Developing quantitative tools for asthma forecast in London using weather and air quality

Ireneous N. SOYIRI, BSc. (Hons), MPhil

Faculty of Medicine, Nursing and Health Sciences

MONASH University

November, 2012

Dedication

To Esther and Tewededew,

for their love, support and inspiration.

“To God be the glory …”

Acknowledgment

I have come through very though and challenging circumstances in conducting my doctoral research work in three countries, transitioning from one institution to another. It has been a great success largely because of the support I have received from various quarters. I therefore sincerely acknowledge this support. My large extended families and network of friends globally have been instrumental in supporting me through various ways and hence my determination and perseverance. I am also grateful to Professor Pascale Allotey for personal support and motivation even in very challenging situations.

I wish to thank my advisor Professor Daniel Reidpath for his guidance and support throughout my doctoral studies. He provided an all round support to suit my career development, and I have learnt a lot from his in-depth experience in the area of Public Health and also from his practical skills in other areas including statistics. I also appreciate the constructive criticism and comments received from both known and anonymous persons (including evaluators/examiners) about various aspects of my research work and during my candidature. I wish to thank Mr Hassan Yusuf, who assisted with literature search and also read my work at various stages. My colleagues and friends particularly Amos, Nana, Keivan, Ginarthini, and members of Global Public Health have also been of immense help in diverse ways. I appreciate the inspiration and personal advice from all.

I acknowledge the Ghana Government Educational Trust Fund for partly funding my doctoral training, and also to the United Kingdom Met Office, which provided the opportunity and support to conduct part of this study by availing data on weather, air quality and asthma daily admissions. I am equally grateful to the Jeffrey Cheah Foundation, which funded my studies at a very critical period.

Table of contents

List of Figures

List of Tables

List of Equations

List of Acronyms

A&E	Accident and Emergency
AIC	Akaike Information Criterion
ANN	Artificial Neural Network
APHEA	Air Pollution and Health: A European Approach
AQ	Air Quality
ARIMA	Autoregressive Integrated Moving Average
AURN	Automatic Urban and Rural Network
COMEAP	Committee on the Medical Effects of Air Pollutants
COPD	Chronic Obstructive Pulmonary Disease
DH	Department of Health
DLM	Distributed Lag Models
DQI	Data Quality Indicator
EPV	Extreme Predictive Value
GINA	Global Initiative for Asthma
GLM	Generalised Linear Model
GP	General Practitioner
HES	Hospital Episode Statistics
HL	Hosmer-Lemeshow
HPA	Health Protection Agency
ICD	International Classification of Diseases
IPCC	Intergovernmental Panel on Climate Change
LOS	Length of Stay
LRM	Logistic Regression Model
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MSE	Mean Squared Error
NAME	Numerical Atmospheric-dispersion Modelling Environment
NBRM	Negative Binomial Regression Model (or NegBin)
NHS	National Health Service
NMMAPS	National Morbidity, Mortality and Air Pollution Study
NPV	Normal Predictive Value
NYC	New York City
ONS	Office for National Statistics
PACF	Partial Autocorrelation Function
PCT	Primary Care Trusts
PM	Particulate Matter
PMAD	Percent Mean Absolute Deviation
QRM	Quantile Regression Model
RMSE	Root Mean Square Error
ROC	Receiver Operating Characteristic Curve
SARIMA	Seasonal Autoregressive Integrated Moving Average
UK	United Kingdom
WHO	World Health Organization
WLSIN	Weighted Least-Squares Iteration Numbers

Summary

The thesis examines approaches to the forecasting of respiratory events, generally hospital admissions for asthma, but also mortality. The focus of the thesis is forecasting accuracy rather than model specification *per se*. The thesis is a compilation of eight papers (seven published, one "under review"), broken in to four sections, with a brief narrative drawing the themes together.

The topic is introduced with a review of asthma - the main condition examined in the thesis – and the known relationships between environmental conditions and asthma events. An empirical study is then presented that examines the factors affecting length of stay (LOS) in a hospital following an asthma admission. The paper relies on National Health Service (NHS), England data for London from 2001 to 2006. The idea was to demonstrate a burden of disease, as measured in this case by LOS, as a motivation for forecasting asthma events. If there is no consequence for the health system of asthma events, then there may be no point proceeding to the forecasting. Negative binomial regression was used to model the effect(s) of demographic, temporal and diagnostic factors on the LOS, taking into account the cluster effect of each patient's hospital attendance in London. The median and mean asthma LOS over the period of study were 2 and 3 days respectively. Admissions increased over the years from 8,308 (2001) to 10,554 (2006), but LOS consistently declined within the same period. Younger individuals were more likely to be admitted than the elderly, but the latter significantly had higher LOS ($p<0.001$). Respiratory related secondary diagnoses, age, and gender of

the patient as well as day of the week and year of admission were important predictors of LOS.

Having established the burden of asthma on the health system, health forecasting as an approach is introduced in a series of three closely related, published papers. In the first paper a general overview of health forecasting is provided (Soyiri and Reidpath, 2012a). In the second paper, there is a greater emphasis on the specific modelling approaches used in forecasting, and the measures of forecasting accuracy (Soyiri and Reidpath, 2012b). The final published paper in this series introduces in a general sense a "semi structured black-box approach" to forecasting. Two modelling techniques are described Negative Binomial Models for modelling the conditional mean, and Quantile Regression Models for modelling more extreme quantiles; and these are illustrated using London data from 2005-2006 (Soyiri and Reidpath, 2012c).

In the next section of the thesis, four empirical studies are presented, each looking at an approach to health forecasting in greater detail. The first paper examines the use of negative binomial regression to forecast asthma related admissions to London hospitals (2005-2006) using weather and air quality as predictive factors (Soyiri et al, 2013[1]). The data were split in two, with one year's data used for model development and the second years data used for cross validation. Three models were contrasted; a historical average model, a seasonal average model, and a model using selected weather and air quality factors. The seasonal model out performed the historical and the weather and air quality

1 Recently published

models. Given the known causal effect of weather and air quality on asthma, this was somewhat surprising, and led to an alternative approach.

The second paper describes the use of humans as animal sentinels in the forecasting of asthma events (Soyiri and Reidpath, 2012d). In effect, the sensitive lung is "the canary in the coal mine" for the less sensitive lung. Without having to measure any particular environmental trigger or determine the causal relationships between environmental exposures and asthma events, the potential exists to use the frequency of asthma events in the population today to predict the frequency of asthma events in the future. The lungs of the population are seen as "processors of the information" about weather and air quality - avoiding the need to independently estimate the effects. Negative binomial regressions were used in the modelling, allowing for non-contiguous autoregressive components. Selected lags of previous days' admissions were based on partial autocorrelation function (PACF) plot with a maximum lag of 7 days. The model was contrasted with naïve historical and seasonal models. All models were cross validated, with a clear indication of the superiority of the lag - human sentinel - model over the seasonal or historical model.

One of the issues with the previous approaches described here is that they rely on modelling the conditional mean, and yet it is often more useful to be able to forecast a more extreme quantile. Knowing the conditional 90th percentile of asthma admissions for instance provides information about the high end of resources that should be made available. The third paper examines the use of quantile regression to forecast asthma higher than expected numbers of asthma events in London (Soyiri et al. 2012). Appropriate lags of weather and air quality factors were selected, and then pooled to form multivariate predictive models, selected through a systematic backward stepwise reduction approach. Models were cross-validated using a hold-out sample of the data, and their respective root mean square error measures, sensitivity, specificity and predictive values compared. The results indicate that associations between asthma and environmental factors, including temperature, ozone and carbon monoxide can be exploited in predicting future events using quantile regression models. Two criticisms of this paper arose - one during the review process, and one after the review process. The criticism that arose during the review process was that the number of years (2005-2006) was small and it would be better to have more years of data. The second criticism was that the quantile regression approach could be improved upon by using the more unusual quantile regression for count data.

The final paper re-examines the QRM methodology taking account of the two criticisms, a larger dataset was identified that contained 70,830 respiratory related deaths that occurred between 1987-2000 in New York City (Soyiri and Reidpath, Under Review). The models showed improvements of quantile regression models with seasonal and weather/air quality predictors over a seasonal models alone.

Health forecasting is in early stages of development; however, the indications are that relatively simple models may be able to provide information to health systems that will improve service delivery and resource allocation. There remains considerable work to be done in this area both in refining the modelling approaches, and in testing the models in different settings.

References

Soyiri IN, Reidpath DD, Sarran C. (2011). Asthma length of stay in hospitals in London 2001-2006: demographic, diagnostic and temporal factors. PloS One. 6(11):e27184

Soyiri IN, Reidpath DD. (2012a). An overview of health forecasting. Environ Health Prev Med. 2013 [DOI: 10.1007/s12199-012-0294-6]

Soyiri IN, Reidpath DD. (2012b). Evolving forecasting classifications and applications in health forecasting. Int J Gen Med. 5:381-9.

Soyiri IN, Reidpath DD. (2012c). Semistructured black-box prediction: proposed approach for asthma admissions in London. Int J Gen Med. 5:693-705

Soyiri IN, Reidpath DD, Sarran C. (2012). Forecasting peak asthma admissions in London: an application of quantile regression models. Int J Biometeorol. 2012 Aug 12. [DOI: 10.1007/s00484-012-0584-0]

Soyiri IN, Reidpath DD. (2012d). Humans as animal sentinels for forecasting asthma events: helping health services become more responsive. PLoS One. 7(10): e47823.

Soyiri IN, Reidpath DD, Sarran C. Forecasting asthma related hospital admissions in London using negative binomial models. Chronic Respiratory Disease; 2013;10(2):85-94. DOI: 10.1177/1479972313482847.

Soyiri IN, Reidpath DD. (Under Review). The use of quantile regression to forecast higher than expected respiratory deaths in a daily time series: a study of New York City data 1987-2000. PLoS One.

General Declarations

In accordance with Monash University Doctorate Regulation 17/ Doctor of Philosophy and Master of Philosophy (MPhil) regulations the following declarations are made:

I declare that the thesis, except with the Graduate Research Committees approval, contains no material which has been accepted for the award of any other degree or diploma in any university or other institution and affirms that to the best of my knowledge the thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis. In addition to the general declaration presented below, the relative contributions of the other contributing authors have been disclosed at the commencement of each related chapter in the 'Declaration for Thesis Chapter'

Name: **Ireneous N. SOYIRI [[redacted]]**

Sign: **Ireneous N. SOYIRI**

Date: **22 November 2012**

Supervisor: **Professor Dr. Daniel D. Reidpath**

Sign: **Daniel D. Reidpath**

Date: **22 November 2012**

Monash University

Monash Research Graduate School

Declaration for thesis based or partially based on conjointly published or unpublished work

General Declaration

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

This thesis includes 8 original papers, five of which have already been accepted for publication/published as review, methodology or original papers in peer reviewed journals in addition to 3 published peer reviewed conference papers. The core theme of the thesis is to investigate quantitative approaches for health forecasting. The ideas, development and writing up of all the papers in the thesis were the principal responsibility of myself (candidate), working within the Global Public Health Research Strength of the School of Medicine and Health sciences, Monash University Sunway Campus under the supervision of Professor Daniel D. Reidpath.

The inclusion of co-authors reflects the fact that the work came from active collaboration between researchers and acknowledges input into team-based research.

In the case of chapters 3, 4, 5, 6, 7, 8, 9 and 10, as well as the three supplement (conference papers); my contribution to the work involved the following:

Thesis chapter	Publication title	Publication status*	Nature and extent of candidate's contribution
3	Asthma length of stay in hospitals in London 2001-2006: demographic, diagnostic and temporal factors (*PLoS One*, 2011. 6(11): p. e27184)	Published	Lead and corresponding author, conceptualized the idea, accessed, organized and analyzed data, discussed preliminary results with stakeholders, drafted initial manuscript for circulation, submitted manuscript to journal and then managed correspondence with editors/reviewers, editorial staff and publishers until the final publication
Supplement	Determinants of asthma length of stay in London hospitals: individual versus area effects. *Emerg Health Threats J.*, 2011. 4(0): p. 143-143)	Published	Lead and corresponding author, conceptualized the idea, accessed, organized and analyzed data, discussed preliminary results with stakeholders, drafted initial manuscript for circulation, submitted manuscript to journal and then managed correspondence with editors/reviewers, editorial staff and publishers until the final publication
4	Semi-structured black-box prediction: proposed approach for asthma admissions in London (Accepted: 16-07-2012; *Int J Gen Med.*)	Published	Lead and corresponding author, conceptualized the idea, accessed, organized and analyzed literature/data, drafted initial manuscript, submitted manuscript to journal and then managed correspondence with editors/reviewers, editorial staff and publishers until the final publication
Supplement	The role of weather and air quality factors in forecasting asthma admissions in London. (Conference paper: summary available at http://elsevier.conference-	Published	Lead and corresponding author, conceptualized the idea, accessed and organized the literature, drafted initial script, submitted paper to conference review team and then managed correspondence with

Thesis chapter	Publication title	Publication status*	Nature and extent of candidate's contribution
	services.net/programme.asp?conferenceID=2205&action=prog_categories)		editors/reviewers, editorial staff and publishers until the final publication in proceedings
5	An overview of health forecasting (2013; DOI: 10.1007/s12199-012-0294-6, *Env Health Prev Med*.)	Published	Lead and corresponding author, conceptualized the idea, accessed and organized the literature, drafted initial manuscript, submitted manuscript to journal and then managed correspondence with editors/reviewers, editorial staff and publishers until the final publication
6	Evolving forecasting classifications and applications in health forecasting (*Int J Gen Med.* 2012. 5(1): p. 381-9)	Published	Lead and corresponding author, conceptualized the idea, accessed and organized the literature, drafted initial manuscript, submitted manuscript to journal and then managed correspondence with editors/reviewers, editorial staff and publishers until the final publication
7	Forecasting asthma related hospital admissions in London using negative binomial models (*Chron Respir Dis.*; 2013;10(2):85-94. DOI: 10.1177/1479972313482847)	Published	Lead and corresponding author, conceptualized the idea, accessed, organized and analyzed data, discussed preliminary results with stakeholders, drafted initial manuscript for circulation, submitted manuscript to journal and then managed correspondence with editors/reviewers, editorial staff and publishers until the final publication
8	Humans as animal sentinels for forecasting asthma events: helping health services become more responsive. (*PLoS One, 2012*. 7(10): e47823.)	Published	Lead and corresponding author, conceptualized the idea, accessed, organized and analyzed data, discussed preliminary results with co-author, drafted initial manuscript, submitted the final manuscript to journal and then managed correspondence with editors/reviewers, editorial staff and publishers until the final publication
9	Forecasting peak asthma admissions in London: an application of quantile regression models (DOI: 10.1007/s00484-012-0584-0; *Int J Biometeorol.*)	Published	Lead and corresponding author, conceptualized the idea, accessed, organized and analyzed data, discussed preliminary results with stakeholders, drafted initial manuscript for circulation, submitted manuscript to journal and then managed correspondence with editors/reviewers, editorial staff and

Thesis chapter	Publication title	Publication status*	Nature and extent of candidate's contribution
			publishers until the final publication
Supplement	Predicting extreme asthma events in London using quantile regression models (*Emerg Health Threats J.*, 2011. 4:s162: p. 39-40)	Published	Lead and corresponding author, conceptualized the idea, accessed, organized and analyzed data, drafted and submitted conference paper and then managed correspondence with editors/reviewers, editorial staff and publishers until the final publication in proceedings
10	The use of quantile regression to forecast higher than expected respiratory deaths in a daily time series: a study of New York City data 1987-2000	Under review: 19-11-2012	Lead author, conceptualized the idea, accessed, organized and analyzed data, and drafted initial manuscript.

I have / have not (circle that which applies) renumbered sections of submitted or published papers in order to generate a consistent presentation within the thesis.

Signed: Ireneous N. SOYIRI

Date: 22 November 2012

List of contributions from thesis research

Key [○: Peer reviewed conference paper; ■: Peer reviewed journal paper; √: Report]

Review papers:

- ■ Soyiri IN, Reidpath DD: An overview of health forecasting (DOI: 10.1007/s12199-012-0294-6, *Env Health Prev Med*. 2013)
- ■ Soyiri, I.N. and D.D. Reidpath, Evolving forecasting classifications and applications in health forecasting. *Int J Gen Med.* 2012. 5(1): p. 381-9.
- o Soyiri, I.N. and D.D. Reidpath, The role of weather and air quality factors in forecasting asthma admissions in London, in Environmental Health 2011: Resetting our Priorities 2011, Elsevier: Salvador, Brazil. (Available: http://elsevier.conference-services.net/programme.asp?conferenceID=2205&action=prog_categories)

Analytical papers:

- o Soyiri, I., D. Reidpath, and C. Sarran, Determinants of asthma length of stay in London hospitals: individual versus area effects (Published: *Emerg Health Threats J.*, 2011. 4(0): p. 143-143).
- ■ Soyiri, I.N., D.D. Reidpath, and C. Sarran, Asthma length of stay in hospitals in London 2001-2006: demographic, diagnostic and temporal factors (Published: *PLoS One*, 2011. 6(11): p. e27184).
- ■ Soyiri IN, Reidpath DD: Semi-structured black-box prediction: proposed approach for asthma admissions in London. *Int J Gen Med.* 2012. 5(1): p. 693 - 705.

