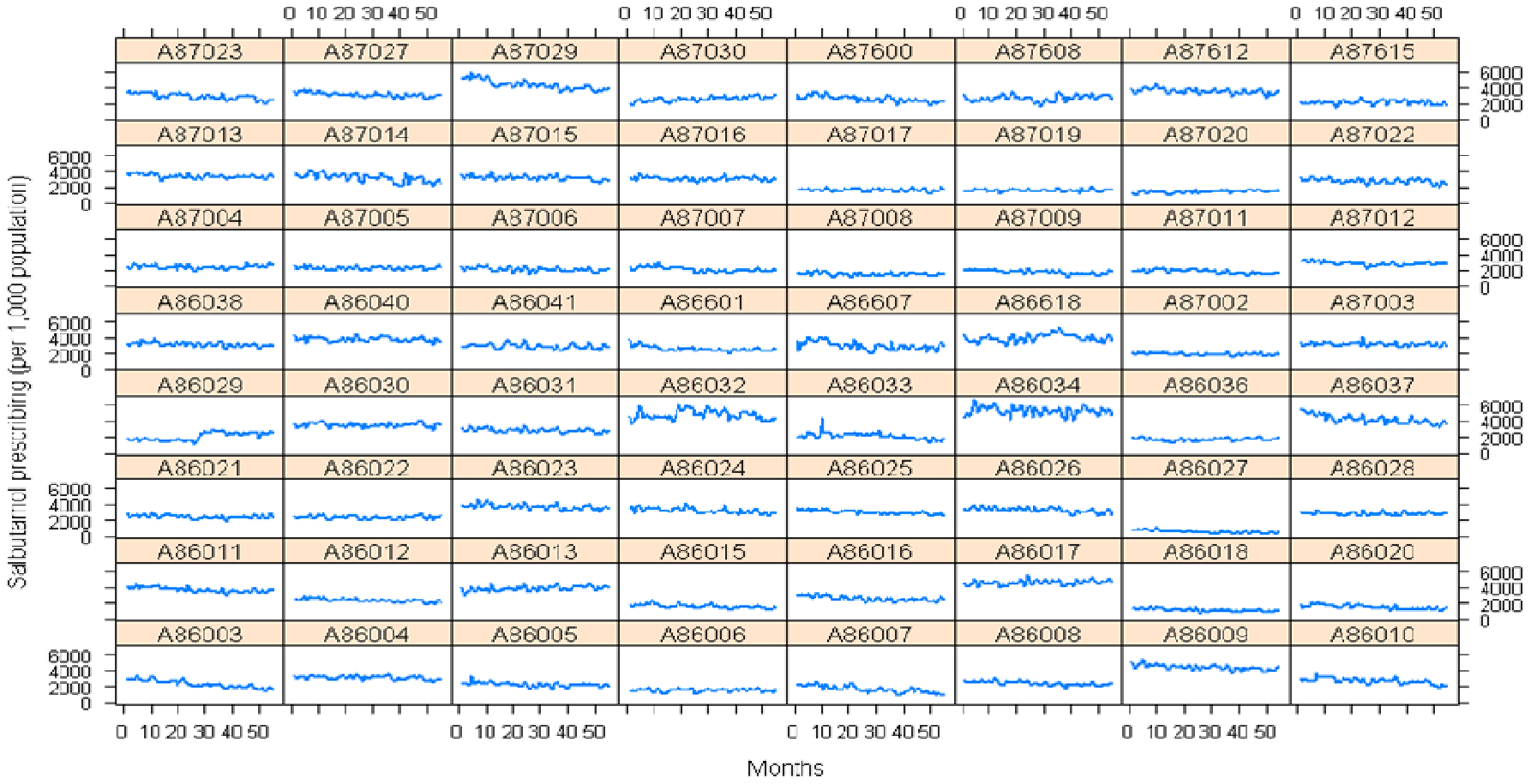


Figure 3-1 Monthly average of salbutamol prescribing per GP Practice.

Figure 3-2 A dotplot of salbutamol prescribing per GP Practice.

An improvement of this initial model was to assume that each practice has its own practice-specific prescribing (in statistical language, intercept). The model now has a double index because I have grouped data: *i* corresponds to the *i*th GP practice (*i*=1,…,*N*) and *j* to the *j*th observation of the *i*th practice (*j*=1,…n_i), where n_i= is the number of observations from the *i*th practice. The total number of observations is $\mathrm{K} = \sum_{i=1}^{N} n_i$. In this study, the number of GP practices was i=1,…,64 and the number of observations per practice was j=1,…55, so K= 3,520.

$$y_{ij} = \mu_i + \alpha x_{ij} + \varepsilon_{ij} \qquad 3\text{-}2$$

Where, y_{ij} is the salbutamol prescribing rate for observations *j* in group *i,* μ_i is the GP practice specific intercept. The ε_{ij} is the error term for *j*th observation of the *i*th practice, again assumed independent and identically distributed, N (0, σ^2). This second model seemed more appropriate for the prescribing data as it included the GP practice effect. However, it assumed that the practice effect takes constant (fixed) values and the only assumption about variation is that the residuals ε_{ij} are normally distributed. As the individual observations are modelled as the sum of $\mu_i+\alpha x_{ij}$, which are all constants, plus the residual term, it follows that the variance of individual observations equals the residual variance. The covariance of any two separate observations y_{ij} and $y_{i'j'}$ is zero since all the residuals are assumed independent (i.e. uncorrelated).This means that data sharing the same grouping factor (e.g. prescribing per GP practice) are considered independent. However, prescribing per GP practice was not expected to be independent due to various random influences, such as the prescribing pattern individual doctors may have.

An improvement of this second model was to set the GP practice as a random parameter to allow the model to capture the by-practice random effects. The new model assumed that the practice effect also arises from independent samples from a normal distribution and is an example of mixed-effects model. The equation 3.2 (fixed-effects model) modelled the specific sample of GP practices used in this study, while the random effects model treated the practices' effect as random variations around a population mean. This central assumption of the mixed models, that intercepts μ_i, *i*=1,…,*N* are random and belong to a general population, can be expressed in the following equation as

$$\mu_i = \mu + b_i \qquad 3\text{-}3$$

where μ is the population-average prescribing of salbutamol (intercept) and b_i is a random variable, or the deviation of the practice-specific prescribing from the population-averaged prescribing for the *ith* GP practice. Taking into account the random part of the intercept expressed by equation 3-7, the model 3-2 could be written as:

$$y_{ij} = \mu + \alpha x_{ij} + b_i + \varepsilon_{ij} \qquad 3\text{-}4$$
$$\varepsilon_{ij} = \mathrm{N}(0, \sigma^2) \;\&\; b_i = \mathrm{N}(0, \sigma_b^2)$$

Where the random variable ε_{ij} is the error term for *j*th observation of the *i*th practice, again assumed independent and identically distributed, N (0, σ^2). The b_i are referred to as random effects also assumed to have come from a normal distribution $\mathrm{N}(0,\sigma^2)$. Joining the two random terms together, where $\eta_{ij} = \varepsilon_{ij} + b_i$ the model can be written as $y_{ij} = \mu + ax_{ij} + \eta_{ij}$. In general, the linear mixed effects model is written as (Laird and Ware, 1982)

$$y_{ij} = \mu + ax_{ij} + Z_i b_i + \varepsilon_{ij} \qquad 3\text{-}5$$

where Z is a matrix giving the values of random effects b_i. It is apparent this model is more complex than the previous model (3-2). Consequently, the mixed model assumed two sources of variation (variability), the ε_{ij} or "within groups" and b_i or "between groups" (Pinheiro and Bates, 2000). The former have the same meaning as in traditional statistics but the latter are random and are estimated as posteriori means, that links such model to Bayesian statistics (Demidenko, 2004, Gelman, 1995), as discussed in section 3.1.1.2.

Furthermore, the variance of individual observations in the mixed effects model is the sum of the variance components:

$$\mathrm{var}(y_{ij}) = \sigma_b^2 + \sigma^2 \qquad 3\text{-}6$$

σ^2_b is the variance for the between group variability (or inter variability) and σ^2 is for the ε_{ij} for the within group variability (or intra variability). The covariance of pairs of observations can then be described as:

$$\text{cov}(y_{ij}, y_{i'j'}) = \text{cov}(\mu + \alpha x_{ij} + b_i + \varepsilon_{ij}, \mu + \alpha x_{i'j'} + b_{i'} + \varepsilon_{i'j'}) \qquad 3\text{-}7$$

When observations of different GP practices are considered (i.e. i≠i') then $\text{cov}(y_{ij}, y_{i'j'}) = 0$ because of the independence of the observations. However, when two samples from the same practice are considered (i.e. i=i') (Brown and Prescott, 2006, Demidenko, 2004, Pinheiro and Bates, 2000), then

$$\text{cov}(y_{ij}, y_{i'j'}) = \text{cov}(b_i + \varepsilon_{ij}, b_{i'} + \varepsilon_{i'j'})$$

$$= \text{cov}(b_i, b_{i'}) = \sigma_b^2 \qquad 3\text{-}8$$

Thus, observations on the same GP practice are correlated and have covariance equal to the practice variance component, while observations on different practices are uncorrelated. This contrasts with the fixed effects models where the covariance of any pair of observations is zero.

3.1.1.4 Discussion/Conclusions

In summary, grouping factors induce a correlation structure in the data but mixed effect models allow the modelling of such data (Faraway, 2006, Demidenko, 2004, Pinheiro and Bates, 2000). Allowing for the interdependence of observations by group was the reason for considering this type of statistical method appropriate to model the prescribing of salbutamol by GP practice. Each practice has characteristics (e.g. age distribution of patients) that I needed to take into account in my model as well. In the following section I developed the mixed-effects model by adding all the explanatory variables of salbutamol prescribing and carrying out tests to assess formally whether this model met the independence assumption required by linear regression.

3.2 Model Formulation

In previous chapters, I described how the health, air quality, deprivation and demographics datasets were converted to suitable spatial and temporal formats to allow them to be used as explanatory variables. I have also discussed the statistical model that in principle is relevant for modelling prescribing data at the primary care level.

My aim was to model the monthly prescribing rate of salbutamol (respiratory medication) in relation to air quality, at primary health care level. I hypothesized that exposure to air pollution increases the frequency and duration of asthma and COPD symptoms, generating a corresponding increase in the use of salbutamol medication and consequently an increase in prescriptions. I evaluated possible time-lags in the response of medication use to air quality. Contextual factors of local areas (income, employment and educational deprivation) and demographic factors (age and sex) were also included as covariates in the statistical models.

The response variable was monthly prescribing of salbutamol per GP practice, standardized per 1,000 population. The primary health care level was the spatial unit of analysis for prescribing and monthly variation was the temporal unit. Not all covariates shared the same level of temporal and spatial resolution. The ambient air quality (PM_{10}) covariate deriving from one monitoring site accounted for monthly temporal but not spatial variation. I attempted to account for within city air quality spatio-temporal variability by analysing traffic flows with both spatial (50 monitoring sites) and temporal (monthly) resolution.

The demographics and the derived deprivation covariate have one year temporal resolution, as the postcodes of patients were provided on the 1st of April for each year of study. No significant change in deprivation and demographics variables was expected on a monthly basis and I considered that these covariates accounted sufficiently for spatio-temporal variation. I summarise below the covariates constructed for use in my statistical model:

$x_{1a\text{-}1b}$: monthly averaged PM_{10} with 0-, 7-, 14-, 21-days and 1-month lag time

$x_{2a\text{-}2e}$: monthly averaged traffic flows per practice with 0- and 1-month lag time

x_3: income deprivation per GP practice, annually

x_4: employment deprivation per GP practice, annually

x_5: education deprivation per GP practice, annually

x_6: average age of registered patients per GP practice, annually

x_7: ratio of males and females registered patients per GP practice, annually

x_8: time (months)

My interest was to explore the potential relevance of covariates, as well as learn and understand better the structure of the data, in order to build a model that I believed to be both parsimonious and adequate. The inference to be made using the p-values, coefficient and confidence intervals of my final model would be relevant to the total population of GP practices that my sample was drawn from, as well as the sample of practices I analysed. In the next section I focus on model formulation, the preliminary stage of model building strategy.

3.2.1 Preliminary Statistical Model

My model building started with the model formulation process that is developing and examining the main structure of the statistical model, which was followed by the refining process. Preparation and examination of the main aspects of the statistical model reduced the risk of the final model not satisfying the main assumptions of regression. As discussed in section 3.1.1, a linear mixed-effects model was considered, in principle, appropriate for addressing the aims of this study. I fitted this model on my data and examined whether this was the case. Firstly, I assessed whether transformation of the data was required and then fitted the preliminary mixed-effects model. I present below the process for deciding on log transforming the data and then the application of a mixed-effects model.

3.2.1.1 Linear Mixed Effects Model

One of the main assumptions that linear regression models have to satisfy is homoscedasticity (constant variance). Logarithmic transformation is a common approach taken to address the problem of heteroscedasticity (Weisberg, 2005). In order to make a decision on whether log transformation of the data was required, I estimated the mean prescribing rate of salbutamol per month and its variance and plotted them against each other.

Figure 3-3 shows that the monthly mean of salbutamol prescribing rate increased when variance was increased. Log transformation approximately stabilised the variance of salbutamol prescribing, as depicted in

Figure 3-4, therefore, I proceeded to use log transformed data to model the salbutamol prescribing rate using linear regression. I checked again that the assumption of homoscedasticity was met by my final fitted model by examining the residuals of my model.

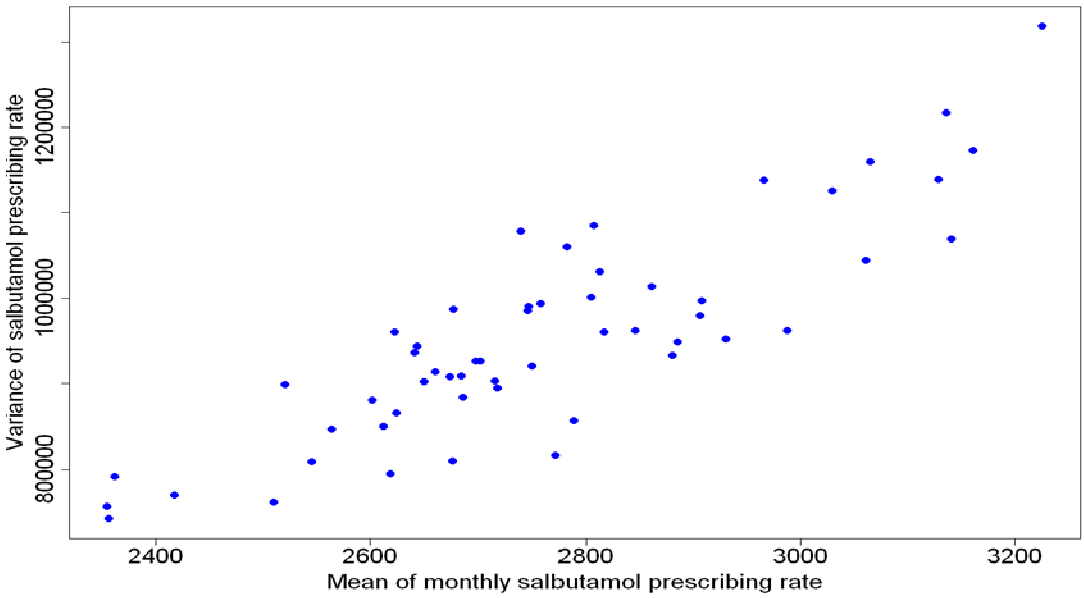


Figure 3-3 Mean of monthly salbutamol prescribing rate against monthly variance.

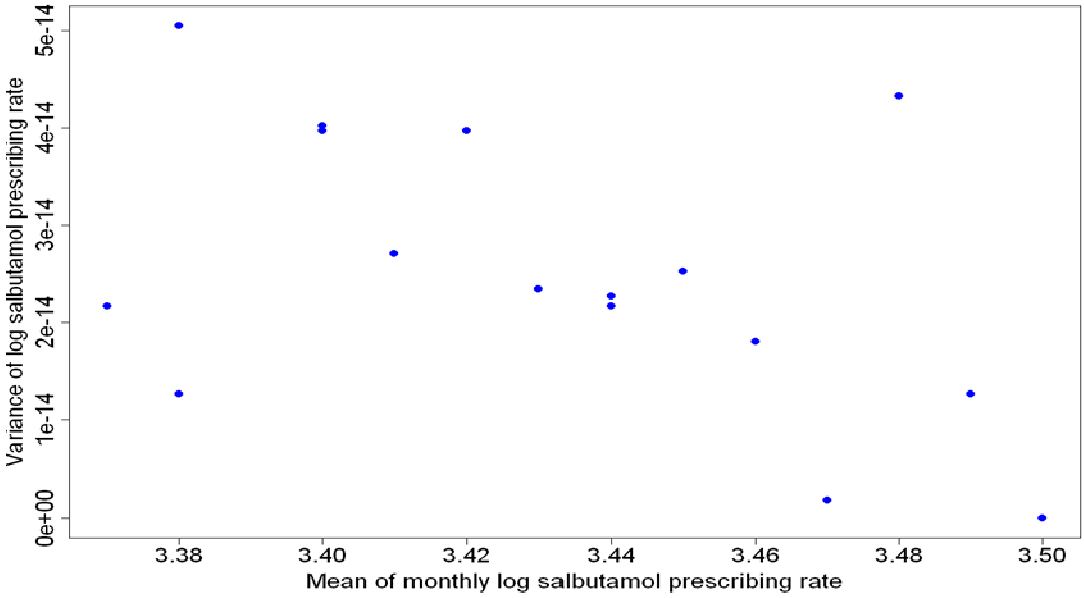


Figure 3-4 Mean monthly log salbutamol prescribing rate against monthly variance.

Having log transformed the data, I denoted Y the log transformed rate of salbutamol prescribing. Figure 3-5 shows the area-wide average salbutamol prescribing, $\overline{Y_t}$, in each of 55 months of the study period, where the original scale is shown on the left vertical axis and the logarithmic scale on the right vertical axis. Figure 3-6 depicts the monthly log transformed rate of salbutamol prescribing per GP practice.

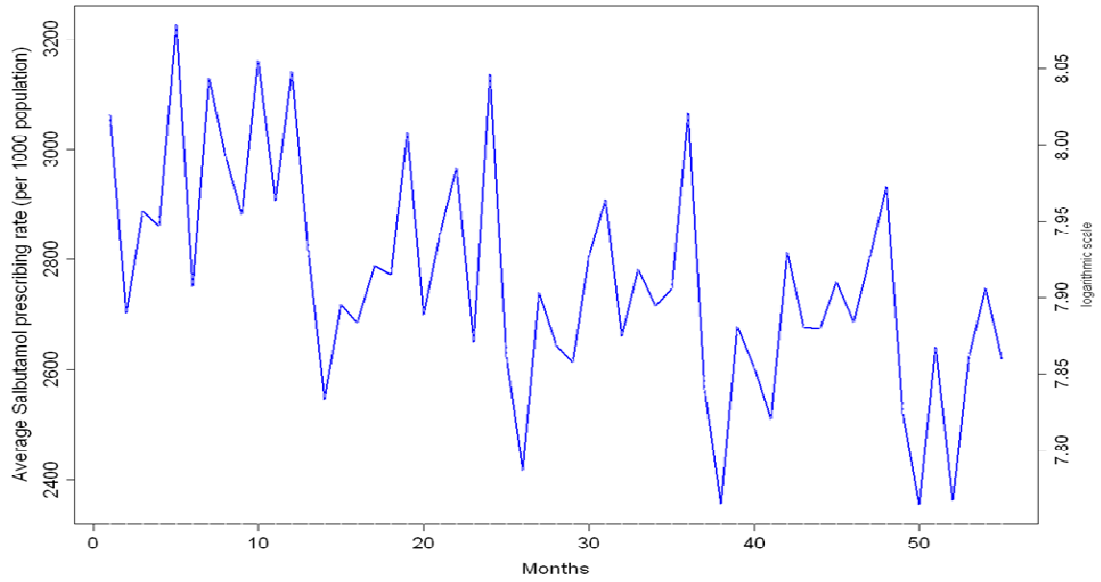


Figure 3-5 Area wide monthly average of salbutamol prescribing rate.

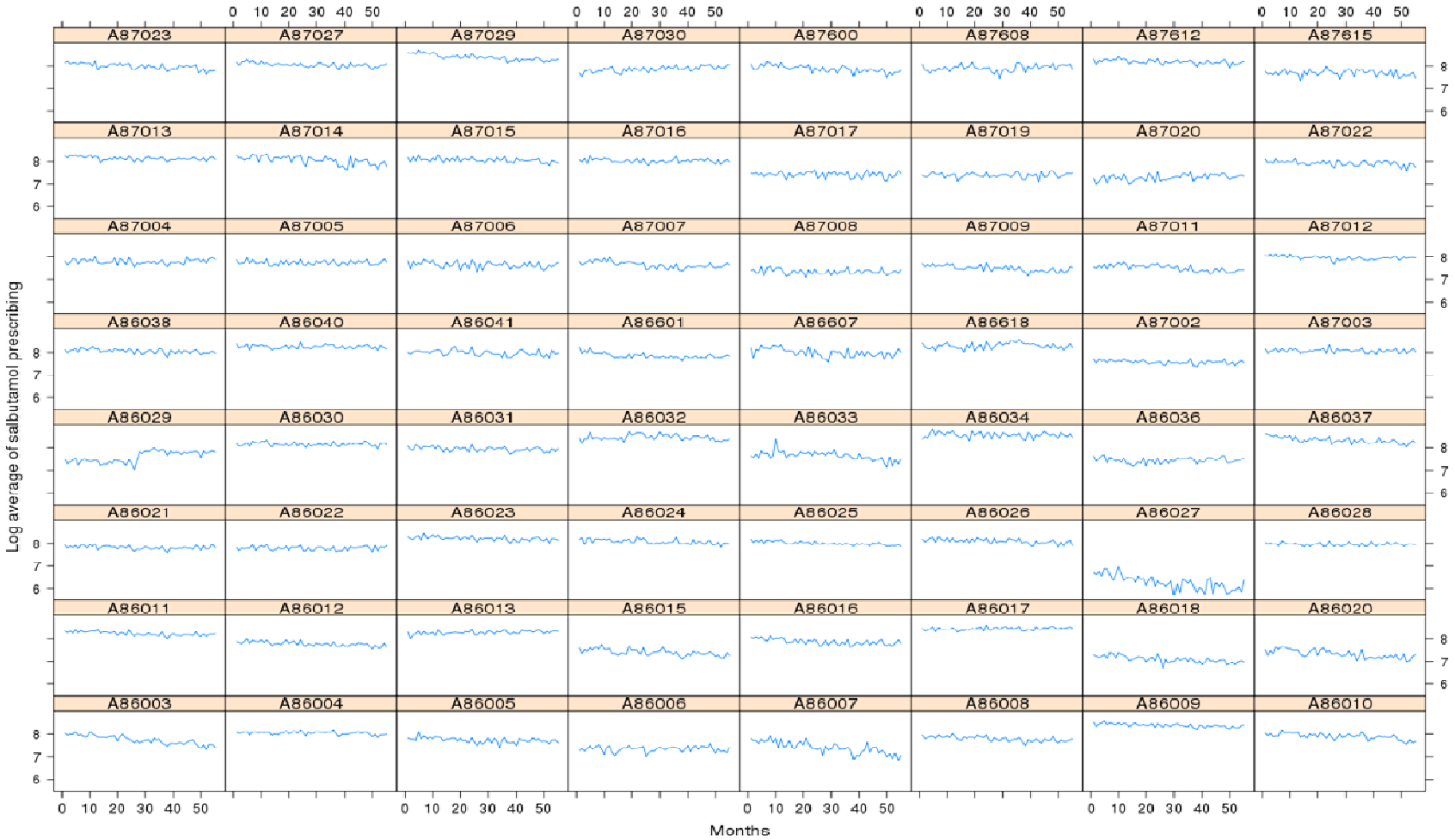


Figure 3-6 Log monthly average of salbutamol prescribing per GP Practice.

I fitted both fixed and mixed effects models adding all the information that I had gathered about individual practices on particulate matter, traffic, deprivation and demographics. All covariates have been listed at the beginning of this chapter. I then examined and compared the spread of the residuals of the models. The examination of residuals is necessary for identifying whether a model's results can be reliable.

The fixed effects model with all the covariates of interest took the form:

$$y_{ij} = \mu + a_1 x_{ij} + \ldots + a_9 x_{ij} + \varepsilon_{ij} \qquad 3\text{-}9$$

as it fitted eight fixed effect parameters relating to x=1,…,8 covariates. The mixed effect model had additionally 64 random effects (q=1,…,64) parameters respective to 64 GP practices:

$$y_{ij} = \mu + a_1 x_{ij} + \ldots + a_8 x_{ij} + b_1 z_{ij} + \ldots + b_{64} z_{ij} + \varepsilon_{ij} \qquad 3\text{-}10$$

3.2.1.2 Model Criticism

Figure 3-7 shows boxplots of the fixed effects linear regression model residuals by subject. The residuals corresponding to the same subject (GP practice) tend to have the same sign, which indicated the need to account for the group effect (Pinheiro and Bates, 2000). The anticipation that the mixed effects model would account successfully for the GP effects was better illustrated by the boxplots of residuals, shown in Figure 3-8. The residuals are centred around zero and have smaller magnitudes than those in Figure 3-7. This indicated that the mixed-effects model could provide more accurate estimates, which allowed for the by-GP practice differences.

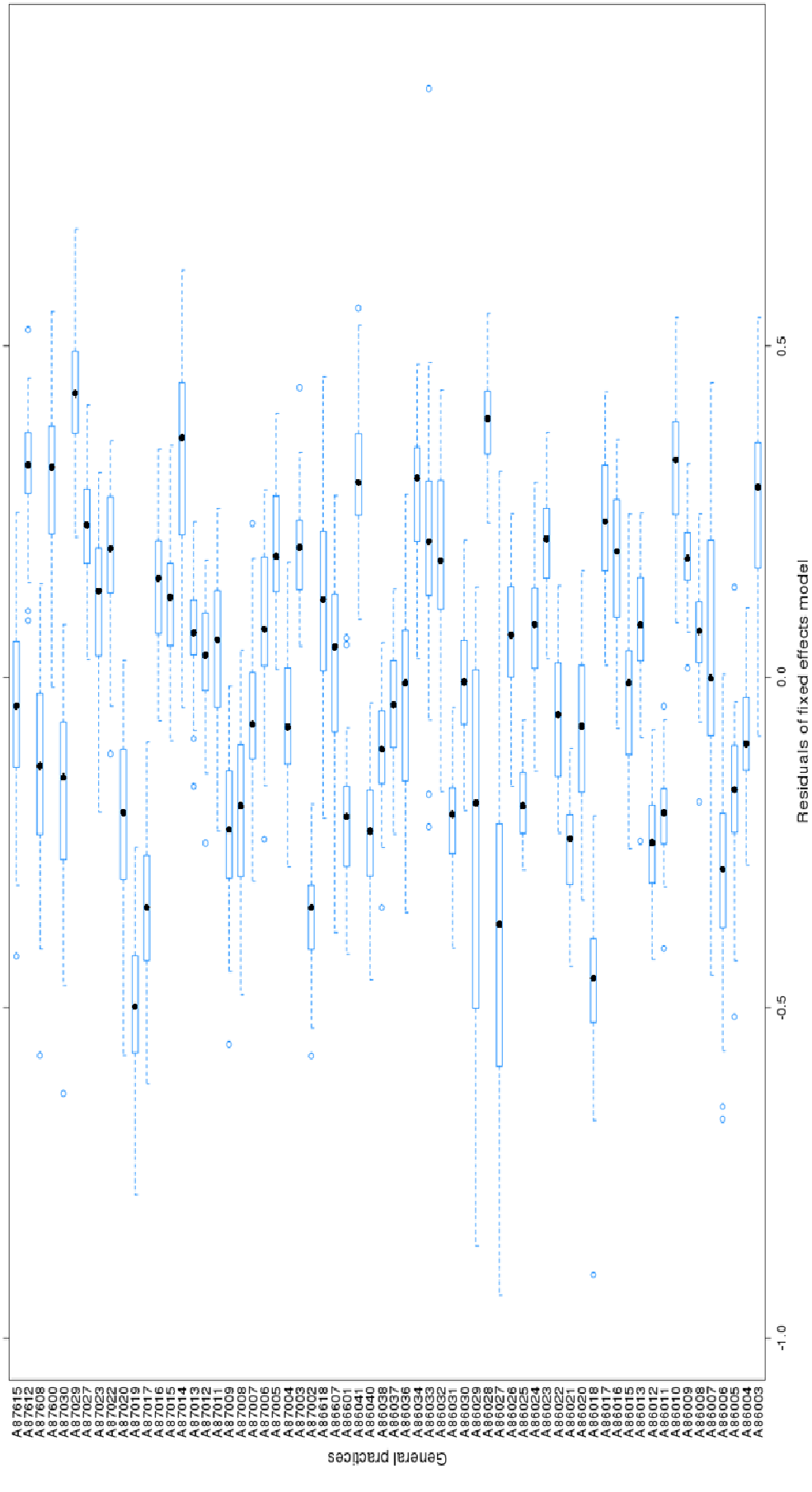


Figure 3-7 Boxplots of the fixed effects model residuals, by GP Practice.

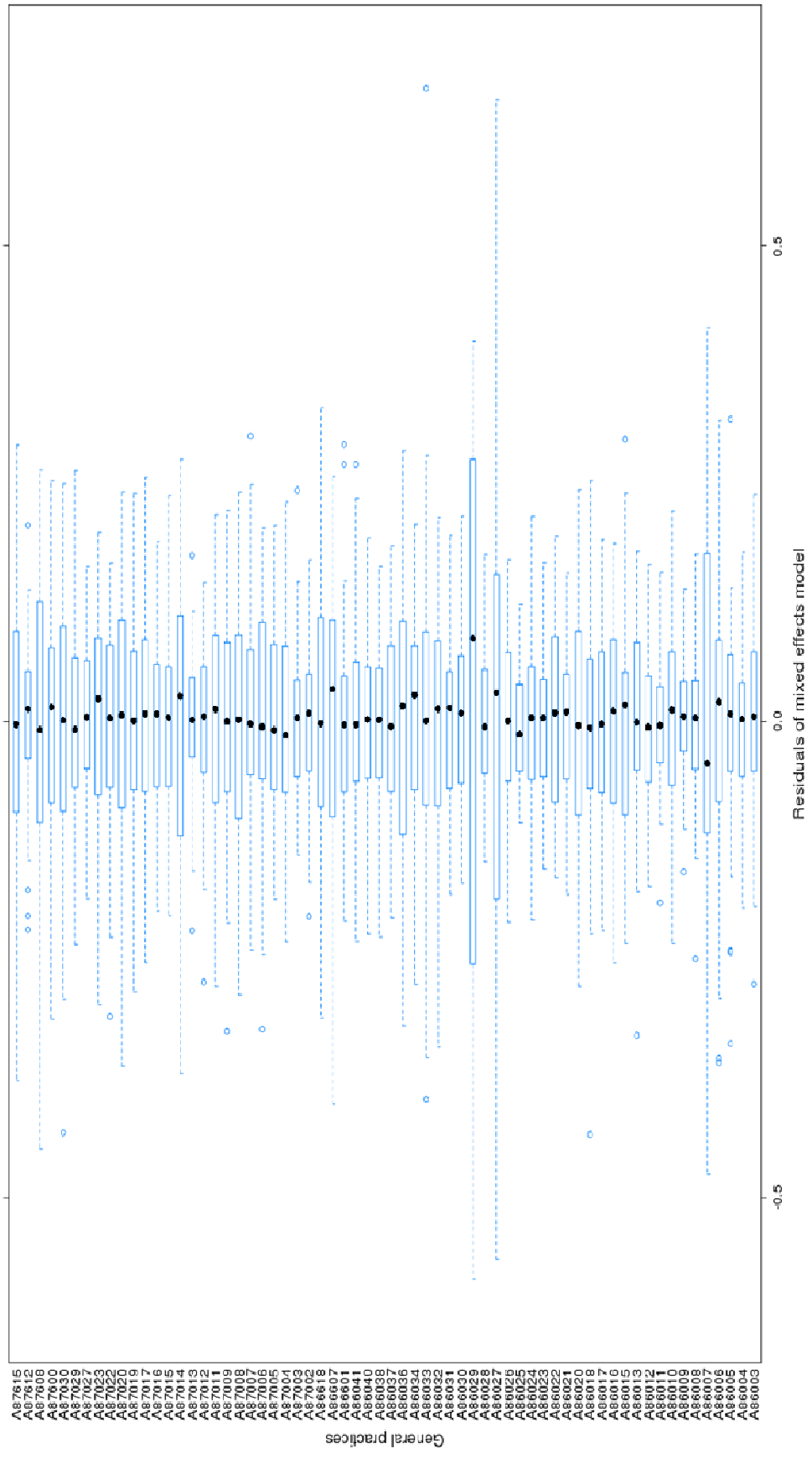


Figure 3-8 Boxplots of the mixed effects model residuals, by GP Practice.

3.2.1.3 Discussion/Conclusions

The residuals of the mixed effects model were centred around zero and had smaller magnitudes than those of the fixed effects model, indicating that the mixed effects model had more accurate estimates, taking into account the by-GP practice differences. Overall, by examining the spread of residuals, I concluded that the reliability of the fixed effects model would be doubted when fitting the salbutamol prescribing data in relation to other covariates, while the mixed effects model appeared to better account for the grouped nature of for the prescribing data.

3.3 Model Refining (Stage A)

Having fitted and examined the preliminary model, I needed to refine and finalise it. A major concern was the covariates that would be included as explanatory variables for salbutamol prescribing. Peaks in salbutamol prescribing occured each winter and spring, albeit with some variation from year to year (Figure 3-5). I have already mentioned that factors like pollen, cold air and respiratory infections can increase the frequency and duration of symptoms of patients with asthma and COPD. The salbutamol prescribing was consequently determined by such seasonal variations.

In an attempt to account for this seasonal effect, my modelling strategy was separated into two stages. Firstly, I modelled the expectation of the area wide monthly average log-transformed salbutamol prescribing rate, $\mu_t = \mathrm{E}\left[\overline{Y_t}\right]$, in relation to monthly average temperature. Temperature was used to capture the seasonal variation of salbutamol prescribing, given the absence of any other covariates to account for seasonal trigger factors (e.g. respiratory infections, pollen).

I then fitted the mixed effects model presented in the previous section, including the monthly average log-transformed salbutamol prescribing rate, $\mu_t = \mathrm{E}\left[\overline{Y_t}\right]$, as an offset. An offset specifies a priori known component to be included in the linear predictor during fitting. Including $\hat{\mu}_t$ as an offset in my model would allow me to assess the relationship of the remaining prescribing rate in relation to particulate matter, traffic, deprivation and demographic characteristics of GP practices. A two stage modelling strategy has been previously implemented by Fanshawe et al., 2008. Implementing a two stage-modelling approach using a mixed effects model proved challenging in terms of software requirements. In the next two sections I will present the two stages of statistical modelling followed by the results and their interpretation.

3.3.1 Modelling seasonal variation of salbutamol prescribing

I modelled the relationship between the log transformed salbutamol prescribing rate and temperature in order to account for the seasonal variation. I accessed the daily mean for temperature data, which I averaged monthly. There was only one monitoring station recording temperature in the study area, which was considered

representative, as no major differences in temperature were expected within the study area. Figure 3-9 graphically depicts the temperature data which followed a distinct seasonal pattern. Associations between salbutamol prescribing and temperature were tested for different time lags. I looked at the relationship of monthly average temperature recorded in the same month as salbutamol prescribing (lag0) as well as that preceding by 7 days (lag7), 14 days (lag 14), 21 days (lag21) and a month (lag30).

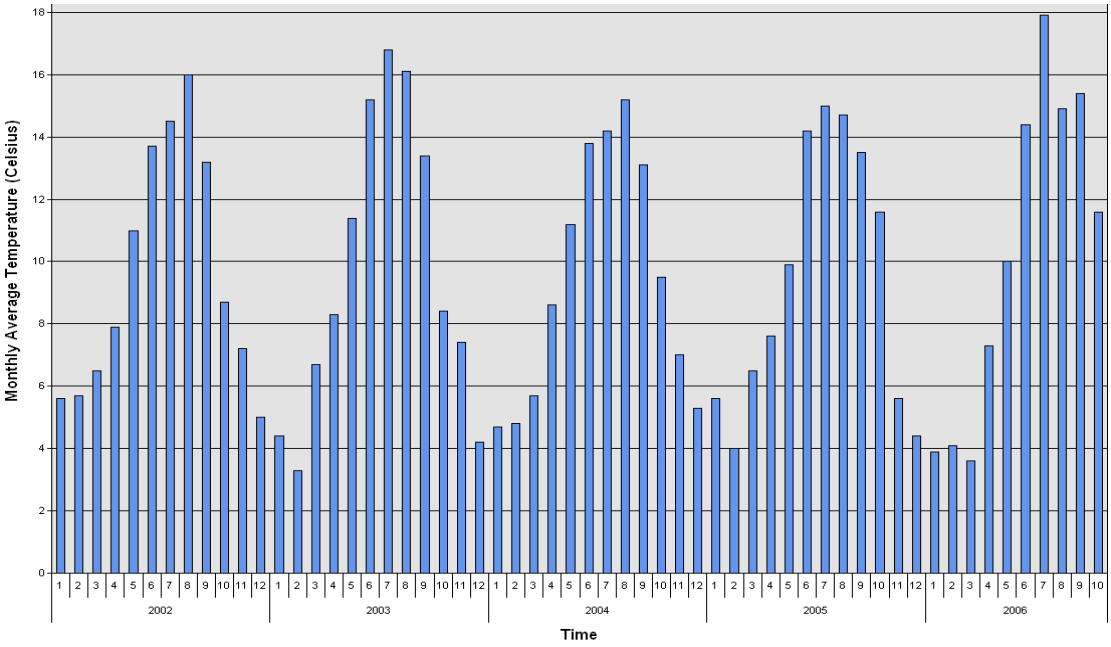


Figure 3-9 Monthly average temperature.

3.3.1.1 Method

Frequency domain models are one of the standard ways to model temporal autocorrelation (Gelman, 1995). Such models analyse data with respect to frequency rather than time, using a mixture of sinusoids at different frequencies. Frequency is the number of occurrences of a repeated observation by any given unit of time. There are alternative perspectives on capturing correlation and one of them, used by such models, is to transform data into vectors and assess the angle between them.

Variables presented as vectors in a space are commonly used in the physical sciences and engineering for plotting forces and visualizing their relationships. Vectors pointing in the same direction have a positive relationship and those pointing in opposite directions have a negative relationship. Applying the angles of

vectors as a measure of correlation can be complicated. The angles take values from 0^o to 360^o degrees, while the correlation coefficient take values between 1.00 for perfect correlation, -1.00 for perfect negative correlation, and zero for no correlation. A transformation can be done by using basic functions of trigonometry, sine and cosine. For instance, the cosine of the angle between vectors will be +1.00 for vectors with an angle of 0^o, -1.00 for an angle of 180^o (completely opposite directionality), and 0 for an angle of 90^o or 270^o.

I have used a type of frequency domain models, called harmonic regression or fixed frequency effects models, to estimate area wide monthly salbutamol prescribing in relation to temperature. Harmonic regression can be applied to stationary or non-stationary time series (Young et al., 1999). More details on frequency domain models and harmonic regression models can be found in relevant literature (Artis M. et al., 2007, Gelman, 1995, Kay and Marple, 1981, Chiu, 1989, Hannan, 1971, Bujosa et al., 2007, Young et al., 1999).

3.3.1.2 Analysis

The co-variation of prescribing between months as captured by sine and cosine was added as a variable in the model. I defined twelve periods as $\omega=2\pi/12$ to represent an annual cycle. The initial version of the model I used is:

$$\bar{Y}_t = \alpha + \beta_t + \gamma d_t + \mathrm{A}\cos(\omega t) + \mathrm{B}\sin(\omega t) + E_t \qquad 3\text{-}11$$

where $\bar{Y}_t$ is the monthly average log-transformed salbutamol prescribing rate for month t. The value d_t is the average of the daily mean temperature readings in month t. α, β, γ, A and B are parameters and E_t are independent residuals $N(0,\sigma_E^2)$ for month t. All the parameters of this model were static (fixed) taking constant values.

An improvement of this model was to replace the A and B parameters by variables that were estimated randomly. This second model was called a dynamic model and differed from the static harmonic regression model in the estimation of parameters in the frequency domain. I used the package "sspir" in R software to fit the dynamic harmonic regression model (Dethlefsen and Lundbye-Christensen, 2006). This method has been previously implemented in an epidemiological context for estimating weekly area-wide black smoke averages (Fanshawe et al., 2008).

In the dynamic model the parameters α, β, γ and E_t were as before, but the parameters A, and B were replaced by variables that were estimated by the random walks $A_t \mid A_{t-1} \sim N(A_{t-1}, \sigma_A^2)$ and $B_t \mid B_{t-1} \sim N(B_{t-1}, \sigma_B^2)$ (Fanshawe et al., 2008):

$$\overline{Y}_t = \alpha + \beta_t + \gamma d_t + A_t \cos(\omega t) + B_t \sin(\omega t) + U_t \qquad 3\text{-}12$$

I evaluated possible time-lags of 1 month and 21, 14 and 7 days for responses of medication prescribing to temperature, as mentioned earlier. Consequently, I modelled the area wide monthly average log-transformed salbutamol prescribing rate, $\mu_t = E\left[\overline{Y}_t\right]$, against temperature using five static and five dynamic frequency domain models. The results of the static models were compared to the results of dynamic harmonic regression models.

3.3.1.3 Results

The rates of prescribing as estimated from the static and dynamic modesl were plotted against the monthly log prescribing. Figure 3-10, Figure 3-11, Figure 3-12, Figure 3-13 and Figure 3-14 depict the results of fitted static and dynamic harmonic regression models for the area-wide average of prescribing in relation to temperature at the following time lags: a) 0 days (lag0), b) 7 days (lag7), c) 14 days (lag14), d) 21 days (lag21) and e) 1 month (lag30). The observed monthly prescribing is illustrated as an orange solid line. The estimation of static models (dotted line) differed substantially from the observed prescribing, indicating that these models were not adequate. In contrast, the dynamic models (solid blue line) captured much better the month-to-month variation. Out of the five dynamic models the one that used the temperature at 7 days lag, provided the best fit to the data, as depicted in Figure 3-11.

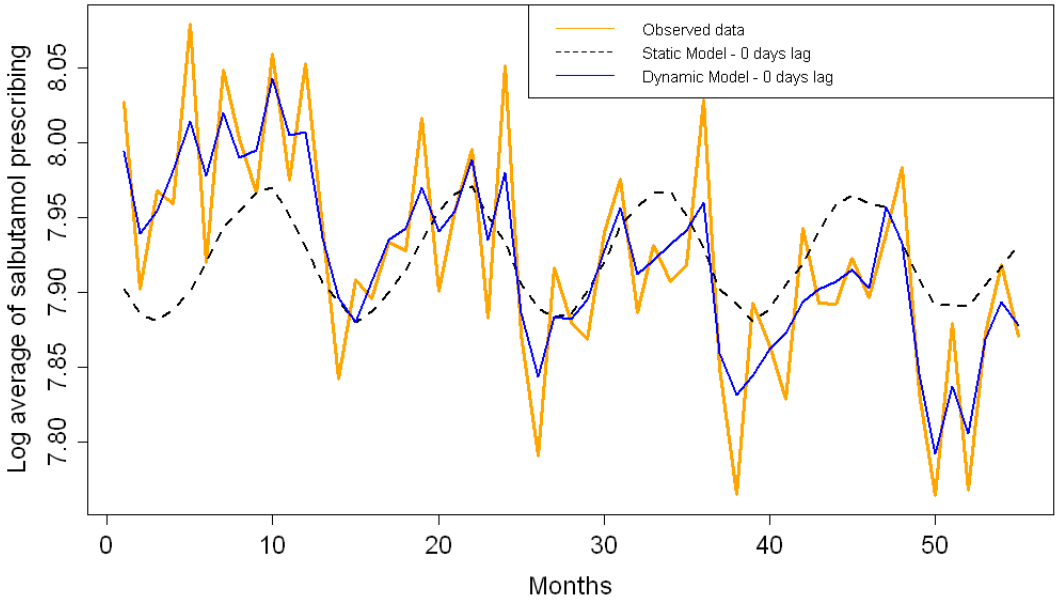


Figure 3-10 Fit of static and dynamic regression models for the area-wide average of prescribing in relation to temperature at 0 days lag.

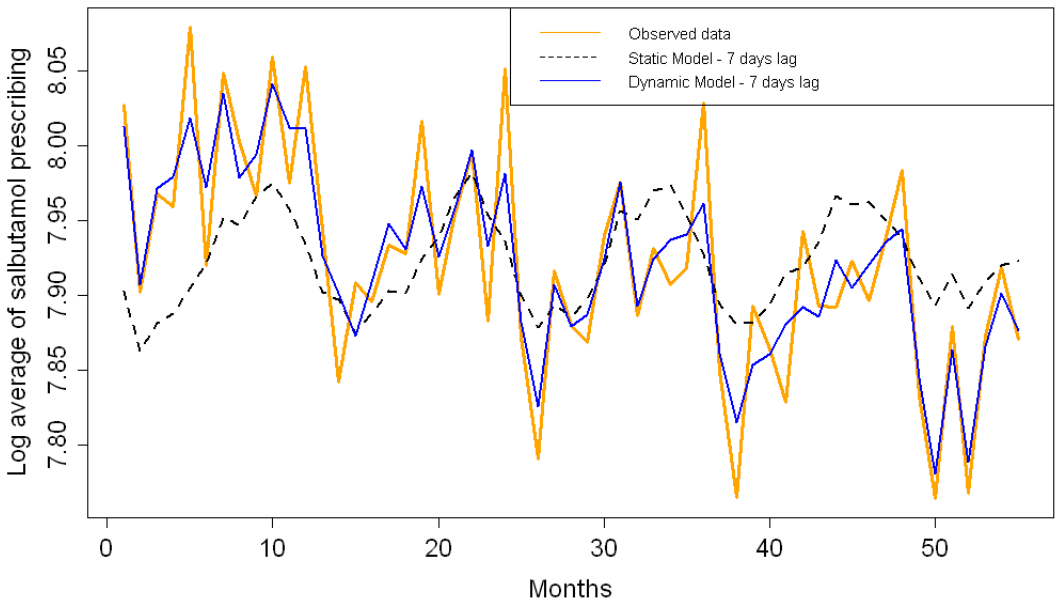


Figure 3-11 Fit of static and dynamic regression models for the area-wide average of prescribing in relation to temperature at 7 days lag.

