The Met Office, Health Forecasting and Public Health

Doctor of Public Health Internship Report:

Developing Tools for Asthma Forecast in London

Period: September 2008 – June 2009

Host Institution: **The Met Office Hadley Centre, Exeter**
Student/ID: **INS/0725657***
Supervisors: **DDR***
Mentors: **PM§, TML§, CS§**

§The Met Office Hadley Centre, Exeter
***Centre for Public Health Research, Brunel University Graduate School**

Declaration

I declare that the work conducted in this study is original. Apart from the citations, which are duly referenced, no part of this document has been reported elsewhere.

The thoughts and ideas or opinions expressed in this report do not necessarily represent those of the Met Office or Brunel University. The primary author takes full responsibility for any misrepresentation or faults in this report.

Acknowledgement

I do express my gratitude and appreciation to all who facilitated in any way to make this study successful. I am particularly grateful for the close contact maintained and technical support provided by Mentors at the Met Office Hadley Centre and academic Supervisors here in Brunel University. Your remarks have been very helpful. To my colleagues and mates, I say thank you for all the criticisms comments and heckling.

Table of Contents

List of Tables

List of Figures

List of acronyms

APHEA	Air Pollution and Health: A European Approach
AQ	Air Quality
AURN	Automatic Urban and Rural Network
COMEAP	Committee
COPD	Chronic Obstructive Pulmonary Disease
DH	Department of Health
DQI	Data Quality Indicator
EPV	Extreme Predictive Value
GINA	Global Initiative for Asthma
GLM	Generalised Linear Model
HES	Hospital Episode Statistics
HL	Hosmer-Lemeshow
HPA	Health Protection Agency
ICD	International Classification of Diseases
IPCC	Intergovernmental Panel on Climate Change
LRM	Logistic Regression Model
NAME	Numerical Atmospheric-dispersion Modelling Environment
NegBin	Negative Binomial
NHS	National Health Service
NMMAPS	National Morbidity, Mortality and Air Pollution Study
NPV	Normal Predictive Value
PM	Particulate Matter
QRM	Quantile Regression Model
RMSE	Root Mean Square Error
UK	United Kingdom
WLSIN	Weighted Least-Squares Iteration Numbers

INTRODUCTION

Background

Asthma is a respiratory illness manifested by an inflammation and or subsequent obstruction of air flow within the respiratory system [1, 2]. It is known that the inflammation of the air ways of an individual may result in asthmatic symptoms. Alternatively, the inflammation may lead to an obstruction of the air flow directly or indirectly by causing a hyperresponsiveness[1] of the air way, which may then cause an obstruction of the air ways. This cycle is partly illustrated in Figure 1 below.

There are two categories of factors that can affect the manifestation of asthmatic symptoms described above. These are factors that are either known to be partly responsible for the development of the disease in an individual, or are risk factors that exacerbate asthma. In either group, they include air quality and weather related factors as well as tobacco consumption (either in a passive or active manner) and the exposure to allergens (e.g. molds, pollen, cockroaches) [3-7]. Other risk factors that exacerbate asthma include respiratory infections (bacterial and viral), physical exercise and its related condition of hyperventilation[2], as well as bronchoconstriction[3] [8-10]. Race, gender, socio-economic status /family size [11-14] and extreme emotional expression [15] have also been strongly associated with asthma events in different studies. Predicting asthma episodes in an ideal situation would require that we account for all the known potential predictors/indicators, however, the availability of data that is useable is a common limitation.

This study focuses on asthma episodes and how they are related to weather and air quality factors. The purpose of the study was to use the existing relationships to predict daily asthma admissions, and to explore the chances of predicting extreme asthma events in London.

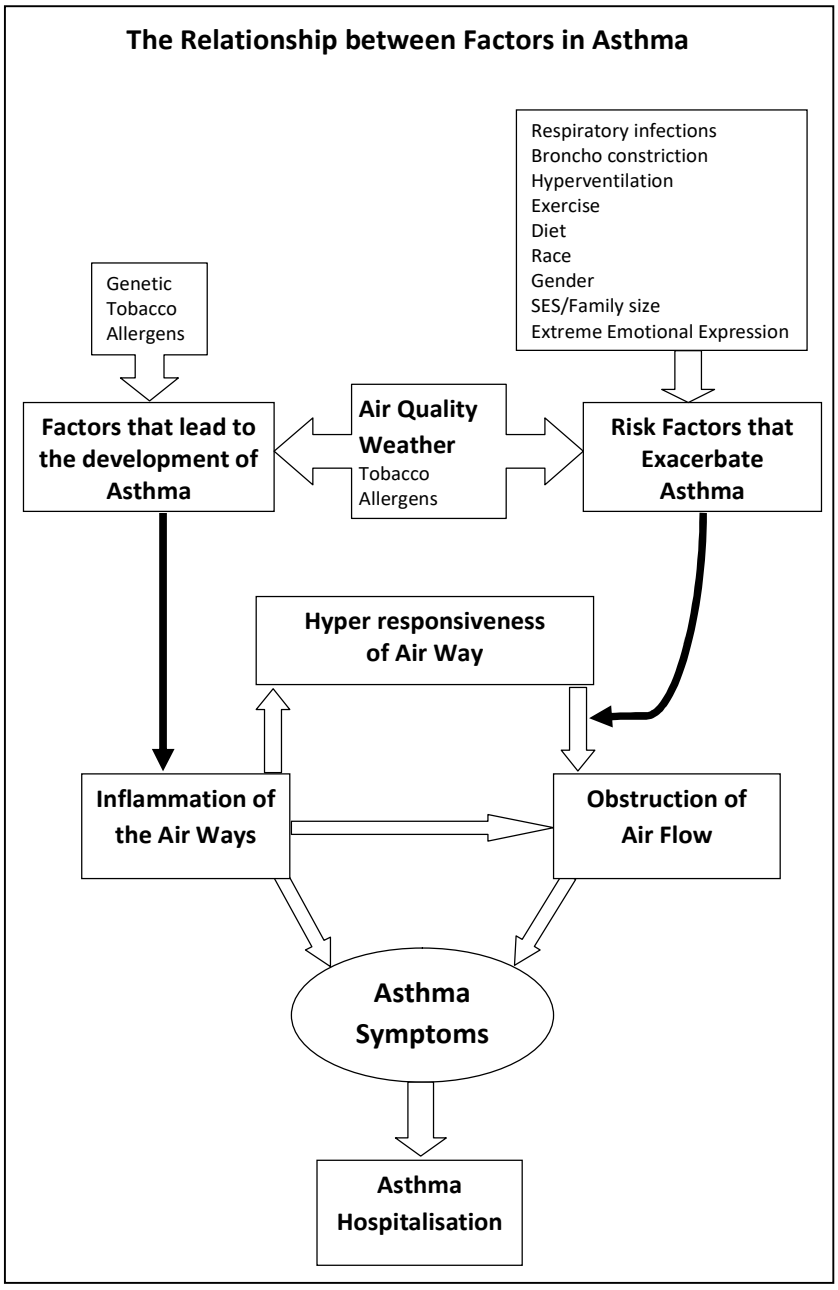
The approach to developing tools for asthma forecasting was wholly quantitative. There are known associations /relationships between environmental (weather and air quality) factors and asthma. There are also available longitudinal datasets on some of these weather and air quality indicators, as well as asthma episodes statistics. Hence the choice of time series modeling, using total asthma daily admissions in London as the dependent variable and all others as potential predictors was appropriate. Distributions of asthma hospital admissions were statistically matched to weather and air quality factors to provide a multivariate description of the daily admissions risks [16].

Suitable statistical models and methods were considered because the anticipated predictive models comprised many variables that respond in different ways to the dependent variable (particularly for categorical variables). The findings from these

[1] A state characterized by easily triggered contraction of the small airways (spasm).
[2] Breathing in excess of what the body needs i.e. rapid or deep breathing that can occur with exercise, anxiety or panic.
[3] The constriction of the airways in the lungs due to the tightening of surrounding smooth muscle, with consequent coughing, wheezing, and shortness of breath.

modeling techniques provide some important tools needed in formulating an asthma forecast.

Figure 1 The Relationship between Factors in Asthma



Asthma

Asthma is a condition of great public health concern here in the UK and globally [17-21]. It is reported that the disease affects people of all races and ethnic groups worldwide, from infancy to old age, but with slightly more boys than girls affected and, after puberty, more women than men [17, 22]. This is consistent with some earlier findings in the United Kingdom (England & Wales), where a similar pattern was reported from data on patients consulting their General Parishioners for asthma [2]. Currently global estimates suggest that as many as 300 million people are affected worldwide [17, 22, 23], and the global burden of the condition is rising particularly among children [17, 20].

Asthma has a predictable prognosis; that is to say, the course of the disease is well understood [24]. However its diagnoses remain a challenge, as the disease is not clearly defined by particular set of conditions, but a mix of several factors [18, 25]. There are numerous and quite unpredictable underlying causes of asthma, some of which include genetic and environmental factors [26, 27]. As a result of the complex nature of the condition, some of the diagnostic techniques commonly employed range from the history and patterns of symptoms, physical examination and lung function measurements including spirometry, to the skin test identification of allergens [28].

Several studies have shown changes in the global epidemiology of asthma. Developed countries have consistently shown dramatic increases in the prevalence and this change has, more recently been observed in some less-developed countries [20, 29-31]. In 2001 the (UK) National Asthma Campaign [32] reported that asthma affected over five million people, about one in five households. Later reports indicated that the United Kingdom had one of the highest prevalence rates of more than 15% [21], with the latest estimates supporting this [17]. In England 67,077 people were hospitalized for asthma between April 2006 and March 2007, of whom more than 40% were children under the age of 15 years [33]. According to the Hospital Episode Statistics of the Department of Health (January 2001 - December 2006), London which is the busiest and most densely populated areas in the UK had about 56,832 asthma related hospital admissions recorded over that period giving a crude annual rate of 9,472. This situation presents asthma as an important condition of public health concern with dimensions not just limited to the individual(s) affected, but also posing a significant burden on the health care resources as well as the society more broadly.

In spite of rising prevalence and a significant socioeconomic burden, the fundamental biology or pathophysiology of asthma remains poorly or inaccurately defined with no clear consensus [23, 34, 35]. Notwithstanding this, the diagnosis of asthma based on physician's assessment and recorded as hospital admissions, is a vital source of data. These data capture those acute, severe episodes requiring hospital attendance. They supplement self reported morbidity [36] and general practice records [37], providing a comprehensive picture of the full scale of asthma morbidity.

Asthma, like most respiratory illnesses, is strongly affected by environmental conditions. The association between asthma and some environmental factors have been modelled to identify some of the underlying causes, and this has been useful for mitigating asthma events particularly in the UK [38]. Nonetheless our

understanding of the exact mechanisms at play in this multi-complex condition are under-developed and do not permit a comprehensive prediction model.

The Biology of Asthma

Asthma is a chronic respiratory condition typified by obstruction and continuously persistent inflammation of the airways [39]. This obstruction to airflow, which is episodic within individuals with early or mild asthma, can cause symptoms of tightness and wheeziness in the chest [2]. Recently British and American asthma education, prevention and management guidelines also include acute or sub-acute episodes of progressively worsening shortness of breath, cough, wheezing and chest tightness or some combination of these symptoms. These symptoms are accompanied by decreases in expiratory airflow and objective measures of lung functioning that employ spirometry and peak flow [28, 40, 41].

Various biological models have been advanced to explain the processes involved in asthma morbidity generally relying on interplay between genetic and environmental factors, and the immune response. The details of this are described elsewhere[1] [2, 23, 42].

The UK Committee on the Medical Effects of Air Pollutants (COMEAP)[2] in 1995, classified asthma as a disease of the lungs in which the airways are unusually sensitive to a wide range of stimuli, including inhaled irritants and allergens. They further elaborated on the role of environmental stimuli, particularly air pollutants in triggering or exacerbating the condition. The inherent interdependence or independent effects of known environmental determinants of ill health, particularly air pollutants and some weather factors has been reported by authors who have looked at the effect of the environment on asthma exacerbations [43-45].

Some other biological changes that result in increasing an individual's vulnerability to asthma exacerbation and are initiated by environmental changes may be of important note. A common one relates to inflammatory and structural changes in the airways in the lung, which contribute to the full manifestation of the chronic form of asthma [46-49]. This remodelling of the airways increases an individual's predisposition to asthma [24, 46, 48, 50], and thus supports the proposition that environmental factors play a critical role in the inception and progression of the disease in genetically susceptible individuals [29, 30].

Asthma is a complex multi-system condition, the appropriate prevention or management strategy of which transcend the known biological causes [51].

The Epidemiology of Asthma

Debates in the scientific literature on the epidemiology of asthma in England and Wales have often focussed on the asthma trends over the past few decades. Are increases or decreases in asthma prevalence due to changes in the environment, and if so what are those changes; are they attributable to changes in the

[1] The "*Ontogeny of atopic asthma*" as illustrated by the relationship between genetic and environmental factors in asthma, DH, 1995 pp16-18

[2] The Department of Health (DH) asked Committee on the Medical Effects of Air Pollutants (COMEAP), 1995 to advise on the possible links between outdoor air pollution and asthma, excluding biological pollutants such as pollen. This constituted the report: "ASTHMA AND OUTDOOR AIR POLLUTION" published by the HSMO

population; or is there another explanation entirely? It has, for instance, been suggested that declines in prevalence of asthma that have sometimes occurred are entirely attributable to variations in diagnosis [52-54]. Arguments have been put forward about changes in diagnostic categories or misdiagnoses that could explain, say, rises or falls in the rates of acute bronchitis compared with asthma. The evidence for this in the literature, however, does not adequately account for the changes in asthma prevalence over the past decades.

In recent years a number of studies published have identified environmental factors that appear to trigger asthma exacerbations or protect against the development of asthma. The "Genetic-Environment" interaction and resultant changes that affect asthma have equally been discussed by many authors. However the striking note highlighted in one of the debates is the fact that the expression of environmental and genetic determinants of a complex disease such as asthma, depends on the context in which this occurs [55]. Local environmental conditions are thus important in determining the likely impact or manifestation of asthma as some of these notable factors such as temperature, humidity, pressure as well as air pollutants do laterally interact and hence do not independently affect asthma/ asthma exacerbation [44, 56-62]. The discussion on asthma epidemiology beyond the genetic or environment link(s) is not exhaustive, hence for the purpose of this study; subsequent sections will highlight the relationship between asthma and some specific weather and air quality indicators.

The Disease Burden of Asthma [Individual / Health Service (Cost and Infrastructure)]

The severity of asthma has been classified into four different categories by the Global Initiative for Asthma [17]. According to its earlier report, these four categories were intermittent, mild persistent, moderate persistent or severe persistent asthma. These asthma classifications were based on the clinical features of the disease as well as the prescribed/ corresponding treatment plan [63].
Lately, the classification has considered the etiology as well as environmental factors [64, 65].

On a per capita basis, the UK has the greatest burden of severe asthma of any country in Europe [66]. The Global Initiative for Asthma (GINA) currently reports that more than 18% of people in Scotland, 17% of people in Wales and 15.3% of people in England experience symptoms of asthma. This compare unfavourably with 8.2% of people in the US [67], 7% in Germany and 7% in France [68]. Of those asthma sufferers in the UK, some studies suggests the patients with severe asthma account for the majority of hospitalisations due to asthma [66]. This group of severe asthma sufferers consist of 2.6 million individuals (i.e. 2.1 million adults and 500,000 children) in the UK [69].

Medical Practitioners in the UK are seeing 20,000 new cases of asthma each week and about 30% of children aged 13-14 years are known to have asthma symptoms [17]. In the UK in 2004, there were 75,000 emergency hospital admissions due to asthma and 1,500 fatalities [17]. It was recently reported that about 5.4 million people in the UK are currently *receiving* treatment for asthma, 1.1 million of these are children [33]. Despite these treatments, work done by Demoly and colleagues to describe the characteristics of "asthma control" in some European countries including the UK suggest, that a substantial portion of asthmatics are "not well-

controlled" and thus need to be educated on the importance of asthma control and adherence to treatments [70].

The burden of Asthma, however, is not solely a health burden; there is also an associated economic burden. In 2004, it was estimated that asthma cost the UK over £2.3 billion a year [19], including £1.2 billion in individual productivity losses [71]. The Office for Health Economics further estimated that the cost to the NHS alone in 2001 totalled £889 million. Most of that was associated with dispensing and prescriptions (£659 million), but around 5.5% of the cost was associated with hospital admissions. In addition, poorly controlled asthma appears to have a considerable impact on health care costs [72].

Relationship between Environmental factors and Asthma events

The *Department of Health* and *Health Protection Agency* published a review of the health effects of climate change in 2008 [16]. In the report the two component issues of the environment with which we are principally interested, weather and air quality, were both considered.

The constituent indicators (e.g. temperature, humidity, vapour pressure, wind, as well as atmospheric aerosols) and their prevailing regimes at some altitudes can promote the formation of mists, fogs and polluted environment [73]. Meanwhile environmental factors do have intricate interrelationships and their collective impact on health may even be more complicated to estimate. Hence when presented with this kind of environment with several emissions of various pollutants, there is a propensity to cause environmental pollution near ground or habitable levels. The environmental dynamics as well as its pollution can exacerbate asthma through many ways and some of the mechanisms involved have been discussed [74].

Health conditions triggered by global environmental changes as well as occupational exposures vary considerably in symptoms and these are known to primarily depend on the individual's susceptibility and level of exposure [75-78]. Vulnerable groups within given populations, particularly children and the elderly tend to be the hardest hit with the former experiencing both the direct and indirect effects of these changes [77, 79]. The evidence for environmental effects on health is based on five main types of empirical studies [80]:

1. Health impacts of individual extreme events (e.g., heat waves/extreme cold, floods, storms, droughts);
2. Spatial studies where climate is an explanatory variable in the distribution of the disease or the disease vector;
3. Temporal studies assessing the health effects of inter-annual climate variability, of short-term (daily, weekly) changes in temperature or rainfall, and of longer-term (decadal) changes in the context of detecting early effects of climate change;
4. Experimental laboratory and field studies of vector, pathogen, or plant (allergen) biology; and
5. Intervention studies that investigate the effectiveness of public-health measures to protect people from climate hazards.

These studies, and many others have demonstrated practically the need to deduce potential health effects from current and past climate and air quality variability

[81]. Thus the dynamic states of the weather and air pollutants, which have demonstrated some effect(s) on asthma and its severity in the past, is useful in predicting future occurrences of the condition. Probabilistic processes taking into account individual and collective effect of very specific modifiers do provide some opportunities for prediction [82, 83].

Asthma and Weather

In 1995 the Intergovernmental Panel on Climate Change (IPCC) formally reported on the likely health effects of the rapidly changing climate and environment [84]. It was anticipated that the expected health risks from the changing global climate systems would transcend all national boundaries and hence affect large scales of populations. The indirect effects of these changes on the entire ecosystem and their resultant effects on diseases, which impact populations, is a matter of great Public Health concern. Forecasting these risks is complex and uncertain, but also requires specific data on a very long-term basis [85].

There is ample evidence on the effect of temperature changes, barometric pressure and relative humidity on the worsening of asthmatic symptoms [86-93]. Many of these studies have used the association of weather components with a range of disease incidence, hospitalization or mortality (of medical data) to examine their effects on the health condition. For instance the effect of temperature was observed with GP consultations for respiratory diseases, and it was evident that there could be up to 15 days delayed effect of cold temperatures on the incidence respiratory illness [94]. Also a constant seasonal variability in asthma admissions among children was found in Athens (Greece), where relative humidity and atmospheric pressure were established as key determinant meteorological factors [95]. This supports the fact that environmental drivers of asthma events could be effective in predetermining the occurrence or severity of the condition.

It has been well established that there is a relationship between weather related factors and asthma events, and this is not a source of debate [56]. It is also well established that the relationship is affected in complex ways by changes in air quality and season, but again worth noting that these seasonal effects have different associated conditions depending on the location. Asthma events in Mexico, for instance, are associated with the rainy season, whilst in England and Wales asthma events are more strongly associated with seasonal change rather than rainfall [90, 93]. Furthermore, it has been observed in the UK and Taiwan that peaks in asthma events occur in the winter /autumn seasons but not in summer [94, 96, 97]. Under these circumstances, it becomes critical to understand the local relationships between asthma, weather, air quality, and season. However, even when local relationships are well understood, it remains difficult to predict extreme asthma events; that is, unusual peaks in asthma events that fall outside the usual fluctuations associated with seasonal changes as well as variations in weather and air quality.

Asthma and Air Quality

The evidence in literature for pollution related health events, particularly for respiratory ailments like asthma is considerable. The association between asthma events and notable air pollutants like nitrogen (iv) oxide, particulate matter, ozone, sulphur dioxide, smoke as well as household or natural environmental allergens is well known [44, 45, 57, 59, 98-100]. For instance, it has been observed that

susceptible individuals in particular get "asthma exacerbations" more frequently when exposed to pollutants than, would happen in initiating allergies among non-susceptible individuals [101-103]. The known relationships between asthma and air quality, however, have not been successfully used to forecast or predict asthma events. This has largely been due to the difficulty in estimating very randomly distributed/dispersed air pollutants whose original sources (of pollution) are unknown [104, 105].

Individual air pollutant effects on asthma are better understood than their collective effect. The interaction between air pollutants and other environmental factors further complicates their likely effect on asthma, and hence less understood. This complex situation makes the prediction and forecasting of asthma using air pollution information even more difficult. As a result of the shortfall in understanding the complexity of pollutants, the idea of associating "increased air pollution" to asthma /allergic symptoms of asthma has been criticized by few studies [106-108], even though some others do agree [109-118].

Modelling and Forecasting (Environmental Determinants of Health)

Forecasting is an estimation process. It relies on modelling relationships in historical data to predict the likelihood and/or magnitude of a future unknown event. The scientific study and reporting of forecasting has been documented in literature dating as far back as the 1930s [119]. It is most strongly associated with the economic and financial literature; however, forecasting is an important policy tool in many areas. Notwithstanding the potential of forecasting, its absence from the health literature is conspicuous. This is particularly surprising given the impact that forecastable environmental changes can have on health.

There are several methods that could be used in developing forecasts and are described as either subjective or objective [120]. The main subjective method is judgemental, whilst the objective methods include time series, causal or multivariate models (including econometric methods) as well as various simulation or probabilistic procedures[121]. The strengths and weakness of these methods have been discussed elsewhere [122, 123]. Forecasting has often been done using analogies, and more recently, structured judgemental procedure where experts list analogies, rate their similarity to the target, and match outcomes with possible target outcomes [124]. A diagrammatic presentation of a forecasting "Selection Tree" is presented in Appendix A, [125].

As discussed in earlier sections, environmental factors (weather and air quality) have significantly affected health conditions. There are several time series studies that examined the distributions of these environmental factors, and a majority of these studies have subsequently described the patterns or trends of decomposition of the series in relation to the patterns of diseases (e.g. respiratory illnesses). For example, some recent studies reported significant health effects of air pollution at various levels, with some remarkable epidemiological models demonstrating consistent associations between respiratory illnesses and exposure to increased concentrations of airborne pollutants [110, 126-128]. These models have contributed in explaining the relationships between the environmental dynamics and the changing patterns of respiratory illnesses within various time frames and geographical areas. The models have equally generated some debates and questioning regarding the robustness of their methodologies, some inherent

inconsistencies in their findings [129]. Meanwhile the observed variations and some inconsistencies in the relationship(s) between weather/air quality factors and diseases have been attributed to poorly defined diagnosis of some particular conditions [58], or as a result of some location characteristics [128].

Very significant developments in air pollution epidemiology has come from multi-centre studies such as the "Air Pollution and Health: A European Approach" project - APHEA-2 [126-128] and the National Morbidity, Mortality and Air Pollution Study – NMMAPS [130] both of which addressed the challenging issues in their study designs. The details of these major studies have been discussed in the scientific literature [126, 127, 129, 131], and is worth noting that these methods have been revaluated using some common protocols which sought to identify possible heterogeneity in the findings of other previous multi-city studies carried out [132]. Even though much is said about the relationships, between environmental/pollution factors and disease occurrence, this has largely been limited to the decomposition of trends over time periods. There has not been sufficient work on forecasting diseases particularly in relation to hospital admissions episodes.

Methodological Challenges in Forecasting

Health forecasting is known to present very complex methodological challenges, given the nature of interdependent determinants of ill health. There are obviously many exposures or risk factors with different dimensions of complexity as well as complex human behaviors to consider in a health forecast model. Thus even for a particular health condition, it is quite difficult to pinpoint its cause(s).

It is difficult to adequately collect data on both exposures and health outcomes in suitable locations [133, 134], or to find historical data of the sort that covers a wide range of time, suitable for forecasting disease episodes. Thus when faced with data insufficiency with regards to the precision and complexity in the distribution of individual indicators, forecasting a health condition can further be limited by the choice of potential analytical methods. Also the challenge of aggregating very widely dispersed indicators across wide geographical areas and times (i.e. a source of variation); could potentially hinder the prediction of disease episodes. Data and methodological shortfalls pose a serious challenge to health forecasting and can thus lead to substantial deviation or worse still, an inability to predict extreme events.

Health Forecasting and Commerce

Modelling and forecasting generally follow approaches that require some historical data (time series data in particular). The time trends of these data are then used to project future events. Even though the reported innovations in forecasting are strongly associated with finance and econometrics [122], their methods can be adapted in health forecasting. However, Health forecasting or weather /air pollution forecasting each bare some distinctive characteristics. Health forecasting assumes the conceptual perspective of disease causality, in the modeling and prediction of the outcome [135]. Though it also largely depends on historical data, the results of a health forecast are not usually shaped by recent events (except for catastrophic events), but by long term trends in the dataset. Health forecast provides a better understanding of the occurrence of a disease condition, and uses predictions to ease any potential extreme/unexpected risks [136].

Commercialization of health forecast is relatively new, but well practiced by few organizations including the Met Office in Exeter. In the next section, the approach to health forecasting is briefly described considering the Met Office example.

The Met Office Scenario: Approaches to Developing a Health Forecast

The concept of Health Forecasting from a commercial perspective involves turning simple scientific concepts /ideas related to a health condition into business solutions for the well being of individuals.

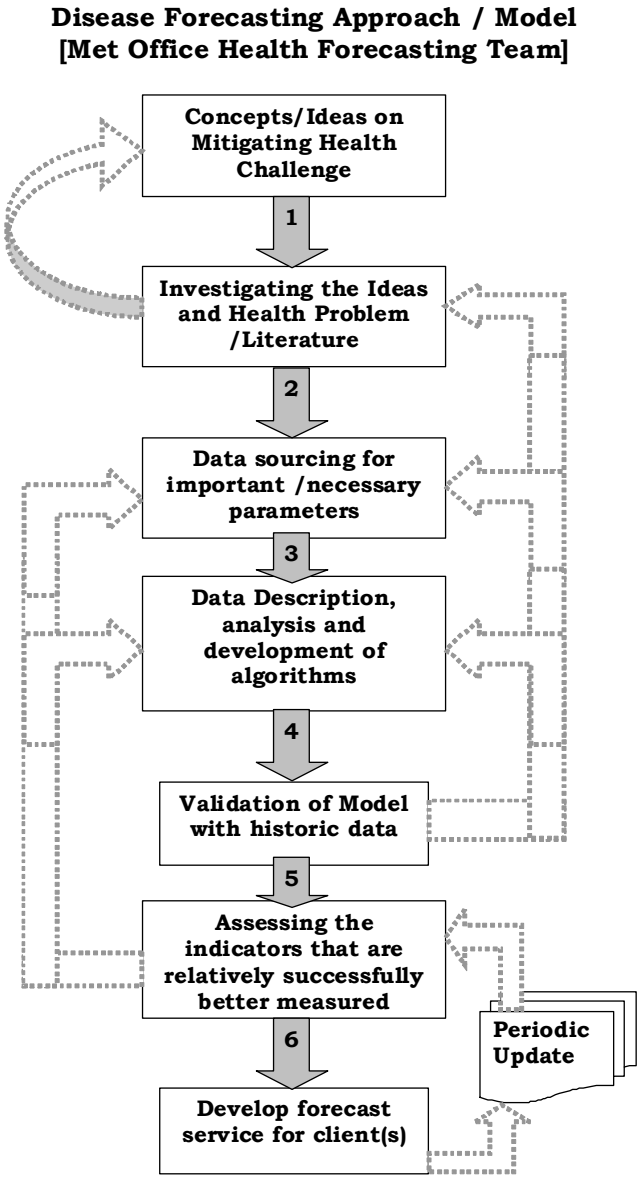
In the process of developing a health forecast, the negative impact of conditions surrounding the operational activities of potential clients is first observed and analyzed to establish the magnitude of the problem. A typical example is the forecast for COPD in the UK. It was observed that symptoms of COPD sufferers get worse during colder weather, and hence exacerbate the condition [137]. The environmental factors which could be directly related to such a health problem (i.e. the different measures of cold temperature) are related to the condition. A forecast is then developed using the relations. This forecast constitutes an environmental solution aimed at minimizing the disease burden, and hence is piloted and then repackaged to suit specific clients.

The process of developing a health forecast service can be summarized in the following steps:

1. Concepts and ideas that address an important health condition of great burden and /causing significant financial cost to the health service are identified.
2. The true impact of this health condition is assessed from records /literature. The ideas are also investigated from the literature and hence a justification of the exercise or study is reached.
3. Data is then sourced for most of the important or necessary parameters shown to be of essence in the scientific literature.
4. The datasets are then prepared for analysis; and these include basic descriptive patterns as well as the development of algorithms
5. The Model obtained is then validated using similar historical data
6. The final list of parameters used in this model is again evaluated for the relative ease of their measure and predictability.
7. Very specific and tailor-made disease forecast services are then developed for the client.

The procedure outlined above is illustrated in Figure 2 below. For the Met Office, this is an important activity, which has corporate management and support. There is also a Health Forecast Team directly responsible for all its activities in health forecasting ranging from scientific research to commerce and clients.

Figure 2 The Met Office Health Forecasting Team Disease Model



Aims and Objectives of study

The Aims of the study

The study on asthma and the environment was commissioned by the Met Office Health Forecast Team as part of the preliminary development of tools which could be used to forecast asthma hospital admissions. The aim of the study was to investigate the extent to which meteorological and air quality factors could be used to forecast hospital admissions. There was a particular interest in forecasting extreme levels of asthma admissions; that is, peaks in admission well above the expected value given seasonal variation. In view of the highly contextual nature of the interaction between weather and air quality to affect asthma events, this preliminary investigation focussed on asthma admissions in the London area (i.e., asthma admissions occurring in the area circumscribed by the M25 motorway).

General Objectives of the study

The general objectives of this study were to:

- Merge independent hospital admissions, weather, and air quality datasets;
- Ensure that the merged dataset was suitable for time series analyses and included appropriate dummy variables to study unmeasured effects of month, seasonality, etc.;
- Undertake a descriptive analysis of key variables;
- Undertake time series modelling, focusing on Generalised Linear regression Models (GLM) including the Poisson and Negative binomial regression count models, and Logistic Regression Models (LRM) as well as Quantile Regression Models (QRM); and
- Validate predictive models, using conventional clinical statistical tests and suggested error measurements.