- Soyiri, I., D. Reidpath, and C. Sarran, Forecasting asthma related hospital admissions in London using negative binomial models **(Under review)**
- Soyiri IN, Reidpath DD: Humans as animal sentinels for forecasting asthma events: helping health services become more responsive. *PLoS One, 2012*. 7(10): e47823.
- Soyiri, I., D. Reidpath, and C. Sarran, Forecasting peak asthma admissions in London: an application of quantile regression models (In press: DOI: 10.1007/s00484-012-0584-0; *Int J Biometeorol.*; Accepted: 26-07-2012)

o Soyiri, I.N. and D.D. Reidpath, Predicting extreme asthma events in London using quantile regression models. (Published: *Emerg Health Threats J.*, 2011. 4:s162: p. 39-40).

- Soyiri, I.N. and D.D. Reidpath, The use of quantile regression to forecast higher than expected respiratory deaths in a daily time series: a study of New York City data 1987-2000 **(Under review)**

Reports:

√ Soyiri I.N., **The Met, Forecasting and Public Health**, *A Report on DrPH Internship Study Visit* to the Health Forecasting Team, Met Office Hadley Centre, Fitzroy Road, Exeter, Devon UK [March, 2009]

√ Soyiri I.N., **Developing tools for asthma forecast in London**, A Doctor of Public Health Internship Report: September 2008-June 2009. Health forecasting and Public Health, The Meteorological Office, Hadley Centre, Exeter Devon, UK.

SECTION I

Chapter 1

1.0 General Introduction

Historically humans have been fascinated with the idea of being able to foretell events and their consequences: births, deaths, plagues, and wars. Crystal ball gazing, augury, prognostication and fortune telling all refer to this idea; and until recently, it has been the almost exclusive purview of crackpots, witches, and fair-ground gypsies. The development of statistical and analytic techniques for forecasting has begun to put the process of telling the future on a scientific footing; where even if forecasting is imperfect, the magnitude of error and the degree of performance improvement are open to investigation.

Over the past 50 years, economic, meteorological, and industrial forecasting have developed significantly. However, other areas have lagged behind, and health forecasting, i.e., predicting future health events or situations such as demands for health services and healthcare needs, remain a novel area of forecasting. When developed, health forecasting has the potential to facilitate preventive medicine and preventive care, and public health planning aimed at facilitating health care service provision in populations (1).

Literature on the subject of health forecasting is limited, and the lack of literature is particularly obvious when looking into specific topics /areas, such as the one that is the focus of this thesis, the forecasting of adverse respiratory health events like asthma exacerbations.

Forecasting asthma events has been the focus of limited research - although it has drawn considerable attention in terms of causal modelling, where researchers seek to identify factors that cause asthma events. Simply transposing forecasting approaches without adequate investigation is too ad hoc. Where tools have been used in health forecasting, the approaches do not necessarily share clear, common definitions and they rely on diverse measures to evaluate accuracy, even though the approaches are mostly adapted forms of statistical procedures used in other areas of forecasting (2). There is no single approach to health forecasting, and various methods have been adopted to forecast aggregate or specific health conditions as discussed subsequently. Health forecasting requires reliable data, and suitable analytical tools for the prediction of health conditions or health events. The aim of this thesis is to develop and test methods for health forecasting.

Health systems rely on information and good judgement to anticipate events, to plan and adequately allocate resources for the future. Health forecasting has the potential to provide tools necessary to inform and support the smooth running and provision of health care services.

This thesis investigates the essential principles in forecasting, which have relevance to health, and are commonly used in forecasting various health conditions and /or situations.

It then evaluates a number of statistical methods and forecasting approaches in predicting anticipated and peak/extreme health events.

By virtue of its nature (thesis by compilation) of papers, this report contains many standalone chapters that are the reproduction of articles that needed to be sufficiently detailed and provide enough background to stand alone. The report consists of a number of studies; including reviews and methodology papers as well as five empirical papers, which collectively, are about an investigation of approaches to health forecasting.

1.1 Problem statement

A number of questions or issues can arise from the concept of health forecasting and how it can be used to meet the needs of health services. Some of these, which are important to this thesis include:

1. *What kind of data can be used in developing a health forecast?*
2. *Are the key principles in health forecasting adequately described to guide the process of health forecasting?*
3. *Are there any defined health forecasting horizons (range of period the forecast is intended to cover) to match the methods often used in health forecasting?*
4. *Approaches for determining accuracy and validity not explicitly presented,*
5. *What are the typologies for health forecasting methods?,* and
6. *What are the strengths and weakness of health forecasting techniques?*

Assuming that asthma admissions represent a significant burden to a health system, can expected asthma admissions be forecast?

> To what extent does knowledge of weather and air quality support forecasts of expected asthma admissions?
>
> Is forecasting of expected admissions improved if one assumes that current asthma admissions predict future admissions?

However, even if forecasting proves to be successful, the expected number of admissions (i.e. the mean) may not be as informative for health systems planning as knowing the likely peaks in daily admissions. This leads to further question:

> Can peaks in the number of respiratory related health events be forecast?

In order to attempt to answer some of these questions, using a sample of a large population dataset of individuals hospitalized for asthma in London, an exploration of aspects of the disease burden and the key determinants of its exacerbation was conducted. Also the classical approaches to forecasting and classification of its typologies which may be applicable to health forecasting were revisited.

A number of studies have demonstrated strategies for forecasting chronic respiratory health events using hospital administrative data and environmental data (3-6). These studies have contributed to the management of health delivery services in some respects (7, 8). Earlier approaches have often used aggregate health conditions rather than specific disease approaches (9-11). For example, they may use "attendance at an accident and emergency centre" generally, rather than asthma related attendance. This is an issue because; such a forecast does not provide sufficient information to guide the management of specific health conditions.

Almost without exception, similar approaches have attempted to use averages, such as the daily mean, but have not considered estimating the tails of the distribution. By modelling the tails of the distribution one can determine the occurrence of extreme or peak events which has the advantage of forewarning service delivery, clinical diagnosis and assessments as well as resource allocation.

Whatever approach has been used in the past there is considerable variation in the determination of the validity. To what extent do these approaches vary?

1.2 The aim and main objectives

This thesis aims at developing and evaluating methods for health forecasting specific disease conditions; primarily using data on asthma admissions in London and secondarily, using respiratory related deaths in New York City.

1.2.1 The specific objectives of the work include

1. To investigate the key principles applicable to health forecasting and develop a framework for conducting a health forecasting scheme;
2. To examine the factors that interplay and lead to asthma related health visits and, to also evaluate practical approaches to health forecasting using these factors;

3. To develop a predictive tool for estimating the burden of asthma related to the Length of stay (LOS) during admission/hospitalisation;
4. To develop and evaluate a predictive model for forecasting asthma daily admissions using negative binomial regression models;
5. To evaluate a lag model for forecasting asthma daily admissions based on previous records;
6. To test an approach for forecasting extreme/peak health events using quantile regression models.

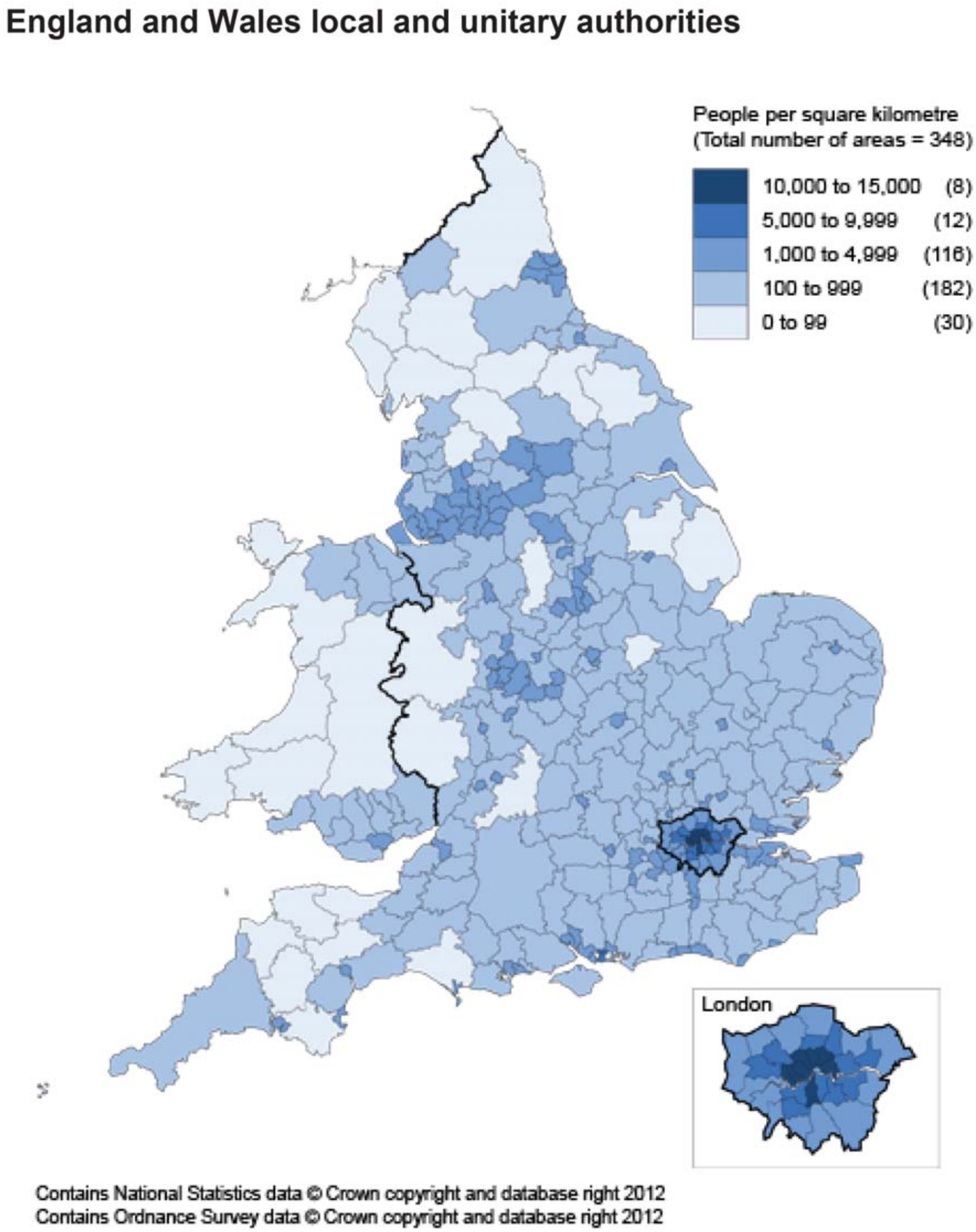
1.3 Study location, data sources and organisation

1.3.1 Main study location: London area

The main study location was London; the region /area bounded by the M25 motorway (Figure 1) which encircles (188 km) Greater London in the United Kingdom. London is the most populous municipality in Europe and among the world's biggest cities. According to a recent population census report released by the Office for National Statistics (ONS), the resident population of London is estimated to be 8.174 million, with a density of 5,200 people per square kilometre (Figure 1) compared to the 321 individuals per square kilometre for the rest of England and Wales (12). The city also has by far the greatest diversity in ethnicity and is a global leader in many areas/sectors including education, research, and health. In London, commercial, industrial and other overbearing human activities have direct or indirect impact on the environment and population health. This situation of London makes it immensely important to public health and surveillance.

London has a publicly funded healthcare system called the National Health Service (NHS) as well as private systems, which are all regulated by the government through the Department of Health. The set up of the health system enables relevant organizations /authorities such as the Hospital Episode Statistics database (HES) to gather and use health data and information.

Figure 1 Map of England and Wales showing local and unitary authorities



Source: Office for National Statistics

1.3.2 Data sources and data organisation

A number of independent datasets were sourced. These included asthma hospital admission data from the Hospital Episode Statistics database sourced through the Met Office; weather and air quality data. Even though these datasets each presented a collection of indicators with a corresponding time (date) record, they were provided as raw data files with little pre-processing. The data were cleaned and organised for analyses. Independent time-series datasets for each of the three sources mentioned above were generated by summarizing and ordering the daily records. These three new time-series datasets were then merged into a single dataset (date-matched) and used in the development of the forecasting models. The Asthma admissions, Weather, and Air quality data sets are each discussed in turn. This process is described in turn for each of the datasets.

1.3.2.1 Asthma emergency hospital 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 emergency hospital admission within London from January 1, 2001 to December 31, 2006[2].

Our operational definition for *Asthma Admission* was any diagnosis with a primary diagnostic ICD-10 code for asthma which is “J 45”. Data from the HES are extracts from

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

routine data flows exchanged between healthcare providers and commissioners via the Secondary Uses Service (13). The data entry and quality checks involved have been described elsewhere (13). Asthma hospital admission in this dataset was indicated by a unique variable, which contains the "anonymised" / "de-identified" 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 (13-20). The HES data is known to have some shortfalls in maternity and psychiatric data for example. A tool, Data Quality Indicator (DQI)[3] is made available, which enables both users and providers of HES data to analyse the data quality at the level of the NHS Trust(13). However, no DQI reports were available for asthma related to our dataset. An issue of generic concern in dealing with data on asthma morbidity has always been the difficulty associated with its diagnosis. Nonetheless, it remains the best available data.

Mindful of these discussions and the nature of some inherent deficiencies in the dataset on asthma hospital admissions, such as difficulty in diagnosis, 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 establishing the disease burden and subsequently predicting / forecasting asthma admissions.

[3] 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

The probability plots (normal distribution) of the key dependent variable (Asthma hospital admissions) and other selected variables in the HES dataset were examined, and checked for outliers as well as the proportion of missing entries. A comprehensive list of the variables in the HES dataset are listed and fully described elsewhere (16). Individuals (patients) were identified by a unique ID number[4] across all data years (13). This unique ID number together with the date enabled us to treat each visit as unique and also identify all repeated number of visits within the study time frame.

We restricted some of the analyses to the available air quality and weather dataset, which only recorded data 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 of all hospitals 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.

1.3.2.2 Meteorological factors

The daily Meteorological factors[5] for all weather monitoring sites (and their respective postcode areas in London), as well as their location coordinates and altitudes were

---

[4] 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.
[5] Meteorological factors: Maximum Temperature (deg C); Minimum Temperature (deg C); Night Minimum Temperature (deg C); Night Maximum Temperature (deg C); Day Maximum Temperature (deg

sourced through the UK Met Office database (Methodology for data collection is described elsewhere[6]). This dataset was matched to all postcodes by their respective nearest (distance) monitoring station postcode.

The key meteorological indicators in this dataset were: *Maximum Temperature* (degrees Celsius), *Minimum Temperature* (degrees Celsius), *Night Minimum Temperature* (degrees Celsius), *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), and *Humidity* (%). These were presented as daily records. The weather stations (Figure 2) 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 observations. Also the daily summaries of the weather measures are however known /referred to as climate observations (21). The detailed description of individual weather elements and how they were quantified over the period has been described (22). Five weather stations within the London area were considered representative (Figure 2 Synoptic and climate stations).

The weather data from each of the five selected weather stations in London (Heathrow, High Wycombe, London Weather Centre, Northolt and South Farnborough) were

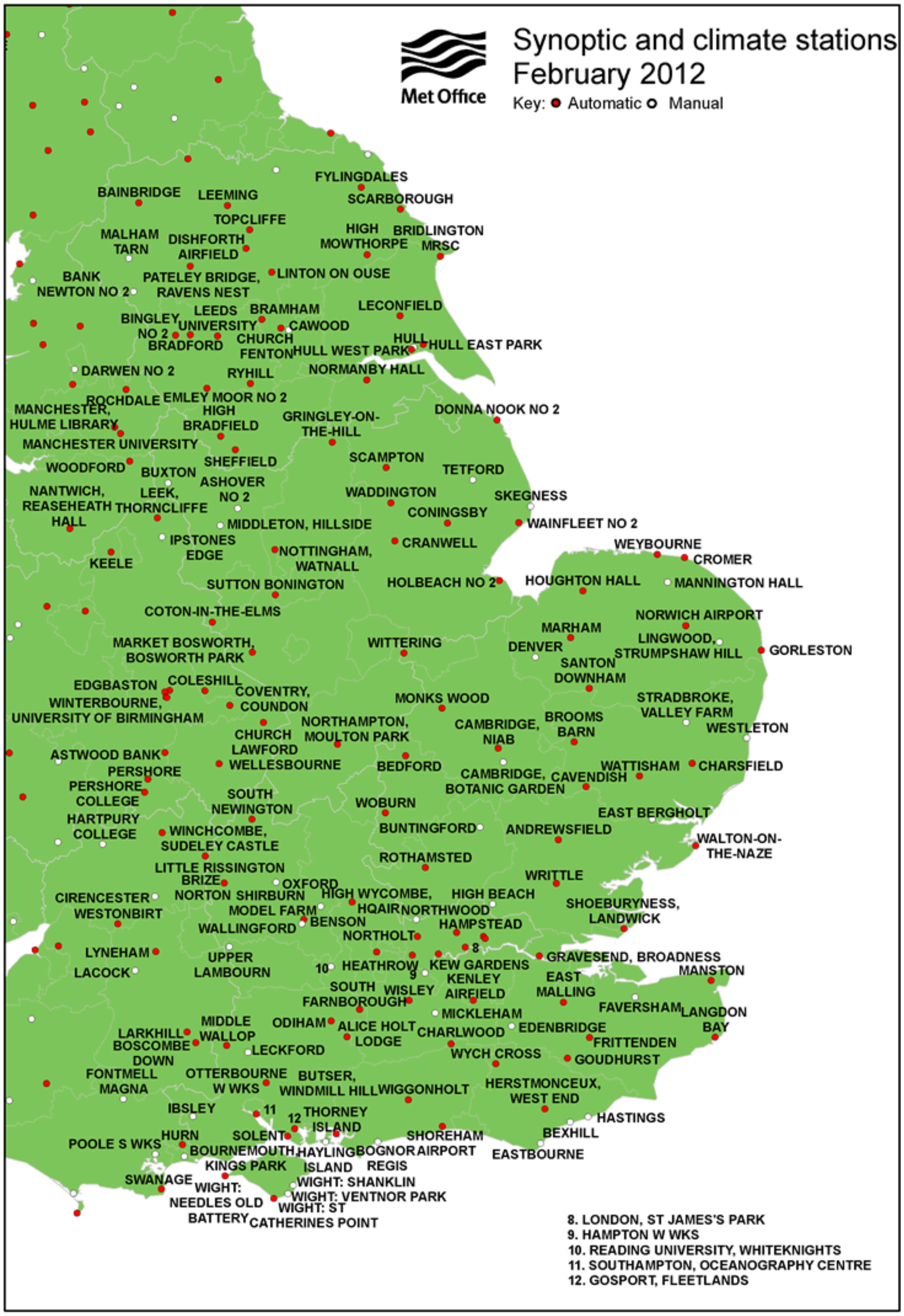
---

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 (%);

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

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 above results in 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., (23)].