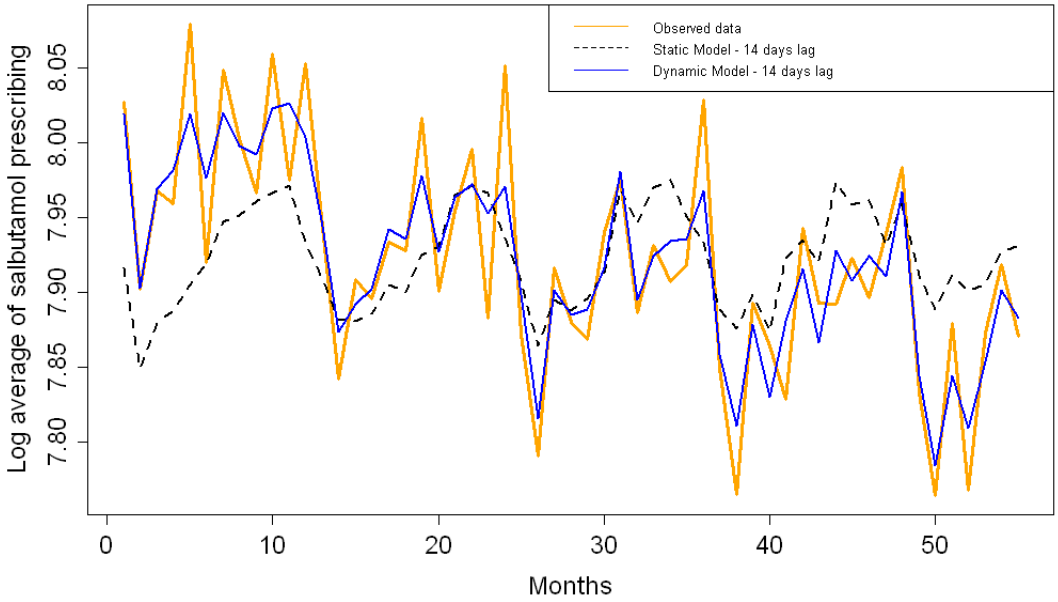


Figure 3-12 Fit of static and dynamic regression models for the area-wide average of prescribing in relation to temperature at 14 days lag.

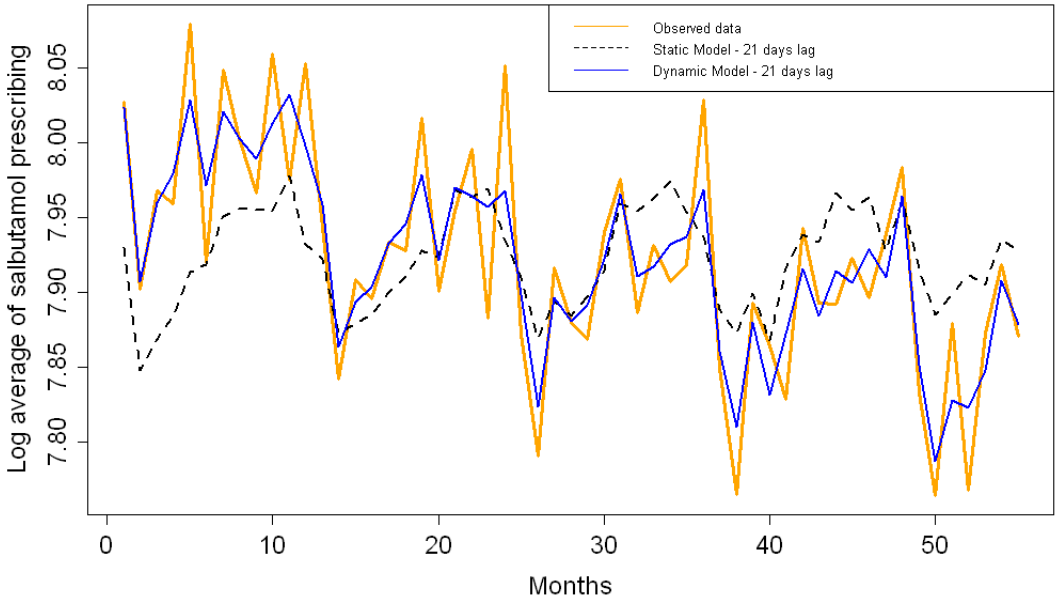


Figure 3-13 Fit of static and dynamic regression models for the area-wide average of prescribing in relation to temperature at 21 days lag.

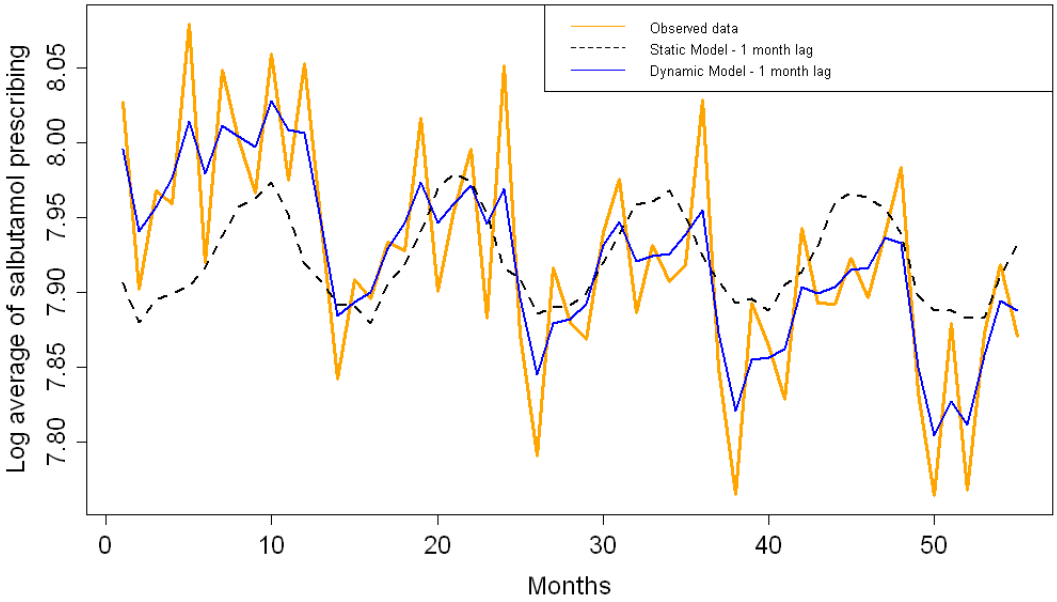


Figure 3-14 Fit of static and dynamic regression models for the area-wide average of prescribing in relation to temperature at 1 month lag.

I calculated the R squared (R^2) coefficient of determination, a statistical measure of how well the regression line approximated the real data points. An R^2 of 1.0 indicates that the regression line perfectly fits the data. Figure 3-15 shows that the model with 7 days lag had the highest R squared value (74,5%). Therefore I considered it to be best out of the five in terms of how well it fitted the original data.

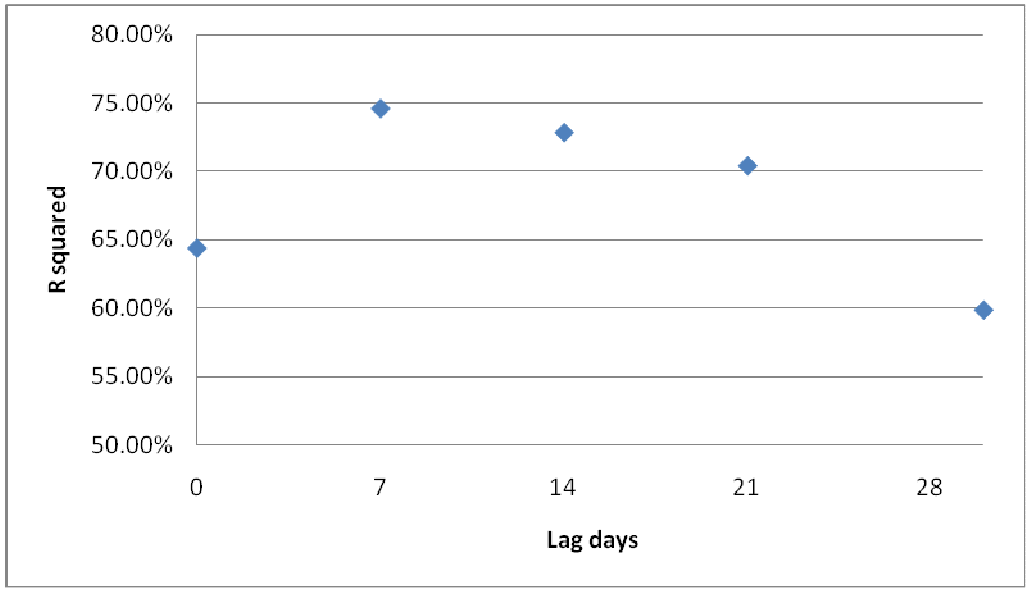


Figure 3-15 R squared for dynamic models.

3.3.1.4 Model Criticism

I examined the residuals of the dynamic model with lag7 (temperature data recorded 7 days preceding the first day of prescribing month), in order to assess the validity of the model.

Normality, Linearity, Homoscedasticity Assumptions

I created a normal Q-Q plot of the residuals, using the dynamic harmonic regression model (lag7). FiFigure 3-17 shows a pattern that is close to a straight line, demonstrating that the normality assumption was met by the model.

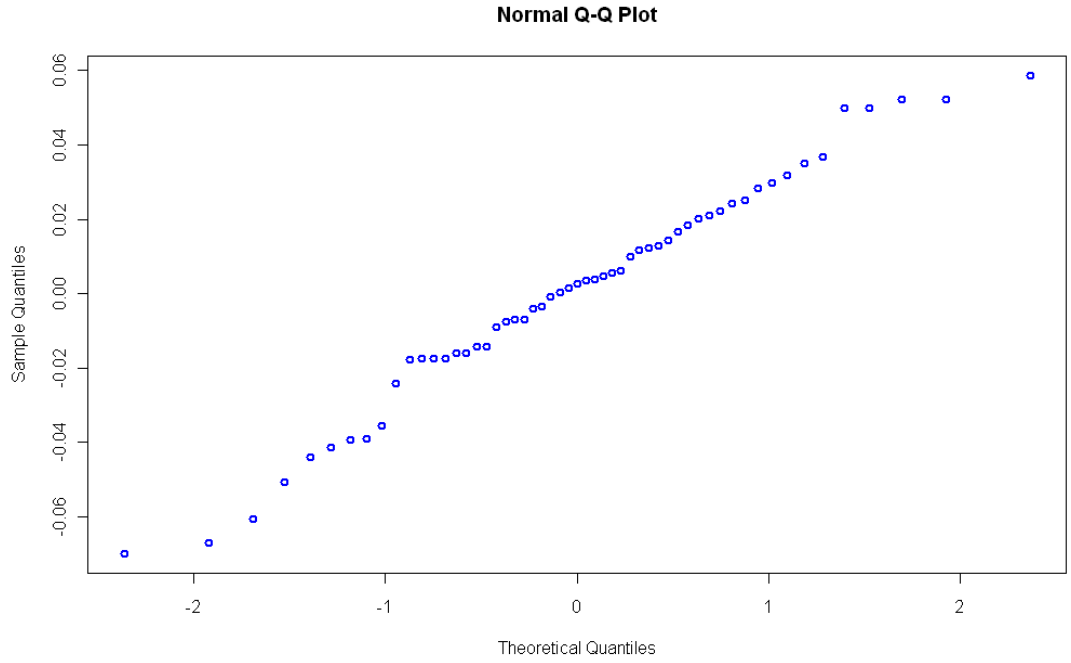
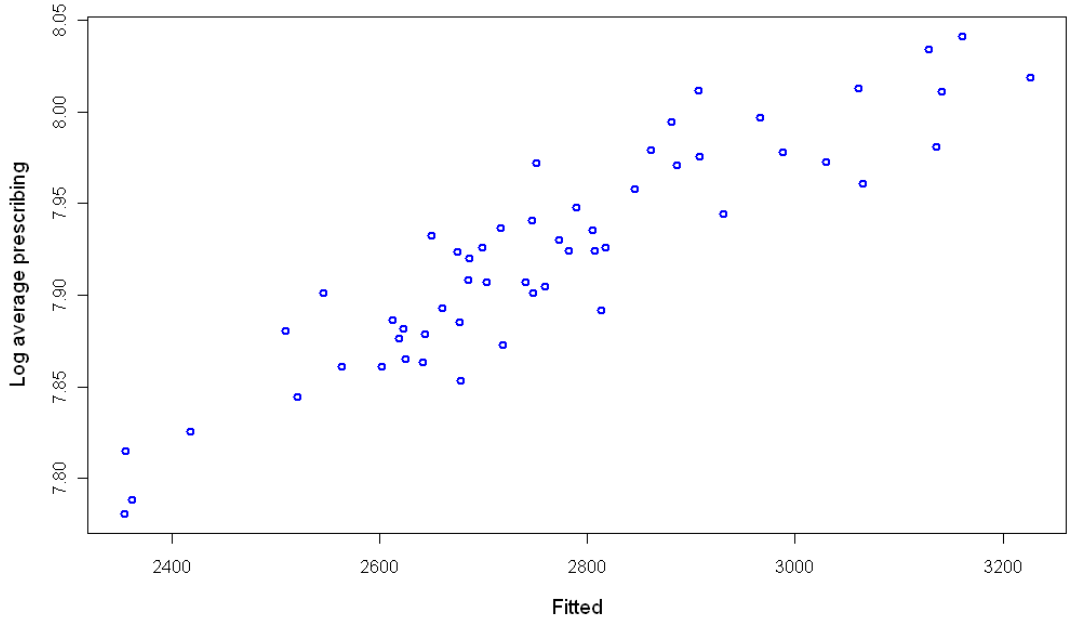


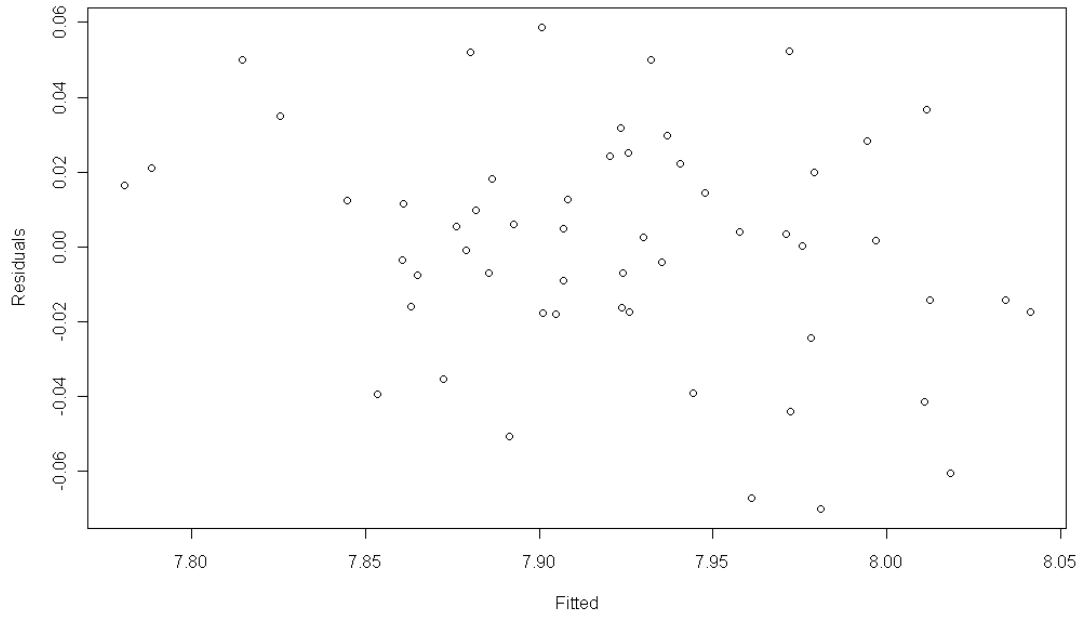
Figure 3-16 Normality Plot of residuals of dynamic harmonic regression model.

I assessed whether the assumption of linearity was met. By plotting observed versus fitted values of salbutamol prescribing (FiFigure 3-17), adequately it was met. The plot of residuals versus fitted values (FiFigure 3-18) was also examined to assess the linearity and homoscedasticity assumptions. Homoscedasticity was also examined by plotting the residuals versus time. The observed versus fitted values (FiFigure 3-17) were symmetrically distributed around a diagonal line showing the linearity assumption was adequately met. FiFigure 3-18 illustrates that the residuals remained constant in relation to fitted values and time (Figure 3- 21) indicating that the homoscedasticity (constant variance) assumption was also adequately met.



Fi

Figure 3-17 Plot of observed versus the fitted values of salbutamol prescribing of dynamic harmonic regression model.



Fi

Figure 3-18 Plot of residuals versus fitted values.

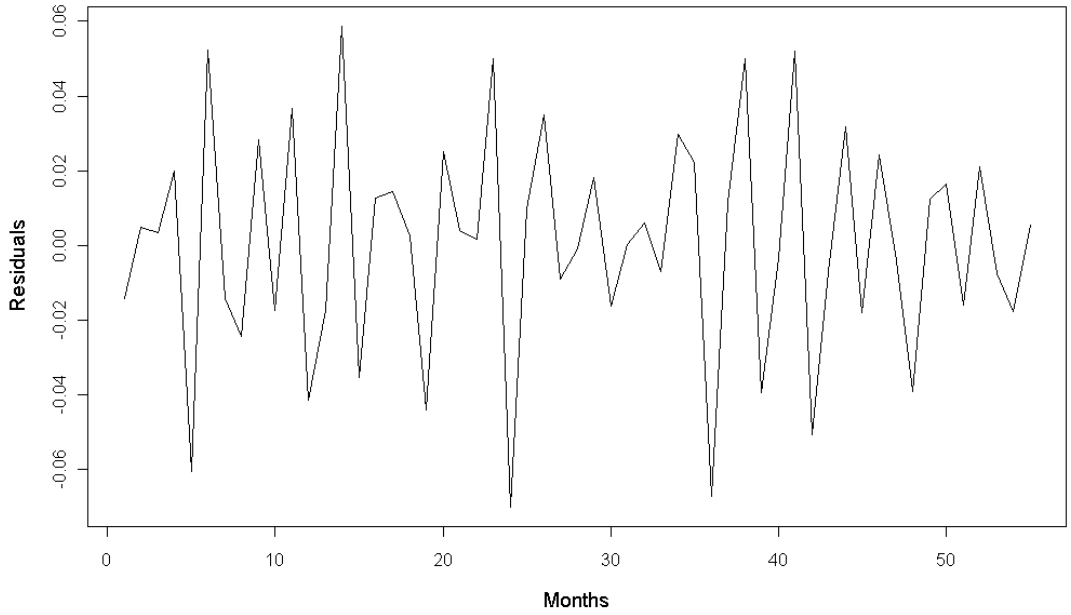


Figure 3-19 Plot of residuals versus time.

Independence Assumption

I examined whether the residuals of the dynamic harmonic regression model (lag7) met the independence assumption and compared that to the respective static model. I plotted the autocorrelation function to compare the temporal autocorrelation of residuals produced by the two models. The concept of autocorrelation functions was explained in Section 3.2.1.2, where I checked the reliability of the preliminary mixed-effects model.

The dynamic model improved the temporal autocorrelation of residuals (Figure 3-21), compared to the static model (Figure 3-20), as the residuals showed an even distribution around zero, within the 95% confidence limit (dotted lines) for most lags.

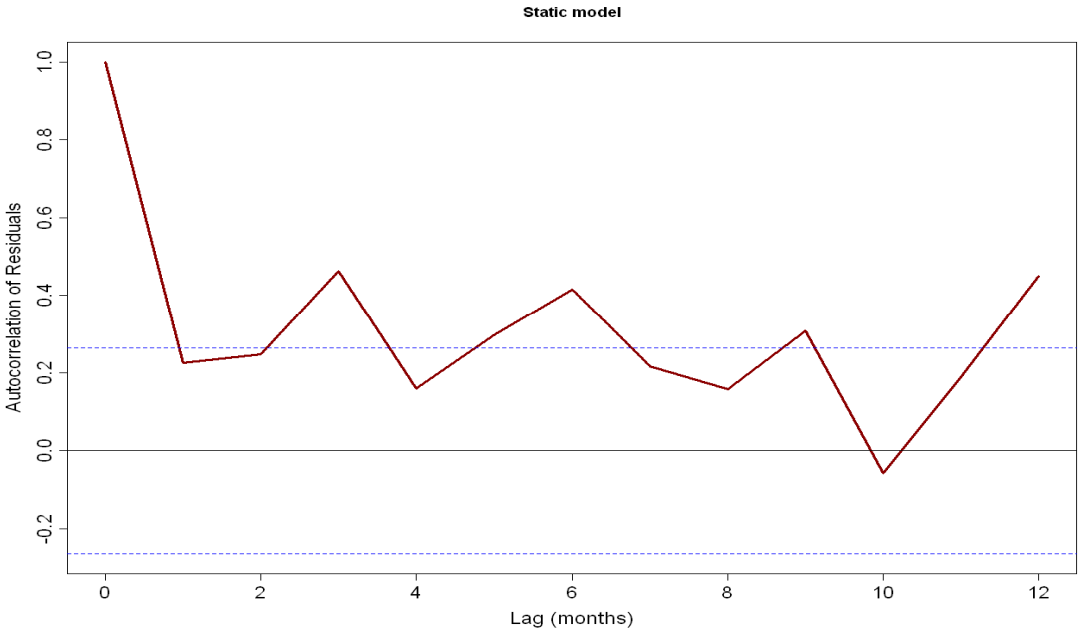


Figure 3-20 Temporal autocorrelation of residuals of the static model.

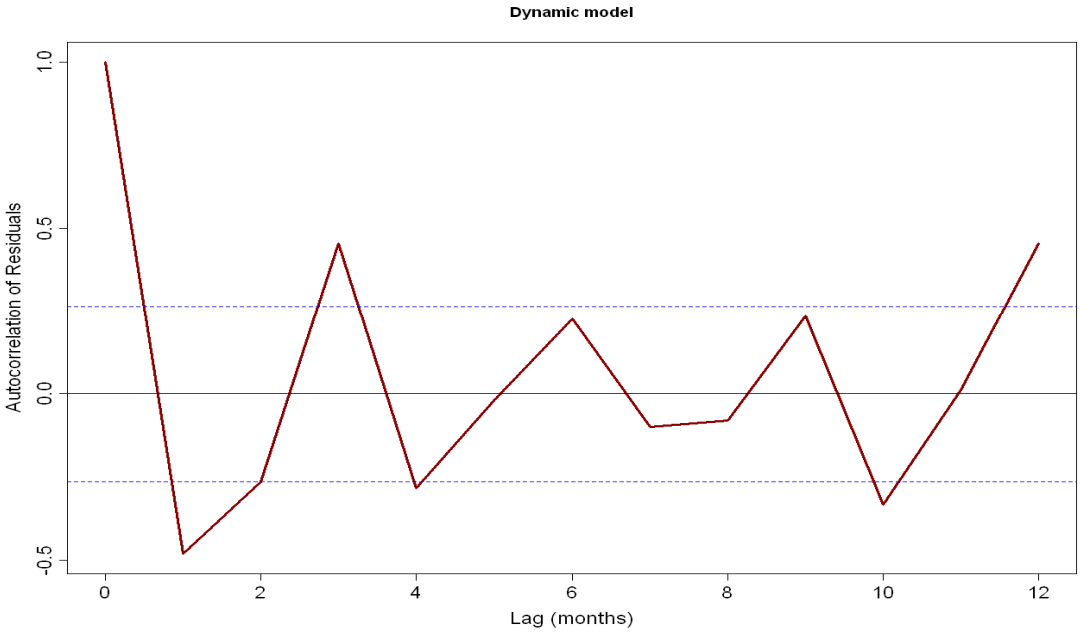


Figure 3-21 Temporal autocorrelation of residuals of the dynamic model.

Overall Figure 3-16, Figure 3-19, Figure 3-21 and Figure 3-23 illustrate that assumptions of linearity, normality, and homoscedasticity were met by the dynamic harmonic regression model.

3.3.1.5 Discussion/Conclusions

Results

I evaluated possible time-lags of 1 month and 21, 14 and 7 days, in response of medication prescribing to temperature. This was due to the fact that a delay was expected between the event that triggered the increased consumption of salbutamol, causing a need for a new inhaler, and thus a visit to the GP for a new prescription. The dynamic harmonic regression model captured the seasonal periodicity of salbutamol prescribing, in contrast to the static model which failed to capture it. Out of the five dynamic models, the one that used temperature at 7 days lag captured best the temporal variation in salbutamol prescribing. The models that used temperature data measured 14 (lag14) and 21 (lag21) days before the prescribing month, also captured salbutamol variation as well as the lag7 model. All of those three models performed well almost indicating that temperature recorded within a time window of 7-21 days preceding the prescribing month was a good explanatory variable of asthma and COPD exacerbations due to seasonal factors.

The dynamic model with lag7 performed better than the other models and therefore was assessed further in order to examine the reliability of it results. The model's residuals met the normality, linearity and homoscedasticity assumptions of linear regression. The residuals were also assessed for interdependence in time (temporal autocorrelation) but not in space (spatial autocorrelation). This is because temperature was considered spatially homogenous in the study and consequently this first stage of the model ignored any spatial variation of the covariate. The second stage of the model, which is presented in the next section, included covariates with both spatial and temporal variability and therefore the model's residuals were checked for both temporal and spatial autocorrelation. Overall, the dynamic harmonic regression model with 7 days lag performed well and was considered suitable for estimating the area-wide average of salbutamol prescribing attributable to seasonality. The output of this model was used as input on the final model that included all covariates, described in the section 3.4.

Potential Future Applications

The output of this model could also be used in isolation from the rest of the analysis. The model captured 75% of the prescribing variation and after some additional robust validation could be used for predicting salbutamol prescribing in relation to seasonal factors. This predictive model could be used for management of resources at the primary care level. Given that the model validation proved reliable,

similar models could also be developed for other types of medication or health outcomes that follow a seasonal pattern.

Other studies that assessed the impact of air pollution on prescribing of short-acting b2-agonists used data on pollen and influenza epidemics in addition to weather conditions in order to account as best possible for the seasonal variation of prescribing. Using a dynamic harmonic regression model may appear complicated. However collecting, editing and manipulating data on pollen and respiratory infections is also a complicated and time consuming process. Overall, I consider that using temperature data with a dynamic harmonic regression model, provided an efficient and pragmatic approach to capturing the seasonal variation of salbutamol prescribing. It might also be useful when analysing other asthma or COPD health indicators, such as hospital admissions and emergency room visits, which are also subject to seasonal variation.

3.4 Model Refining (Stage B)

In the last stage of statistical model development, I fitted linear mixed-effects models with an offset. The offset captured the seasonal variation of log salbutamol prescribing rate as described in previous chapter.

3.4.1 Modelling Spatio-temporal Variation of Salbutamol Prescribing

The final model included all covariates: 1) PM_{10}, 2) Traffic index, 3) Income deprivation, 4) Employment deprivation, 5) Educational Deprivation, 6) Average age of patients registered with GP practices and 6) sex (males/females) ratio of patients registered with GP practices.

I tested the influence of air quality indicators (mean 24-h concentrations for PM_{10} and monthly traffic flows per GP service area) observed in the same month as the prescribing (lag0), and then with a variety of lag periods. I evaluated monthly average PM_{10} levels observed 7 days (lag 7), 14 days (lag 14), 21 days (lag 21) and 1 month (lag 30) before the first day of each prescribing month. The traffic flow data were provided with a monthly temporal resolution, therefore it was not possible to create weekly time lags as in the case of air pollutants. I created time-lag of one month (lag30) based on data observed in the month before the prescribing month. The estimation of air quality covariates for different time points was important because a lag period was expected between exposure of individuals and prescribing. I was interested to examine whether my models would capture that. I was also interested to examine which if any lag period would have a statistically significant association with prescribing.

A summary of the descriptive statistics is presented in Table 3-1 for all covariates at different lag times. The mean, minimum and maximum values were estimated. However the median, 1^{st} and 3^{rd} quintile were more informative about the data distribution.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
PM10_lag0days	13.2	16.1	18.0	18.9	20.9	29.0
PM10_lag7days	11.7	16.0	18.8	18.9	20.8	33.9
PM10_lag14days	12.1	16.0	17.7	18.6	20.1	35.5
PM10_lag21days	12.5	16.0	18.0	18.7	20.8	31.0
PM10_lag30days	13.2	16.1	17.7	18.7	20.8	29.0
Traffic_lag0days	191.9	496.4	781.8	905.0	1268.5	3235.0
Traffic_lag30days	191.9	495.4	781.8	903.9	1267.5	3235.0
Income Deprivation	0.1	0.1	0.2	0.2	0.3	0.5
Employment Deprivation	0.1	0.1	0.2	0.2	0.2	0.3
Educational Deprivation	4.1	17.5	28.9	30.2	42.4	70.8
Average age	29.0	38.0	39.0	39.0	41.0	46.0
Sex ratio	0.8	0.9	1.0	1.0	1.0	1.8

Table 3-1. Descriptive statistics of covariates.

3.4.1.1 Method

I fitted the models with all covariates. The mixed effects model took the form:

$$y_{ij} = \mu + a_1 x_{ij} + \ldots + a_8 x_{ij} + b_1 z_{ij} + \ldots + b_{64} z_{ij} + \varepsilon_{ij}$$

$b_i = N(0, \sigma^2_b)$ *and*

$$\varepsilon_{ij} = \mathrm{N}(0, \sigma^2)$$

where y_{ij} is the log transformed salbutamol prescribing rate for observations j in group i, μ is the population-average prescribing of salbutamol (intercept), α are eight fixed effect parameters respective to x=1,…,8 covariates. The random variable ε_{ij} is the error term for jth observation of the *ith* practice, assumed independent and identically distributed, N (0, σ^2). The b_i are the random effects also assumed to have come from a normal distribution N (0, σ^2).

In cases where both GP practices and PCTs were introduced as random effects the vector of random effects b consisted of the 64 practice parameters plus 128 PCT-practice interaction parameters b_{65}, b_{66}, …, b_{192}. I fitted models with both types of random effects and compared them.

3.4.1.2 Analysis

I fitted separate models for each combination of time lags for which the air quality indicators were estimated. These combinations produced 10 mixed-effects models which I denoted as:

1) **L30PM_30TR**: 1 month lag for PM_{10} and 1 month for traffic flows

2) **L21PM_30TR**: 21 days lag for PM_{10} and 1 month for traffic flows

3) **L14PM_30TR**: 14 days lag for PM_{10} and 1 month for traffic flows

4) **L7PM_30TR**: 7 days lag for PM_{10} and 1 month for traffic flows

5) **L0PM_30TR**: 0 days lag for PM_{10} and 1 month for traffic flows

6) **L30PM_0TR**: 1 month lag for PM_{10} and 0 days for traffic flows

7) **L21PM_0TR**: 21 days lag for PM_{10} and 0 days for traffic flows

8) **L14PM_0TR**: 14 days lag for PM_{10} and 0 days for traffic flows

9) **L7PM_0TR**: 7 days lag for PM_{10} and 0 days for traffic flows

10) **L0PM_0TR**: 0 days lag for PM_{10} and 0 days for traffic flows

Random effects in the above models were specified for GP practices. I also considered grouping GP practices within PCTs, as PCTs are the administrative bodies of GPs and I wanted to examine whether there was an effect on prescribing due to such grouping. Consequently, I fitted 10 additional models by grouping the data both by GP practices and PCTs. I added the ending "pct" to the model descriptions (i.e. L30PM_30TR_pct) to denote that they had random effects assigned to both GP practice and PCT.

3.4.1.3 Results

The results of the final models with GP practice random effects are summarised in Table 3-2. A few models returned identical results in cases where PM_{10} was not significant in specific time-lags, the results of which are also presented in Table 3-2. For instance the models L30PM_30TR, L7PM_30TR and L0PM_30T produced identical results as PM_{10} was not statistically significant in 30-, 7- and 0- days lag, and consequently, the coefficients of the statistically significant predictors were identical.

TRAFFIC 30 DAYS LAG				
L21PM_30TR	Estimate	Std.Error	t value	Pr(>|t|)
(Intercept)	-2.01821	0.15007	-13.447986	<0.0001
PM_{10}_lag21days	0.00102	0.00047	2.185799	0.0288
Traffic_lag30days	0.00007	0.00002	3.032136	0.0024
Time (months)	-0.00066	0.00012	-5.629881	<0.0001
Income Deprivation	2.81133	0.30695	9.158998	<0.0001
Average Age of Patients	0.03296	0.00316	10.439382	<0.0001
L14PM_30TR	Estimate	Std.Error	t value	Pr(>|t|)
(Intercept)	-2.02196	0.15014	-13.46709	<0.0001
PM_{10}_lag14days	0.00109	0.00045	2.411585	0.0159
Traffic_lag30days	0.00007	0.00002	3.138812	0.0017
Time (months)	-0.00067	0.00012	-5.660241	<0.0001
Income Deprivation	2.80937	0.30721	9.144675	<0.0001
Average Age of Patients	0.03297	0.00316	10.44511	<0.0001
L30PM_30TR - L7PM_30TR - L0PM_30TR	Estimate	Std.Error	t value	Pr(>|t|)
(Intercept)	-1.99854	0.14987	-13.33504	<0.0001
Traffic_lag30days	0.00007	0.00002	3.030806	0.0024
Time (months)	-0.00067	0.00012	-5.679115	<0.0001
Income Deprivation	2.81650	0.30700	9.174114	<0.0001
Average Age of Patients	0.03292	0.00316	10.42185	<0.0001
TRAFFIC 0 DAYS LAG				
L21PM_0TR	Estimate	Std.Error	t value	Pr(>|t|)
(Intercept)	-2.02922	0.15001	-13.52753	<0.0001
PM_{10}_lag21days	0.00093	0.00047	1.975763	0.0482
Traffic_lag0days	0.00009	0.00002	3.677388	0.0002
Time (months)	-0.00067	0.00012	-5.68998	<0.0001
Income Deprivation	2.81647	0.30843	9.131577	<0.0001
Average Age of Patients	0.03294	0.00316	10.43984	<0.0001
L14PM_0TR	Estimate	Std.Error	t value	Pr(>|t|)
(Intercept)	-2.02735	0.14994	-13.52078	<0.0001
PM_{10}_lag14days	0.00090	0.00045	1.976244	0.0481
Traffic_lag0days	0.00009	0.00002	3.623905	0.0003
Time (months)	-0.00067	0.00012	-5.700385	<0.0001
Income Deprivation	2.81550	0.30830	9.13223	<0.0001
Average Age of Patients	0.03294	0.00316	10.43937	<0.0001
L30PM_0TR & L7PM_0TR & L0PM_0TR	Estimate	Std.Error	t value	Pr(>|t|)
(Intercept)	-2.01411	0.14991	-13.43565	<0.0001
Traffic_lag0days	0.00009	0.00002	3.795247	0.0001
Time (months)	-0.00068	0.00012	-5.750147	<0.0001
Income Deprivation	2.82119	0.30879	9.136207	<0.0001
Average Age of Patients	0.03291	0.00316	10.42655	<0.0001

Table 3-2. Results of p-values and coefficients' values from models with GP practice grouping as random effect.

The results of the models with GP practice and PCT-GP practice random effects are presented in Table 3-3. Again models that produced identical results, due to identical statistically significant covariates are presented only once.

TRAFFIC 30 DAYS LAG				
L21PM_30TR_pct	Estimate	Std.Error	t value	Pr(>|t|)
(Intercept)	-2.04707	0.16230	-12.61292	<0.0001
PM_{10}_lag21days	0.00102	0.00047	2.17111	0.0299
Traffic_lag30days	0.00008	0.00002	3.26859	0.0011
Time (months)	-0.00067	0.00012	-5.71480	<0.0001
Income Deprivation	2.95462	0.31390	9.41270	<0.0001
Average Age of Patients	0.03290	0.00315	10.43121	<0.0001
L14PM_30TR_pct	Estimate	Std.Error	t value	Pr(>|t|)
(Intercept)	-2.05158	0.16268	-12.61130	<0.0001
PM_{10}_lag14days	0.00110	0.00045	2.42894	0.0151
Traffic_lag30days	0.00008	0.00002	3.37794	0.0007
Time (months)	-0.00068	0.00012	-5.74695	<0.0001
Income Deprivation	2.95500	0.31411	9.40743	<0.0001
Average Age of Patients	0.03291	0.00315	10.43707	<0.0001
L30PM_30TR_pct & L7PM_30TR_pct & L0PM_30TR_pct	Estimate	Std.Error	t value	Pr(>|t|)
(Intercept)	-2.02768	0.16221	-12.50014	<0.0001
Traffic_lag30days	0.00008	0.00002	3.26898	0.0011
Time (months)	-0.00068	0.00012	-5.76545	<0.0001
Income Deprivation	2.96043	0.31396	9.42927	<0.0001
Average Age of Patients	0.03286	0.00316	10.41354	<0.0001
TRAFFIC 0 DAYS LAG				
L30PM_0TR_pct & L21PM_0TR_pct& L14PM_0TR_pct & L7PM_0TR_pct & L0PM_0TR_pct	Estimate	Std.Error	t value	Pr(>|t|)
(Intercept)	-2.04608	0.16450	-12.43783	<0.0001
Traffic_lag0days	0.00010	0.00002	4.04616	0.0001
Time (months)	-0.00069	0.00012	-5.84281	<0.0001
Income Deprivation	2.98172	0.31551	9.45056	<0.0001
Average Age of Patients	0.03284	0.00315	10.41621	<0.0001

Table 3-3. Results of log salbutamol prescribing rate from models with PCT and GP grouping as random effect.

The small p-values associated with the output of the models (Table 3-2 & Table 3-3) indicated that PM_{10} had a positive significant association with prescribing of salbutamol medication at primary care health level. PM_{10} levels on 14 and 21 days preceding the prescribing month were related significantly to respiratory prescribing. No influence of salbutamol prescribing rate was observed for PM_{10} concentrations occurring in the same month (lag 0) as salbutamol prescribing, or one month (lag30) or 7 days (lag7) earlier than prescribing.

The results also indicated that GP practices with different traffic conditions in their service area (defined as the area where 98% of their patients are expected to live) had significantly different prescribing patterns. The traffic flow index based on data observed in the month of prescribing as well as a month earlier to prescribing were statistically significant.

The small p-values also indicated that prescribing was negatively influenced by time, meaning that a decrease of salbutamol prescribing occurred during the study period. In addition, GP practices with different levels of income deprivation had significantly different prescribing. An increase in salbutamol prescribing was observed with increase in deprivation. Employment and educational deprivation had no significant relation to salbutamol prescribing.

The average age of patients registered with each GP practice also appeared to be associated with different prescribing patterns by GPs. The older the population a GP served, the higher the rate of salbutamol prescribing appeared to be. The ratio of males to females that GP practices served had a negative relationship with salbutamol prescribing (more male patients - more salbutamol prescribing) but it was not statistically significant.

The same covariates were identified as significant by the models that assigned GP practice or both PCT \ GP practice as random effects. The main difference was that only PM_{10} levels 21 days before prescribing (lag21), were significant for models with the PCT\GP grouping factor, while PM_{10} with both 21 and 14 days lag was significant in models with only GP grouping.

I referred to the models without PM_{10}, as reduced models because had fewer significant covariates (traffic, income, time, average age of patients) compared to those that included PM_{10} as a statistically significant covariate. In the next section, I compared the models presented in Table 3-2 and Table 3-3, in order to examine if they were statistically different from each other.

3.4.1.4 Model Comparison – Parsimonious Model

There is no absolute "correct" model. However, the best model is the simplest one out of those that most closely achieve the objectives of the study. Comparison of the statistical models I had developed was required, in order to identify the most parsimonious. Particulate matter appeared to be significant in some models but not in others. I compared only models that included particulate matter as an explanatory variable for estimating salbutamol prescribing, because that was the main measure of air pollution exposure in my analysis.

I carried out anova tests to compare mixed-effects models (Pinheiro and Bates, 2000). I grouped my models into those that had one level of grouping (GP) and those that had two levels (PCT\GP). I firstly assessed whether the models that shared the same grouping factor were significantly different from each other. Then, I compared the best models from each category to assess whether the grouping by GP or PCT\GP made them statistically different.

If the p-values on the anova output were high, the models were not statistically different. In such cases, it is suggested that the Akaike Information Criterion (AIC) (Akaike 1973, Skamoto et al, 1976) be used to identify the most parsimonious model (Pinheiro and Bates, 2000). The chosen model is the one for which the $AIC=-2l_{max}+2k$ reaches a minimum, where l_{max} is the log-likelihood maximum and k is the number of unknown parameters. The smaller the AIC, the better the model (Demidenko, 2004, Pinheiro J. et al., 2008).

Table 3-4 presents the anova tests conducted for models that had been assigned random effects on GP level, while Table 3-5 shows the results of anova tests for models with PCT\GP random effects. The small p-values show that the first model should be rejected in favour of the others. I repeated the process after removing the rejected model until there were no further models to reject as significantly different. Table 3-4 shows that the most parsimonious model when grouping the data by GP was the one that included the following covariates: a) PM_{10} 21 days before salbutamol prescribing, and b) traffic index based on traffic flows observed a month before prescribing. The same model emerged as the most parsimonious from the class of models that had two levels of random effects (PCT\GP), as shown in Table 3-5.

The models with 21 days and 1 month latency period for PM_{10} and traffic index respectively proved more parsimonious under both types of groupings used: i) GP level (model L21PM_30TR) and ii) PCT\GP level (model L21PM_30TR_pct). I finally used the anova test again to assess if those two final models were significantly different. High p-values in anova output (Table 3-6) indicated that I could not reject the L21PM_30TR model in favour of the L21PM_30TR_pct, meaning there was no significant difference between them.

A judgment on which was more parsimonious should be based on AIC criterion since the p-values were high. The AIC was slightly smaller for the latter model. However, the difference between those values was still marginal. In the next section (3.4.2), I examined those two models in detail and critically evaluated them in order to reach conclusions regarding the validity of their results (Table 3-2, Table 3-3).

Models with GP grouping factor	Df	AIC	BIC	logLik	Chisq Chi	Df		Pr(>Chisq)	
L14PM_0TR	8	-5374.1	-5324.8	2695.1					
L14PM_30TR	8	-5370.9	-5321.6	2693.4	0	0	<	2.20E-16	***
L21PM_0TR	8	-5374.1	-5324.8	2695.1	3.2189	0	<	2.20E-16	***
L21PM_30TR	8	-5369.9	-5320.5	2692.9	0	0	<	2.20E-16	***
L14PM_30TR	8	-5370.9	-5321.6	2693.4					
L21PM_0TR	8	-5374.1	-5324.8	2695.1	3.2189	0	<	2.20E-16	***
L21PM_30TR	8	-5369.9	-5320.5	2692.9	0	0	<	2.20E-16	***
L21PM_0TR	8	-5374.1	-5324.8	2695.1					
L21PM_30TR	8	-5369.9	-5320.5	2692.9	0	0	<	2.20E-16	***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1									

Table 3-4. Comparison of models grouped per GP.

Models with PCT\GP grouping factor	Df	AIC	BIC	logLik	Chisq Chi	Df		Pr(>Chisq)	
L14PM_30TR_pct	9	-5373,7	-5318,2	2695,8					
L21PM_30TR_pct	9	-5372,5	-5317	2695,2	0	0	<	2,20E-16	***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1									

Table 3-5. Comparison of models grouped per PCT and GP.

Models with GP vs PCT\GP grouping factor	Df	AIC	BIC	logLik	Chisq Chi	Df	Pr(>Chisq)
L21PM_30TR	8	-5369.9	-5320.5	2692.9			
L21PM_30TR_pct	9	-5372.5	-5317	2695.2	0	1	1

Table 3-6. Comparison of models grouped per GP versus model grouped per PCT & GP.

3.4.2 Model Criticism

My analysis had produced two statistical models which I needed to examine critically, in order to be able to comment on the reliability of their results. Firstly, I inspected the structure of their random effects. I then examined the fixed effects by checking whether the residuals of the model met the four assumptions of linear regression: 1) normality, 2) linearity, 3) homoscedasticity and 4) independence. I examined both temporal and spatial independence. Model evaluation was a very important part of the statistical modelling process as it allowed assessment of the reliability of results produced by those models.

3.4.2.1 Random Effects Structure

The two final models included particulate matter with 21 days lag and the traffic index with 30 days lag. All fixed effects (covariates) were common to these two models but they had different random effects. The first model had GP level as a random effect (L21PM_30TR) while the second had two levels of random effect; PCT and GP (L21PM_30TR_pct). Based on anova results (Table 3-6) these two models were not statistically different and I could not reject the former in favour of the latter or vice versa.

Model with PCT and GP random effects

I then examined the structure of their random effects. I extracted the random effects for each grouping factor and summarised the random effects structure in graphical format. Figure 3-22 shows the value of random effects by PCT as well as the 95% prediction intervals associated with this value.

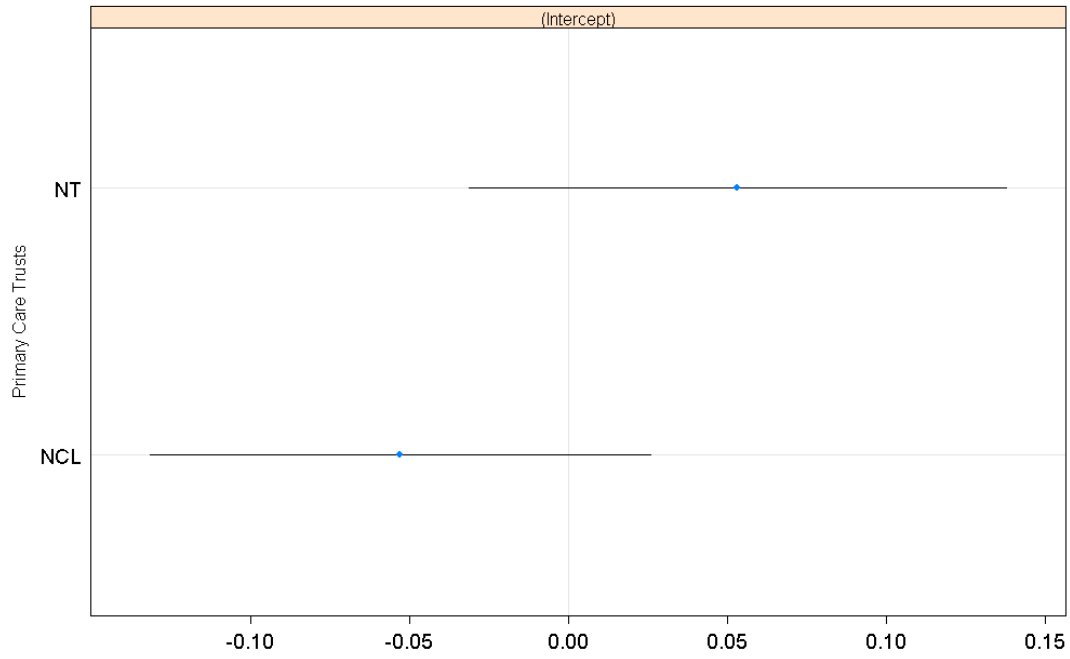


Figure 3-22 Plot of the random effects for Newcastle (NCL) and North Tyneside (NT) PCTs and their respective confidence intervals

Each set of prediction intervals (Figure 3-22) had constant width because of the balance of the data used in this study. The prediction intervals for the PCTs' random effects overlapped zero and each other, denoting that the effect of PCTs was not statistically significant. This meant that the effect of PCTs in salbutamol prescribing was likely to be zero. The fact that the intervals were wide also denoted great uncertainty in this prediction, showing that the model had not satisfactorily captured those effects. This was related to the very small number of PCTs used in this study.