Specific Activities

- We examined the data type and distribution of all continuous variables as well as their probability plots for normality and missing values. Ordinary frequency tables were examined for nominal qualitative or categorical variables.
- Time series datasets were generated by collapsing data records of individual hospital admissions into records of daily counts of hospital admissions. Independent weather and air quality data collected from multiple stations were combined to produce average daily weather and air quality.
- Descriptive analyses were conducted to identify longitudinal trends in the continuous weather and AQ data and in the distribution of daily hospital admissions.
- A pool of target independent variables was developed using Negative Binomial Regression to explore the individual relationship between each of the available weather and AQ data and hospital admissions. The same procedure was used to establish suitable time-series lags that might be required for subsequent multivariate modelling.
- Preliminary multivariate analyses were then conducted using Negative Binomial Regression (NegBin) to pre-select temperature and air quality variables best suited for modelling, since the NegBin regression model was more suitable compared to the Poisson.

- Three separate methodologies to predict daily asthma hospital admissions were contrasted. Two of the modelling techniques (Poisson and Logistic regression) model expected values. Quantile regression can be used to model expected values (50^{th} percentile), as well as values at any arbitrary percentile, and can be used to model extreme events. Standard model tests (Link test, HL Goodness of fit) were used to assess the fitness of the models.
- We applied a simple conventional clinical statistical test to determine the predictive value of each model. The reduction in Root Mean Square Error was used in the cross validation of each predictive model.

Rationale for the study

The global rise in the incidence and/ prevalence of asthma is well known [17, 20-23, 29-31, 33, 138-141]. Even though there are notable differences in the diagnoses of asthma regionally [142], it is still notable that the number of susceptible individuals have increased globally. This has subsequently led to an increased demand for limited health care resources like hospital beds.

Studies that have reported on the likely causes or determinants of asthma globally have associated asthma to atopic factors [143], ethnic and domestic factors [11-14, 144-147], allergens and allergies [3-7] as well as the weather and air quality factors [2, 44, 45, 57, 59, 98-103, 148]. In the United Kingdom, the disease burden of asthma and its management disproportionately affects the vulnerable populations and demographics [144, 149].

The effects of the weather and air quality factors on asthma have been extensively discussed but the exact mechanism and role they play to either cause or exacerbate the condition is not clearly understood. Another conspicuous gap in the literature is the forecast of the disease episodes using these environmental components. There isn't sufficient information to guide health care/service providers on the likely occurrence of asthma events; be they normal or extreme/unexpected events. Some commissioned studies (by the Department of Health) on air pollution and asthma in the UK recommends among other things that further investigation on the epidemiology of asthma in relation to air pollution and weather, be conducted to close the knowledge gap [2].

This study is thus important because it provides some tools for developing the asthma forecast using the weather and air quality factors. The research carefully describes suggested methodologies that could be used in creating or selecting suitable predictors for asthma hospital admissions based on weather and air quality datasets as well as asthma hospital records. Furthermore, innovative ways of testing and validating forecast models have been provided. This study provides information that may be useful to health service providers as well as policy makers to optimize services for asthma patients.

METHODS

The effect of meteorological factors on hospital admission for asthma was investigated. We specifically sought to develop tools for predicting and forecasting hospital admissions for asthma. Our key questions related to the feasibility of forecasting asthma based admissions on weather and air quality factors. The motivating questions were:

- To what extent can meteorological and air quality factors be used in forecasting asthma hospital admissions? and
- Can daily hospital admissions for normal and extreme asthma episodes be predicted from the changes in the meteorological and air quality mediating factors?

We examine three methodologies for predicting and forecasting asthma hospital admissions for normal and extreme events, and evaluate the fit and predictive capacity of each method.

Literature

The review of literature on this section was limited to the extent that it informed the key factors (i.e. the daily hospital admissions, daily weather and air quality indicators) under investigation in the dataset. The nature and distribution of these measures guided our selection of the literature, in support of the statistical methods used. Our key dependent variable (daily hospital admissions) consisted of count data entries, which were not normally distributed. The other independent variables were quite varied in both their data types and distributions. These variables are discussed in detail below.

Generalised Linear Models (GLM), specifically count models (poisson and negative binomial regression), and Quantile Regression Models (QRM) were used in different approaches to modelling the predictors of asthma daily admissions. These methods and their reference literature are subsequently discussed.

Data Sources

Three independent datasets were sourced. These included asthma hospital admission data from the Hospital Episode Statistics database sourced through the Met Office; Weather data and estimates of Air Pollutants. Even though these datasets each presented a collection of indicators with a corresponding time (date) record, they were not organised in the usual format of time-series data. Hence we generated independent time-series datasets for each of the three sets mentioned above by summarizing and ordering the daily records. These three new time-series datasets were then combined into one dataset and used in the development of the forecasting models. The Asthma admissions, Weather, and Air quality data sets are each discussed in turn.

Asthma Admissions

The Hospital Episode Statistics (HES) is a record-level data warehouse managed by the NHS Information Centre for Health and Social Care Data. The data included a record of all asthma admission in hospitals within London from January 1, 2005

to December 31, 2006[1]. Our operational definition for *Asthma Admission* was any diagnosis with a primary diagnostic ICD-10 code. Data from the HES are extracts from routine data flows exchanged between healthcare providers and commissioners via the Secondary Uses Service[150]. The data entry and quality checks involved have been described elsewhere[150]. Asthma hospital admission in this dataset was indicated by a unique variable, which contains the "anonymised" personal identity of the individual hospitalized.

The strengths and weaknesses of the HES data source and similar hospital admissions /episode statistics have been discussed extensively in the literature; in particular issues have been highlighted regarding the compilation and purpose of the dataset [150-157]. The HES data is known to have some shortfalls in maternity and psychiatric data for example, but again provides a tool to handle any lapses. This tool is the Data Quality Indicator (DQI)[2] which enables both users and providers of HES data to analyse the data quality at the level of the NHS Trust[150]. Even though DQI reports on asthma for our dataset were not available, an issue of generic concern in dealing with data on the morbidity has always been the difficulty associated with its diagnosis. Nonetheless, it is no reason for withholding the data on asthma hospital admissions, or going ahead to using it in developing tools for forecasting.

Mindful of these discussions and the nature of some inherent deficiencies in the dataset on asthma hospital admissions, we proceeded to examine dataset and explore the possible associations between asthma admissions as a key dependent variable and other independent factors within other datasets. This data assessment was focussed on predicting and forecasting asthma admissions.

We examined the probability plots (normal distribution) of the key dependent variable (Asthma hospital admissions) and other selected variables in the HES dataset, and checked for outliers as well as the proportion of missing entries. A comprehensive list of the variables in the HES dataset are listed in the attached codebook (Appendix C) and fully described elsewhere[153]. Individuals (patients) were identified by a unique ID number[3] across all data years [150]. This unique ID number did not identify any individual's number of visits within the study time frame and hence we treated each hospital visit on any occasion as unique.

Climate 2005 & 2006

The daily Meteorological factors[4] for all weather monitoring sites (and their respective postcode areas in London), as well as their location coordinates and altitudes were sourced through the UK Met Office database (Methodology for data

[1] The HES data was procured by the Met Office Health Forecasting Team

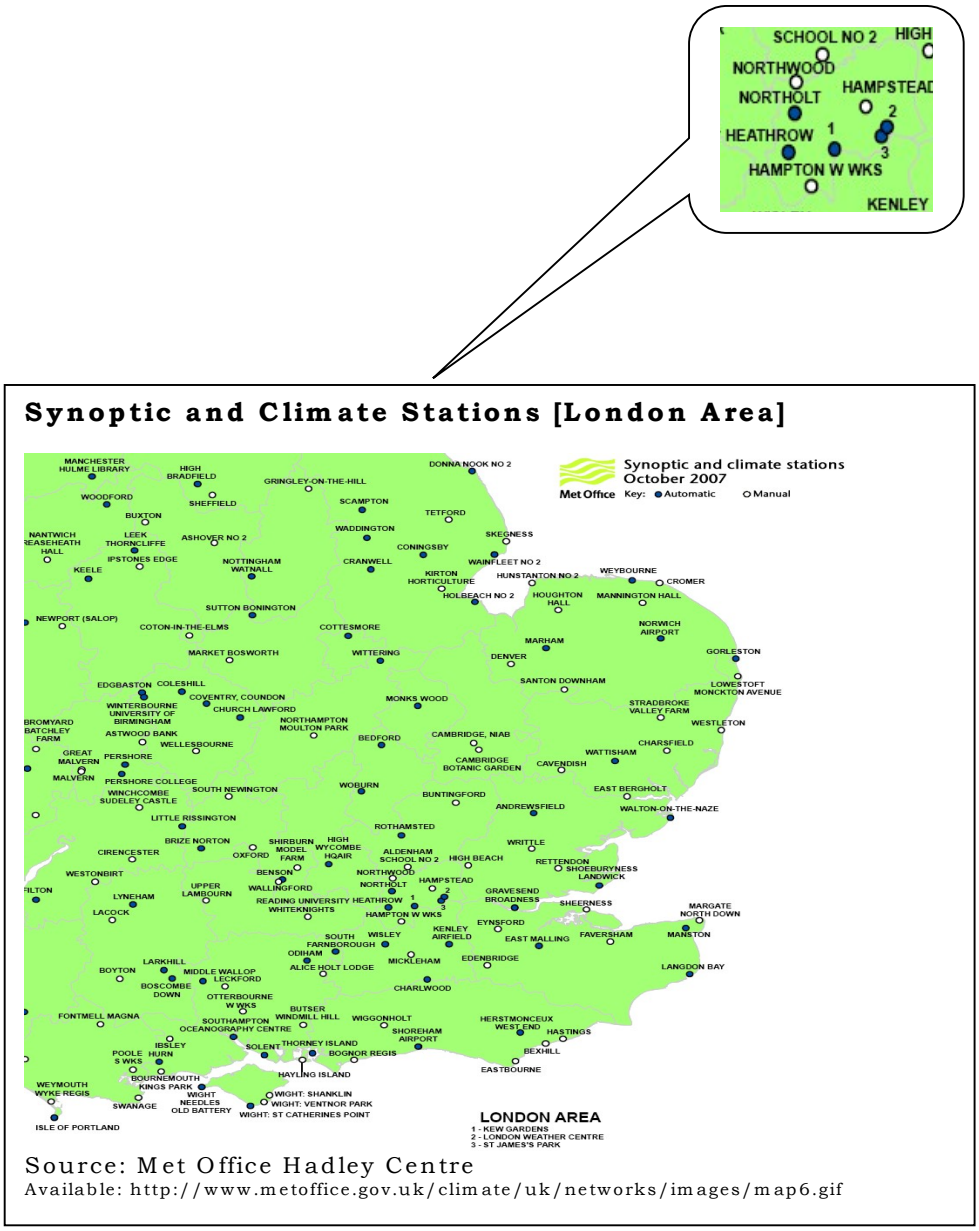
[2] The Data Quality Indicator provide a summary of HES data quality, and should identify issues that need to be addressed by data providers, and taken account of by analysts

[3] It is generated by matching records for the same patient using a combination of NHS Number and local patient identifier, plus the patients' postcode, sex and date of birth; but maintaining anonymity.

[4] Meteorological factors: Maximum Temperature (deg C); Minimum Temperature (deg C); Night Minimum Temperature (deg C); Night Maximum Temperature (deg C); Day Maximum Temperature (deg C); Day Minimum Temperature (deg C); Mean Wind Speed (m/s); Wind direction; Ambient Air Temperature (deg C); Wet Bulb Temperature (deg C); Dew Point Temperatures (deg C); Vapour Pressure (HPa); Humidity (%);

collection is described elsewhere[1]). This dataset was matched to all postcodes by their respective nearest (distance) monitoring station postcode.

Figure 3 The synoptic and climate stations within the London area



Source: Met Office Hadley Centre
Available: http://www.metoffice.gov.uk/climate/uk/networks/images/map6.gif

The key meteorological indicators in this dataset were: *Maximum Temperature, Minimum Temperature, Night Minimum Temperature, Night Maximum Temperature, Day Maximum Temperature, Day Minimum Temperature, Mean Wind Speed, Ambient Air Temperature, Wet Bulb Temperature, Dew Point Temperatures, Vapour Pressure, and Humidity.* These were presented as daily records. The weather stations (Fig 3) in the UK as a whole report a mixture of snapshot hourly observations of the weather condition and this is known /referred to as synoptic

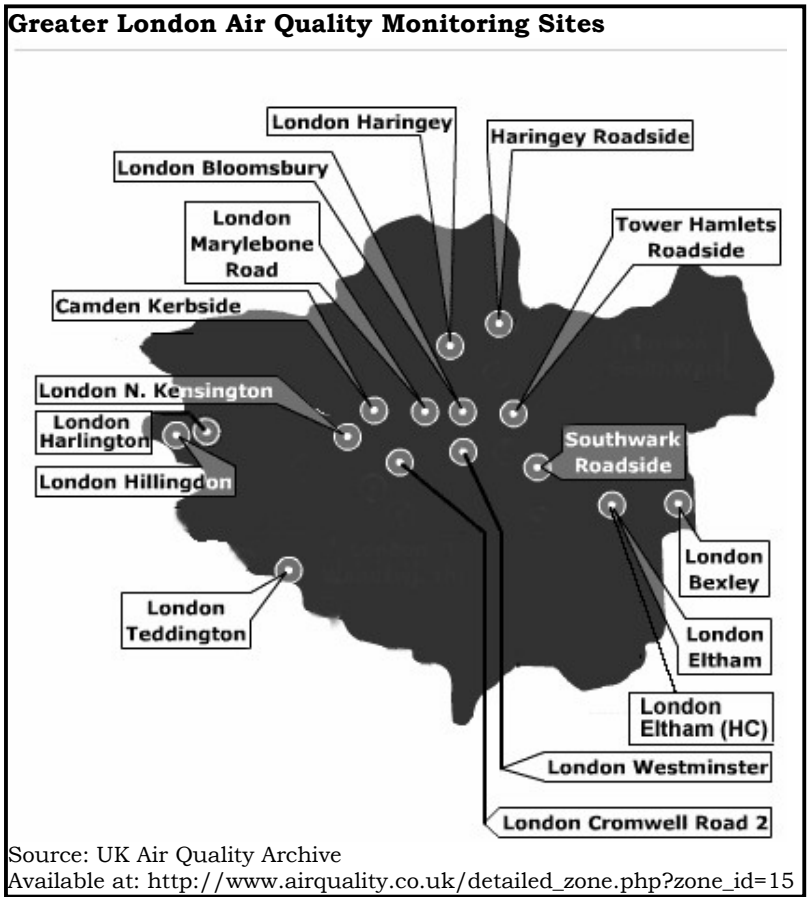
[1] http://www.metoffice.gov.uk; http://badc.nerc.ac.uk/home

observations. Also the daily summaries of the weather measures are however known /referred to as climate observations[158]. The detailed description of individual weather elements and how they were quantified over the period as been described [159].

Air Quality Estimates

Air quality is monitored across the UK through a variety of sites at strategic locations which continuously capture ambient air quality levels for selected pollutants. In London, about 16 functional sites located across the region (Fig. below) provide air quality measures for various pollutants. The UK Air Quality Data Archive provides this information, and details on the location and characteristic nature of each site as well as the measures they provide (http://www.airquality.co.uk/detailed_zone.php?zone_id=15). This additional information also includes a description of the mode and frequency of quantification of all the respective pollutant measures.

Figure 4 The Greater London air quality monitoring sites



Source: UK Air Quality Archive
Available at: http://www.airquality.co.uk/detailed_zone.php?zone_id=15

In this study we had access to two air quality datasets, and these were provided as:
(1) Daily values derived from the Air Quality Archive (AURN) in situ measurements for 2001-2006, matched to postcode districts by closeness, up to 50 km;
(2) Daily values from the Met Office NAME atmospheric dispersion model for postcode districts for 2005-2006 [160].

Our analysis was based on the second dataset, from the NAME which consisted of daily estimates of Carbon monoxide, Formaldehyde (HCHO), Nitrogen dioxide, Nitrogen oxide, Ozone, Sulphur dioxide, and Particulate matter (pm10). These modelled daily air quality estimates[1] for respective postcode areas also do account for both accident and episode analysis, and as well used for pollution forecasting [160].

Additional Predictor Variables

We generated additional potential predictor variables to account for monthly and seasonal variations as well as the rate of temperature drop. The monthly variation indicator was generated form the date variable whilst the seasonal variable was created by categorizing the days of the year into the four known astronomical seasons (spring summer autumn winter) [161]. We created variables to represent the rate of temperature drop by evaluating the temperature differences (i.e. for day, night and maximum/minimum daily temperatures).

Creating the Time Series

The use of the term *time series*, in this investigation, refers to a sequence of observations that are ordered in time. Three time series datasets were generated from the three datasets described above. For each original dataset, the corresponding time series dataset generated consisted of a *time* variable, which was designated by the date as well as independent variables representing daily averages of various measuring stations.

Time Series for Asthma Hospital Admissions

London is by far the most densely populated regions of England. The impact of various human and natural activities on the environment and /health is considerably high and hence an issue of great public health importance. We restricted the analysis to the available dataset, which recorded events in London from January 1, 2005 to December 31, 2006.

A new time-series variable was created from the HES data. This was a count of all unique, hospital admissions with a primary diagnosis of asthma for each day from January 1 2005 to December 31 2006 to hospital in London. This was done by collapsing the total number of daily admissions (of all individuals) for asthma. Hence the new time series dataset from this manipulation mainly consisted of the *time* indicator and *new asthma* variables for the period described above.

Time Series for Climate

The Weather dataset sourced from the Met Office database consisted of several indicators recorded daily over the period (January 1, 2005 to December 31, 2006). These indicators included:

Maximum Temperature (degrees Celsius),
Minimum Temperature (degrees Celsius),
Night Minimum Temperature (degrees Celsius),

[1] Daily mean values of Carbon monoxide; Nitrogen Dioxide; Nitrogen oxide; Ozone; Particulate matter10; Formaldehyde and Sulphur dioxide in SI units; Estimates used by UK Met Office in accordance with the National Statistics Code of Practice (See appendix for further information on NAME)

Night Maximum Temperature (degrees Celsius),
Day Maximum Temperature (degrees Celsius),
Day Minimum Temperature (degrees Celsius),
Mean Wind Speed (knots)
Ambient Air Temperature (degrees Celsius),
Wet Bulb Temperature (degrees Celsius),
Dew Point Temperatures (degrees Celsius),
Barometric Vapour Pressure (hectoPascals),
Humidity (%),

Even though the weather dataset included a total of eight weather stations, only five were actually within the London[1] area and thus representative.

The weather data from each of the five selected weather stations in London (Heathrow, High Wycombe, London Weather Centre, Northolt and South Farnborough) were averaged, to produce a single daily summary of London weather. In the analysis of time series data there is often a trade-off between creating usable data sets and information loss. Averaging data as described here necessarily results in a loss of information. The strategy, however, was supported by a preliminary analysis, in which generally high correlations in the weather indicators data was observed between the different weather stations. There is precedent for this kind of approach [e.g., [162]].

Time Series for Air Quality

Air Quality data was obtained from the Met Office's Numerical Atmospheric-dispersion Modelling Environment (NAME) database, which accounts for both accident and episode analysis, and also used for pollution forecasting [160]. The listed indicators included:

Carbon monoxide (kgm^{-3}),
Formaldehyde (kgm^{-3}),
Nitrogen dioxide (kgm^{-3}),
Nitrogen oxide (kgm^{-3}),
Ozone (kgm^{-3}),
Particulate Matter [PM10] (kgm^{-3}) and
Sulphur dioxide (kgm^{-3})

Some air quality readings, such as PM2.5, black smoke etc, that are known to be causally related to asthma events were not available in this dataset. Preliminary investigations carried out to examine the relationship between the patterns of distributions of air quality measures across several stations in London showed very wide variations. These variations largely reflected the different types and locations of the measuring stations. *Urban Background* measures are known to account for urban locations that are distanced from potential sources of direct emissions (pollutants), and therefore broadly representative of city-wide background conditions [163]. Hence we sought to use only recognised "*Urban Background*" measuring stations for the purpose of comparing sites and generalising some area measures.

1 See appendix for "London Weather Stations Records"

In the London Air Quality dataset, a process, which converts the dataset in memory into a dataset of means or sums and described by the Stata statistical package as *collapse,* was used in producing our time series dataset. We collapsed daily average values and subsequently generated a new dataset, which included daily average air quality measures for all the representative areas. Hence we generate a new representative air quality variable for the entire London region.

Combined Time Series Dataset

The final time series data set on which all analyses were based comprised the count of daily asthma admissions (dependent variable), the averaged daily weather data, and the averaged daily air quality data. This represented a complete dataset with no missing data for any day between 1 January 2005 and 31 December 2006. The combined time series dataset was created by merging the three separate time series datasets described earlier. Additional dummy variables, which were generated to account for seasonality, rate of daily temperature change, monthly effects as well as other categorical variables formed part of this combined dataset. Although the approach taken here to the data analysis is not entirely for the strict purpose of hypothesis testing, nor is it strictly parameter estimation, we nonetheless use the general convention of referring to the count of asthma admissions as the dependent variable; all other variables collectively are called the independent variables.

Data Analysis

Descriptive Analysis

General Distribution of Parameters, Summary Statistics by Categories

The distributions of probability plots of the continuous variables were examined. Summary statistics were also calculated by suitably categorizing each indicator based on either known recommended categories, percentiles or a suitable fit.

We examined the probability plots (normal distribution) of the key dependent variable (i.e. Asthma hospital admissions) and other selected variables in the HES dataset, and checked for outliers. The same analysis was done for the other datasets (Weather and Air Quality). We outlined variables according to the scales of data type[1] and on this basis identified the appropriate bivariate statistical tools [164, 165] for preliminary descriptive analysis.

Distribution of meteorological indicators across measuring stations

The patterns and correlations in the distribution of meteorological and air quality indicators across measuring stations in London were observed by scatter plots and correlation matrices. Thus we examined the individual relationships between the dependent variable and each of the independent variable (graphical).

Basic associations for daily Asthma hospital admissions

Basic associations for daily Asthma hospital admissions with meteorological and other indicators were presented in categories. This was done using cross tabulation

[1] see appendix on "*Data Types and Methods*"

of basic frequencies suitably categorized. Also separate binomial models for each exposure variable (uncategorized) were used for bivariate comparisons.

Distributions of Asthma Admissions by Demographics & Spell related factors

We examined the basic associations between individual daily hospital admissions for asthma and the categories of demographic factors using cross tabulations and/ correlations. The same statistical analyses were conducted for categories of asthma spell related factors (time and duration of hospitalization, related secondary diagnosis, type of health facility, etc.)

The Lag Effects of Exposures

In order to understand the nature of the compensatory period required to fully experience the cumulative effect of an exposure we compared the lag properties of independent variables in the dataset. We followed an approach to generate single lag models that provide estimates for the effect of a unit increase in an exposure over a single day. Hence the lag properties of individual meteorological and air quality factors were explored for modelling asthma daily hospital admissions in London for up to a 21-day lag period. The output of each of these bivariate tests for meteorological and air quality factors were further examined and the suitable lags selected based on their relative *p-values* and regression coefficients. Hence for each parameter (e.g. regression coefficients), the general distribution was observed and the peak estimate selected. Other parameters (like the *Alpha, Loglikelihood and PseudoR2*) were only used for the purpose of comparing lags for each individual variable.

The process of generating, testing and selecting lags is necessary because it provides an additional resource for the cross validation stage for the dataset, which we will describe shortly.

Time Series Forecast Models

The forecasting of daily asthma hospitalisations in London given meteorological and air quality factors was investigated using generalised linear modelling (GLM) techniques[166], and quantile regression [167, 168]. The GLM techniques include a range of statistical linear models, which have non-normal probability distributions, such as the Poisson, Binomial, Multinomial, etc. These set of models easily fit the data because they do not usually require the variance to be constant / equal to the mean, in hypothesis testing. The selected GLM techniques used as well as the general approach to health forecasting are discussed shortly.

Forecasting

Forecasting, as opposed to traditional hypothesis testing and causal analysis, is principally concerned with the prediction of future events, rather than explaining the relationships between variables. This is a distinctly instrumental approach to data, which in the domain of health is usually directed towards practical outcomes such as early warning for peaks in service demand. In the selection of a good model, some important criteria and tests that have often been referred to and used include attributes like the predictive power, theoretical consistency, goodness of fit, "identifiability" and parsimony [169]. However causal modelling is not at issue, concerns with confounding, or the perfect choice of data model or parameterisation of variables can be relaxed. The value of a forecasting model is based on (a) its

predictive rather than its explanatory power, and (b) the simplicity/cost of its implementation.

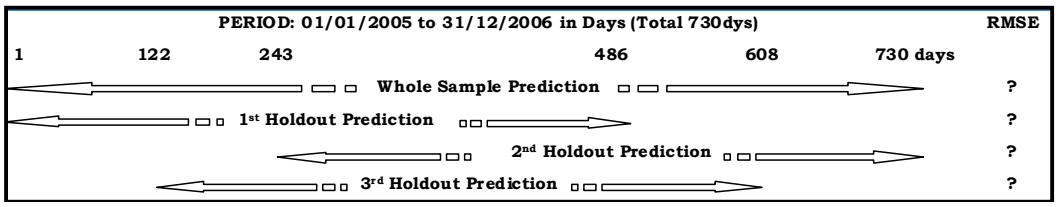
Predictive power of a forecast model is related to the forecast error, a measure of the difference between the actual value and the forecast value for a corresponding period [170]. Forecast error can be estimated by a number of methods[1] including the Root Mean Squared Error (RMSE). The RMSE is the square root of the average squared error. That is, the squared difference between each actual and predicted value is calculated; the square root of the mean of the squared differences is then calculated. RMSE estimates the accuracy of a single predictive measure. Equation (1) below illustrates the estimation of RMSE.

$$\mathbf{RMSE} = \{(\sum^{N}_{t=1}\mathbf{E}^2_t)/\mathbf{N}\}^{1/2}. \quad . \quad . \quad . \quad . \quad \textit{(1)}$$

$$E_t = Y_t - F_t$$

Where: E is the forecast error in a given period of time t
Y is the actual value at a given period of time t
F is the forecast value at a given period of time t
N is the number of observations

The suitability of the predictive model generated for asthma hospital admissions was tested using a hold-out sample of the available dataset. We predicted asthma daily hospital admissions for London from the dataset holding out a third for cross validation. Subsequently, we cross-validated the predicted model with a second holdout sample of similar proportion of observations. Thus the errors in these predictions as determined by the means were used in the estimation of RMSE. The figure bellow illustrates this process



Forecasting asthma (hospital admission) events, however, is not simply a question of estimating a daily figure. It is also potentially a matter of alerting services about days of peak demand. In this case one is making a binary forecast: a day of peak demand or a day of normal demand. The value of this approach to forecasting can be examined with a traditional analysis of clinical -test accuracy; that is, the positive (normal) and negative (extreme) predictive value of the test. Predictive values, sensitivity and specificity tests have been used extensively in many different ways to assess the accuracy of determining an event [26, 171].

Taking into consideration the actual and predicted distributions of asthma admission, we dichotomized both indicators. This was done with the assumption that the upper 10% of the distribution of actual asthma hospital admissions were extreme events. Hence our cut-off for dichotomizing both the predicted and actual asthma admissions was the corresponding entry of daily admissions at the 90th

[1] Mean squared error (MSE), Percent mean absolute deviation (PMAD), Mean absolute percentage error (MAPE), Forecast skill, Mean absolute error and the Root Mean Squared Error (RMSE)

percentile. We then cross tabulated the two categorized indicators to obtain the predictive validities. This test illustrates the positive (normal) and negative (extreme) predictive values for asthma admissions in the given model, as illustrated in the matrix below.

Table 1 Matrix for estimating the predictive values

		Actual Asthma Admissions		**Predictive Value**
		<40	**≥40**	
Predicted Asthma Admission	**<40**	**A**	**B**	**Normal = A/[A+B]**
	≥40	**C**	**D**	**Extreme = D/[D+C]**
		Sensitivity = A/[A+C]	**Specificity = D/[D+B]**	

A, B, C&D are estimates of the number of daily hospital admissions based on the cross tabulation of the dichotomised actual and predicted models.

The following parameters (α, β and likelihood ratios) were then estimated from the above matrix as part of the diagnostic test to provide further information on the probability that the test will give the correct diagnosis [172-174]:

False Normal rate (α) = 1-Specificity = B/ [B+D]
False Extreme rate (β) = 1-Sensitivity = C/ [C+A]
Normal Likelihood-ratio = Sensitivity / [1-Specificity]
Extreme Likelihood-ratio = [1-Sensitivity] / Specificity
NB: Power = Sensitivity = 1-β

Generalised Linear Models with Count Models

Poisson regression and negative binomial regression were used as the techniques of choice for modelling the asthma (hospital admissions) events data. Poisson regression is well suited to the modelling of count data, and one of the most common techniques used for modelling asthma events [166, 175-177]. However, in causal modelling and hypothesis testing, it is not suitable when there is over-dispersion in the data - that is when the variance exceeds the rate of daily asthma events. In these circumstances, negative binomial regression is the preferred modelling technique. This is discussed in greater detail shortly.

The Statistical analysis conducted thus involved regression analysis of the daily time-series data in which we evaluated the likely associations between the probability of asthma-related hospitalisation and the various exposure variables. We explored the short- and relatively longer-term effect(s) of these exposure variables using their lags. Core adjusted models involving ALL the likely covariates with significance at 95% CI were thus generated.

Count Models

The total number of daily hospital admissions for asthma was generated from the HES dataset. These episodes were *count* records of non-negative integers ranging

from 6 admissions per day to 130 admissions per day. Considering the entire range of this dependent variable, its distribution was observed to be slightly skewed.

Poisson regression (equation 2) is one of the basic parameterised count models. It predicts the expected number of hospital admissions for asthma assuming that the variance equals the mean ($p^2 = u$) [178]. The predicted rate of daily admissions can estimated as:

$$\mathbf{y = Pr(Y_i = A \mid E) = exp(E\beta)}$$

where the probability of observing a specific count (of total daily hospital admissions for asthma), given y (i.e. the predicted rate of daily admissions) is computed as:

$$\mathbf{Pr(Asthma_i = A \mid y) = e^{-y}y^A / A!\ ,\ A = 0, 1, 2, \quad . \quad . \quad . \ (2)}$$

Where:

Pr *($Asthma_i$) = probability of asthma admission for a given day, i*
E= *given exposure for i*
β = *coefficient of a given exposure measure*
y = *predicted rate of daily admissions*

Assuming there is no over dispersion or under-dispersion (i.e., assuming that p^2=u), the expected value of y is determined by the coefficient of the exposure variables (β). That is, β explains the marginal change in the number of hospital admissions given a one-unit change in the exposure variable.