Figure 2 Synoptic and climate stations within region 6 of the UK (including the London area)



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

1.3.2.3 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 (2008), about 16 functional sites located across the region (Figure 3 below) provided 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.

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's Numerical Atmospheric-dispersion Modelling Environment (NAME) database for postcode districts (2005-2006), which accounts for both accident and episode analysis, and also used for pollution forecasting (24).

Our analysis was based on the second dataset, from the NAME which consisted of daily estimates of *Carbon monoxide* (kgm^{-3})*, Formaldehyde* (kgm^{-3})*, Nitrogen dioxide* (kgm^{-3})*, Nitrogen oxide* (kgm^{-3})*, Ozone* (kgm^{-3})*, Particulate Matter* [PM_{10}] (kgm^{-3}) and *Sulphur* *dioxide* (kgm^{-3}). These modelled daily air quality estimates[7] do account for both accident and episode analysis (24).

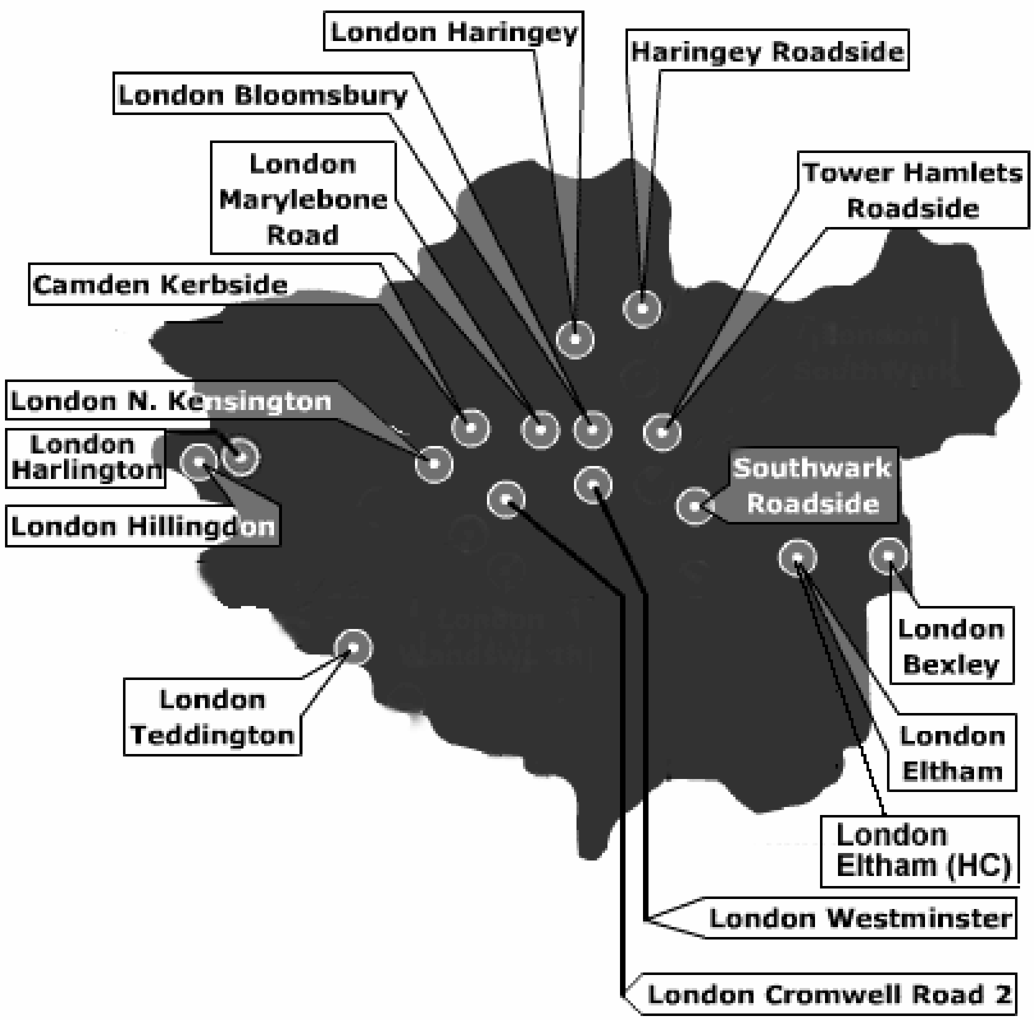
Some other air quality readings, such as $PM_{2.5}$, black smoke etc, that are known to be causally related to asthma events were not available in this dataset. Preliminary investigations carried out to examine the relationship between the patterns of distributions of air quality measures across several 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 are broadly representative of city-wide background conditions (25). Hence we sought to use only recognised "*Urban Background*" measuring stations for the purpose of comparing sites and generalising some area measures.

In the London air quality 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 generated a new representative air quality variable for the entire London region.

[7] 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)

Figure 3 Greater London air quality monitoring sites

Greater London Air Quality Monitoring Sites



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

1.3.2.4 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 "day of the week" and "monthly" variation indicators were generated form the date variable whilst the "seasonal" dummy variable was created by categorizing the days of the year into the four known astronomical seasons (spring, summer, autumn and winter) (26). 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).

1.3.2.5 Creating time series dataset

The use of the term *time series*, in this investigation, refers to a sequence of observations that are ordered in time (1). Three time series datasets were generated from the three datasets described above. In order to create a time series from the asthma admissions data, a process which converts the dataset in memory into a dataset of means or sums, described by the Stata statistical package as *collapse,* was used in producing our time series dataset. This process was repeated for the air quality and weather datasets. The series were then combined by date-linking all the three datasets. The final dataset generated therefore consisted of a *time* variable, which was designated by the date as well as independent variables representing daily averages of air quality and weather measures of the various measuring stations.

The final time series dataset, on which all analyses of hospital asthma admissions were based, comprised the count of daily admission (dependent variable), the averaged daily weather data, and the averaged daily air quality data. This represented a complete dataset with only 3% of missing data (24days) for only humidity and temperature between 1 January 2005 and 31 December 2006.

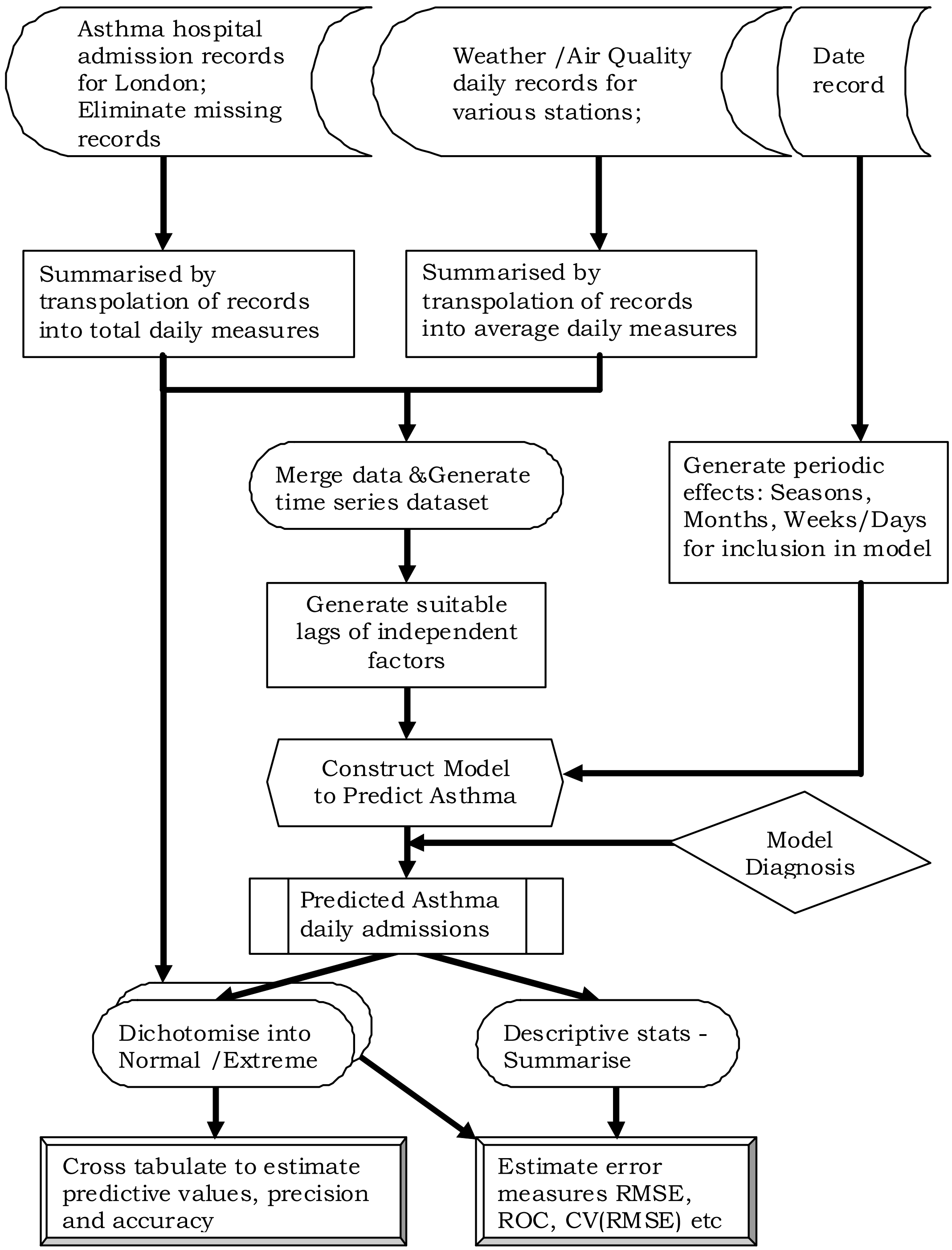
Although the approach taken here to the data analysis is not entirely for the strict purpose of hypothesis testing, nor is it strictly parameter estimation, we nonetheless use the general convention of referring to the count of asthma admissions as the dependent variable; all other variables collectively are called the independent variables. Figure 4 summarises the steps involved data management, modelling and forecasting.

1.3.3 NMMAPS data of New York City (1987-2000)

In order to extend the application of an approach to forecasting extreme/peak health events, we sourced additional data from the National Morbidity, Mortality, and Air Pollution Study (NMMAPS) data of New York City (27). The data are publicly available through the Health and Air Pollution Surveillance System website (http://www.ihapss.jhsph.edu), and, in our case, was accessed using the NMMAPS package in the R statistical environment (28). The daily count of respiratory deaths was the outcome measure of interest. The data included 70,830 respiratory deaths over 5,114 days of surveillance.

The dataset also included a range of daily weather and air quality measures which were used as predictors in the modelling. The predictors included daily mean air temperature, dew point, ozone (O_3), sulphur dioxide (SO_2), nitrogen dioxide (NO_2), and carbon monoxide (CO). Measures of particulate matter were not included because of the levels of missing data. In addition to the measures of weather and air quality, cosinor values representing a yearly and a half yearly cycle (29, 30), and dummy variables representing the days of the week were also used as predictors.

Figure 4 Framework for managing data and developing asthma forecast model(s)



1.4 Preliminary analyses

Tools for describing aspects of the disease burden related to the LOS and for predicting and forecasting hospital admissions for asthma were developed. Generalised Linear Models (GLM), specifically count models (Poisson and negative binomial regression), and Quantile Regression Models (QRM) were used to model the asthma daily admissions. The methods and their reference literature are subsequently discussed.

The preliminary data exploration involved basic descriptive analysis and an examination of the general distribution(s) of the variables. This also included some tests of association between variables. Some of these preliminary exploratory analyses are presented in the appendix.

1.4.1 General distribution of variables, summary statistics by categories

In order to understand the nature of the distribution of the data, the probability plots of the continuous variables were examined. Subsequently, summary statistics were calculated according to the categories of major percentiles in their distribution or known recommended categories.

We also further inspected the probability plots 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 examined variables according to the scales (e.g. interval, ordinal, nominal and dichotomous) and types (e.g. categorical, continuous, ratio, discrete) of data, and on this basis identified the appropriate bivariate statistical tools (31, 32) for preliminary descriptive analysis.

1.4.2 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.

1.4.3 Basic associations of asthma admissions and independent factors

The basic associations of asthma admissions and environmental factors were tested using cross tabulations (with suitably categorized frequencies of variables). Also separate negative binomial models for each exposure variable (uncategorized) were used for bivariate comparisons. These initial /preliminary analyses helped us to understand the strengths of association between asthma daily admissions in London and environmental effects.

1.4.4 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.) (33)

1.4.5 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, the lag properties of independent weather and air quality variables in the dataset were compared. 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.

1.4.6 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 (34), and quantile regression (35, 36). The GLM techniques include a range of statistical linear models, which have non-normal probability distributions, such as the Negative binomial regression. These models correctly fit the data because they do not usually require the variance to be constant / equal to the mean, in hypothesis testing.

1.5 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 /other fits like R^2 or AIC, "identifiability" and parsimony (2, 37). 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 (38). Forecast error can be estimated by a number of methods[8] which are described in later section (39).

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/high demand. In this case one is making a binary forecast: a day of peak/high 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 (40, 41).

[8] 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)

1.5.1 Generalized 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 (34, 42-44). 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, and this is discussed in Chapter 4 (45).

We illustrate this point further using the total number of daily hospital admissions for asthma, which was generated from the HES dataset; these episodes (*count* records of non-negative integers) range 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 ($\boldsymbol{p}^2 = \boldsymbol{u}$) (46). Meanwhile, the predicted rate of daily admissions can be estimated as:

Equation I Rate of admissions

$$\mathbf{Y = Pr(Y_{\mathit{i}} = A \mid E) = exp(E\beta)} \quad (1)$$

Where:

Y	is the predicted rate of daily admissions
E	is the given exposure for *i*
β	is the coefficient of a given exposure measure

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

Equation II Poisson regression

$$\mathbf{Pr(Asthma_{\mathit{i}} = A \mid y) = \frac{e^{-y} y^{A}}{A!}, \qquad A = 0, 1, 2, \ldots} \quad (2)$$

NB: Pr (Asthma$_i$) is the probability of asthma admission for a given day, *i*

Assuming there is no over or under-dispersion (i.e. ρ^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 in the data. Under such circumstances, the negative binomial regression model is preferred (46-50). In a negative binomial model, the probability of observing a specific count of asthma events estimated by:

Equation III Negative binomial model [A]

$$Pr(Y = y|\lambda, \alpha) = \frac{\Gamma(y + \alpha^{-1})}{y!\,\Gamma(\alpha^{-1})}\left(\frac{\alpha^{-1}}{\alpha^{-1} + \lambda}\right)^{\alpha^{-1}}\left(\frac{\lambda}{\alpha^{-1} + \lambda}\right)^{y} \qquad (3)$$

Where:

λ	is the mean of the distribution;
α	is the over dispersion parameter;
y	is the component of the dependant variable represented by the counts of say daily asthma admissions: 0, 1, 2, …;
Γ	is a gamma function.

Alternatively, equation III may also be presented in the form:

Equation IV Negative binomial model [B]

$$P(Y = y_i \mid X_i) = \frac{\Gamma(y_i + 1/\alpha)}{\Gamma(y_i + 1)\Gamma(1/\alpha)} \cdot \left(\frac{1}{1 + \alpha\mu}\right)^{1/\alpha} \cdot \left(\frac{\alpha\mu}{1 + \alpha\mu}\right)^{y_i} \qquad (4)$$

Where:

y_i represents the number of admissions;

μ = $\exp(X_i\beta)$;

β is the vector of coefficients;

X_i is the vector of predictor variables (in this case "1" for the historical model, the dummy variables of three seasons for the seasonal model, and the admissions counts for the lagged days 1, 2, 3, 6 and 7 for the lags model);

α is the overdispersion parameter; and

Γ is a gamma function.

The predictor variable parameters (β) were estimated via maximum likelihood estimation.

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. The preference for the *negative binomial model over* Poisson *model* is largely determined by the value of the dispersion parameter (α). If α is significantly greater than zero ($\alpha>0$) then the negative binomial model is preferred (46, 48). The post estimation tests (*vuong* test and *robust* options in Stata statistical software) provide better estimates of the marginal effects for the standard error terms in the final model (46, 48). The robust standard errors adjust for the heterogeneity in the model, and also provide better estimates of the standard errors 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 (45). The link test (51) is a test of whether the hypothesized link function – in this case a negative binomial link function – is correctly specified in a GLM model. In the Stata software, 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 correctly specified.