Based on examination of random effects of the "L21PM_30TR_pct" model (Figure 3-22), I decided that the model failed to predict the random effects on PCT level reliably, and consequently, I could not accept this model's output.

Model with GP Random Effects

The structure of random effects of the remaining final model (L21PM_30TR) was then examined further by plotting again the random effects in the form of 95% prediction intervals, arranged in increasing order of conditional mean.

Figure 3-23 shows that the model satisfactorily captured the random effects of GP practices, with small prediction intervals associated with each one. Most of the

prediction intervals for the random effects did not overlap zero, meaning that the effect of GP practices was statistically significant. In other words, effect of GP practices was unlikely to be zero, or more accurately the effect found in this sample of 64 GP practices was not the sort of effect one would expect to see if there was no effect in the population from which this sample was drawn.

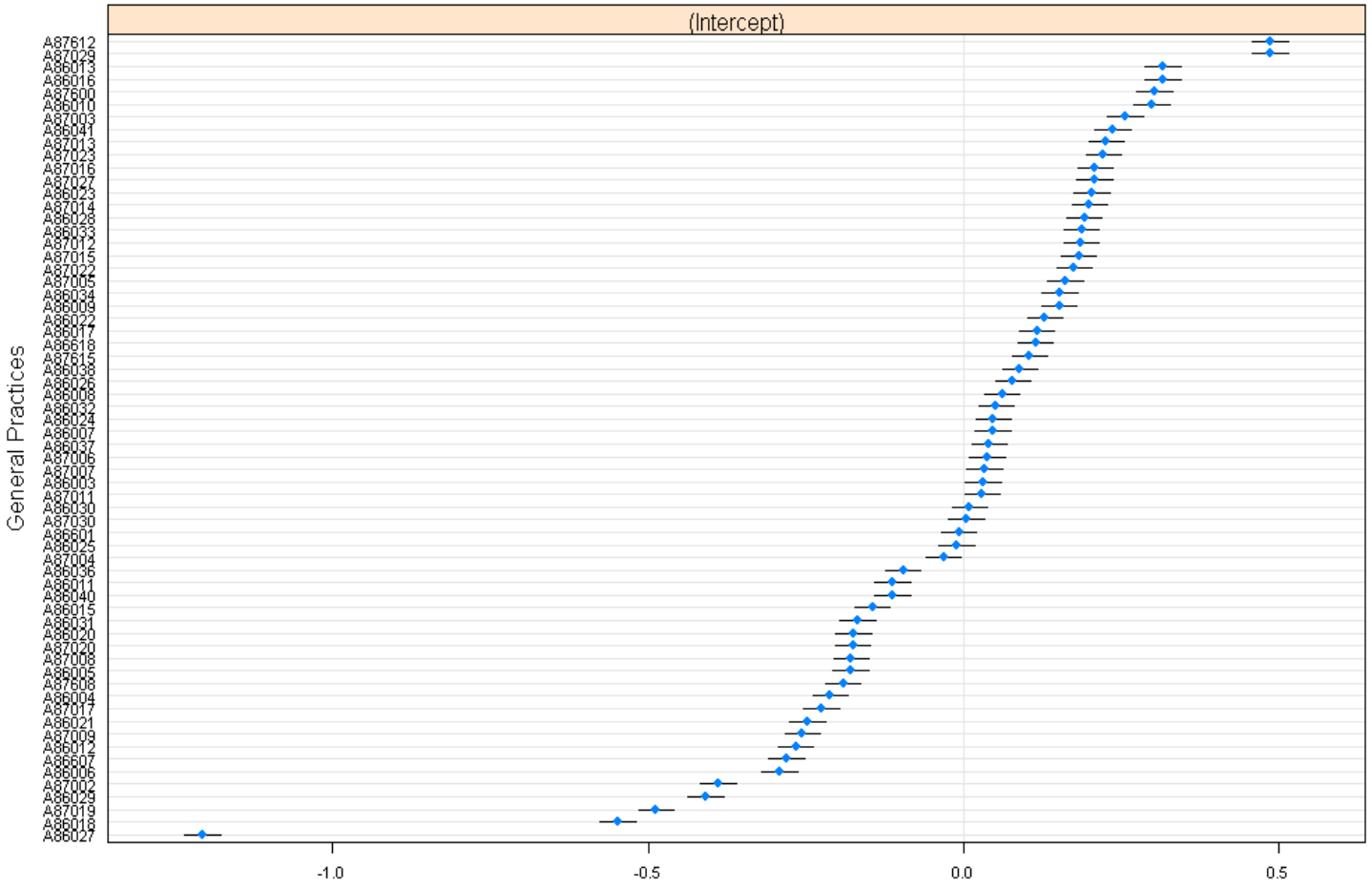


Figure 3-23. Plot of the random effects for 64 General practices and their respective confidence intervals.

For linear mixed models the conditional distribution of the random effects, given the data, written (B/Y=y), is a multivariate Gaussian distribution with zero mean. Since the random effects were assumed to follow a Gaussian distribution I created a QQ-plot to assess the normality of random effects in the final "L21PM_30TR" model.

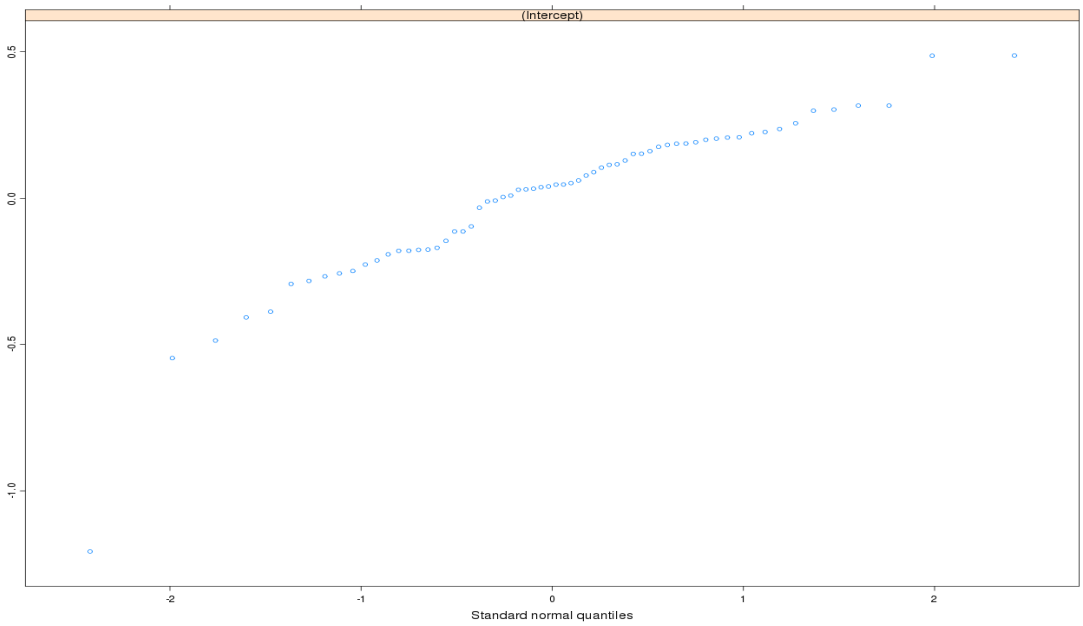


Figure 3-24. Normality plot of random effects of final model.

Figure 3-24 shows that the random effects satisfactorily approximated a straight line. An outlier was observed at the bottom of the graph that refers to the GP practice A86027, which is located within the campus of Newcastle University. This practice had the lowest prescribing rate of salbutamol and also a different demographic profile compared to other practices. A large number of its registered patients were young people and a significantly larger proportion of population of the registered patients were male. This was explained due to its proximity to the university, making it one of the two practices within Newcastle that have large numbers of students registered with them.

The very low prescribing rate of salbutamol in this practice was not considered to be an outlier caused by error in data recording. According to the NHS Information Centre data, this practice scores lower than the regional and national average for all main chronic diseases, including asthma and COPD (NHS The Information Centre for Health and Social Care, 2008).

3.4.2.2 Fixed Effects Structure

I checked the fixed effects as well, in order to ascertain whether the most parsimonious model "L21PM_30TR" I arrived at, provided a satisfactory fit to the data. I examined the model for normality, linearity, homoscedasticity and independence.

Normality, Linearity, Homoscedasticity Assumption

The best test for normally distributed errors is a normal probability plot of the residuals.

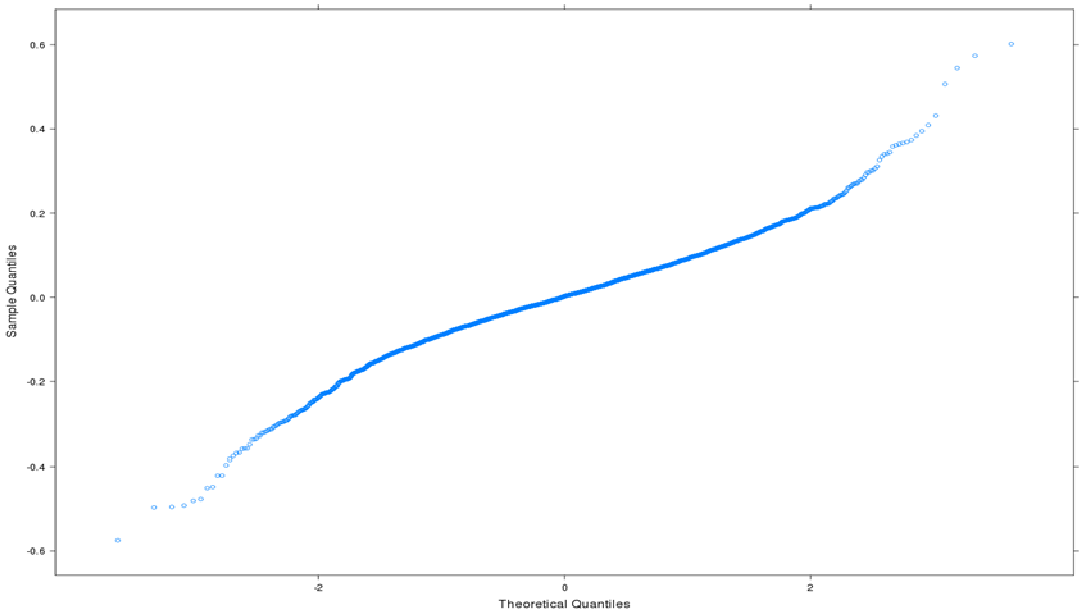


Figure 3-25. Normal probability plot of the residuals of the final model.

The normal probability plot of residuals was close to a straight line with slight curves at the edges. This indicated a longer tail for the lower and upper levels of salbutamol prescribing, suggesting the model was not adequate to capture variation at these levels. However because of the large number of observations (3,520 records) this deviation from normality was considered to have mild impact on the fitted values.

The standard plots for detecting linearity were plots of observed versus fitted values, or residuals versus fitted values. The latter plot was used to assess whether the homoscedasticity (constant variance) assumption was met, together with plots of residuals versus time. These three graphs are presented below.

.

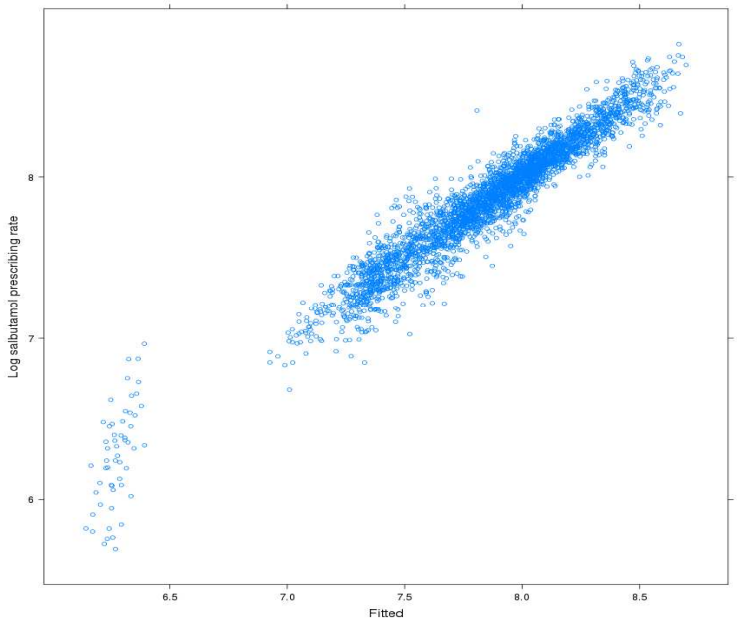


Figure 3-26. Plot of observed against fitted values of the final model.

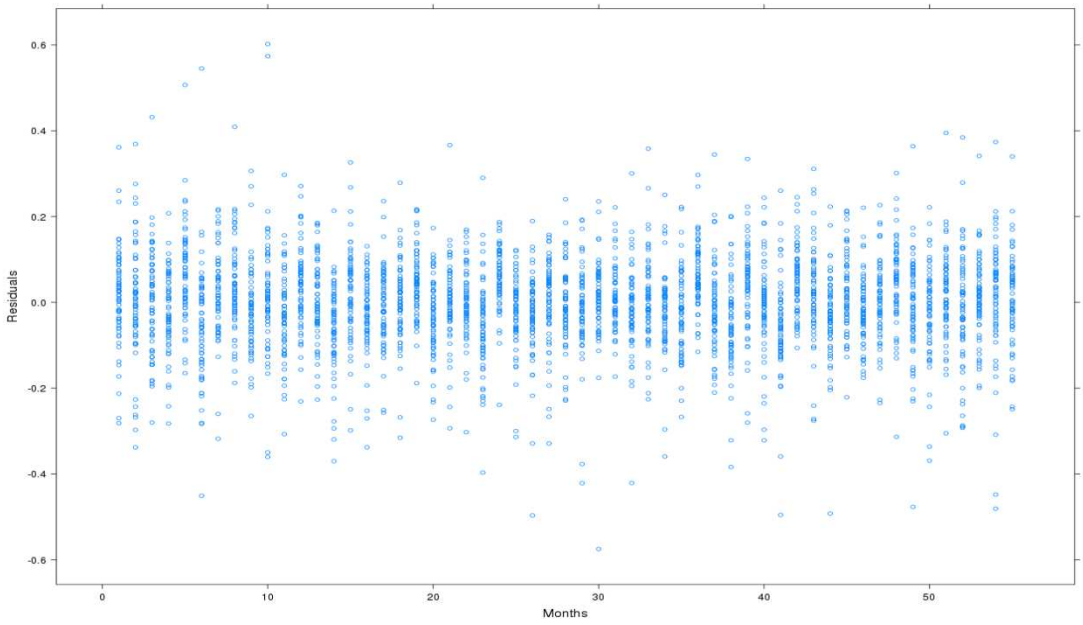


Figure 3-27. Residuals against time of the final model.

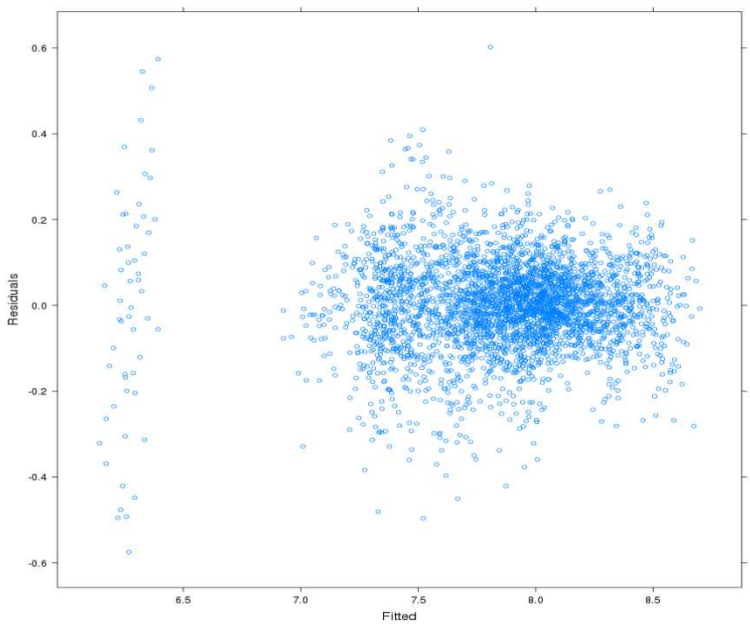


Figure 3-28 Residuals against fitted values of the final model.

The observed and fitted values of salbutamol prescribing in Figure 3-26 were symmetrically distributed around a diagonal line. Exceptions formed a number of observations at the left bottom corner of the graph, indicating that the model made unusually small estimations. Figure 3-27 illustrates that residuals against time formed a horizontal line as they should. Figure 3-28 shows that residuals remained constant in relation to fitted values with the exception of a few residuals that increased in relation to very small fitted values. In order to further investigate the exceptions on those graphs I created the same plots by GP practice Figure 3-28, Figure 3-29 and Figure 3-30. Because of the big number of practices is difficult to look in detail the diagnostic per practice therefore the three graphs are also presented in detail in Appendix H.

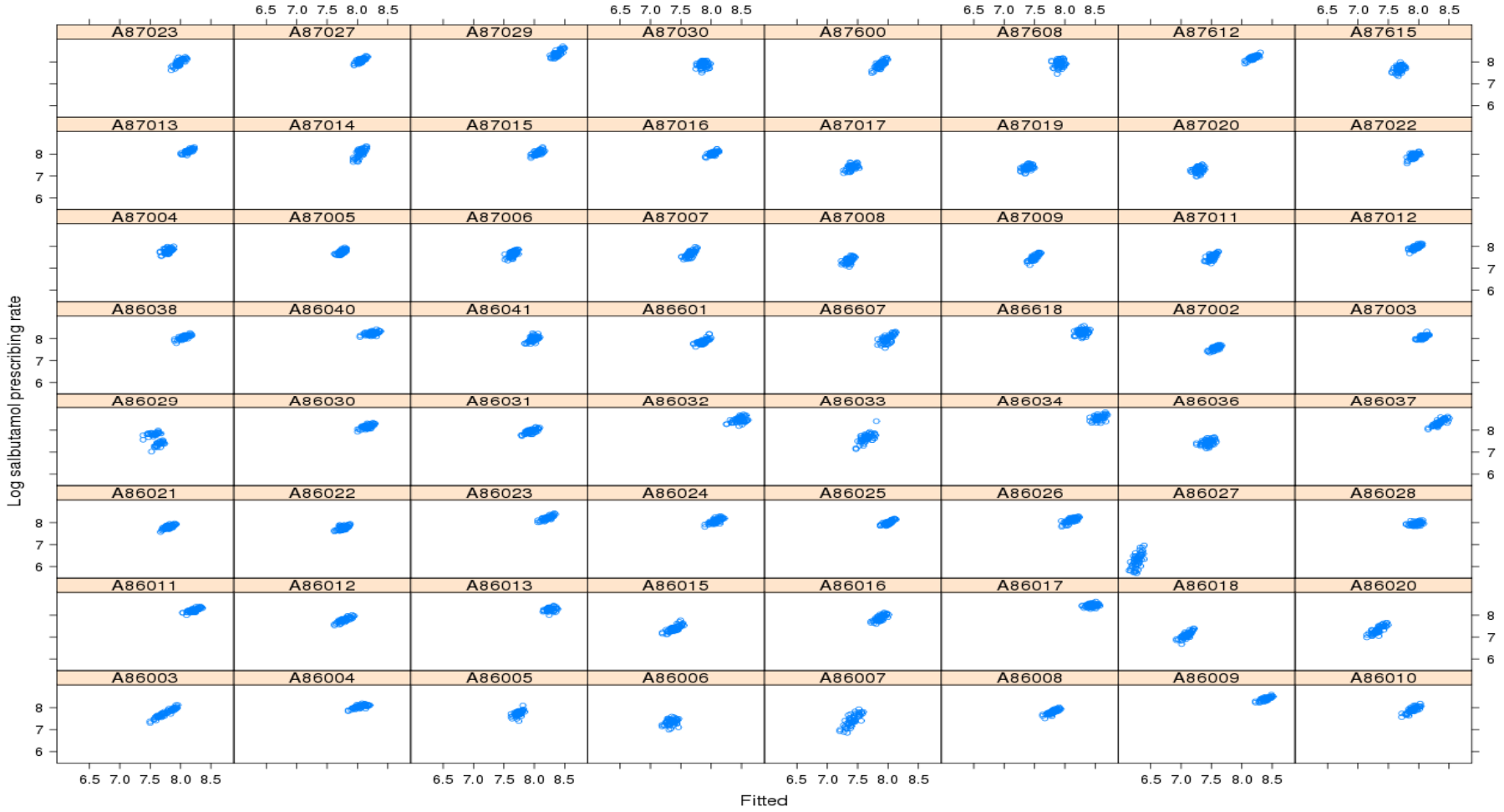


Figure 3-29. Observed against fitted values of the final model by GP practice.

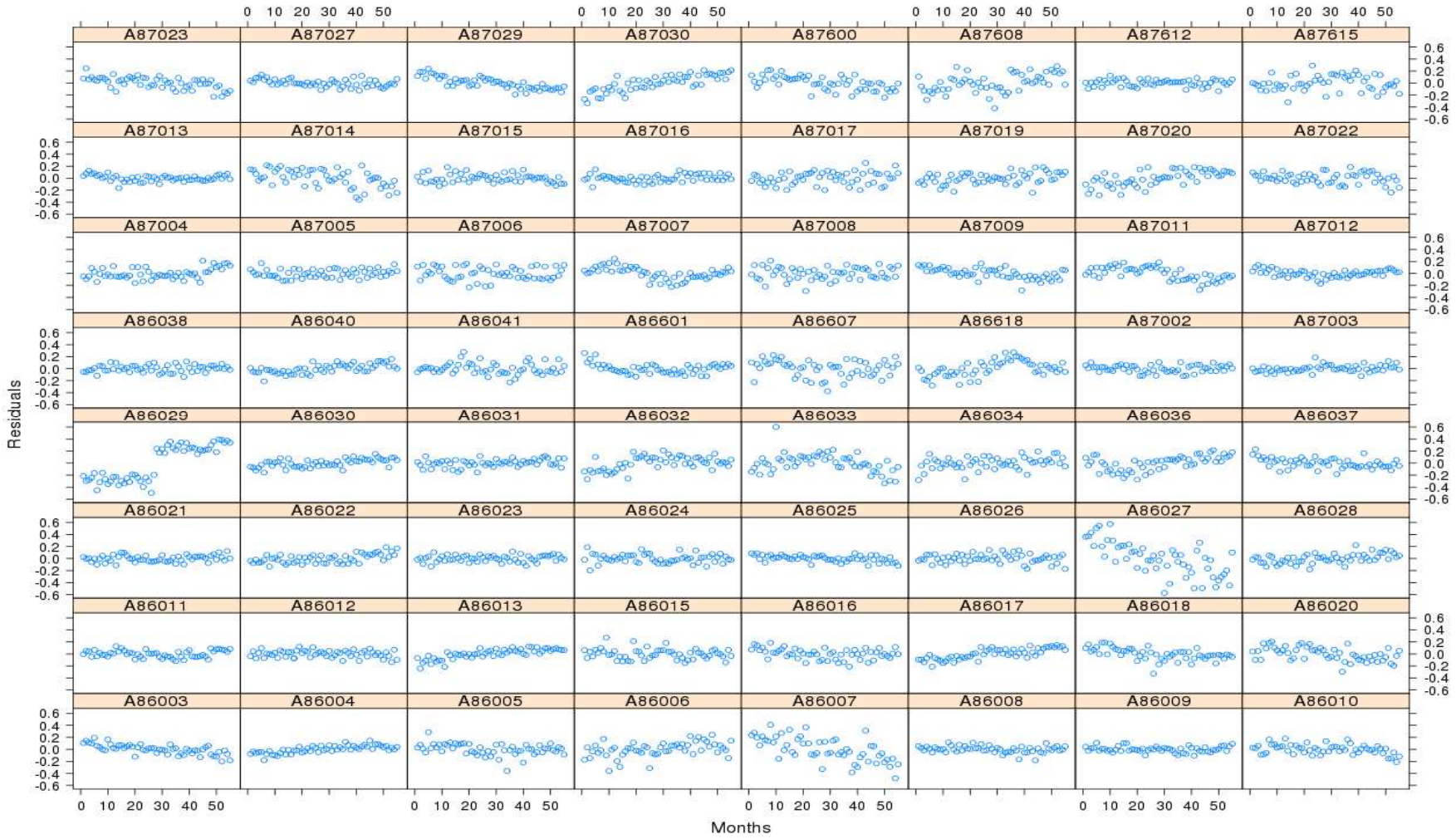


Figure 3-30. Residuals against time of the final model by GP practice.

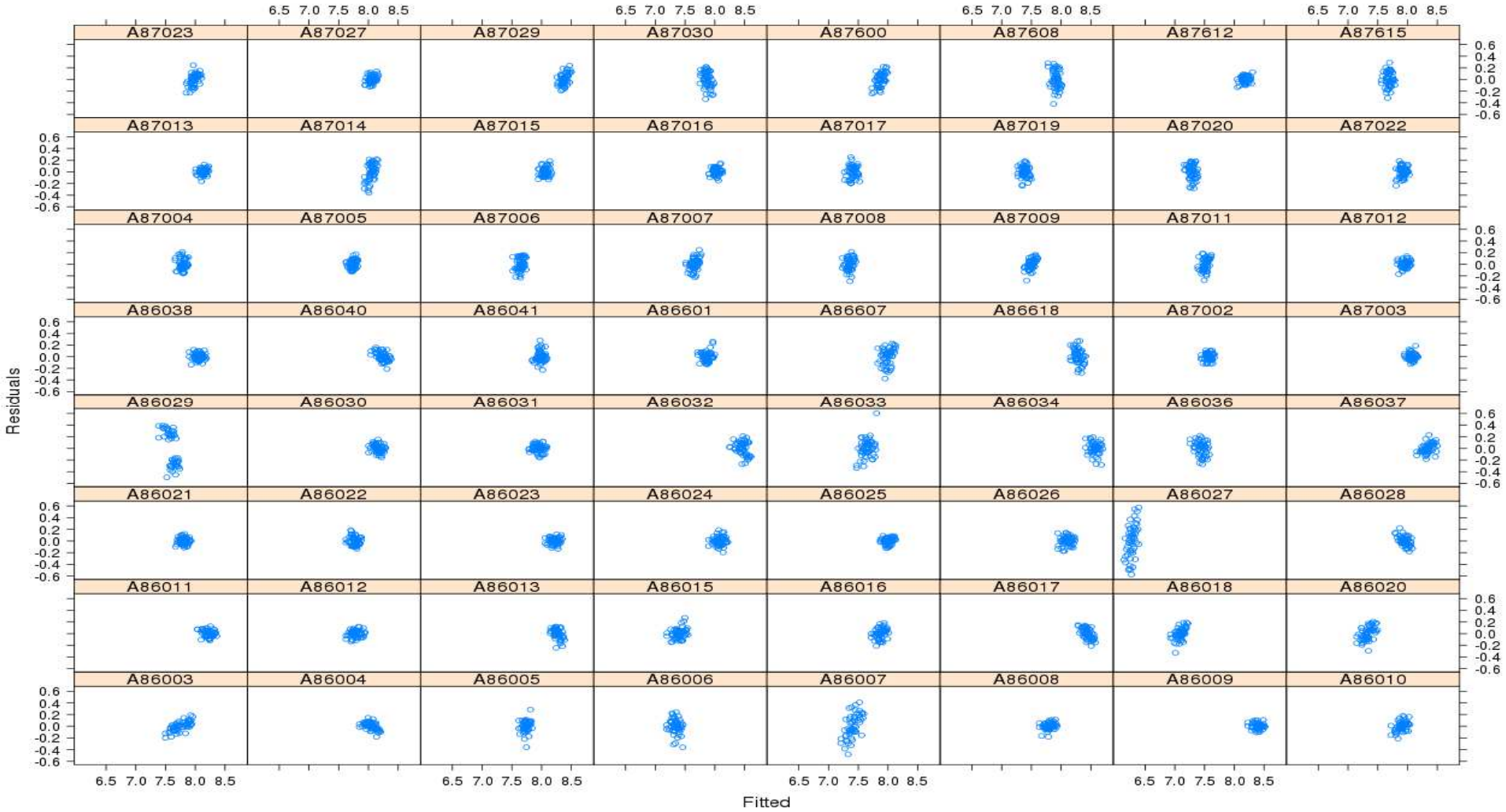


Figure 3-31. Residuals against fitted values of salbutamol prescribing by GP practice.

Figure 3-29, Figure 3-30 and Figure 3-31 demonstrate the symmetrical distribution of GP-specific residuals, with the exception of mainly two GP practices. One of them was practice A86027 which had the lowest prescribing rate and an unusual distribution of age and sex of patients registered within it, as discussed earlier. Unusual patterns also emerged from the residuals of practice A86029, for which an abrupt increase in salbutamol prescribing occurred approximately in the middle of the study period. Overall, the distribution of observed values versus fitted, as well as residuals versus time, and predicted values, were symmetrical for the vast majority of GP practices, thus implying that the assumptions of linearity and homoscedasticity were met.

Interdependence in Time - Temporal Autocorrelation

Interdependence (correlation) between observations was anticipated for the same GP practices, due to unknown influences, varying randomly over practices. For example, the prescribing patterns that individual General Practitioners may have, could cause the prescribing in a given month to correlate with prescribing in the next or previous month. When a variable correlates to itself the term autocorrelation is used to describe this interdependence. As a next step, I examined the temporal autocorrelation of the residuals.

An important guide to correlation of temporal data was given by the autocorrelation function (acf) that measured the correlation between observations at different times (Chatfield, 2004). I estimated and plotted the set of autocorrelation coefficients, arranged as a function of separation in time that I defined to be 12 monthly intervals or lags (Figure 3-32). By presenting the coefficients at increasing lags (e.g., l=1,2,…12) I could assess whether dependence was restricted to one or more lags.

When the correlation coefficient at any given lag fell within the 95% confidence interval, there was no temporal autocorrelation on this lag. An approximation of the 95% confidence limit was calculated with the equation:

$$r_{.95} \cong 0 \pm 2/\sqrt{N} \qquad \textit{3-13}$$

where N is the number of observations. This is based on the assumption that if a time series is completely random, and the sample size is large, the lagged-correlation coefficient is approximately normally distributed with mean 0 and variance $1/N$ (Chatfield, 2004). For a sample of 55 months the confidence interval

is approximately ± 0.27. The correlation coefficients for the mixed effects model were around 0.3 (Figure 3-32) at some lags, which was considered minor positive temporal autocorrelation. The persistent temporal autocorrelation indicated that some monthly variation was not accounted sufficiently over time. This could be possible related to the measurement of monthly air pollution exposure, however I consider that the reliability of the model output is sufficient as the observed temporal autocorrelation is minor.

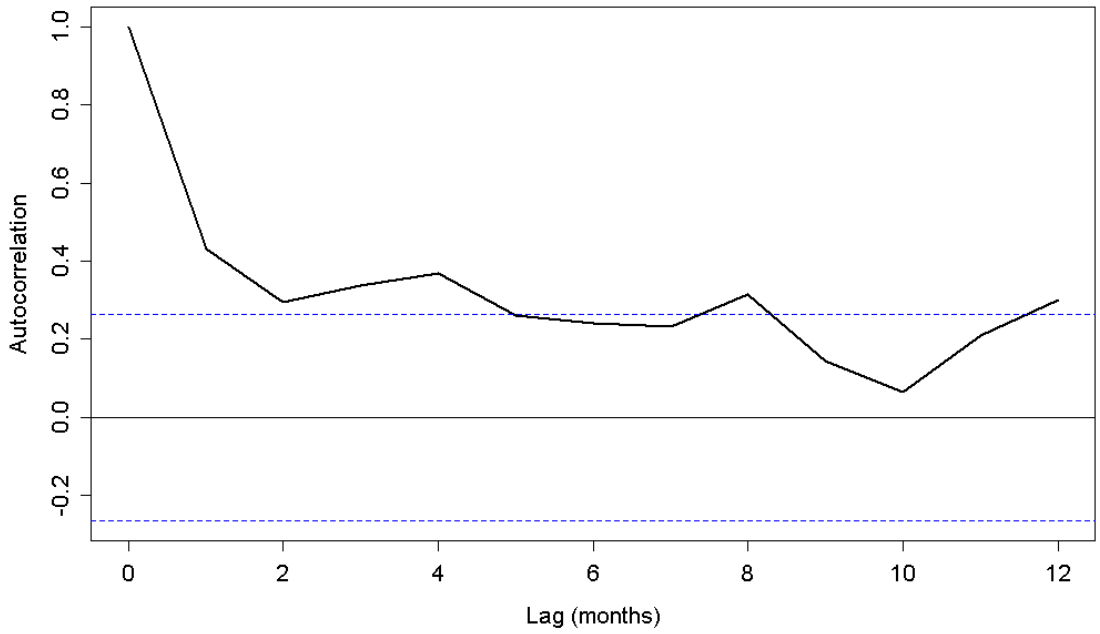


Figure 3-32 Interdependence of residuals in time (temporal autocorrelation)

Interdependence in Space - Spatial Autocorrelation

Spatial autocorrelation is a mode of spatial statistics that has been developed mainly over the past five decades. Models have to be checked as to whether the interdependence assumption of observations in space is satisfied. Not meeting this assumption again calls into question the reliability of a model's output. I needed to examine whether the random effects were correlated spatially. I assessed the spatial interdependence of random effects, using geoR that is a package for geostatistical data analysis using the R software (Ribeiro JR. and Diggle, 2001).

If the random effects were not independent that would indicate that the model did not account sufficiently for some explanatory process or event that exhibits spatial correlation (e.g. air pollution, deprivation). Random effects can be a combination of factors that depend on facilities within the practice, as well as the training, experience and prescribing pattern of individual General Practitioners. Such random influences should not be spatially correlated.

I used empirical variograms to assess the spatial interdependence of random effects. The empirical variogram (Figure 3-37) shows an increasing trend over distances up to 3000m, suggesting that there might be some positive spatial correlation over this range. I compared the variogram with the computed envelope of variograms. Figure 3-34 indicates no spatial autocorrelation as the empirical variogram fall within the upper and lower simulation envelope.

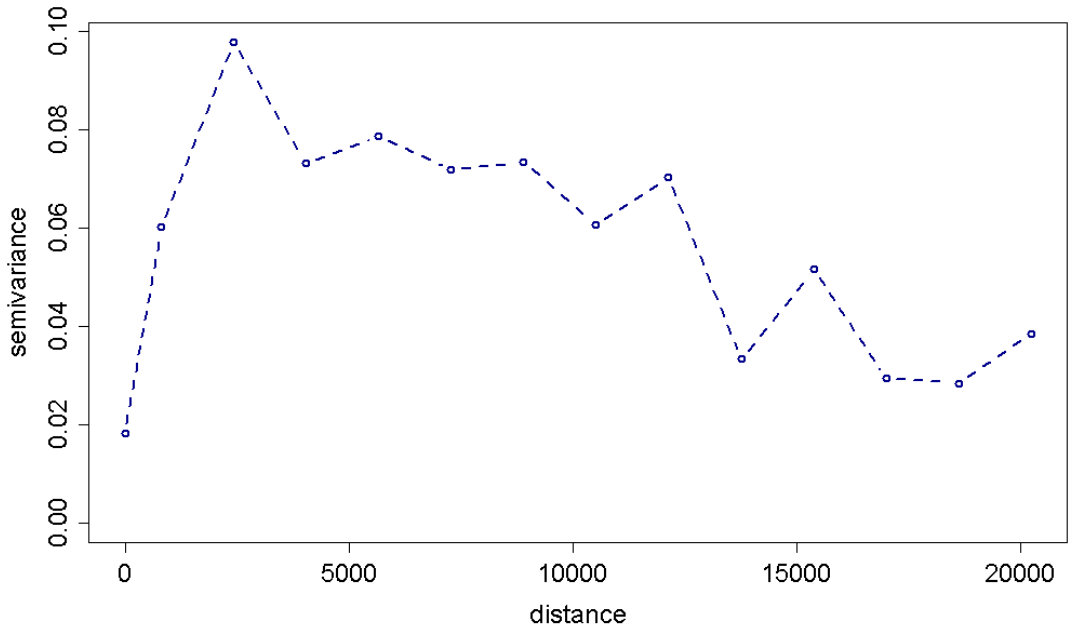


Figure 3-33 Empirical variogram for random effects

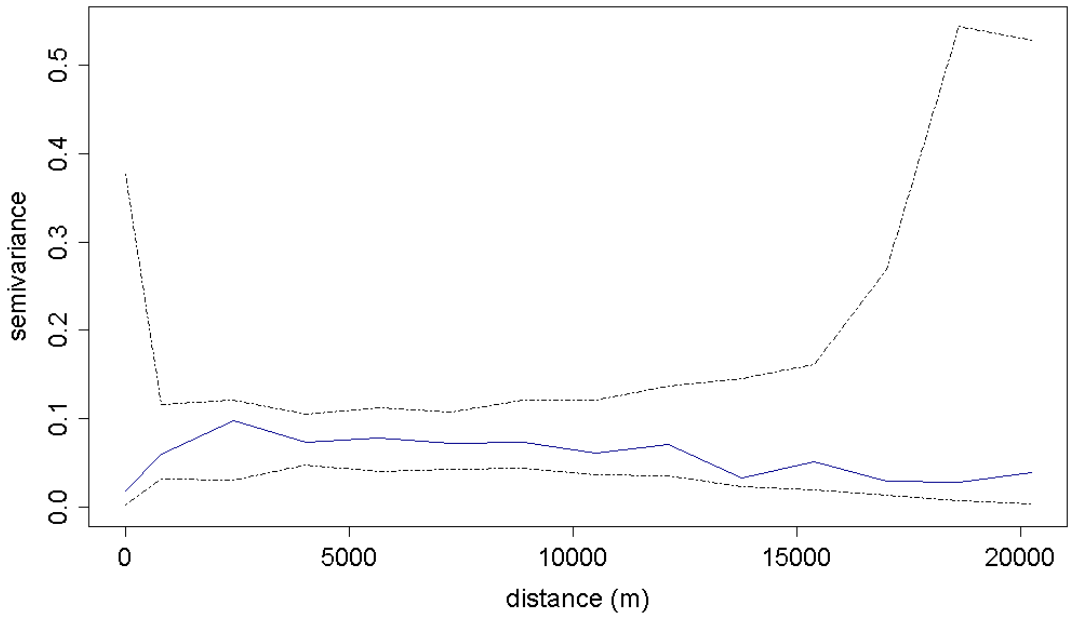


Figure 3-34 MonteCarlo envelop for the variogram of random effects

3.4.3 Final Results

So far I have screened covariates for significant p-values and have gradually developed a model (L21PM_30TR, section 3.4.1.3) that I believe to be both parsimonious and adequate. I needed to evaluate the p-values associated with various explanatory variables. The p-values of the final model were informative, but it was of equal importance to examine how exact they were. Therefore, I examined the confidence intervals associated with the coefficients of the model. The confidence intervals provided information of how precise the estimates of the model were and it can be argued that they are at least as important as p-values.

In Table 3-7, I present the p-values, coefficients and their confidence intervals produced by the final model "L21PM_30TR". The columns L_CL and U_CL, in Table 3-7, show the 95% lower confidence limits and upper confidence limits respectively. The confidence intervals were not wide indicating a sufficiently precise estimate of the coefficients.

Model L21PM_30TR						
	Estimate	Std.Error	t value	Pr(>|t|)	L_CL	U_CL
(Intercept)	-2.01821	0.15007	-13.447986	<0.0001	-2.31235	-1.72406
PM_{10}_lag21days	0.00102	0.00047	2.185799	0.0288	0.00011	0.00194
Traffic_lag30days	0.00007	0.00002	3.032136	0.0024	0.00003	0.00012
Time (months)	-0.00066	0.00012	-5.629881	<0.0001	-0.00089	-0.00043
Income Deprivation	2.81133	0.30695	9.158998	<0.0001	2.20972	3.41295
Average Age of Patients	0.03296	0.00316	10.439382	<0.0001	0.02677	0.03915

Table 3-7 Coefficients and confidence intervals for L21PM_30TR model

I then examined the strength of association between salbutamol prescribing and covariates. I assessed the impact of an increase on fixed effects (covariates) to salbutamol prescribing. I quantified the impact on salbutamol prescribing rate in the case of: 1) an increase of 10μg/m^3 in ambient PM_{10}, 2) an increase of one standard deviation (s.d.) of the traffic index (513 on traffic index score), 3) one year elapsing 4) an increase of one standard deviation (s.d.) of income deprivation (equals 0.1 income deprivation score) and 5) an increase of 3 years in average age of patients registered per GP practice.

The association of explanatory variables with salbutamol prescribing is presented in Table 3-8. I transformed the log salbutamol prescribing rate back to a normal scale, in order to be able to interpret the results meaningfully.

Table 3-8 presents the percentage change in salbutamol prescribing with given increases in predictors' units. An increase of 10 μg/m^3 of PM_{10} was associated with a 1% (95% CI, 0.1 to 2.0%) increase in salbutamol prescribing. An increase of one standard deviation (513) in traffic index score was associated with an increase of 3.8% (95% CI, 1.6 to 5.8%). One year in elapse of time was related to a 0.6% reduction (95% CI, 0.4% to 1.1%) indicating a marginal decrease in salbutamol prescribing over the study period. An increase of one standard deviation (0.1) in income deprivation was associated with a much larger increase in salbutamol prescribing rate of 32.5% (95% CI, 24.7 to 40.7%) to salbutamol prescribing rate. Finally, three years increase in the average age of patients registered with a GP practice was associated with a 10.4% (95% CI, 8.4 to 12.5%) increase in salbutamol prescribing rate.

Model L21PM_30TR				
	Covariates' increase	**Change on salbutamol prescribing rate**	**Lower CL**	**Upper CL**
PM_{10} (lag 21 days)	10μg/m^3	1.0%	0.1%	2.0%
Traffic (lag 1 month)	1 s.d. (513)	3.8%	1.6%	5.8%
Time (months)	12 months	-0.6%	-1.1%	-0.4%
Income deprivation	1 s.d. (0.1)	32.5%	24.7%	40.7%
Age	3 years	10.4%	8.4%	12.5%

Table 3-8 Association of explanatory variables to salbutamol prescribing rate

3.4.3.1 Parameter Evaluation using Markov Chain Monte Carlo simulations

As a last step, I evaluated the parameters and p-values of the final statistical model (L21PM_30TR) using Markov chain Monte Carlo (MCMC) simulations. I evaluated p-values for fixed effects from the MCMC sample using a function within the "languageR" extension in R software. This function for estimating the p-values based on MCMC sampling was not stable at the time I conducted the analysis (July 2009). This analysis was to great extent motivated by my interest to test the most recent algorithms and techniques in the highly active field of mixed effects models, in an attempt to evaluate the results of my final model. Therefore, I did not intend to accept the results as absolutely correct, but rather to examine whether its output would be close to that of my final model, as would be expected in theory.

Method

I generated MCMC sampling from the posterior distribution of my final model's parameters. I plotted the densities of the sample in Figure 3-35, which shows that the posterior density of the fixed-effect parameters was reasonably symmetric and close to a normal distribution.

I fitted the final model using a Markov chain Monte Carlo (MCMC) technique and compared the p-values of the new model to my final model. I denoted the new model, created by MCMC analysis, as "L21PM_30TR_mcmc". The fixed effects, as

well as respective prediction intervals were also estimated, based on repeated-sampling properties.

The MCMC model (L21PM_30TR_mcmc) would be expected to produce similar p-values to the final model (L21PM_30TR), where the sample is not small (Baayen et al., 2008). The sample in this study (55 observations from 64 practices) could not be considered small; therefore I expected the MCMC technique to confirm the p-values of my final model.

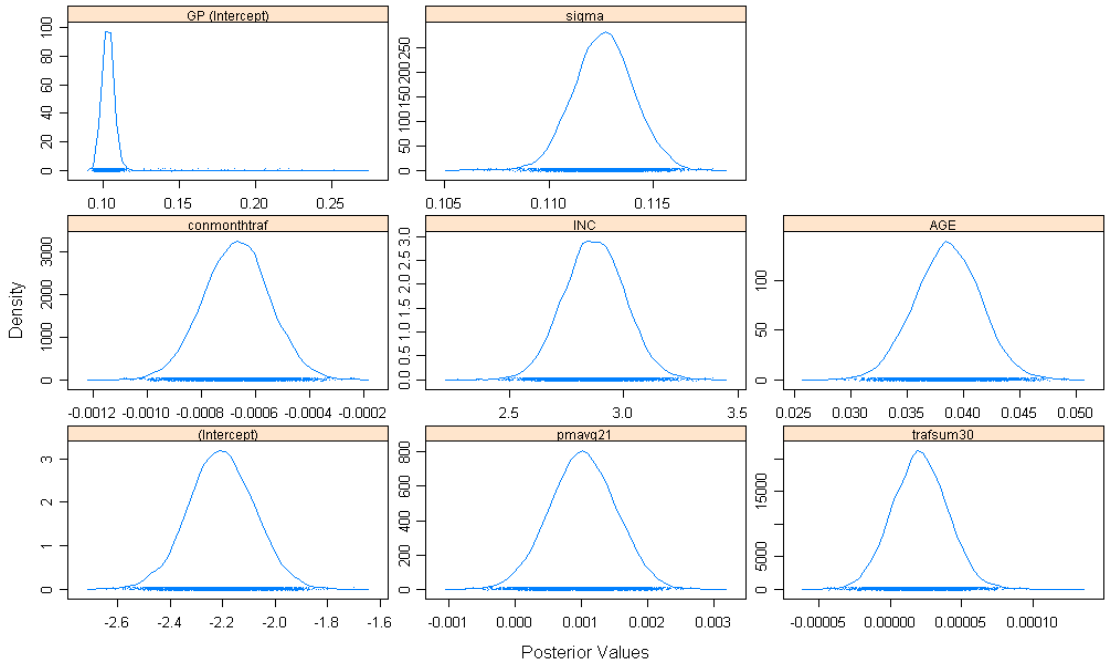


Figure 3-35 Empirical density estimates for the MCMC sample for the posterior distribution of parameters in the final model

Results

The results of L21PM_30TR_mcmc are presented in Table 3-9. The last column (Pr(>|t|)) of the table shows the p-values of the final model "L21PM_30TR" (Table 3-7). Comparing the p-values based on the posterior distribution (pMCMC) and on the t-distribution (Pr(>|t|)), a difference appeared in the p-value associated with the covariate "Traffic_lag30days". The high p-value of 0.294, estimated by the new model, meant that the traffic flow index had no statistically significant relationship to salbutamol prescribing. The p-values related to the other explanatory variables remained significant.