It is, however, not uncommon to observe over dispersion of the variance (i.e. p^2>u) in these kinds of models. Under these circumstances, the negative binomial regression model is preferred [178-182]. In a negative binomial model, the probability of observing a specific count of asthma events estimated by:

$$\mathbf{Pr(Asthma_i = A \mid x_i) = \{\Gamma(A+\alpha^{-1})/A!\Gamma(\alpha^{-1})\}(\alpha^{-1}/\alpha^{-1}+y)^{\alpha^{-1}}\ (y/[\alpha^{-1}+y])^k, \ldots . (3)}$$
$$\mathbf{A=0, 1, 2 \ldots}$$

Where:

α = *increased variance in the predicted counts*
Γ = *mathematical function "gamma"*
x_i = *independent variables (exposure)*
y = *predicted rate of daily admissions*

In the Negative binomial regression model, the count dependent variable is generated by a Poisson-like process, except there is an additional parameter to account for variation that is greater than in a Poisson model. This additional variation is referred to as over-dispersion. The preference for the *negative binomial model over* poisson *model* is largely determined by the value of the dispersion parameter (α). If α is statistically, significantly greater than zero (α>0) then the negative binomial model is preferred [178, 180]. The post estimation tests (*vuong* test and *robust* options in STATA) provide better estimates of the marginal effects for the standard error terms in the final model [178, 180]. The robust standard errors adjust for the heterogeneity in the model, and also provide better estimation

coefficients for the model. Hence this additional model diagnostic procedure is useful in further identifying and eliminating predictors that are not significant in the model.

Further tests involving the "goodness of fit" and "link test" in STATA, were used for the purpose of checking the model specification. The procedure for the goodness of fit test provides the deviance statistic, which is used in deciding on the preference for the poisson regression or the negative binomial regression. In the case of the latter, the Link test performs a link specification of the model, and this is preceded by the negative binomial regression single-equation estimation. The significance of the Link test is determined by the p-values of both the predicted variable (hat) as well as the square of the predicted variable (hatsq). Hence the hat should be significant since it is the predicted value. Meanwhile, the p-value for the hatsq should not be significant, (since the squared predictions should not have much explanatory power) when the model is specified correctly.

Quantile Regression

Quantile regression technique was introduced by Koenker and Bassett in 1978 as an extension of the linear-regression model. The quantile regression does not assume normality of the dependent variable and it models the conditional quantiles as functions of predictors; specifying changes in any conditional quantile [167, 183]. Unlike the linear-regression, quantile regression models have the ability to characterize the relationship between the dependent variable and the independent variable(s) particularly in the extremes of the distribution.

In theory, the n^{th} quantile of the dependent variable Y is the value, Q(n), for which its given probability is P[Y<Q(n)]=n. This given probability is assumed to have a distribution with corresponding quantile estimates for n, which exclusively range from zero to one (i.e. 0<n<1) [184]. The corresponding quantile regression model which explains the relationship between the dependent variable, Y can then be expressed as

$$\mathbf{Y_i} = \boldsymbol{\beta^{(p)}_0} + \boldsymbol{\beta^{(p)}_1 x_i} + \boldsymbol{\varepsilon^{(p)}_i} \quad . \quad . \quad . \quad . \quad . \quad . \quad \textbf{(4)}$$

Where:

Y_i is Asthma hospital admissions for a given day, i
$\beta^{(p)}_0$ is a constant term
$\beta^{(p)}_1$ is the coefficient of exposure term
x_i is the exposure term
$\varepsilon^{(p)}$ is the error term

Quantile regression techniques have been used for estimating several extreme outcomes and they include modelling the effect of meteorological factors on some environmental pollutants [185], describing the sea level trends at different tides [186], modelling the factors that affect ecological processes [187-189], and even the effect of school quality on student performance [190].

In this study, we used quantile regressions to estimate extreme variations in asthma hospital admissions resulting from the changing patterns of selected meteorological and air quality indicators in London. We compared the default median quantile regression estimates to estimates at other percentiles (0.80, 0.85

quantiles). The optimization process for the QRM was controlled by selecting the best **w**eighted **l**east-**s**quares **i**teration **n**umbers (**wlsin**), which matched the quantile (these were: 278 & 280). The *wlsin* number is the estimator of the response variable (in this case the Daily asthma admissions). The weighted least squares iteration identifies the best estimator for the regression before generating the output. It is obtained by minimizing the sum of the squares of the weighted residuals, taking into account that each residual is weighted by the inverse of the local variance of the response variable. Hence this was done to be able to achieve a convergence in the estimation of the maximum-likelihood model.

The goodness of fit test was used to assess the model fitness [191].

Logistic Regression

Logistic Regression Model (LRM) was used in this case because the outcome variable was dichotomised. In principle the *logit* of a proportion $\boldsymbol{p}$ is the log odds, described in the relation below in equation 5a [165, 166]. Given the assumption that the relationships being investigated are linear on the logistic scale (equation 5b), the effect(s) of the predictor variable(s) are found as log odds ratios in the output.

$$\mathbf{logit}(\boldsymbol{p}) = \log_e (\boldsymbol{p}/1-\boldsymbol{p}) \quad \textit{(5a)}$$

$$\log_e (\boldsymbol{p}/1-\boldsymbol{p}) = \boldsymbol{b}_0 + \boldsymbol{b}_1\boldsymbol{x}_1 + \boldsymbol{b}_2\boldsymbol{x}_2 + \ldots + \boldsymbol{b}_m\boldsymbol{x}_m \quad \textit{(5b)}$$

Where:

$\boldsymbol{x_1, \ldots, x_m}$ *are predictor variables*
$\boldsymbol{p}$ *is the proportion to be predicted*

We dichotomized the continuous data on daily asthma hospital admissions to study the extreme events. Extreme events were classified as those which occurred above and within the upper 10th quartile. Hence we examined the effects of the independent variables on the two categories of the dependent variable using a logistic regression, since a linear model was not necessary in demonstrating the relation between the outcome and predictor variable [192]. For this purpose, we regenerated non-time series version of the lagged variables in the model. This was necessary because of the choice of the post-estimation model fitness test (Hosmer-Lemeshow).

The Hosmer and Lemeshow goodness-of-fit test and p-value were used in determining the overall fitness of the model. The significance of the p-value for the goodness-of-fit test as well as the relative proportions of the number of covariate patterns to the number of observations was used to inform this choice.

RESULTS

This section describes time series pattern of the total daily asthma admissions in London and compares it to the patterns of some potential predictors (month, seasons, weather, air quality).

Asthma daily admissions are associated with the weather, air quality as well as monthly and seasonal effects. These associations can be used to predict and forecast normal and extreme asthma events.

Selected lags of the independent weather and air quality factors, which were obtained from bivariate analysis, are presented. The selection includes twenty four lagged independent variables, however, only four of these (i.e. 9-day lag temperature; 7-day lag humidity; 14-day lag ozone; and 3-day lag nitrogen oxide), in addition to the monthly and seasonal variations, were considered suitable and used in modelling.

Five different multivariate models from the selected variables were used to estimate both normal and extreme asthma events. These were then compared for their predictive values. The findings show that the Negative binomial regression model had the least RMSE (0.16%) compared to the quantile regression models (QRMs). On the other hand, the QRMs had higher predictive coefficients as well as predictive values for both normal and extreme events. Air temperature and humidity were significant weather factors that were associated to asthma events. Also Nitrogen oxide and ozone were significant in most of the models.

Daily Asthma Hospital Admissions in London (2005-2006)

The total daily asthma admission in London is illustrated in the time series plot in Figure 5. Two major extreme events were observed; around the end of spring 2005 and same period in 2006. Generally the summer periods appear to have the lowest admission rates whilst the highest were reported in the autumn months. Both the winter and spring periods had moderate rates of admissions. Even though these observations do not clearly show any peaks and troughs as may be suggested by the illustration above, they do attempt to present the distribution of asthma daily admissions seasonally. There are very wide variations in the autumn, compared to the winter and spring periods.

Figure 5 Daily Asthma hospital admissions in London (2005-2006)

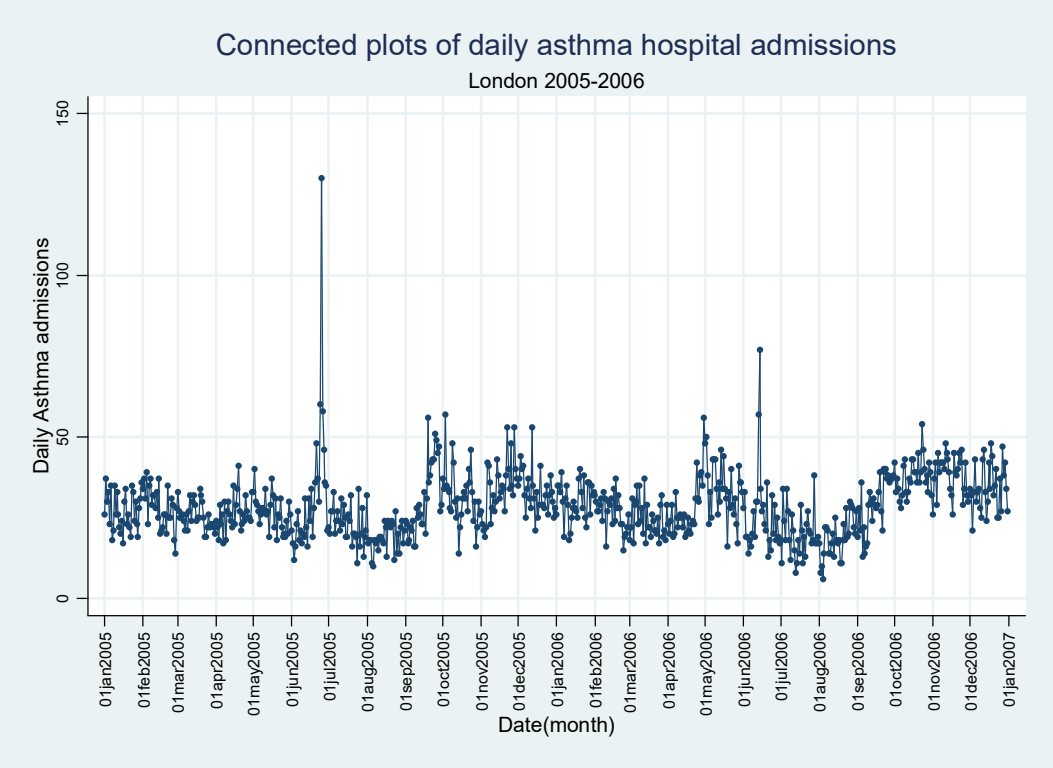
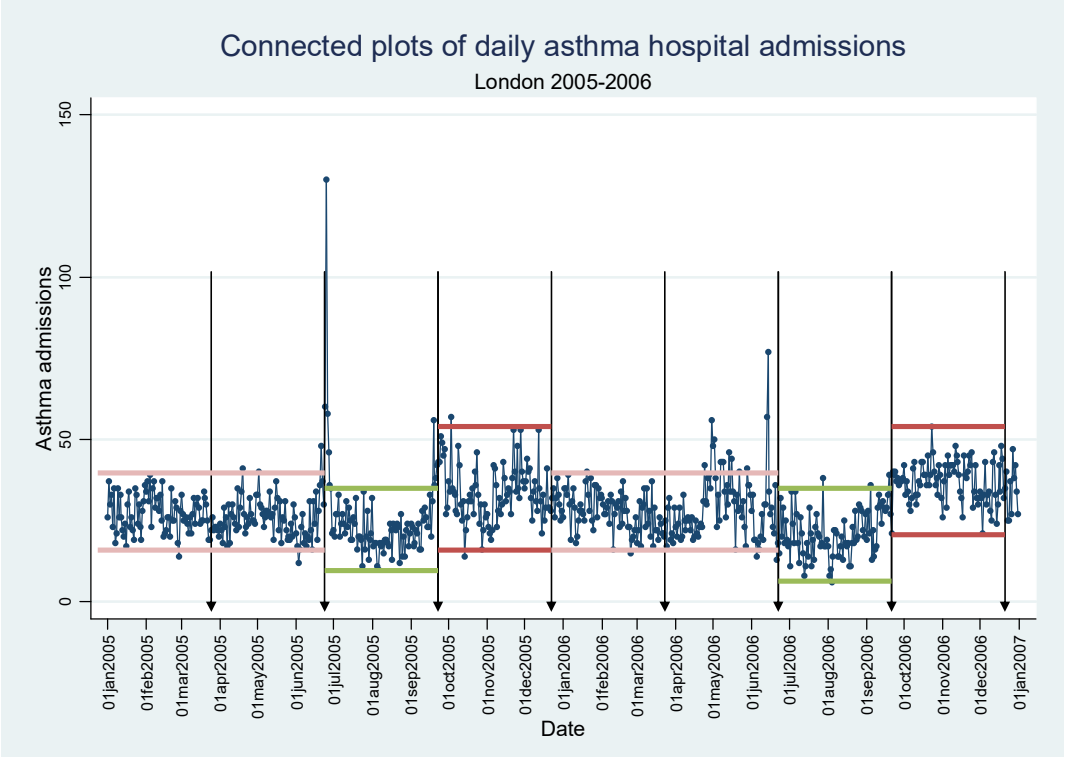


Figure 6 Daily Asthma hospital admissions in London (2005-2006)

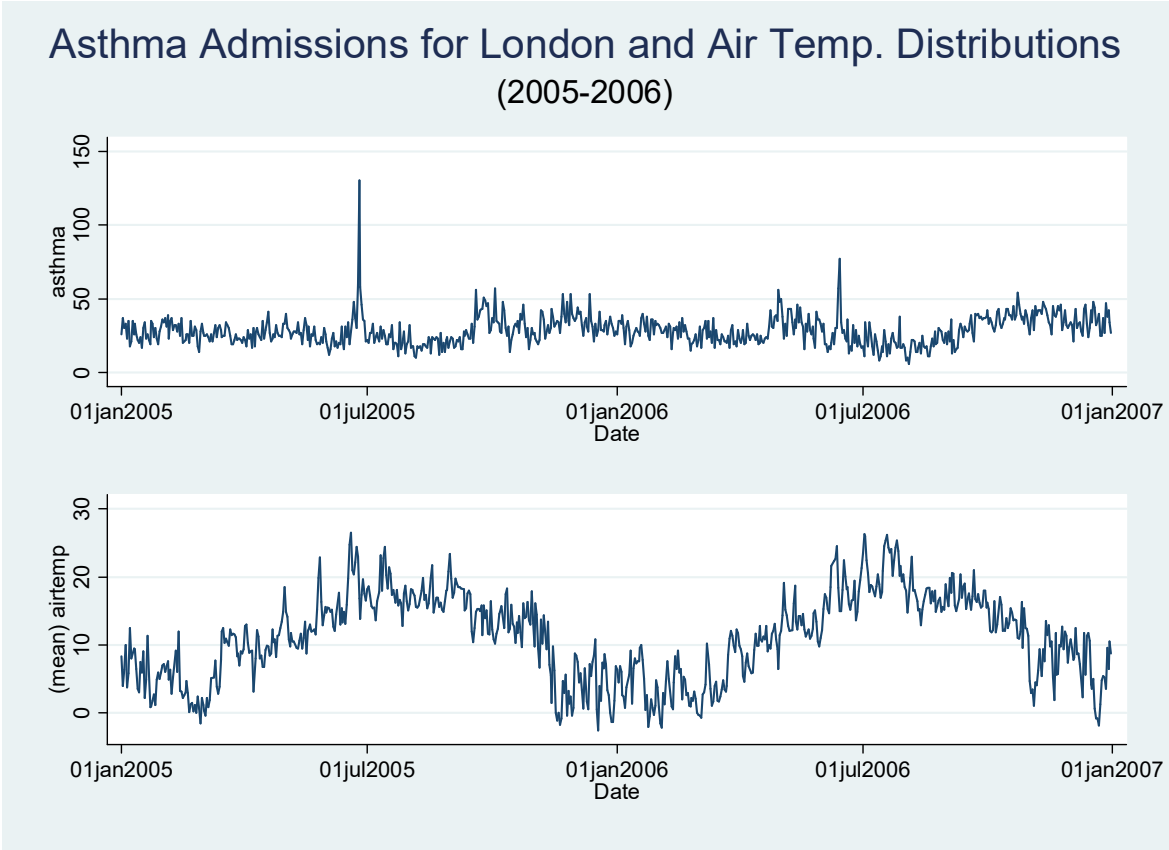


Arrow marks /divisions represent the seasons in a year

Asthma Admissions and Temperature (London, 2005-2006)

The mean air temperature distribution is presented alongside the asthma daily hospital admissions for London in Figure 7. The higher temperatures were recorded in summer whilst the lower temperatures were recorded in the winter months. The other specific temperature readings like the daily maximum/minimum temperature, dew point or wet bulb temperatures followed the same patterns (see Figures 18-21 in the appendix).

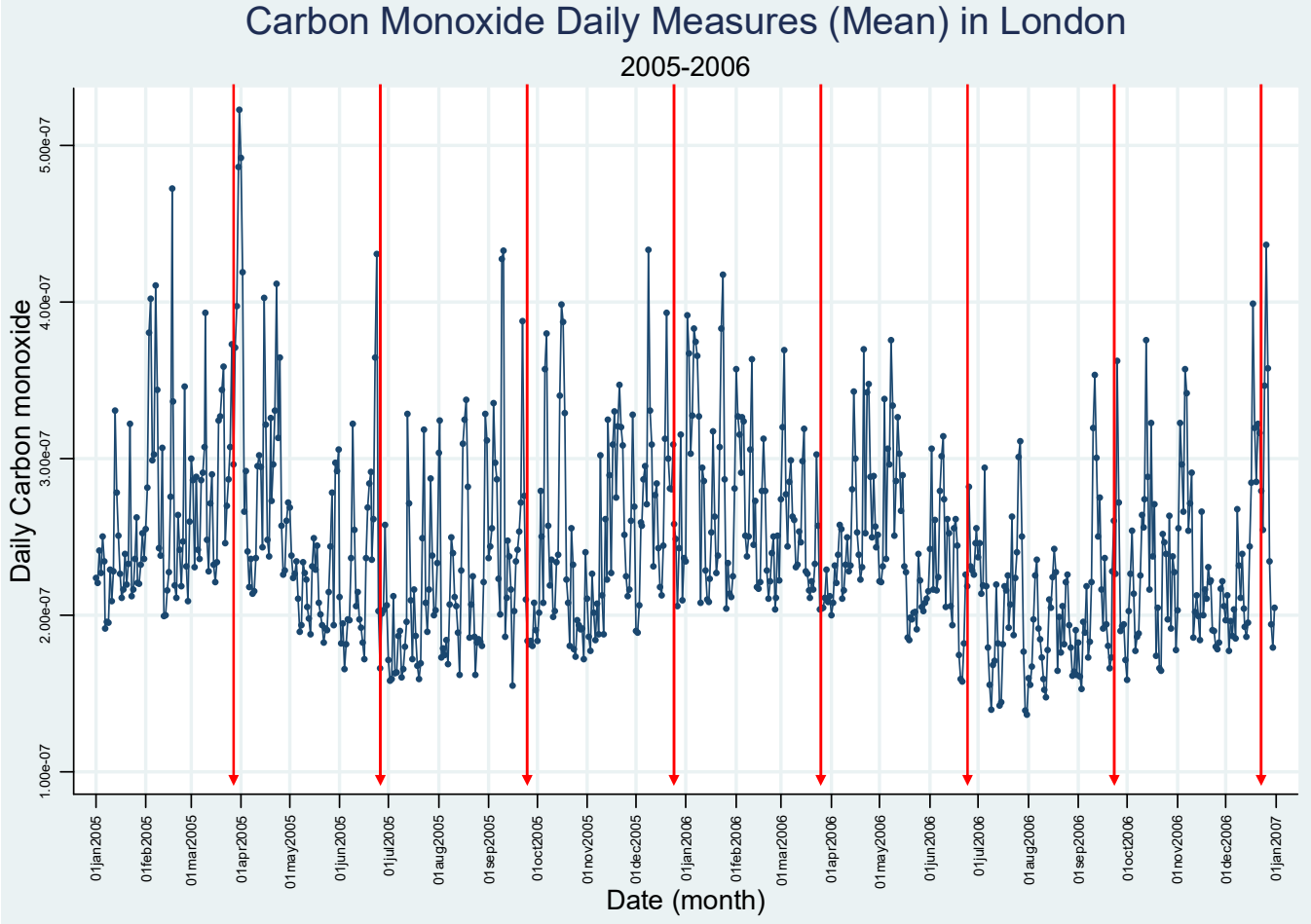
Figure 7 Asthma Admissions and Mean Air Temperature Distribution (London, 2005-2006)



Asthma Admissions and Air Quality (London, 2005-2006)

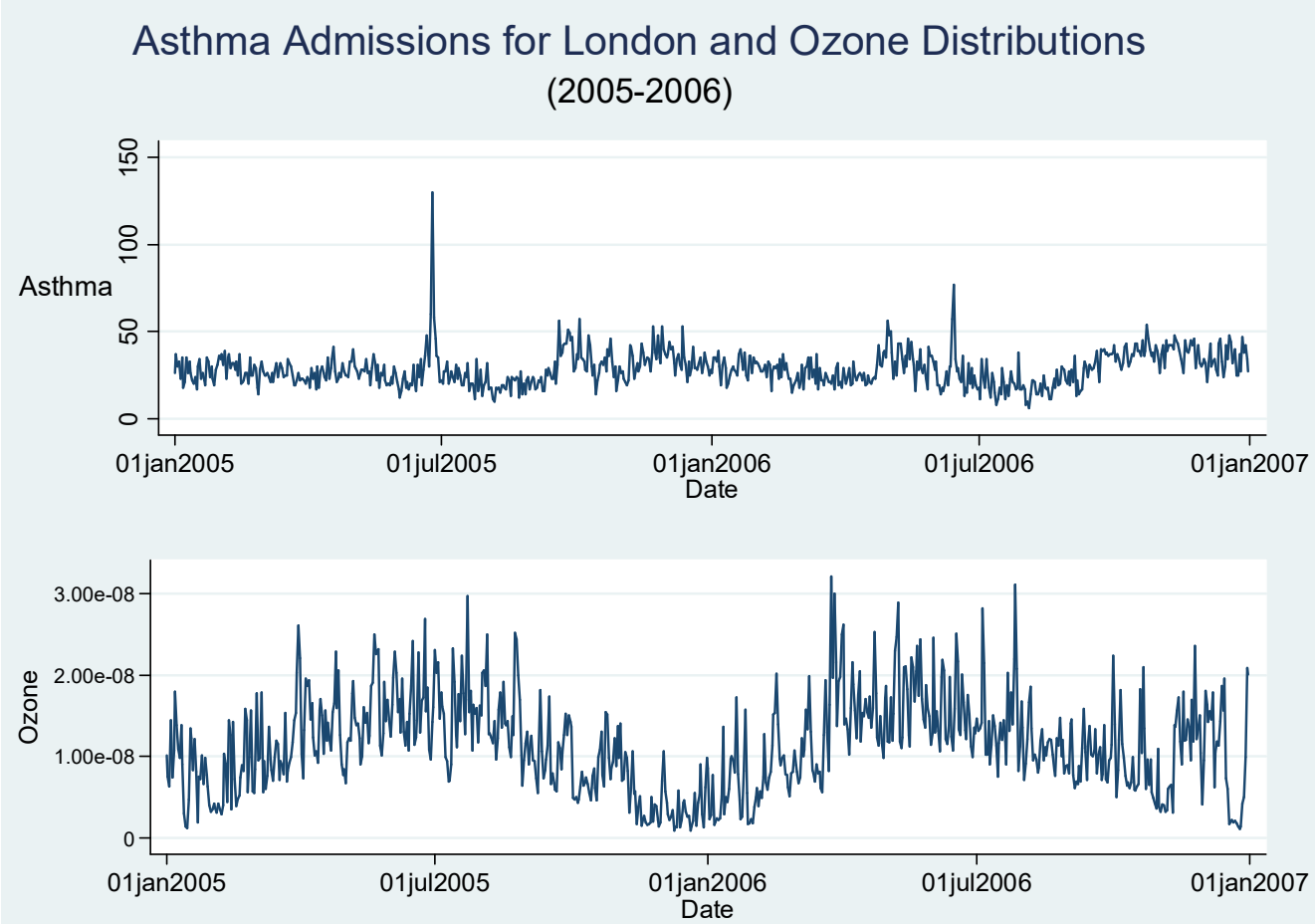
Apart from Ozone all the other air pollutants do not appear to have any regular seasonal or occasional trend(s). Taking Carbon monoxide as an example, the pattern distribution with reference to the seasonal marks (red lines) with vertical monthly grids (Figure 8), do not show any regular seasonal or monthly effect. This is same for the other pollutants (Nitrogen dioxide, Nitrogen oxide, Sulphur dioxide Formaldehyde and PM10) except ozone. Ozone shows some seasonal trends with notable peaks in the summer period and low records in the winter (See Figure 9).

Figure 8 Mean Daily Carbon monoxide distribution (London, 2005-2006)



Arrow marks /divisions represent the seasons in a year

Figure 9 Asthma Admissions and Ozone Distribution (London, 2005-2006)



Asthma Admissions and Other Indicators (London, 2005-2006)

The other indicators whose distribution patterns have been plotted with that of the Asthma daily admissions are presented in Figures 10 to Figure 12. These include Humidity, Barometric vapour pressure and Wind speed. Both Humidity and Pressure show some seasonal patterns whilst Wind speed does not.

Figure 10 Asthma Admissions and Humidity Distributions (London, 2005-2006)

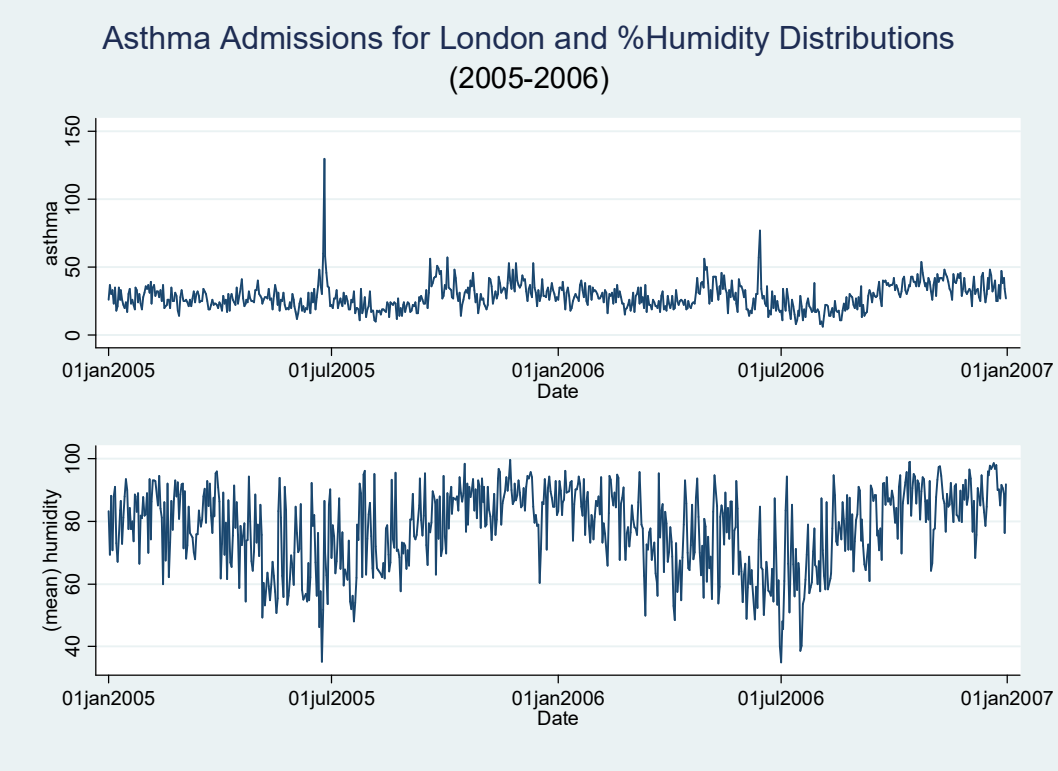


Figure 11 Asthma Admissions and Mean Wind speed Distributions (London, 2005-2006)