1.5.2 Quantile regression

The thesis proposes the use of quantile regression techniques to forecast health events in a time series setting beyond the mean of the outcome of the distribution, i.e., at the high extremes of the data. 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 (35, 52-54). Unlike the linear-regression, quantile regression models have the ability to characterize the relationship between the dependent variable and the independent variable(s) on any quantile especially including the median (50^{th} percentile) or some more extreme quantiles such as the 95^{th} percentile.

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) (55). The corresponding quantile regression model which explains the relationship between the dependent variable, Y can then be expressed as

Equation V Quantile regression model

$$Y_i = \beta_0^{(p)} + \beta_1^{(p)} x_i + \varepsilon_i^{(p)} \tag{5}$$

Where:

Y_i	is asthma hospital admissions for a given day, *i*
$\beta_0^{(p)}$	is a constant term
$\beta_1^{(p)}$	is the coefficient of the exposure term
x_i	is the exposure term
$\varepsilon_i^{(p)}$	is the error tem

Quantile regression techniques have been used for estimating several extreme/peak outcomes and they include modelling the effect of meteorological factors on some environmental pollutants (56), describing the sea level trends at different tides (57), modelling the factors that affect ecological processes (58-60), and even the effect of school quality on student performance (61). In health related research, quantile regression has been used to investigate differential body mass index of children (62).

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. However, because of the paucity of data at our disposal (inability to achieve a convergence in the estimation of the maximum-likelihood model), we further explored the same technique using a different dataset involving respiratory related deaths in a similar large urban population (New York City). This latter analysis was largely as a further proof of the concept and not directly linked to asthma admissions although it did relate to respiratory outcomes. The analyses for QRM are presented in section IV of the thesis.

The idea of using quantile regression techniques in a time series is quite innovative and it adds modest novelty in the focus of forecasting for biomedical research. There is a great potential for the proposed practice of using QRMs in forecasting health events at extremes of the data

1.6 The Thesis Structure (Section and Chapter outline)

The remainder of the thesis is developed in the following sections and chapters:

The Section and Chapter outline of the thesis follows:

<u>SECTION I</u>

Chapter 1: Introduction and outline

SECTION II

Chapter 2: A discussion paper on Asthma and the environment

Chapter 3: Asthma Length of Stay in Hospitals in London 2001–2006

SECTION III

Chapter 4: Semi-structured black-box prediction: proposed approach for asthma admissions in London

Chapter 5: An overview of health forecasting

Chapter 6: Evolving forecasting classifications and applications in health forecasting

SECTION IV

Chapter 7: Forecasting asthma related hospital admissions with negative binomial models

Chapter 8: Humans as animal sentinels for forecasting asthma events

Chapter 9: Forecasting peak asthma admissions in London: an application of quantile regression models

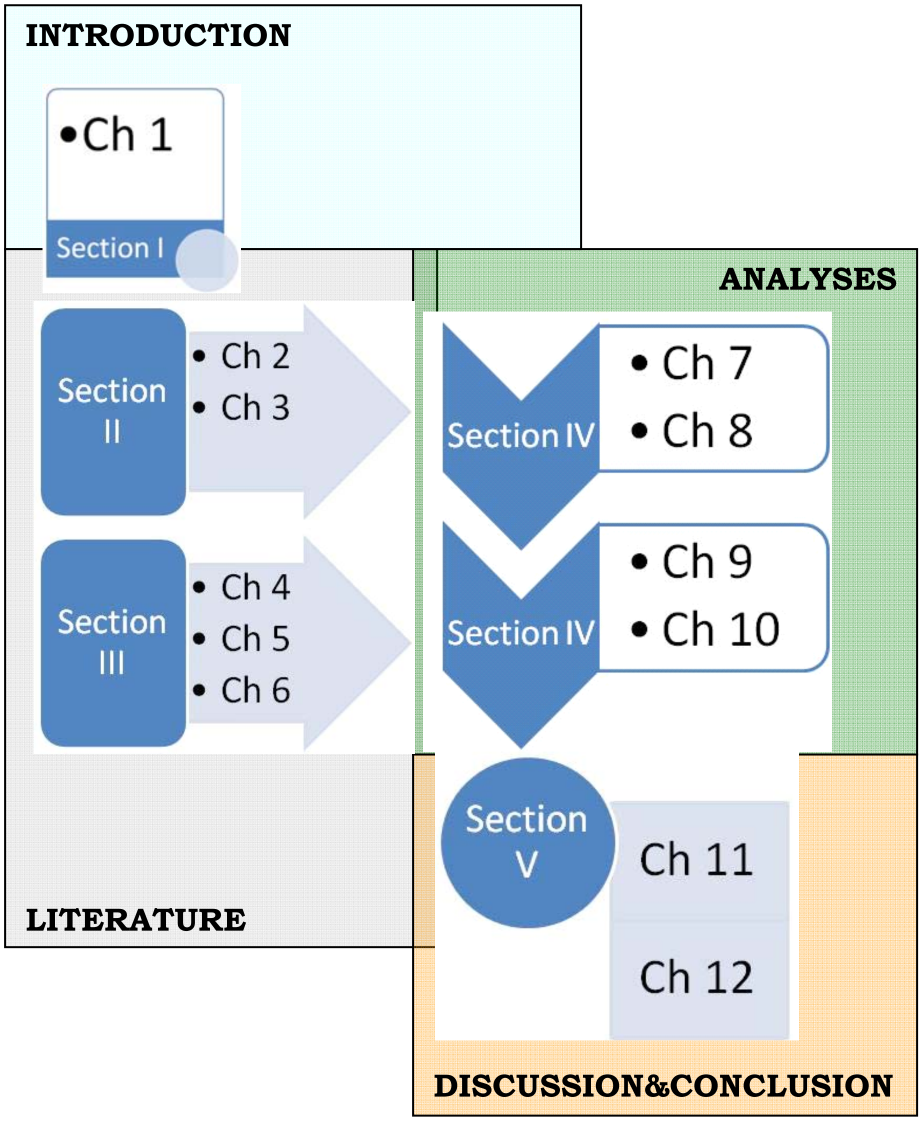
Chapter 10: The use of quantile regression to forecast higher than expected respiratory deaths in a daily time series: a study of New York City data 1987-2000

SECTION V

Chapter 11: General Discussion

Chapter 12: Conclusions and Contributions

Figure 5 The Thesis structure and outline



Section I is the introduction. Sections II and III present the background literature on asthma and health forecasting. Section IV presents various analytical approaches which investigate modelling and forecasting with negative binomial regression and quantile regression. The final section V contains the discussion, conclusion and recommendations of the thesis.

In Section I, which is also Chapter 1, the Introduction to the thesis includes extended highlights on the preliminary data analyses and methods used. Further outputs are referenced in the Appendix.

Chapter 2 consists of a discussion paper on asthma and the environment. It provides a broad snapshot of the disease burden, particularly in the United Kingdom (UK), as well as summaries of some key social and environmental factors known to cause and/or exacerbate the condition. It sets the stage for the subsequent analyses.

A preliminary and critical issue was to establish the point for the forecasting. If there is no disease burden, then there is no need for a forecast. Chapter 3 therefore investigated an analytical approach to estimating a component of the disease burden in the form of Length of Stay (LOS) during admission. This paper models LOS as a function of demographic, diagnostic and temporal factors using a fixed effect model of a negative binomial regression. Important predictors of LOS are thus identified, and their collective and individual effects explained.

Chapter 4 extends the literature on the factors associated with asthma, into forecasting daily events. The paper then proposes approaches for forecasting, and this is exemplified using administrative health records of asthma admissions in London between 2001 and 2006.

Chapters 5 and 6 both critically review the literature on forecasting and health forecasting. Chapter 5 provides an overview of health forecasting, which includes theoretical analysis of health forecasting as well as descriptions of some uncommon principles. Further insights on matters related to the value of health forecasting in health services provision are also discussed. Building on this knowledge base, Chapter 6 then examines the evolving forecasting classifications and their potential applications in health forecasting. It identifies previous forecasting typologies and also discusses the strengths and weaknesses of these methods.

Chapters 7 and 8 advances one of the analytical approaches to health forecasting which was proposed in chapter 4 (45). The two chapters use negative binomial models to develop health forecasting models using environmental predictors on one hand (Chapter 7) and univariate lag models of asthma daily admissions on the other hand (Chapter 8).

Chapters 9 and 10 delve further into analytical approaches that could be adapted to forecast peak /extreme health events. These chapters use quantile regression models and involve environmental predictors.

Chapter 11 is a general but brief discussion of all the key findings presented by the various chapters.

In the concluding Chapters, 12, the salient contributions of the thesis as well as its limitations are reiterated. Areas of future research are also identified and highlighted.

SECTION II

Chapter 2

2.0 A discussion paper on asthma and the environment

This chapter presents a brief background on asthma and its environmental influences. The review of the literature on asthma focuses on its epidemiology, potential environmental mediating factors, and disease burden – highlighting the socio-demographic, economic, and health services management related perspectives. The reason for conducting this review was to provide the background literature on asthma, and to help in identifying aspects of the disease burden.

The review is intended as an overview of the concepts and factors that interplay to determine the disease prognoses, highlighting the approaches that could be exploited in managing the condition through health forecasting. For instance, reliable longitudinal data on some of the important factors mentioned (weather and air quality factors like temperature and ozone pollution) could potentially be useful in predicting future asthma events to help ease health care provision and improve disease burden.

2.1 Introduction

Asthma is a global public health problem and is also well known as the most common chronic disease among children (45). It is under-diagnosed and under-treated, and constitutes a significant burden to individuals, societies and institutions (45, 63-69). Recent global estimates suggest that as many as 300 million people are affected worldwide (64, 70, 71). Meanwhile, the overall global burden of the condition is rising, with children being the hardest hit (64, 67, 69, 72, 73). The World Health Organization (WHO) lists Asthma as *"a chronic disease characterized by recurrent attacks of breathlessness and wheezing, which vary in severity and frequency from person to person. Symptoms may occur several times in a day or week in affected individuals, and for some people become worse during physical activity or at night. … During an asthma attack, the lining of the bronchial tubes swell, causing the airways to narrow and reducing the flow of air into and out of the lungs. Recurrent asthma symptoms frequently cause sleeplessness, daytime fatigue, reduced activity levels and school and work absenteeism"* (63).

Asthma and other closely associated conditions such as wheezing have been well recognised in history. In very early times, asthma was perceived as an act of God to cleanse the body of evil spirits (74). Various reports and historical findings suggest that the condition must have been recognised and treated in different ways even earlier than the time of Hippocrates (460-360 BC). However in the medical literature, Hippocrates is known to have provided the first description of asthma (*Corpus Hippocraticum*)[9]; even though it is not clear whether he meant *asthma* as a clinical entity or as merely a set of symptoms (75-78). Notwithstanding the significant developments in our clinical understanding since then, which have greatly improved the knowledge base and understanding of the morbidity and its prognosis, there remains some grey areas.

Asthma is reported to affect 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 (64, 70). 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 Practitioners for asthma (79).

The natural history of asthma is well understood, and the prognosis is generally predictable (45, 80). The diagnoses however remains a challenge, as the disease is not clearly defined by a particular set of conditions, but a mix of several dynamic factors (45, 65, 69, 81). There are also numerous and quite unpredictable underlying causes of asthma, including genetic and environmental factors (40, 69, 82, 83). Given the complex nature of the condition, the diagnostic techniques commonly employed in detection include the clinical history and patterns of symptoms, physical examination and lung function measurements including spirometry, as well as skin tests for allergens (84).

[9] The **Hippocratic Corpus** - a collection of early medical works from ancient Greece associated with Hippocrates and his teachings. They are known to vary in content, age and style with unknown authorship.

A number of studies have identified a change in the global epidemiology of asthma; and it is well documented that developed countries have consistently shown dramatic increases in its prevalence (67, 82, 85, 86). This change has more recently been observed in some less-developed countries (64, 68).

In 2001, the United Kingdom National Asthma Campaign (87) reported that asthma affected over five million people, about one in five households. Meanwhile the HES showed a 6.0% increase over a 10-year period (between 1999 and 2008) in the number of admissions to hospital in England alone with asthma and allergies. The majority of these patients were either young males or older females (88). However, some other reports showed that the United Kingdom as a whole had prevalence rates of more than 15%, which was considered one of the highest in Europe (64, 68). According to Asthma UK, 67,077 people in England were hospitalized for asthma between April 2006 and March 2007, of whom more than 40% were children under the age of 15 years (89).

2.2 The disease burden of Asthma

On a per capita basis, the United Kingdom has the greatest burden of severe asthma of any country in Europe (90). The Global Initiative for Asthma (GINA) 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 United States (91), 7% in Germany and 7% in France (92). Of those asthma sufferers in the United Kingdom, some studies suggests that the patients with severe asthma account for the majority of hospitalisations due to asthma (90). This group of severe asthma sufferers consist of 2.6 million individuals (i.e. 2.1 million adults and 500,000 children) in the United Kingdom (93).

Medical Practitioners in the United Kingdom 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 (64). In the United Kingdom in 2004, there were 75,000 emergency hospital admissions due to asthma and 1,500 fatalities (64). It was recently reported that about 5.4 million people in the United Kingdom are currently *receiving* treatment for asthma, 1.1 million of these are children (89). Despite treatment, a substantial portion of asthmatics are "not well-controlled" (94).

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 United Kingdom over £2.3 billion a year (66), including £1.2 billion in individual productivity losses (95). The Office for Health Economics further estimated that the cost to the National Health Service (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 (96). Furthermore, poorly controlled asthma appears to have a considerable impact on health care costs (97).

2.2.1 Asthma events and related spell durations

Asthma spell duration or Length of stay (LOS) refers to the duration of a hospital admission (i.e. the difference in days between the date of admission and the date of discharge). It is an important indicator of health care cost and management, particularly as it is associated with hospital bed occupancy and related services (98-103). Length of stay can be used as an indirect estimator of resource consumption and efficiency within the settings of a hospital, and has direct implications for overall healthcare planning and policy (98, 101, 104). It is therefore seen as one of the measures which can be used to estimate a part of the total disease burden of asthma in a population (33, 105).

Analyses of length of stay associated with asthma, and its correlates with demographic, clinical and temporal factors are relatively unusual (106-108). The standard procedures for decision making in the case of asthma hospitalizations vary widely during its diagnosis; but this ultimately affects the variability associated with its management in relation to the Length of stay (109-111).

The study proposed by Arnold and colleagues, which sought to identify the key clinical predictors for acute asthma exacerbations in paediatric patients (110), also touched on the clinical determinants of Length of stay. This study however had a sharp clinical focus. Meanwhile, investigations on the combined demographic, clinical and temporal determinants of asthma Length of stay in large populations are not common.

Some other earlier studies on Length of stay for asthma hospitalisation in the United Kingdom (Scotland) focussed on the trends in asthma admissions and how changes in hospital bed occupancy could measure resource use (104). In this study, and others similar to it (112-114), they described the trends in the percentage change in asthma admission rates with respect to temporal and demographic explanatory variables (e.g., year of admission, aggregate ages and gender). These studies have however not been able to predict or explain the Length of stay for asthma suffers, whilst accounting for variations in demographic, temporal and clinical factors. In more recent studies, Length of stay was better predicted by demographic, diagnostic and temporal characteristics using the multilevel effect of the individual asthma sufferers (33), compared to the same effect of their area of residence (105).

2.2.2 Socio-demographic factors/economic

Socio-demographic and economic factors have been shown collectively to play a role in asthma prevalence (115). Ethnicity/race, community vitality and social capital contribute significantly to asthma variation. Asthma UK reports of minority ethnic groups bearing the greater burden of the disease, though in terms of absolute numbers, the majority ethnic groups (i.e. white) are admitted to hospitals in the United Kingdom more than any other ethnic group. The situation is not different from reports of other countries including the United States, Australia and Italy (115-118). For instance the study conducted by Adams *et al*. in Australia on the factors associated with asthma hospital admissions in adult populations showed that higher income earners i.e. >A$50,000 per annum were less (8%) compared to the lowest income group (<A$8,000 per annum) which was 32% of the individuals affected (115). Meanwhile gender and age are the most frequently identified factors in the literature in the socio-demographics of asthma s(108, 116, 119-123).

2.3 The Biology of Asthma

Asthma is a chronic respiratory condition typified by obstruction and continuously persistent inflammation of the airways (124). This obstruction to airflow, which is episodic within individuals with early or mild asthma, can cause symptoms of tightness and wheeziness in the chest (79). 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. The symptoms are accompanied by decreases in expiratory airflow shown by objective measures of lung functioning that employ spirometry and peak flow (84, 125, 126).

The United Kingdom Committee on the Medical Effects of Air Pollutants (COMEAP)[10] 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 (79). The inherent interdependence or independent effects of known environmental determinants of ill health, particularly air pollutants and some weather factors have been reported by authors who have looked at the effect of the environment on asthma exacerbations (127-129).

Other biological changes that result in increasing individual vulnerability to asthma exacerbation and are initiated by environmental changes may be an important note. A

[10] 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

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 (130-134). This remodelling of the airways increases an individual's predisposition to asthma (80, 130, 132, 135), and thus supports the proposition that environmental factors play a critical role in the inception and progression of the disease in genetically susceptible individuals (82, 85, 134).