Model L21PM_30TR_mcmc						
	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC	Pr(>|t|)
(Intercept)	-2.0182	-2.2035	-2.4446	-1.96	0.0001	<0.0001
PM_{10}_lag21days	-0.0007	-0.0007	-0.0009	-0.0004	0.0001	0.0289
Traffic_lag30days	0.001	0.001	0.0001	0.002	0.0324	**0.0024**
Time (months)	0.0001	0	0	0.0001	0.2774	<0.0001
Income deprivation	2.8113	2.8703	2.6102	3.1309	0.0001	<0.0001
Average age of patients	0.033	0.0386	0.0328	0.0441	0.0001	<0.0001

Table 3-9 Results of L21PM_30TR model based on Markov chain Monte Carlo analysis

I then examined whether the magnitudes of the coefficients had changed. The columns L_CI and U_CI, in Table 3-7, show the 95% lower confidence limits and upper confidence limits, respectively. The confidence intervals were not wide and showed a sufficiently precise estimate of the coefficients. The same held for the equivalent MCMC intervals observed in Table 3-9, where the columns "*HPD95lower"* and "*HPD95upper*" show the lower and upper 95% Highest Posterior Density (HPD) intervals respectively, for the parameters in the MCMC sample. A 95% prediction interval is an estimate of an interval in which future observations will fall, with a 95% probability, given what has already been observed.

In order to investigate the mcmc model further, I plotted the MCMC estimates of covariates in association with salbutamol prescribing rate (Figure 3-36). In Figure 3-36, the dotted lines depict the 95% prediction intervals associated with salbutamol prescribing. The estimates related to large values of traffic flow index (diagram "Traf_30lag" in Figure 3-36), fell outside the 95% prediction interval, making this relationship not statistically significant according the MCMC analysis as discussed earlier. However, I observed that the prediction interval was wider towards the large values of traffic index, showing increased uncertainty in this prediction. Overall, I accepted the result of the final model (L21PM_30TR). This was because the function for estimating the p-values based on MCMC sampling was not stable at the time I conducted the analysis. At the same time, I would re-visit the calculation of that index, taking into account the issues I discussed in Section 2.2.2.4.

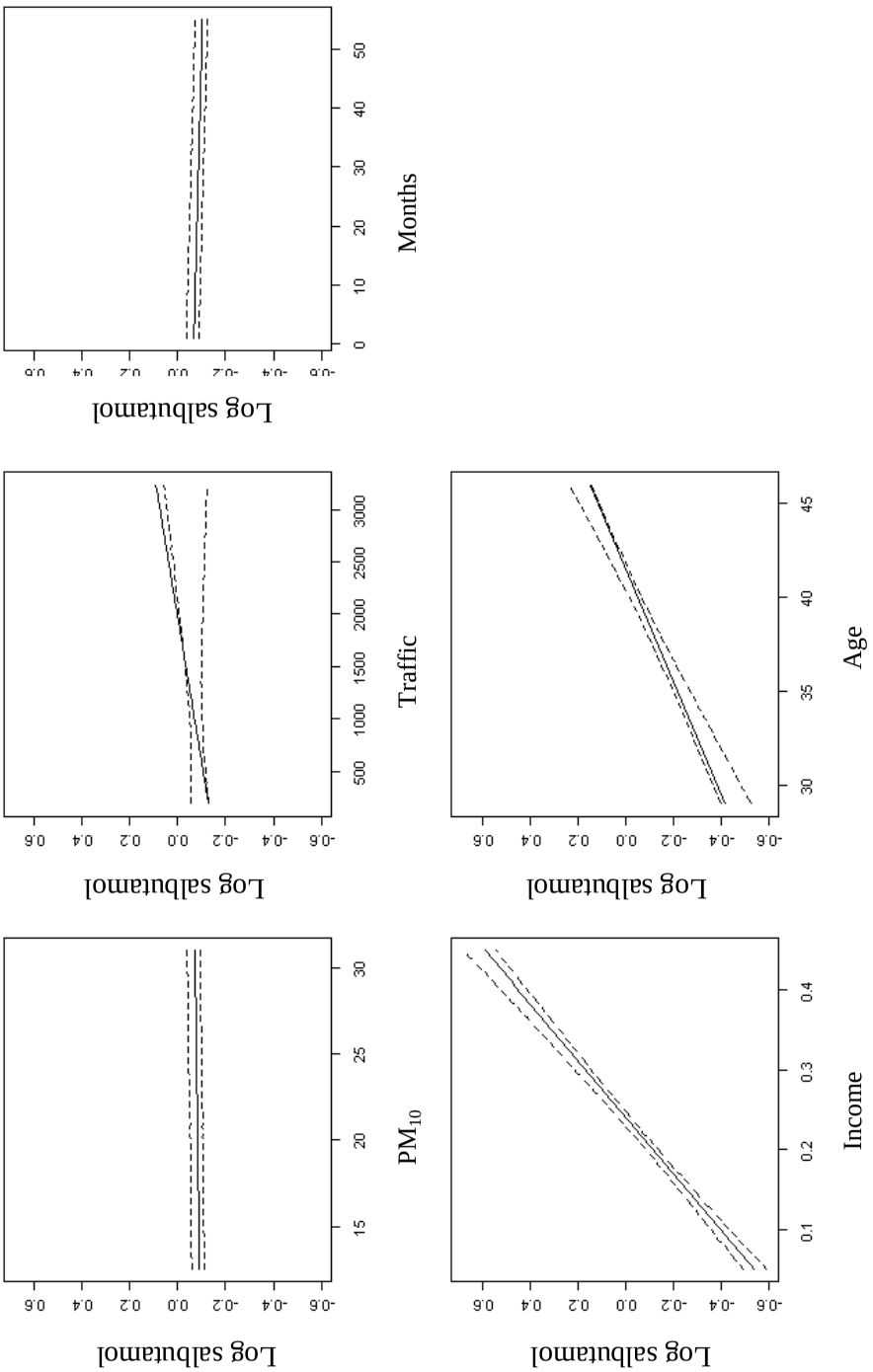


Figure 3-36 MCMC estimates of predictors in association to log average salbutamol prescribing

3.4.3.2 Goodness-of-fit

I created a graph (Figure 3-37) of the predicted values of log salbutamol prescribing based on the final model (L21PM_30TR) against the observed values by GP practice. Figure 3-37, Figure 3-38, Figure 3-39 and Figure 3-40 illustrate the goodness-of-fit of the final model. The predicted value captured the GP practice specific prescribing (intercept), while the within GP practice variation was not captured in an equally successful manner in some practices. The highest discrepancy occurred in practices with high within group variability, showing that the model was not adequate to capture the variation of salbutamol prescribing on the edges.

The coefficient of determination (R-squared) shows the proportion of variability in a data set, which is accounted for by the statistical model and provides a measure of how well future outcomes are likely to be predicted by models. R-squared is appropriate for models that use the Ordinary Least Squares (OLS) approach to calculate the minimum variance. The estimates of a mixed-effects model are maximum likelihood estimates though, so the OLS approach as a measure of goodness-of-fit is not appropriate. To evaluate the goodness-of-fit of mixed-effects model, other approaches have been developed. These are called pseudo R-squares, the usefulness of which is debatable (Orelien and Edwards, 2008). I used a pseudo R-square which is meant to be appropriate for mixed-effects model with random intercepts. I extracted the total variance explained by a mixed-effects models and the total variance explained by the same mixed-effects model when fitted only with the random effects. More details on this estimation are presented in Appendix I. The pseudo R-squared value for the final statistical model (L21PM_30TR) was 48.3%.

In the absence of a widely accepted statistic such as R^2 of traditional linear regression for the linear mixed model, I used correlation as a last step to gain a quantitative estimate of how well the model fitted the data. The correlation coefficient was 0.96 between the observed and the fitted values.

Overall, the final model captured the GP practice specific prescribing satisfactorily. The within GP practice variability was captured satisfactorily for the vast majority of GP practices but not for the few that exhibited very high prescribing variability. The model performed reasonably well but robust external validation of its outcome will be required if there is any interest to predict salbutamol prescribing in the future.

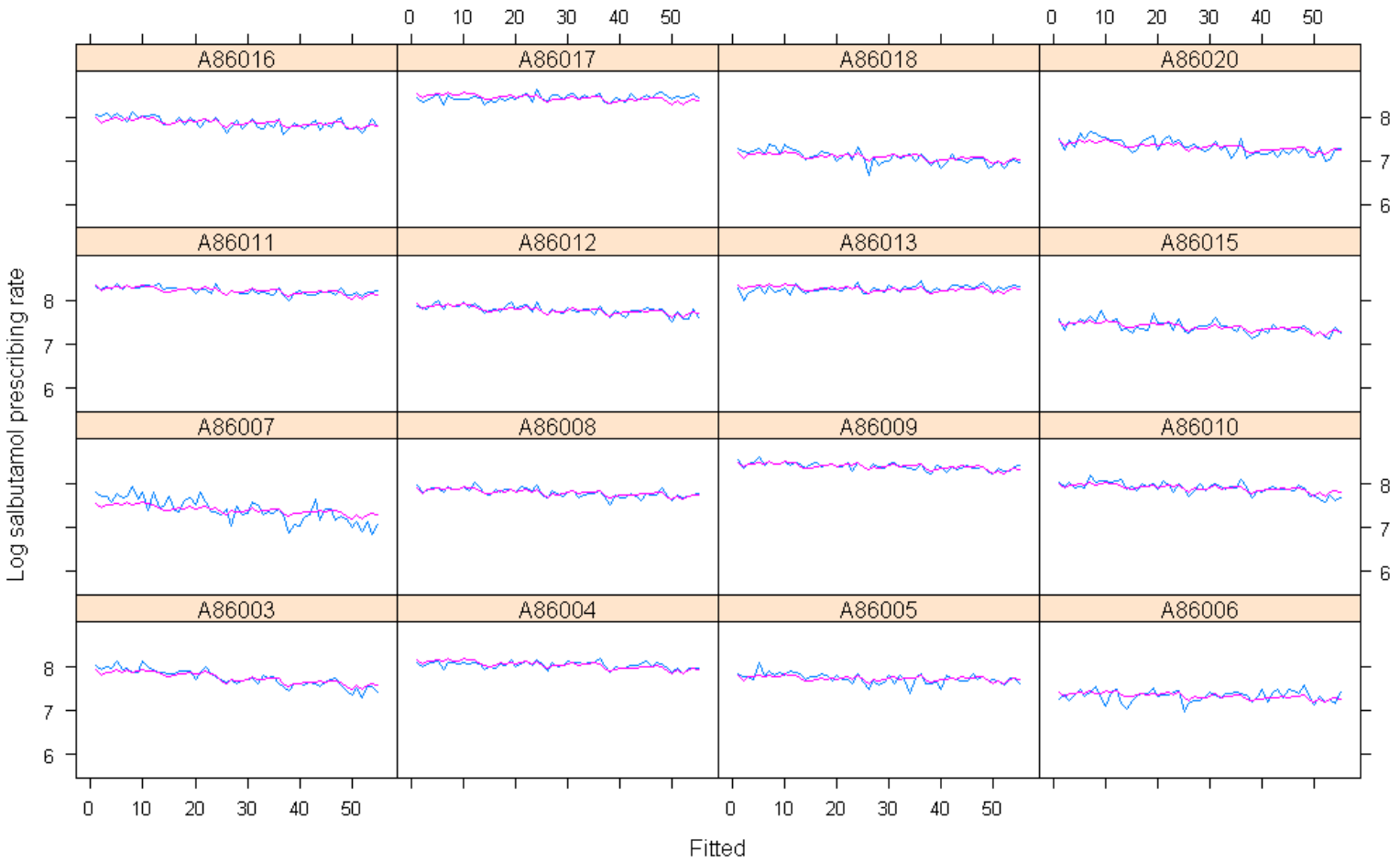


Figure 3-37 Final model: Predicted values (pink line) of log salbutamol prescribing against the observed (blue line) per GP practice

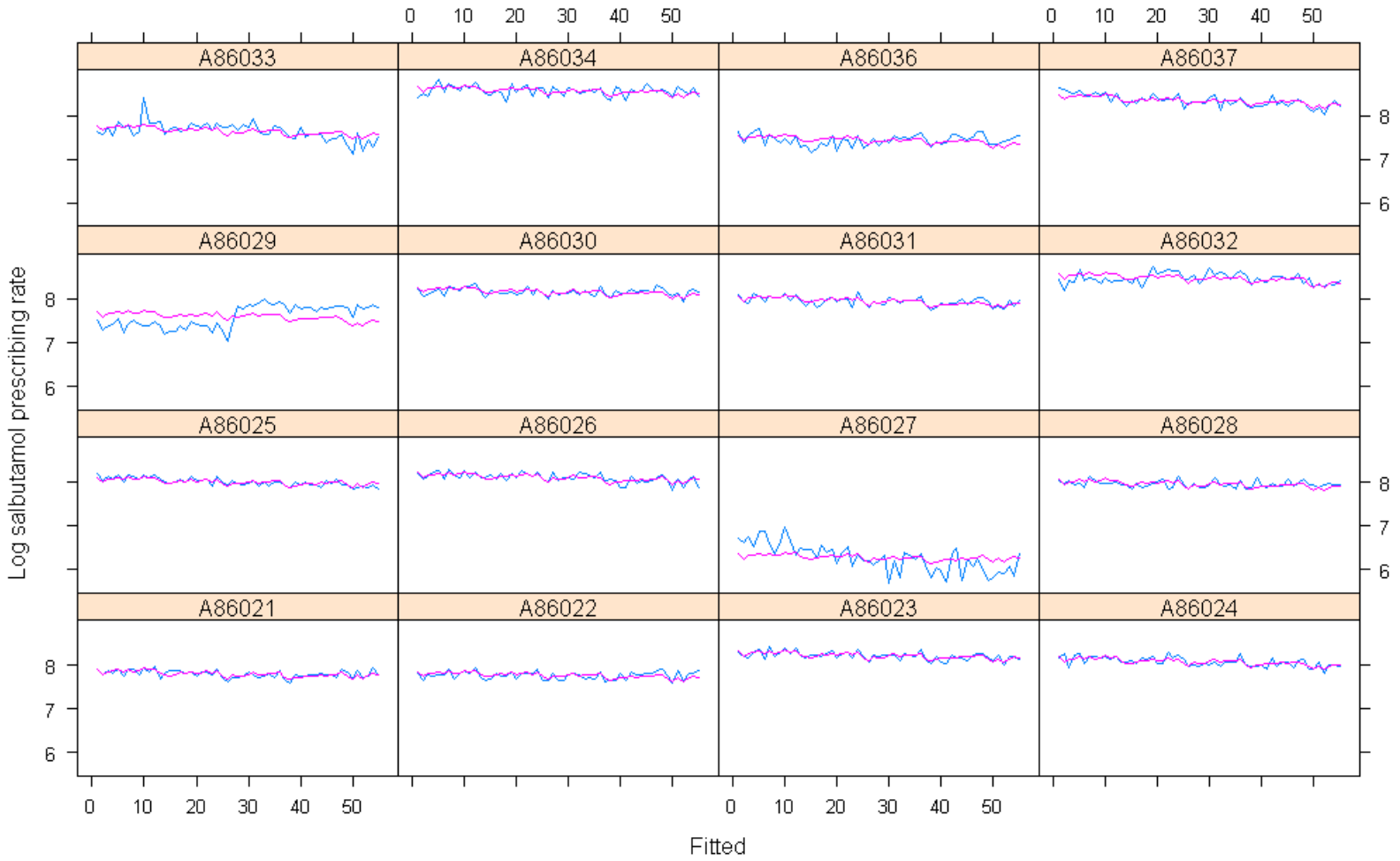


Figure 3-38 Final model: Predicted values (pink line) of log salbutamol prescribing against the observed (blue line) per GP practice

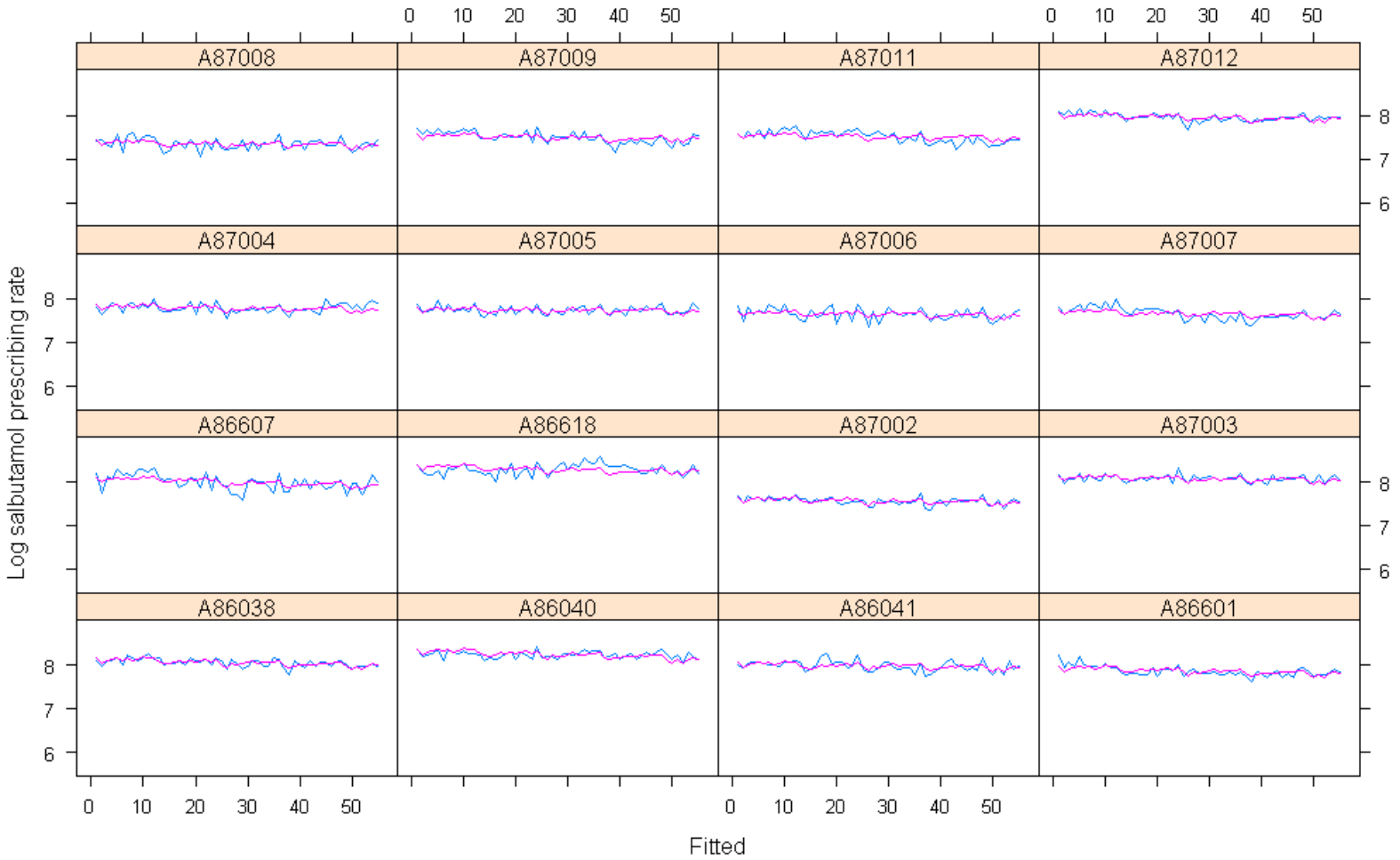


Figure 3-39 Final model: Predicted values (pink line) of log salbutamol prescribing against the observed (blue line) per GP practice

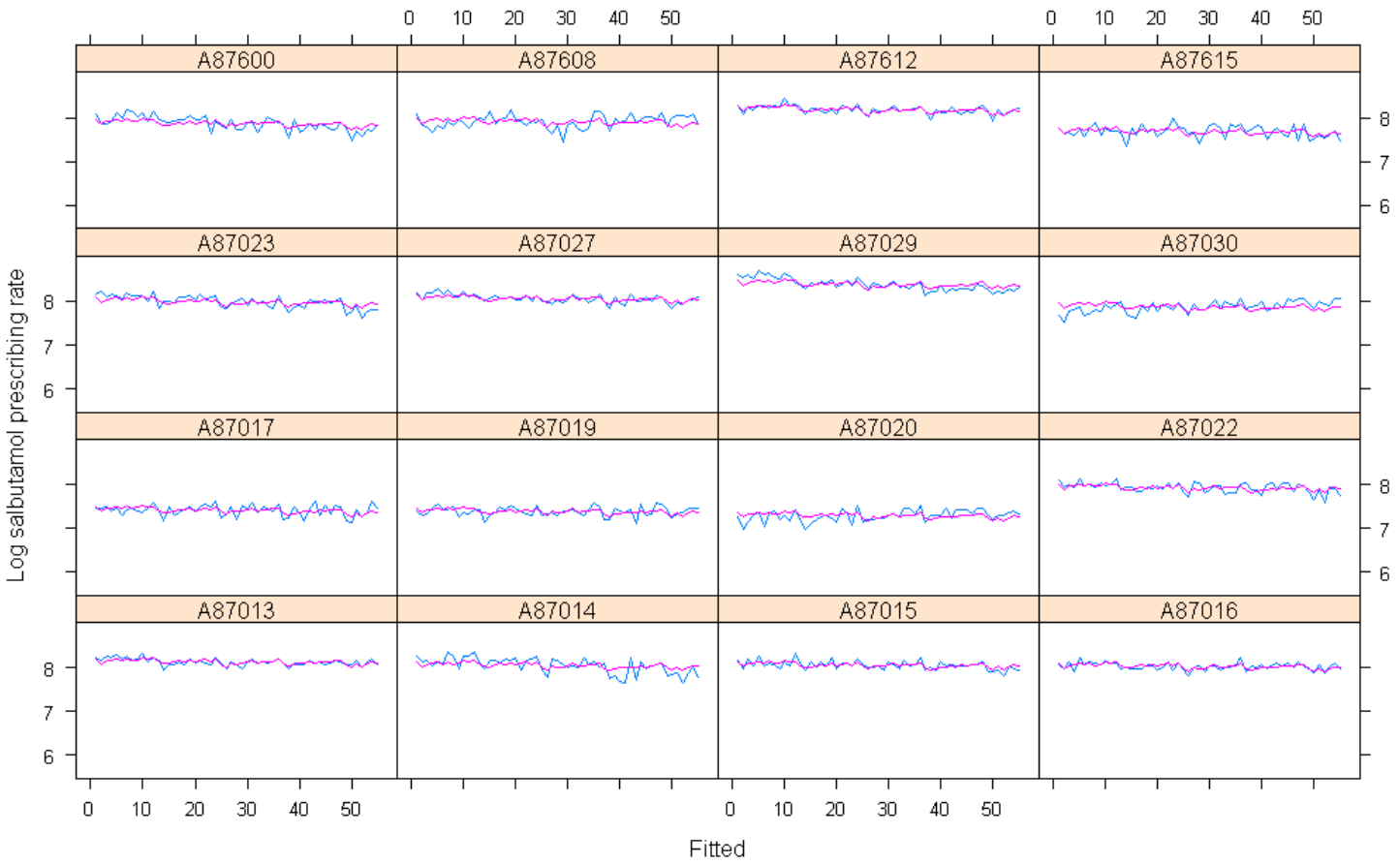


Figure 3-40 Final model: Predicted values (pink line) of log salbutamol prescribing against the observed (blue line) per GP practice

Chapter 4. Discussion

This study demonstrated an association between patients' exposure to ambient air pollution and the risk of increased quick-relief medication usage by asthma and COPD patients at primary care level. These findings, added to current understanding of the link between air quality, asthma & COPD exacerbation and prescribing, can ultimately increase the scope of primary care management. The study demonstrated that respiratory prescribing can be a useful indicator of respiratory health outcomes with great scope for use in epidemiological surveillance of air pollution health effects. The study found that an increase of 10μg/m^3 in monthly ambient PM_{10} concentrations was associated with increases of 1% in salbutamol prescribing. This is the first study in the UK to quantify the relationship of respiratory prescribing and its predictors, in primary health care.

4.1 Ecological Study

This study had an ecological design. An ecological study is defined as "a study in which the units of analyses are populations or groups of people, rather than individuals" (Last, 1995). Often, the primary reason for an ecological design is not a specific interest in effects of contextual, ecological variables but rather the lack of individual data. This study had as its central interest the effects of contextual, ecological variables that occur at primary care level. In the UK, a consistent priority for NHS reform is the shift from provision of hospital-based acute care to proactive care delivered in primary care (Department of Health, 2005, Department of Health, 2006, Department of Health, 2010). Appropriate tools and studies, on this level of care, are required to provide information to help make decisions to support effective and efficient management. Ecological studies are useful tools for

surveillance of disease as well as risk prediction. Surveillance analyses are often ecological, since they describe trends in groups of individuals (Teutsch and Churchill, 2000).

The concept of ecological studies is covered by all major epidemiological textbooks, explaining at the same time the major concern linked to these studies, called ecological fallacy. The ecological fallacy (bias) stems from making inferences on individual risks based on group-level associations. In the context of environmental or air pollution epidemiology, making inference of exposure-outcome associations on individuals rather on the group level, would create an ecological fallacy. Brenner et al. (1992) as well as Richardson et al. (1987) made the point that inference at the level of the individual, based on ecological studies, can only be judged as posterior, meaning after the establishment of the individual-level cause-effect association. Their work compared relative risks for lung, bladder and oesophageal cancer derived from ecological and individual studies, with smoking as the risk factor. Their study demonstrated that ecological estimates of smoking-related lung cancer risk are similar to those derived from individual (cohort) studies, but the same did not hold for the cases of bladder and oesophageal cancer (Richardson et al., 1987).

This ecological study aimed to capture the association of respiratory prescribing and air pollution in primary care. The study described correctly the direction of the relationship of air quality and asthma/COPD exacerbations that has been previously documented by individual level studies on asthma and COPD. The study's results confirmed that respiratory prescribing can be a useful indicator of respiratory health outcomes with great scope for use in epidemiological surveillance of air pollution health effects.

4.2 Latency Periods

A delayed response of respiratory prescribing to ambient pollution was investigated and evidence was found on the statistical models with different PM_{10} lag times. Long-term exposure to air pollution makes patients more susceptible to asthma/COPD triggers leading to exacerbation of symptoms, which in turn causes an increase in medication use. The increase in respiratory medication use eventually leads to the issue of a new prescription. Because the urgency by which a prescription would be issued would vary by individuals, and with their initial stock of medication, short-term effects are more difficult to observe, using prescribing data.

I used long periods (months) of prescribing, and long air pollution latency periods (weeks). As the prescribed data were averaged monthly, examining their relationship to daily latency periods would not be appropriate. During those longer periods of time, increased air pollution would be expected to be related to increased prescriptions, over and above the short term fluctuations driven by individuals' needs for medication replacement or opening hours of practices and pharmacies. The studies discussed in the literature review section (Section 1.3.2.2), used daily respiratory medication data looking at short term associations, the only exception being the study by Vegni et al. (2005) that used weekly data. Additional issues were revealed in studies examined the short-term associations, such as days that practices or pharmacies were closed (e.g. weekend, bank holidays), which added to the complexity of their statistical models. Vegni et al. (2005) used 7-days average of respiratory medication, in order to avoid over-complication of his statistical model.

The PM_{10} concentrations recorded 21 and 14 days before the prescribing month were positively associated with the salbutamol prescribing rate. This meant that there was a time window between 14 and 21 days for this relationship. The final statistical model assessed this relationship, after I had firstly accounted for the prescribing variation attributable to seasonal factors, captured by temperature at various latency periods. This initial analysis, described as Stage A of the statistical model (section 3.3), showed a significant association of salbutamol prescribing and temperature on 7-, 14- and 21-days lag. The results indicated a time window between 7 and 21 days. Therefore, even though the relationship between PM_{10} and prescribing appeared to be significant at approximately 14-21 days, the time window increased to 7-21 days when taking into account all explanatory variables of monthly salbutamol prescribing.

Previous similar studies have focused on latency periods that did not exceed 14 days. The study by Zeghnoun et al. (1999) examined the longest latency periods (14 days) compared to other studies. Their study did not use PM_{10} but BS, for which associations were found for 1-day lag as well as for 8-days lag. The most recent study by Laurent et al. (2009) was the most comparable to mine, as they assessed the relationship between short-acting β2-agonists and PM_{10}, among other pollutants. Their study observed statistically significant associations involving latency periods of 4-10 days.

The latency period for respiratory prescribing in relation to trigger factors is complex process, being partly related to management of medicine supplies and partly to pathophysiological response. Based on the exploration of prescribing data,

the time windows observed were plausible, in terms of medication consumption. I examined further studies that reported latency periods on air pollution and prescribing, aiming to find evidence as to whether the longer latency periods I found as significant could be biologically plausible. To assess the pathophysiological responses I looked at studies based on individuals that assessed whether the impact of air pollution can induce respiratory disorders expressed by use of short-acting β2-agonist consumption. I found that air pollution can induce respiratory disorders within a few hours (Rabinovitch et al., 2006), to a few days (Schildcrout, Sheppard et al. 2006) and a few weeks (von Klot et al., 2002).

A latency of several hours was observed in the study by Rabinovitch et al. (2006), but their subjects had severe asthma and therefore the findings are not relevant for the majority of asthmatics and are of limited relevance to my study which included asthmatics at all levels of severity. The few weeks latency period was reported by von Klot's et al. (2002) using 5-, 10- and 14-day running averages for the relationship between ultrafine particles and short-acting β2-agonist consumption. A key difference between von Klot's et al. (2002) study and most previous investigations was that their study examined effects of air pollution with lags of up to 14 days, rather than just a few days. Von Klot et al. (2002) concluded that the full range of air pollution effects would not have been presented if their study had used exposure on the same or the previous few days only. This could be explained by a provocation of an inflammatory reaction in association with exposure to accumulated ultrafine particle concentration. This was in line with Neukirch et al. (1998) who had found effects of air pollution on symptoms that continued for several days after the exposure, and who had suggested that the inflammatory process in the airways was the reason for this.

One of the findings of my ecological study is that longer latency periods (over 14 days), in addition to the most frequently used (1 - 14 days) should be examined. The results of this ecological study are in line with those studies that have used individual data. My overall conclusion is that the model successfully captured the presence of delayed responses between salbutamol prescribing and asthma/COPD exacerbations trigger factors.

4.3 Random Effects

I analyzed data at the Primary Care level for Newcastle and North Tyneside. The differences between practices were modelled jointly by means of random effects and the effect of GP practices was found to be statistically significant. In other

words, the effect of GP practices was unlikely to be zero, or more accurately the effect found in this sample of 64 GP practices was not the sort of effect one would expect to see if there was no effect in the population from which this sample was drawn. An important aspect of mixed-effects models is that the question of interest is for the whole population and not only for the population being sampled (Pinheiro and Bates, 2000). Consequently, inferences can be made for the population of practices that the practices I used came from rather only for those specific practices. When analysing health data requiring ethical and confidentiality considerations, using as little patients' data as possible to make inferences with same level of accuracy is desirable.

A characteristic of mixed effects models is that they have two sources of variation, both within and between groups. The final model predicted the prescribing variability between GP practices (random effects) satisfactorily, while the within GP practice variability (fixed effects) was predicted satisfactorily for the majority of GP practices but not for all. I would expect that a variable that I have missed including in my model would improve the prediction value of the model on some practices with high within practice variability (e.g. A86027, A87615), as presented in Figure 3-37. The fact that the variance between GP practices was better captured than the within group variance, illustrated the importance of random effects in this analysis. The findings showed that the variation of salbutamol prescribing was subject not only to health needs caused by deprivation and air quality, but also practice specific random effects. Random effects can be a combination of factors that depend on facilities within the practice, as well as the training, experience and prescribing pattern of individual General Practitioners. None of these are captured by any accessible data source.

Chapter 5. Concluding Chapter

5.1 Overall Discussion and Conclusions

Asthma and COPD are the two most common respiratory diseases (World Health Organization, 2009a). Asthma had been classed as epidemic during the 1970s and 1980s world-wide, with Western countries having some of the highest prevalence rates (World Health Organization, 2000). Over the past 10-15 years the findings on time trends of asthma have been conflicting. Even though some ISAAC centres in Western countries have reported either no increase or even a decrease, an increase of asthma prevalence has been reported in Spanish and Portuguese speaking centres (Asher et al., 2006). The ISAAC study also found an increase in the prevalence of reported symptoms in children in younger age groups, in Western countries with high prevalence (Asher et al., 2006). A few studies in England and Switzerland have reported that asthma prevalence has reduced, probably because allergic asthma has declined (Anderson et al., 2004, Bollag et al., 2009, Braun-Fahrlander et al., 2004). **It appears that prevalence of asthma is still rising in low and middle income countries, therefore on a worldwide scale asthma prevalence is still increasing (World Health Organization, 2009a). At the same time, it is commonly accepted that COPD prevalence is increasing and it is predicted to be the third leading cause of death in coming decades (World Health Statistics, 2008).**

It is estimated that 90% of deaths from asthma are preventable and 75% of hospital admissions for asthma are avoidable (Asthma UK, 2010). No similar data are available for COPD. In the UK, a consistent priority for NHS reform is the shift from provision of hospital-based acute care to proactive care delivered in primary care (Department of Health, 2006, Department of Health, 2005, Department of Health, 2010). This ecological study focused on data from a primary health care

setting, linking different datasets and developing a model that would be able to address the main methodological issues related to a statistical model at primary care level. The approach to summarising GP catchment areas allowed me to reduce the degree of overlap that proved almost prohibitive for spatial analysis. In addition, the formation of GP service areas that did not conform to administrative boundaries increased the scope to link GP level data to determinants of ill health that are not provided within administrative boundaries, such as air pollution. Finally, I demonstrated that mixed-effects models can address, to a great extent, the temporal autocorrelation issues associated with data grouped by GP practice. **The study's model of the primary care setting contributes to the development of evidence based research in primary health care as well as strengthening of health information systems.**

The majority of primary care consultations in the UK are for patients with respiratory disease (Pinnock and Sheikh, 2009, British Thoracic Society, 2006). I am arguing that this emphasises the importance of exploiting respiratory data at this level and the need for evidence at this level. Most epidemiological studies use data from secondary and tertiary care such as hospitals and emergency care (Walters et al., 1994, Schwartz, 1994, Atkinson et al., 2001a, Tolbert et al., 2000, Arbex M A et al., 2009, Sunyer et al., 1993, Sunyer et al., 1997, Medina-Ramon et al., 2006), even though the vast majority of patients are fully treated in primary care. Medication prescribing data in primary care has been suggested for monitoring and surveillance of chronic diseases (World Health Organization, 2008a). I used prescribing for respiratory disease as an indicator of asthma and COPD exacerbations. This outcome has the advantage that it can capture patients with any level of severity of the disease from mild to severe. In contrast, traditional indicators such as hospital admissions and emergency care tend to capture events in patients who suffer severe symptoms, representing only a minority of the population of asthma and COPD patients. A few studies in France, Italy and USA have reported respiratory prescribing as a useful indicator of respiratory mortality as well air pollution health effects (Laurent et al., 2009, Naureckas et al., 2005, Pitard et al., 2004, Vegni et al., 2005, Zeghnoun et al., 1999). **This study was the first in the UK to use respiratory prescribing as a marker of health outcome for asthma and COPD, and demonstrates that salbutamol prescribing can be useful for epidemiological studies and surveillance of air pollution health effects, in a primary care setting.**

The relationship between prescribing rate of salbutamol and air pollution was quantified and latency effects on this relationship were assessed. I found that an increase of 10μg/m^3 in ambient PM_{10} was associated with a 1% increase (95% C.I.

0.1% to 2%) in salbutamol prescribing rate. The association of PM_{10} levels 14 and 21 days before the prescribing month were both found to be statistically significant, but the stronger association was found for the 21 days latency period. This was the first ecological study to assess the relationship between a short-acting β2-agonist medication and air pollution with such long latency periods. A similar study of individuals had examined effects of air pollution on consumption of asthma medication using lags up to 14 days (von Klot et al., 2002) and suggested that long latency periods should also be examined. The latency periods for respiratory prescribing and air pollution are complex, being partly explained by the management of medicine supplies and partly by provocation of an inflammatory reaction in association with exposure to air pollution (Neukirch et al., 1998). **An increase of 10 μg/m^3 of PM_{10} was associated with a 1% (95% CI, 0.1 to 2.0%) increase in salbutamol prescribing. The effect size of exacerbation of respiratory symptoms triggered by daily variation of air pollution has been found to be between 1% (Atkinson et al., 2001; Donaldson, K. Medina-Ramon, M., et al., 2006; Pope C.A. et al., 1995) and 7% (Laurent et al., 2009) per 10μg/m^3. One of the findings of this ecological study was that longer latency periods over 14 days should also be examined. Fitting the same model to new datasets would be required in order to confirm the results of this ecological study.**

The modelling process was separated into two stages. The first stage captured the area-wide seasonal variation in prescribing. Its results were then used as an offset input to the second stage of the model. The second stage comprised a mixed-effects model, which assessed the remaining unexplained spatio-temporal variation of salbutamol prescribing in relation to air quality, deprivation and demographic variables. In the first stage, I modelled the seasonal variation of salbutamol prescribing using just temperature data. This was possible by employing a dynamic harmonic regression model. Previous studies on respiratory prescribing have used seasonal data such as pollen counts and respiratory infections as well as weather data to account for seasonal variation (Zeghnoun et al., 1999, Vegni et al., 2005, Pitard et al., 2004, Laurent et al., 2009). To my knowledge this is the first study of asthma or COPD epidemiology that has used such methods to account for seasonal variation. **In the absence of pollen or respiratory infections data this method is a pragmatic way to capture seasonal variation. Even in cases where pollen and respiratory infections data exist, the use of temperature data with a dynamic harmonic regression model is an efficient way to model seasonal variation of prescribing or other health outcomes that follow a seasonal pattern. In addition, the two stage modelling strategy can be useful for studies that aim to disentangle the effect of seasonal**

trigger factors from the effect of air pollution or other explanatory variables.

5.2 Limitations and Strengths

5.2.1 Exposure Misclassification

One of the main limitations of this study was the unavailability of a dense air pollution monitoring network. I therefore had to assume spatially homogenous exposure to ambient PM_{10} concentrations. The lack of any variation on ambient air quality within the study area was partially addressed by the construction of a traffic flow index by GP practice.

Exposure misclassification also arose from a lack of indoor air quality data. Exposure to indoor air pollution is associated with life style, such as smoking, housing conditions and occupation. Such data is difficult to include in an ecological study. However, such factors are correlated with income deprivation; therefore I consider that indoor air quality has been accounted for to some extent by the inclusion of the income deprivation index.

5.2.2 Respiratory Prescribing as Indicator of Health Outcome

In this study I used respiratory prescribing data as the health indicator of the outcome. This has been used by very few epidemiological studies as a proxy measure of respiratory health outcomes. Based on the results of studies from other countries, this indicator has appeared to be a useful proxy of respiratory morbidity when compared to traditional indicators such as hospital admissions and emergency room visits. Its main advantage is that it captures patients with any severity, while traditional indicators mainly capture patients who suffer from relatively severe symptoms. In the case of asthmatic patients, only 20% suffer from severe symptoms. In the case of COPD it is also estimated that the patients who suffer from severe symptoms are not the majority.

Population based prescribing data are not available in every country, their availability depends on the organisation of the health system. For example, in the United Kingdom and France it is possible to know the population that the

prescribing refers to, while in other countries such as Germany this is not known, making it impossible to use this as a health indicator of a population.

It was not possible to disentagle the prescribing data for asthmatic and COPD patients, due to limitations of the Prescribing Unit database. The RDTC had mentioned that the database would be developed further, so in the future it would be possible to access prescribing data by diagnosis.

5.2.3 Model in Primary Health Care

The majority of primary care consultations in the UK are for patients with respiratory disease (Pinnock and Sheikh, 2009, British Thoracic Society, 2006). This emphasizes the important role of primary care in managing respiratory diseases. The World Health Assembly set out an action plan in 2008, to prevent and control chronic non-communicable diseases, including asthma and COPD (World Health Organization, 2008a). Part of the action plan was to strengthen the management of the diseases at primary care level. They also suggest that accessibility to medication in primary care can be used as an indicator to monitor progress (World Health Organization, 2008a). This ecological study focused on data in primary care, linking them to environmental and lifestyle datasets and developing a model that would be able to address some of the methodological issues related to a statistical model in primary care. The approach to summarise the GP catchment areas allowed a reduction in the degree of overlap between practice areas. In addition, I demonstrated that the mixed effects model was an appropriate choice of statistical model that could address the autocorrelation issues linked to data grouped by GP practice.

5.2.4 Methodological Application

I used data that was collected routinely by different government bodies. This study's methodology demonstrated that such existing data can be subject to new tools and techniques, allowing extraction of new information and evidence from them. Current policies such as INSPIRE, support the spatial analysis of existing data (Commission of the European Communities, 2004). This has been followed by a pan-government initiative to improve the sharing and re-use of public sector location information and the implementation of INSPIRE, called UK Location (DEFRA, 2010).

To my knowledge this is the first study in asthma and COPD epidemiology to employ a harmonic regression model to account for the seasonal variation of the health outcome indicator. The output of the harmonic regression was used as input to a mixed effects model. The application of mixed effects models has increased in public health and epidemiology over the last decade. This study contributed to this field by expanding their applicability, through using a mixed model with offset. Implementing a mixed effects model with offset was one of the most recent developments in R statistical software. I initially faced problems executing the mixed models with offset (library lme4 in R software). I achieved fitting the statistical models after communicating with the developer of the algorithm for mixed effects models in lme4 and an updated version was released on the R website.

Chapter 6. Appendices

6.1 Appendix A

Asthma Insights and Reality in Europe

The AIRE findings are based upon nearly 75,000 households selected by random digit dialling in seven Western European countries (UK, France, Germany, Netherlands, Sweden, Italy and Spain) forming the largest survey in Europe to assess control of asthma. The estimated asthma prevalence from the AIRE study showed the UK to have the highest asthma prevalence rate and it was at least two fold greater than that in any other European country (Vermeire et al., 2002). The sample of the AIRE study has been found to be representative of the asthmatic population in the countries surveyed, after comparing AIRE's estimated asthma prevalence with prevalence rates found by the International Study of Asthma and Allergy in Childhood (ISAAC) and European Community Respiratory Health Survey (ECHRS) (Vermeire et al., 2002). More details on the results of the ISAAC and ECHRS are presented in section 1.1.1.1 and 1.1.1.2 respectively.

6.2 Appendix B

Prescription of Salbutamol Prescribing

Page 1.
Patient ID : [redacted]

PLEASE TICK BOX FOR THE MEDICINE(S) YOU REQUIRE.
PLEASE ALLOW 48 HOURS BEFORE COLLECTION.