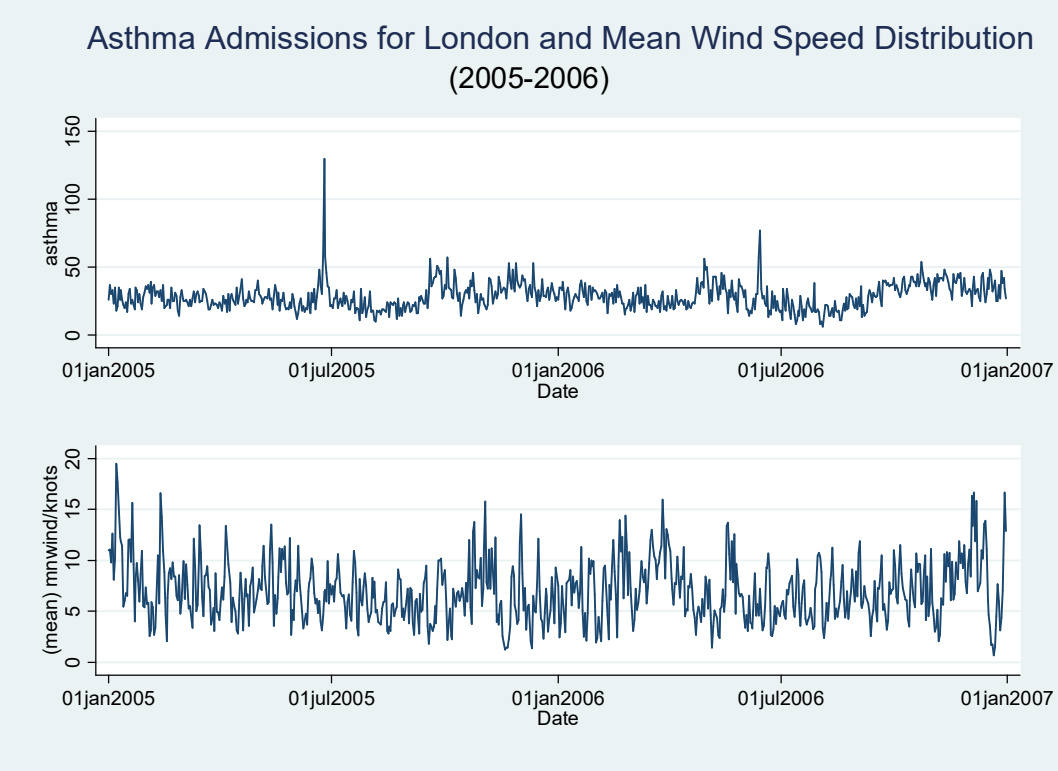
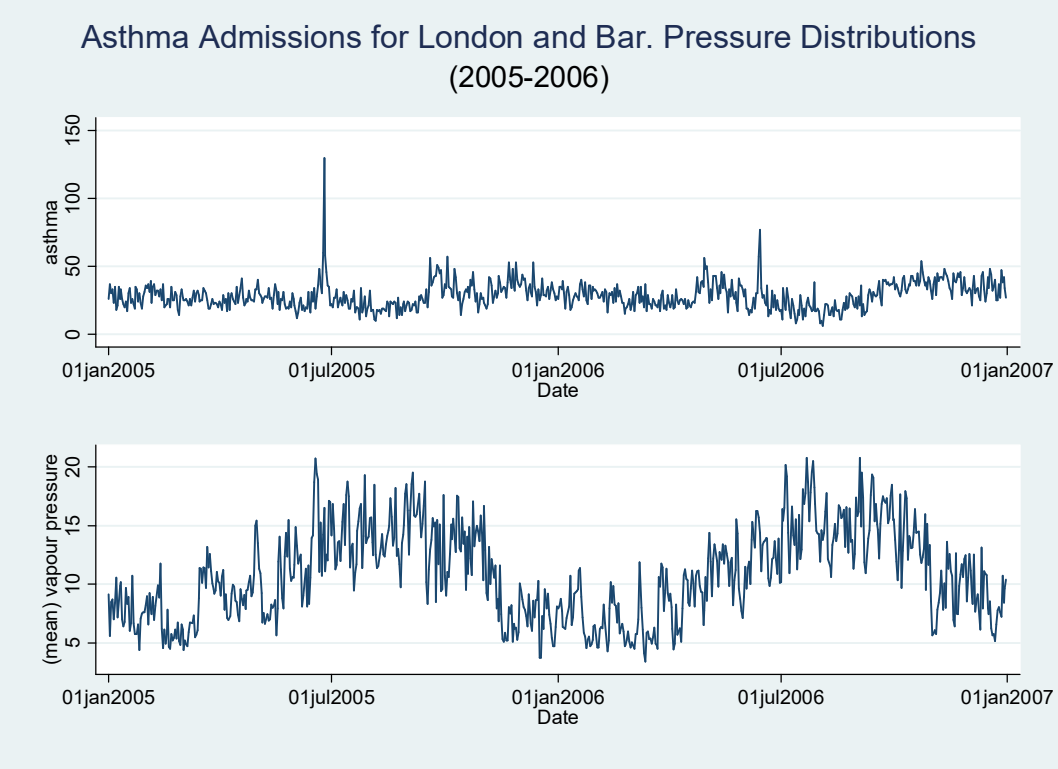


Figure 12 Asthma Admissions and Barometric vapour pressure Distributions (London, 2005-2006)



Bivariate Analysis

Exploring Lag Days for the Explanatory Factors

We explored lags from 1 to 21 days for each key independent variable through a bivariate test with asthma hospital admission as a dependent variable. The most appropriate lag-day was selected for modelling the effect(s) of the independent variables. The following were selected for consideration:

Table 2 List of best selected lag days of the bivariate analysis (independent variables) generated from the NegBin model given that the alpha coefficient of each >0

Variable	LagDay	Coef.	P|>|z	[95% Conf.	Interval]
Maximum temperature	L15.	-0.00943	0.000	-0.01306	-0.0058
Minimum temperature	L8.	-0.01004	0.000	-0.01453	-0.00556
Night minimum temperature	L8.	-0.00878	0.000	-0.01333	-0.00423
Night maximum temperature	L8.	-0.00942	0.000	-0.01345	-0.00539
Day Maximum temperature	L15.	-0.0094	0.000	-0.013	-0.0058
Day Minimum temperature	L9.	-0.01062	0.000	-0.01463	-0.00661
Night temperature drop	L19.	-0.03297	0.000	-0.04536	-0.02057
Day Temperature drop	L13.	-0.02492	0.000	-0.03811	-0.01174
Temperature drop	L19.	-0.01841	0.000	-0.02589	-0.01093
Mean wind speed	L2.	-0.01145	0.006	-0.01956	-0.00334
Air temperature	L9.	-0.01032	0.000	-0.01419	-0.00645
Wet bulb temperature	L9.	-0.00886	0.000	-0.01334	-0.00438
Dew point temperature	L2.	-0.00601	0.008	-0.01049	-0.00153
Barometric vapour pressure	L2.	-0.00937	0.004	-0.01569	-0.00305
Humidity	L7.	0.007207	0.000	0.005318	0.009095
Humidity	L19.	0.008059	0.000	0.006146	0.009972
Carbon monoxide	L1.	969867	0.000	578761.6	1360972
Formaldehyde	L2.	1.85E+07	0.000	1.13E+07	2.57E+07
Nitrogen dioxide	L1.	5674741	0.000	2683235	8666247
Nitrogen oxide	L3.	4742168	0.000	2720752	6763584
Ozone	L14.	-1.31E+07	0.000	-1.73E+07	-8981818
Sulphur dioxide	L3.	4143018	0.012	919360.7	7366675
Particulate Matter (PM10)	L2.	2568230	0.056	-68198.6	5204659
Particulate Matter (PM10)	L21.	-4248289	0.003	-7058080	-1438497

Multivariate Analysis

Comparison and Selection of Temperature Related Indicator(s):

We included all the temperature related variables including some generated potential predictors in a single model to determine the one(s) best associated with asthma hospitalization. These independent temperature related variables were Maximum temperature, Night minimum temperature, Night maximum temperature, Day maximum temperature, Air temperature, Dew point temperature, Wet bulb temperature, Day temperature drop, Temperature drop, Minimum temperature, Day minimum temperature, and Night temperature drop. The Minimum temperature, Day minimum temperature, and Night temperature drop were dropped from the model because of collinearity. The effects demonstrated by these potential predictors (independent variables) represent their respective effect on the day of hospitalization. At a p-value of 0.05, those found to be significant were Night maximum temperature (p=0.002), Day maximum temperature (p=0.01), Air temperature (p=0.025), Temperature drop (p=0.036).

The effect(s) of Environmental factors on health impacts like asthma hospitalization are cumulative, and not frequently instantaneous. We thus investigated the lagged effects of the above selected temperature related variables. The model was constructed with the best selected lags for each indicator. For the lagged model, at a p-value of 0.05, the significant predictors included the 8-day lag night maximum temperature (p=0.033), 9-day lag day minimum temperature (p=0.049), 9-day lag air temperature (p=0.001), 2-day lag dew point temperature (p=0.09) and a 9-day lag wet bulb temperature (p<0.0001). The NegBin output illustrates the expected change in log count for a one-unit increase in temperature. Thus for example in the un-lagged NegBin model (Table 3), every one degree increase in the “Night max temperature” there is an expected increase of 1.041 (~exp 0.04) daily asthma hospital admissions, given all other factors are held constant. Similarly, for every one degree increase in “Air temperature” there is an expected drop of 1.083 (~exp -0.08) daily asthma hospital admissions. However in the lagged model (Table 4) one degree increase in the “Night max temperature” results in an increase of 1.031 (~exp 0.03) daily asthma hospital admissions, accounting for all the other listed variables

Table 3 Negative binomial regression Asthma Model comparing the temperature related independent variables [alpha>0]

Variable	Log change[Ψ]	Robust Std. Err
Maximum temperature	0.03	0.03
Night min. temperature	0.02	0.02
Night max. temperature	0.04**	0.01
Day max. temperature	-0.09**	0.03
Air temperature	-0.08*	0.03
Dew point temperature	-0.01	0.03
Wet bulb temperature	0.09	0.06
Day temperature drop	0.02	0.01
Temperature drop	0.03*	0.01

[Ψ]Expected change in log count for a one-unit increase in temperature
Coefficient: * p<0.1 ** p<0.01; Log psuedo-likelihood: -2508.86; **Chi2**: 101.

Table 4 Negative binomial regression Asthma Model comparing the lagged(L) day temperature related independent variables [alpha>0]

Variable	Log change[Ψ]	Robust Std. Err
L15.Maximum temperature	0.01	0.05
L8.Minimum temperature	-0.02	0.02
L8.Night min. temperature	0.01	0.02
L8.Night max. temperature	0.03*	0.01
L15.Day max. temperature	-0.02	0.04
L9.Day min. temperature	-0.04	0.02
L19.Night temperature drop	-0.01	0.01
L13.Day temperature drop	-0.01	0.01
L19.Temperature drop	-0.00	0.01
L9.Air temperature	-0.05**	0.02
L9.Wet bulb temperature	0.07***	0.01
L2.Dew point temperature	0.01	0.00

[Ψ]Expected change in log count for a one-unit increase in variable
Coefficient: * p<0.1 ** p<0.01; *** p<0.001; Log psuedo-likelihood: -2035.88; **Chi2**: 86.03.

Comparison and Selection of Air Pollutant Related Indicator(s):

In selecting the appropriate air pollutants, we included all air pollutants (available in our dataset) in the negative binomial regression base model. These variables were Carbon monoxide, Formaldehyde (HCHO), Nitrogen dioxide, Nitrogen oxide, Ozone, Sulphur dioxide, and Particulate matter (pm10). Hence they were all treated as potential predictors so that we could subsequently determine the one(s) most strongly associated with asthma hospitalization. We observed Nitrogen dioxide (p=0.043) Ozone (p=0.001) Sulphur dioxide (p=0.002) were significantly associated with daily asthma hospital admissions.

Table 5 Negative binomial regression Asthma Model comparing the air pollutant related independent variables [alpha>0]

Variable	Log change$^{\Psi}$ in kgm^{-3} (μm^{-3})		Robust Std. Err
Carbon monoxide	6.5e+05	(0.00065)	4.10E+05
Formaldehyde	1.0e+07	(0.01)	1.50E+07
Nitrogen dioxide	7.4e+06*	(0.0074)	3.70E+06
Nitrogen oxide	-5.2e+05	(-0.00052)	4.20E+06
Ozone	-9.7e+06***	(-0.009.7)	2.80E+06
Sulphur dioxide	-1.2e+07**	(-0.012)	3.80E+06
Particulate matter	-2.3e+06	(-0.0023)	2.20E+06

$^{\Psi}$Expected change in log count for a one-unit increase in variable
Coefficient: * p<0.1 ** p<0.01; *** p<0.001; Log psuedo-likelihood: -2626.79; **Chi2**: 44.41.

We proceeded to assess the lagged effects of these pollutants. The lagged model was constructed with the best selected lags for each indicator. These lags were 1-day lag Carbon monoxide 2-day lag HCHO, 1-day lag Nitrogen dioxide, 3-day lag Nitrogen oxide, 14-day lag Ozone, 3-day lag Sulphur dioxide, 2-day lag PM10, and a 21-day lag PM10. At a p-value of 0.05, 14-day lag Ozone (p<0.0001), 3-day lag Nitrogen oxide (p=0.012), 2-day lag PM10 (p=0.024), and a 21-day lag PM10 (p=0.03) were significant. This preliminary predictive model also took into account other potential predictors e.g. the astronomical seasonal effect, monthly variations, humidity and wet bulb temperature.

Table 6 Negative binomial regression Asthma Model comparing the lagged(L) day air pollutant related independent variables [alpha>0]

Variable	Log change$^{\Psi}$ in kgm^{-3} (μm^{-3})		Robust Std.Err
L1. Carbon monoxide	3.5e+05	(0.00035)	3.20E+05
L2.Formaldehyde	9.9e+06	(0.0099)	5.50E+06
L1.Nitrogen dioxide	3.8e+06	(0.0038)	2.30E+06
L3.Nitrogen oxide	4.8e+06*	(0.0048)	1.90E+06
L14.Ozone	-1.0e+07***	(-0.01)	2.20E+06
L3.Sulphur dioxide	-2.3e+06	(-0.0023)	2.90E+06
L2.Particulate matter	-4.4e+06*	(-0.0044)	2.00E+06
L21.Particulate matter	-3.0e+06*	(-0.003)	1.40E+06

$^{\Psi}$Expected change in log count for a one-unit increase in variable
Coefficient: * p<0.1 ** p<0.01; *** p<0.001; Log psuedo-likelihood: --2538; **Chi2**: 79.99.

Modelling

Negative Binomial Regression Predictive Model

We constructed the first model; a negative binomial regression model for asthma admissions accounting for a 9-day lag air temperature, 14-day lag ozone, 13-day lag nitrogen oxide and 7-day lag humidity as well as seasonality and monthly variation. The output as shown in the Table 7 below indicates Asthma daily admissions within February-March and July-August are not significant in the model (i.e. p-value > 0.05).

Table 7 An illustration of the predicted output for the Negative binomial regression model for Asthma hospital admissions in London (2005-2006)

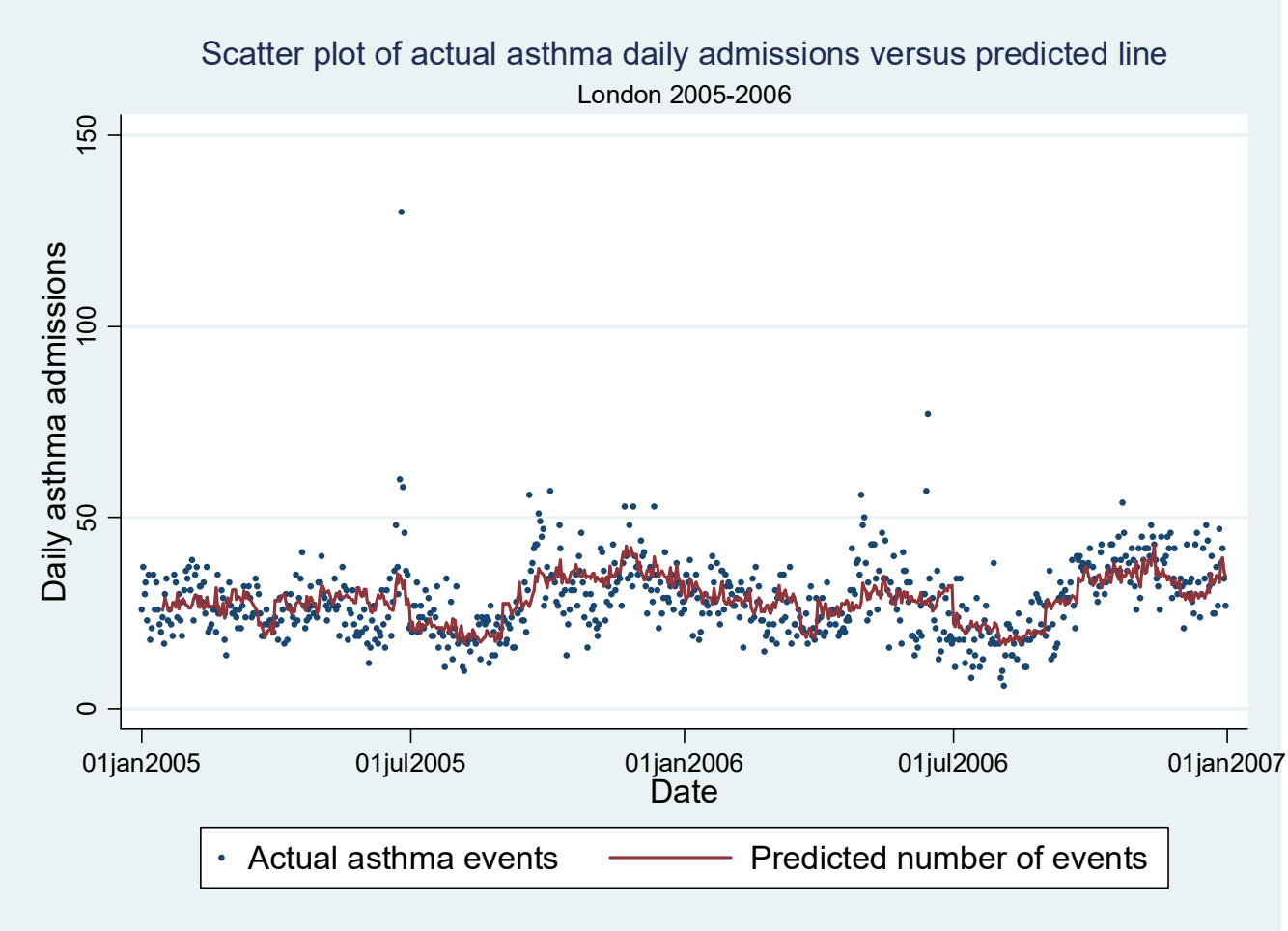
Variable	**Log change$^{\Psi}$**	**Robust Std. Err**
Summer_2	0.14	-0.13
Autumn_3	0.35***	-0.09
Winter_4	0.32***	-0.07
Feb_2	-0.06	-0.04
Mar_3	0.02	-0.05
Apr_4	0.37***	-0.09
May_5	0.50***	-0.09
Jun_6	0.47***	-0.11
Jul_7	0.09	-0.11
Aug_8	-0.07	-0.1
Sep_9	0.27**	-0.09
Oct_10	0.23***	-7.00E-02
Nov_11	0.23***	-0.05
Dec_12	0.13**	-4.00E-02
L9.airtemp	-0.01*	0.0
L14.Ozone	5.5e+06* (0.0055§)	-2.40e+06 (-0.0024§)
L3.Noxide	3.5e+06*** (0.0035§)	-7.70e+05 (-0.00077§)
L7.humidity	0.00**	0.0

$^{\Psi}$Expected change in log count for a one-unit increase in variable. §μm^{-3}
Coefficient * p<0.1 ** p<0.01 *** p<0.001; Log psuedo-likelihood: -2290.01; **Chi2**: 473.38

The negative binomial regression model for daily asthma hospital admissions from the listed variables in Table 7 was statistically significant (chi-squared = 473.38, df = 18, p<.0001). All the predictors were statistically statically significant excluding the months February, March, July and August as well as the summer season. For all the data, the expected change in log count for a one-unit increase in each variable is indicated in the Table 7. The expected log count comparator for season and month variables are spring and January respectively; holding other variables constant. For instance, a one-unit (μm^{-3}) increase in "L3.Noxide" (3-day lag exposure of nitrogen oxide) would result in a 1.008 asthma admissions, adjusting for all the other listed variables, and similarly, the months of May and June had expected log counts of 0.50 and 0.47 less than the month of January, holding all other variables constant.

The illustration below (Fig. 13) shows the distribution of the actual asthma daily admissions alongside the negative binomial regression predicted curve. Even though peak events failed to be captured by the predictive model, the average (expected daily admissions) appears to be reasonably well captured. This graphical presentation demonstrates the degree of accuracy to which the dependent factor could be predicted with this model. The nature of the predicted line of fit in relation to the original distribution is characteristic of any model, which predicts based on the mean. Hence this predicted model is useful mainly in estimations that are particularly interested in the normal events.

Figure 13 Scatter plot of actual asthma daily admissions versus NegBin predicted line



The NegBin Model Diagnostic Test (Link test)

The fitness of the negative binomial model was assessed using the Link test (Table 8) and the fit of the model was determined by the p-value of "the square of the predicted variable" being >0.05 given that the entire model is significant as well as the p-value of the predicted variable. We further predicted asthma admissions on the basis of this model and this illustrated alongside a scatter plot of the actual daily asthma admissions for London (Fig. 13)

Table 8 The NegBin Model specification test after estimation (Link test)

Test Variable	Coef.	Std. Err.	P>z	[95% Conf.	Interval]
Predicted	0.05269	1.488789	0.972	-2.86528	2.970662
(Predicted)²	0.14395	0.226094	**0.524**	-0.29918	0.587088

The NegBin Model Predictive Value (Accuracy & Specificity)

The negative binomial regression models expected values. Nonetheless, if the predictors fit the data well, even apparently extreme events will be well modelled. We assessed the predictive validity of the model by contrasting the number of days with predicted extreme numbers of asthma admissions against the actual number of days with extreme numbers of asthma admission. "Extreme" in this context used the 90th percentile of asthma admissions as the cut-off. This percentile corresponded to a maximum of 40 daily asthma admissions for London. Actual and predicted daily admissions were represented by dichotomous variables representing ≥40 or <40 admissions per day. A cross tabulation of these two indicators yielded the predictive values of asthma hospital admission. The results of the cross tabulation (Table 9) show 8 days of extreme events were predicted from the data, whilst 62 days were falsely predicted as days of extreme asthma events. On the other hand, 75 days of extreme asthma admissions were missed. Hence the predictive values for both the normal and extreme asthma events are 88.64% and 11.43% respectively.

Table 9 The Predictive Value of Asthma Daily Admissions using a Negative binomial regression mode.

		Actual Asthma Admissions		Predictive Value
		<40	≥40	
Predicted Asthma Admission	**<40**	585	75	(Normal): 88.64%
	≥40	62	8	(Extreme): 11.43%
		Sensitivity: 90.42%	Specificity: 9.64%	

False positive (α): 11.36% False negative (β): 9.58%

Forecast Error and Cross validation of NegBin Model

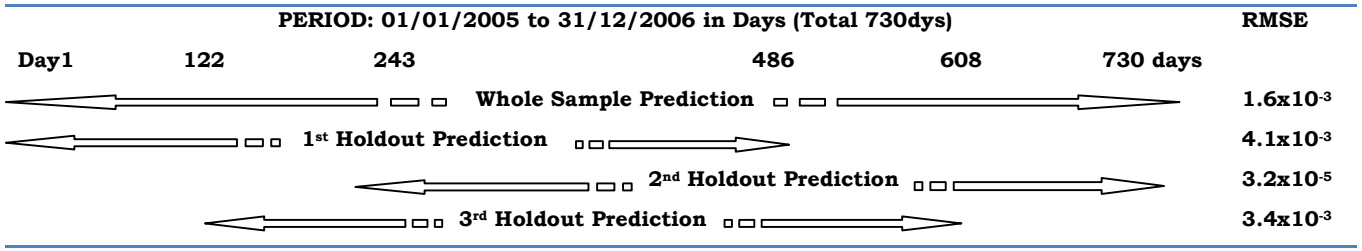
The forecast error was determined by estimating the RMSE for the negative binomial regression model. This was found out to be **$1.6x10^{-3}$** (Table 10).

A 3-fold cross validation of the model was done to examine the variation in RMSE value given a sub-sample of the dataset. This was done holding 2/3rd of the dataset as observations to predict for asthma events. The Root Mean Square Error (RMSE) was used as an indicative measure of the models accuracy. The results from this showed very wide variations in RMSE (Table 10). The first holdout sample had a RMSE of $4.1x10^{-3}$, which shows 156.25% increase compared to the original/total sample and the third holdout sample had a RMSE of $3.4x10^{-3}$ (i.e. 112.5% increase). However, the second holdout sample (RMSE of $3.2x10^{-5}$) had 98% drop compared to the whole sample.

Table 10 Summary Statistics of Actual and Predicted Asthma Admissions for various window periods using the NegBin Model

Variable	Window Period (Dates)	Obs	Mean	Std. Dev.	Range		RMSE
					Min	Max	
Whole Sample							**$1.6x10^{-3}$**
Asthma admissions (actual)	01/01/05-31/12/06	730	28.48493	9.846025	6	130	
Predicted asthma from actual	01/01/05-31/12/06	668	28.52901	5.490202	16.80289	42.76237	
1st Holdout Sample							**$4.1x10^{-3}$**
Asthma admissions from 1st holdout	01/01/05-01/05/06	486	27.8786	9.347786	10	130	
Predicted asthma from 1st holdout	01/01/05-01/05/06	440	27.96990	5.196354	16.49666	46.23573	
2nd Holdout Sample							**$3.2x10^{-5}$**
Asthma admissions from 2nd holdout	01/09/05-31/12/06	486	29.78807	9.614548	6	77	
Predicted asthma from 2nd holdout	01/09/05-31/12/06	454	29.78737	6.132507	16.64241	46.06879	
3rd Holdout Sample							**$3.4x10^{-3}$**
Asthma admissions from 3rd holdout	02/05/05-31/08/06	486	27.26543	10.49681	6	130	
Predicted asthma from 3rd holdout	02/05/05-31/08/06	454	27.34119	5.70974	15.26246	43.3106	

RE: ILLUSTRATION



Quantile Regression Predictive Model(s)

Quantile Regression Models (QRM) were constructed for asthma admissions accounting for a 9-day lag air temperature, 14-day lag ozone, 13-day lag nitrogen oxide and 7-day lag humidity as well as seasonality and monthly variation. Based on the scale of the dataset we selected some appropriate quantiles (0.80 & 0.85) for modelling the extreme asthma events. The optimization process for the QRM was controlled by the selecting the best **w**eighted **l**east-**s**quares **i**teration **n**umbers (**wlsin**). The wlsin number is the iteration that identifies the best estimator for the regression output. Two wlsin were identified in the simulations (these were: 278 & 280). Following these selection criteria three feasible QRMs with a mix of 0.80/0.85 quantiles and 278/280 wlsin were developed and used in the subsequent analysis and graphical presentations.

As a result of some methodological limitations in the application of the earlier proposed cross validation technique, the procedure was not carried out in the QRMs. For each particular predictive QRM, a specific wlsin, based on the entire dataset was identified used to generate the regression output. Since the wlsin parameter changes with changes in the dataset, it was not possible to cross validate using different sub-samples of the dataset.

The QRM at 85th quantile and 278 wlsin

The QRM output for the 85th quantile (with a wlsin 278) is presented in Table 11. The last three months of the year (i.e. October November and December) as well as Air temperature, Ozone and Nitrogen oxide were highly significant.

The Table 12 below illustrates the predictive values of the model for both normal (91.86%) and extreme (16%) asthma events. It also presents the sensitivity (61.05%) and specificity (57.83%) as well as the False normal (α=8.14%) and False extreme (β=38.95%) prediction of this model.

The graphical presentation of the both the actual and predicted asthma admissions in Figure 14 show the contrast between the extremely distributed records and the predicted /fitted line. The predicted line is for extreme values beyond the 85th quantile, but visibly does not ideally capture all the extreme points, and hence the error estimations. A test of the RMSE of this model, i.e. comparing the actual asthma admissions to the predicted asthma admissions gave an error of 0.23074.

Table 11 Quantile regression estimates for 0.85 quantile (at wlsin 278)

Variable	Log change[Ψ]	Robust Std. Err
Summer_2	6.07	4.12
Autumn_3	-7.13	4.52
Winter_4	1.18	3.53
Feb_2	8.73*	4.18
Mar_3	-0.33	4.25
Apr_4	-7.45	5.80
May_5	2.18	5.61
Jun_6	4.32	5.80
Jul_7	-14.72*	7.31
Aug_8	-12.55	7.10
Sep_9	16.97*	6.62
Oct_10	26.96***	5.43
Nov_11	24.38***	4.62
Dec_12	20.83***	4.02
L9.airtemp	-0.91***	0.23
L14.Ozone	1.3e+09*** (1.3[§])	1.3e+08 (0.13[§])
L3.Noxide	-4.0e+08*** (0.4[§])	7.8e+07 (0.078[§])
L7.humidity	-0.05	0.07

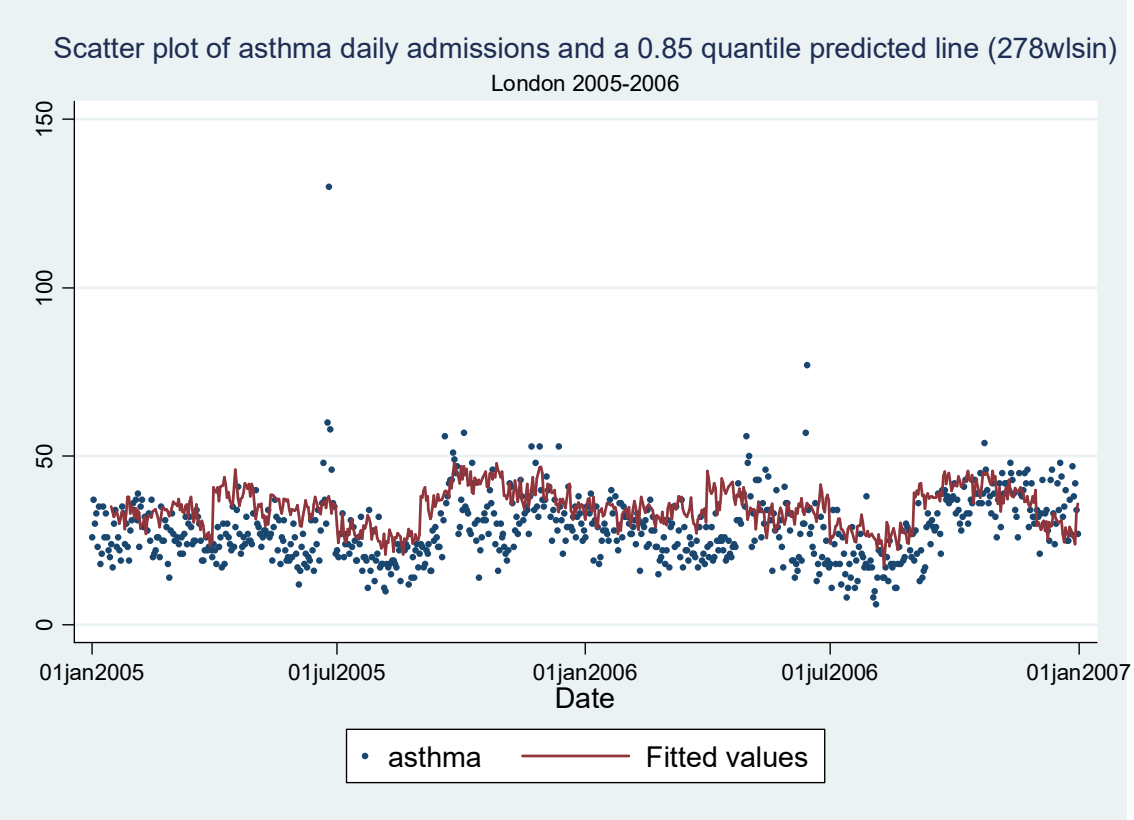
[Ψ]Expected change in log count for a one-unit increase in variable OR: * p<0.1 ** p<0.01 *** p<0.001; [§]μm^{-3}
[Ψ]Expected change in log count for a one-unit increase or a change from 1-0 in the case of the dummy variable;

Table 12 Predictive values of the Quantile regression estimates for 0.85 quantile (at wlsin 278)

		Actual Asthma Admissions		Predictive Value
		<40	≥40	
Predicted Asthma Admission	<40	395	35	(Normal): 91.86%
	≥40	252	48	(Extreme): 16%
		Sensitivity: 61.05%	Specificity: 57.83%	

False normal (α): 8.14% False extreme (β): 38.95%

Figure 14 Scatter plot of asthma daily admissions and a 0.85 quantile fitted values at a 278 weighted least-squares iteration number (London 2005-2006)



The QRM at 85th quantile and 280 wlsin

The second QRM model for the 85th quantile but with a slightly higher wlsin (280) presented a different result (compared to the 278 wlsin) as illustrated in Table 13. The month of April and a 14-day lag Ozone exposure had very significant p-values compared to the other variables. Thus when compared to the month of January, April is likely to have 13 times more daily admissions for asthma, given that all other factors are held constant. Also a one unit (μm^{-3}) increase /change in a 14-day lag exposure to ozone, is likely to cause a 0.37 reduction in daily asthma hospital admissions given that all other factors are held constant.