2.4 The Epidemiology of Asthma

Asthma has already been highlighted as a substantial health problem among all demographic and population groups globally. Its prevalence rates in many countries are increasing - leading to higher hospital admission rates (136). In England and Wales, discussions on the epidemiology of asthma have focussed on trends over the past few decades and there are questions arising from these trends: 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 population; or is there another explanation entirely? It has, for instance, been suggested that when declines in the prevalence of asthma are observed they are entirely attributable to variations in diagnosis (137-139). 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 (66, 79, 136, 140).

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. It is well noted that occupational exposures constitute a common risk factor for adult asthma (69). 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 (141). This argument is similar to the one by Subbarao and colleagues, which attributes the wide variation in the prevalence of asthma worldwide to the results of variations due to the gene-by-environment interactions (69).

Local environmental conditions are thus likely to be important in determining the impact or manifestation of asthma where factors such as temperature, humidity, air pressure as well as air pollutants interact and do not have independent effects on asthma (128, 142-148).

2.5 Asthma events and environmental factors

Environmental factors do have complex interrelationships, and their collective impact on health is not clearly understood (149-151). A review of the health effects of climate change published in 2008 (152) highlighted weather and air quality as the two component issues of interest to the environment. The constituent indicators of temperature, humidity, vapour/atmospheric pressure, wind, and atmospheric aerosols are known to produce polluted environments, i.e. mists, fogs or smog (151, 153). There are many pathways through which these environmental pollution and dynamics can exacerbate asthma and lead to primary care provider visits (154, 155) and some of the mechanisms involved in this have been discussed in subsequent sections below (45).

Health conditions triggered by local environmental changes, including indoor conditions, as well as occupational exposures vary considerably in their effects and symptoms and also across various spatial settings (45, 156). There is overwhelming evidence for the role of these environmental factors; which includes several classical experimental laboratory and field studies, interventions and health impact studies (142, 156-166). These environmental effects primarily depend on the individual's susceptibility and level of epidemiological exposure (167-170). Vulnerable groups within given populations, particularly children (73, 171, 172) and the elderly tend to be the hardest hit with the former experiencing both the direct and indirect effects of these changes (169, 173).

2.5.1 Asthma and Weather

There is ample evidence on the effect of temperature changes, barometric pressure and relative humidity on the exacerbation of asthmatic symptoms (142, 151, 159-166). A number of studies have also used the association of weather and disease incidence, hospitalization or mortality to examine the nature of the relationships (45, 174, 175).

It is well established that the relationship between asthma and environmental conditions is affected in complex ways by changes in weather and season (176-179). It is also worth noting that seasonal effects vary geographically. In Mexico, for instance, asthma is associated with the rainy season, whereas in England and Wales asthma is more strongly associated with temperature change rather than rainfall (45, 163, 166). In the United Kingdom and Taiwan, peaks in asthma events occur in the winter /autumn seasons but not in summer (174, 180, 181). These underlying circumstances makes it imperative to understand the local relationships between asthma and the weather /air quality and season (45). In the light of the geographical variation, forecasting will need to take account of local conventional effects.

2.5.1.1 Thunderstorms, allergens and asthma

Thunderstorms are known to cause rapid environmental changes leading to the release of high levels of asthma allergens such as moulds and plant pollen, which can exacerbate asthma. Short-term increases in the concentrations of grass pollen and spores were found to be associated with emergency department visits from asthma among children in Montreal, Canada (171, 172). Other related exposure effects like dust material from various emission sources are also released into the atmosphere, and these could all increase the susceptibility of individual asthma sufferers (182-184). In the United Kingdom, peaked asthma events have often been preceded by heavy thunderstorms and thunder activities, and these have been discussed along with other exposures (allergens) and environmental mediating factors (183, 185-188). Higham, Venables, *et al*. for instance reported on the association between thunderstorms and asthma events leading to hospitalizations in Britain (188). The effect of thunderstorm on grass pollen counts and their role on asthma events/hospital admissions in England was also reported by Newson *et al* (183). However, consistent and reliable data on thunderstorms is limited, and has not been used effectively in forecasting respiratory events.

2.5.2 Asthma and Air Quality (Pollution)

The evidence in literature for pollution-related health events, particularly for respiratory conditions like asthma is substantial (189). The association between asthma events and air pollutants like nitrogen (IV) oxide, particulate matter, ozone, sulphur dioxide, smoke, as well as household or natural environmental allergens is well known (128, 129, 143, 145, 190-192). It has been observed that susceptible individuals in particular respond more frequently when exposed to pollutants than, would happen in initiating allergies among non-susceptible individuals (193-195). The known relationships between asthma and air quality, however, have not been successfully used to forecast asthma events (196, 197).

The effect of individual air pollutants on asthma is better understood than the collective effect of multiple simultaneous, pollutants. The interaction between air pollutants and other environmental factors further complicates their likely effect on asthma, and are 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 (198-200), even though others support the idea (201-210).

2.6 Summary

This chapter provided an overview of asthma, its biology, disease burden and epidemiology. The complex role of environmental conditions as triggers for asthma was also discussed, particularly along the areas of weather and air quality.

One issue that comes out strongly in the literature is the contextual variation in the disease burden. Most of the forecasting work that is developed in this thesis relies on a set of hospital admissions data from London, England. Before developing the forecasting, however, it seemed important to establish the actual burden of asthma within the population and the period we were working. Because the forecasting work related to hospital admissions, it seemed natural to establish the burden of disease within the context of hospital admissions – Length of stay was a suitable proxy.

Chapter 3

3.0 Introduction

This chapter reposts a study published in PLoS One journal on the length of stay of asthma patients in hospitals within London.

Given the constraints of space, and the focussed nature of papers published in health journals, the paper does not account for other approaches and techniques that were explored in the preliminary analysis leading to the development of the paper. These additional approaches provide important backgrounds which are discussed briefly here.

Asthma LOS can be modelled with geographical, socio-demographic, temporal and clinical factors using count models on hospital admission data. This study showed that age, gender, and co-morbidity were important predictors of LOS, and the procedure may be a useful tool for planning and resource allocation in health service provision. It is also worth noting that in using asthma admissions to estimate and analyse LOS, the nature of the distribution of LOS data was considered in determining the choice of statistical technique (which was in this case negative binomial model).

3.1 Establishing the burden of asthma in relation to LOS

In a preliminary statistical model we assumed that there could be a patient area effect. That is some geographical areas might have longer LOS than others. Specifically, this could arise if hospital admission policies in some areas were different from those in other areas, or if the populations in some areas were more prone to severe asthma events requiring longer LOS. To account for this possibility a random effects model of the patients' residential postcode was fitted. This allowed us to account for the lack of independence between admissions associated with similar patient residential areas in London.

In the literature on asthma admissions in general, there is evidence to suggest that repeated hospital visits is a significant predictor of total admissions (115, 211), and by extension could also apply to LOS. Therefore a multilevel (3 level) effects model accounting for area, individual and visit effects would be an ideal way of modelling LOS. But at the time of conducting this analysis, there was no generally available statistical software to support this analysis. Most packages (e.g. Stata, R, SPSS) could accommodate a two-level, but not a three-level model (212, 213). One option in this situation was to separately model the area and patient effects and then select the more robust model based on a common model diagnostic tool (AIC). The result of this comparative study was presented at an international conference (105). There were marginal differences between the two models in terms of the individual variable parameters, but on the whole the individual random effect model (accounts for repeat visits) was a better fit than the area effect model. This was reproduced in PLoS One article.

3.1.1 Summary discussion on LOS

The research report show an aspect of the burden of asthma from a health services perspective.

Methodologically, the use of a negative binomial model and a random effect multilevel model presented a more robust approach to model LOS compared to traditional GLM approaches. AIC, was an ideal model for comparing different predictive models and selecting the most suitable one.

One weakness in this research was that it only observed the associations between the factors and the length of stay, and we cannot attribute any causal links. The paper also points out limitations associated with the lack of data on J46 *Status asthmaticus* and of Length of Stay (which underestimates those admitted for less than 24hrs).

3.2 Asthma Length of Stay in Hospitals in London 2001–2006

3.2.1 Declarations for Thesis Chapter 3

Monash University

Declaration for Thesis Chapter 3

Declaration by candidate

In the paper: *Asthma length of stay in hospitals in London 2001-2006: demographic, diagnostic and temporal factors (Published: PLoS One, 2011. 6(11): p. e27184)* /Chapter 3 of thesis, the nature and extent of my contribution to the work was the following:

Nature of contribution	Extent of contribution (%)
Lead and corresponding author, conceptualized the idea, accessed, organized and analyzed data, discussed preliminary results with stakeholders, drafted initial manuscript for circulation, submitted manuscript to journal and then managed correspondence with editors/reviewers, editorial staff and publishers until the final publication	100%

The following co-authors contributed to the work. Co-authors who are students at Monash University must also indicate the extent of their contribution in percentage terms:

Name	Nature of contribution	Extent of contribution (%) for student co-authors only
Daniel D. REIDPATH	Supervised all activities leading to the publication of the paper; critiqued conceptual ideas, analyses and all drafts of manuscript	N/A
Christophe SARRAN	Assisted in accessing data, provided expert interpretation to the data and took part in reviewing and critiquing manuscript prior to submission	N/A

Candidate's Signature	Ireneous N Soyiri	**Date:** 22-11-2012

Declaration by co-authors

The undersigned hereby certify that:

(1) the above declaration correctly reflects the nature and extent of the candidate's contribution to this work, and the nature of the contribution of each of the co-authors.

(2) they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;

(3) they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;

(4) there are no other authors of the publication according to these criteria;

(5) potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and

(6) the original data are stored at the following location(s) and will be held for at least five years from the date indicated below:

Location(s)	**Global Public Health, School of Medicine & Health Sciences, Monash University Sunway campus**

Signature 1	Daniel D. Reidpath	**Date:** 22-11-2012
Signature 2	Christophe Sarran	**Date:** 04-08-2012

- -

Asthma Length of Stay in Hospitals in London 2001–2006: Demographic, Diagnostic and Temporal Factors

Ireneous N. Soyiri[1,2]*, Daniel D. Reidpath[2], Christophe Sarran[3]

1 School of Public Health, University of Ghana, Accra, Ghana, **2** Global Public Health, School of Medicine & Health Sciences, Monash University, Sunway Campus, Malaysia, **3** Met Office, Exeter, United Kingdom

Abstract

Asthma is a condition of significant public health concern associated with morbidity, mortality and healthcare utilisation. This study identifies key determinants of length of stay (LOS) associated with asthma-related hospital admissions in London, and further explores their effects on individuals. Subjects were primarily diagnosed and admitted for asthma in London between 1st January 2001 and 31st December 2006. All repeated admissions were treated uniquely as independent cases. Negative binomial regression was used to model the effect(s) of demographic, temporal and diagnostic factors on the LOS, taking into account the cluster effect of each patient's hospital attendance in London. The median and mean asthma LOS over the period of study were 2 and 3 days respectively. Admissions increased over the years from 8,308 (2001) to 10,554 (2006), but LOS consistently declined within the same period. Younger individuals were more likely to be admitted than the elderly, but the latter significantly had higher LOS ($p<0.001$). Respiratory related secondary diagnoses, age, and gender of the patient as well as day of the week and year of admission were important predictors of LOS. Asthma LOS can be predicted by socio-demographic factors, temporal and clinical factors using count models on hospital admission data. The procedure can be a useful tool for planning and resource allocation in health service provision.

Citation: Soyiri IN, Reidpath DD, Sarran C (2011) Asthma Length of Stay in Hospitals in London 2001–2006: Demographic, Diagnostic and Temporal Factors. PLoS ONE 6(11): e27184. doi:10.1371/journal.pone.0027184

Editor: Marco Idzko, University Hospital Freiburg, Germany

Received June 2, 2011; **Accepted** October 12, 2011; **Published** November 2, 2011

Funding: The authors have no support or funding to report.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: soyiriin@yahoo.com

Introduction

Globally, the morbidity and mortality associated with asthma places a high burden on health care infrastructure and services [1,2,3,4,5]. Asthma is a chronic condition which in the United Kingdom alone affects over 5.2 million people including 1.1 million children [6]. The UK together with the Republic of Ireland have the highest prevalence of asthma in the world [3]: it is the leading cause of hospital admissions particularly among children [7], and disproportionately affects certain ethnic groups and demographics [2]. The total cost of asthma is estimated to be £2.5 billion in the UK and Ireland and results in millions of lost working days [3]. Notwithstanding a significant literature on asthma and its impacts, most of the researches are focused on the clinical presentation of the disease.

Length of stay (LOS) refers to the duration of a hospital admission (i.e. the difference in days between the date of admission and the date of discharge). It reflects several aspects of hospital care including the complexity of the case, the efficiency of hospital care, and the nature of hospital policies on admission and discharge [8,9,10,11,12,13,14]. LOS can be used as an indirect estimator of resource consumption and efficiency within the settings of a hospital, and has direct implications for overall healthcare planning and policy [5,9,12].

Analyses of length of stay associated with asthma, and its correlates with demographic, hospital, and temporal factors are relatively unusual [7,15,16]. The standard procedures for decision making in the case of asthma hospitalizations vary widely during its diagnosis; but this ultimately affects the variability associated with its management in relation to the LOS [17,18,19].

The study proposed by Arnold and colleagues, which sought to identify the key clinical predictors for acute asthma exacerbations in paediatric patients [18], may also identify the clinical determinants of LOS. This study however has a sharp clinical focus. Meanwhile, investigations on the combined demographic, diagnostic and temporal determinants of asthma LOS in large populations are not common.

Some other earlier studies on LOS for asthma hospitalisation in the UK (Scotland) focussed on the trends in asthma admissions and how changes in hospital bed occupancy could measure resource use [5]. In that study, and others similar to it [20,21,22], they described the trends in the percentage change in asthma admission rates with respect to temporal and demographic explanatory variables (e.g., year of admission, aggregate ages and gender). Such studies have, however, not been sufficient in predicting or explaining the LOS for asthma sufferers, whilst accounting for variations in demographic and temporal factors.

The aim of this study was to extend earlier work by examining the independent effect of demographic, hospital, and temporal factors associated with asthma LOS, using hospital admission records from London (2001 2006).

Methods

Data

This study involved a secondary analysis of hospital administrative data from London, England. The data covered 56,832 emergency asthma hospital admissions from January 1st, 2001 to December 31st, 2006. After removing data with missing patient sex, 56,768 admissions were available for analysis. The data relate to 40,359 unique individuals.

Data were sourced from the nationally recorded Hospital Episode Statistics (HES) maintained by the National Health Service, England [23]. HES data do not include all fields for which individual hospitals may collect data, such as which attending staff were involved in patient management; and not all data were readily available for secondary analysis, such as any variation post-admission in the diagnosis.

Asthma admissions were defined as any hospital admission with a primary diagnosis of asthma; i.e., an International Classification of Diseases (ICD-)10 code of J45. The mis-coding rate of J45 is not known, although it is known that J45 and ICD-10 coded J46 (*status asthmaticus*) admissions do have a diagnostic overlap [24]. Unfortunately, the J46 data were not available. Previous research has indicated that the J45 coded admissions in the UK cover around 82% of all asthma admissions. Furthermore, in absolute numbers, there are more, severe cases among the J45 coded admissions than among the J46 admissions [24]. The implications of this are discussed further, towards the end of the article.

The length of stay was estimated as the difference in days between the date of admission and the date of discharge. All stays of less than 24 hours (i.e., <1 day) were recorded as zero days admission. This has implications for the underestimation of LOS, and is discussed later in the article. LOS is the total length of stay without regard to whether a patient may remain in hospital for reasons unrelated to asthma. It is, thus, a measure of the length of stay, for all causes, given admission for asthma.

Following other research,[2] solely for the presentation of descriptive data, LOS was categorised into short (less than 24 hours), medium (1 3 days), long (4 7 days) and very long (more than a week) stays.

The explanatory variables from the HES dataset that were included in the analyses could be broadly described as demographic (age, sex, ethnicity), hospital (primary diagnosis, secondary diagnosis, method of admission/discharge) and temporal (day of week, season and year of admission) variables. Data on sex (male and female) were used as recorded in the dataset. Age was categorised as: 0 4, 5 14, 15 44, 45 59, 60 74 and more than 75 years. The approach to age categorization in asthma research is not consistent; however, the categories used here account for differences between younger and older children as well as young adults and older adults. Ethnicity was represented by five categories of (i) White (-Irish, -British, -Any other white), (ii) Black (all black i.e. -African, -Caribbean, -American), (iii) Asian (Indian, Pakistani, Bangladeshi, Chinese), (iv) Mixed and (v) Unknown/not indicated. This grouping conforms broadly to the categories used by the Office of National Statistics, but collapses some of the categories in which there were small numbers [25].

The ICD-10, J45 diagnosis of asthma is sub-coded into four categories; however, 95% of all admissions fell within a single category (J45.9 "Asthma, unspecified" on their hospital records). Notwithstanding the uncertainty about the accuracy of the sub-coded diagnoses, primary diagnosis was retained as a separate explanatory variable in the model [16]. All the secondary diagnoses were categorised according to their respective ICD-10 codes and were included in the analysis. These included: (i) Acute upper respiratory infections; (ii) Other diseases of upper respiratory tract; (iii) Influenza and pneumonia; (iv) Other acute lower respiratory infections; (v) Suppurative and necrotic conditions of lower respiratory tract; (vi) Chronic lower respiratory diseases; (vii) Lung diseases due to external agents; (viii) Other respiratory diseases principally affecting the interstitium; (ix) Other diseases of pleura; (x) Other diseases of the respiratory system; (xi) Other non-respiratory system diseases; and (xii) Missing values. The categories viii x of the secondary diagnosis, which had very few counts were grouped and reclassified as one category "Other diseases of the respiratory system".