Usual Doctor : [redacted]
Date Printed : 29/07/2010.
Date Printed :29/07/2010

Betamethasone Dipropionate And Salicylic Acid Ointment 0.05 % + 3 % [ ]
APPLY BD
Quant : 100 gram

Calcichew D3 Forte Chewable tablets [ ]
take two daily
Quant : 112 tablet

Dermol 500 Lotion [ ]
AS DIRECTED
Quant : 500 ml

Fluticasone Propionate Cfc-free inhaler 250 micrograms/puff [ ]
4 PUFFS BD
Quant : 4 INHALER

Formoterol Dry Powder Inhaler 12 micrograms/actuation, 60 dose [ ]
2 PUFFS TWICE DAILY
Quant : 4 INHALER

Ramipril Capsules 10 mg [ ]
TAKE ONE DAILY
Quant : 56 capsule

Salbutamol Cfc-free inhaler 100 micrograms/puff [ ]
inhale 2 doses as needed
Quant : 2 inhaler

PATIENTS – please read the notes overleaf

6.3 Appendix C

GP Practices in the Study Area

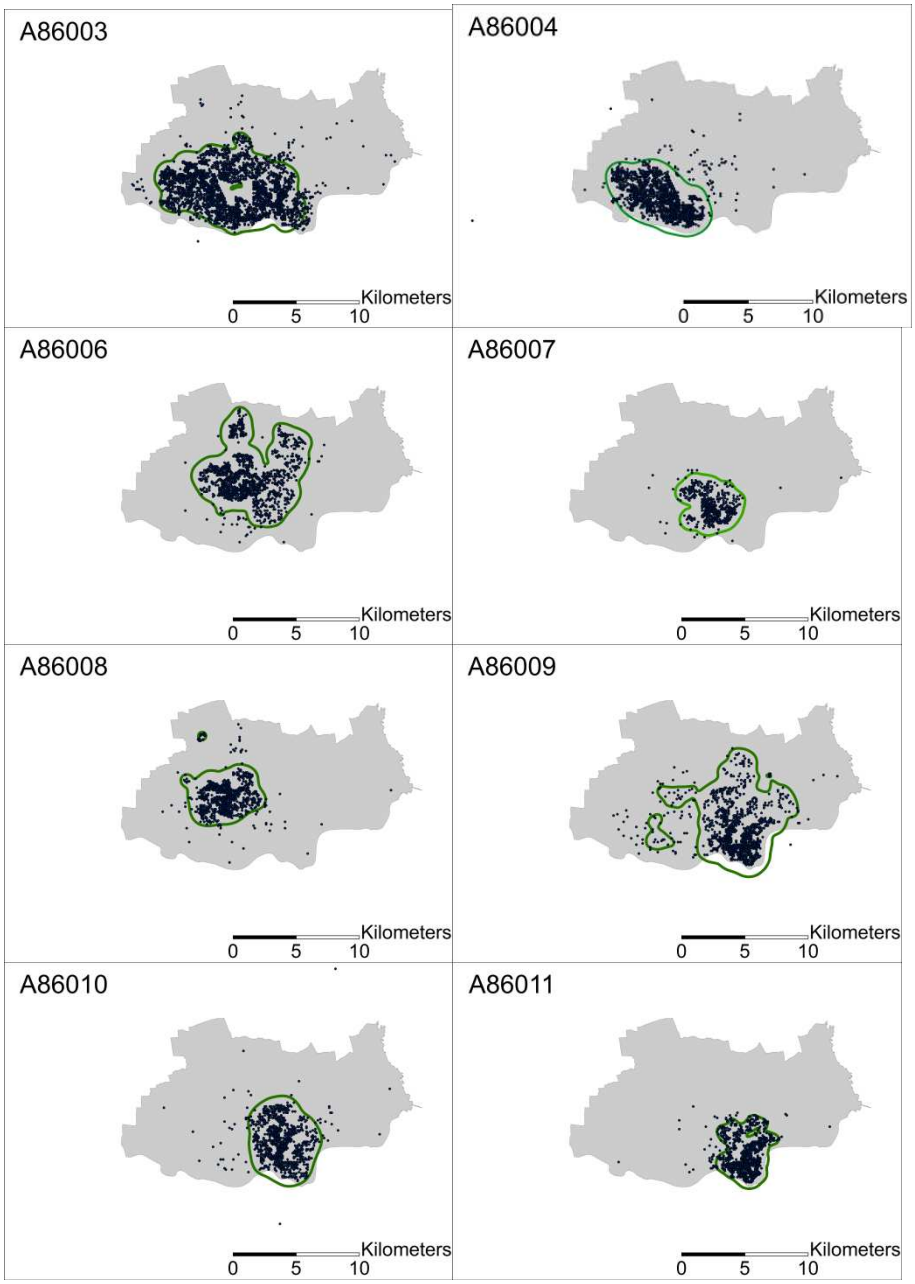
ID	GP practice code	GP practice name	Primary Care Trust	Easting	Northing
1	A86003	SAVILLE MEDICAL GROUP	NEWCASTLE	425050	564650
2	A86004	PROSPECT MEDICAL CENTRE	NEWCASTLE	422750	564450
3	A86006	ROSEWORTH SURGERY	NEWCASTLE	424450	567450
4	A86007	AVENUE MEDICAL PRACTICE	NEWCASTLE	425450	566050
5	A86008	PARK MEDICAL GROUP	NEWCASTLE	422750	568850
6	A86009	FALCON HOUSE	NEWCASTLE	426950	564950
7	A86010	BIDDLESTONE HEALTH GROUP	NEWCASTLE	427150	566150
8	A86011	WALKER MEDICAL GROUP	NEWCASTLE	429250	564350
9	A86012	WEST ROAD MEDICAL CENTRE	NEWCASTLE	422050	564650
10	A86013	DENTON PARK MEDICAL GROUP	NEWCASTLE	419350	566550
11	A86015	HOLLY MEDICAL GROUP	NEWCASTLE	425350	566150
12	A86017	CRUDDAS PARK SURGERY	NEWCASTLE	423750	563750
13	A86018	THE GROVE MEDICAL GROUP	NEWCASTLE	424450	567550
14	A86020	THE SURGERY-OSBORNE ROAD	NEWCASTLE	425050	566950
15	A86021	HOLMSIDE MEDICAL GROUP	NEWCASTLE	421850	563850
16	A86022	PARKWAY MEDICAL CENTRE	NEWCASTLE	418150	566850
17	A86023	37A MEDICAL CENTRE	NEWCASTLE	426950	565050
18	A86024	42 HEATON ROAD	NEWCASTLE	426950	565050
19	A86025	WESTERHOPE MEDICAL GROUP	NEWCASTLE	420150	567050
20	A86026	THROCKLEY PRIMARY CARE CENTRE	NEWCASTLE	415550	566550
21	A86027	NEWCASTLE MEDICAL CENTRE	NEWCASTLE	424650	565150
22	A86028	ELMFIELD HEALTH GROUP	NEWCASTLE	424250	567550
23	A86029	THORNFIELD MEDICAL GROUP	NEWCASTLE	427050	564950
24	A86030	BETTS AVENUE MEDICAL GROUP	NEWCASTLE	420850	564350
25	A86031	FENHAM HALL SURGERY	NEWCASTLE	421850	565550
26	A86032	ETHEL STREET SURGERY	NEWCASTLE	421950	563950
27	A86033	BRUNTON PARK	NEWCASTLE	424150	570450
28	A86034	ARMSTRONG ROAD HEALTH CENTRE	NEWCASTLE	420650	564050
29	A86036	GOSFORTH MEMORIAL MED.CTR	NEWCASTLE	424450	568150
30	A86037	ADELAIDE MEDICAL CENTRE	NEWCASTLE	422150	564050
31	A86038	NEWBURN SURGERY	NEWCASTLE	416550	565750
32	A86040	ST.ANTHONY'S HEALTH CENTRE	NEWCASTLE	428550	563550
33	A86601	DENTON TURRET MEDICAL CENTRE	NEWCASTLE	420150	565550
34	A86607	ELSWICK HEALTH CENTRE	NEWCASTLE	423150	563950

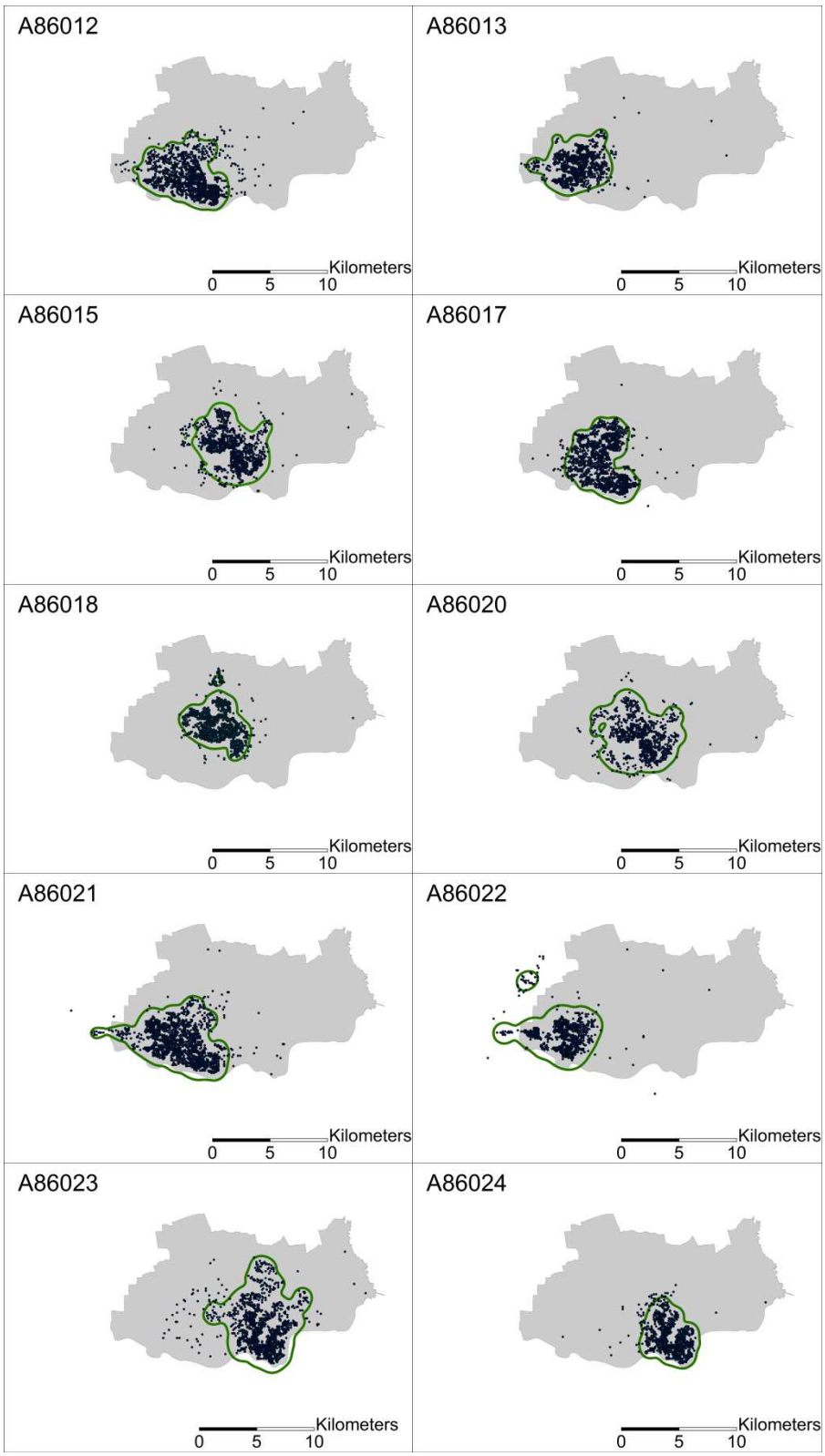
35	A86618	WELBECK ROAD MEDICAL CENTRE	NEWCASTLE	428550	564550
36	A86005	WEST FARM SURGERY	NORTH TYNESIDE	426650	568650
37	A86016	LANE END SURGERY	NORTH TYNESIDE	427150	568350
38	A86041	SWARLAND AVENUE SURGERY	NORTH TYNESIDE	427050	568050
39	A87002	SPRING TERRACE HEALTH CENTRE	NORTH TYNESIDE	435250	568650
40	A87003	PORTUGAL PLACE HEALTH CTR	NORTH TYNESIDE	429650	566250
41	A87004	COLLINGWOOD SURGERY	NORTH TYNESIDE	434850	568650
42	A87005	WHITLEY BAY HEALTH CENTRE	NORTH TYNESIDE	435850	572050
43	A87006	49 MARINE AVENUE	NORTH TYNESIDE	435150	572550
44	A87007	FOREST HALL HEALTH CENTRE	NORTH TYNESIDE	427750	569650
45	A87008	MARINE AVENUE MEDICAL CTR	NORTH TYNESIDE	434850	572350
46	A87009	PRIORY MEDICAL GROUP	NORTH TYNESIDE	435250	568550
47	A87011	BEAUMONT PARK SURGERY	NORTH TYNESIDE	433750	572750
48	A87012	WIDEOPEN MEDICAL CENTRE	NORTH TYNESIDE	424050	572750
49	A87013	BEWICKE MEDICAL CENTRE	NORTH TYNESIDE	432150	567050
50	A87014	EARSDON PARK MEDICAL PRACTICE	NORTH TYNESIDE	431650	571650
51	A87015	APPLEBY SURGERY	NORTH TYNESIDE	434850	568650
52	A87016	THE VILLAGE GREEN SURGERY	NORTH TYNESIDE	430250	566950
53	A87017	WOODLANDS PARK HEALTH CTR	NORTH TYNESIDE	423850	572450
54	A87019	THE WESTGARTH PRACTICE	NORTH TYNESIDE	435150	568450
55	A87020	MONKSEATON MEDICAL CENTRE	NORTH TYNESIDE	433850	571550
56	A87022	THE BOWMAN PRACTICE	NORTH TYNESIDE	431650	571650
57	A87023	THE SMITH PRACTICE	NORTH TYNESIDE	431650	571650
58	A87027	GARDEN PARK SURGERY	NORTH TYNESIDE	432250	567550
59	A87029	PARK ROAD MEDICAL PRACT	NORTH TYNESIDE	430050	566550
60	A87030	WALLSEND ROAD SURGERY	NORTH TYNESIDE	434050	568050

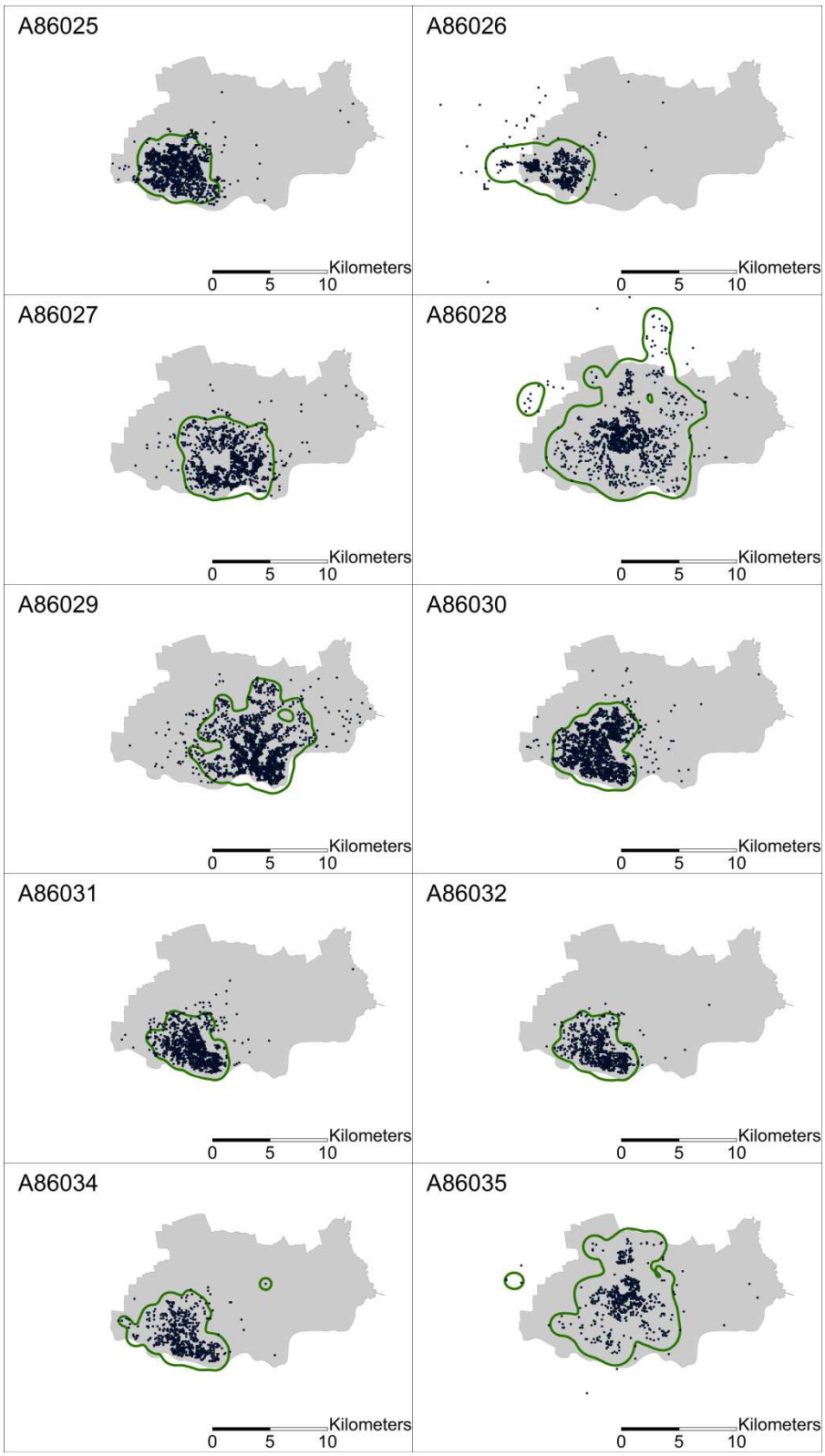
61	A87600	PARK PARADE PRACTICE	NORTH TYNESIDE	435250	572150
62	A87608	FRIARSLEIGH HEALTH CENTRE	NORTH TYNESIDE	426550	568650
63	A87612	WELLSPRING MEDICAL PRACT.	NORTH TYNESIDE	427950	571450
64	A87615	PRESTON & AUSTIN PRACTICE	NORTH TYNESIDE	427950	571450

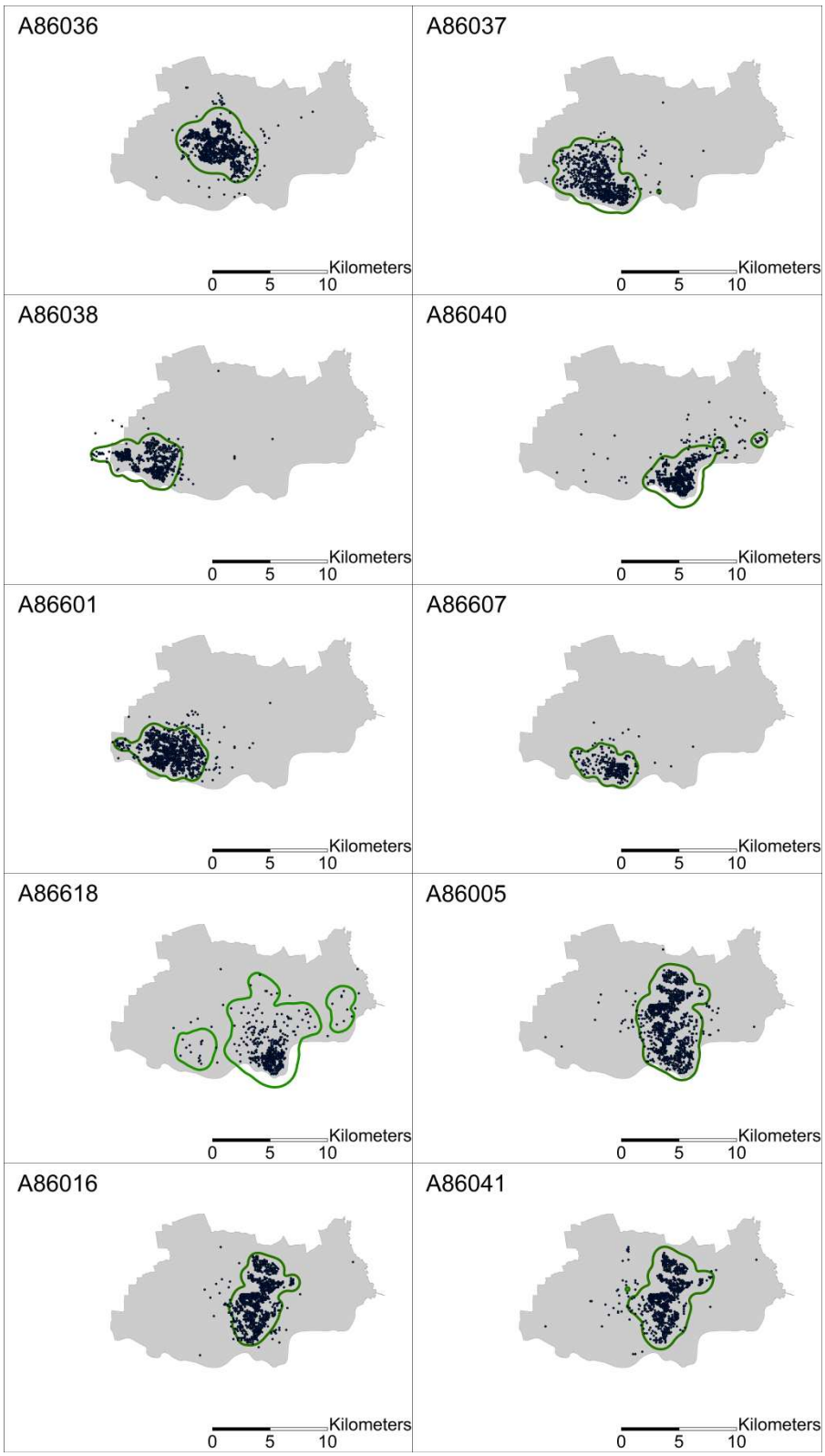
6.4 Appendix D

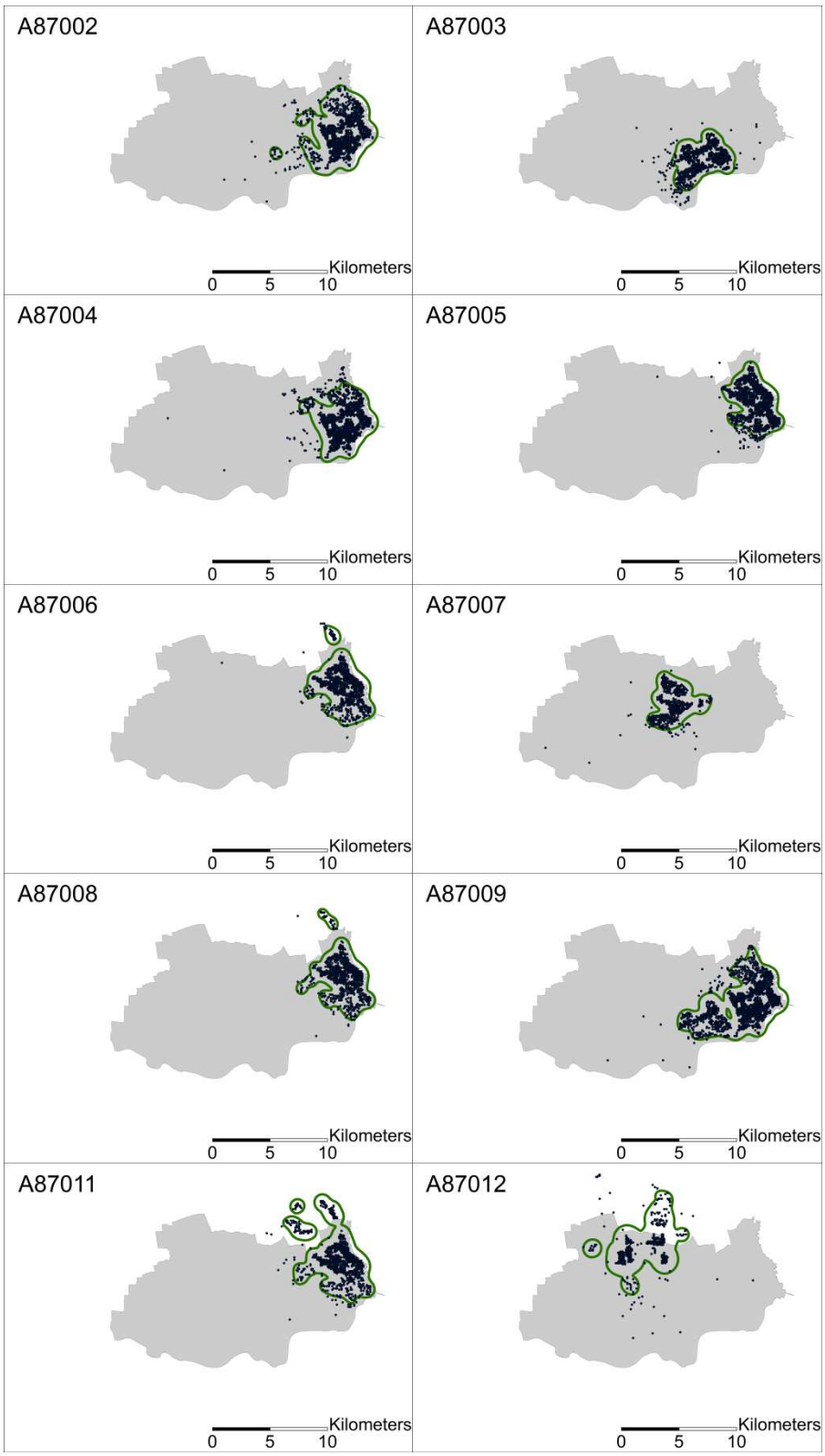
Service areas by GP practice

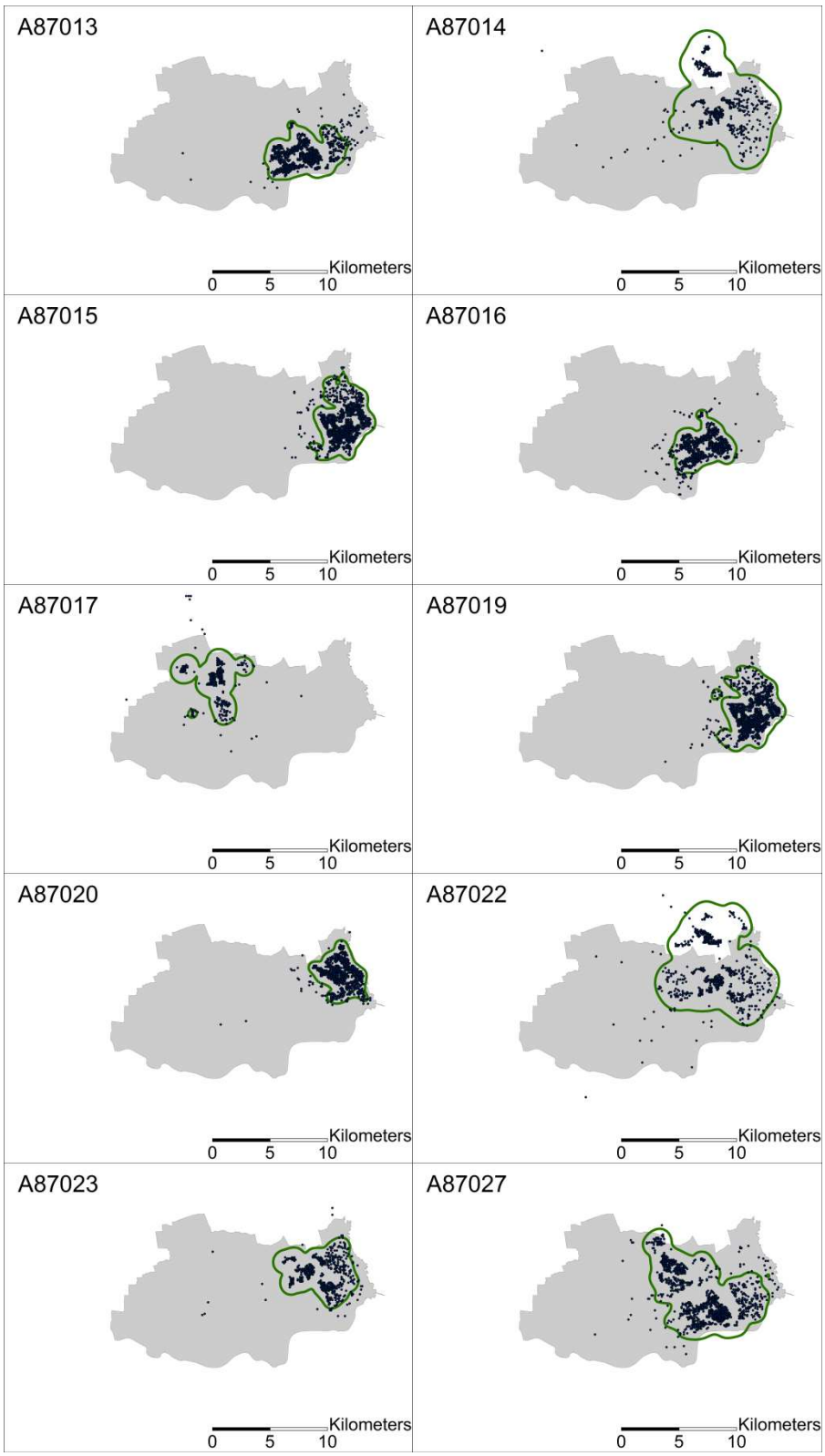


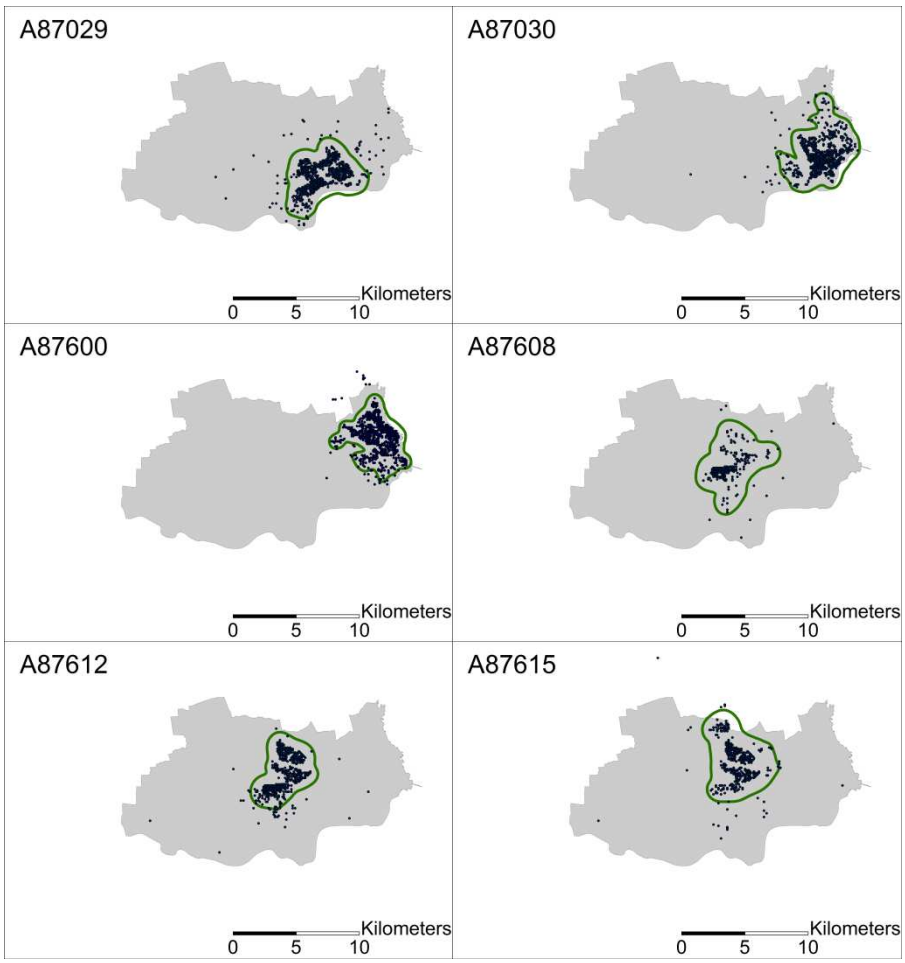

A86012
Kilometers
0 5 10
A86013
Kilometers
0 5 10
A86015
Kilometers
0 5 10
A86017
Kilometers
0 5 10
A86018
Kilometers
0 5 10
A86020
Kilometers
0 5 10
A86021
Kilometers
0 5 10
A86022
Kilometers
0 5 10
A86023
Kilometers
0 5 10
A86024
Kilometers
0 5 10


A86025
Kilometers
0 5 10
A86026
Kilometers
0 5 10
A86027
Kilometers
0 5 10
A86028
Kilometers
0 5 10
A86029
Kilometers
0 5 10
A86030
Kilometers
0 5 10
A86031
Kilometers
0 5 10
A86032
Kilometers
0 5 10
A86034
Kilometers
0 5 10
A86035
Kilometers
0 5 10


A86036
Kilometers
0 5 10
A86037
Kilometers
0 5 10
A86038
Kilometers
0 5 10
A86040
Kilometers
0 5 10
A86601
Kilometers
0 5 10
A86607
Kilometers
0 5 10
A86618
Kilometers
0 5 10
A86005
Kilometers
0 5 10
A86016
Kilometers
0 5 10
A86041
Kilometers
0 5 10

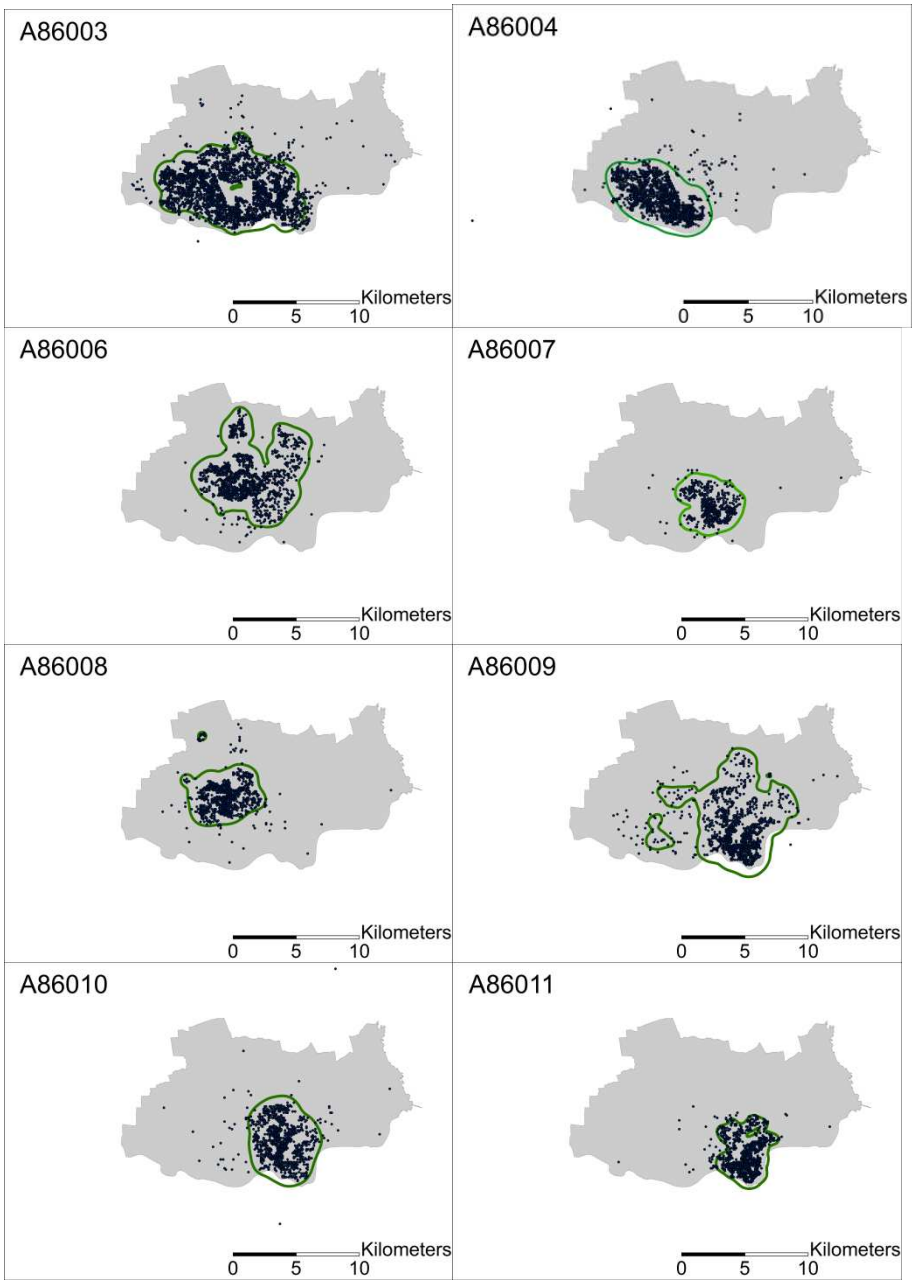

A87002
Kilometers
0 5 10
A87003
Kilometers
0 5 10
A87004
Kilometers
0 5 10
A87005
Kilometers
0 5 10
A87006
Kilometers
0 5 10
A87007
Kilometers
0 5 10
A87008
Kilometers
0 5 10
A87009
Kilometers
0 5 10
A87011
Kilometers
0 5 10
A87012
Kilometers
0 5 10

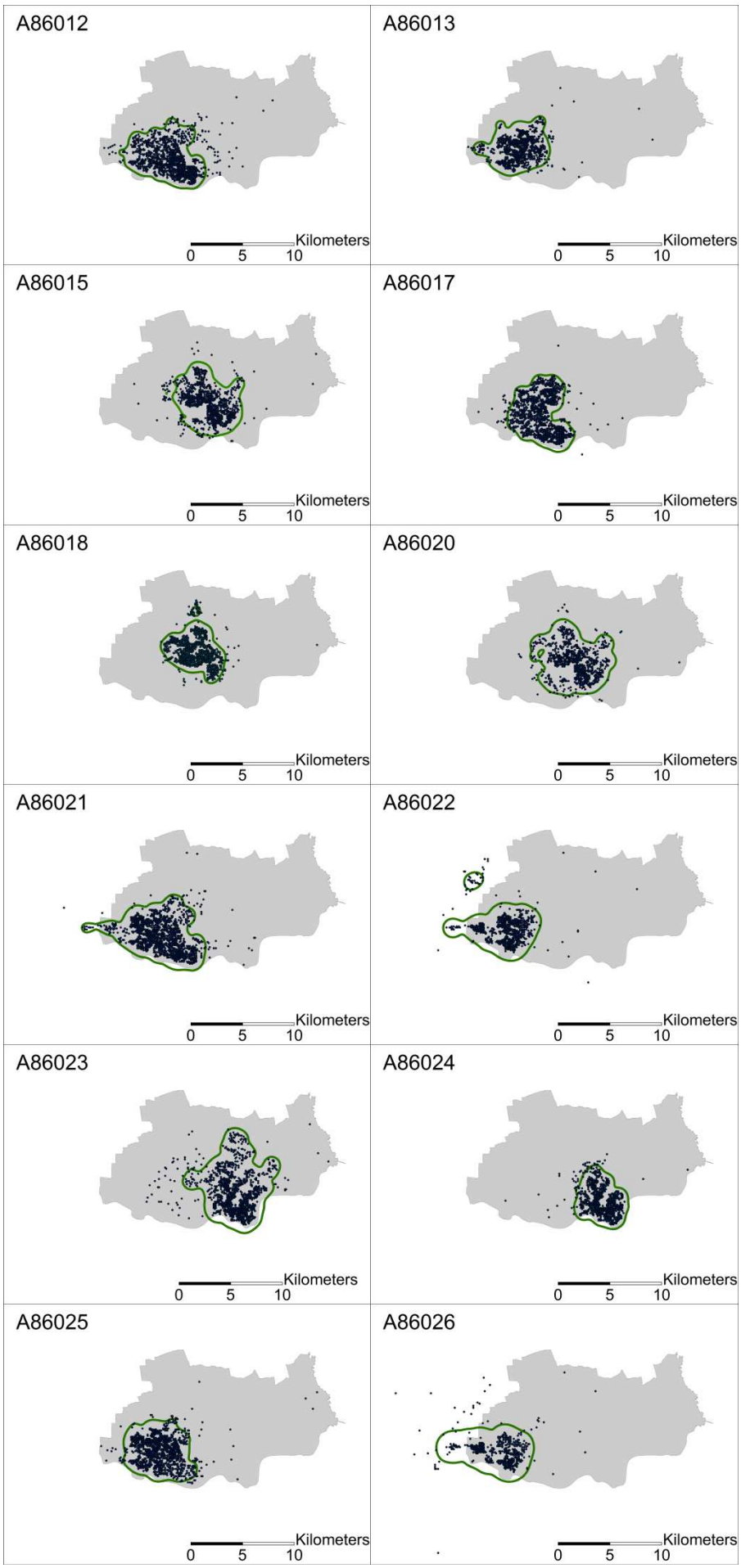

A87013
Kilometers
0 5 10
A87014
Kilometers
0 5 10
A87015
Kilometers
0 5 10
A87016
Kilometers
0 5 10
A87017
Kilometers
0 5 10
A87019
Kilometers
0 5 10
A87020
Kilometers
0 5 10
A87022
Kilometers
0 5 10
A87023
Kilometers
0 5 10
A87027
Kilometers
0 5 10

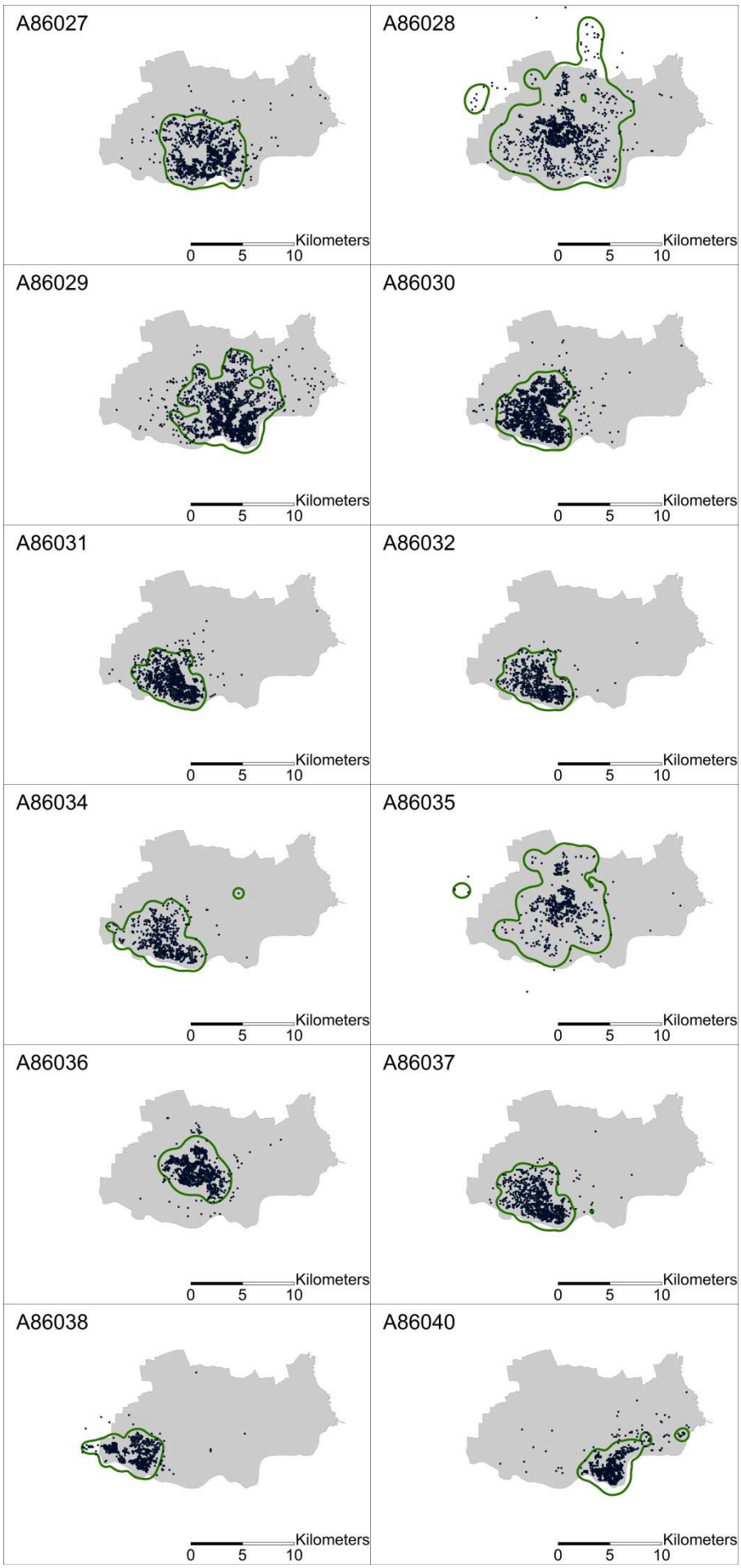

A87029
Kilometers
0 5 10
A87030
Kilometers
0 5 10
A87600
Kilometers
0 5 10
A87608
Kilometers
0 5 10
A87612
Kilometers
0 5 10
A87615
Kilometers
0 5 10

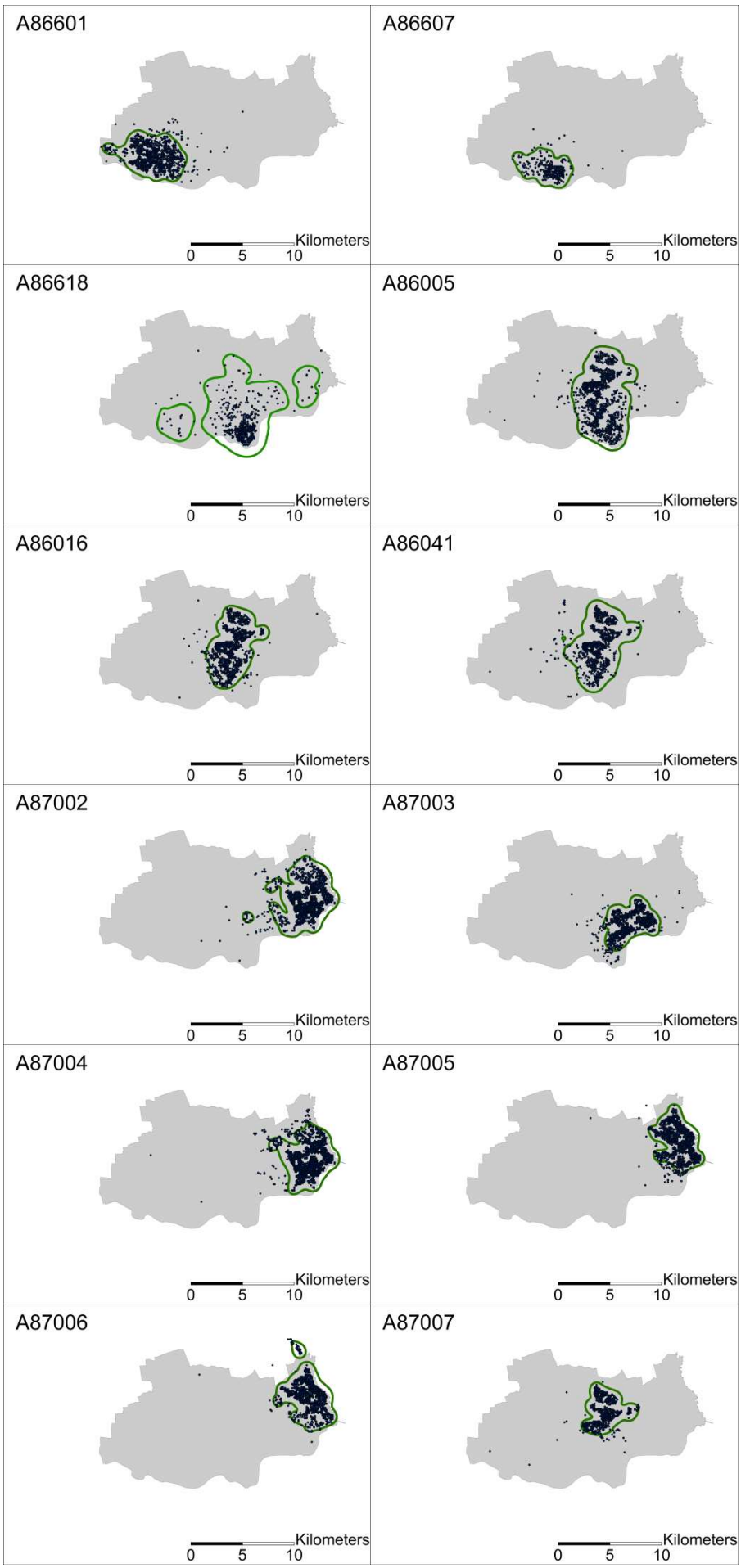
6.5 Appendix E

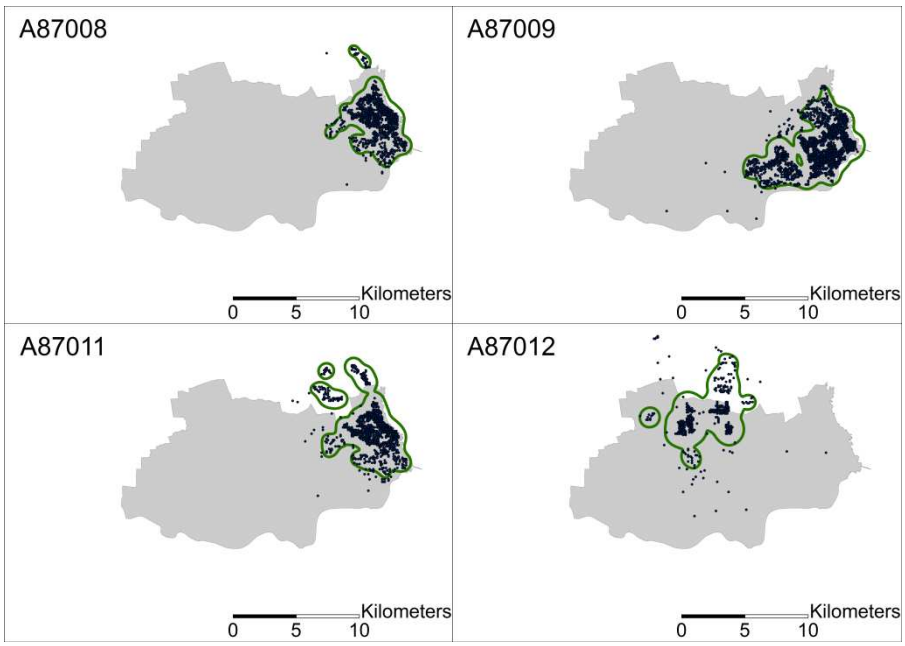
Service areas per year (2002-2006), by GP practice