The predictive values of the model for both extreme (19.2%) and normal (91.21%) asthma events are presented in Table 14. The model's sensitivity and specificity are respectively 76.97% and 42.17%.below illustrates the. The model demonstrates a False normal (α=8.79%) and False extreme (β=23.03%) predictions.

The graphical illustration of the predicted asthma admissions alongside the actual distribution is shown in Figure 15. The predicted line for the extreme events still misses the two outstanding peaks and a few isolated extreme events, even though many others are captured. In the estimation of the forecast error for this model, the RMSE was 0.22354.

Table 13 Quantile regression estimates for 0.85 quantile (at wlsin 280)

Variable	Log change[Ψ]	Robust Std. Err
Summer_2	2.91	3.36
Autumn_3	8.20*	3.38
Winter_4	4.15	2.22
Feb_2	-1.58	3.06
Mar_3	0.20	2.75
Apr_4	13.23***	3.62
May_5	8.08	4.13
Jun_6	10.09*	4.37
Jul_7	2.00	5.18
Aug_8	-0.30	4.87
Sep_9	11.35*	4.74
Oct_10	9.76*	4.55
Nov_11	5.29	4.27
Dec_12	-0.63	3.65
L9.airtemp	-0.40*	0.18
L14.Ozone	-3.7e+08** (0.37§)	1.3e+08 (0.13§)
L3.Noxide	-1.0e+08 (0.1§)	6.7e+07 (0.067§)
L7.humidity	-0.10	0.05

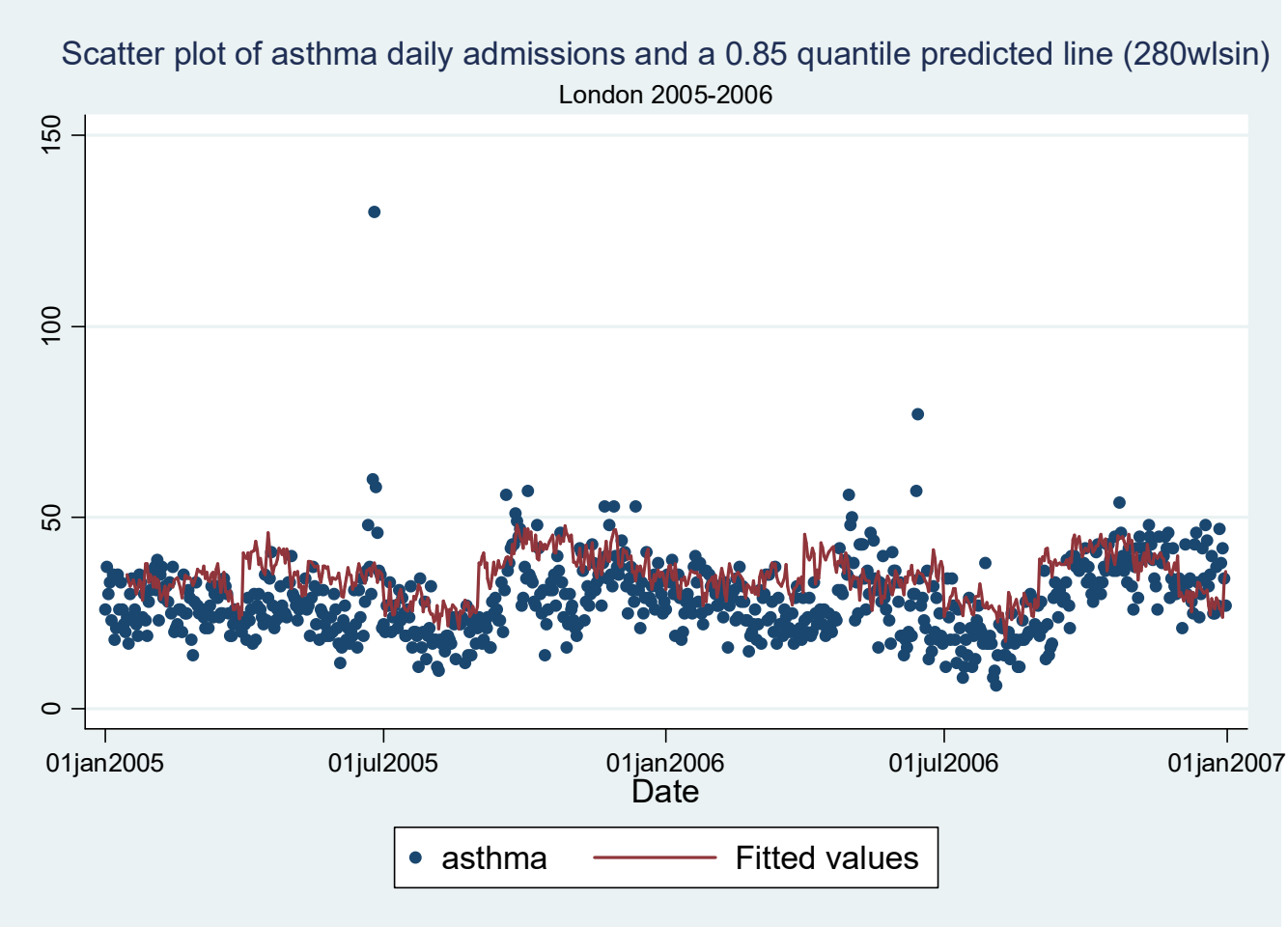
[Ψ]Expected change in log count for a one-unit increase in variable; OR: * p<0.1 ** p<0.01 *** p<0.001; §μm^{-3}

Table 14 Predictive values of the Quantile regression estimates for 0.85 quantile (at wlsin 280)

		Actual Asthma Admissions		Predictive Value
		<40	**≥40**	
Predicted Asthma Admission	**<40**	498	48	(Normal): 91.21%
	≥40	149	35	(Extreme): 19.02%
		Sensitivity: 76.97%	Specificity: 42.17%	

False normal (α): 8.79% False extreme (β): 23.03%

Figure 15 Scatter plot of asthma daily admissions and a 0.85 quantile fitted values at a 280 weighted least-squares iteration number (London 2005-2006)



The QRM at 80th quantile and 280 wlsin

This QRM model was constructed for the 80th quantile with a wlsin of 280. As illustrated in Table 15, but for the seasonal effect, most of the variables displayed significant effect in the model. Though the effect size for the pollutants ozone and nitrogen oxide appeared to be small, the latter was quite significant. For a one unit (μm^{-3}) increase in nitrogen oxide, there is a 0.3 times likely increase of asthma admission.

At the 80th quantile and 280 wlsin, the QRM has a predictive value of 21% for extreme events and 94% for normal asthma events. The model is quite sensitive (69%) but has a beta-error (false prediction of extreme events) of about 31% (see Table 16).

The graphical illustration of the predicted asthma admissions alongside the actual distribution is shown in Figure 16. The predicted line represents the 80th quantile, and has a RMSE of 0.28717.

Table 15 Quantile regression estimates for 0.80 quantile (at wlsin 280)

Variable	Log change[Ψ]	Robust Std. Err
Summer_2	-1.71	3.34
Autumn_3	5.59	2.97
Winter_4	2.79	2.50
Feb_2	-7.86**	2.51
Mar_3	-3.57	2.43
Apr_4	7.50*	3.32
May_5	10.84**	3.41
Jun_6	9.63**	3.56
Jul_7	8.78*	4.16
Aug_8	18.32***	3.87
Sep_9	12.76***	3.72
Oct_10	8.80**	3.04
Nov_11	11.58***	2.45
Dec_12	2.49	2.26
L9.airtemp	-0.69***	0.14
L14.Ozone	5.3e+07 (0.053[§])	8.5e+07 (0.085[§])
L3.Noxide	2.6e+08*** (0.26[§])	3.8e+07 (0.038[§])
L7.humidity	0.18***	0.04

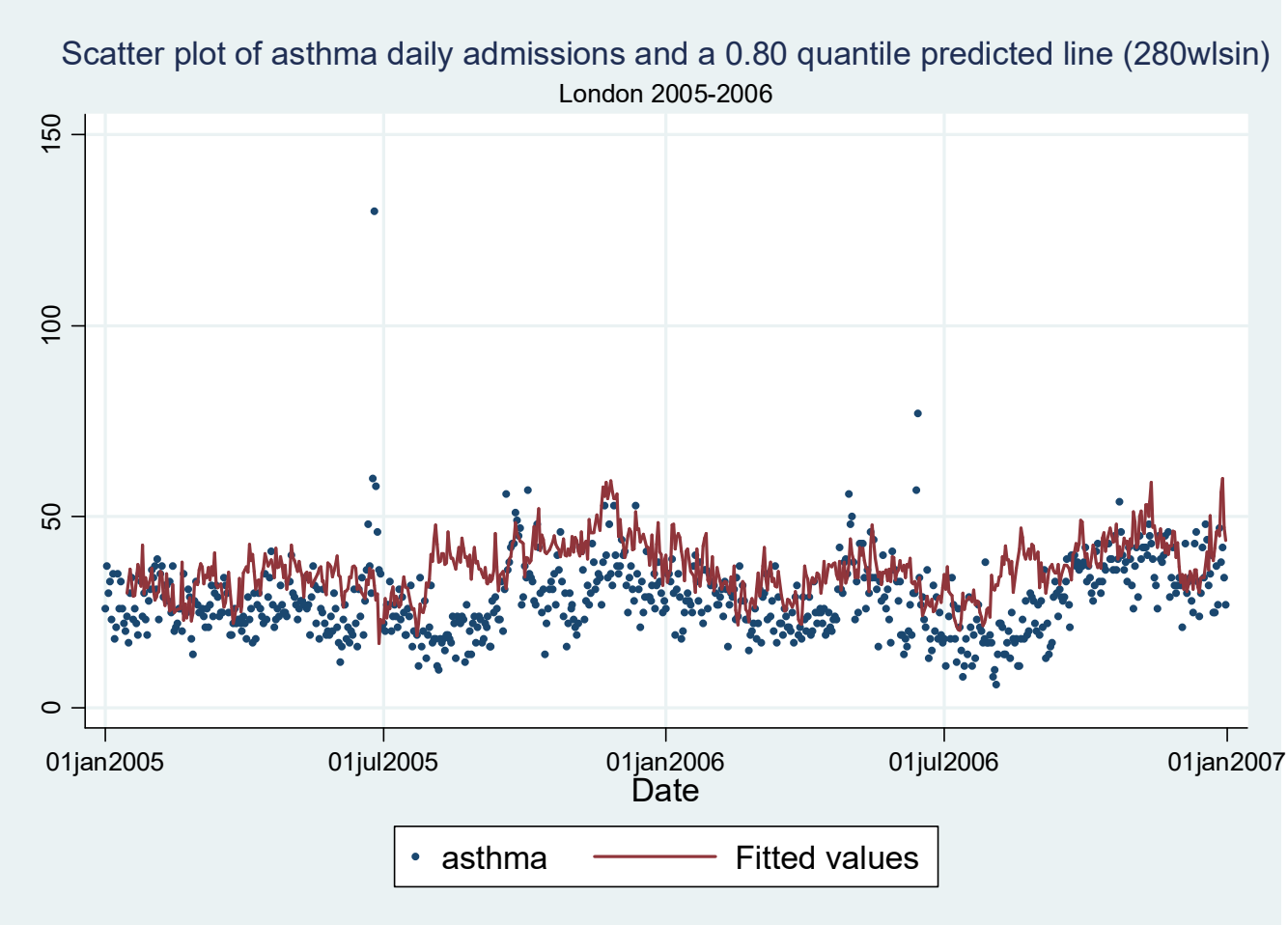
[Ψ]Expected change in log count for a one-unit increase in variable; OR: * p<0.1 ** p<0.01 *** p<0.001; [§]μm^{-3}

Table 16 Predictive values of the Quantile regression estimates for 0.80 quantile (at wlsin 280)

		Actual Asthma Admissions		Predictive Value
		<40	**≥40**	
Predicted Asthma Admission	**<40**	447	29	(Normal): 93.91%
	≥40	200	54	(Extreme): 21.26%
		Sensitivity: 69.09%	Specificity: 65.06%	

False normal (α): 6.09% False extreme (β): 30.91%

Figure 16 Scatter plot of asthma daily admissions and a 0.80 quantile fitted values at a 280 weighted least-squares iteration number (London 2005-2006)



Extreme Asthma Events [Logistic Regression Model]

We dichotomized the actual asthma admissions at the 90th percentile to separate the upper 10th quartile as a category of extreme events. Thus the proportion of extreme to normal days was given as 70:660 days. A logistic regression output of the model constructed with this modified dependent variable (i.e. the dichotomised asthma daily admissions) together with robust error terms are presented in Tables 34-35 (appendix). Table 17 further illustrates the coefficients for interpreting the relationships.

The goodness-of-fit test gave a p-value (p=0.8274), which initially suggests that the model fits reasonably well (Table 18). However, because the number of observations is equal to the number of covariate patterns the Pearson test is not appropriate for these data[1]. Hence we sought to test the Hosmer and Lemeshow chi-square using 2 degrees of freedom (i.e. on assumption that the table collapsed 4 quantile groups of estimated probabilities). The results displayed in Table 19, indicate a significant probability of the prediction involving 4 quantile groups (chi2=1.81 and p>0.1 [p=0.4051]). These tests suggest that the data may not be sufficient to reliably predict the upper 10th quantile as an extreme event using the method of logistic regression. However, the results from the above analysis may suggest that categories of the dependent variable, at most 4, could be simultaneously modelled to predict potential outcome for each category.

The predictive values and other statistics/parameters of the logistic regression model are summarised in Table 20. The model has a very low predictive value for extreme events (1.74%) and equally low model specificity.

[1] In the estimation of the Goodness of Fit (Pearson), it is usually assumed that the deviance statistic has a chi-square distribution, and this indirectly implies that the predictor(s) are categorical. Hence groups are defined by covariate patterns, which are not equal to the number of observations. If however the number of observations is equal to the number of covariate patterns, the Hosmer-Lemeshaw Goodness of Fit test allows us to group the data using predicted values with the group option (with the data table collapsed on quantiles of estimated probabilities).

Table 17 Asthma daily hospital admissions Model: prediction of Extreme versus Normal events

Variable	Log change[Ψ]	Robust Std. Err
Summer2	-2.00*	0.91
Autumn3	-4.40***	1.27
Winter4	-4.58***	1.37
Apr	-21.00***	1.46
May	-22.04***	1.41
Jun	-21.17***	1.51
Sep	-19.15***	1.85
Oct	-18.27***	1.99
Nov	-18.53***	1.94
Dec	-17.33***	1.99
L9.airtemp	0.04	0.04
L14.Ozone	-3.1e+07 (0.031[§])	-3.20e+07 (0.032[§])
L3.Noxide	-3.1e+07** (0.031[§])	1.10e+07 (0.011[§])
L7.humidity	-0.01	0.02

[Ψ]Expected change in log count for a one-unit increase in variable; [§]μm^{-3}; OR: * p<0.1 ** p<0.01 *** p<0.001;
Log Likelihood: -156.71; **Chi2:** 55.72

Table 18 Pearson Goodness-of-fit test: Logistic model for extreme asthma events

Pearson Goodness-of-fit test	
Number of observations	442
Number of covariate patterns	442
Pearson chi2(427)	399.36
Prob > chi2	0.8274

Table 19 Hosmer-Lemeshow Goodness-of-fit test: Logistic model for extreme asthma events 4 quantiles of estimated probabilities

Hosmer-Lemeshow Goodness-of-fit test	
Number of observations	442
Number of groups	4
Hosmer-Lemeshow chi2(2)	1.81
Prob > chi2	0.4051

Table 20 Predictive values of the Logistic regression estimates

		Actual Asthma Admissions		Predictive Value
		<40	**≥40**	
Predicted Asthma Admission	**<40**	364	78	(Normal): 82.35%
	≥40	283	5	(Extreme): 1.74%
		Sensitivity: 56.26%	Specificity: 6.02%	

False normal (α): 17.65% False extreme (β):43.74%

Summary Predictions from All Models

Table 21 presents a summary of some important indicators that were used to compare five different predictive models. It compares the parameters for the expected change in log count for a one-unit increase in a variable for each model. Also the predictive values of each model, sensitivity and specificity are indicated in the table.

Table 21 Summary Predictions from All Models

Variable/ Parameter	NegBin		LRM		QRM 1 [0.80q /wlsin 280]		QRM2 [0.85q /wlsin 280]		QRM3 [0.85q /wlsin 278]	
	Log change[Ψ]	Robust Std. Err	Log change[Ψ]	Robust Std. Err	Log change[Ψ]	Robust Std. Err	Log change[Ψ]	Robust Std. Err	Log change[Ψ]	Robust Std. Err
Summer_2	0.14	-0.13	-2.00*	0.91	-1.71	3.34	2.91	3.36	6.07	4.12
Autumn_3	0.35***	-0.09	-4.40***	1.27	5.59	2.97	8.20*	3.38	-7.13	4.52
Winter_4	0.32***	-0.07	-4.58***	1.37	2.79	2.5	4.15	2.22	1.18	3.53
Feb_2	-0.06	-0.04	-	-	-7.86**	2.51	-1.58	3.06	8.73*	4.18
Mar_3	0.02	-0.05	-	-	-3.57	2.43	0.20	2.75	-0.33	4.25
Apr_4	0.37***	-0.09	-21.00***	1.46	7.50*	3.32	13.23***	3.62	-7.45	5.80
May_5	0.50***	-0.09	-22.04***	1.41	10.84**	3.41	8.08	4.13	2.18	5.61
Jun_6	0.47***	-0.11	-21.17***	1.51	9.63**	3.56	10.09*	4.37	4.32	5.80
Jul_7	0.09	-0.11	-	-	8.78*	4.16	2.00	5.18	-14.72*	7.31
Aug_8	-0.07	-0.1	-	-	18.32***	3.87	-0.30	4.87	-12.55	7.10
Sep_9	0.27**	-0.09	-19.15***	1.85	12.76***	3.72	11.35*	4.74	16.97*	6.62
Oct_10	0.23***	-7.00e-02	-18.27***	1.99	8.80**	3.04	9.76*	4.55	26.96***	5.43
Nov_11	0.23***	-0.05	-18.53***	1.94	11.58***	2.45	5.29	4.27	24.38***	4.62
Dec_12	0.13**	-4.00e-02	-17.33***	1.99	2.49	2.26	-0.63	3.65	20.83***	4.02
L9.airtemp	-0.01*	0.0	0.04	0.04	-0.69***	0.14	-0.40*	0.18	-0.91***	0.23
L14.Ozone	5.5e+06* (0.0055[§])	-2.40e+06 (-0.0024[§])	-3.1e+07 (0.031[§])	-3.20e+07 (0.032[§])	5.3e+07 (0.053[§])	8.50e+07 (0.085[§])	-3.7e+08** (0.37[§])	1.3e+08 (0.13[§])	1.3e+09*** (1.3[§])	1.3e+08 (0.13[§])
L3.Noxide	3.5e+06*** (0.0035[§])	-7.70e+05 (-0.00077[§])	-3.1e+07** (0.031[§])	1.10e+07 (0.011[§])	2.6e+08*** (0.26[§])	3.80e+07 (0.038[§])	-1.0e+08 (0.1[§])	6.7e+07 (0.067[§])	-4.0e+08*** (0.4[§])	7.8e+07 (0.078[§])
L7.humidity	0.00**	0.0	-0.01	0.02	0.18***	0.04	-0.10	0.05	-0.05	0.07
RMSE	0.0016		NA		0.28717		0.22354		0.23074	
EPV	11.43%		1.74%		21.26%		19.02%		16%	
NPV	88.64%		82.35%		93.91%		91.21%		91.86%	
α	90.36%		17.65%		6.09%		8.79%		8.14%	
β	9.58%		43.74%		30.91%		23.03%		38.95%	
Sensitivity	90.42%		56.26%		69.09%		76.97%		61.05%	
Specificity	9.64%		6.02%		65.06%		42.17%		57.83%	
Linktest	0.524									
Pearson GoF			0.8274							
HL GoF: ch2/pvalue			1.81/0.4051							

[Ψ]Expected change in log count for a one-unit increase in variable; OR: * p<0.1 ** p<0.01 *** p<0.001; [§]μm^{-3}

EPV – Extreme predictive value; NPV – Normal predictive value; α - False normal prediction; β - False extreme prediction

DISCUSSION

This study provides tools for developing a health forecast for asthma in London. Quantitative methodologies were developed to study the relationships between asthma episodes and weather/air quality factors. Five different approaches of modelling and predicting normal and extreme asthma events were explored using seasonal, weather and air quality factors. It was identified that asthma daily admissions in London varied widely throughout the year (monthly and seasonally) and was associated with air temperature, humidity, ozone and nitrogen oxide. Developing statistically significant models was unproblematic; however, the practical value of these models for predicting normal and extreme asthma events would need further work and would depend ultimately on the policy goals to be achieved.

Distribution of Parameters

The distribution of daily asthma admissions over the study period appeared to have some monthly and seasonal patterns which we present in Figures 4.2.1a&b. Even though these patterns do not exactly match the patterns described by other studies [96], there are clear seasonal effects. Daily asthma admissions also had some relationship with the distributions of the temperature related measures as well as humidity and a few of the pollutants (Ozone, Nitrogen and PM10), which themselves had regular seasonal and annual patterns (Figure 4.2.2). However, for the wind speed and pollutants like Carbon monoxide and Formaldehyde, there were no noticeable relationships in their patterns and equally no seasonal, monthly or annual trends. These initial observations were important because, they gave a clue on the likely statistical effects and effect sizes of the selected indicators on their respective model output.

In the preliminary bivariate analysis for the selection of suitable individual lag days, it was observed that for asthma episodes, the *dose rate*[1] for exposure to lower temperature-related[2] indicators was lower than for higher temperature-related[3] measures. As a result lower temperature-related measures (e.g. Night minimum temperature) had lesser number of lag days compared to the higher temperature-related measures (e.g. Day maximum temperature), though on the whole, the average air temperature measure was most suitable for modelling (See Table 4.1.0). With the exception of Ozone and Particular matter, which had relatively higher lag days, all the other air pollutants had very low lags in the range of 1-3 days lag (Table 4.1.0).

It is well known that weather factors may modify air quality factors and hence the effects of these modified factors may be different from anticipated [193]. Our proposed models for asthma events, accounted for the most significant weather and air quality measures that reasonably predict asthma. The most significant weather indicators were the 9-day lag for temperature and the 7-day lag for

[1] This term is used to mean the strength of an exposure to independent weather or air quality factor over a period of time

[2] Lower temperature-related measures include: Minimum temperature, Day Minimum temperature, Night minimum temperature, Wet bulb temperature, Dew point temperature

[3] Higher temperature-related measures include: Maximum temperature, Day Maximum temperature, Night maximum temperature

humidity, whilst that for air quality were the 3-day lag for nitrogen oxide and the 14-day lag for ozone. These were significant in predicting asthma events.

Lagged models in time series analysis have consistently been used to describe the patterns of diseases based on previous exposures to effects [194-199]. These models are equally useful in forecasting anticipated disease episodes or trends. Depending on the duration of the lag, the model(s) guarantee a more reliable forecast, since already available data can be used in predicting /forecasting. Hence the choice of variables depends critically on the amount of warning that is desired.

Both lagged and un-lagged indicators were compared in multivariate models to determine the most statistically significant indicator for temperature (see Tables 4.1.1b and 4.1.2b). The same was done for Air pollutants and the output is illustrated in Tables 4.1.3b and 4.1.3d. Lagged predictive models have a relative advantage compared to un-lagged models, and are more useful in predicting events because they make use of reliably measured / known indicators. These are available in databases of daily measures of weather and air quality factors (e.g. AURN datasets).

In this study, we determined and included the best possible lags for temperature, humidity, ozone and nitrogen oxide in our multivariate predictive models. The lag days for the weather and air quality variables of the model ranged from 3 (nitrogen oxide) to 14 (ozone). This implies it is possible to predict asthma events with these models when given at least, a 3-day foreknowledge of data on nitrogen oxide measures (and similarly for the other indicators).

Pulmonary function is sensitive ozone and the resultant effect is also known to increase bronchial hyperresponsiveness [175]. It appeared to be negatively associated with asthma events in two models and positive in the other three models. Some previous studies acknowledged the negative association of ozone with asthma particularly in the winter months and only positive in the summer months [97, 148]. This therefore suggests that a further segregated analysis may provide better parameters for seasonal forecast of asthma events.

Modelling

We used five different models to explore and establish the best predicted asthma daily admissions in London. These included one count model (i.e. a Negative binomial regression model); a logistic regression model and three quantile regression models selected for the 80^{th} and 85^{th} quantiles with a weighted least squares iteration numbers of 278 or 280, interchangeably. The study thus led to the development of testable methodologies that could be used in health forecasting. These methods and tests show that the meteorological and air quality datasets can be statistically matched to asthma daily admission records with multivariate analyses, and hence used in predicting the condition. The approach is consistent with some suggested procedures for modelling disease distributions [16], and could be adapted for health forecasting.

Despite the weakness of the data in terms the relatively short span of the time series (2 year period , the study re-established some of the relationships between asthma episodes and our selected independent factors like temperature, humidity, seasonality, Ozone and Nitrogen oxide exposures that were consistent with the

literature. Significant relationships were observed in all the five proposed statistical models and these are presented in Table 4.6.1.

In the negative binomial regression analysis, we observed significant relationships between daily asthma admissions and all the listed indicators in the model, except for a few responses to the seasonal and monthly (categorical) variables. The Negative binomial regression model has frequently been used in estimating the effect of air quality, meteorological and other factors on hospital admissions for asthma [200-203]. The 3-fold cross validation conducted on this model gave very wide variations. This appears to be an indication of the relatively small sample size of the data used for the purpose or the choice of the 3-fold procedure.

The logistic regression model presents a dichotomised dependent variable in the model. One of the groups represented *extreme* asthma events (i.e. >90th percentile), whilst the other group was defined as *normal*. Unlike most logistic regressions which distinguish between the categories of the dependent variable, the choice of this method was to assess the chances of having an *extreme* event compared to a *normal* one. This procedure of dichotomising continuous variables has some shortfalls including the risks of generating spuriously significant results [192]. However, given the instrumental purpose for which we carried out this procedure, this was not an issue.

Quantile regression was chosen because it is a robust statistical method that makes no assumptions about the distribution of the dependent variable in the population [183]. The analysis revealed that, for at the 80th quantile (QRM1), but for the seasons and ozone, all the predictors in the model were significant. Again, in comparing the two models generated for the 85th quantile with weighted least squares iteration numbers of 280 (QRM2) and 278 (QRM3), the latter model predicted the independent variables better than the former model.

The very wide variation in the coefficients of the predictors observed for all the five different models could be attributed to some inherent deficiencies in the dataset as well as methodological limitation in the modelling. For all the predictions made, it is obvious that there is a trade-off between the effect sizes predicted and the accuracy or error margin (i.e. RMSE and predictive values) of each model. This may have some commercial implications in forecasting.

Strengths

This is unique in so far as it seeks to develop methodologies that could facilitate the forecasting of asthma events resulting in hospitalisation. The study used a wide range of data from different sources and employed methodologies suitably defined by the type, nature and distribution of the dataset used. This was developed in a manner that would make simplify the process of application/validation with multiple datasets

Weakness

In this study there was a need to aggregate the meteorological and air quality indicators. This was done to generate a single shared daily exposure measure per individual/location. One could, therefore, presume that the association of asthma episodes to exposure may have been over or under estimated in some cases.

There are obvious biases in using routinely collected data for which there is no control over data quality checks during data collection. The study may have been limited by some inherent biases of the data due to our inability to validate data quality.

Implications for Policy and Research /Recommendation

The forecasting of meteorological and air quality indicators is important, but even more important to health care providers /users is the forecasting of disease episodes using these measures. Health forecasting services can help in the management of fluctuating demand for health care.

Asthma daily admissions in London are associated with monthly and seasonal effects as well as temperature, humidity, ozone and Nitrogen oxide exposures. Given a fore knowledge of the measure of and exposure (9-day lag temperature; 7-day lag humidity; 14-day lag ozone; 3-day lag nitrogen oxide), it is possible to predict with some degree of certainty, the likely occurrence of extreme/normal asthma events.

Some of the deficiencies identified from these methodologies are amenable to further analysis using different datasets and possibly adding some more indicators that are well known to cause or trigger the onset of asthmatic symptoms. Some of these have been referred to already in the initial discussions (see Figure 1). Future research should examine the additional effect(s) of lightening / thunderstorms and allergens.

Following on from the success of the COPD forecast in the UK [137], further research to sharpen the tools for predicting and forecasting asthma will be of great service to health care providers as both conditions are closely related [34] and often mixed up in their diagnosis

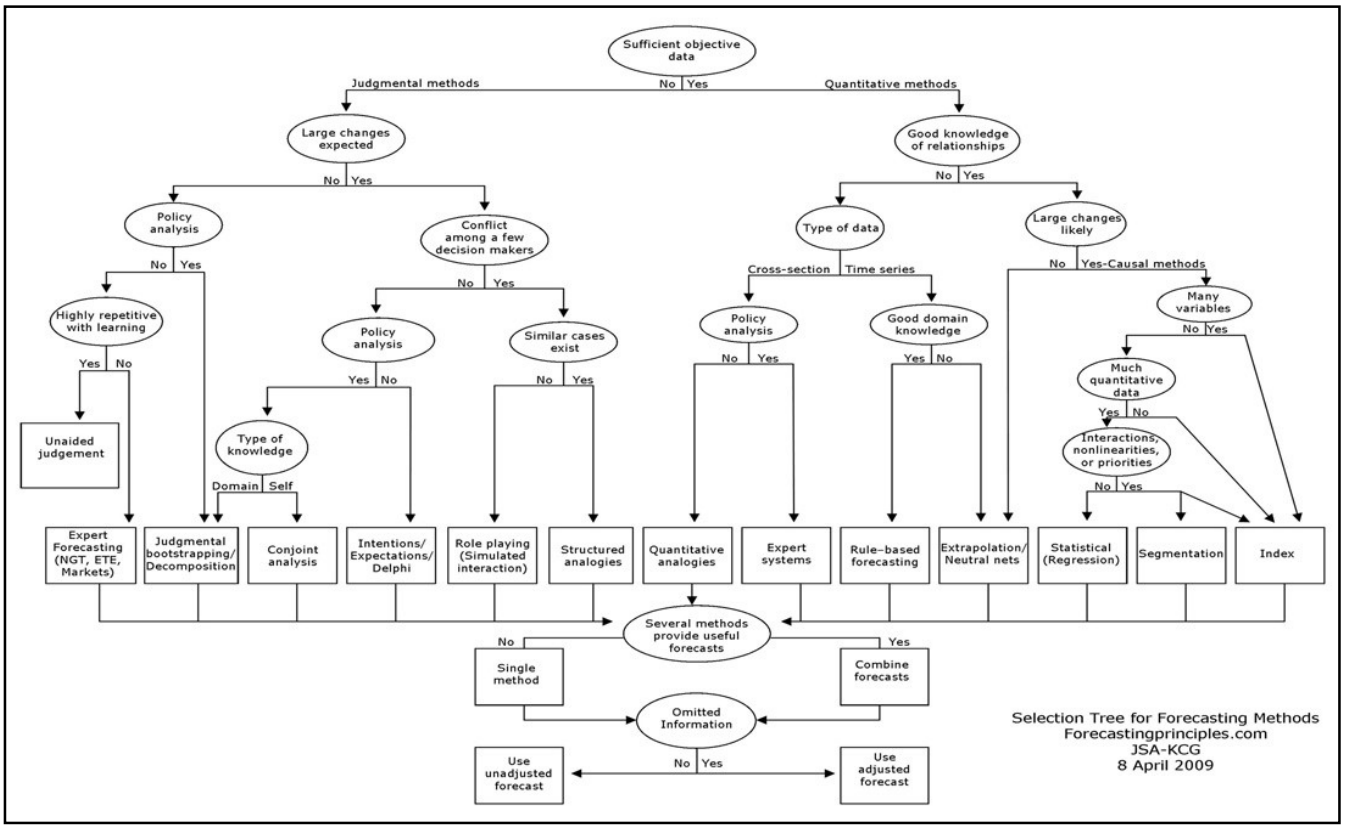
APPENDIX

Appendix A

The Forecast Selection Tree

The *Selection Tree* developed by Armstrong and Green (illustrated below) provides an effective scheme for selecting the best forecasting methods for a problem [125].