Some of the key derived variables created for the analysis included "day of the week", "season", and "year of admission". The meteorological seasons were Spring (1st March 31st May), Summer (1st June 31st August), Autumn (1st September 30th November) and Winter (1st December February end). The year of admission was extracted from the date of admission.

Data Analysis

Length of stay was modelled using negative binomial regression, with a random effect to take account of the lack of independence between admissions associated with repeat admissions of the same patient. Poisson regression is generally well suited for modelling count data. A negative binomial model, however, is preferred when there is over dispersion; that is, when the mean and the variance are not equal. In the context of LOS, this can occur if there are more 0 days of admission than anticipated under a Poisson model [26,27,28,29]. This was formally tested using the likelihood approach suggested by Long and Freese [30].

For the expected LOS, the negative binomial regression can be presented in the form:

$$\Pr(Y=y|\lambda,\alpha)=[\Gamma(y+\alpha^{-1})/y!\Gamma(\alpha^{-1})].$$
$$[\alpha^{-1}/(\alpha^{-1}+\lambda)]^{\alpha^{-1}}.[\lambda/(\alpha^{-1}+\lambda)]^{y}$$

Where:

λ is the mean of the distribution;

α is the over dispersion parameter;

y is the LOS component represented by 0, 1, 2, ...;

Γ is the gamma function.

A positive coefficient in the regression output indicates that a factor will increase the LOS relative to its reference category and conversely a negative coefficient will decrease the LOS relative to its reference category. The exponent of the coefficient can be interpreted, all other things being equal, as the proportionate increase (for values greater than 1) or decrease (for values between 0 and 1) of LOS associated with a one unit increase in the explanatory variable [29].

A positive coefficient in the regression output indicates that a factor will increase the LOS relative to its reference category and conversely a negative coefficient will decrease the LOS relative to its reference category. The exponent of the coefficient is a ratio and it can be interpreted, all other things being equal, as the ratio of length stay at one level of the covariate to length of stay at one level less than this on the covariate [29].

In the first instance, a full model with all explanatory variables was developed. This then led to a reduced model that excluded month of birth. The improved fit of the reduced model was established by the reduction in the Akaike Information Criterion (AIC) [31,32,33].

All analyses were conducted using Stata SE version 10.1 statistical packages (Stata Corporation, Texas, USA). Exemption from ethical review for the secondary analysis of hospital administrative data was obtained from the Monash University Human Research Ethics Committee (Number: 2011001092).

Results

The median recorded length of stay (LOS) following admission was 2 days with a mean LOS of 3 days (sd = 6 days). The distribution had a heavy right tail with a range of 0 367 days. The majority (56.7%) of all admissions related to a person being admitted only once during the period 2001 2006; 17.4% of admissions were associated with a person being admitted twice, and 8.2% were associated with a person being admitted three times. This percentage declined rapidly, although one person was admitted 77 times during the period.

Table S1 shows the distribution of admissions across the key explanatory variables. Males and females each had close to 50% of all asthma admissions. Nearly half of all admissions were whites; with the other ethnic groups each contributing about a tenth. Meanwhile a fifth (20%) had no record of their ethnicity in the dataset. Most admissions were made directly through the hospitals' Accident and Emergency (A&E) departments (92%), and most discharges from hospitals were based on clinical advice or clinical consent (97%). The great majority of admissions were diagnosed as "Asthma, unspecified" (94%). The autumn months recorded the most admissions (30%) and the least were recorded during the summer months (22%). From 2001 to 2006 the number of admissions steadily increased 27% over the 6 year period. This increase was well in excess of the 5.9% growth in the population that occurred in London between 2001 and 2009 [34].

In the multivariable analysis of the explanatory variables, a full model was developed that included all explanatory variables (Table S2). Month of birth showed weak effects, with only one month significantly different from the base month. A reduced model, removing month of birth as an explanatory variable was developed. The reduced model was retained on the basis of an improved AIC (239405.5 in the full model compared with 239394.8 in the reduced model), and little variation in the parameter estimates of the remaining explanatory variables.

All other things being equal, Females had a slightly longer LOS than males (1.11 times longer; 95%CI: 1.09 1.13). The association with age and LOS increased monotonically, with each age category over 0 4 years of age having a significantly longer LOS, from 1.07 (95%CI: 1.04 1.11) times longer for 5 14 year olds up to 3.43 times longer (95%CI: 3.31 3.55) for those over 75. The confidence intervals also suggest a fairly clear separation across the age groups, although this was not formally tested. All other things being equal, there was a small, but significantly longer LOS associated with being Black (1.05; 95%CI: 1.02 1.08) compared with White, and a slightly shorter LOS associated with having no recorded ethnicity (0.93; 95%CI:.91 .95). A primary diagnosis of "Predominantly Allergic Asthma" was associated with a significantly shorter LOS than the most common diagnosis of "Asthma, unspecified" (0.83, 95%CI:.79 .87). A number of the secondary diagnoses were associated with a longer LOS than those with an acute upper respiratory tract infection. Influenza, pneumonia, or other respiratory diseases, were associated with an increased LOS between 1.3 and 1.7 times longer.

Method of admission was significantly associated with LOS. Compared with admission through attendance at Accident and Emergency, referral by a General Practitioner was associated with a shorter LOS (0.90; 95%CI: 0.86 0.93), while referral by a Consultant Physician was associated with a significantly longer LOS (1.20; 95%CI: 1.12 1.28). All other things being equal, days of the week except Saturday were associated with a significantly longer LOS than Sunday admissions. Monday to Friday admissions were associated with an LOS about 1.31 times longer; Saturdays were associated with an LOS 1.09(95%CI: 1.05 1.13) times longer. Autumn, Winter, and Spring were all significantly associated with slightly longer (1.04 to 1.07 times longer) LOS than Summer.

Though the number of yearly admissions increased over the time period, the length of stay consistently reduced over the same time, with every year after 2002 being associated with a significantly shorter LOS. All other things being equal, patients admitted in 2006 had an LOS 0.71 times as long as someone admitted in 2001 (95% C.I. 0.68 0.73).

Discussion

Asthma hospital admissions and their associated lengths of stay place a substantial burden on the health services, carers, and individual asthma sufferers [1,3,5]. Between 2001 and 2006, there were 56,832 asthma admissions in London associated with 40,359 individuals, accounting for around 170,500 days of hospitalised care; a finding consistent with earlier research [20,35,36]. The number of hospital admissions increased from 8,308 to 10,554 between 2001 and 2006, while the actual length of stay associated with each admission reduced significantly.

All the demographic, hospital, and temporal factors investigated were found to have statistically significant associations with the length of stay. The statistical association, however, does not necessarily translate into what might be regarded as variations in the length of stay with significant clinical impact. All other things being equal patient sex, ethnicity, primary diagnosis, and season of admission resulted in no more than a 10% variation in the expected length of stay. The effect of ethnicity was perhaps most surprising given the existing literature [38, 40].

The association of sex on LOS is of some interest because it is known that males are more likely than females to suffer from asthma [15]; but all other things being equal, once admitted, females appear more likely to have a longer stay. Speculatively, this could be explained either by the fact that asthma events in females who are admitted are more severe than those events in males, or males recover more quickly once admitted, or there are unobserved social or system artefacts interacting with patients' sex to vary length of stay.

The remaining factors had "effect sizes" that were sufficiently large to suggest clinical significance. Perhaps unsurprisingly, diagnosis was significantly associated with length of stay. As a primary diagnosis, predominantly allergic asthma was associated with a reduced LOS compared with "asthma, unspecified". Among the secondary diagnoses, co-morbidities of the lower respiratory tract were associated with an LOS up to 1.82. It was also interesting to observe the variation in the length of stay associated with the mode of admission. Admission based on a Consultant in an out-patient clinic, all other things being equal, was associated with an LOS 1.20 times longer, while General (Family) Practice admissions were associated with shorter stays (0.90). Explaining the former seems straightforward: a clinical specialist sees a patient and recognises someone in need of acute care, and those patients are on average more clinically acute than those who attend a hospital Accident & Emergency department without having first seen a doctor. Why admissions by the generalist medical practitioner, however, should on average require shorter stays than those who attend a hospital Accident & Emergency department directly is less clear.

It is well known that the numbers attending hospital Accident & Emergency departments vary by the day of the week [37]. The relationship between the day of the week and the length of stay is less frequently recorded, but has been observed [38]; and may relate to unobserved social and health service factors including bed occupancy rates or staffing levels.

From a health services perspective, large effects that involve relatively few patients are less important than smaller effect involving larger numbers of patients. Bearing this in mind, the explanatory variables of greatest health services interest appear to be sex (females have increased LOS), age (monotonically increasing LOS with increasing age), secondary diagnosis involving a disease of the lower respiratory tract (increased LOS), General practitioner referral (decreased LOS), day of the week (decreased LOS on Sundays), and year (monotonically decreasing LOS).

There are a number of limitations associated with this study. Methodologically, all that can be observed are associations, and the attribution of a causal link (just from these data) between the factors and the length of stay is impossible; although one would anticipate a straightforward causal link between, say, clinical diagnosis and length of stay. Notwithstanding the lack of clear causal pathways within the design of the research, the results can still be useful from a health services perspective, particularly when the results are interpreted within the wider body of knowledge about asthma.

One limitation of this study relates to the definition of asthma. Only ICD-10, J45 (*asthma*) coded admissions were in the data set, excluding J46 (*status asthmaticus*) coded admissions. This will affect the gerneralisability of the results. However, from previous research we know that the majority (90%) of asthma admissions in the UK are J45 coded [24]. Although J46 coded admissions tend, on average, to be more serious, there is considerable diagnostic overlap between the two codes, and in absolute numbers many J45 cases are more serious than J46 coded cases [24]. On balance, these results are likely to point to the correct direction of the relationship between the explanatory variables and LOS, although the specific estimates may need to be adjusted.

The second limitation is with definition of "length of stay" itself. There were a substantial number of admissions of zero days (17%). Zero days of admission, however, do not imply zero time or zero cost, and the measure does not reflect admissions of nearly 24 hours (underestimating LOS), or the high costs of monitoring and testing that occur early in the admission cycle. The full burden on hospitals is, therefore, likely to be higher than might be reflected here [20]. The interpretation of the results, therefore need to separate issues of bed occupancy from the total cost (and time distribution) of management and care. Nonetheless, appropriately weighted LOS results can be factored into a total cost analysis.

Finally, there are likely also to be unobserved factors, or factors for which HES data are unavailable, contributing to length of stay. Changes in the healthcare system, clinical management, and hospital culture for instance may explain some of the variation in the length of stay accounting for the annual decline, variations by day of the week, or sex differences. Capturing these factors when using routinely collected hospital data can be challenging, but the lurking limitation needs to be recognised.

Notwithstanding these limitations, the data are near to a complete record of asthma (J45) admission from 2001 to 2006 and the reported relationships do reflect what are in the data in that time period. For health services, this is useful. The demographic factors can be used directly to project hospital length of stay, and therefore project bed occupancy and some level of health service utilisation. The disparity between the length of stay associated with Consultant and General Practitioner admissions raise interesting clinical questions worthy of further investigation, and they are also informative for health services in projecting utilisation. Future research will ideally capture J46 admissions as well.

Conclusion

Asthma admissions continue to be an important source of hospital admissions and account for a substantial number of bed days. Although the number of admissions has steadily risen over time, the average length of stay has steadily declined. Demographic, temporal, and diagnostic factors independently explain the variation in the length of stay. The identification of these factors is of clinical and health services interest; pointing to potential areas of future research, but also providing a basis for projecting health service utilisation.

Supporting Information

Table S1 Summary statistics of asthma and related indicators of hospital admission records in London, 2001-2006.
(DOC)

Table S2 Multivariable base and reduced models of length of stay in asthma related hospital admissions in London, 2001-2006.
(DOC)

Acknowledgments

We acknowledge the support by United Kingdom Meteorological Office's Health Forecast team, who provided the original data used in this analysis. The work here, however, does not reflect the views of the United Kingdom Meteorological Office or any of its staff. We would also like to thank the reviewers for their input in strengthening this paper.

Author Contributions

Conceived and designed the experiments: INS. Performed the experiments: INS. Analyzed the data: INS. Contributed reagents/materials/analysis tools: INS DDR. Wrote the paper: INS DDR CS.

References

1. Rodrigo GJ, Rodrigo C, Hall JB (2004) Acute asthma in adults: a review. Chest 125: 1081 1102.
2. Asthma-UK (2007) The Asthma Divide. Inequalities in emergency care for people with asthma in England. London.
3. Masoli M, Fabian D, Holt S, Beasley R (2004) The global burden of asthma: executive summary of the GINA Dissemination Committee report. Allergy 59: 469 478.
4. Hodder R, Lougheed MD, Rowe BH, FitzGerald JM, Kaplan AG, et al. (2010) Management of acute asthma in adults in the emergency department: nonventilatory management. CMAJ 182: E55 67.
5. Morrison DS, McLoone P (2001) Changing patterns of hospital admission for asthma, 1981-97. Thorax 56: 687 690.
6. Asthma-UK (2010) Key facts & statistics Asthma UK. London: Asthma UK.
7. Callery P, Kyle RG, Campbell M, Banks M, Kirk S, et al. (2009) Readmission in children's emergency care: an analysis of hospital episode statistics. Archives of Disease in Childhood 95: 341 346.
8. Poulos LM, Toelle BG, Marks GB (2005) The burden of asthma in children: an Australian perspective. Paediatric Respiratory Reviews 6: 20 27.
9. Xiao J, Lee AH, Vemuri SR (1999) Mixture distribution analysis of length of hospital stay for efficient funding. Socio-Economic Planning Sciences 33: 39 59.
10. Graves N, Weinhold D, Tong E, Birrell F, Doidge S, et al. (2007) Effect of healthcare-acquired infection on length of hospital stay and cost. Infect Control Hosp Epidemiol 28: 280 292.
11. Paradis AR, Stewart VT, Bayley KB, Brown A, Bennett AJ (2009) Excess cost and length of stay associated with voluntary patient safety event reports in hospitals. Amŏ Jŏ Med Qual 24: 53 60.
12. Atienza N, García-Heras J, Muñoz-Pichardo JM, Villa R (2008) An application of mixture distributions in modelization of length of hospital stay. Statistics in Medicine 27: 1403 1420.
13. Mulla ZD, Gibbs SG, Aronoff DM (2007) Correlates of length of stay, cost of care, and mortality among patients hospitalized for necrotizing fasciitis. Epidemiol Infect 135: 868 876.

14. Sobota A, Graham DA, Heeney MM, Neufeld EJ (2009) Corticosteroids for acute chest syndrome in children with sickle cell disease: Variation in use and association with length of stay and readmission. American Journal of Hematology 85: 24 28.
15. Lyratzopoulos G, Havely D, Gemmell I, Cook GA (2005) Factors influencing emergency medical readmission risk in a UK district general hospital: A prospective study. BMC Emerg Med 5: 1.
16. Watson L, Turk F, Rabe KF (2007) Burden of asthma in the hospital setting: an Australian analysis. Int J Clin Pract 61: 1884 1888.
17. Arnold DH, Gebretsadik T, Minton PA, Higgins S, Hartert TV (2008) Assessment of severity measures for acute asthma outcomes: a first step in developing an asthma clinical prediction rule. Am J Emerg Med 26: 473 479.
18. Arnold DH, Gebretsadik T, Abramo TJ, Sheller JR, Resha DJ, et al. (2011) The Acute Asthma Severity Assessment Protocol (AASAP) study: objectives and methods of a study to develop an acute asthma clinical prediction rule. Emergency Medicine Journal.
19. Schenkel S (2000) Promoting Patient Safety and Preventing Medical Error in Emergency Departments. Academic Emergency Medicine 7: 1204 1222.
20. Kuan V, Quantrill SJ (2010) S7 Annual trends in asthma emergency attendance, hospital admission and readmission. Thorax 65: A6 A7.
21. Rowe, Brian H, Villa R, Cristina, Abu L, et al. (2010) Admissions to Canadian hospitals for acute asthma: A prospective, multicentre study. Oakville, ON, CANADA: Pulsus. 6 p.
22. Scott IA (2010) Public hospital bed crisis: too few or too misused? Aust 34: 317 324.
23. HES (2008) HES User Guide. Health and Social Care Information Centre. Leeds: Health and Social Care Information Centre.
24. Watson L, Turk F, James P, Holgate ST (2007) Factors associated with mortality after an asthma admission: a national United Kingdom database analysis. Respir Med 101: 1659 1664.
25. Gerrish K (2000) Researching ethnic diversity in the British NHS: methodological and practical concerns. Journal of Advanced Nursing 31: 918 925.
26. Long JS (2001) Predicted probabilities for count models. Stata Journal 1: 51 57.
27. Singh C, Ladusingh L (2010) Inpatient length of stay: a finite mixture modeling analysis. The European Journal of Health Economics 11: 119 126.
28. Alt JE, King G, Signorino C (1999) Aggregation Among Binary, Count, and Duration Models: Estimating the Same Quantities from Different Levels of Data. Political Analysis Vol. 9: 21 44.
29. Hilbe JM (2011) Negative binomial regression. Cambridge: Cambridge University Press.
30. Long JS, Freese J (2003) Regression Models for Categorical Dependent Variables Using Stata. TX: Stata Press.
31. Burg JG (2001) Seasonal activity and spatial distribution of host-seeking adults of the tick Dermacentor variabilis. Medical and Veterinary Entomology 15: 413 421.
32. Gray B, Haro R, Rogala J (2009) Addressing among-group variation in covariate effects using multilevel models. Environmental and Ecological Statistics 17: 573 591.
33. Cameron AC, Trivedi PK (2007) Essentials of Count Data Regression: Blackwell Publishing Ltd. pp 331 348.
34. GLA (2011) 2010 Mid-Year Population Estimates. In: Demography-Team, editor. Greater London Authority Intelligence Update.
35. DiSantostefano RL, Baxter R, Dale P, Smith HT (2011) Emergency Inpatient Admissions For Asthma In England Based On Hospital Episode Statistics (HES) 2005-2008. Am J Respir Crit Care Med 183: A4751-.
36. Holgate ST, Commins BT, Anderson HR (1995) Asthma and outdoor air pollution. HMSO, London: Department of Health Committee on the Medical Effects of Air Pollutants. Available: http://www.dh.gov.uk/ab/COMEAP/DH_108448. Accessed 1 November 2010.
37. Stewart B, Fairhurst R, Markland J, Marzouk O (2006) Review of calls to NHS Direct related to attendance in the paediatric emergency department. Emergency Medicine Journal 23: 911 914.
38. Moloney ED, Bennett K, O'Riordan D, Silke B (2006) Emergency department census of patients awaiting admission following reorganisation of an admissions process. Emergency Medicine Journal 23: 363 367.