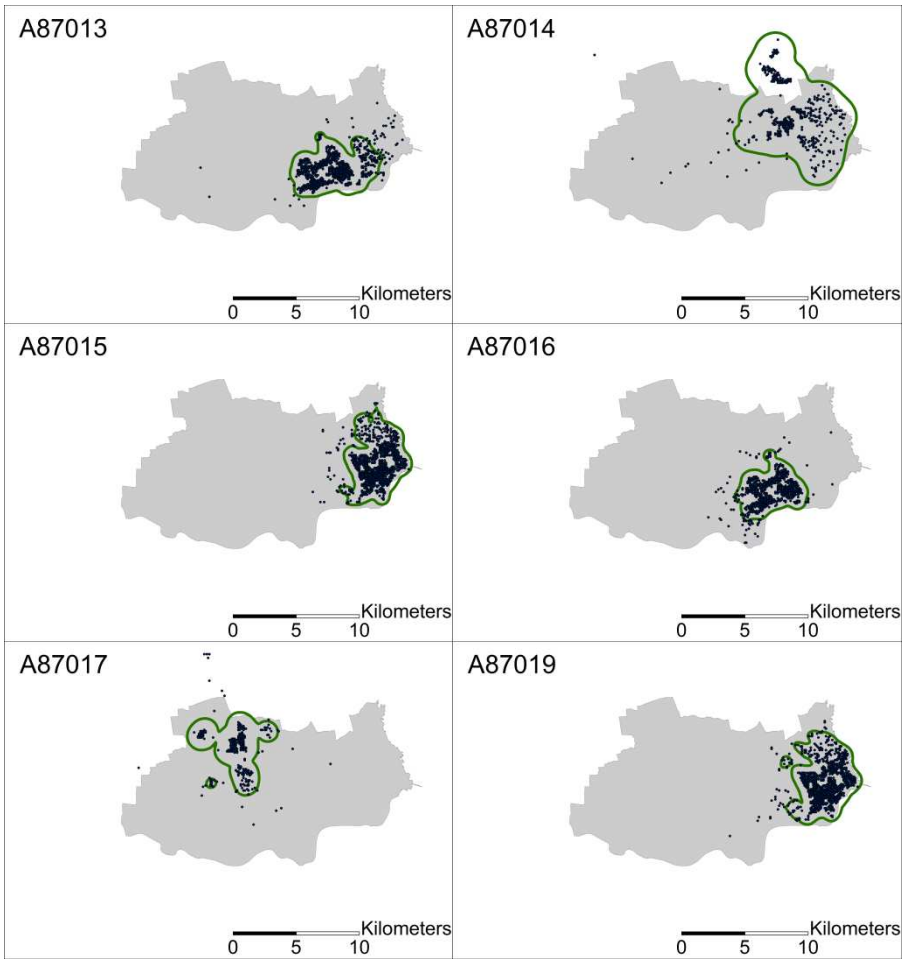


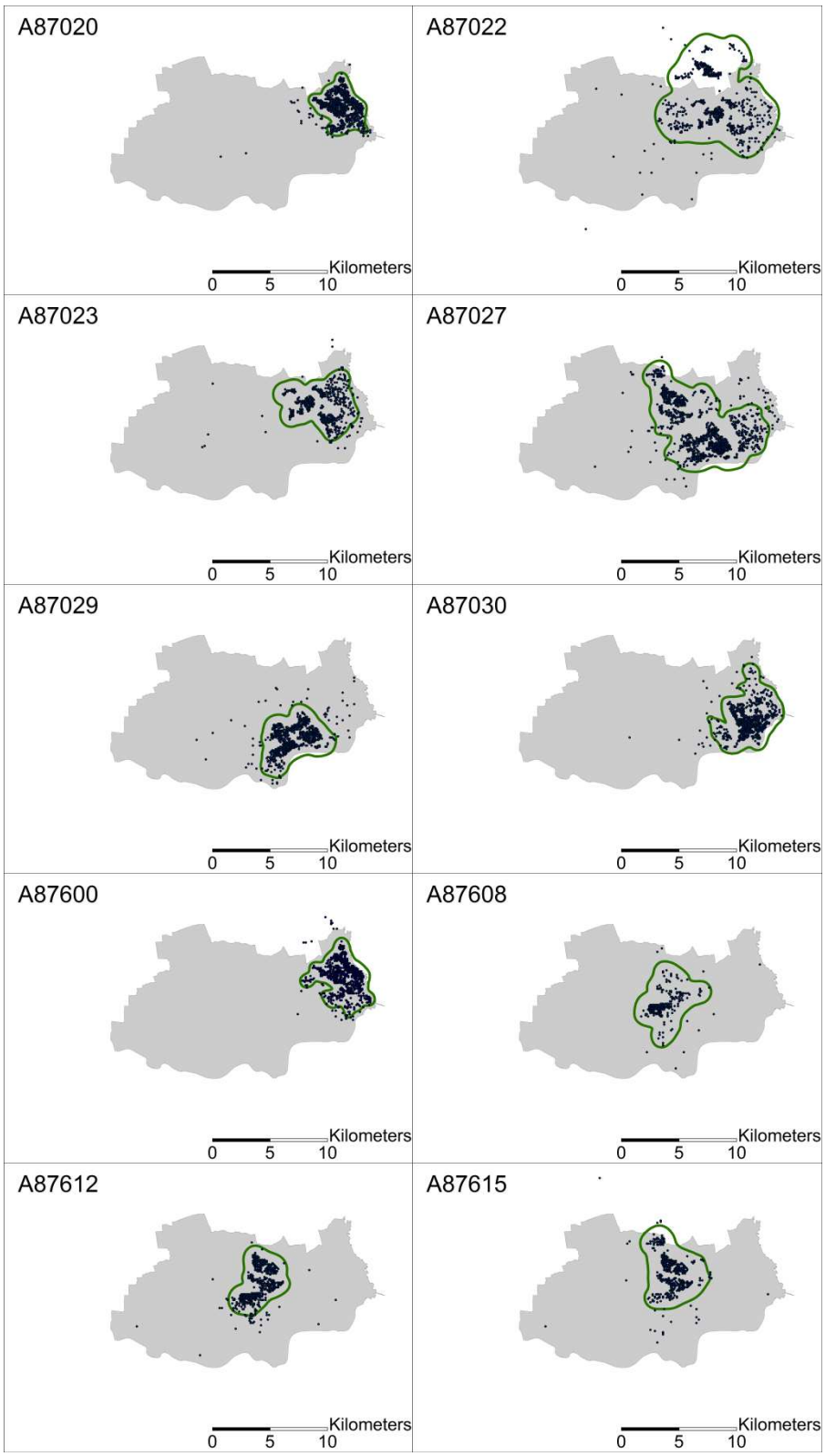

A86012
Kilometers
0 5 10
A86013
Kilometers
0 5 10
A86015
Kilometers
0 5 10
A86017
Kilometers
0 5 10
A86018
Kilometers
0 5 10
A86020
Kilometers
0 5 10
A86021
Kilometers
0 5 10
A86022
Kilometers
0 5 10
A86023
Kilometers
0 5 10
A86024
Kilometers
0 5 10
A86025
Kilometers
0 5 10
A86026
Kilometers
0 5 10


A86027
Kilometers
0 5 10
A86028
Kilometers
0 5 10
A86029
Kilometers
0 5 10
A86030
Kilometers
0 5 10
A86031
Kilometers
0 5 10
A86032
Kilometers
0 5 10
A86034
Kilometers
0 5 10
A86035
Kilometers
0 5 10
A86036
Kilometers
0 5 10
A86037
Kilometers
0 5 10
A86038
Kilometers
0 5 10
A86040
Kilometers
0 5 10


A86601
Kilometers
0 5 10
A86607
Kilometers
0 5 10
A86618
Kilometers
0 5 10
A86005
Kilometers
0 5 10
A86016
Kilometers
0 5 10
A86041
Kilometers
0 5 10
A87002
Kilometers
0 5 10
A87003
Kilometers
0 5 10
A87004
Kilometers
0 5 10
A87005
Kilometers
0 5 10
A87006
Kilometers
0 5 10
A87007
Kilometers
0 5 10


A87008
Kilometers
0 5 10
A87009
Kilometers
0 5 10
A87011
Kilometers
0 5 10
A87012
Kilometers
0 5 10


A87013
Kilometers
0 5 10
A87014
Kilometers
0 5 10
A87015
Kilometers
0 5 10
A87016
Kilometers
0 5 10
A87017
Kilometers
0 5 10
A87019
Kilometers
0 5 10


A87020
Kilometers
0 5 10
A87022
Kilometers
0 5 10
A87023
Kilometers
0 5 10
A87027
Kilometers
0 5 10
A87029
Kilometers
0 5 10
A87030
Kilometers
0 5 10
A87600
Kilometers
0 5 10
A87608
Kilometers
0 5 10
A87612
Kilometers
0 5 10
A87615
Kilometers
0 5 10

6.6 Appendix F

Traffic Monitoring Sites

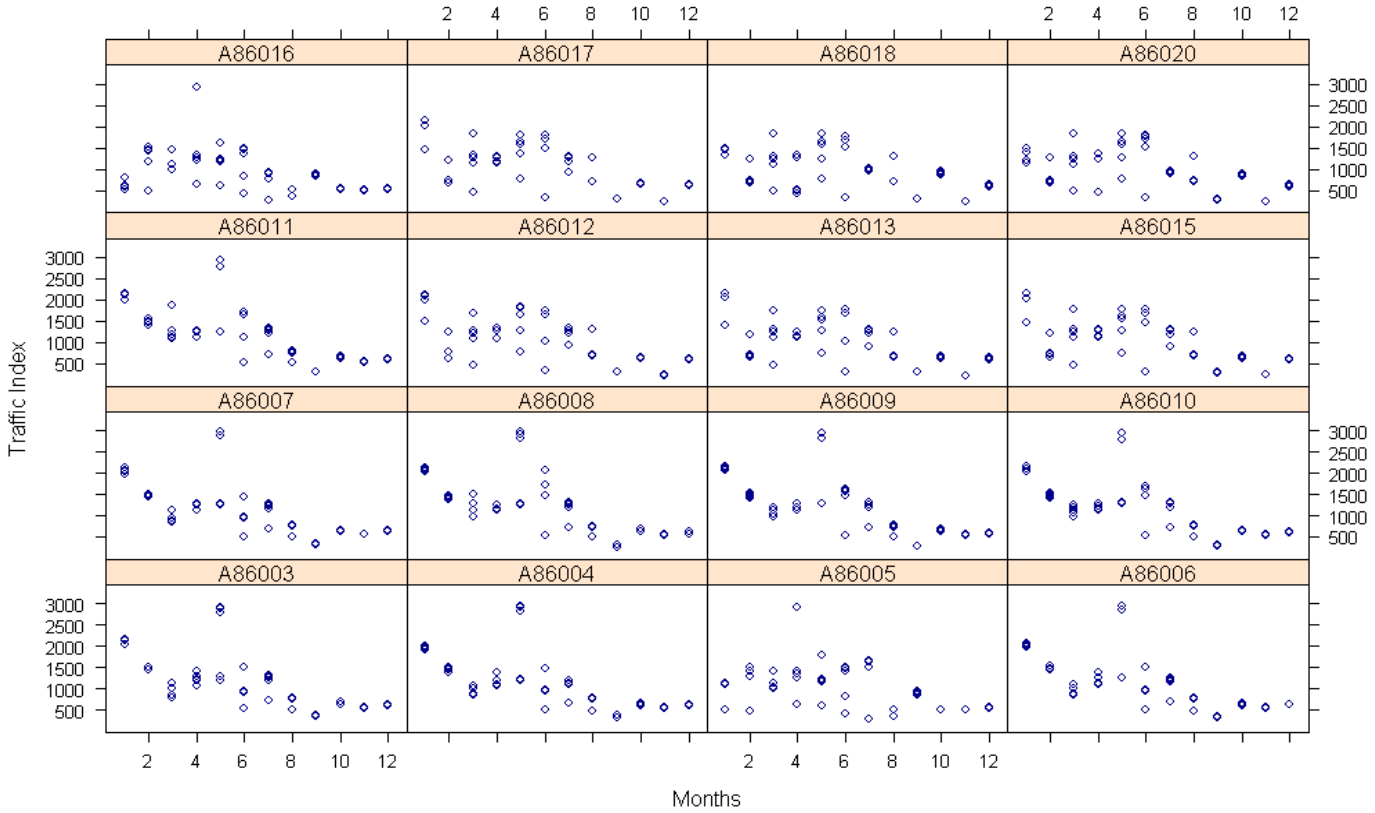
ID	Site code	Location	Easting	Northing	Road Type	Road Name
1	1	B1318 GT NORTH ROAD N OF FORSYTH RD	424690	566610	Broads	B1318
2	2	A189 REDHEUGH BRIDGE (ON NORTH SIDE)	424220	563500	Aroads	A189
3	5	A167 TYNE BRIDGE (ON NORTH SIDE)	425180	564040	Aroads	A167
4	7	A1058 COAST ROAD W OF STATION RD. WALLSEND	429000	567240	Aroads	A1058
5	8	B1318 Great North Road South of Brunton Lane	424220	570400	Broads	B1318
6	14	A1 NORTH OF SEATON BURN INT.	423140	575000	Aroads	A1
7	16	A1058 COAST ROAD W OF NORHAM ROAD	432990	568680	Aroads	A1058
8	29	A167 NORTH WEST RADIAL S.E. GRANSTAND RD	423820	565750	Aroads	A167
9	35	A193 BYKER BRIDGE 20M E OF STODDART STREET	425970	564570	Aroads	A193
10	59	A189 SPINE ROAD N OF SANDY LANE	426160	571900	Aroads	A189
11	69	A19 NORTH OF TYNE TUNNEL APPROACH R/BOUT	433070	567020	Aroads	A19
12	70	A19 SOUTH OF COAST ROAD INTERCHANGE	432440	568040	Aroads	A19
13	71	A19 NORTH OF COAST ROAD INTERCHANGE	431710	568970	Aroads	A19
14	73	A19 NORTH OF B1322/B1318 - BACKWORTH LANE	428200	573440	Aroads	A19
15	80	A1148 MONKSEATON DRIVE W OF REDHOUSE DRIVE	433300	572350	Aroads	A1148
16	88	A1058 BEACH ROAD W OF PRESTON NORTH ROAD	434280	569520	Aroads	A1058
17	89	A192 PRESTON NORTH ROAD S OF A191 RAKE LANE	434700	570533	Aroads	A192
18	205	A19 Tyne Tunnel (south portal)	433181	565125	Aroads	A19
19	206	A1058 (CRADLEWELL) JESMOND RD. S.W. OSBORNE AVE.	426000	565850	Aroads	A1058
20	401	A1056 SANDY LANE E OF M.O.T. STATION	425160	571830	Aroads	A1056
21	402	A1056 OLD GT. NORTH RD. N OF GOSFORTH PK HOTEL	424180	571470	Aroads	A1056
22	403	A191 HILLHEADS ROAD W OF MARDEN ROAD	435340	571840	Aroads	A191

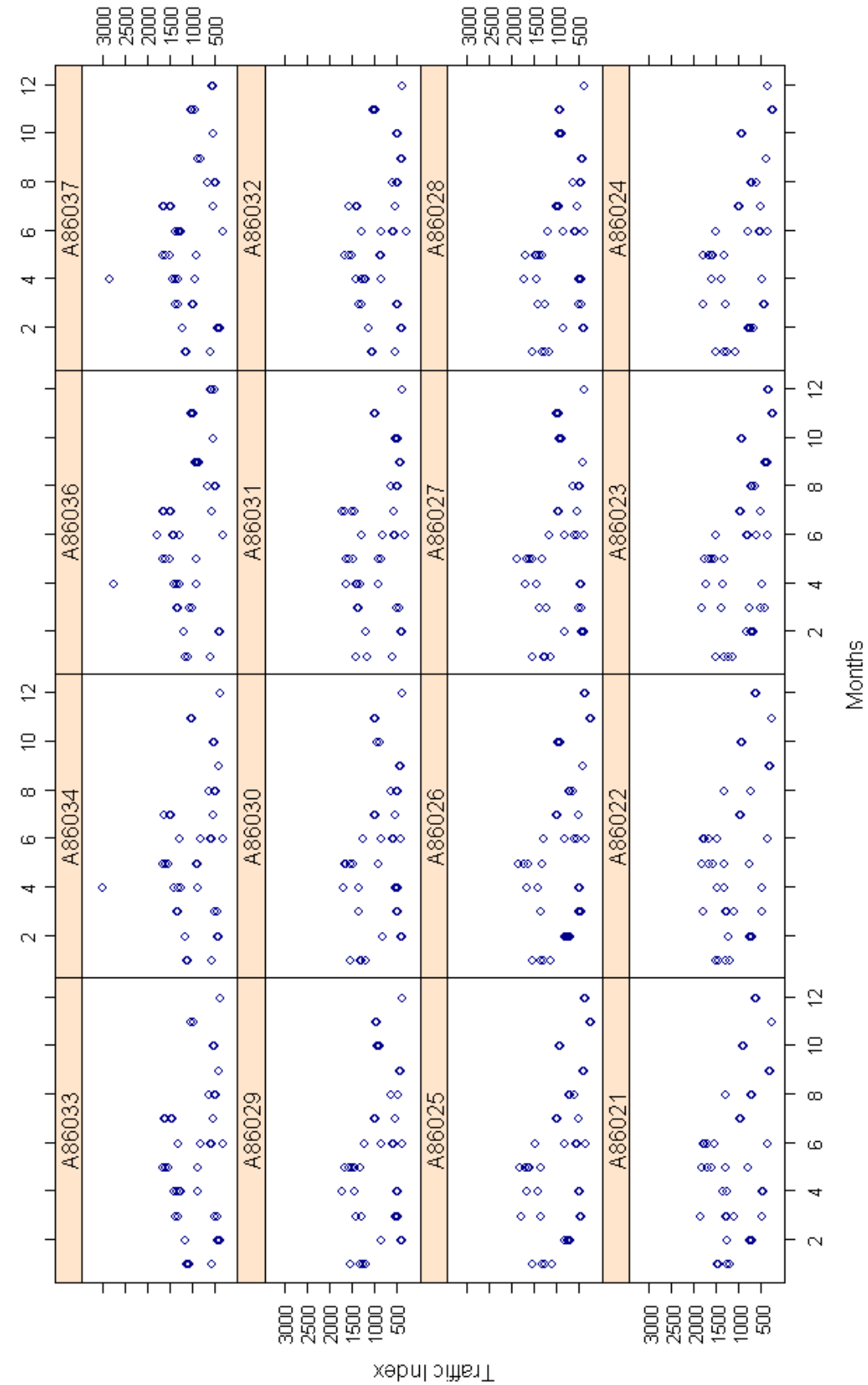
23	404	A189 SPINE ROAD S OF SANDY LANE	426170	571470	Aroads	A189
24	405	A1056 KILLINGWORTH RD	426690	571710	Aroads	A1056
25	407	A186 EARSDON ROAD S.W. OF PARK LANE SHIREMOOR	431630	571740	Aroads	A186
26	408	A186 STATION ROAD NORTH N OF HOTSPUR ROAD	429000	568370	Aroads	A186
27	409	B1505 GREAT LIME ROAD E OF NORTHFIELD DRIVE	426920	570500	Broads	B1505
28	410	A191 WHITLEY ROAD E OF STATION ROAD NORTH	429270	569350	Aroads	A191
29	415	A193 WALLSEND ROAD E OF A1 INTERCHANGE	433090	567410	Aroads	A193
30	423	A1056 CAMPERDOWN E. OF NORTHGATE	427350	571850	Aroads	A1056
31	424	A191 RAKE LANE 330M S.W. PRESTON NORTH ROAD	434300	570480	Aroads	A191
32	429	A1058 COAST ROAD 270M W OF ST. PETER'S ROAD	430690	567850	Aroads	A1058
33	430	A193 THE LINKS 200M S OF MONKSEATON DRIVE	435170	573140	Aroads	A193
34	458	A189 N. NORTH FARM AVENUE	425650	568900	Aroads	A189
35	461	A190 30M E. A189 (DUDLEY) R/BOUT	427270	573580	Aroads	A190
36	611	A1 WESTERN BY-PASS N OF BLAYDON BRIDGE	419430	564730	Aroads	A1
37	612	A1 WESTERN BY-PASS N OF A69 INTERCHANGE	419870	565900	Aroads	A1
38	617	A696 S.W. OF BLACK CALLERTON LANE	418900	570300	Aroads	A696
39	618	A696 WOOLSINGTON BY-PASS S OF AIRPORT R/BOUT	418430	571050	Aroads	A696
40	619	A696 WOOLSINGTON BY-PASS N OF AIRPORT R/BOUT	418230	571570	Aroads	A696
41	628	A189 HADDRICKSMILL ROAD 50M N OF DENE CRESCENT	425440	567720	Aroads	A189
42	632	B1307 SANDYFORD ROAD 40M N.E. OF GOLDSPINK LANE	425840	565550	Broads	B1307
43	650	A695 SCOTSWOOD ROAD 350M E. OF SCOTSWOOD BRIDGE	420450	563650	Aroads	A695
44	660	A695 SCOTSWOOD BRIDGE INTERCHANGE GATESHEAD SIDE	419970	563510	Aroads	A695
45	1071	A189 WEST CENTRAL ROUTE N. OF SUNDERLAND ST	424227	563977	Aroads	A189
46	1072	A189 WEST CENTRAL ROUTE N. WESTGATE ROAD	424280	564120	Aroads	A189
47	1073	A189 WEST CENTRAL ROUTE S. GALLOWGATE	424287	564300	Aroads	A189

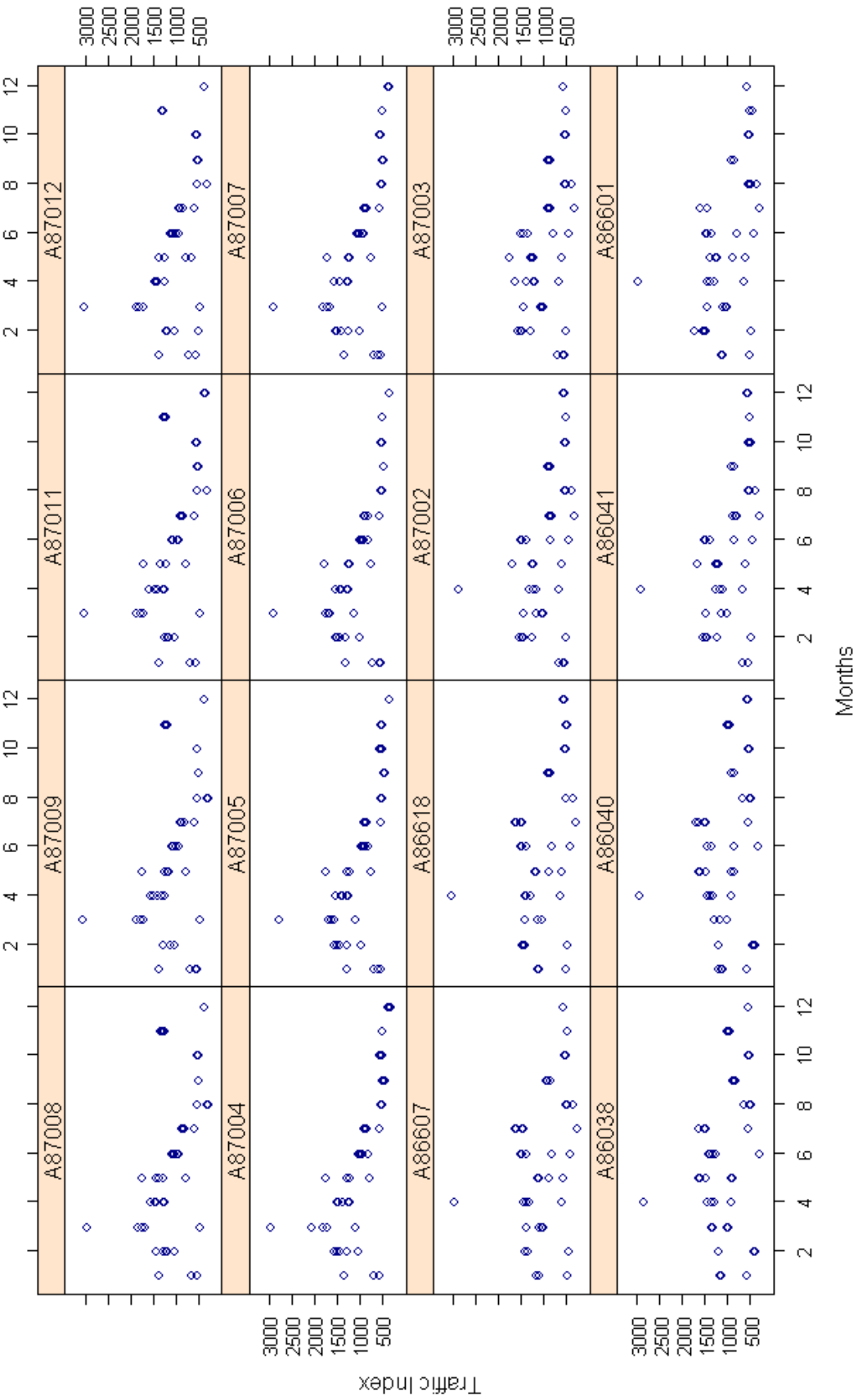
48	1075	A189 WEST CENTRAL ROUTE NORTH OF WALTER TER	423757	564909	Aroads	A189
49	1076	A189 WEST CENTRAL ROUTE NORTH OF BRIGHTON GROVE	423273	565394	Aroads	A189

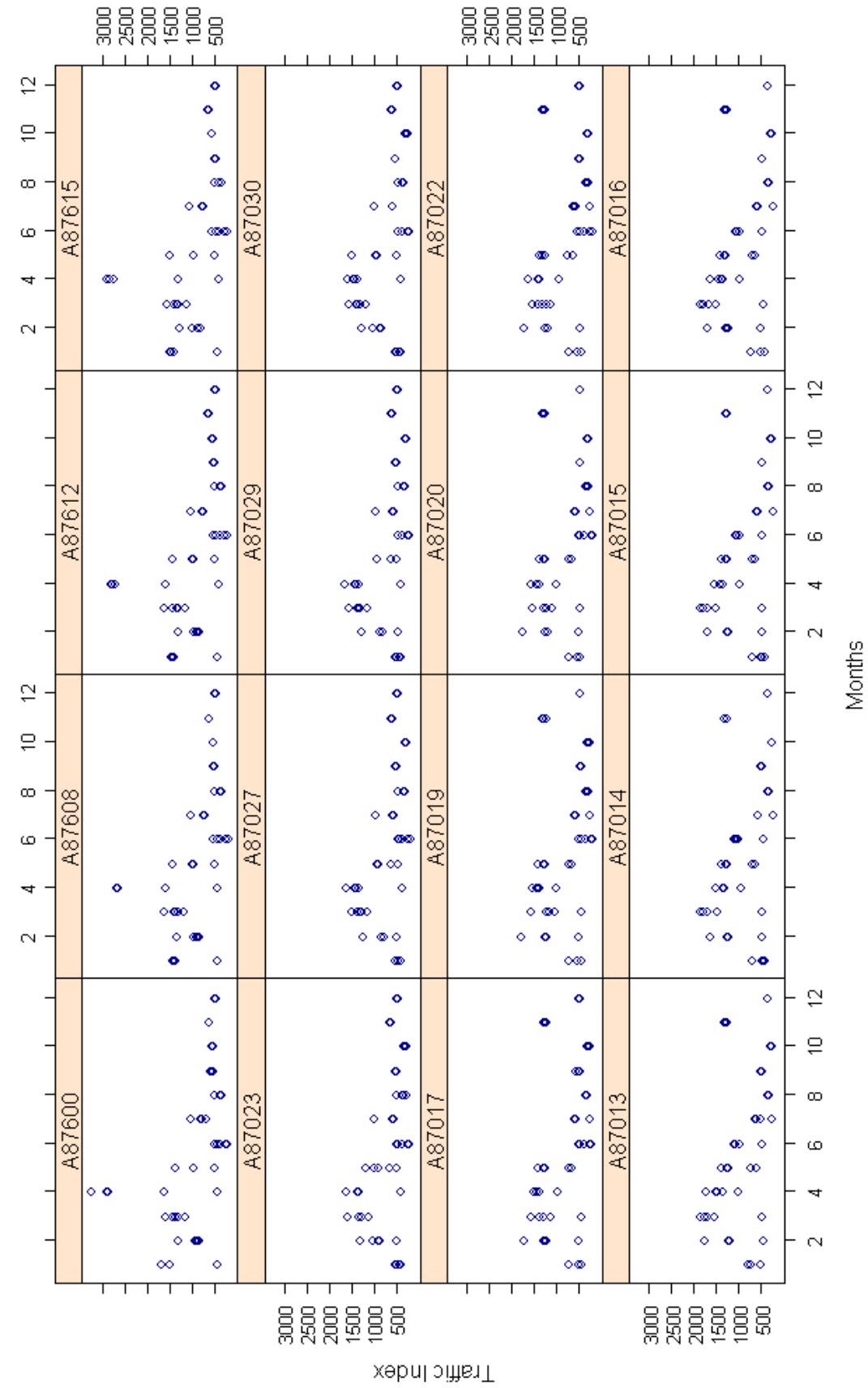
6.7 Appendix G

Traffic index per month over the 5 years study period, by GP practice



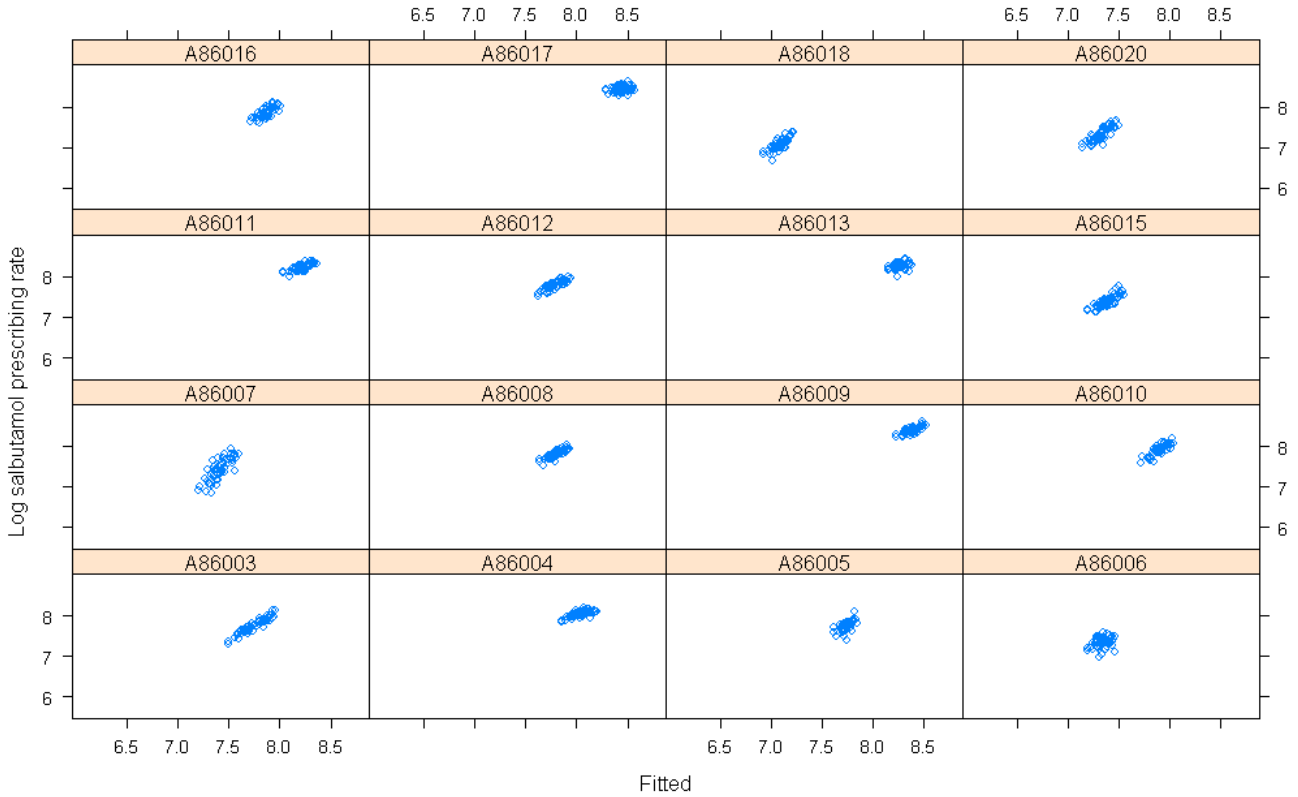

A86033
A86034
A86036
A86037
A86029
A86030
A86031
A86032
A86025
A86026
A86027
A86028
A86021
A86022
A86023
A86024
Traffic Index
Months
500
1000
1500
2000
2500
3000
2
4
6
8
10
12


A87008
A87009
A87011
A87012
A87004
A87005
A87006
A87007
A86607
A86618
A86002
A86003
A86038
A86040
A86041
A86601
Traffic Index
Months
500
1000
1500
2000
2500
3000
2
4
6
8
10
12


A87600
A87608
A87612
A87615
A87023
A87027
A87029
A87030
A87017
A87019
A87020
A87022
A87013
A87014
A87015
A87016
Traffic Index
Months
500
1000
1500
2000
2500
3000
2
4
6
8
10
12

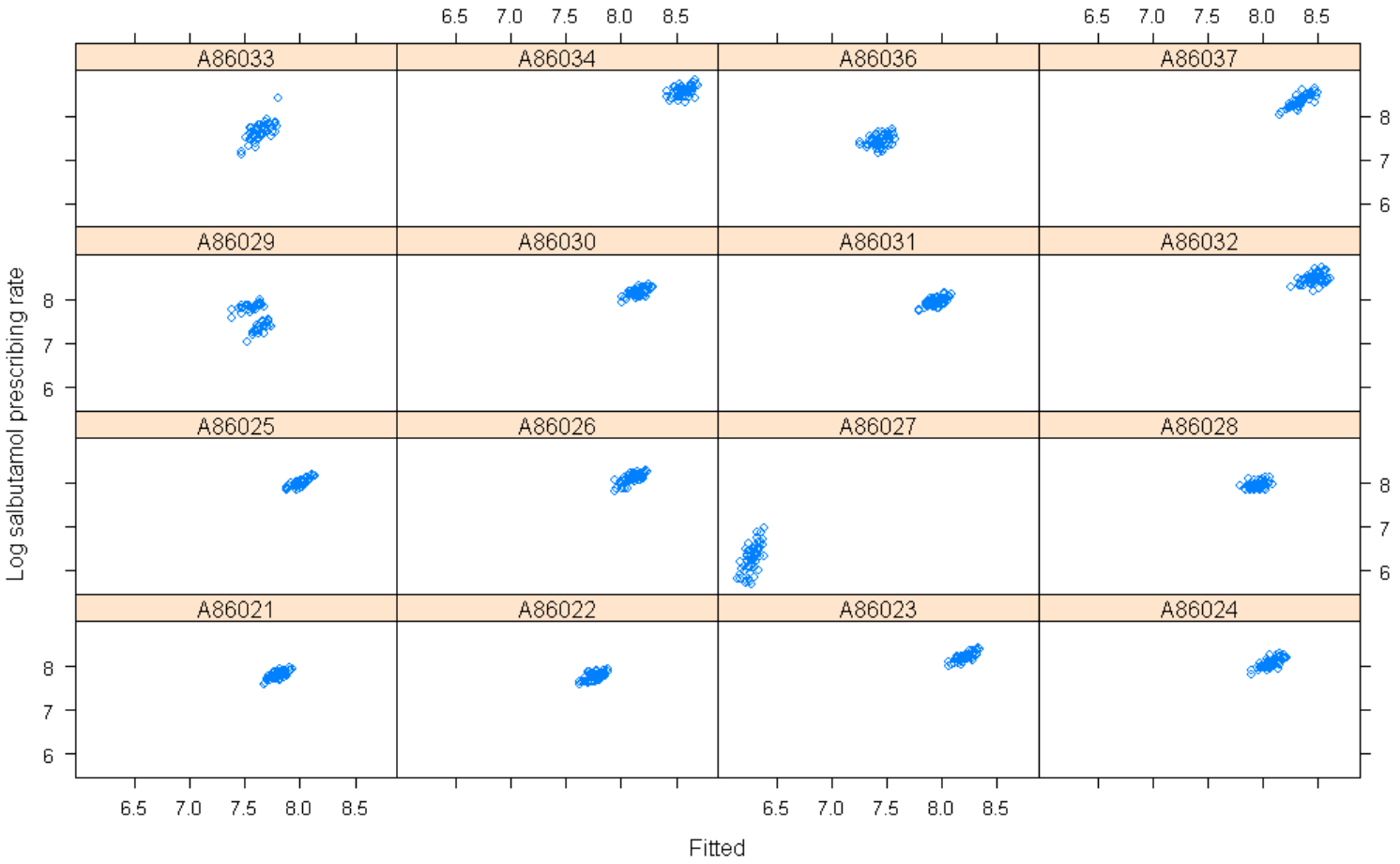
6.8 Appendix H

Diagnostics of the final model –
Observed against fitted values of the final model by GP practice

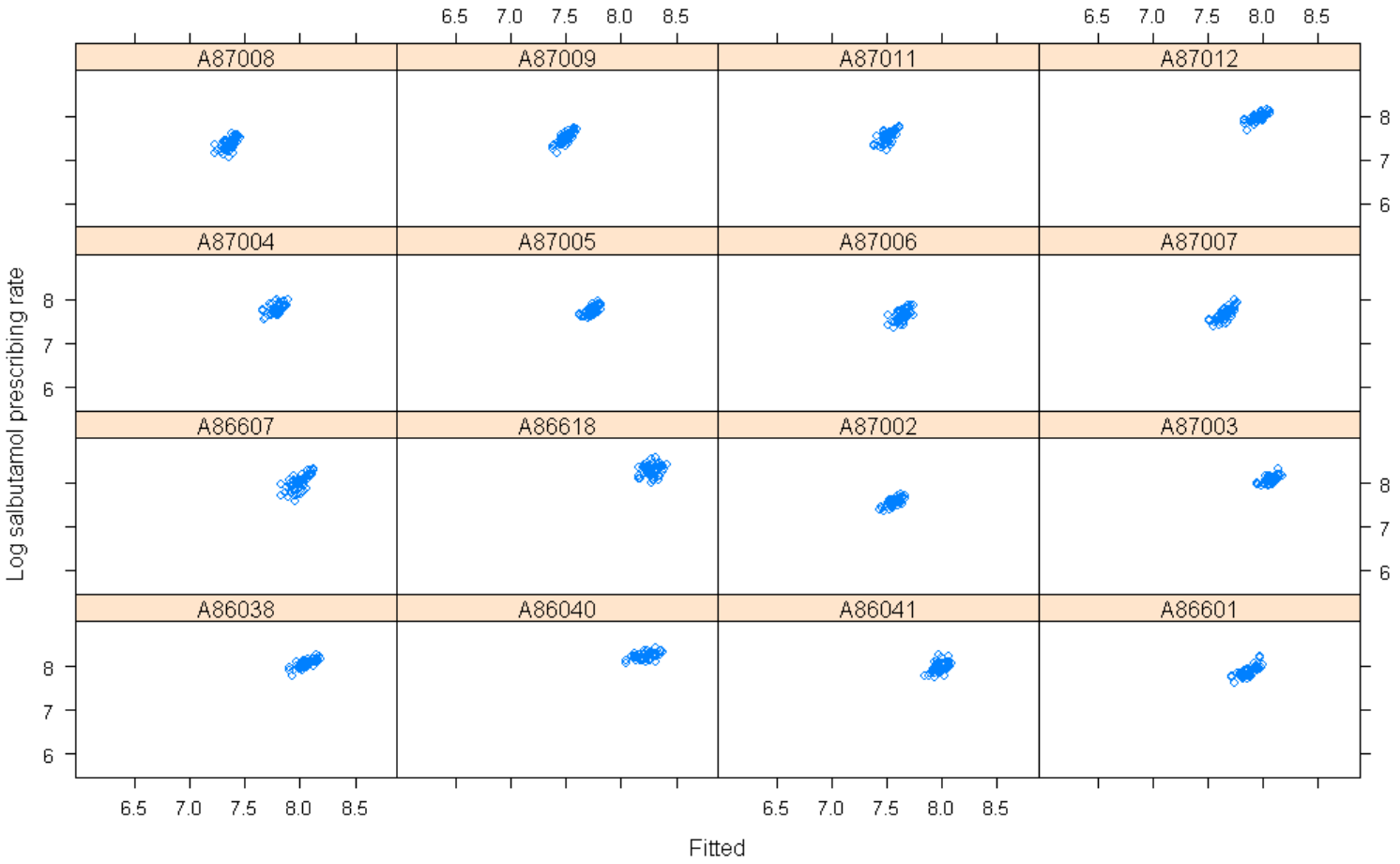


(a)

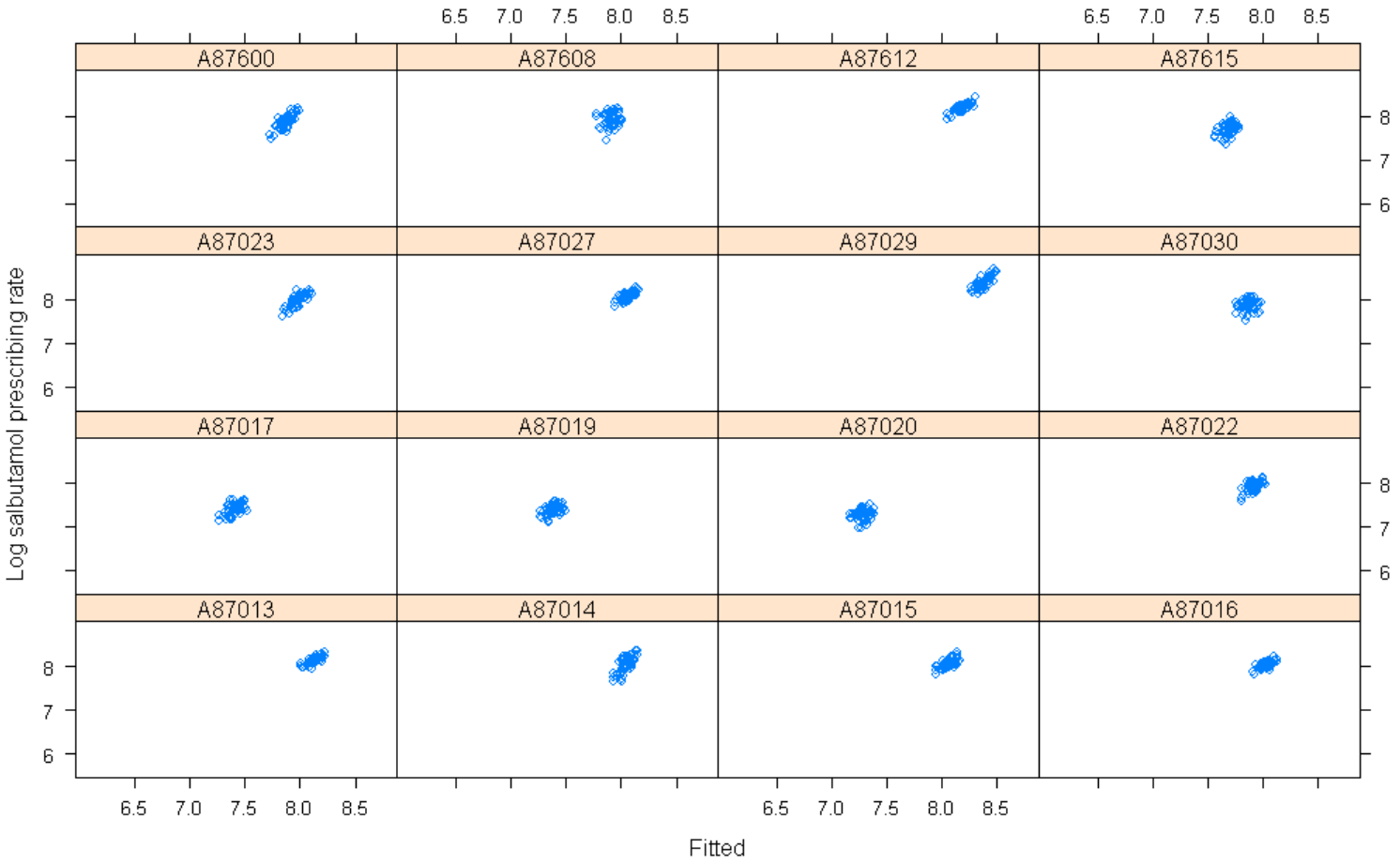
Observed against fitted values of the final model by GP practice (b)



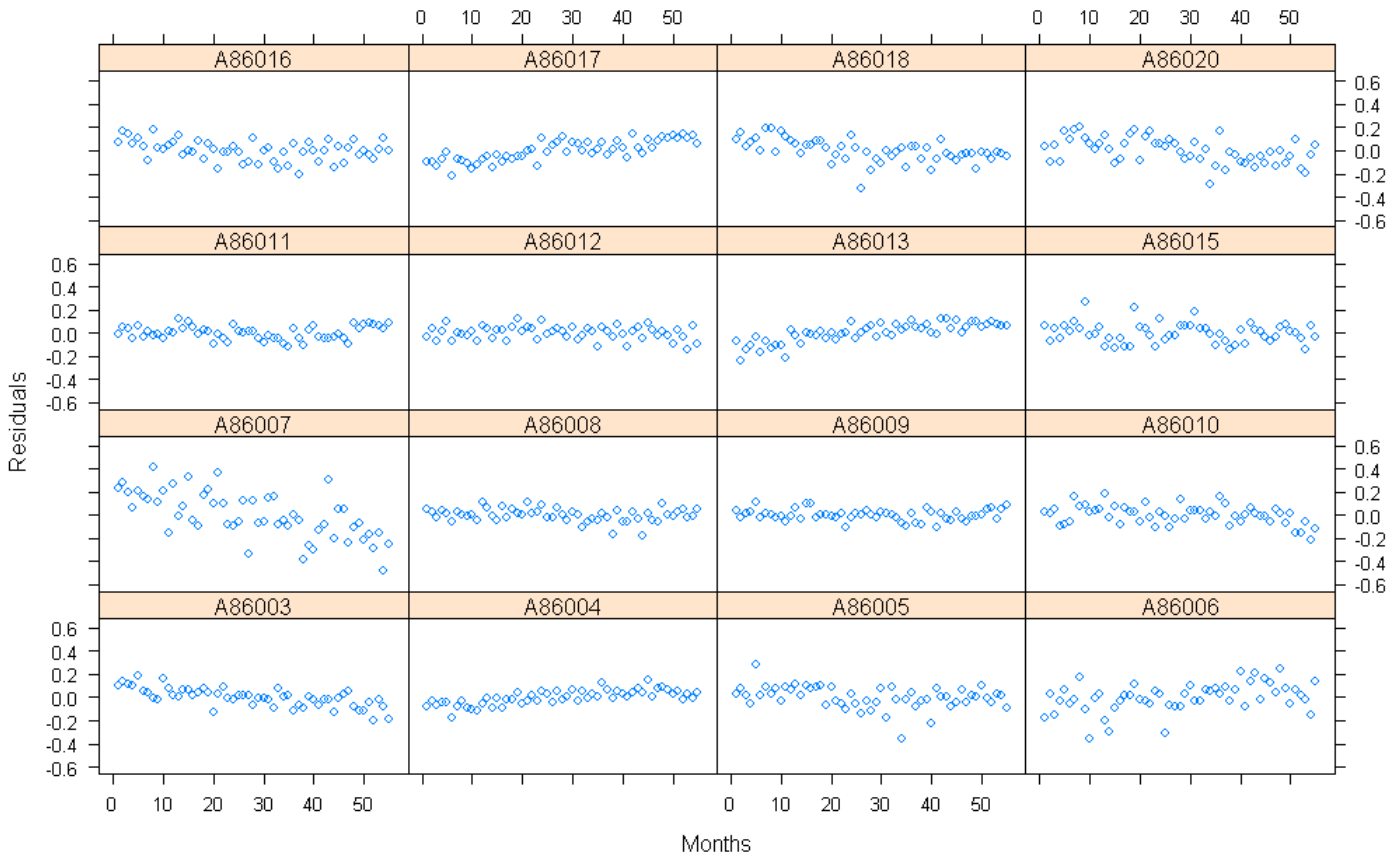
Observed against fitted values of the final model by GP practice (c)