Figure 17 The Selection Tree scheme for selecting forecast methods



Appendix B

Synoptic and Climate Stations in London Area

1. Heathrow
2. Northolt
3. Kew Gardens
4. London Weather Centre
5. St. James's Park

Met Office pollution model: NAME dispersion model

Reference / Source: http://www.metoffice.gov.uk/environment/name.html

This world-renowned atmospheric pollution dispersion model is an invaluable and versatile tool for accident and episode analysis, and for pollution forecasting.

NAME can:

- *Forecast air quality*
- *Assess the cause of pollution incidents*
- *Produce long-term impact assessments*
- *Understand and predict long-standing air pollution problems, like acid rain*
- *Forecast the international movement of volcanic ash*

NAME lies at the heart of the Met Office's air quality forecasting system, and is widely used by industry and government to help solve pollution problems.

- *Making unique use of 3D global weather data, NAME is the result of many years of development and includes enhancements in response to the Chernobyl disaster*
- *Applications covered include: plume rise, realistic boundary layer simulation and upper level transport*
- *All spatial scales are catered for, and it includes a powerful suite of diagnostic tools*
- *3D trajectories of air parcels are used to compute air concentrations and ground deposits*

Appendix C

Codebook and data summary

Table 22 Codebook and data summary

Variable in Codebook	Obs	Mean	Std. Dev.	Min	Max
Asthma Daily admissions	730	28.48493	9.846025	6	130
Number of records	730	1097.577	632.7526	5	2191
Date of record	730	16801.5	210.8771	16437	17166
Maximum temperature	706	15.42558	6.941063	-0.02	34.98
Minimum temperature	706	7.832748	5.479739	-4.98	19.26
Night minimum temperature	706	8.143165	5.402302	-4.98	19.26
Night maximum temperature	706	12.24009	6.144914	-1.32	26.5
Day Maximum temperature	706	15.34045	7.001238	-0.36	34.98
Day Minimum temperature	706	10.21478	6.144866	-2.6	25.28
Mean wind speed	706	7.009241	3.002575	0.7	19.5
Air temperature	706	11.22771	6.365106	-2.54	26.48
Wet bulb temperature	706	9.282508	5.498092	-3.25	20.54
Dew point temperature	706	7.271619	5.506778	-6.7	18.1
Barometric vapour pressure	706	10.82407	3.886903	3.425	20.75
Humidity	706	77.88305	12.71152	35	99.5
Carbon monoxide	730	2.45E-07	6.06E-08	1.37E-07	5.23E-07
Formaldehyde	730	6.50E-09	3.25E-09	1.67E-09	1.90E-08
Nitrogen dioxide	730	2.23E-08	7.91E-09	9.24E-09	5.63E-08
Nitrogen oxide	730	1.72E-08	1.14E-08	2.16E-09	7.31E-08
Ozone	730	1.14E-08	5.85E-09	8.48E-10	3.22E-08
Particulate Matter (PM10)	730	1.12E-08	9.00E-09	1.46E-09	6.00E-08
Sulphur dioxide	730	1.26E-08	7.39E-09	2.90E-09	4.28E-08
Night temperature drop*	706	4.096929	2.010862	0.7	10.38
Day temperature drop*	706	5.125673	1.872485	0.86	10.46
Temperature drop*	706	7.592833	3.320577	0.78	17.74
Month*	730	6.526027	3.450215	1	12
Seasonality*	730	2.539726	1.164959	1	4
Dichotomised “asthma”*	730	0.90411	0.294643	0	1

*Derived variables

Appendix D

Asthma Admissions and Temperature (London, 2005-2006)

Figure 18 Asthma Admissions and Mean Air Temperature Distribution (London, 2005-2006)

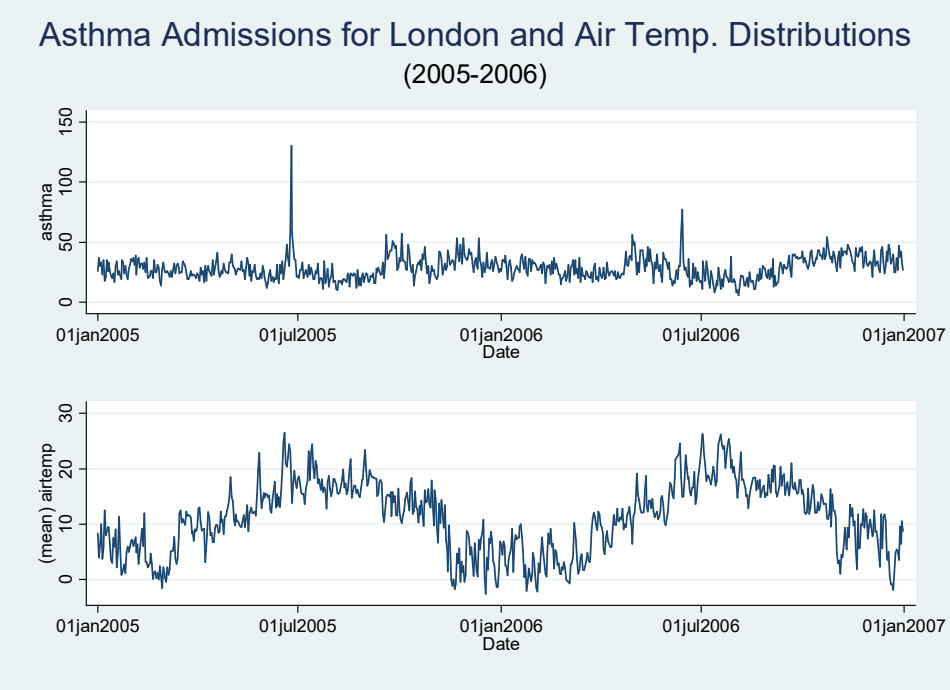


Figure 19 Asthma Admissions and Mean Daily Minimum/ Daily Maximum Temperature Distributions (London, 2005-2006)

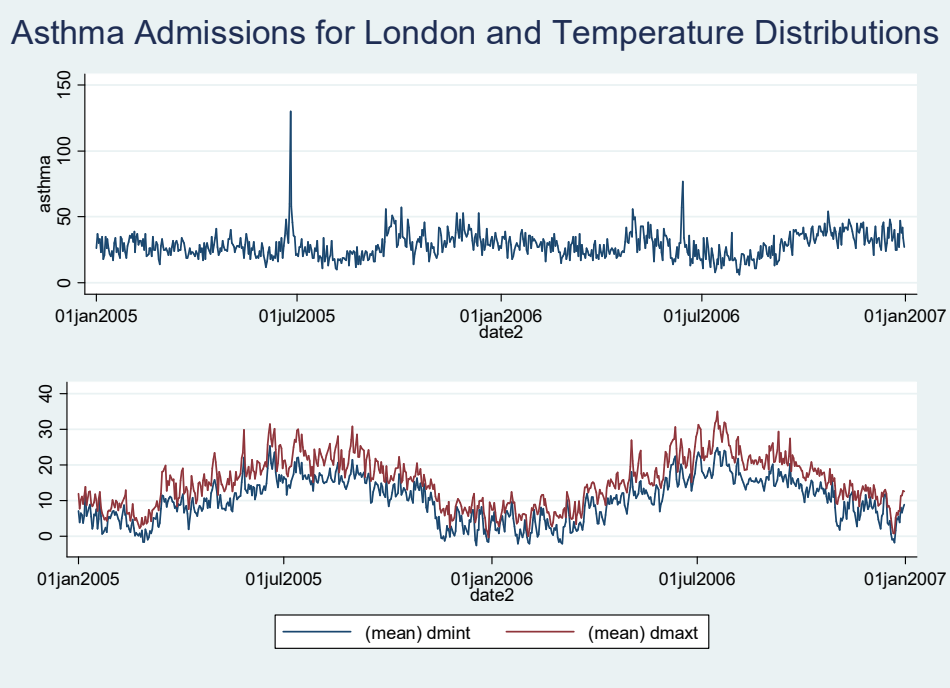


Figure 20 Asthma Admissions and Mean Night Minimum/ Night Maximum Temperature Distributions (London, 2005-2006)

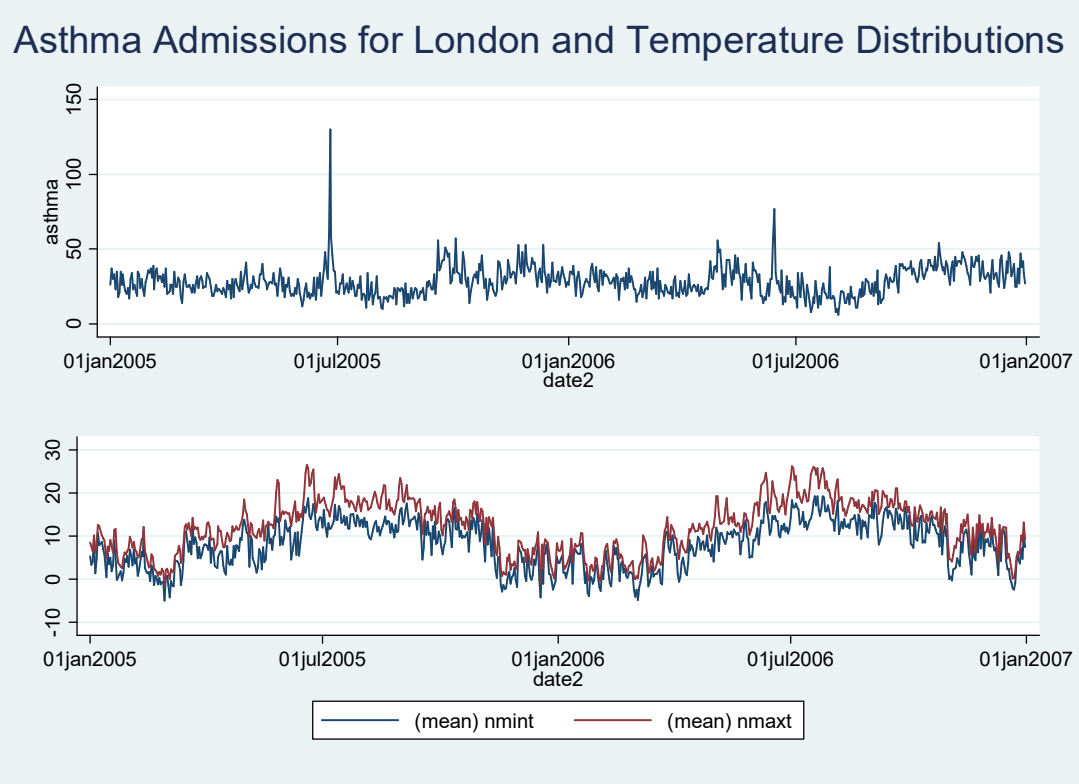


Figure 21 Asthma Admissions and Mean Dew point/ Wet bulb Temperature Distributions (London, 2005-2006)

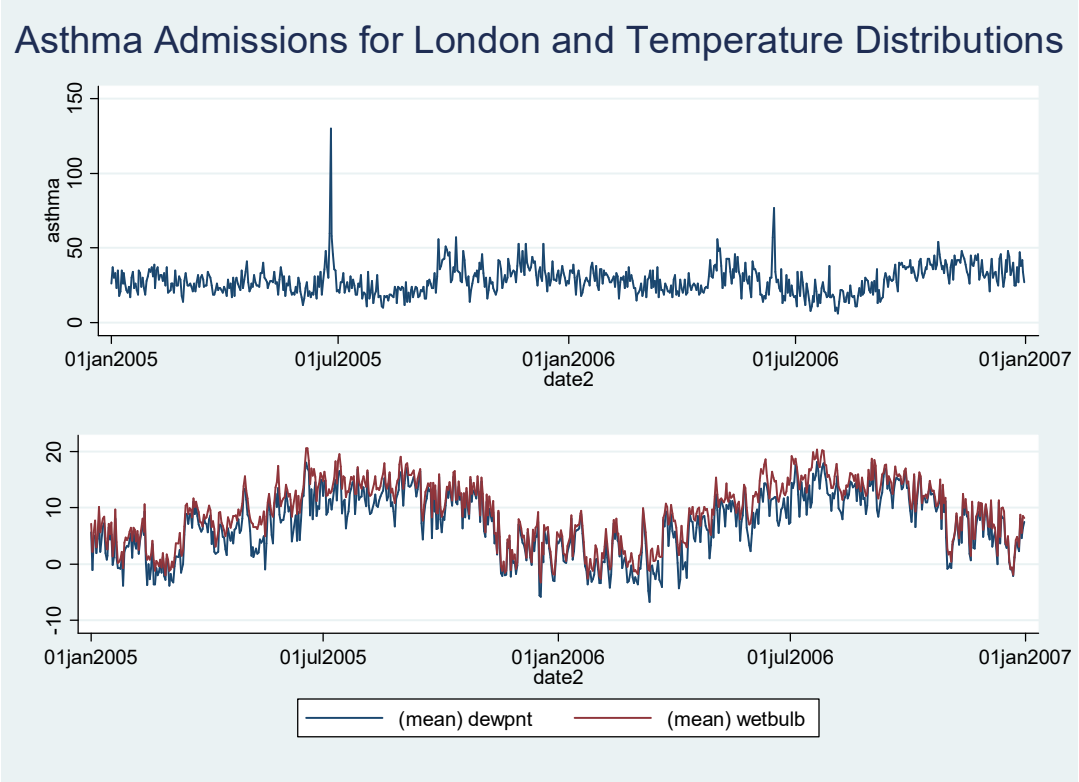
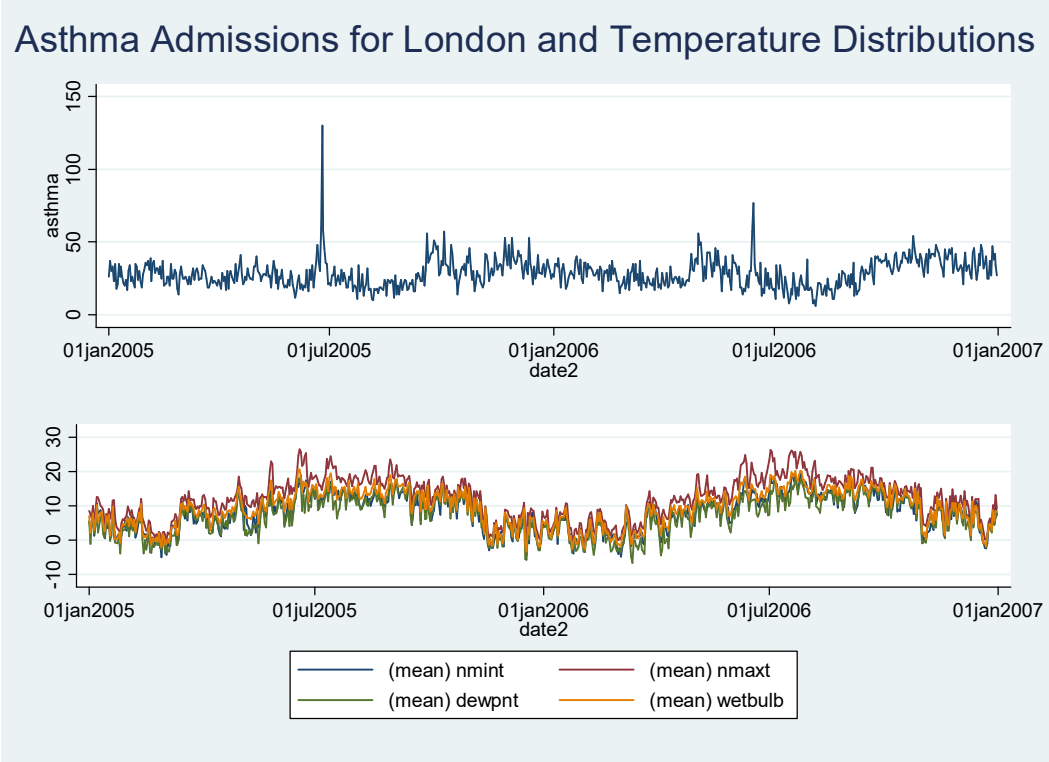


Figure 22 Asthma Admissions and Mean Night Minimum/ Night Maximum and Dew point/ Wet bulb Temperature Distributions (London, 2005-2006)



Asthma Admissions and Air Quality (London, 2005-2006)

Figure 23 Asthma Admissions and NO Distribution (London, 2005-2006)

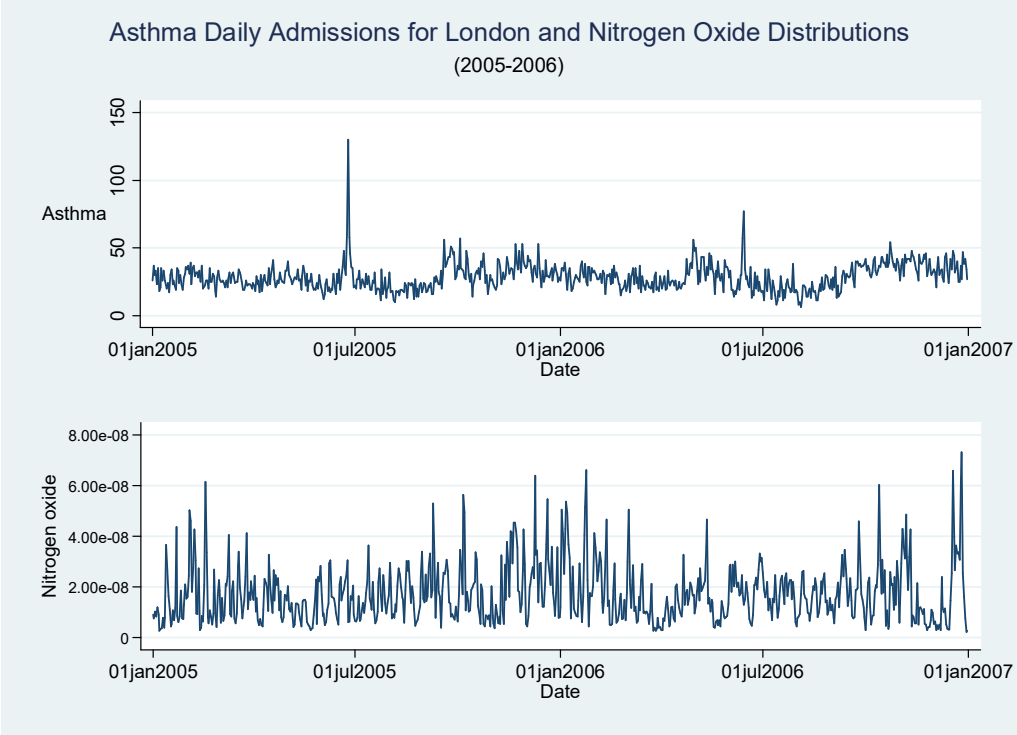


Figure 24 Asthma Admissions and Nitrogen dioxide Distribution (London, 2005-2006)

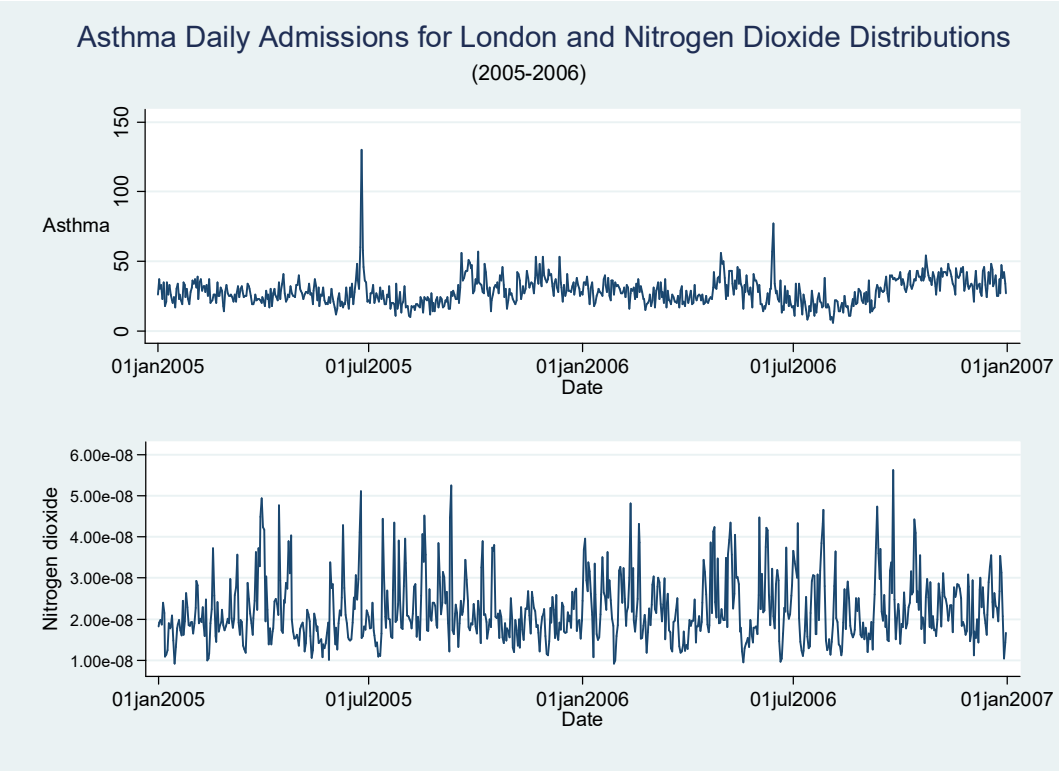


Figure 25 Asthma Admissions and Particulate matter Distribution (London, 2005-2006)

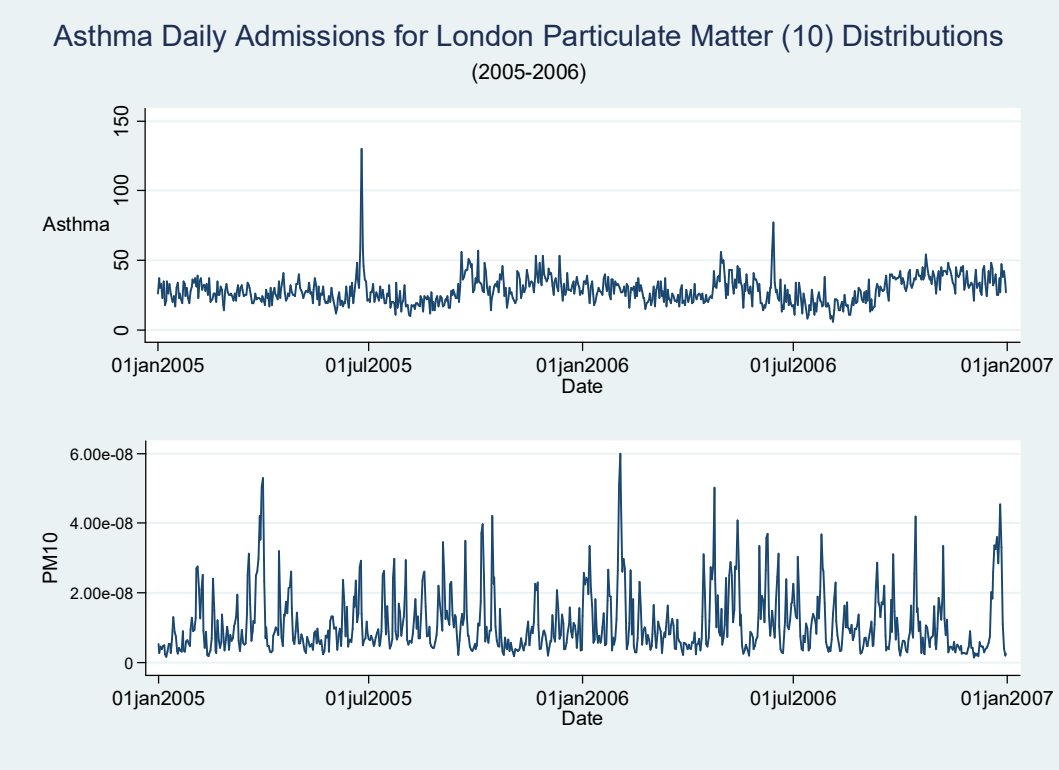


Figure 26 Asthma Admissions and Sulphur dioxide Distribution (London, 2005-2006)

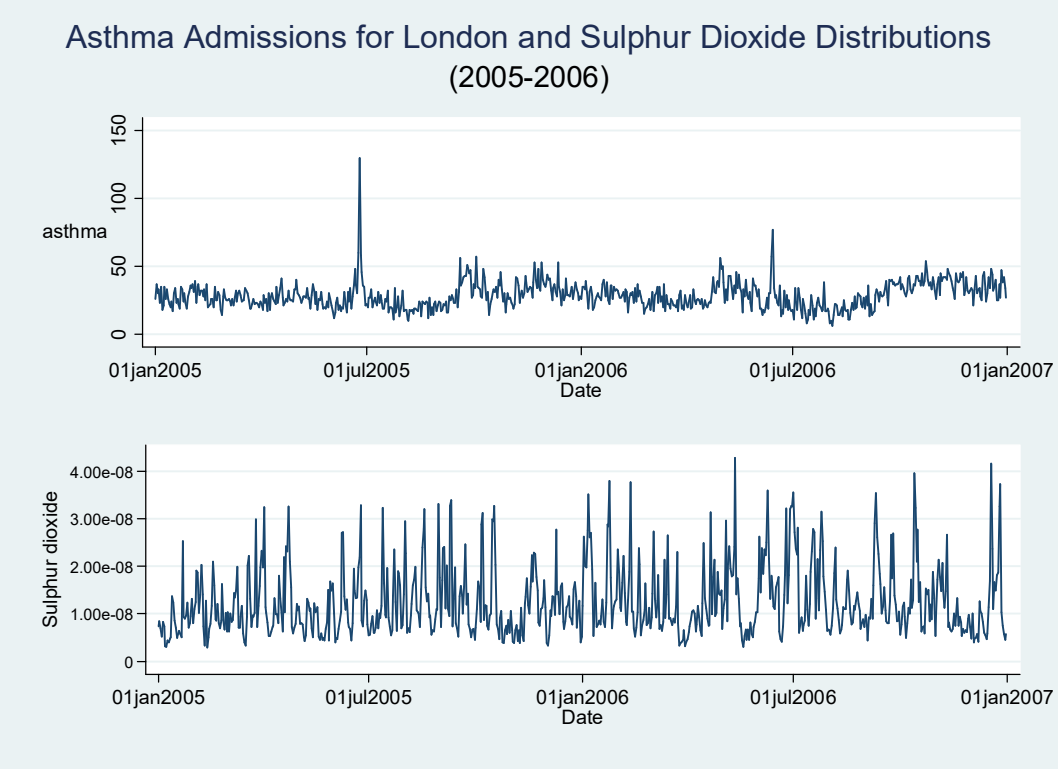


Figure 27 Asthma Admissions and Ozone Distribution (London, 2005-2006)

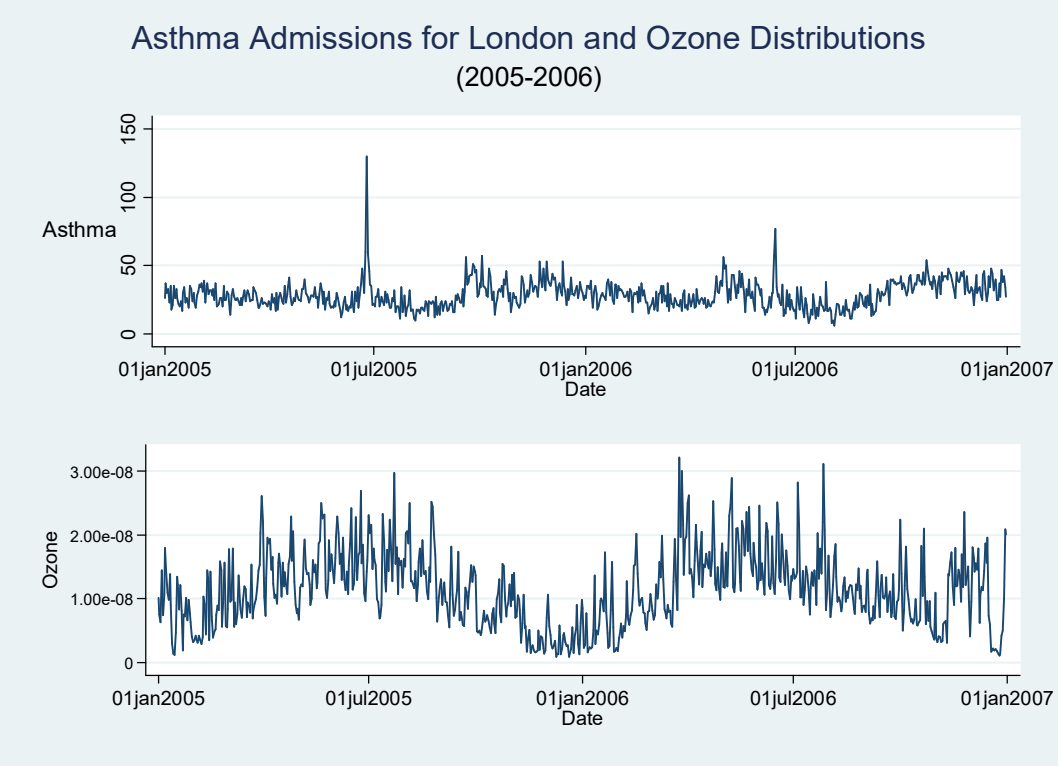


Figure 28 Asthma Admissions and Carbon monoxide Distribution (London, 2005-2006)