Tables:

Asthma length of stay in hospitals in London 2001-2006: demographic, diagnostic and temporal factors

Ireneous N. Soyiri[1,2*], Daniel D. Reidpath[2], Christophe Sarran[3]

[1]School of Public Health, University of Ghana, Accra, Ghana; [2]Global Public Health, School of Medicine & Health Sciences, Monash University, Sunway Campus, Malaysia; [3]Met Office, Fitzroy Road, Exeter EX1 3PB, United Kingdom

*Correspondence: soyiriin@yahoo.com

Table S1. Summary statistics of asthma and related indicators of hospital admission records in London, 2001-2006

Characteristics	Frequency (%)
Sex	
Male	27,329 (48.1)
Female	29,449 (51.9)
Age (years)	
Under 5	12,420 (12.4)
5-14	10,700 (10.7)
15-44	16,612 (16.6)
45-59	7,029 (7.0)
60-74	5,698 (5.7)
Over 75	4,309 (4.3)
Ethnic Group	
White	26,230 (46.2)
Black	6,604 (11.6)
Asian	6,382 (11.2)
Mixed/Other	5,780 (10.2)
Not stated	11,782 (20.8)
Primary Diagnosis	
Asthma, unspecified	53,637 (94.5)
Non-allergic asthma	182 (0.3)
Mixed asthma	54 (0.1)
Predominantly allergic	2,905 (5.1)
Secondary Diagnosis	
Other diseases of upper respiratory tract	25,053 (44.1)
Influenza and Pneumonia	692 (1.2)
Other acute lower respiratory infections	6,256 (11.0)
Acute upper respiratory infections	70 (0.1)
Chronic lower respiratory infections	1,207 (2.1)
Lung diseases due to external agents	1,519 (2.7)
Other diseases of respiratory system	378 (0.7)
Other non-respiratory system diseases	15,227 (26.8)
Missing Values	6,376 (11.2)

Characteristics	Frequency (%)
Method of Admission	
Accident and emergency services	52,074 (91.7)
General Practitioner (GP)	2,602 (4.6)
Bed bureau	41 (0.1)
Consultants out patient clinic	577 (1.0)
Other means	1,484 (2.6)
Day of the week	
Sunday	5,369 (9.5)
Monday	8,708 (15.3)
Tuesday	9,740 (17.2)
Wednesday	9,060 (16.0)
Thursday	8,705 (15.3)
Friday	9,163 (16.1)
Saturday	6,033 (10.6)
Meteorological Season	
Summer	12,340 (21.7)
Spring	13,453 (23.7)
Autumn	16,800 (29.6)
Winter	14,185 (25.0)
Year of admission	
2001	8,308 (14.6)
2002	8,196 (14.4)
2003	9,141 (16.1)
2004	10,340 (18.2)
2005	10,239 (18.0)
2006	10,554 (18.6)
Birth month	
January	5,350 (9.4)
February	4,384 (7.7)
March	5,045 (8.9)
April	4,644 (8.2)
May	4,472 (7.9)
June	4,591 (8.1)
July	4,700 (8.3)
August	4,637 (8.2)
September	4,748 (8.4)
October	4,803 (8.5)
November	4,629 (8.2)
December	4,775 (8.4)

Tables:

Asthma length of stay in hospitals in London 2001-2006: demographic, diagnostic and temporal factors

Ireneous N. Soyiri[1,2*], Daniel D. Reidpath[2], Christophe Sarran[3]

[1]School of Public Health, University of Ghana, Accra, Ghana; [2]Global Public Health, School of Medicine & Health Sciences, Monash University, Sunway Campus, Malaysia; [3]Met Office, Fitzroy Road, Exeter EX1 3PB, United Kingdom

*Correspondence: soyiriin@yahoo.com

Table S2. Multivariable base and reduced models of length of stay in asthma related hospital admissions in London, 2001-2006

Characteristics	Base Model			Reduced Model		
	§Ratio	[95% C.I.]		§Ratio	[95% C.I.]	
Sex						
Male#	1.00			1.00		
Female	1.11***	1.09	1.13	1.11***	1.09	1.13
Age (years)						
Under 5#	1.00			1.00		
5-14	1.07***	1.04	1.11	1.07***	1.04	1.11
15-44	1.70***	1.66	1.75	1.70***	1.66	1.75
45-59	2.25***	2.18	2.32	2.25***	2.18	2.32
60-74	2.79***	2.70	2.88	2.79***	2.70	2.88
Over 75	3.43***	3.31	3.55	3.43***	3.31	3.55
Ethnic Group						
White#	1.00			1.00		
Black	1.05**	1.02	1.08	1.05**	1.02	1.08
Asian	1.01	0.99	1.04	1.02	0.99	1.04
Mixed/Other	1.00	0.97	1.03	1.00	0.97	1.03
Not stated	0.93***	0.91	0.95	0.93***	0.91	0.95
Primary Diagnosis						
Asthma, unspecified#	1.00			1.00		
Non-allergic asthma	0.98	0.86	1.10	0.97	0.86	1.10
Mixed asthma	0.88	0.70	1.11	0.88	0.70	1.10
Predominantly allergic	0.83***	0.79	0.87	0.83***	0.79	0.87
Secondary Diagnosis						
Other diseases of upper respiratory tract#	1.00			1.00		
Influenza and Pneumonia	1.82***	1.72	1.93	1.82***	1.72	1.93
Other acute lower respiratory infections	1.44***	1.41	1.48	1.44***	1.41	1.48
Acute upper respiratory infections	1.07	0.83	1.38	1.08	0.84	1.38
Chronic lower respiratory infections	1.33***	1.27	1.39	1.33***	1.27	1.39
Lung diseases due to external agents	1.02	0.97	1.08	1.02	0.97	1.08
Other diseases of respiratory system	1.69***	1.57	1.82	1.69***	1.57	1.82
Other Non-respiratory system diseases	1.24***	1.21	1.26	1.24***	1.21	1.26
Missing Values	1.02	0.98	1.05	1.02	0.98	1.05

Characteristics	Base Model			Reduced Model		
	§Ratio	[95% C.I.]		§Ratio	[95% C.I.]	
Method of Admission						
Accident and emergency services#	1.00			1.00		
General Practitioner (GP)	0.90***	0.86	0.93	0.90***	0.86	0.93
Bed bureau	1.15	0.89	1.48	1.14	0.89	1.48
Consultants out patient clinic	1.20***	1.12	1.28	1.20***	1.12	1.28
Other means	0.98	0.93	1.03	0.98	0.93	1.03
Day of the week						
Sunday#	1.00			1.00		
Monday	1.32***	1.28	1.37	1.32***	1.28	1.37
Tuesday	1.37***	1.33	1.42	1.37***	1.33	1.42
Wednesday	1.30***	1.25	1.34	1.30***	1.25	1.34
Thursday	1.31***	1.26	1.35	1.31***	1.26	1.35
Friday	1.28***	1.24	1.33	1.28***	1.24	1.33
Saturday	1.09***	1.05	1.13	1.09***	1.05	1.13
Meteorological Season						
Summer#	1.00			1.00		
Spring	1.05***	1.02	1.07	1.05***	1.02	1.07
Autumn	1.04***	1.02	1.06	1.04***	1.02	1.06
Winter	1.07***	1.05	1.09	1.07***	1.05	1.09
Year of admission						
2001#	1.00			1.00		
2002	0.98	0.95	1.01	0.98	0.95	1.01
2003	0.92***	0.89	0.95	0.92***	0.89	0.95
2004	0.83***	0.80	0.86	0.83***	0.80	0.86
2005	0.78***	0.75	0.80	0.78***	0.75	0.80
2006	0.71***	0.68	0.73	0.71***	0.68	0.73
Birth month						
January#	1.00					
February	0.99	0.95	1.03			
March	0.98	0.95	1.02			
April	0.98	0.94	1.02			
May	0.99	0.95	1.03			
June	1.00	0.96	1.04			
July	0.96*	0.92	1.00			
August	0.97	0.94	1.01			
September	0.98	0.94	1.02			
October	0.96	0.92	1.00			
November	0.98	0.94	1.02			
December	1.01	0.97	1.05			
Akaike Information Criterion (AIC)	*239405.5*			*239394.8*		

§Exponent of the coefficient, which is the expected change in log count for a one-unit increase in a "Characteristic", interpreted as a ratio to the #Reference category; C.I. Confidence Interval; * p<0.05; **p<0.01; *** p<0.001

2

Having established the burden of asthma within the principal population within which we hoped to develop the forecasting models, it is now appropriate to turn the focus towards those models.

The following section (Chapters 4-6) introduces the forecasting models and the application to health forecasting.

SECTION III

III Reviews on health forecasting and potential approaches

Individuals and institutions often desire to know the future, so that informed choices can be made. These choices may involve making alternative plans, reallocating resources to meet unexpected demand, or continue on an already chosen path. A reliable health forecast is thus important for health service delivery because it can: (1) enhance preventive health care/services; (2) create alerts for the management of patient overflows (in situations of peak or reduced demand for health care services); and (3) reduces costs associated with supplies and staff redundancy (54). There are various approaches to predicting the future, including the use of anecdotal evidence, previous experiences, expert intuition, or through formal analytical procedures (214). However, there are hardly any reviews on **health forecasting**, and practically none focused on the methodological aspects.

Section III presents reviews of the forecasting and health forecasting literature with three published papers (chapters 4, 5 & 6). The chapters provide both broad overviews and specific methodological aspects of adaptable forecasting techniques. Chapter 4 serves as a bridge between sections II and III. Within the context of a published review paper it presents the background information on asthma and also proposes two methodological approaches for forecasting the condition. The reviews (Chapters 5 & 6) discuss the types of forecasting & types of health data, principles of health forecasting, as well as the methods (of evaluating error and comparing forecasts) and models/techniques in forecasting. There is necessarily some repletion in the presentation of the ideas, because each of these papers did need to stand alone in their original published form.

Chapter 4

4.0 Non-causal modelling and prediction of asthma admissions

This chapter presents a synopsis of asthma and its global burden, and then discusses two statistical approaches for forecasting the condition. The ideas discussed herein, will support the development of the approaches used in the empirical works reported in section IV.

4.1 Semi-structured black-box prediction: proposed approach for asthma admissions in London

4.1.1 Declarations for Thesis Chapter 4

Monash University

Declaration for Thesis Chapter 4

Declaration by candidate

In the paper: Semi-structured black-box prediction: proposed approach for asthma admissions in London. *Int J Gen Med.* 2012. 5(1): p. 693 - 705, the nature and extent of my contribution to the work was the following:

Nature of contribution	**Extent of contribution (%)**
Lead and corresponding author, conceptualized the idea, accessed, organized and analyzed literature/data, drafted initial manuscript, submitted manuscript to journal and then managed correspondence with editors/reviewers, editorial staff and publishers until the final publication	100

The following co-authors contributed to the work. Co-authors who are students at Monash University must also indicate the extent of their contribution in percentage terms:

Name	**Nature of contribution**	**Extent of contribution (%) for student co-authors only**
Daniel D. REIDPATH	Supervised all activities leading to the publication of the paper; critiqued conceptual ideas, analyses and all drafts of manuscript	N/A

Candidate's Signature	Ireneous N. Soyiri	**Date:** 22-11-2012

Declaration by co-author

The undersigned hereby certify that:

(1) the above declaration correctly reflects the nature and extent of the candidate's contribution to this work, and the nature of the contribution of each of the co-authors.

(2) they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;

(3) they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
(4) there are no other authors of the publication according to these criteria;
(5) potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and
(6) the original data are stored at the following location(s) and will be held for at least five years from the date indicated below:

Location(s)	**Global Public Health, School of Medicine & Health Sciences, Monash University Sunway campus**

Signature 1	Daniel D. Reidpath	**Date:** 22-11-2012

- -

International Journal of General Medicine

Dovepress
open access to scientific and medical research

Open Access Full Text Article

METHODOLOGY

Semistructured black-box prediction: proposed approach for asthma admissions in London

This article was published in the following Dove Press journal:
International Journal of General Medicine
17 August 2012

Ireneous N Soyiri[1,2]
Daniel D Reidpath[1]

[1]Global Public Health, School of Medicine and Health Sciences, Monash University, Malaysia; [2]School of Public Health, University of Ghana, Accra, Ghana

Abstract: Asthma is a global public health problem and the most common chronic disease among children. The factors associated with the condition are diverse, and environmental factors appear to be the leading cause of asthma exacerbation and its worsening disease burden. However, it remains unknown how changes in the environment affect asthma over time, and how temporal or environmental factors predict asthma events. The methodologies for forecasting asthma and other similar chronic conditions are not comprehensively documented anywhere to account for semistructured noncausal forecasting approaches. This paper highlights and discusses practical issues associated with asthma and the environment, and suggests possible approaches for developing decision-making tools in the form of semistructured black-box models, which is relatively new for asthma. Two statistical methods which can potentially be used in predictive modeling and health forecasting for both anticipated and peak events are suggested. Importantly, this paper attempts to bridge the areas of epidemiology, environmental medicine and exposure risks, and health services provision. The ideas discussed herein will support the development and implementation of early warning systems for chronic respiratory conditions in large populations, and ultimately lead to better decision-making tools for improving health service delivery.

Keywords: asthma, health care, black-box forecast, chronic, epidemiology, environment, respiratory risk

Introduction

Asthma is a global public health problem and the most common chronic disease among children. It is underdiagnosed and undertreated, and constitutes a huge burden on individuals, societies, and institutions.[1–5] Current estimates suggest that as many as 300 million people are affected worldwide,[1,6,7] and the burden of this chronic respiratory condition is rising, particularly among children.[1,3,5,8,9] Recent reviews on asthma reaffirm the highly heterogeneous nature of the disease, which is also influenced by a number of complex genetic and environmental factors.[10] Many of these reviews have comprehensively addressed key factors which contribute to the manifestation and progression of asthma in individuals, as well as laboratory-based experiments.[11–13] However, there is little information on the forecasting of asthma events with the purpose of providing early warning systems to help in the management of the condition at the population level.[14]

Asthma has a predictable prognosis.[15] However, its diagnosis remains a challenge because the disease is not clearly defined by a particular set of conditions, but by a mix of several dynamic factors.[2,5,16] These include numerous and quite unpredictable

Correspondence: Ireneous N Soyiri
Global Public Health, School of Medicine and Health Sciences, Monash University, 46150 Bandar Sunway, Selangor DE, Malaysia
Tel +603 5514 6300 ext 61569
Fax +603 5514 6323
Email soyiriin@yahoo.com

submit your manuscript | www.dovepress.com
Dovepress
http://dx.doi.org/10.2147/IJGM.S34647

underlying genetic and environmental factors,[5,17–19] as well as occupational factors.[20,21] As a result of the complex nature of the condition, some of the diagnostic techniques commonly used range from history and patterns of symptoms, physical examination, and lung function measurements including spirometry, through to skin test identification of allergens.[22]

Several studies have shown changes in the global epidemiology of asthma. Developed countries have consistently shown dramatic increases in prevalence, and this change has more recently been observed in some less developed countries as well.[3,17,23,24] The United Kingdom National Asthma Campaign[25] reported that asthma affected over five million people, ie, about one in every five households. Subsequent reports indicated that the United Kingdom had one of the highest prevalence rates, at over 15%.[4] 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.[26] According to the Hospital Episode Statistics of the Department of Health (January 2001–December 2006), London, which is the busiest and most densely populated area in the United Kingdom, recorded about 57,000 asthma-related hospital admissions over that period, giving a crude annual rate of around 9500 (ie, an absolute estimate).[9] 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 health care resources as well as society at large, and therefore a need for preparedness. A significant factor in making preparation is the capacity to forecast what to expect in terms of health systems demands.