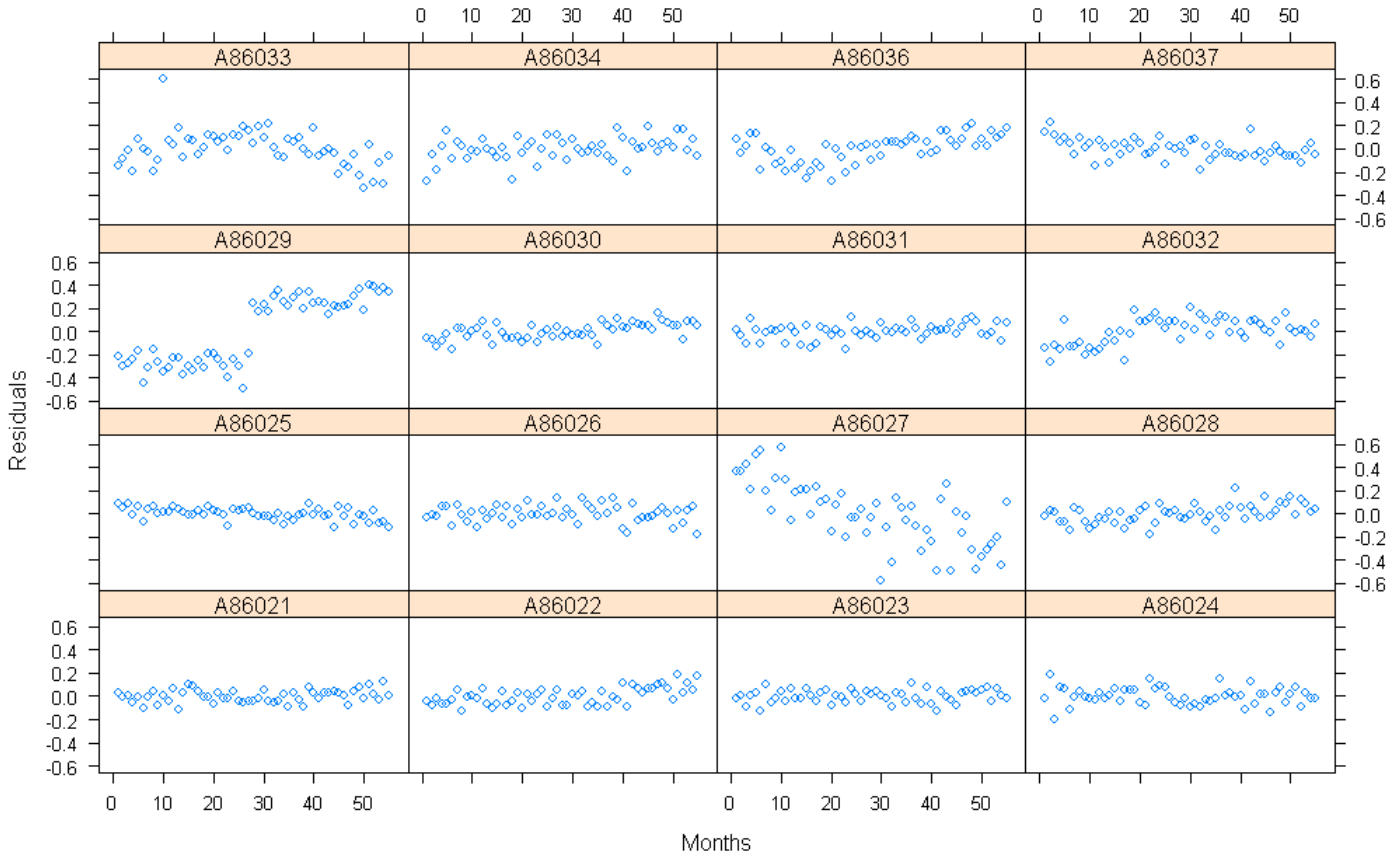
Observed against fitted values of the final model by GP practice (d)



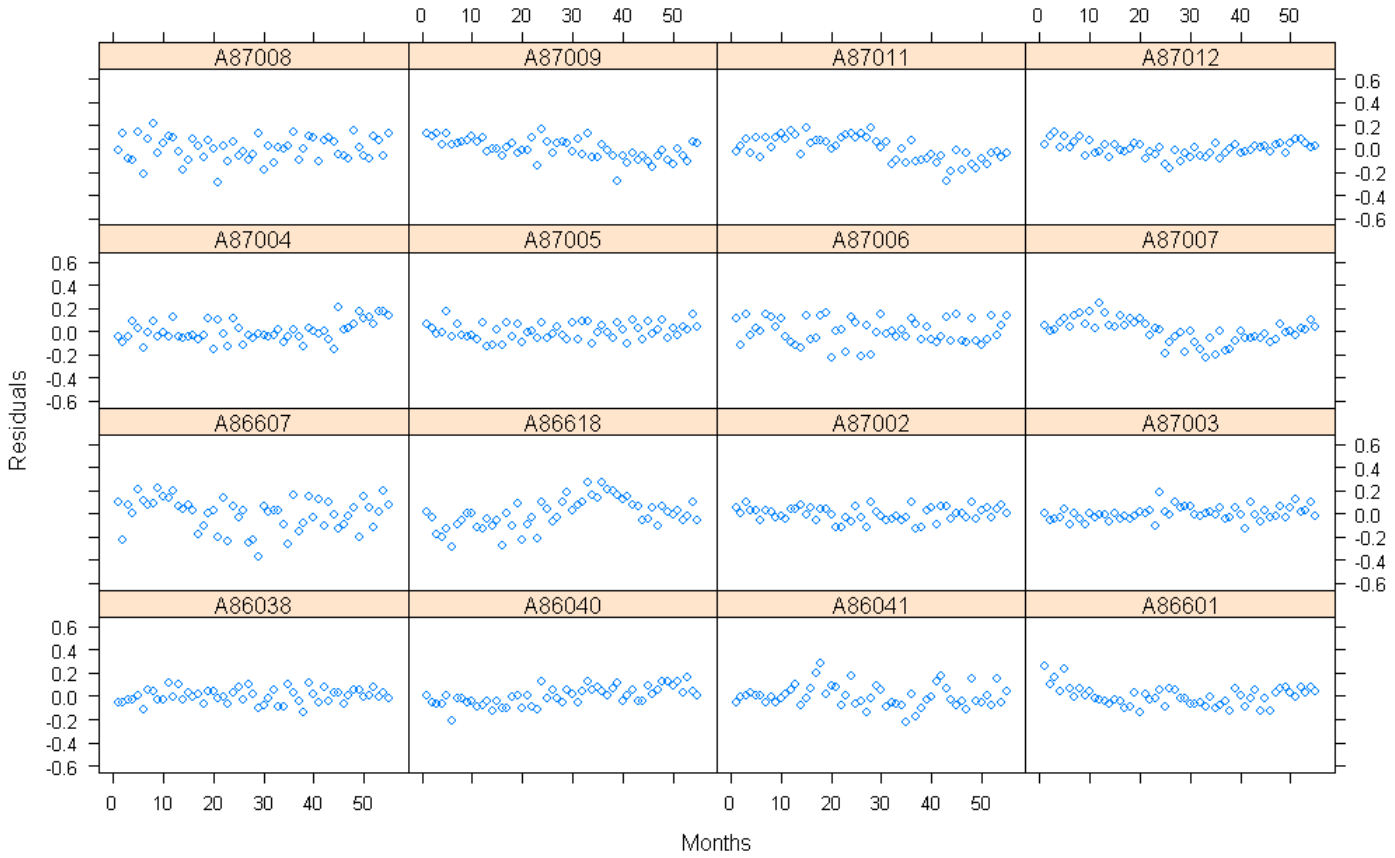
Residuals against time of the final model by GP practice (a)



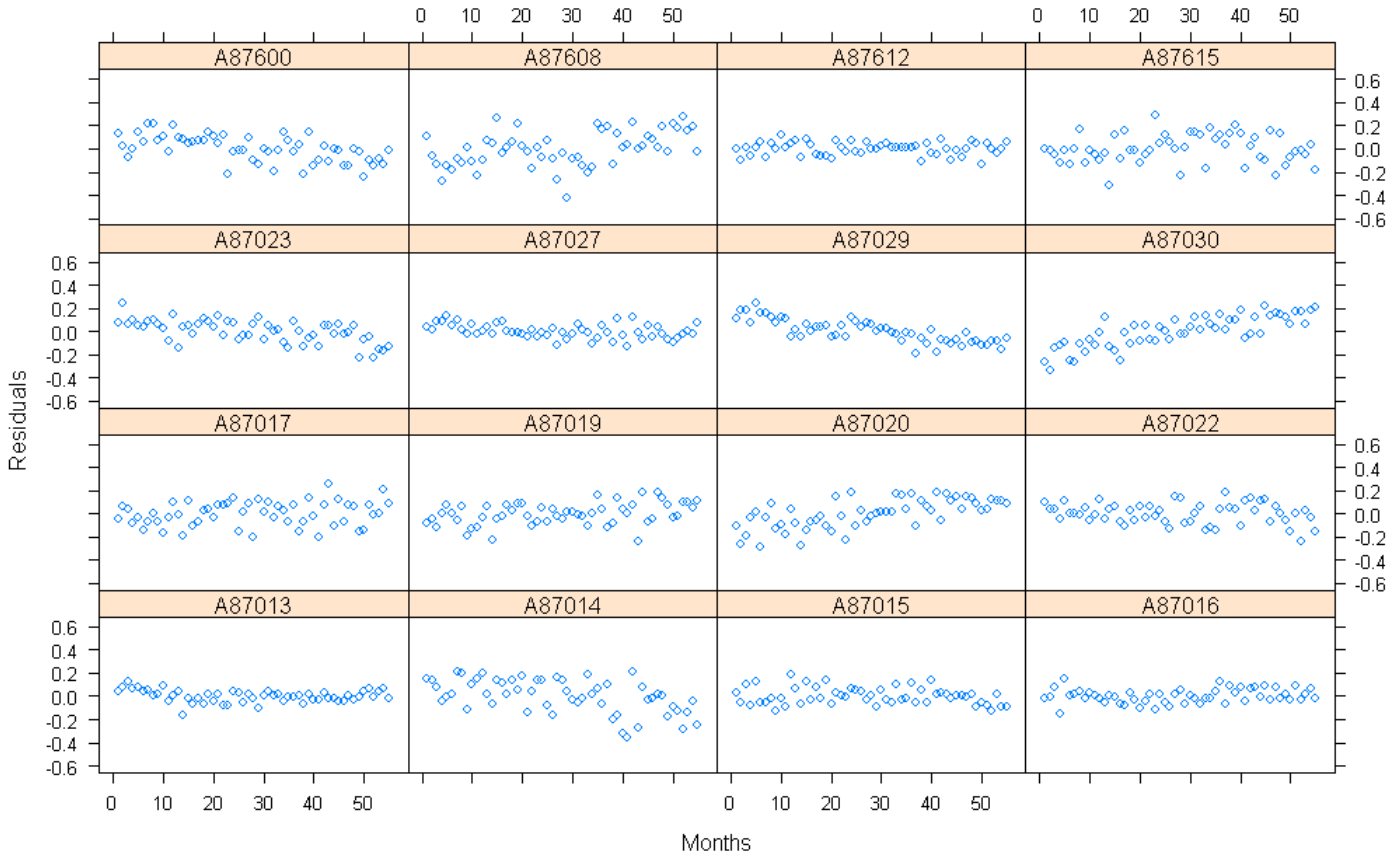
Residuals against time of the final model by GP practice (b)



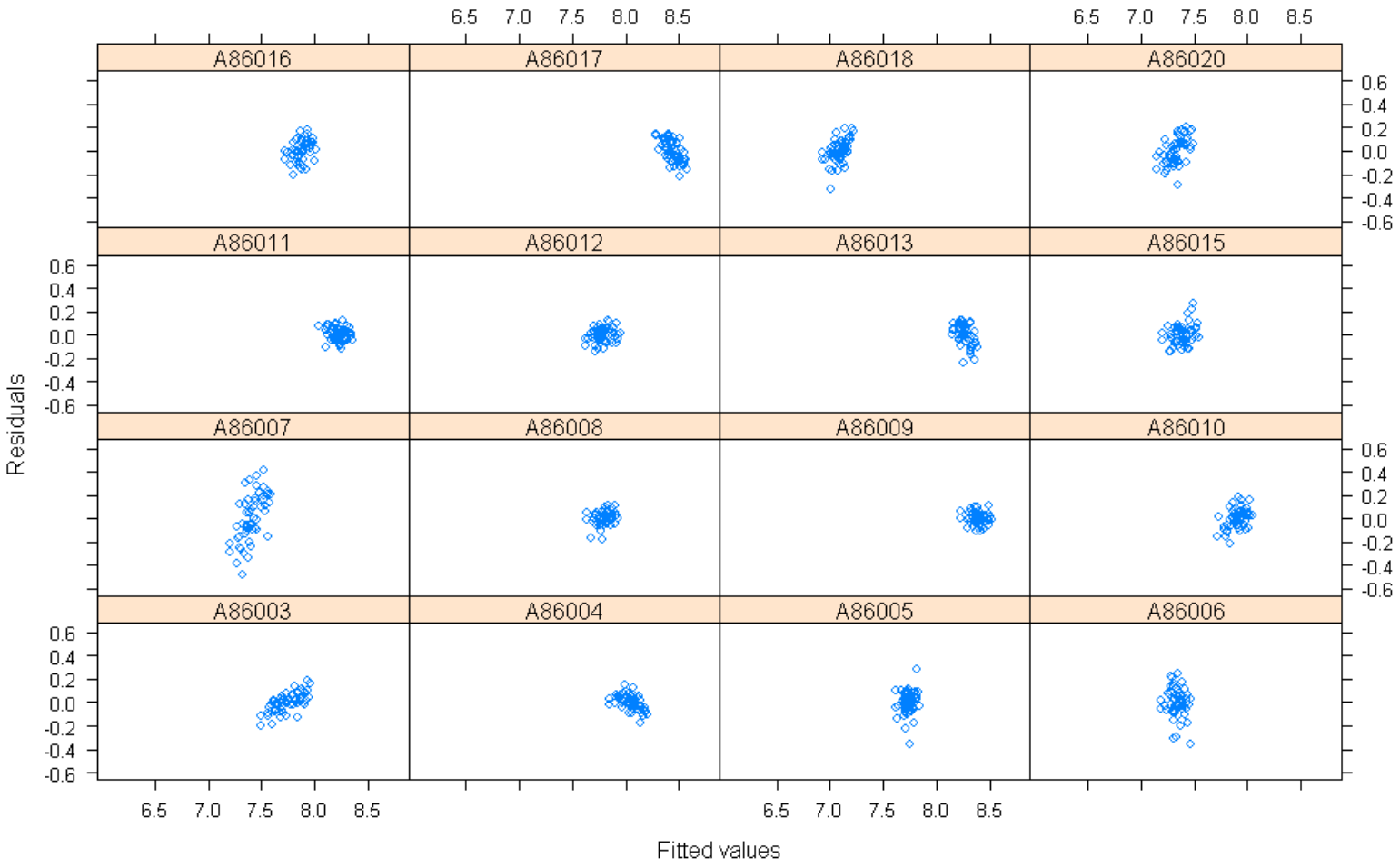
Residuals against time of the final model by GP practice (c)



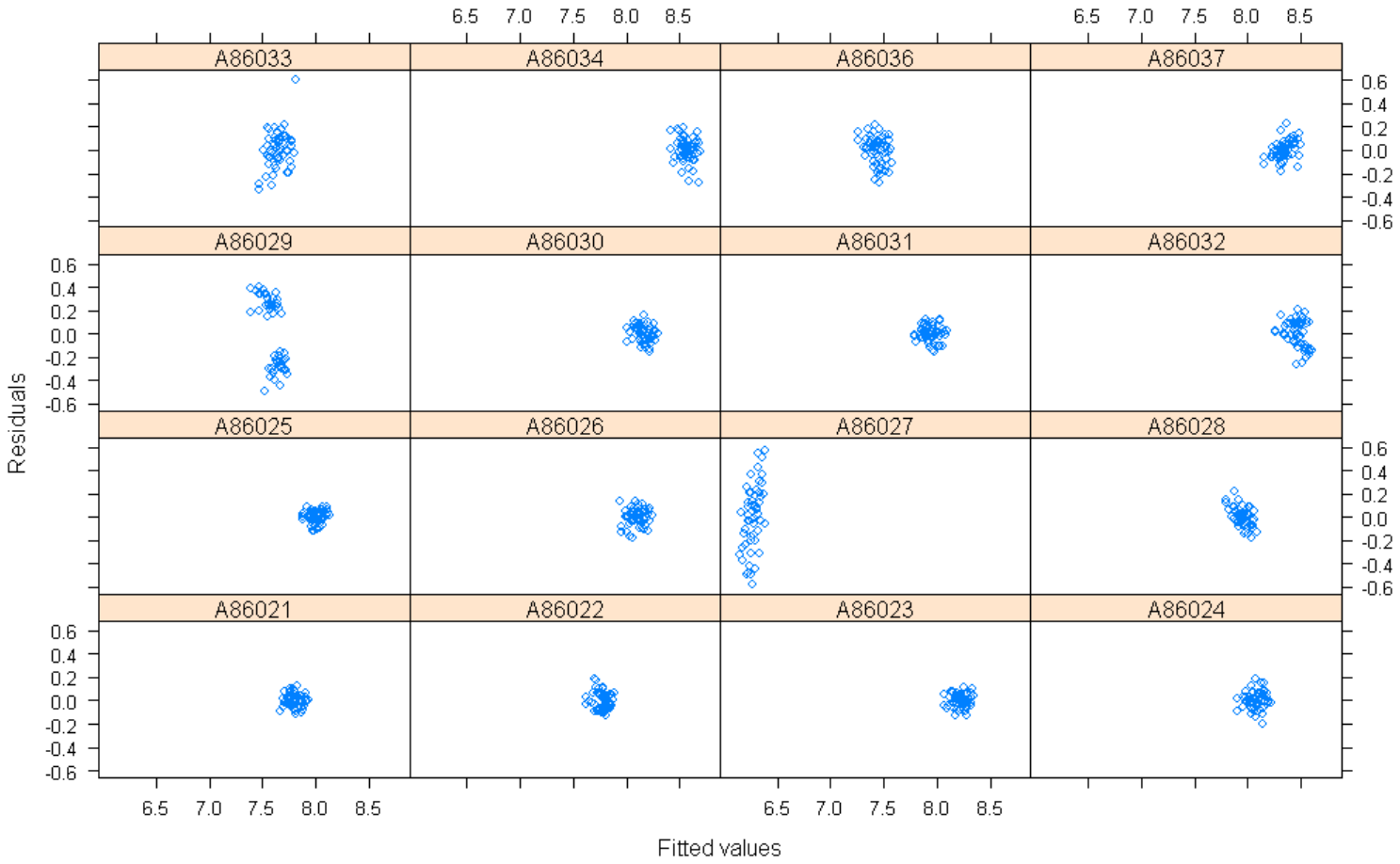
Residuals against time of the final model by GP practice (d)



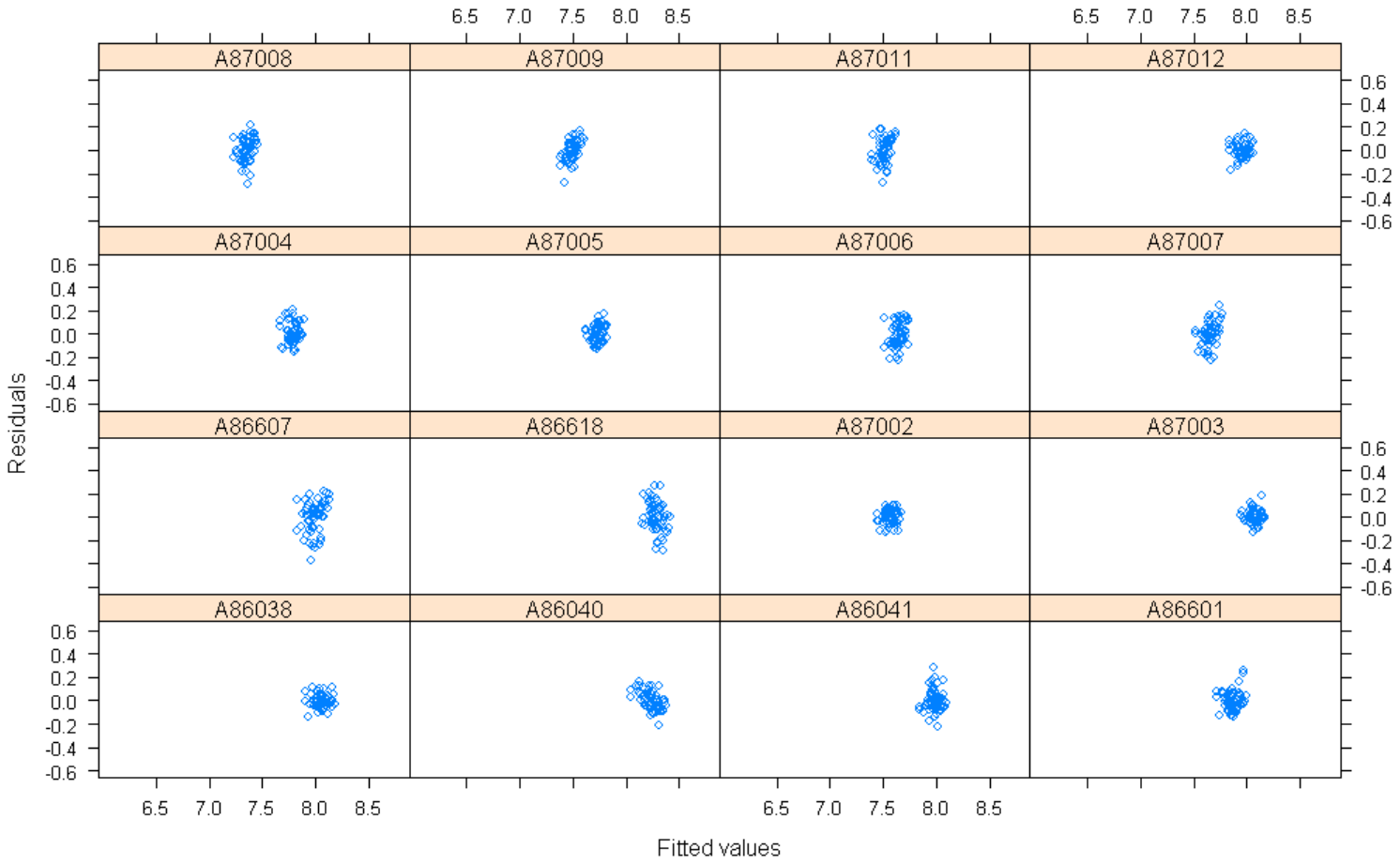
Residuals against fitted values of the final model (a)



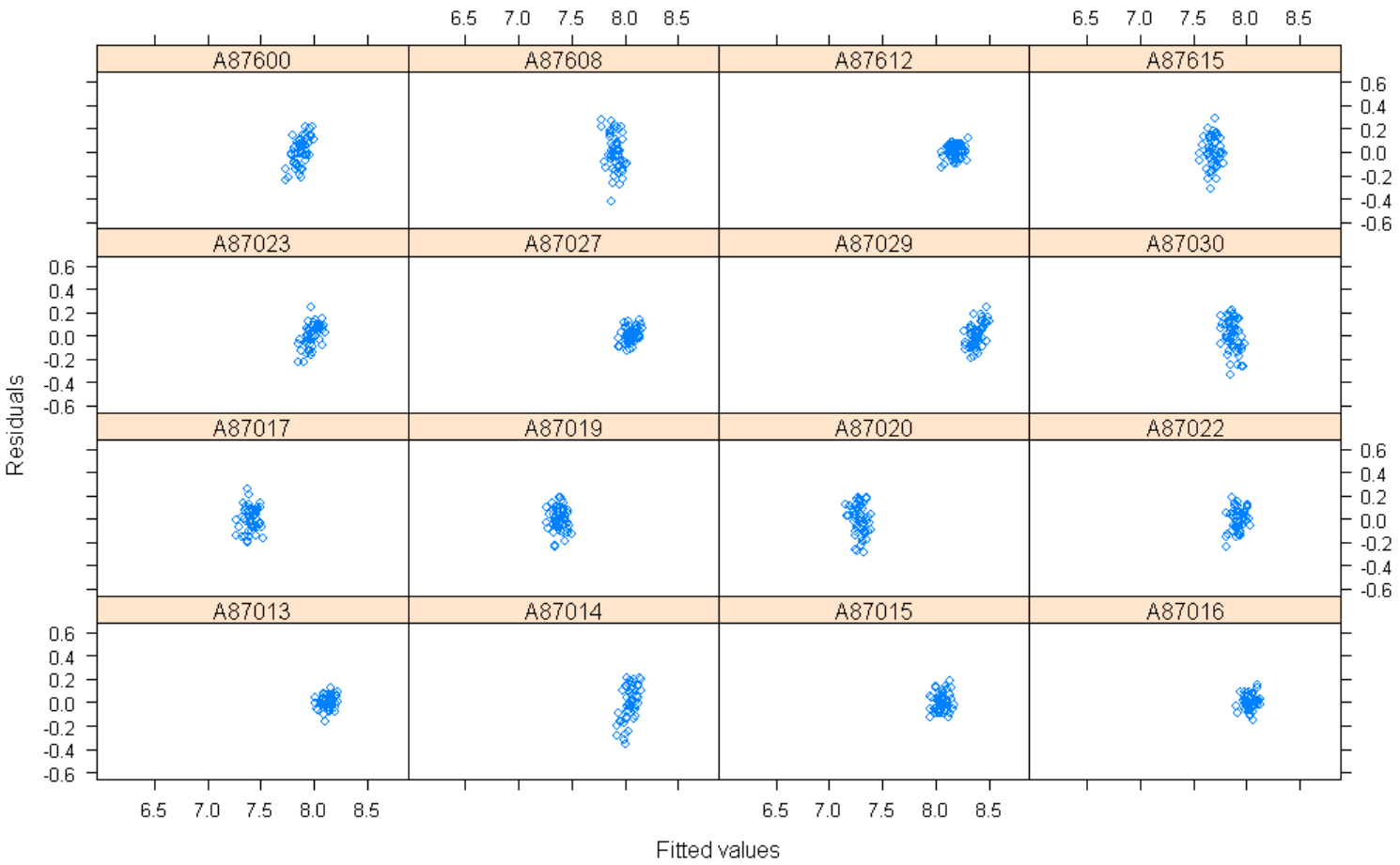
Residuals against fitted values of the final model (b)



Residuals against fitted values of the final model (c)



Residuals against fitted values of the final model (d)



6.9 Appendix I

R squared for mixed models

The coefficient of determination (R-squared) provides a measure of how well future outcomes are likely to be predicted by models. R-squared is appropriate for models that use the ordinary least squares (OLS) approach to calculate the minimum variance. The estimates of a mixed-effects model are maximum likelihood estimates though, so the OLS approach as a measure of goodness-to-fit is not appropriate. To evaluate the goodness-of-fit of mixed-effects model other approaches have been developed, called pseudo R-squares, the usefulness of which is debatable (Orelien and Edwards, 2008).

I have used the following equation to estimate pseudo R-squared that is appropriate for mixed-effects model with random intercept like the final statistical model (Snijders and Bosker, 1999).

$$R^2 = 1 - \frac{\sigma^2_{total\_Full}}{\sigma^2_{total\_Intercept}}$$

where $\sigma^2_{total\_Full}$ is the total variance explained by a mixed-effects model and $\sigma^2_{total\_Intercept}$ is the total variance explained by an the same mixed-effects model when fitted only with the random effects. As a result, this pseudo R-squared measure can only capture the predictive power of fixed effects on a model and not the random effects. I fitted a model with only GP practice random effects and compared that to the final model (L21PM_30TR). The pseudo R-squared value for the final model was 48.3%.

6.10 Appendix J

Conference Abstracts

Sofianopoulou, E., Pless-Mulloli, T. and Rushton, S. (2009) 'Estimating Traffic Exposure in Primary Care Service Areas', *Epidemiology,* 20, (6), pp. S203-S203.

Sofianopoulou, E., Rushton, S. and Pless-Mulloli, T. (2009) 'Analysis of Spatio-Temporal Patterns of Short-Acting beta(2) Prescribing', *Epidemiology,* 20, (6), pp. S204-S204.

Sofianopoulou, E., Pless-Mulloli, T. and Rushton, S. (2009) 'Estimating traffic conditions within GP practice service areas', GISRUK, 1 - 3 April 2009.

Sofianopoulou, E., Rushton, S. and Pless-Mulloli, T. (2010) 'Spatio-temporal Analysis in Environmental Health: Respiratory Medication in Relation to Air Pollution and Deprivation', INSPIRE conference 2010, 22 -25 June 2010

Other presentations

I presented the final results of my work "Spatio-temporal analysis – respiratory prescribing in relation to air pollution and deprivation" at:

- Imperial College, Department of Occupational and Environmental Medicine, London, 27th May 2010
- Durham University, Wolfson Research Institute, Durham, 25th January 2010
- NHS Regional Drug & Therapeutics Centre, Newcastle, 14th July 2009
- Colt Foundation Day, King's College London, London, December 2009

I also had the chance to present interim results at:

- Colt Foundation Day, King's College London, London, December 2008
- Lancaster University, School of Health and Medicine, Lancaster, 1st October 2008
- North Tyneside Primary Care Trust, Meeting of Public Health Directors, 05th February 2008
- Colt Foundation Day, King's College London, London, December 2007

References

ALLEGRA, L., CORDARO, C. I. & GRASSI, C. 1996. Prevention of acute exacerbations of chronic obstructive bronchitis with carbocysteine lysine salt monohydrate: A multicenter, double-blind, placebo-controlled trial. *Respiration,* 63**,** 174-180.

ANDERSON, H. R., RUGGLES, R., STRACHAN, D. P., AUSTIN, J. B., BURR, M., JEFFS, D., STANDRING, P., STERIU, A. & GOULDING, R. 2004. Trends in prevalence of symptoms of asthma, hay fever, and eczema in 12-14 year olds in the British Isles, 1995-2002: Questionnaire survey. *British Medical Journal,* 328**,** 1052-1053.

ANDERSON, H. R., SPIX, C., MEDINA, S., SCHOUTEN, J. P., CASTELLSAGUE, J., ROSSI, G., ZMIROU, D., TOULOUMI, G., WOJTYNIAK, B., PONKA, A., BACHAROVA, L., SCHWARTZ, J. & KATSOUYANNI, K. 1997. Air pollution and daily admissions for chronic obstructive pulmonary disease in 6 European cities: Results from the APHEA project. *European Respiratory Journal,* 10**,** 1064-1071.

APSIMON, H. M., GONZALEZ DEL CAMPO M.T. & H.S., A. 2000. Modelling long-range transport of primary particulate material over Europe *Atmospheric Environment,* 35.

ARBEX M A , DE SOUZA CONCEIÇÃO G M , S P CENDON, F F ARBEX, A C LOPES, E P MOYSÉS, S L SANTIAGO, P H N SALDIVA, L A A PEREIRA & BRAGA, A. L. F. 2009. Urban air pollution and chronic obstructive pulmonary disease-related emergency department visits. *Journal of Epidemiology and Community Health,* 63**,** 777-783.

ARTIS M., CLAVEL G.J., HOFFMANN M. & NACHANE D. 2007. Harmonic Regression Models: A Comparative Review with Applications, Working Paper No. 333. Institute for Empirical Research in Economics, University of Zurich.

ASHER MI, MONTEFORT S, BJÖRKSTÉN B, LAI CKW, STRACHAN DP, WEILAND SK, H, W. & GROUP, T. I. P. T. S. 2006. Worldwide time trends in the prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and eczema in childhood: ISAAC Phases One and Three repeat multicountry cross-sectional surveys. *The Lancet,* 368**,** 733-743.

ASHER, M. I., MONTEFORT, S., BJORKSTEN, B., LAI, C. K. W., STRACHAN, D. P., WEILAND, S. K. & WILLIAMS, H. 2006. Worldwide time trends in the prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and eczema in childhood: ISAAC Phases One and Three repeat multicountry cross-sectional surveys. *Lancet,* 368**,** 733-743.

ASTHMA UK 2004. Where do we stand? Asthma in UK today. Asthma UK.

ASTHMA UK. 2010. *Key facts and statistics* [Online]. Asthma UK. Available: http://www.asthma.org.uk/news_media/media_resources/for_journalists_key.html 2010].

ATKINSON, R. W., ANDERSON, H. R. & JORDI SUNYER, J. A., MICHELA BACCINI, JUDITH M. VONK, AZZEDINE BOUMGHAR, FRANCESCO FORASTIERE, BERTIL FORSBERG, GIOTA TOULOUMI, JOEL SCHWARTZ, AND KLEA KATSOUYANNI 2001a. Acute Effects of Particulate Air Pollution on Respiratory Admissions *American Journal of Respiratory and Critical Care Medicine,,* 164**,** 1860-1866.

ATKINSON, R. W., ANDERSON, H. R., SUNYER, J., AYRES, J., BACCINI, M., VONK, J. M., BOUMGHAR, A., FORASTIERE, F., FORSBERG, B., TOULOUMI, G., SCHWARTZ, J. & KATSOUYANNI, K. 2001b. Acute effects of particulate air pollution on respiratory admissions: Results from APHEA 2 project. *American Journal of Respiratory and Critical Care Medicine,* 164**,** 1860-1866.

ATKINSON, R. W., STRACHAN, D. P. & H R ANDERSON, S. H., AND J EMBERLIN 2006. Temporal associations between daily counts of fungal spores and asthma exacerbations. *Occup. Environ. Med,* 63**,** 580-590.

BAAYEN, R. H., DAVIDSON, D. J. & BATES, D. M. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language,* 59**,** 390-412.

BAILEY, T. C. & GATRELL, A. C. 1995. *Interactive spatial data analysis,* Harlow Essex, England: Longman Scientific & Technical, 1995.

BEYER, H. L. 2004. *Hawth's Analysis Tools for ArcGIS* [Online]. Available: http://www.spatialecology.com/htools

BLEECKER, E. R. 2004. Similarities and differences in asthma and COPD: The Dutch hypothesis. *Chest,* 126.

BOLLAG, U., CAPKUN, G., CAESAR, J. & LOW, N. 2005. Trends in primary care consultations for asthma in Switzerland, 1989-2002. *International Journal of Epidemiology,* 34**,** 1012-1018.

BOLLAG, U., GRIZE, L. & BRAUN-FAHRLANDER, C. 2009. Is the ebb of asthma due to the decline of allergic asthma? A prospective consultation-based study by the Swiss Sentinel Surveillance Network, 1999-2005. *Family Practice,* 26**,** 96-101.

BRAUN-FAHRLANDER, C., GASSNER, M., GRIZE, L., TAKKEN-SAHLI, K., NEU, U., STRICKER, T., VARONIER, H. S., WÃ¼THRICH, B. & SENNHAUSER, F. H. 2004. No further increase in asthma, hay fever and atopic sensitisation in adolescents living in Switzerland. *European Respiratory Journal,* 23**,** 407-413.

BRENNER, H., GREENLAND, S. & SAVITZ, D. A. 1992. The Effects of Nondifferential Confounder Misclassification in Ecologic Studies. *Epidemiology,* 3**,** 456-459.

BRESLOW, N. E. & CLAYTON, D. G. 1993. Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association,* 88**,** 9-25.

BRITISH NATIONAL FORMULARY 2009. Bronchodilators. *BNF*, , 57.

BRITISH THORACIC SOCIETY 2006. Burden of lung disease. *In:* EDITION, N. (ed.). London: British Thoracic Society.

BRONARS, S. & JANSEN, D. 1987. The geographic distribution of unemployment rates in the US. *Journal of Econometrics*, 36**,** 251-79.

BROWN, H. & PRESCOTT, R. 2006. *Applied mixed models in medicine*, Chichester, England ; Hoboken, NJ, John Wiley.

BUJOSA, M., GARCIA-FERRER, A. & YOUNG, P. C. 2007. Linear dynamic harmonic regression. *Computational Statistics & Data Analysis*, 52**,** 999-1024.

BURR, M. L. & HOLLIDAY, R. M. 1987. Why is chest disease so common in South Wales? Smoking, social class, and lung function: a survey of elderly men in two areas. *Journal of Epidemiology and Community Health*, 41**,** 140-144.

BURTON, P., GURRIN, L. & SLY, P. 1998. Extending the simple linear regression model to account for correlated responses: An introduction to generalized estimating equations and multi-level mixed modelling. *Statistics in Medicine*, 17**,** 1261-1291.

CABINET OFFICE. 2009. *UK Goverment Data Standards Catalogue* [Online]. Available: http://www.govtalk.gov.uk/gdsc/html/frames/Postcode.htm [Accessed 30 April 2009 2009].

CAGNEY, K. A. & BROWNING, C. R. 2004. Exploring Neighborhood-level Variation in Asthma and other Respiratory Diseases. *Journal of General Internal Medicine*, 19**,** 229.

CALVERLEY, P. M. A. & SONDHI, S. 1998. The burden of obstructive lung disease in the UK - COPD and asthma. *Thorax*, 53.

CARLIN, B. P. & LOUIS, T. A. 1996. *Bayes and empirical Bayes methods for data analysis*, London, Chapman & Hall.

CARLOS, H. A., SHI, X., SARGENT, J., TANSKI, S. & BERKE, E. M. 2010. Density estimation and adaptive bandwidths: a primer for public health practitioners. *Int J Health Geogr*, 9**,** 39.

CAZZOLA, M., BETTONCELLI, G., SESSA, E. & CRICELLI, C. 2009. Primary care of the patient with chronic obstructive pulmonary disease in Italy. *Respiratory Medicine*, 103**,** 582-588.

CELLI, B. R., MACNEE, W., AGUSTI, A., ANZUETO, A., BERG, B., BUIST, A. S., CALVERLEY, P. M. A., CHAVANNES, N., DILLARD, T., FAHY, B., FEIN, A., HEFFNER, J., LAREAU, S., MEEK, P., MARTINEZ, F., MCNICHOLAS, W., MURIS, J., AUSTEGARD, E., PAUWELS, R., RENNARD, S., ROSSI, A., SIAFAKAS, N., TIEP, B., VESTBO, J., WOUTERS, E. & ZUWALLACK, R. 2004. Standards for the diagnosis and treatment of patients with COPD: A summary of the ATS/ERS position paper. *European Respiratory Journal*, 23**,** 932-946.

CENTRE FOR DRUG EVALUATION AND RESEARCH 2001. Statistical approaches to establishing bioequivalence. U.S. Department of Health and Human Services

CERVERI, I., ACCORDINI, S., CORSICO, A., ZOIA, M. C., CARROZZI, L., CAZZOLETTI, L., BECCARIA, M., MARINONI, A., VIEGI, G., DE MARCO, R., GINESU, F., PIRINA, P., OSTERA, S., PINNA, G. P., FARRE, A., IMPARATO, S., DALLARI, R., TURRINI, E., FOGLIA, M., GIAMMANCO, G., PIGNATO, S., ROTONDO, A., CUSPILICI, A., BUGIANI, M., PICCIONI, P., CAROSSO, A., AROSSA, W., CARIA, E., CASTIGLIONI, G., MIGLIORE, E., ROMANO, C., FABBRO, D., CICCONE, G., MAGNANI, C., DALMASSO, P., BONO, R., GIGLI, G., GIRAUDO, A., BRUSSINO, M. C., BUCCA, C., ROLLA, G., STRUZZO, P., OREFICE, U., SCHNEIDER, M., CHITTARO, F., PERESSON, D., DE MARCO, R., VERLATO, G., ACCORDINI, S., ZANOLIN, M. E., LOCATELLI, F., CAZZOLETTI, L., BATTISTI, L., PATTARO, C., POLI, A., DORIGO, N., CANTARELLI, S., CIRESOLA, D., LO CASCIO, V., OLIVIERI, M., FERRARI, M., BIASIN, C., LAURIOLA, P., DANIELLI, G., SESTI, D., GHIGLI, E., NATALE, P., GROSA, M., TACCONI, A., FRONTERO, P. & SALOMONI, A. 2003. Chronic cough and phlegm in young adults. *European Respiratory Journal,* 22**,** 413-417.

CHATFIELD, C. 2004. *The analysis of time series : an introduction,* London, Chapman & Hall/CRC.

CHINN S, JARVIS D, BURNEY P, LUCZYNSKA C & ACKERMANN-LIEBRICH U, A. J., CERVERI I, DE MARCO R, GISLASON T, HEINRICH J, JANSON C, KÜNZLI N, LEYNAERT B, NEUKIRCH F, SCHOUTEN J, SUNYER J, SVANES C, VERMEIRE P, WJST M. 2004. Increase in diagnosed asthma but not in symptoms in the European Community Respiratory Health Survey. *Thorax,* 59**,** 646-51.

CHIU, S. 1989. Detecting Period Components in a White Gaussian Time Series. *Journal of the Royal Statistical Society,* Ser. B, 52**,** 249-259.

CLARK, L. C., COMBS, G. F., TURNBULL, B. W., SLATE, E. H., CHALKER, D. K., CHOW, J., DAVIS, L. S., GLOVER, R. A., GRAHAM, G. F., GROSS, E. G., KRONGRAD, A., LESHER, J. L., PARK, H. K., SANDERS, B. B., SMITH, C. L. & TAYLOR, J. R. 1996. Effects of selenium supplementation for cancer prevention in patients with carcinoma of the skin a randomized controlled trial - A randomized controlled trial. *Jama-Journal of the American Medical Association,* 276**,** 1957-1963.

CLAYTON, R. R., CATTARELLO, A. M. & JOHNSTONE, B. M. 1996. The effectiveness of Drug Abuse Resistance Education (Project DARE): 5-year follow-up results. *Preventive Medicine,* 25**,** 307-318.

CLIFF, A. D. & ORD, J. K. 1973. *Spatial autocorrelation,* London,, Pion.

CLIFF, A. D. & ORD, J. K. 1981. *Spatial processes: models & applications,* London, Pion.

COGDON, P. 1994. Spatiotemporal analysis of area mortality. *Statistician,* 43**,** 513-528.

COHEN, B. H., BALL, W. C., BRASHEARS, S., DIAMOND, E. L., KREISS, P., LEVY, D. A., MENKES, H. A., PERMUTT, S. & TOCKMAN, M. S. 1977. Risk-Factors in Chronic Obstructive Pulmonary-Disease (Copd). *American Journal of Epidemiology,* 105**,** 223-231.

COMMISSION OF THE EUROPEAN COMMUNITIES 2004. Proposal of the European Parliament and of the Council - establishing an infrastructure for Spatial Information in the Community (INSPIRE).

COMMUNITIES AND NEIGHBOURHOODS 2004. The English Indices of Deprivation 2004. London: Office of the Deputy Prime Minister.

COMMUNITIES AND NEIGHBOURHOODS 2008. The English Indices of Deprivation 2007. London: Department for Communities and Local Government.

CONGDON, P. & BEST, N. 2000. Small area variation in hospital admission rates: Bayesian adjustment for primary care and hospital factors. *Journal of the Royal Statistical Society. Series C: Applied Statistics,* 49**,** 207-226.

CORINAIR - CORE INVENTORY OF AIR EMISSIONS METHODOLOGY 1996. EMEP/CORINAIR Atmospheric Emission Inventory Guidebook. *In:* EUROPEAN ENVIRONMENT AGENCY (ed.). Copenhagen.

COULTAS, D. B. & MAPEL, D. W. 2003. Undiagnosed airflow obstruction: Prevalence and implications. *Current Opinion in Pulmonary Medicine,* 9**,** 96-103.

CRAMPTON, J. 2005. *Mapping: a Critical Introduction to Gis and Cartography,* Oxford, Blackwell Publishing Ltd.

CRESSIE, N., FREY, J., HARCH, B. & SMITH, M. 2006. Spatial prediction on a river network *Journal of Agricultural, Biological, and Environmental Statistics* 11.

CRESSIE, N. A. C. 1993. *Statistics for Spatial Data,* John Wiley and Sons, Inc.

CULLINAN P. & A., N. T. 2003. Asthma: environmental and occupational factors. *British Medical Bulletin,* 68**,** 227-242.

DE MARCO, R., ACCORDINI, S., CERVERI, I., CORSICO, A., SUNYER, J., NEUKIRCH, F., KÃ¼NZLI, N., LEYNAERT, B., JANSON, C., GISLASON, T., VERMEIRE, P., SVANES, C., ANTO, J. M. & BURNEY, P. 2004. An international survey of chronic obstructive pulmonary disease in young adults according to GOLD stages. *Thorax,* 59**,** 120-125.

DECRAMER, M., RUTTEN-VAN MALKEN, M., DEKHUIJZEN, P. N. R., TROOSTERS, T., VAN HERWAARDEN, C., PELLEGRINO, R., VAN SCHAYCK, C. P. O., OLIVIERI, D., DEL DONNO, M., DE BACKER, W., LANKHORST, I. & ARDIA, A. 2005. Effects of N-acetylcysteine on outcomes in chronic obstructive pulmonary disease (Bronchitis Randomized on NAC Cost-Utility Study, BRONCUS): A randomised placebo-controlled trial. *Lancet,* 365**,** 1552-1560.

DEFRA. 2010. *UK Location* [Online]. Available: http://location.defra.gov.uk/inspire/ [Accessed 15th March 2010 2010].

DEMIDENKO, E. 2004. *Mixed models : theory and applications,* Hoboken, N.J., Wiley-Interscience.

DEMISSIE, K., ERNST, P., HANLEY, J. A., LOCHER, U., MENZIES, D. & BECKLAKE, M. R. 1996. Socioeconomic status and lung function among primary school children in Canada. *American Journal of Respiratory and Critical Care Medicine,* 153**,** 719-723.

DEPARTMENT OF HEALTH 1995. Understanding Asthma. *In:* NUMBER OF PRESCRIPTIONS FOR ASTHMA PREPARATIONS FOR ENGLAND BETWEEN 1980-1993 (ed.). Institute for Environment and Health.

DEPARTMENT OF HEALTH 2004. Prescriptions used in the prevention and treatment of respiratory disease. England 2004.

DEPARTMENT OF HEALTH 2005. Health reform in England: update and next steps. *In:* DOH (ed.). London.

DEPARTMENT OF HEALTH 2006. Chief executive's report to the NHS. London: Department of Health.

DEPARTMENT OF HEALTH 2010. White paper: Equity and excellence - Liberating the NHS.