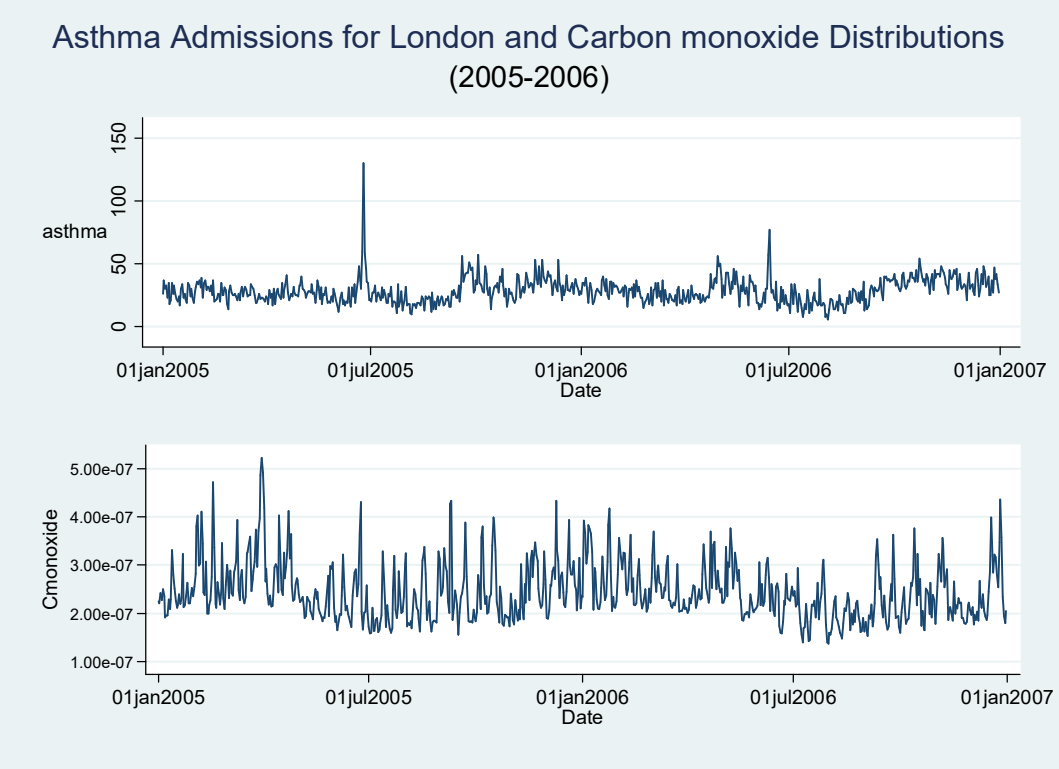


Figure 29 Asthma Admissions and Formaldehyde Distribution (London, 2005-2006)

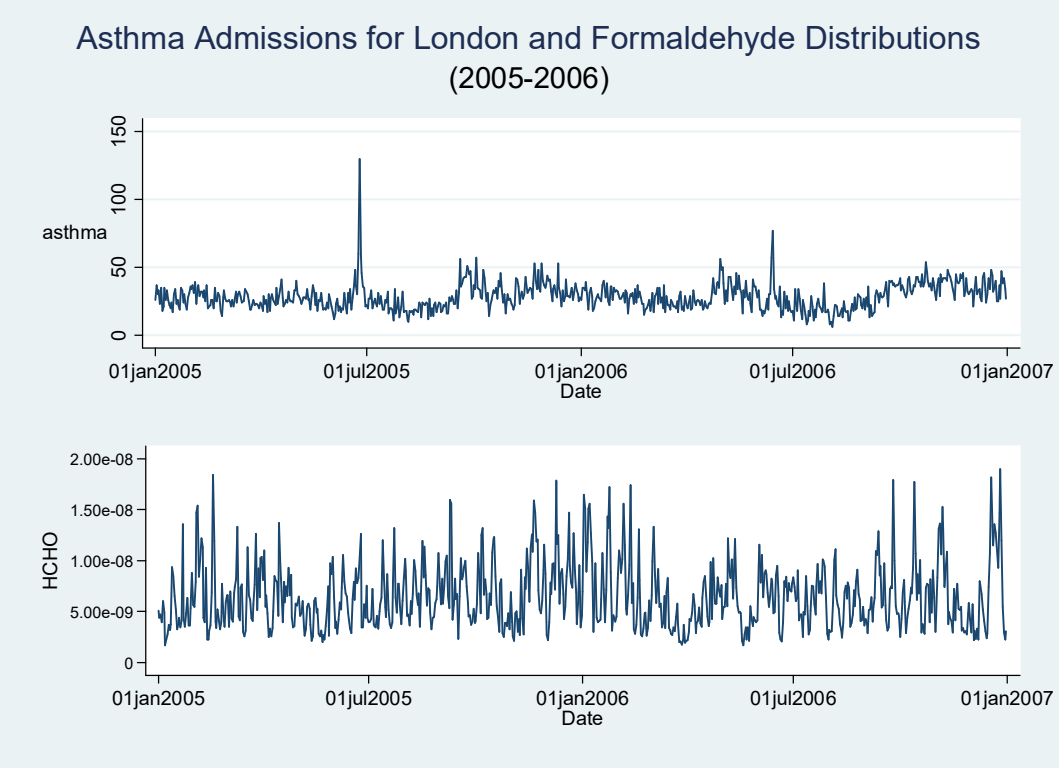


Table 23 Bivariate analysis (lagged independent variables) generated from the NegBin model given that the alpha coefficient of each >0

Variable	Lag Day	Coef.	Std.Err.	z	P|>|z	[95% Conf.	Interval]
Maximum temperature	L15.	-0.00943	0.001851	-5.09	0.000	-0.01306	-0.0058
Minimum temperature	L8.	-0.01004	0.002288	-4.39	0.000	-0.01453	-0.00556
Night minimum temperature	L8.	-0.00878	0.00232	-3.78	0.000	-0.01333	-0.00423
Night maximum temperature	L8.	-0.00942	0.002058	-4.58	0.000	-0.01345	-0.00539
Day Maximum temperature	L15.	-0.0094	0.001836	-5.12	0.000	-0.013	-0.0058
Day Minimum temperature	L9.	-0.01062	0.002046	-5.19	0.000	-0.01463	-0.00661
Night temperature drop	L19.	-0.03297	0.006322	-5.21	0.000	-0.04536	-0.02057
Day Temperature drop	L13.	-0.02492	0.006725	-3.71	0.000	-0.03811	-0.01174
Temperature drop	L19.	-0.01841	0.003818	-4.82	0.000	-0.02589	-0.01093
Mean wind speed	L2.	-0.01145	0.004137	-2.77	0.006	-0.01956	-0.00334
Air temperature	L9.	-0.01032	0.001977	-5.22	0.000	-0.01419	-0.00645
Wet bulb temperature	L9.	-0.00886	0.002285	-3.88	0.000	-0.01334	-0.00438
Dew point temperature	L2.	-0.00601	0.002284	-2.63	0.008	-0.01049	-0.00153
Barometric vapour pressure	L2.	-0.00937	0.003225	-2.9	0.004	-0.01569	-0.00305
Humidity	L7.	0.007207	0.000964	7.48	0.000	0.005318	0.009095
Humidity	L19.	0.008059	0.000976	8.26	0.000	0.006146	0.009972
Carbon monoxide	L1.	969867	199547.2	4.86	0.000	578761.6	1360972
Formaldehyde	L2.	1.85E+07	3656317	5.06	0.000	1.13E+07	2.57E+07
Nitrogen dioxide	L1.	5674741	1526307	3.72	0.000	2683235	8666247
Nitrogen oxide	L3.	4742168	1031354	4.6	0.000	2720752	6763584
Ozone	L14.	-1.31E+07	2116023	-6.2	0.000	-1.73E+07	-8981818
Sulphur dioxide	L3.	4143018	1644753	2.52	0.012	919360.7	7366675
Particulate Matter (PM10)	L2.	2568230	1345142	1.91	0.056	-68198.6	5204659
Particulate Matter (PM10)	L21.	-4248289	1433593	-2.96	0.003	-7058080	-1438497

Negative binomial regression Asthma Models

Table 24 Negative binomial regression Asthma Model output: comparing the temperature related independent variables [alpha>0]

Variable	Coef.	P|>|z	[95% Conf.	Interval]
Maximum temperature	0.031711	0.346	-0.03426	0.097681
Night min. temperature	0.021357	0.157	-0.00821	0.050924
Night max. temperature	0.035668	0.002	0.012706	0.05863
Day max. temperature	-0.0883	0.01	-0.15548	-0.02112
Air temperature	-0.07678	0.025	-0.14391	-0.00965
Dew point temperature	-0.00836	0.77	-0.06431	0.047602
Wet bulb temperature	0.086987	0.157	-0.03337	0.20734
Day temperature drop	0.017915	0.139	-0.00583	0.041662
Temperature drop	0.027737	0.036	0.001878	0.053595

Table 25 Negative binomial regression Asthma Model output: comparing the lagged(L) day temperature related independent variables [alpha>0]

Variable	Coef.	Std. Err.	P>z	[95% C.I.]	
L15.Maximum temperature	0.012215	0.045319	0.788	-0.07661	0.101038
L8.Minimum temperature	-0.01757	0.021748	0.419	-0.06019	0.025058
L8.Night min. temperature	0.010055	0.019787	0.611	-0.02873	0.048835
L8.Night max. temperature	0.02636	0.01163	0.023	0.003566	0.049154
L15.Day max. temperature	-0.01821	0.044994	0.686	-0.10639	0.069978
L9.Day min. temperature	-0.03569	0.018825	0.058	-0.07259	0.001206
L19.Night temperature drop	-0.01449	0.010316	0.16	-0.0347	0.005733
L13.Day temperature drop	-0.00792	0.00814	0.331	-0.02387	0.008034
L19.Temperature drop	-0.00211	0.006273	0.737	-0.0144	0.010183
L9.Air temperature	-0.05285	0.016218	0.001	-0.08464	-0.02107
L9.Wet bulb temperature	0.072636	0.013127	0	0.046907	0.098365
L2.Dew point temperature	0.005769	0.003566	0.106	-0.00122	0.012758

Table 26 Negative binomial regression Asthma Model output: comparing the air pollutant related independent variables [alpha>0]

Variable	Coef.	P>z	[95% Conf.	Interval]
Carbon monoxide	654161.9	0.107	-141899.8	1450224
Aldehyde	9999356	0.491	-1.85E+07	3.85E+07
Nitrogen dioxide	7420196	0.043	247980.1	1.46E+07
Nitrogen oxide	-520672.4	0.9	-8658836	7617491
Ozone	-9726108	0.001	-1.52E+07	-4231523
Sulphur dioxide	-1.17E+07	0.002	-1.91E+07	-4377266
Particulate matter	-2294866	0.287	-6515491	1925760

Table 27 Negative binomial regression Asthma Model output: comparing the lagged(L) day air pollutant related independent variables [alpha>0]

Variable	***Coef.***	***Std. Err.***	***z***	***P>z***	***[95% Conf.***	***Interval]***
L1. Carbon monoxide	348301.3	316345.6	1.1	0.271	-271725	968327.3
L2.Aldehyde (HCHO)	9873557	5499699	1.8	0.073	-905655	2.07E+07
L1.Nitrogen dioxide	3836785	2326530	1.65	0.099	-723131	8396701
L3.Nitrogen oxide	4830577	1925435	2.51	0.012	1056793	8604361
L14.Ozone	-1.01E+07	2190761	-4.63	0.000	-1.44E+07	-5853814
L3.Sulphur dioxide	-2313985	2869959	-0.81	0.42	-7939001	3311032
L2.Particulate matter	-4406147	1957636	-2.25	0.024	-8243043	-569251
L21.Particulate matter	-3003781	1381040	-2.18	0.03	-5710570	-296991

Table 28 Negative binomial regression Asthma predictive Model I: output

Variable	**Coef.**	**P>|z|**	**[95% Conf.**	**Interval]**
Summer_2	0.140692	0.027	0.0163488	0.265036
Autumn_3	0.350927	0.000	0.21291	0.488944
Winter_4	0.315164	0.000	0.1879445	0.442383
Feb_2	-0.05612	0.318	-0.166344	0.054105
Mar_3	0.017414	0.779	-0.104232	0.13906
Apr_4	0.366958	0.000	0.1933843	0.540532
May_5	0.50403	0.000	0.330006	0.678054
Jun_6	0.465243	0.000	0.2893184	0.641167
Jul_7	0.090724	0.363	-0.104808	0.286256
Aug_8	-0.073616	0.45	-0.264809	0.117578
Sep_9	0.266811	0.002	0.1010651	0.432557
Oct_10	0.228422	0.003	0.0758018	0.381043
Nov_11	0.234457	0.000	0.1088046	0.360109
Dec_12	0.125736	0.025	0.016077	0.235394
L9.airtemp	-0.006746	0.041	-0.013204	-0.00029
L14.Ozone	-5487578	0.02	-1.01E+07	-870439
L3.Noxide	3511809	0.000	1683269	5340349
L7.humidity	0.002768	0.009	0.000685	0.004851

Table 29 b: Negative binomial regression Asthma predictive Model I with robust standard errors: output

Variable	**Coef.**	**Robust Std. Err***	**P>|z|**	**[95% Conf. Interval]**	
Summer_2	0.140692	0.132355	0.027	0.0163488	0.265036
Autumn_3	0.350927	0.090399	0.000	0.21291	0.488944
Winter_4	0.315164	0.069943	0.000	0.1879445	0.442383
Feb_2	-0.05612	0.042349	0.318	-0.166344	0.054105
Mar_3	0.017414	0.051614	0.779	-0.104232	0.13906
Apr_4	0.366958	0.090871	0.000	0.1933843	0.540532
May_5	0.50403	0.090036	0.000	0.330006	0.678054
Jun_6	0.465243	0.110563	0.000	0.2893184	0.641167
Jul_7	0.090724	0.105931	0.363	-0.104808	0.286256
Aug_8	-0.073616	0.103272	0.45	-0.264809	0.117578
Sep_9	0.266811	0.085022	0.002	0.1010651	0.432557
Oct_10	0.228422	0.064893	0.003	0.0758018	0.381043
Nov_11	0.234457	0.048614	0.000	0.1088046	0.360109
Dec_12	0.125736	0.042966	0.025	0.016077	0.235394
L9.airtemp	-0.006746	0.00303	0.041	-0.013204	-0.00029
L14.Ozone	-5487578	2315343	0.02	-1.01E+07	-870439
L3.Noxide	3511809	741015.7	0.000	1683269	5340349
L7.humidity	0.002768	0.000967	0.009	0.000685	0.004851

*The robust standard errors attempt to adjust for heterogeneity in the model.

Table 30 Summary Statistics of Actual and Predicted Asthma Admissions for various window periods

Variable	Window Period (Dates)	Obs	Mean	Std. Dev.	Range		RMSE
					Min	Max	
Whole Sample							**$1.6x10^{-3}$**
Asthma admissions (actual)	01/01/05-31/12/06	730	28.48493	9.846025	6	130	
Predicted asthma from actual	01/01/05-31/12/06	668	28.52901	5.490202	16.80289	42.76237	
1st Holdout Sample							**$4.1x10^{-3}$**
Asthma admissions from 1st holdout	01/01/05-01/05/06	486	27.8786	9.347786	10	130	
Predicted asthma from 1st holdout	01/01/05-01/05/06	440	27.96990	5.196354	16.49666	46.23573	
2nd Holdout Sample							**$3.2x10^{-5}$**
Asthma admissions from 2nd holdout	01/09/05-31/12/06	486	29.78807	9.614548	6	77	
Predicted asthma from 2nd holdout	01/09/05-31/12/06	454	29.78737	6.132507	16.64241	46.06879	
3rd Holdout Sample							**$3.4x10^{-3}$**
Asthma admissions from 3rd holdout	02/05/05-31/08/06	486	27.26543	10.49681	6	130	
Predicted asthma from 3rd holdout	02/05/05-31/08/06	454	27.34119	5.70974	15.26246	43.3106	

Quantile Regression Predictive Model(s)

Table 31 Quantile regression output for 0.85 quantile (at wlsin 278)

Variable	Coef.	Std. Err.	P|>|t	[95%Conf.Interval]	
Summer_2	6.065154	4.117885	0.141	-2.02083	14.15114
Autumn_3	-7.13253	4.518804	0.115	-16.0058	1.740716
Winter_4	1.181752	3.53139	0.738	-5.75258	8.116081
Feb_2	8.73403	4.182277	0.037	0.521602	16.94646
Mar_3	-0.32566	4.250593	0.939	-8.67224	8.020915
Apr_4	-7.44991	5.797112	0.199	-18.8333	3.933445
May_5	2.181703	5.610754	0.698	-8.83572	13.19913
Jun_6	4.32394	5.799203	0.456	-7.06353	15.71141
Jul_7	-14.7178	7.305733	0.044	-29.0635	-0.37204
Aug_8	-12.552	7.101053	0.078	-26.4959	1.391769
Sep_9	16.96719	6.622045	0.011	3.963968	29.97041
Oct_10	26.96345	5.425173	0.000	16.31044	37.61647
Nov_11	24.38224	4.622532	0.000	15.30532	33.45916
Dec_12	20.83072	4.01577	0.000	12.94525	28.71619
L9.airtemp	-0.90871	0.22517	0.000	-1.35086	-0.46656
L14.Ozone	1.33E+09	1.32E+08	0.000	1.07E+09	1.59E+09
L3.Noxide	-3.99E+08	7.77E+07	0.000	-5.52E+08	-2.47E+08
L7.humidity	-0.04601	0.072998	0.529	-0.18935	0.097331

Table 32 Quantile regression output for 0.85 quantile (at wlsin 280)

Variable	Coef.	Std. Err.	P|>|t	[95% Conf. Interval]	
Summer_2	2.91201	3.36135	0.387	-3.68843	9.512444
Autumn_3	8.201833	3.375693	0.015	1.573236	14.83043
Winter_4	4.150939	2.224718	0.063	-0.21758	8.519454
Feb_2	-1.58447	3.056581	0.604	-7.58645	4.417511
Mar_3	0.200902	2.751461	0.942	-5.20194	5.603743
Apr_4	13.23104	3.615096	0.000	6.132341	20.32973
May_5	8.080963	4.128435	0.051	-0.02574	16.18767
Jun_6	10.09169	4.370655	0.021	1.509364	18.67403
Jul_7	2.00479	5.179938	0.699	-8.16667	12.17625
Aug_8	-0.29917	4.871955	0.951	-9.86587	9.267524
Sep_9	11.34894	4.738983	0.017	2.043349	20.65453
Oct_10	9.760338	4.54528	0.032	0.835108	18.68557
Nov_11	5.292576	4.265714	0.215	-3.08369	13.66884
Dec_12	-0.62673	3.650765	0.864	-7.79546	6.542012
L9.airtemp	-0.39775	0.180555	0.028	-0.75229	-0.04321
L14.Ozone	-3.74E+08	1.34E+08	0.006	-6.39E+08	-1.10E+08
L3.Noxide	-1.00E+08	6.74E+07	0.138	-2.33E+08	3.21E+07
L7.humidity	-0.09972	0.051796	0.055	-0.20143	0.001989

Table 33 Quantile regression output for 0.80 quantile (at wlsin 280)

Variable	Coef.	Std. Err.	P|>|t	[95% Conf.	Interval]
Summer_2	-1.71133	3.34239	0.609	-8.27453	4.851879
Autumn_3	5.589849	2.967582	0.06	-0.23737	11.41707
Winter_4	2.794592	2.499297	0.264	-2.11309	7.702276
Feb_2	-7.85716	2.514817	0.002	-12.7953	-2.919
Mar_3	-3.56955	2.429087	0.142	-8.33937	1.200264
Apr_4	7.497079	3.31661	0.024	0.984498	14.00966
May_5	10.83918	3.412476	0.002	4.138356	17.54001
Jun_6	9.633121	3.560632	0.007	2.641371	16.62487
Jul_7	8.778227	4.158664	0.035	0.612166	16.94429
Aug_8	18.31678	3.870361	0.000	10.71684	25.91673
Sep_9	12.7591	3.718452	0.001	5.457451	20.06075
Oct_10	8.797581	3.037281	0.004	2.833498	14.76166
Nov_11	11.58352	2.44834	0.000	6.775896	16.39114
Dec_12	2.489721	2.264858	0.272	-1.95761	6.937055
L9.airtemp	-0.68771	0.137153	0.000	-0.95703	-0.41839
L14.Ozone	5.29E+07	8.55E+07	0.536	-1.15E+08	2.21E+08
L3.Noxide	2.60E+08	3.76E+07	0.000	1.86E+08	3.34E+08
L7.humidity	0.183685	0.041725	0.000	0.101752	0.265618

Extreme Asthma Events [Logistic Regression Model]

Table 34 Asthma daily hospital admissions Model: prediction of extreme versus normal events.

Variable	Coef.	Std. Err.	P>z	[95% Conf.	Interval]
Summer_2	-2.004092	0.913941	0.028	-3.79538	-0.2128
Autumn_3	-4.399244	1.27402	0.001	-6.89628	-1.90221
Winter_4	-4.583926	1.368844	0.001	-7.26681	-1.90104
Apr_4	-20.99707	1.463867	0.000	-23.8662	-18.128
May_5	-22.03651	1.413758	0.000	-24.8074	-19.2656
Jun_6	-21.17399	1.505728	0.000	-24.1252	-18.2228
Sep_9	-19.14742	1.853461	0.000	-22.7801	-15.5147
Oct_10	-18.272	1.992764	0.000	-22.1777	-14.3663
Nov_11	-18.53182	1.944342	0.000	-22.3427	-14.721
Dec_12	-17.33444	1.992402	0.000	-21.2395	-13.4294
L9.airtemp	0.0416198	0.041649	0.318	-0.04001	0.123251
L14.Ozone	-3.14E+07	3.20E+07	0.327	-9.42E+07	3.14E+07
L3.Noxide	-3.08E+07	1.14E+07	0.007	-5.32E+07	-8385940
L7.humidity	-0.0122357	0.01547	0.429	-0.04256	0.018084

February, March, July & August dropped from the model because the extreme category (0) predicts success perfectly

Table 35 Asthma daily hospital admissions Model: prediction of extreme versus normal events, with Odds Ratios.

Variable	Odds Ratio	Std. Err.	P>z	[95% Conf.	Interval]
Summer_2	0.134783	0.123183	0.028	0.022474	0.808317
Autumn_3	0.012287	0.015653	0.001	0.001012	0.149238
Winter_4	0.010215	0.013982	0.001	0.000698	0.149413
Apr_4	7.60E-10	1.11E-09	0.000	4.32E-11	1.34E-08
May_5	2.69E-10	3.80E-10	0.000	1.68E-11	4.30E-09
Jun_6	6.37E-10	9.59E-10	0.000	3.33E-11	1.22E-08
Sep_9	4.83E-09	8.96E-09	0.000	1.28E-10	1.83E-07
Oct_10	1.16E-08	2.31E-08	0.000	2.34E-10	5.77E-07
Nov_11	8.95E-09	1.74E-08	0.000	1.98E-10	4.04E-07
Dec_12	2.96E-08	5.90E-08	0.000	5.97E-10	1.47E-06
L9.airtemp	1.042498	0.043419	0.318	0.960779	1.131168
L14.Ozone	0.0000	0.0000	0.327	0.0000	.
L3.Noxide	0.0000	0.0000	0.007	0.0000	0.0000
L7.humidity	0.987839	0.015281	0.429	0.958338	1.018248

February, March, July & August dropped from the model because the extreme category (0) predicts success perfectly

Table 36 Summary of the measures used in estimating RMSE for the three QRMs

Predicted asthma at Quantile	WLSIN	Obs	Mean	Std. Dev.	Min	Max	RMSE
Asthma-actual	-	*730*	*28.48493*	*9.846025*	*6*	*130*	*0*
asthma@85q	278	668	34.71906	12.28903	0.8463322	75.76692	0.23074
asthma@85q	280	668	34.52453	5.762216	17.62017	48.08187	0.22354
asthma@80q	280	668	36.24383	7.217736	16.89369	59.95979	0.28717

REFERENCES

1. GINA. *Global Burden of Asthma, Global Initiative for Asthma (GINA).* 2004 [cited 01/02/2009]; Electronic]. Available from: http://www.ginasthma.com/download.asp?intid=29.
2. DH, *Asthma and Outdoor Air Pollution - COMEAP*, in *Committee on the Medical Effects of Air Pollutants (COMEAP).* 1995, HSMO: Department of Health: London.
3. Taylor, P.E., et al., *Release of allergens as respirable aerosols: A link between grass pollen and asthma.* J Allergy Clin Immunol, 2002. **109**(1): p. 51-6.
4. Burney, P.G., et al., *The effects of allergens in outdoor air on both atopic and nonatopic subjects with airway disease.* Allergy, 2008. **63**(5): p. 542-6.
5. Buckley, D.A., et al., *Atopy and contact allergy to fragrance: allergic reactions to the fragrance mix I (the Larsen mix).* Contact Dermatitis, 2008. **59**(4): p. 220-5.
6. Gupta, R., et al., *Time trends in allergic disorders in the UK.* Thorax, 2007. **62**(1): p. 91-6.
7. Newson, R., et al., *Acute asthma epidemics, weather and pollen in England, 1987-1994.* Eur Respir J, 1998. **11**(3): p. 694-701.
8. Rundell, K.W. and J.B. Slee, *Exercise and other indirect challenges to demonstrate asthma or exercise-induced bronchoconstriction in athletes.* Journal of Allergy and Clinical Immunology, 2008. **122**(2): p. 238-246.
9. Anderson, S.D., et al., *Asthma Provoked by Exercise, Hyperventilation and the Inhalation of Non-isotonic Aerosols*, in *Asthma (Third Edition).* 1998, Academic Press: London. p. 569-587.
10. Kinnula, V.L. and A.R.A. Sovijärvi, *Hyperventilation during exercise: independence on exercise-induced bronchoconstriction in mild asthma.* Respiratory Medicine, 1996. **90**(3): p. 145-151.
11. Woods, S.E., et al., *Young Adults Admitted for Asthma: Does Gender Influence Outcomes?* Journal of Women's Health, 2003. **12**(5): p. 481-485.
12. Almqvist, C., M. Worm, and B. Leynaert, *Impact of gender on asthma in childhood and adolescence: a GA²LEN review.* Allergy, 2008. **63**(1): p. 47-57.
13. Gupta, R.S., et al., *The protective effect of community factors on childhood asthma.* Journal of Allergy and Clinical Immunology, 2009. **123**(6): p. 1297-1304.e2.
14. Neidell, M.J., *Air pollution, health, and socio-economic status: the effect of outdoor air quality on childhood asthma.* Journal of Health Economics, 2004. **23**(6): p. 1209-1236.
15. Bray, M.A., et al., *Written emotional expression as an intervention for asthma.* Psychology in the Schools, 2003. **40**(2): p. 193-207.
16. DH/HPA, *Health effects of climate change in the UK 2008: An update of the Department of Health report 2001/2002*, Ed:-S.Kovats, Editor. 2008, Department of Health, Health Protection Agency
17. GINA. *Global Strategy for Asthma Management and Prevention, Global Initiative for Asthma (GINA).* 2008 [cited 01/02/2009]; Electronic]. Available from: http://www.ginasthma.com/download.asp?intid=29.
18. Kaur, B., et al., *Prevalence of asthma symptoms, diagnosis, and treatment in 12-14 year old children across Great Britain (international study of asthma and allergies in childhood, ISAAC UK).* BMJ, 1998. **316**(7125): p. 118-124.

19. Asthma-UK, *'Where Do We Stand?' Asthma in the UK today* 2004.
20. Pearce, N., et al., *Worldwide trends in the prevalence of asthma symptoms: phase III of the International Study of Asthma and Allergies in Childhood (ISAAC).* Thorax, 2007. **62**(9): p. 758-66.
21. Masoli, M., et al., *The global burden of asthma: executive summary of the GINA Dissemination Committee report.* Allergy, 2004. **59**(5): p. 469-78.
22. Fanta, C.H., *Asthma.* N Engl J Med, 2009. **360**(10): p. 1002-14.
23. Dougherty, R.H. and J.V. Fahy, *Acute exacerbations of asthma: epidemiology, biology and the exacerbation-prone phenotype.* Clin Exp Allergy, 2009. **39**(2): p. 193-202.
24. Panettieri Jr, R.A., et al., *Natural history of asthma: Persistence versus progression--does the beginning predict the end?* Journal of Allergy and Clinical Immunology, 2008. **121**(3): p. 607-613.
25. Fishwick, D. and A.D. Curran, *Variability in the diagnosis of occupational asthma and implications for clinical practice.* Current Opinion in Allergy and Clinical Immunology, 2008. **8**(2): p. 140-144 10.1097/ACI.0b013e3282f60f75.
26. Galant, S.P., et al., *Predictive Value of a Cross-Cultural Asthma Case-Detection Tool in an Elementary School Population.* Pediatrics, 2004. **114**(3): p. e307-316.
27. Gaffin, J.M. and W. Phipatanakul, *The role of indoor allergens in the development of asthma.* Current Opinion in Allergy and Clinical Immunology, 2009. **9**(2): p. 128-135 10.1097/ACI.0b013e32832678b0.
28. British Thoracic Society Scottish Intercollegiate Guidelines, N., *British Guideline on the Management of Asthma.* Thorax, 2008. **63**(Suppl_4): p. iv1-121.
29. Holgate, S.T., *The epidemic of asthma and allergy.* J R Soc Med, 2004. **97**(3): p. 103-10.
30. Holgate, S.T., *Lessons learnt from the epidemic of asthma.* Qjm, 2004. **97**(5): p. 247-57.
31. Nowak, D., C. Suppli Ulrik, and E. von Mutius, *Asthma and atopy: has peak prevalence been reached?* Eur Respir J, 2004. **23**(3): p. 359-360.
32. NAC, *Out in the Open: a true picture of asthma in the United Kingdom today.* The Asthma Journal 2001. **6 (3 Special Supplement) National Asthma Campaign**(6): p. 3-14.
33. Asthma-UK, *'Wish you were here?'* 2008, Asthma UK.
34. Marklund, B., A. Tunsater, and C. Bengtsson, *How often is the diagnosis bronchial asthma correct?* Fam. Pract., 1999. **16**(2): p. 112-116.
35. Welte, T. and D.A. Groneberg, *Asthma and COPD.* Experimental and Toxicologic Pathology, 2006. **57**(Supplement 2): p. 35-40.
36. CDC, *Self-Reported Asthma Prevalence and Control Among Adults--United States, 2001.* JAMA, 2003. **289**(20): p. 2639-2640.
37. Punekar, Y.S. and A. Sheikh, *Establishing the incidence and prevalence of clinician-diagnosed allergic conditions in children and adolescents using routinely collected data from general practices.* Clinical & Experimental Allergy, 2009(Epub ahead of print).
38. Aathma-UK. *Asthma triggers - Weather.* 2009 [cited 01/03/2009]; Available from: http://www.asthma.org.uk/all_about_asthma/asthma_triggers_az/weather.html.

39. Jeffery, P.K., *Structural and inflammatory changes in COPD: a comparison with asthma.* Thorax, 1998. **53**(2): p. 129-36.
40. NIH, *National Asthma Guidelines Updated*, U.S.D.o.H.a.H.S.I.o. Health, Editor. 2007, NIH News.
41. Busse, W.W. and R.F. Lemanske Jr, *Expert Panel Report 3: Moving forward to improve asthma care.* Journal of Allergy and Clinical Immunology, 2007. **120**(5): p. 1012-1014.
42. Holgate, S.T., *The epidemic of allergy and asthma.* Nature, 1999. **402**(6760 Suppl): p. B2-4.
43. Ho, W.C., et al., *Air pollution, weather, and associated risk factors related to asthma prevalence and attack rate.* Environ Res, 2007. **104**(3): p. 402-9.
44. Diette, G.B., et al., *Environmental issues in managing asthma.* Respir Care, 2008. **53**(5): p. 602-15; discussion 616-7.
45. Mortimer, K.M., et al., *The effect of air pollution on inner-city children with asthma.* Eur Respir J, 2002. **19**(4): p. 699-705.
46. Benayoun, L., et al., *Airway Structural Alterations Selectively Associated with Severe Asthma.* Am. J. Respir. Crit. Care Med., 2003. **167**(10): p. 1360-1368.
47. Boulet, L.P. and P.J. Sterk, *Airway remodelling: the future.* Eur Respir J, 2007. **30**(5): p. 831-834.
48. James, A.L. and S. Wenzel, *Clinical relevance of airway remodelling in airway diseases.* Eur Respir J, 2007. **30**(1): p. 134-155.
49. Beasley, R., C. Page, and L. Lichtenstein, *Airway remodelling in asthma.* Clinical & Experimental Allergy Reviews, 2002. **2**(4): p. 109-116.
50. Chung, K.F. and I.M. Adcock, *Pathophysiological mechanisms of asthma - Application of cell and molecular biology techniques.* Molecular Biotechnology, 2001. **Volume 18**(Number 3): p. 213-232.
51. Frey, U., *Predicting asthma control and exacerbations: chronic asthma as a complex dynamic model.* Current Opinion in Allergy and Clinical Immunology, 2007. **7**(3): p. 223-230.
52. Fleming, D.M., et al., *Comparison of the seasonal patterns of asthma identified in general practitioner episodes, hospital admissions, and deaths.* Thorax, 2000. **55**(8): p. 662-5.
53. Campbell, M.J., et al., *Age specific trends in asthma mortality in England and Wales, 1983-95: results of an observational study.* Bmj, 1997. **314**(7092): p. 1439-41.
54. Campbell, M.J., S.T. Holgate, and S.L. Johnston, *Trends in asthma mortality. Data on seasonality of deaths due to asthma were omitted from paper but editorial's author did not know.* Bmj, 1997. **315**(7114): p. 1012.
55. Patino, C.M. and F.D. Martinez, *Interactions between genes and environment in the development of asthma.* Allergy, 2001. **56**(4): p. 279-86.
56. Johnston, N.W. and M.R. Sears, *Asthma exacerbations {middle dot} 1: Epidemiology.* Thorax, 2006. **61**(8): p. 722-728.
57. Ma, L., et al., *Effects of airborne particulate matter on respiratory morbidity in asthmatic children.* Journal of Epidemiology, 2008. **18**(3): p. 97-110.
58. Parker, J.D., L.J. Akinbami, and T.J. Woodruff, *Air Pollution and Childhood Respiratory Allergies in the United States.* Environmental Health Perspectives, 2009. **117**(1): p. 140-147.
59. Walters, S., M. Phupinyokul, and J. Ayres, *Hospital admission rates for asthma and respiratory disease in the West Midlands: their relationship to air pollution levels.* Thorax, 1995. **50**(9): p. 948-54.

60. Lin, M., et al., *Effect of short-term exposure to gaseous pollution on asthma hospitalisation in children: a bi-directional case-crossover analysis.* J Epidemiol Community Health, 2003. **57**(1): p. 50-5.
61. Walters, S., R.K. Griffiths, and J.G. Ayres, *Temporal association between hospital admissions for asthma in Birmingham and ambient levels of sulphur dioxide and smoke.* Thorax, 1994. **49**(2): p. 133-40.
62. Gent, J.F., et al., *Association of low-level ozone and fine particles with respiratory symptoms in children with asthma.* Jama-Journal of the American Medical Association, 2003. **290**(14): p. 1859-1867.
63. GINA. *Global strategy for asthma management and prevention, Global Initiative for Asthma (GINA).* NIH publication no. 02-3659 2002 [cited 04/03/2008]; Electronic].
64. NIH, *Expert Panel Report 3: Guidelines for the Diagnosis and Management of Asthma*, L. National Institutes of Health /National Heart, and Blood Institute, Editor. 2007, NIH Publication No. 07-4051.
65. GINA. *Global Strategy for Asthma Management and Prevention, Global Initiative for Asthma (GINA).* 2006 [cited 01/02/2009]; Electronic]. Available from: http://www.theipcrg.org/resources/nov2006/gina_final_06.pdf.
66. Partridge, M.R., *Examining the unmet need in adults with severe asthma.* EUROPEAN RESPIRATORY REVIEW, 2007. **16**(104): p. 67-72.
67. CDC. *Adult asthma data: prevalence tables and maps.* Centers for Disease Control and Prevention 2007 [cited 19 June 2009]; Available from: http://www.cdc.gov/asthma/brfss/07/current/tableC1.htm.
68. Peters, S.P., et al., *Uncontrolled asthma: A review of the prevalence, disease burden and options for treatment.* Respiratory Medicine, 2006. **100**(7): p. 1139-1151.
69. Asthma-UK. *Living on a Knife Edge.* 2004 [cited 01/05/2009]; Available from: http://www.asthma.org.uk/how_we_help/publishing_reports/.
70. Demoly, P., et al., *Prevalence of asthma control among adults in France, Germany, Italy, Spain and the UK.* EUROPEAN RESPIRATORY REVIEW, 2009. **18**(112): p. 105-112.
71. OHE, *Compendium of Health Statistics*, O.o.H. Economics, Editor. 2004.
72. Hoskins, G., et al., *Risk factors and costs associated with an asthma attack.* Thorax, 2000. **55**(1): p. 19-24.
73. Met-Office. *Climate science explained.* Climate scientists discussants: Pope, V. and Haywood, J. 2009 [cited 01/02/2009]; Available from: http://www.metoffice.gov.uk/climatechange/science/explained/index.html .
74. Lin, S., et al., *Self-Reported Home Environmental Risk Factors for Childhood Asthma: A Cross-Sectional Study of Children in Buffalo, New York.* Journal of Asthma, 2008. **45**(4): p. 325 - 332.
75. Zeka, A., et al., *Risk of upper aerodigestive tract cancers in a case-cohort study of autoworkers exposed to metalworking fluids.* Occup Environ Med, 2004. **61**(5): p. 426-31.
76. Fouillet, A., et al., *Has the impact of heat waves on mortality changed in France since the European heat wave of summer 2003? A study of the 2006 heat wave.* Int J Epidemiol, 2008. **37**(2): p. 309-17.
77. Kovats, R.S. and S. Hajat, *Heat stress and public health: a critical review.* Annu Rev Public Health, 2008. **29**: p. 41-55.

78. Koskela, H.O., *Cold air-provoked respiratory symptoms: the mechanisms and management.* Int J Circumpolar Health, 2007. **66**(2): p. 91-100.
79. O'Shea, S.D., N.F. Taylor, and J.D. Paratz, *...But watch out for the weather: factors affecting adherence to progressive resistance exercise for persons with COPD.* J Cardiopulm Rehabil Prev, 2007. **27**(3): p. 166-74; quiz 175-6.
80. Menne, B., S. Kovats, and J. Bell, *Climate change and health: international research agenda*, in *Draft background paper for the European Preparatory Meeting for Bamako 2008* 2008: Copenhagen.
81. Sauerborn, R., *Climate change: an agenda for research and teaching in public health.* Scand J Public Health, 2007. **35**(6): p. 561-3.
82. Schonwetter, R.S., et al., *Predicting emergency room visits and hospitalizations among hospice patients with cardiac disease.* J Palliat Med, 2008. **11**(8): p. 1142-50.
83. Tobias, A., et al., *Sensitivity analysis of common statistical models used to study the short-term effects of air pollution on health.* Int J Biometeorol, 2003. **47**(4): p. 227-9.
84. IPCC, I.P.o.C.C., *Assessing the health impacts of climate change*, in *IPCC Impacts Assessment WMO/UNEP*. 1995, WMO/UNEP: Geneva (Switzerland).
85. Kim, J., et al., *The weather watch/warning system for stroke and asthma in South Korea.* International Journal of Environmental Health Research, 2008. **18**(2): p. 117-127.
86. Storr, J. and W. Lenney, *School holidays and admissions with asthma.* Arch Dis Child, 1989. **64**(1): p. 103-7.
87. Rossi, O.V., et al., *Association of severe asthma attacks with weather, pollen, and air pollutants.* Thorax, 1993. **48**(3): p. 244-8.
88. Epton, M.J., et al., *Climate and aeroallergen levels in asthma: a 12 month prospective study.* Thorax, 1997. **52**(6): p. 528-34.
89. Hales, S., et al., *Prevalence of adult asthma symptoms in relation to climate in New Zealand.* Environ Health Perspect, 1998. **106**(9): p. 607-10.
90. Weiland, S.K., et al., *Climate and the prevalence of symptoms of asthma, allergic rhinitis, and atopic eczema in children.* Occup Environ Med, 2004. **61**(7): p. 609-15.
91. Chavarria, J.F., *Short report: Asthma admissions and weather conditions in Costa Rica.* Arch Dis Child, 2001. **84**(6): p. 514-5.
92. Ivey, M.A., D.T. Simeon, and M.A. Monteil, *Climatic variables are associated with seasonal acute asthma admissions to accident and emergency room facilities in Trinidad, West Indies.* Clin Exp Allergy, 2003. **33**(11): p. 1526-30.
93. Khot, A., et al., *Biometeorological triggers in childhood asthma.* Clin Allergy, 1988. **18**(4): p. 351-8.
94. Hajat, S. and A. Haines, *Associations of cold temperatures with GP consultations for respiratory and cardiovascular disease amongst the elderly in London.* Int J Epidemiol, 2002. **31**(4): p. 825-30.
95. Priftis, K.N., et al., *Association of weather conditions with childhood admissions for wheezy bronchitis or asthma in Athens.* Respiration, 2006. **73**(6): p. 783-90.
96. HESonline. *Asthma, Health and seasons.* Health and Social Care Information Centre 2009 [cited 2009 01/01/2009]; Available from: http://www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=1937&categoryID=972.

97. Yang, C.-Y., et al., *Air Pollution and Hospital Admissions for Asthma in a Subtropical City: Taipei, Taiwan.* Journal of Toxicology and Environmental Health Part A, 2007. **70**: p. 111-117.
98. Sarnat, J.A. and F. Holguin, *Asthma and air quality.* Curr Opin Pulm Med, 2007. **13**(1): p. 63-6.
99. Babin, S., et al., *Medicaid patient asthma-related acute care visits and their associations with ozone and particulates in Washington, DC, from 1994-2005.* International Journal of Environmental Health Research, 2008. **18**(3): p. 209-221.
100. Lin, M., et al., *Gaseous air pollutants and asthma hospitalization of children with low household income in Vancouver, British Columbia, Canada.* Am J Epidemiol, 2004. **159**(3): p. 294-303.
101. Bartra, J., et al., *Air pollution and allergens.* J Investig Allergol Clin Immunol, 2007. **17 Suppl 2**: p. 3-8.
102. D'Amato, G., et al., *Environmental risk factors and allergic bronchial asthma.* Clin Exp Allergy, 2005. **35**(9): p. 1113-24.
103. von Mutius, E., *The environmental predictors of allergic disease.* Journal of Allergy and Clinical Immunology, 2000. **105**(1, Part 1): p. 9-19.
104. Saxon, A. and D. Diaz-Sanchez, *Air pollution and allergy: you are what you breathe.* Nat Immunol, 2005. **6**(3): p. 223-226.
105. Katsouyanni, K., *Ambient air pollution and health.* Br Med Bull, 2003. **68**(1): p. 143-156.
106. Nicolai, T., et al., *Urban traffic and pollutant exposure related to respiratory outcomes and atopy in a large sample of children.* Eur Respir J, 2003. **21**(6): p. 956-63.
107. Ramadour, M., et al., *Prevalence of asthma and rhinitis in relation to long-term exposure to gaseous air pollutants.* Allergy, 2000. **55**(12): p. 1163-9.
108. Wang, H.C. and E. Yousef, *Air Quality and Pediatric Asthma-Related Emergencies.* Journal of Asthma, 2007. **44**(10): p. 839-841.
109. Annesi-Maesano, I., et al., *Residential proximity fine particles related to allergic sensitisation and asthma in primary school children.* Respir Med, 2007. **101**(8): p. 1721-9.
110. Hajat, S., et al., *Association between Air Pollution and Daily Consultations with General Practitioners for Allergic Rhinitis in London, United Kingdom.* Am. J. Epidemiol., 2001. **153**(7): p. 704-714.
111. Hwang, B.F., et al., *Relation between air pollution and allergic rhinitis in Taiwanese schoolchildren.* Respir Res, 2006. **7**: p. 23.
112. Janssen, N.A., et al., *The relationship between air pollution from heavy traffic and allergic sensitization, bronchial hyperresponsiveness, and respiratory symptoms in Dutch schoolchildren.* Environ Health Perspect, 2003. **111**(12): p. 1512-8.
113. Kramer, U., et al., *Traffic-related air pollution is associated with atopy in children living in urban areas.* Epidemiology, 2000. **11**(1): p. 64-70.
114. Lee, Y.L., et al., *Climate, traffic-related air pollutants and allergic rhinitis prevalence in middle-school children in Taiwan.* Eur Respir J, 2003. **21**(6): p. 964-70.
115. Morgenstern, V., et al., *Respiratory health and individual estimated exposure to traffic-related air pollutants in a cohort of young children.* Occup Environ Med, 2007. **64**(1): p. 8-16.

116. Penard-Morand, C., et al., *Long-term exposure to background air pollution related to respiratory and allergic health in schoolchildren.* Clin Exp Allergy, 2005. **35**(10): p. 1279-87.
117. Wong, T.W., et al., *Associations between daily mortalities from respiratory and cardiovascular diseases and air pollution in Hong Kong, China.* Occup Environ Med, 2002. **59**(1): p. 30-5.
118. Yu, J.H., et al., *The relationship of air pollution to the prevalence of allergic diseases in Taichung and Chu-Shan in 2002.* J Microbiol Immunol Infect, 2005. **38**(2): p. 123-6.
119. Armstrong, J.S., K.C. Green, and W. Soon, *Polar Bear Population Forecasts: A Public-Policy Forecasting Audit.* Marketing Papers, 2008.
120. Armstrong, J.S., F. Callopy, and K.C. Green. *Forecasting Principles* Evidence-based Forecasting 2006 [cited 20/01/2009]; Available from: http://www.forecastingprinciples.com/index.php?option=com_content&task=view&id=3&Itemid=3#field_1.
121. Fildes, R., *The forecasting journals and their contribution to forecasting research: Citation analysis and expert opinion.* International Journal of Forecasting, 2006. **22**(3): p. 415-432.
122. Armstrong, J.S., *Principles of Forecasting: A Handbook for Researchers and Practitioners*. 2001: Kluwer Academic Publishers, Norwell, MA.
123. Blatt, D. *BOOK REVIEW Principles of Forecasting Edited by J. Scott Armstrong*. FUTURECASTS online magazine 2009 [cited 20/05/2009]; Vol. 5, No. 2, 2/1/03.:[Available from: http://www.futurecasts.com/Default.htm.
124. Green, K.C. and J.S. Armstrong, *Structured analogies for forecasting.* International Journal of Forecasting, 2007. **23**(3): p. 365-376.
125. Armstrong, J.S. and K.C. Green. *Selection Tree*. Forecasting Principles - Evidence-based Forecasting 2009 [cited 09/06/2009]; Available from: http://www.forecastingprinciples.com/index.php?option=com_content&task=view&id=17&Itemid=17.
126. Touloumi, G., et al., *Seasonal confounding in air pollution and health time-series studies: effect on air pollution effect estimates.* Stat Med, 2006. **25**(24): p. 4164-78.
127. Katsouyanni, K., et al., *Short term effects of air pollution on health: a European approach using epidemiologic time series data: the APHEA protocol.* J Epidemiol Community Health, 1996. **50 Suppl 1**: p. S12-8.
128. Atkinson, R.W., et al., *Acute effects of particulate air pollution on respiratory admissions: results from APHEA 2 project. Air Pollution and Health: a European Approach.* Am J Respir Crit Care Med, 2001. **164**(10 Pt 1): p. 1860-6.
129. Schwartz, J., et al., *Methodological issues in studies of air pollution and daily counts of deaths or hospital admissions.* J Epidemiol Community Health, 1996. **50 Suppl 1**: p. S3-11.
130. Samet, J.M., et al., *The National Morbidity, Mortality, and Air Pollution Study. Part II: Morbidity and mortality from air pollution in the United States.* Res Rep Health Eff Inst, 2000. **94**(Pt 2): p. 5-70; discussion 71-9.
131. Katsouyanni, K., et al., *Short-term effects of air pollution on health: a European approach using epidemiological time-series data. The APHEA project: background, objectives, design.* Eur Respir J, 1995. **8**(6): p. 1030-8.

132. Samoli, E., et al., *Acute effects of ambient particulate matter on mortality in Europe and North America: results from the APHENA study.* Environ Health Perspect, 2008. **116**(11): p. 1480-6.
133. Horan, J.M. and S. Mallonee, *Injury Surveillance.* Epidemiol Rev, 2003. **25**(1): p. 24-42.
134. Dominici, F., J.I. Levy, and T.A. Louis, *Methodological Challenges and Contributions in Disaster Epidemiology.* Epidemiol Rev, 2005. **27**(1): p. 9-12.
135. Benito, V.F., *Causality and Causal Models: A Conceptual Perspective*.* International Statistical Review, 2006. **74**(3): p. 305-334.
136. Sekhri, N., et al. *Principles for Forecasting Demand for Global Health Products.* Global Health Forecasting Working Group, Background Paper 2006 [cited 01/02/2009]; Available from: http://www.cgdev.org/doc/DemandForecasting/Principles.pdf.
137. Marno, P., et al., *How different measures of cold weather affect chronic obstructive pulmonary disease (COPD) hospital admissions in London.* Eur Respir Rev 2006, 2006. **15**(101): p. 185-186.
138. Soriano, J.B., et al., *Increasing prevalence of asthma in UK primary care during the 1990s.* Int J Tuberc Lung Dis, 2003. **7**(5): p. 415-21.
139. Rees, J., *ABC of asthma. Prevalence.* Bmj, 2005. **331**(7514): p. 443-5.
140. McDonald, J.C., et al., *Incidence by occupation and industry of acute work related respiratory diseases in the UK, 1992-2001.* Occup Environ Med, 2005. **62**(12): p. 836-42.
141. Burney, P., *The changing prevalence of asthma?* Thorax, 2002. **57 Suppl 2**: p. II36-II39.
142. Strachan, D.P., *The epidemiology of childhood asthma.* Allergy, 1999. **54 Suppl 49**: p. 7-11.
143. Court, C.S., D.G. Cook, and D.P. Strachan, *Comparative epidemiology of atopic and non-atopic wheeze and diagnosed asthma in a national sample of English adults.* Thorax, 2002. **57**(11): p. 951-7.
144. Walters, P., M. Ashworth, and A. Tylee, *Ethnic density, physical illness, social deprivation and antidepressant prescribing in primary care: ecological study.* Br J Psychiatry, 2008. **193**(3): p. 235-9.
145. Panico, L., et al., *Ethnic variation in childhood asthma and wheezing illnesses: findings from the Millennium Cohort Study.* Int J Epidemiol, 2007. **36**(5): p. 1093-102.
146. Netuveli, G., B. Hurwitz, and A. Sheikh, *Ethnic variations in incidence of asthma episodes in England & Wales: national study of 502,482 patients in primary care.* Respir Res, 2005. **6**: p. 120.
147. Smeeton, N.C., et al., *Parental attitudes towards the management of asthma in ethnic minorities.* Arch Dis Child, 2007. **92**(12): p. 1082-7.
148. Hajat, S., et al., *Association of air pollution with daily GP consultations for asthma and other lower respiratory conditions in London.* Thorax, 1999. **54**(7): p. 597-605.
149. Calam, R., et al., *Childhood asthma, behavior problems, and family functioning.* J Allergy Clin Immunol, 2003. **112**(3): p. 499-504.
150. HES. *HES User Guide.* Health and Social Care Information Centre 2008 [cited 2009 01/01/2009]; Available from: http://www.hesonline.nhs.uk
151. Hansell, A., et al., *Accessing and using hospital activity data.* J Public Health, 2001. **23**(1): p. 51-56.
152. Peled, J.U., et al., *Do Electronic Health Records Help or Hinder Medical Education?* PLoS Med, 2009. **6**(5): p. e1000069.

153. HESonline. *Hospital Episode Statistics HESonline*. Health and Social Care Information Centre 2009 [cited 2009 01/01/2009]; Available from: http://www.hesonline.nhs.uk
154. Jha, A.K., et al., *Use of Electronic Health Records in U.S. Hospitals.* N Engl J Med, 2009. **360**(16): p. 1628-1638.
155. Westaby, S., et al., *Comparison of hospital episode statistics and central cardiac audit database in public reporting of congenital heart surgery mortality.* BMJ, 2007. **335**(7623): p. 759-.
156. Davies, S., *Summary of other responses.* BMJ, 2007. **335**(7625): p. 839-b-840.
157. Pullen, I. and J. Loudon, *Improving standards in clinical record-keeping.* Adv Psychiatr Treat, 2006. **12**(4): p. 280-286.
158. Met-Office. *Synoptic and climate stations*. 2009 [cited 01/03/09]; Available from: http://www.metoffice.gov.uk/climate/uk/networks/index.html.
159. Met-Office. *Methods and analysis*. Met Office historic UK climate records 2009 [cited 01/03/2009]; Available from: http://www.metoffice.gov.uk/climate/uk/about/methods.html.
160. Met-Office. *NAME III dispersion model* 2009 [cited 01/05/2009]; Available from: http://www.metoffice.gov.uk/environment/name_iii.html.
161. Met-Office. *Changing seasons*. 2009 [cited 26/04/2009]; Available from: http://www.metoffice.gov.uk/climatechange/guide/effects/seasons.html.
162. Eisenberg, D., *The mixed effects of precipitation on traffic crashes.* Accident Analysis & Prevention, 2004. **36**(4): p. 637-647.
163. AEA. *Monitoring Site Environment Type*. UK Air Quality Archive 2009 [cited 25/03/2009]; Available from: http://www.airquality.co.uk/archive/site_type.php.
164. Peacock, J. and S. Kerry, *Presenting Medical Statistics from Proposal to Publication: A Step-by-Step Guide.* 2007, Oxford: Oxford University Press.
165. Bland, J.M., *An Introduction to Medical Statistics* 3rd Edition ed. 2006, Oxford: Oxford University Press.
166. Collett, D., *Modelling Binary Data*. Second Edition ed. Texts in Statistical Science. 2003, London: CHAPMAN & HALL/CRC Press. 387.
167. Hao, L. and D.Q. Naiman, eds. *Quantile Regression*. Quantitative Applications in the Social Sciences. . Vol. 149. 2007, Sage Publications, Inc. 139.
168. Koenker, R. and K.F. Hallock, *Quantile Regression.* The Journal of Economic Perspectives, 2001. **15**(4): p. 143-156.
169. Gujarati, D.N., *Essentials of Econometrics*. Third ed. 2005: McGraw Hill Higher Education.
170. Armstrong, J.S. and F. Collopy, *Error measures for generalizing about forecasting methods: Empirical comparisons : International Journal of Forecasting, 8 (1), 69-80 (June 1992).* Long Range Planning, 1993. **26**(1): p. 150-150.
171. Sistek, D., et al., *Clinical diagnosis of current asthma: predictive value of respiratory symptoms in the SAPALDIA study.* Eur Respir J, 2001. **17**(2): p. 214-219.
172. Altman, D.G. and J.M. Bland, *Statistics Notes: Diagnostic tests 1: sensitivity and specificity.* BMJ, 1994. **308**(6943): p. 1552-.
173. Altman, D.G. and J.M. Bland, *Statistics Notes: Diagnostic tests 3: receiver operating characteristic plots.* BMJ, 1994. **309**(6948): p. 188-.

174. Altman, D.G. and J.M. Bland, *Statistics Notes: Diagnostic tests 2: predictive values.* BMJ, 1994. **309**(6947): p. 102-.
175. Sheppard, L., et al., *Effects of ambient air pollution on nonelderly asthma hospital admissions in Seattle, Washington, 1987-1994.* Epidemiology, 1999. **10**(1): p. 23-30.
176. Hajat, S., et al., *Effects of air pollution on general practitioner consultations for upper respiratory diseases in London.* Occup Environ Med, 2002. **59**(5): p. 294-9.
177. Erbas, B., et al., *Air pollution and childhood asthma emergency hospital admissions: estimating intra-city regional variations.* Int J Environ Health Res, 2005. **15**(1): p. 11-20.
178. Long, J.S., *Predicted probabilities for count models.* Stata Journal, 2001. **1**(1): p. 51-57.
179. Zorn, C. *Evaluating Zero-Inflated and Hurdle Poisson Specifications.* The Society for Political Methodology ad indicia spectate 1998 [cited 01/11/2008]; Working Papers]. Available from: http://polmeth.wustl.edu/workingpapers.php?text=use+of+force&searchkeywords=T&order=dateposted.
180. Anon. *A BRIEF OVERVIEW OF COUNT DATA COMMANDS IN STATA.* 2008 [cited 01/11/2008]; Available from: http://www.sts.uzh.ch/past/hs08/count/stata_glossar.pdf.
181. Alt, J.E., G. King, and C. Signorino, *Aggregation Among Binary, Count, and Duration Models: Estimating the Same Quantities from Different Levels of Data.* Political Analysis, 1999. **Vol. 9**(No. 1(Winter, 2001)): p. Pp. 21-44.
182. Mitchell, S.M. and W.H. Moore. *A New Look at Cold War Presidents' Use of Force: Aggregation Bias, Truncation, and Temporal Dynamic Issues.* The Society for Political Methodology ad indicia spectate 2000 [cited 01/11/2008]; Working Papers]. Available from: http://polmeth.wustl.edu/workingpapers.php?text=use+of+force&searchkeywords=T&order=dateposted.
183. Koenker, R., *Quantile regression.* 2005, New York: Cambridge Press.
184. Taylor, J.W. and D.W. Bunn, *Investigating improvements in the accuracy of prediction intervals for combinations of forecasts: A simulation study.* International Journal of Forecasting, 1999. **15**(3): p. 325-339.
185. Baur, D., M. Saisana, and N. Schulze, *Modelling the effects of meteorological variables on ozone concentration--a quantile regression approach.* Atmospheric Environment, 2004. **38**(28): p. 4689-4699.
186. Barbosa, S.M., *Quantile trends in Baltic sea level.* Geophys. Res. Lett., 2008. **35**.
187. Vaz, S., et al., *Modelling species distributions using regression quantiles.* Journal of Applied Ecology, 2008. **45**(1): p. 204-217.
188. Cade, B.S. and Q. Dong, *A quantile count model of water depth constraints on Cape Sable seaside sparrows.* J Anim Ecol, 2008. **77**(1): p. 47-56.
189. Cade, B.S. and B.R. Noon, *A gentle introduction to quantile regression for ecologists.* Frontiers in Ecology and the Environment, 2003. **1**(8): p. 412-420.
190. Eide, E. and M.H. Showalter, *The effect of school quality on student performance: A quantile regression approach.* Economics Letters, 1998. **58**(3): p. 345-350.

191. Koenker, R. and J.A.F. Machado, *Goodness of Fit and Related Inference Processes for Quantile Regression.* Journal of the American Statistical Association, 1999. **94**(448): p. 1296-1310.
192. Altman, D.G. and P. Royston, *The cost of dichotomising continuous variables.* BMJ, 2006. **332**(7549): p. 1080-.
193. Boutin-Forzano, S., et al., *Visits to the emergency room for asthma attacks and short-term variations in air pollution. A case-crossover study.* Respiration, 2004. **71**(2): p. 134-7.
194. Zanobetti, A., et al., *The temporal pattern of mortality responses to air pollution: a multicity assessment of mortality displacement.* Epidemiology, 2002. **13**(1): p. 87-93.
195. Arbex, M.A., et al., *Air pollution from biomass burning and asthma hospital admissions in a sugar cane plantation area in Brazil.* J Epidemiol Community Health, 2007. **61**(5): p. 395-400.
196. Wilson, A.M., et al., *Air pollution, weather, and respiratory emergency room visits in two northern New England cities: an ecological time-series study.* Environmental Research, 2005. **97**(3): p. 312-321.
197. Lincoln, D., et al., *Childhood asthma and return to school in Sydney, Australia.* Public Health, 2006. **120**(9): p. 854-862.
198. Neuberger, M., D. Rabczenko, and H. Moshammer, *Extended effects of air pollution on cardiopulmonary mortality in Vienna.* Atmospheric Environment, 2007. **41**(38): p. 8549-8556.
199. Leitte, A.M., et al., *Respiratory health, effects of ambient air pollution and its modification by air humidity in Drobeta-Turnu Severin, Romania.* Science of The Total Environment, 2009. **407**(13): p. 4004-4011.
200. Regens, J.L., *Ambient Air Pollution and Daily Pediatric Hospitalizations for Asthma (5 pp).* Environmental Science and Pollution Research, 2007. **14**(1): p. 19-23.
201. Jacquemin, B., et al., *Association between modelled traffic related air pollution (NO2) and asthma score in ECHRS.* Eur Respir J, 2009: p. 09031936.00138208.
202. Grineski, S.E., *Marginalization and health: children's asthma on the Texas-Mexico border.* International Journal of Sociology and Social Policy, 2009. **29** (5/6): p. 287 - 304.
203. Chen, A.Y. and J. Escarce, *Family Structure and the Treatment of Childhood Asthma.* Medical Care, 2008. **46**(2): p. 174-184.