DEPARTMENT OF THE ENVIRONMENT TRANSPORT AND THE REGIONS 2000. Indices of Deprivation 2000. Regeneration Research Summary.

DETHLEFSEN, C. & LUNDBYE-CHRISTENSEN, S. 2006. Formulating state space models in R with focus on longitudinal regression models. *Journal of Statistical Software,* 16.

DIEZ-ROUX, A. V. 1998. Bringing context back into epidemiology: Variables and fallacies in multilevel analysis. *American Journal of Public Health,* 88**,** 216-222.

DIEZ-ROUX, A. V., JAVIER NIETO, F., CARLES MUNTANER, HERMAN A. TYROLER, GEORGE W. COMSTOCK, EYAL SHAHAR, LAWTON S. COOPER, ROBERT L. WATSON & SZKLO, M. 1997. Neighborhood Environments and Coronary Heart Disease: A Multilevel Analysis *American Journal of Epidemiology* 149**,** 48-63.

DIGGLE, P., FAREWELL, D. & HENDERSON, R. 2007. Analysis of longitudinal data with drop-out: Objectives, assumptions and a proposal. *Journal of the Royal Statistical Society. Series C: Applied Statistics,* 56**,** 499-550.

DIGGLE, P. & KENWARD, M. 1994. Informative drop-out in longitudinal data analysis. *Journal of the Royal Statistical Society. Series C,* 1**,** 49-93.

DIGGLE, P. J. 1983. *Statistical Analysis of Spatial Point Patterns,* London, Academic Press.

DIGGLE, P. J. & RIBEIRO JR, P. J. 2007. *Model-based Geostatistics*.

DONALDSON, K., GILMOUR, M. I. & MACNEE, W. 2000. Asthma and PM10. *Respir Res,* 1**,** 12-5.

DRAKE, R. E., MCHUGO, G. J., BEBOUT, R. R., BECKER, D. R., HARRIS, M., BOND, G. R. & QUIMBY, E. 1999. A randomized clinical trial of supported employment for inner-city patients with severe mental disorders. *Archives of General Psychiatry,* 56**,** 627-633.

DUNN, C. E., WOODHOUSE, J., BHOPAL, R. S. & ACQUILLA, S. D. 1995. Asthma and Factory Emissions in Northern England - Addressing Public Concern by Combining Geographical and Epidemiologic Methods. *Journal of Epidemiology and Community Health,* 49**,** 395-400.

EBELT, S., BRAUER, M. & WILSON, W. E. 2004. A comparison of health effects from exposure to ambient and non-ambient particles. *Epidemiology,* 15**,** S48.

ELLIOTT, P., HILLS, M., BERESFORD J, KLEINSCHMIDT I, PATTENDEN S, JOLLEY D, RODRIGUES L, WESTLAKE AJ & G., R. 1992. Incidence of cancer of the larynx and lung near incinerators of waste solvents and oils in Great Britain. *Lancet,* 339**,** 854-858.

ELLIOTT, P., WAKEFIELD, J., BEST, N. & BRIGGS, D. 2000. *Spatial Epidemiology: Methods and Applications* Oxford University Press.

ELLIOTT, P. & WARTENBERG, D. 2004. Spatial Epidemiology: Current Approaches and Future Challenges. *Environmental Health Perspectives,* 112**,** 998-1006.

ELMHAGEN, B. & RUSHTON, S. P. 2007. Trophic control of mesopredators in terrestrial ecosystems: Top-down or bottom-up? *Ecology Letters,* 10**,** 197-206.

EUROPEAN COMMISSION 2004. "Environment & Health indicators" Options for actions and recommendations. *the framework of European Environment & Health Strategy.* Technical Working Group on Indicators and Priority Diseases.

EUROPEAN FEDERATION OF ALLERGY AND AIRWAYS DISEASES PATIENTS ASSOCIATIONS. 2003. *Asthma* [Online]. Available: http://www.efanet.org/asthma/index.html [Accessed Jan 2007.

EUROPEAN FEDERATION OF ALLERGY AND AIRWAYS DISEASES PATIENTS ASSOCIATIONS. 2009a. *Chronic Obstructive Pulmonary Disease (COPD)* [Online]. Available: http://www.efanet.org/copd/index.html [Accessed Jun 2009.

EUROPEAN FEDERATION OF ALLERGY AND AIRWAYS DISEASES PATIENTS ASSOCIATIONS. 2009b. *WHO and EFA Call for Global Action on Chronic Respiratory Disease* [Online]. Available: http://www.efanet.org/enews/press.html.

FANSHAWE, T. R., DIGGLE, P. J., RUSHTON, S., SANDERSON, R., LURZ, P. W. W., GLINIANAIA, S. V., PEARCE, M. S., PARKER, L., CHARLTON, M. & PLESS-MULLOLI, T. 2008. Modelling spatio-temporal variation in exposure to particulate matter: a two-stage approach. *Environmetrics,* 19**,** 549-566.

FARAWAY, J. J. 2006. *Extending the linear model with R : generalized linear, mixed effects and nonparametric regression models,* Boca Raton, Chapman & Hall/CRC.

FLEMING, D. M., SUNDERLAND, R., CROSS, K. W. & ROSS, A. M. 2000. Declining incidence of episodes of asthma: a study of trends in new episodes presenting to general practitioners in the period 1989-98. *Thorax,* 55**,** 657-661.

FLORAX, R. J. G. M. & VAN DER VLIST, A. J. 2003. Spatial Econometrics Data Analysis: Moving Beyond Traditional Models. *International Regional Science Review,* 26**,** 223-243.

FOTHERINGHAM, S. & ROGERSON, P. (eds.) 1994. *Spatial analysis and GIS,* London Washington, DC : Taylor & Francis.

FUKUCHI, Y., NISHIMURA, M., ICHINOSE, M., ADACHI, M., NAGAI, A., KURIYAMA, T., TAKAHASHI, K., NISHIMURA, K., ISHIOKA, S., AIZAWA, H. & ZAHER, C. 2004. COPD in Japan: The Nippon COPD Epidemiology study. *Respirology,* 9**,** 458-465.

GARDNER, M. J., SNEE, M. P., HALL, A. J., POWELL, C. A., DOWNES, S. & TERRELL, J. D. 1990. Results of case-control study of leukaemia and lymphoma among young people near Sellafield nuclear plant in West Cumbria. *BMJ* 300**,** 423-9.

GATRELL, A. C., BAILEY, T. C., DIGGLE, P. J. & ROWLINGSON, B. S. 1996. Spatial point pattern analysis and its application in geographical epidemiology. *Transactions of the Institute of British Geographers,* 21**,** 256-274.

GEHRIG, R. & BUCHMANN, B. 2003. Characterising seasonal variations and spatial distribution of ambient PM10 and PM2.5 concentrations based on long-term Swiss monitoring data. *Atmospheric Environment,* 37**,** 2571-2580.

GELMAN, A. 1995. *Bayesian data analysis,* London, Chapman & Hall.

GELMAN, A., CARLIN L.B., STERN H.S. & D.B., R. 2009. *Bayesian Data Analysis*, Chapman & Hall/CRC

GLIKSON, M., RUTHERFORD, S., SIMPSON, R. W., MITCHELL, C. A. & YAGO, A. 1995. Microscopic and submicron components of atmospheric particulate matter during high asthma periods in Brisbane, Queensland, Australia. *Atmospheric Environment,* 29**,** 549-562(14).

GLOBAL INITIATIVE FOR CHRONIC OBSTRUCTIVE LUNG DISEASE (GOLD) 2008. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease.

GOLDSTEIN, H. 1995. *Multilevel Statistical Models* London, Edward Arnold.

GOLDSTEIN, H. 2003. *Multilevel statistical models,* London, Arnold.

GOVERNMENT OFFICE FOR THE NORTH EAST. 2007. *Indices of Deprivation 2004* [Online]. Available: http://www.gos.gov.uk/gone/news/imd_update/ [Accessed 03 Sep 2007.

GRIFFITHS, C., SIMON, B. N. & WOODRUFF, T. A Cross-Sectional Analysis of Asthma Medication Use and Air Pollution: A Preliminary Analysis. Valuing Environmental Health Risk Reductions to Children. Session 4: Air Pollution and Asthma in Children, 20-21 October 2003 Washington.

GRIFFITHS, C., SIMON, N. B. & WOODRUFF, T. 2009. Asthma Medication Use and Air Pollution In California: A Cross-Sectional Analysis. *Working Papers Series* Washington: U.S. Environmental Protection Agency and National Centre for Environmental Economics.

HAJAT, S., HAINES, A., GOUBET, S. A., ATKINSON, R. W. & ANDERSON, H. R. 1999. Association of air pollution with daily GP consultations for asthma and other lower respiratory conditions in London. *Thorax,* 54**,** 597-605.

HAJAT S., HAINES A., GOUBET S.A., ATKINSON R.W. & H.R., H. R. A. 1999. Association of air pollution with daily GP consultations for asthma and other lower respiratory conditions in London. *Thorax,* 54**,** 597-605.

HALBERT, R. J., NATOLI, J. L., GANO, A., BADAMGARAV, E., BUIST, A. S. & MANNINO, D. M. 2006. Global burden of COPD: systematic review and meta-analysis. *European Respiratory Journal,* 28**,** 523-532.

HANNAN, E. J. 1971. Non-linear Time Series Regression. *Journal of Applied Probability,* 8**,** 767-780.

HEYWOOD, I., CORNELIUS, S. & CARVER, S. 1998. *An introduction to geographical information systems,* New York :, Addison Wesley Longman.

HIGGINS, M. W., KELLER, J. B. & METZNER, H. L. 1977. Smoking, socioeconomic status, and chronic respiratory disease. *American Review of Respiratory Disease,* 116**,** 403-410.

HIPPOCRATES 1881. *Hippocrates on airs, waters, and places / the received Greek text of Littre, with Latin, French, and English translations by eminent scholars,* London, Printed--not for sale -- by Wyman & Sons.

INSIPRE FRAMEWORK DEFINITION SUPPORT WORKING GROUP 2003. Contribution to the extended impact assessment of INSPIRE. *In:* CONTRACTOR CRAGLIA MAX IN SHEFFIELD UNIVERSITY (ed.). Environment Agency for England and Wales.

ISAAC STEERING COMMITTEE 1998a. Worldwide variation in prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and atopic eczema: ISAAC. *In:* QUESTIONNAIRES, -. M. P. O. S.-R. A. S. F. W. (ed.).

ISAAC STEERING COMMITTEE 1998b. Worldwide variation in prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and atopic eczema: ISAAC. *Lancet,* 351**,** 1225-32.

J.C. PINHEIRO & BATES, D. M. 2000. *Mixed-Effects Models in S and S-PLUS,* Springer Verlag New York, LLC.

JANSON, C., ANTO, J. & P. BURNEY3, S. C., R. DE MARCO4, J. HEINRICH5, D. JARVIS3, N. KUENZLI6, B. LEYNAERT7, C. LUCZYNSKA3, F.

NEUKIRCH7, C. SVANES8, J. SUNYER2 AND M. WJST5 ON BEHALF OF THE EUROPEAN COMMUNITY RESPIRATORY HEALTH SURVEY II 2001. The European Community Respiratory Health Survey: what are the main results so far? . *European Respiratory Journal,* 18**,** 598-611.

JANSON, C., CHINN, S., D JARVIS & BURNEY, P. 1997. Physician-diagnosed asthma and drug utilization in the European Community Respiratory Health Survey *European Respiratory Journal,* 10**,** 1795-1802.

JERRETT, M. & FINKELSTEIN, M. 2005. Geographies of Risk in Studies Linking Chronic Air Pollution Exposure to Health Outcomes. *Journal of Toxicology and Environmental Health Part A,* 68**,** 1207-1242.

JOHNSON, D. L. & THOMPSON, R. 1995. Restricted Maximum-Likelihood-Estimation of Variance-Components for Univariate Animal-Models Using Sparse-Matrix Techniques and Average Information. *Journal of Dairy Science,* 78**,** 449-456.

KATSOUYANNI, K., SCHWARTZ, J., SPIX, C., TOULOUMI, G., ZMIROU, D., ZANOBETTI, A., WOJTYNIAK, B., VONK, J. M., TOBIAS, A., PONKA, A., MEDINA, S., BACHAROVA, L. & ANDERSON, H. R. 1996. Short term effects of air pollution on health: A European approach using epidemiologic time series data: The APHEA protocol. *Journal of Epidemiology and Community Health,* 50**,** S12-S18.

KATSOUYANNI, K., TOULOUMI, G., SAMOLI, E., GRYPARIS, A., LE TERTRE, A., MONOPOLIS, Y., ROSSI, G., ZMIROU, D., BALLESTER, F., BOUMGHAR, A., ANDERSON, H. R., WOJTYNIAK, B., PALDY, A., BRAUNSTEIN, R., PEKKANEN, J., SCHINDLER, C. & SCHWARTZ, J. 2001. Confounding and effect modification in the short-term effects of ambient particles on total mortality: Results from 29 European cities within the APHEA2 project. *Epidemiology,* 12**,** 521-531.

KATSOUYANNI, K., TOULOUMI, G., SPIX, C., SCHWARTE, J., BALDUCCI, F., MEDINA, S., ROSSI, G., WOJTYNIAK, B., SUNYER, J., BACHAROVA, L., SCHOUTEN, J. P., PONKA, A. & ANDERSON, H. R. 1997. 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. *British Medical Journal,* 314**,** 1658-1663.

KAY, S. M. & MARPLE, S. L. Spectrum Analysis - A Modern Perspective. Proceedings of the IEEE, 69, 1380-1419, 1981.

KENNEDY, B., KAWACHI, I. & PROTHROW-SMTH, D. 1996. Income distribution and mortality: test of the Robin Hoos Index in the United States. *British Medical Journal,* 47**,** 7-17.

KENNEDY, M. 2006. *Introducing Geographic Information Systems with ArcGIS : featuring GIS software from environmental systems,* Hoboken, N.J., J. Wiley.

KIM, D. S., KIM, Y. S., JUNG, K. S., CHANG, J. H., LIM, C. M., LEE, J. H., UH, S. T., SHIM, J. J. & LEW, W. J. 2005. Prevalence of chronic obstructive

pulmonary disease in Korea: A population-based spirometry survey. *American Journal of Respiratory and Critical Care Medicine,* 172**,** 842-847.

KING, A. M. & DORLING, S. 1997. PM10 particulate matter--The significance of ambient levels. *Atmospheric Environment,* 31**,** 2379-2381.

LAIRD, N. M. & WARE, J. H. 1982. Random-effects models for longitudinal data. *Biometrics,* 38**,** 963-974.

LANGE, P., PARNER, J., VESTBO, J., SCHNOHR, P. & JENSEN, G. 1998. A 15-Year Follow-Up Study of Ventilatory Function in Adults with Asthma. *The New England Journal of Medicine,* 339**,** 1194-1200.

LANGFORD, I. H. 1994. Using empirical Bayes estimates in the geographical analysis of disease risk. *Area,* 26**,** 142-149.

LAST, J. M. 1995. *A Dictionary of Epidemiology,* Oxford , New York, Oxford University Press.

LAURENT, O., PEDRONO, G., FILLEUL, L., SEGALA, C., LEFRANC, A., SCHILLINGER, C., RIVIÃ¨RE, E. & BARD, D. 2009. Influence of socioeconomic deprivation on the relation between air pollution and Î²-agonist sales for asthma. *Chest,* 135**,** 717-723.

LAWSON, A. 2001. *Statistical Methods in Spatial Epidemiology* John Wiley & Sons, Ltd.

LAWSON, A. 2009. *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology* Chapman & Hall/CRC

LEBOWITZ, M. D., HOLBERG, C. J. & MARTINEZ, F. D. 1990. A longitudinal study of risk factors in asthma and chronic bronchitis in childhood. *European Journal of Epidemiology,* 6**,** 341-347.

LIN, M., CHEN, Y., VILLENEUVE, P. J., BURNETT, R. T., LEMYRE, L., HERTZMAN, C., MCGRAIL, K. M. & KREWSKI, D. 2004. Gaseous Air Pollutants and Asthma Hospitalization of Children with Low Household Income in Vancouver, British Columbia, Canada. *American Journal of Epidemiology,* 159**,** 294-303.

LINDBERG, A., BJERG-BACKLUND, A., RANMARK, E., LARSSON, L. G. & LUNDBACK, B. 2006a. Prevalence and underdiagnosis of COPD by disease severity and the attributable fraction of smoking: Report from the Obstructive Lung Disease in Northern Sweden Studies. *Respiratory Medicine,* 100**,** 264-272.

LINDBERG, A., ERIKSSON, B., LARSSON, L. C., RÃ¶NMARK, E., SANDSTRÃ¶M, T. & LUNDBÃ¤CK, B. 2006b. Seven-year cumulative incidence of COPD in an age-stratified general population sample. *Chest,* 129**,** 879-885.

LINDBERG, A., JONSSON, A. C., RÃ¶NMARK, E., LUNDGREN, R., LARSSON, L. G. & LUNDBÃ¤CK, B. 2005. Prevalence of chronic obstructive pulmonary disease according to BTS, ERS, GOLD and ATS criteria in relation to doctor's

diagnosis, symptoms, age, gender, and smoking habits. *Respiration,* 72**,** 471-479.

LINDLEY, D. V. & SMITH, A. F. M. 1972. Bayes Estimates for the Linear Model. *Journal of the Royal Statistical Society. Series B (Methodological),* 34**,** 1-41.

LU, H., QIU, F. & CHENG, Y. Temporal and Spatial Relationship of Ozone and Asthma. Valuing Environmental Health Risk Reductions to Children. Session 4: Air Pollution and Asthma in Children, 2003 Washington.

LURZ, P. W. W., GARSON, P. J. & WAUTERS, L. A. 1997. Effects of temporal and spatial variation in habitat quality on red squirrel dispersal behaviour. *Animal Behaviour,* 54**,** 427-435.

MAGUIRE, D. J., BATTY, M. & GOODCHILD, M. F. 2005. *GIS, spatial analysis, and modeling,* Redlands, Calif., ESRI Press.

MALCOLM, A. L., DERWENT, R. G. & MARYON, R. H. 2000. Modelling the long-range transport of secondary PM10 to the UK. *Atmospheric Environment,* 34**,** 881-894.

MARITZ, J. S. & LWIN, T. 1989. *Empirical Bayes methods,* London, Chapman and Hall.

MARMOT, M. G., SHIPLEY, M. J. & ROSE, G. 1984. Inequalities in Death - Specific Explanations of a General Pattern. *Lancet,* 1**,** 1003-1006.

MARTIN, D. & WRIGHT, J. A. 2009. Disease prevalence in the English population: A comparison of primary care registers and prevalence models. *Social Science and Medicine,* 68**,** 266-274.

MARTINEZ, F. D., WRIGHT, A. L., TAUSSIG, L. M., HOLBERG, C. J., HALONEN, M., MORGAN, W. J., BEAN, J., BIANCHI, H., CURTISS, J., EY, J., SANGUINETI, A., SMITH, B., VONDRAK, T., WEST, N. & MCLELLAN, M. 1995. Asthma and Wheezing in the First 6 Years of Life. *New England Journal of Medicine,* 332**,** 133-138.

MASOLI M., FABIAN D., HOLT S. & BEASLEY R. 2004a. World map of the asthma case mortality rates, . *In:* GLOBAL INITIATIVE FOR ASTHMA (GINA) (ed.).

MASOLI M., FABIAN D., HOLT S. & BEASLEY R. 2004b. World map of The Prevalence of Asthma,. *In:* GLOBAL INITIATIVE FOR ASTHMA (GINA) (ed.).

MASOLI M., FABIAN D., HOLT S. & R., B. 2004c. The global burden of asthma: executive summary of the GINA Dissemination Committee Report. *Allergy***,** 469-478.

MEDINA-RAMON, M., ZANOBETTI, A. & SCHWARTZ, J. 2006. The effect of ozone and PM10 on hospital admissions for pneumonia and chronic obstructive pulmonary disease: A national multicity study. *American Journal of Epidemiology,* 163**,** 579-588.

MEDLINEPLUS - A.D.A.M. INC 2006. Normal versus asthmatic bronchiole.

MELVILLE A.M. , PLESS-MULLOLI T., AFOLABI O.A. & S.C., S. 2010. COPD prevalence and its association with occupational exposures in a general population. *European Respiratory Journal online,* doi:10.1183/09031936.00038309.

MENEZES, A. M. B., PEREZ-PADILLA, R., HALLAL, P. C., JARDIM, J. R., MUIÃ±O, A., LOPEZ, M. V., VALDIVIA, G., PERTUZE, J., MONTES DE OCA, M. & TÃ©LAMO, C. 2008. Worldwide burden of COPD in high- and low-income countries. Part II. Burden of chronic obstructive lung disease in Latin America: The PLATINO study. *International Journal of Tuberculosis and Lung Disease,* 12**,** 709-712.

MENEZES, A. M. B., PEREZ-PADILLA, R., JARDIM, J. R. B., MUIÃ±O, A., LOPEZ, M. V., VALDIVIA, G., DE OCA, M. M., TALAMO, C., HALLAL, P. C. & VICTORA, C. G. 2005. Chronic obstructive pulmonary disease in five Latin American cities (the PLATINO study): A prevalence study. *Lancet,* 366**,** 1875-1881.

MIRAVITLLES, M., LLOR, C., DE CASTELLAR, R., IZQUIERDO, I., BARÃ³, E. & DONADO, E. 2009. Validation of the COPD severity score for use in primary care: The NEREA study. *European Respiratory Journal,* 33**,** 519-527.

MURTAGH, E., HEANEY, L., GINGLES, J., SHEPHERD, R., KEE, F., PATTERSON, C. & MACMAHON, J. 2005. The prevalence of obstructive lung disease in a general population sample: The NICECOPD study. *European Journal of Epidemiology,* 20**,** 443-453.

MUTIUS E., FRITZSCH C., WEILAND S.K. & G., R. 1992. Prevalence of asthma and allergic disorders among children in united Germany: a descriptive comparison. *British Medical Journal,* 305**,** 1395.

NATIONAL ASTHMA CAMPAIGN. *Asthma Audit 1999/2000* [Online]. Available: http://www.asthma.org.uk/news/press08.php.

NATIONAL CENTER FOR HEALTH STATISTICS. 2005. *GIS and Public Health* [Online]. Available: http://www.cdc.gov/nchs/gis.htm [Accessed Jan 2007.

NAUENBERG, E. & BASU, K. 1999. Effect of insurance coverage on the relationship between asthma hospitalizations and exposure to air pollution. *Public Health Reports,* 114**,** 135-148.

NAURECKAS, E. T., DUKIC, V., BAO, X. & RATHOUZ, P. 2005. Short-acting Î²-agonist prescription fills as a marker for asthma morbidity. *Chest,* 128**,** 602-608.

NEIDELL, M. J. 2004. Air pollution, health, and socio-economic status: The effect of outdoor air quality on childhood asthma. *Journal of Health Economics,* 23**,** 1209-1236.

NETELER, M. & MITASOVA, H. 2004. *Open source GIS : a GRASS GIS approach,* Boston, Kluwer Academic Publishers.

NEUKIRCH, F., SEGALA, C., LE MOULLEC, Y., KOROBAEFF, M. & AUBIER, M. 1998. Short-term effects of low-level winter pollution on respiratory health of asthmatic adults. *Archives of Environmental Health,* 53**,** 320-328.

NHS THE INFORMATION CENTRE FOR HEALTH AND SOCIAL CARE. 2008. *Quality and Outcomes Framework - On line GP practice results database* [Online]. Available: http://www.qof.ic.nhs.uk/ [Accessed May 2010 2010].

NHS THE INFORMATION CENTRE FOR HEALTH AND SOCIAL CARE. 2010. *Quality and Outcomes Framework (QOF) for April 2009 - March 2010, England, Raw Prevalence Rates* [Online]. Available: http://www.ic.nhs.uk/webfiles/QOF/2009-10/Prevalence%20tables/QOF0910_National_Prevalence.xls [Accessed Oct 2010.

NORTH EAST FAMILY HEALTH SERVICES AGENCY. 2008. *RE: Matching the GP codes between the prescribing and GP population datasets.* Type to SOFIANOPOULOU, E.

OFFICE OF PUBLIC SECTOR INFORMATION. 2005. *Public Sector Information 2005 No.1515* [Online]. Available: http://www.opsi.gov.uk/si/si2005/20051515.htm Jan 2007].

ORELIEN, J. G. & EDWARDS, L. J. 2008. Fixed-effect variable selection in linear mixed models using R-2 statistics. *Computational Statistics & Data Analysis,* 52**,** 1896-1907.

PAUWELS, R. A., LOFDAHL, C. G., LAITINEN, L. A., SCHOUTEN, J. P., POSTMA, D. S., PRIDE, N. B., OHLSSON, S. V. & EUROPEAN RESP SOC STUDY CHRONIC OBSTRUCT PULM, D. 1999. Long-term treatment with inhaled budesonide in persons with mild chronic obstructive pulmonary disease who continue smoking. *New England Journal of Medicine,* 340**,** 1948-1953.

PEARCE, N. & DOUWES, J. 2005. Commentary: Asthma time trends - Mission accomplished? *International Journal of Epidemiology,* 34**,** 1018-1019.

PELKONEN, M., NOTKOLA, I. L., NISSINEN, A., TUKIAINEN, H. & KOSKELA, H. 2006. Thirty-year cumulative incidence of chronic bronchitis and COPD in relation to 30-year pulmonary function and 40-year mortality: A follow-up in middle-aged rural men. *Chest,* 130**,** 1129-1137.

PETERSEN, J., ATKINSON, P., PETRIE, S., GIBIN, M., ASHBY, D. & LONGLEY, P. 2009. Teenage pregnancy-New tools to support local health campaigns. *Health & Place,* 15**,** 300-307.

PETERSEN J., ATKINSON P., PETRIE S., GIBIN M., ASHBY D. & P., L. 2008. Teenage pregnancy-new tools to support local health campaigns. *Health & Place,* 15**,** 300-7.

PETTY, T. L. 1990. The National Mucolytic Study - Results of a Randomized, Double-Blind, Placebo-Controlled Study of Iodinated Glycerol in Chronic Obstructive Bronchitis. *Chest,* 97**,** 75-83.

PICKETT, K. E. & PEARL, M. 2001. Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review. *Journal of epidemiology and community health.,* 55**,** 111-122.

PINHEIRO, C. J. & BATES, M. D. 2000. *Mixed Effects Models in S and S-Plus* Springer Verlag, New York

PINHEIRO J., BATES D., DEBROY S. & SARKAR D. 2008. The nlme Package. *In:* R-CORE (ed.) *R-project.*

PINNOCK, H. & SHEIKH, A. 2009. Primary care research and clinical practice:respiratory disease. *Postgraduate Medical Journal,* 85**,** 74-79.

PITARD, A., ZEGHNOUN, A., COURSEAUX, A., LAMBERTY, J., DELMAS, W., FOSSARD, J. L. & VILLET, H. 2004. Short-term associations between air pollution and respiratory drug sales. *Environmental Research,* 95**,** 43-52.

PITT, B., BYINGTON, R. P., FURBERG, C. D., HUNNINGHAKE, D. B., MANCINI, G. B. J., MILLER, M. E., RILEY, W. & INVESTIGATORS, P. 2000. Effect of amlodipine on the progression of atherosclerosis and the occurrence of clinical events. *Circulation,* 102**,** 1503-1510.

POPE C.A., DOCKERY D.W. & J., S. 1995. REVIEW OF EPIDEMIOLOGICAL EVIDENCE OF HEALTH-EFFECTS OF PARTICULATE AIR-POLLUTION. *Inhal Toxicol,* 7**,** 1-18.

POPE III, C. A., BURNETT, R. T., THURSTON, G. D., THUN, M. J., CALLE, E. E., KREWSKI, D. & GODLESKI, J. J. 2004. Cardiovascular Mortality and Long-Term Exposure to Particulate Air Pollution: Epidemiological Evidence of General Pathophysiological Pathways of Disease. *Circulation,* 109**,** 71-77.

PRESCRIBING SUPPORT UNIT. 2000. *Average Daily Quantities & STAR(97)-PUS* [Online]. Available: http://www.ic.nhs.uk/webfiles/Services/PSU/ADQ%20and%20STAR(97)PUs%20Report.pdf [Accessed Jan 2007.

PRESCRIBING SUPPORT UNIT. 2005. *ADQs* [Online]. Available: http://www.ic.nhs.uk/psu/measure/presmeas/copy_of_adqs/ [Accessed Jan 2007.

PRESCRIPTION PRICING AUTHORITY 2006. Drugs Used in Asthma and COPD. *In:* NATIONAL HEALTH SYSTEM (NHS) (ed.).

PRESCRIPTION PRICING AUTHORITY 2007. Trends in prescribing of drugs for asthma and COPD in General Practices in England National Health System - NHS.

PRESCRIPTION PRICING DIVISION. 2008. *Prescribing Review: Drugs used in Asthma and COPD* [Online]. Available: http://www.ppa.org.uk/news/pact-122005.htm [Accessed 18 Nov 2008].

RABINOVITCH, N., STRAND, M. & GELFAND, E. W. 2006. Particulate levels are associated with early asthma worsening in children with persistent disease. *American Journal of Respiratory and Critical Care Medicine,* 173**,** 1098-1105.

RAUDENBUSH, S. W. & BRYK, A. S. 2002. *Hierarchical linear models : applications and data analysis methods,* Thousand Oaks, Calif., Sage Publications.

REVELT, D. & TRAIN, K. 1998. Mixed logit with repeated choices: Households' choices of appliance efficiency level. *Review of Economics and Statistics,* 80**,** 647-657.

RIBEIRO JR., P. J. & DIGGLE, P. J. G. A. P. F. G. A. 2001. geoR: A package for geostatistical analysis. R-NEWS Vol 1, No 2. ISSN 1609-3631.

RICHARDSON, S., ABELLAN, J. J. & BEST, N. 2006. Bayesian spatio-temporal analysis of joint patterns of male and female lung cancer risks in Yorkshire (UK) *Statistical Methods in Medical Research,* 15**,** 385-407.

RICHARDSON, S., STUCKER, I. & HEMON, D. 1987. Comparison of Relative Risks Obtained in Ecological and Individual Studies - Some Methodological Considerations. *International Journal of Epidemiology,* 16**,** 111-120.

ROBERT, S. A. 1998. Community-Level Socioeconomic Status Effects on Adult Health *Journal of Health and Social Behavior,* 39**,** 18-37.

ROBERTSON, C. F., ROBERTS, M. F. & KAPPERS, J. H. 2004. Asthma prevalence in Melbourne schoolchildren: Have we reached the peak? *Medical Journal of Australia,* 180**,** 273-276.

ROMEO, E., DE SARIO, M., FORASTIERE, F., COMPAGNUCCI, P., STAFOGGIA, M., BERGAMASCHI, A. & PERUCCI, C. A. 2006. PM 10 exposure and asthma exacerbations in pediatric age: a meta-analysis of panel and time-series studies. *Esposizione a PM10 e riacutizzazioni dell'asma in etÃ pediatrica: una metanalisi degli studi di panel e di serie temporale.,* 30**,** 245-254.

ROSSI, O. V., KINNULA, V.L.,, TIENARI, J. & HUHTI, E. 1993. Association of severe asthma attacks with weather, pollen, and air pollutants. *Thorax,* 48**,** 244-248.

RUSHTON, S. P., LURZ, P. W. W., FULLER, R. & GARSON, P. J. 1997. Modelling the distribution of the red and grey squirrel at the landscape scale: A combined GIS and population dynamics approach. *Journal of Applied Ecology,* 34**,** 1137-1154.

RUSHTON, S. P., LURZ, P. W. W., GURNELL, J., NETTLETON, P., BRUEMMER, C., SHIRLEY, M. D. F. & SAINSBURY, A. W. 2006. Disease threats posed by alien species: The role of a poxvirus in the decline of the native red squirrel in Britain. *Epidemiology and Infection,* 134**,** 521-533.

SALMOND, C., CRAMPTON, P., S HALES, S LEWIS & N PEARCE 1999. Asthma prevalence and deprivation: a small area analysis. *Journal of Epidemiology and Community Health,* 53**,** 476-480.

SAUERZAPF, V., JONES, A. P. & CROSS, J. 2009. Environmental factors and hospitalisation for chronic obstructive pulmonary disease in a rural county of England. *Journal of Epidemiology and Community Health,* 63**,** 324-328.

SCHILDCROUT, J. S., SHEPPARD, L., LUMLEY, T., SLAUGHTER, J. C., KOENIG, J. Q. & SHAPIRO, G. G. 2006. Ambient air pollution and asthma exacerbations in children: An eight-city analysis. *American Journal of Epidemiology,* 164**,** 505-517.

SCHUURMAN, N. 2004. *GIS : a short introduction,* Oxford, Blackwell.

SCHWARTZ, J. 1994. Air Pollution and Hospital admissions for the Elderly in Birmingham, Alabama. *American Journal of Epidemiology,* 139**,** 589-98.

SCOTT, D. W. 1992. *Multivariate density estimation : theory, practice, and visualization,* Wiley.

SEARS, M. R., GREENE, J. M., WILLAN, A. R., WIECEK, E. M., TAYLOR, D. R., FLANNERY, E. M., COWAN, J. O., HERBISON, G. P., SILVA, P. A. & POULTON, R. 2003. A longitudinal, population-based, cohort study of childhood asthma followed to adulthood. *New England Journal of Medicine,* 349**,** 1414-1422.

SHI X 2009. A Geocomputational Process for Characterizing the Spatial Pattern of Lung Cancer Incidence in New Hampshire *Annals of the Association of American Geographers,* 99**,** 521-533.

SIAFAKAS, N. M., VERMEIRE, P., PRIDE, N. B., PAOLETTI, P., GIBSON, J., HOWARD, P., YERNAULT, J. C., DECRAMER, M., HIGENBOTTAM, T., POSTMA, D. S. & REES, J. 1995. Optimal assessment and management of chronic obstructive pulmonary disease (COPD). *European Respiratory Journal,* 8**,** 1398-1420.

SILVERMAN, B. W. 1986. *Density estimation for statistics and data analysis,* London, Chapman and Hall.

SMITH, A., CULLIS, B. & THOMPSON, R. 2001. Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. *Biometrics,* 57**,** 1138-1147.

SMITH W.H.F. & WESSEL P. 1990. Gridding with continuous curvature splines in tension. *Geophysics,* 55**,** 293-305.

SNIJDERS, T. & BOSKER, R. 1999. *Multilevel Analysis: an introduction to basic and advanced multilevel modeling,* London: Sage Publications.

SORIANO, J. B., DAVIS, K. J., COLEMAN, B., VISICK, G., MANNINO, D. & PRIDE, N. B. 2003. The proportional venn diagram of obstructive lung disease: Two approximations from the United States and the United Kingdom. *Chest,* 124**,** 474-481.

SORIANO, J. B., MAIER, W. C., EGGER, P., VISICK, G., THAKRAR, B., SYKES, J. & PRIDE, N. B. 2000. Recent trends in physician diagnosed COPD in women and men in the UK. *Thorax,* 55**,** 789-794.

SPEIZER, F. E., FAY, M. E., DOCKERY, D. W. & FERRIS JR, B. G. 1989. Chronic obstructive pulmonary disease mortality in six U.S. cities. *American Review of Respiratory Disease,* 140.

SPEIZER, F. E. & TAGER, I. B. 1979. Epidemiology of chronic mucus hypersecretion and obstructive airways disease. *Epidemiol Rev,* 1**,** 124-42.

STARFIELD B, SHI LY & MACINKO J 2005. Contribution of primary care to health systems and health. *MILBANK QUARTERLY,* 83**,** 457-502.

STARFIELD B. 1994. Is primary care essential ? *Lancet,* 344**,** 1129-1133.

STEDMAN, J. R. 1997. A U.K.-wide episode of elevated particle (PM10) concentration in March 1996. *Atmospheric Environment,* 31**,** 2381-2383.

STRACHAN, D. P., BUTLAND, B. K. & ANDERSON, H. R. 1996. Incidence and prognosis of asthma and wheezing illness from early childhood to age 33 in a national British cohort. *British Medical Journal,* 312**,** 1195-1199.

SUNYER, J., SAEZ, M., MURILLO, C., CASTELLSAGUE, J., MARTINEZ, F. & ANTO, J. M. 1993. Air pollution and emergency room admissions for chronic obstructive pulmonary disease: A 5-year study. *American Journal of Epidemiology,* 137**,** 701-705.

SUNYER, J., SPIX, C., QUÃ©NEL, P., PONCE-DE-LEÃ³N, A., PÃ¶NKA, A., BARUMANDZADEH, T., TOULOUMI, G., BACHAROVA, L., WOJTYNIAK, B., VONK, J., BISANTI, L., SCHWARTZ, J. & KATSOUYANNI, K. 1997. Urban air pollution and emergency admissions for asthma in four European cities: The APHEA project. *Thorax,* 52**,** 760-765.

TERZOPOULOS, D. 1988. The computation of visible-surface representations. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 10**,** 417-438.

TEUTSCH, S. M. & CHURCHILL, R. E. 2000. *Principles and practice of public health surveillance,* Oxford, Oxford University Press.

TIMMRECK, T. C. 2002. *An introduction to epidemiology,* Sudbury, Mass., Jones and Bartlett Publishers.

TOBLER, W. R. 1970. A computer Movie Simulation Urban Growth in the Detroit Region. *Economic Growth,* 46**,** 234-240.

TOLBERT, P. E., MULHOLLAND, J. A. & DAVID L. MACINTOSH1, FAN XU1, DANNI DANIELS1, OWEN J. DEVINE4, BRADLEY P. CARLIN5, MITCHEL KLEIN1, ANDRE J. BUTLER2, DALE F. NORDENBERG6, HOWARD FRUMKIN1, P. BARRY RYAN1 AND MARY C. WHITE1 2000. Air Quality and Paediatric Emergency Room Visits for Asthma and Atlanta, Georgia. *American Journal of Epidemiology,* 151**,** 798-810.

TYNE AND WEAR AIR QUALITY INFORMATION. 2010. Available: http://tyneandwearair.sunderland.ac.uk/ [Accessed 01 April 2010 2010].

UK NATIONAL AIR QUALITY ARCHIVE. 2010. Available: http://www.airquality.co.uk/ [Accessed 1 April 2010 2010].

UNUTZER, J., KATON, W., CALLAHAN, C. M., WILLIAMS, J. W., HUNKELER, E., HARPOLE, L., HOFFING, M., DELLA PENNA, R. D., NOEL, P. H., LIN, E. H. B., AREAN, P. A., HEGEL, M. T., TANG, L. Q., BELIN, T. R., OISHI, S., LANGSTON, C. & INVESTIGATORS, I. 2002. Collaborative care management of late-life depression in the primary care setting - A randomized controlled trial. *Jama-Journal of the American Medical Association,* 288**,** 2836-2845.

VAN DER LINDE, A., WITZKO, K.-H. & JOCKEL, K.-H. 1995. Spatio-temporal analysis of mortality using splines. *Biometrics,* 51**,** 1352-1360.

VEGNI, F. E., CASTELLI, B., AUXILIA, F. & WILKINSON, P. 2005. Air pollution and respiratory drug use in the city of Como, Italy. *European Journal of Epidemiology,* 20**,** 351-358.

VERMEIRE, P. A., RABE, K. F., SORIANO, J. B. & MAIER, W. C. 2002. Asthma control and differences in management practices across seven European countries. *Respiratory Medicine,* 96**,** 142-149.

VESTBO, J. 2004. COPD in the ECRHS. *Thorax,* 59**,** 89-90.

VESTBO, J. & LANGE, P. 2002. Can GOLD stage 0 provide information of prognostic value in chronic obstructive pulmonary disease? *American Journal of Respiratory and Critical Care Medicine,* 166**,** 329-332.

VIEGI, G., MATTEELLI, G., ANGINO, A., SCOGNAMIGLIO, A., BALDACCI, S., SORIANO, J. B. & CARROZZI, L. 2004. The proportional Venn diagram of obstructive lung disease in the Italian general population. *Chest,* 126**,** 1093-1101.

VIEGI, G., PEDRESCHI, M., PISTELLI, F., DI PEDE, F., BALDACCI, S., CARROZZI, L. & GIUNTINI, C. 2000. Prevalence of airways obstruction in a general population: European Respiratory Society vs American Thoracic Society definition. *Chest,* 117.

VIEGI, G., PISTELLI, F., SHERRILL, D. L., MAIO, S., BALDACCI, S. & CARROZZI, L. 2007. Definition, epidemiology and natural history of COPD. *European Respiratory Journal,* 30**,** 993-1013.

VON KLOT, S., WOLKE, G., TUCH, T., HEINRICH, J., DOCKERY, D. W., SCHWARTZ, J., KREYLING, W. G., WICHMANN, H. E. & PETERS, A. 2002. Increased asthma medication use in association with ambient fine and ultrafine particles. *European Respiratory Journal,* 20**,** 691-702.

WALLER, L. A. & GOTWAY, C. A. 2004a. *Applied spatial statistics for public health data,* Hoboken, N.J. ; [Great Britain], John Wiley & Sons.

WALLER, L. A. & GOTWAY, C. A. 2004b. *Applied Spatial Statistics for Public Health Data* John Wiley & Sons, Inc.

WALLEY, T. & ROBERTS, D. J. 2000. Average Daily Quantities - a tool for measuring prescribing volume in England. *Pharmacoepidemiology and Drug Safety* 9**,** 55-58.

WALTER, S. D. 2000. Disease mapping: a historical perspective. *In:* ELLIOT, P., WAKEFIELD, J., BETS, N. & BRIGGS, D. (eds.) *Spatial Epidemiology: Methods and Applications.* Oxford: Oxford University Press.

WALTERS, J. A., HANSEN, E. C., JOHNS, D. P., BLIZZARD, E. L., WALTERS, E. H. & WOOD-BAKER, R. 2008. A mixed methods study to compare models of spirometry delivery in primary care for patients at risk of COPD. *Thorax,* 63**,** 408-414.

WALTERS, S., GRIFFITHS, R. K. & AYRES, J. G. 1994. Temporal Association Between Hospital Admissions for Asthma in Birmingham and Ambient Levels of Sulphur Dioxide and Smoke. *Thorax,* 49**,** 133-140.

WATSON, J., COWEN, P. & LEWIS, R. 1996. The relationship between asthma admission rates, routes of admission, and socioeconomic deprivation. *European Respiratory Journal,* 9**,** 2087-2093.

WEISBERG, S. 2005. *Applied linear regression,* Hoboken, N.J., Wiley-Interscience.

WHITE, D., BURTON, M. & DOW, M. 1981. Sexual Division of labor in African agriculture: a network autocorrelation analysis. *American Anthropologist,* 84**,** 824-49.

WHITESIDE, H., SCRIVENER, G., ROBERTS, D. & WALLEY, T. 2001. ADQs: a new measure of prescribing volume. 12**,** 23-29.

WILKINSON, P., ELLIOTT, P., GRUNDY, C., SHADDICK, G., THAKRAR, B., WALLS, P. & FALCONER, S. 1999. Case-control study of hospital admission with asthma in children aged 5-14 years: relation with road traffic in north west London. *Thorax,* 54**,** 1070-1074.

WILKINSON, R. G. 1996. *Unhealthy Societies: the afflictions of inequality,* London, Routledge

WILSON, W. 2004. The personal exposure fallacy in particulate matter (PM) time-series epidemiology. *Epidemiology,* 15**,** S56.

WOODRUFF, P. G. & FAHY, J. V. 2001. Asthma: Prevalence, pathogenesis and prospects for novel therapies. *Journal of American Medical Association,* 286**,** 395.

WORLD HEALTH ORGANIZATION. Declaration of Alma-Ata: International Conference on Primary Health Care. 6-12 September 1978 1978 USSR, Geneva.

WORLD HEALTH ORGANIZATION 1994. Public Health surveillance: historical origins, methods and evaluation. Declich, S.,

Carter, A.O., ed.

WORLD HEALTH ORGANIZATION. 2000. *Bronchial asthma* [Online]. Available: http://www.who.int/respiratory/asthma/en/index.html [Accessed Jan 2006.

WORLD HEALTH ORGANIZATION 2003. A Global Review of Primary Health Care: Emerging Messages. Geneva.

WORLD HEALTH ORGANIZATION. 2008a. *61st World Health Assembly (19-24 May 2008) - Chronic respiratory diseases* [Online]. Available: http://www.who.int/respiratory/asthma/WHA_2008/en/index.html [Accessed 22 June 2009 2009].

WORLD HEALTH ORGANIZATION. Prevention and control of noncommunicable diseases: implementation of the global strategy. 61st World Health Assembly (19-24 May 2008), 2008b.

WORLD HEALTH ORGANIZATION. 2009a. *Chronic respiratory diseases* [Online]. Available: http://www.who.int/respiratory/en/ [Accessed 16 June 2009.

WORLD HEALTH ORGANIZATION. 2009b. *Global Alliance against Chronic Respiratory Diseases (GARD)* [Online]. Available: http://www.who.int/respiratory/gard/en/ [Accessed 16 June 2009 2009].

WORLD HEALTH STATISTICS 2008. World Health Organization. *In:* WHO PRESS (ed.). Geneva.

XIA, H. & CARLIN, B. P. 1998. Spatio-temporal models with errors in covariates: mapping Ohio lung cancer mortality. *Statistics in Medicine,* 17**,** 2025-2043.

YEN, I. H. & SYME, S. L. 1999. The social environment and health: a discussion of the epidemiologic literature. *Annual Review of Public Health,* 20**,** 287-308.

YOUNG, P. C., PEDREGAL, D. J. & TYCH, W. 1999. Dynamic harmonic regression. *Journal of Forecasting,* 18**,** 369-394.

ZEGHNOUN, A., BEAUDEAU, P., CARRAT, F., DELMAS, V., BOUDHABHAY, O., GAYON, F., GUINCETRE, D. & CZERNICHOW, P. 1999. Air pollution and respiratory drug sales in the city of Le Havre, France, 1993-1996. *Environmental Research,* 81**,** 224-230.

ZHONG, N., WANG, C., YAO, W., CHEN, P., KANG, J., HUANG, S., CHEN, B., WANG, C., NI, D., ZHOU, Y., LIU, S., WANG, X., WANG, D., LU, J., ZHENG, J. & RAN, P. 2007. Prevalence of chronic obstructive pulmonary disease in China: A large, population-based survey. *American Journal of Respiratory and Critical Care Medicine,* 176**,** 753-760.

ZHOU, S. Y., DING, Y., CHEN, Y. & SHI, J. J. 2003. Diagnosability study of multistage manufacturing processes based on linear mixed-effects models. *Technometrics,* 45**,** 312-325.

ZHU, L., CARLIN, B. P. & A.E., G. 2003. Hierarchical regression with misaligned spatial data: relating ambient ozone and pediatric asthma ER visits in Atlanta. *Environmetrics,* 14**,** 537-557.

ZHU, L., CARLIN, B. P., ENGLISH, P. & SCALF, R. 2000. Hierarchical modeling of spatio-temporally misaligned data: relating traffic density to pediatric asthma hospitalizations. *Environmetrics,* 11**,** 43-61.

ZORASTER, S. 2003. A surface modeling algorithm designed for speed and ease of use with all petroleum industry data. *Computers & Geosciences,,* 29**,** 175-182.