The use of algorithms in forecasting, also known as black-box forecasting, provides a novel approach to achieve a better outcome compared with most traditional structural/causal modeling techniques.[27] Black-box forecasting involves a theoretical association of predictors with outcomes, the utility of which is strictly based on forecasting performance.[28] Because it is theoretical, and can be a computationally exhaustive process, this sometimes leads to overfitting and poor performance on unseen datasets.[27,28] On the other hand, structural models account for specific indicators/variables and require substantive knowledge of the subject matter in order to construct an intelligent model.[27] They also overemphasize specific causes in an environment in which the complete causal process is poorly understood. Given the two approaches, a balanced semistructured black-box approach can be useful mid path in developing predictive models for health forecasting, where there is some prior knowledge of the relationship between a health condition and its environmental mediators.

We provide a brief overview of asthma and its environmental causes, with perspectives from the United Kingdom pertinent to other countries with similar populations. One of the key aims of the review is the association between asthma and environmental factors which have potential roles in health forecasting. The other is on developing semistructured black-box approaches that are predictive and can hypothetically forecast asthma events.

Literature search and study approach

In preparing this discussion paper, a scoping of the literature on asthma and associated environmental factors, as well as approaches adopted for managing the condition, was conducted using medical-related databases including PubMed (Medline), Web of Science, and Google Scholar. In addition, citation mapping was used to search and retrace the literature from the initially selected key papers and documents. All the papers and documents identified were synthesized and summarized according to the objective of this paper.

The paper presents reviewed literature on asthma, with a focus on environmental triggers which have been reported, particularly for their effects on respiratory health, including highlights from studies that support the links between environment and health. It further illustrates a wider scope of factors associated with asthma and the manifestation of its symptoms in a framework. The second part of the paper focuses on describing proposed statistical approaches for developing predictive forecasting models. One of these approaches (the negative binomial regression predictive model) is exemplified using data on daily admissions for asthma in London (2001–2006) as well as synthetically generated temporal dummy variables. The paper also briefly discusses a variety of selection strategies for predictive modeling.

Asthma and the environment

Local environmental conditions are important in determining the impact or manifestation of asthma. Factors such as temperature, humidity, and air pressure, as well as air pollutants, all interact to affect the occurrence of asthma, but do not have exclusively independent effects on the condition.[12,29–32] The impact of these complex environmental factors and their interrelationships in health has never been fully understood, even though understanding the key pathways of some individual factors has played an important role in developing a number of therapies.[11,12] The Department of Health and the then Health Protection Agency in the United Kingdom published a review of the health effects of climate change in 2008.[33] In the report, two component issues of the environment, which are principally of interest, are weather and air quality, and these were both considered. It is noted that the constituent indicators of temperature, humidity, vapor/atmospheric pressure, wind, and atmospheric aerosols can produce polluted environments, which are usually recognized as mist, fog, or smog.[34] Hence, environmental pollution and dynamics can exacerbate asthma in many ways,[35,36] and some of the mechanisms involved in this have been discussed subsequently.

Health conditions triggered by local environmental changes, including indoor conditions, as well as occupational exposures vary considerably in their effects and symptoms. These effects are known to depend primarily on the individual's susceptibility and level of exposure to environmental conditions.[20,21,37–40] Vulnerable groups within given populations, particularly children[41,42] and the elderly, tend to be the hardest hit, with the former experiencing both direct and indirect effects of these environmental changes.[39,43] The evidence for environmental effects on health is based on five main types of study:[44]

- health impacts associated with extreme events (eg, heat waves/extreme cold, floods, storms, droughts)
- spatial studies where climate is an explanatory variable in the distribution of the disease or the disease vector
- temporal studies assessing the health effects of change in climate or weather
- experimental laboratory and field studies of vector, pathogen, or plant (allergen) biology
- intervention studies that investigate the effectiveness of public health measures to protect people from environmental exposures.[45]

These types of studies have demonstrated the need to understand fully the health effects of weather and air quality.[46] Thus, dynamic states of weather and air pollutants, which have demonstrated some effect(s) on asthma and its severity in the past, are useful in predicting future occurrences of the condition. Various quantitative procedures have been used to estimate some of these known relationships between a given health condition, such as asthma, and its potential effects.[47,48]

Climate generally affects health,[49] and there is ample evidence of the effect of temperature changes, barometric pressure, and relative humidity on the worsening of asthma symptoms.[50–58] Many of these studies have used the association between weather and disease incidence, hospitalization, or mortality to examine the relationship. For instance, the effect of temperature on general practitioner consultations for respiratory disease was observed, and it was found that there could be up to 15 days of delayed effect of cold temperatures on the incidence of respiratory illness.[59] Also, constant seasonal variability in asthma admissions among children was found in Athens, Greece, where relative humidity and atmospheric pressure were established as key determinants.[60]

The relationship between asthma and environmental conditions is affected in complex ways, and it is worth noting that these effects have different associations depending on 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 temperature change rather than rainfall.[54,57] Furthermore, it has been observed in the United Kingdom and Taiwan that peaks in asthma events occur in the winter and autumn seasons, but not in summer.[59,61,62] Given the importance of context, it becomes critical to understand 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, ie, unusual peaks in asthma events that fall outside the usual fluctuations associated with seasonal changes and variations in weather and air quality. Forecasting these risks is complex and uncertain, but also requires specific data on a very long-term basis.[63] Meanwhile, the use of semistructured black-box approaches in forecasting routine and/or extreme asthma events has not been comprehensively explored.

The issues discussed above are quite global in many respects. The hypothetical flow diagram (Figure 1) illustrates the relationship between asthma symptoms and immediate or underlying causes. Asthma is manifested by an inflammation and/or subsequent obstruction of air flow within the respiratory system.[64,65] It is known that inflammation of the airways in an individual can result in asthmatic symptoms. Alternatively, inflammation may lead to obstruction of air flow directly or indirectly by causing hyperresponsiveness of the airways, ie, a state characterized by easily triggered contraction of the small airways (spasm), which may then cause obstruction of the airways.

Predicting asthma episodes in an ideal situation would require that we account for all the "known" potential predictors/indicators, which of course includes all immediate and underlying mediating factors. However, the availability

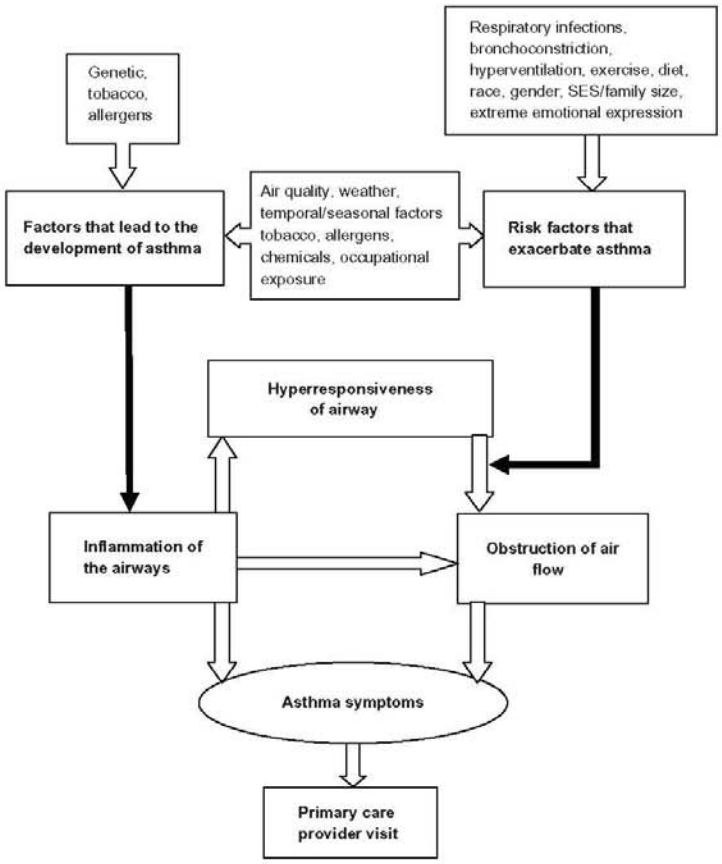


Figure 1 Factors involved in asthma manifestation.
Abbreviation: SES, Socioeconomic status.

of data that is usable is a common limitation. These factors and interrelationships, as illustrated in Figure 1, may present some clues to data sources that can be mined and used for forecasting asthma.

Semistructured black-box modeling

Among the many health issues that can be forecast, the need for emergency care is the commonest form of health forecasting.[66–71] This is particularly related to hospital bed occupancy or number of visits to the emergency room. Although popular, there have been some challenges associated with forecasting the demand for emergency department services. However, this paper does not necessarily focus on these lapses, given that they have been discussed more elaborately elsewhere.[72]

The number of daily asthma admissions or routine measures for similar health conditions can be presented as integer value indicators (also referred to as rate data). This type of count data is common in many disciplines, can form a time series, and thus be used for causal or predictive modeling and forecasting.

Causal models are constructed to provide an explanation of model parameters. Hence, for a given outcome variable, Y will be defined by a function of the variables (X) known to have causal links with the dependent (outcome) variable Y, plus random noise and then the parameter error (Equation 1). In relation to asthma (Figure 1), Y could be represented as "primary care provider visit" for affected individuals, whilst X could be any one or more underlying causes (eg, air quality, weather, and temporal factors) associated with the affected individuals. The autoregressive integrated moving average (ARIMA) is the commonest technique used in this kind of health forecasting.

$$\text{Outcome (Y)} = \text{Function of} \\ (\text{Xs} + \text{random noise} + \text{parameters error}) \quad (1)$$

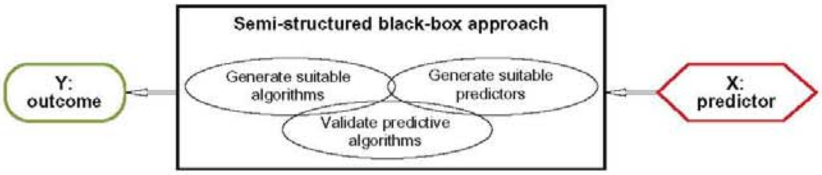


Figure 2 Schematic presentation of semistructured black-box modeling.

In the black-box approach to modeling, formulation of the predictive model does not require prior knowledge of causal links. As illustrated in Figure 2, the process of predicting an outcome involves generation of suitable predictors and models, which are then validated before use in predictions.

There are more studies on causal (structural) modeling than on predictive (black-box) modeling, but the focus of this paper is on semistructured black-box modeling. For the latter approach, even though selection of variables is based originally on prior knowledge, it may also include important predictive factors/variables that have no causal relationship merely because those that end up being used in forecasting are based on their predictive capacity and not just on conformance to a particular theoretical relationship(s). This means that the approach is data-driven. Although data-driven approaches have sometimes created quite tense disagreements between causal modelers and predictive modelers, both approaches have their roles, and in the empirical forecasting and data mining areas, data-driven approaches are generally regarded as superior for the purposes of forecasting and out-of-sample prediction.[28] In some specific context (as in the use of negative binomial models which is described subsequently), their outputs could still have an extended or additional use for describing relationships in past data.

In modeling, count regression models, such as the Poisson or negative binomial model, are most suitable. This is because they have the advantages of being able to handle time series data and their autocorrelations, whilst adjusting for any potential intercorrelation between dependent and independent predictors.[73] Generally, count models are estimated using the maximum likelihood, which computationally proceeds iteratively until there is a convergence of the log likelihood.[74,75] The exact choice of an appropriate count model depends on many factors which are directly related to the properties of the primary variable, such as the skewness of the distribution (kernel density) and the proportion and distribution of "0s" within the dataset. Figure 3 illustrates two major pathways for selecting a suitable count model, ie, a step-by-step approach and a one-stop test selection criterion, which involves the likelihood approach suggested by Long and Freese,[75] and is also available as an application in Stata statistical software.[76]

Using a sample of hospital admission data on asthma, two statistical methods are proposed, ie, a negative binomial model and a quantile regression model, for the development of predictive forecasting models that are aimed at predicting a future/anticipated event(s) and at predicting peak events. The asthma dataset has already been described elsewhere in the nationally recorded hospital episode statistics maintained by the National Health Service in the United Kingdom,[77] and has also been used in some preliminary studies.[9,78] Other data sources from which potential predictors could be extracted or derived include environmental data containing routine measures of weather and air quality indicators. Such data are accessible from the databases of the United Kingdom Meteorological Office. Additional variables (eg, temporal effects like day of the week or month of the year) can also be generated in addition to these data, to help in further investigations.

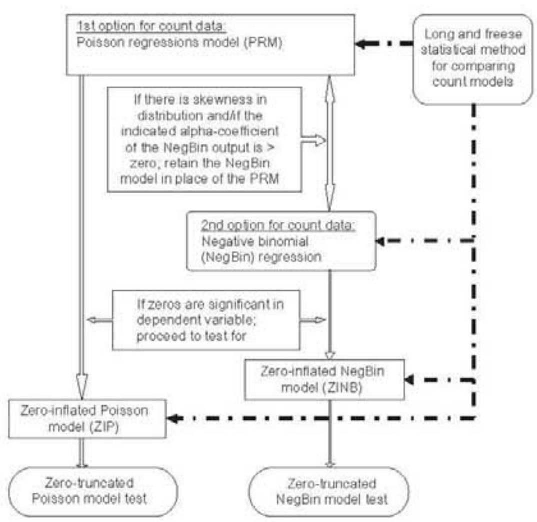


Figure 3 Decision tree for selecting an appropriate count model(s).

Predictive modeling with negative binomial models

Negative binomial models are applicable in developing both univariate and multivariate forecasting models. Given an expected number of daily admissions for asthma, the negative binomial regression can be presented in the form:

$$\Pr(Y = y \mid \lambda, \alpha) = [\Gamma(y + \alpha^{-1})/y!\,\Gamma(\alpha^{-1})] \cdot [\alpha^{-1}/(\alpha^{-1} + \lambda)]^{\alpha^{-1}} \cdot [\lambda/(\alpha^{-1} + \lambda)]^{y} \quad (2)$$

where λ is the mean of the distribution, α is the over dispersion parameter, y is the number of daily asthma admissions, and Γ is the gamma function.

A positive coefficient in the regression output indicates that a factor will increase the number of daily asthma admissions relative to its reference category. Conversely, a negative coefficient will decrease the number of daily asthma admissions relative to its reference category. The exponent of the coefficient can be interpreted, all other things being equal, as the proportionate increase (for values > 1) or decrease (for values between 0 and 1) in number of daily asthma admissions associated with a one unit increase in the explanatory variable.[9,79] Obtaining an improved fit for a model in this situation can be established by inspecting the Akaike information criterion.[9,14,80–82]

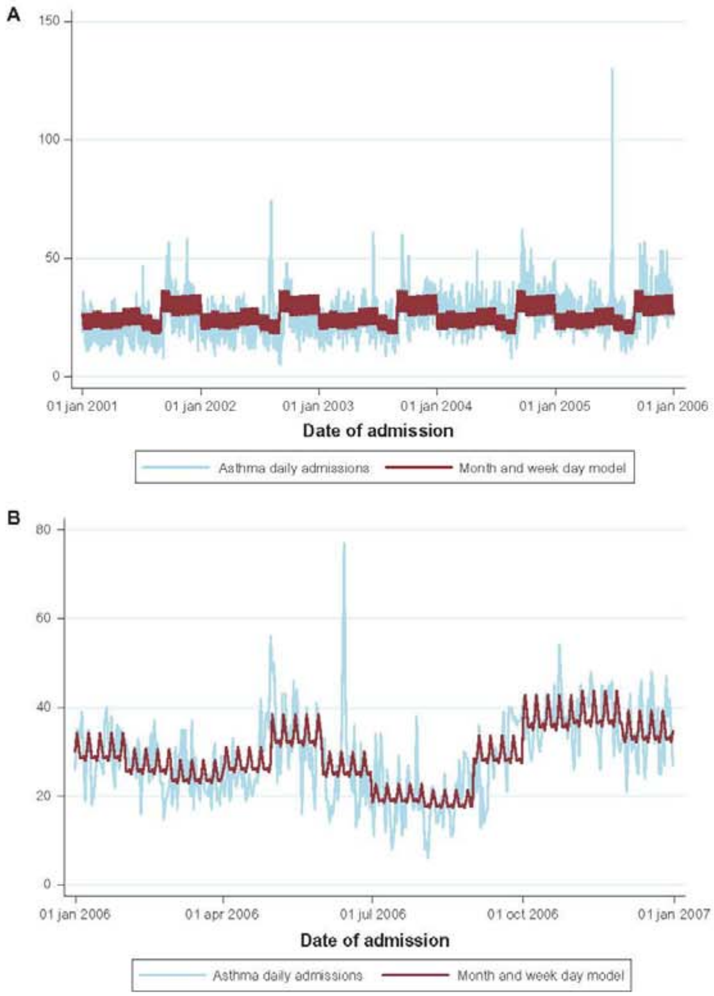


Figure 4 Asthma daily admissions and predictive model based on month and week day. (**A**) Model development sample (hold-in dataset). (**B**) Model validation sample (hold-out dataset).

To exemplify this proposed approach, we present an analysis and outputs based on hospital episode statistics data on asthma in London, which have been described elsewhere.[9] The dataset, which contains nonidentifiable individual records of asthma patients visiting their general practitioner or the emergency department, was transposed by daily sums to provide a time series dataset of total daily attendance. For the purpose of convenience, and also because of the lack of compatible and adequate data on other potential predictors of asthma already mentioned in the literature above (Figure 1), we only illustrate with a temporal predictive model (ie, multivariable model based on temporal factors). The temporal factors we considered included seasons (spring, summer, autumn, and winter), month of the year (January, February through December) and day of the week (Sunday, Monday through Saturday), which were synthetically generated from the date of attendance record using statistical software.[76]

A hold-in sample of the data was used in model development, and validation was done with a hold-out sample

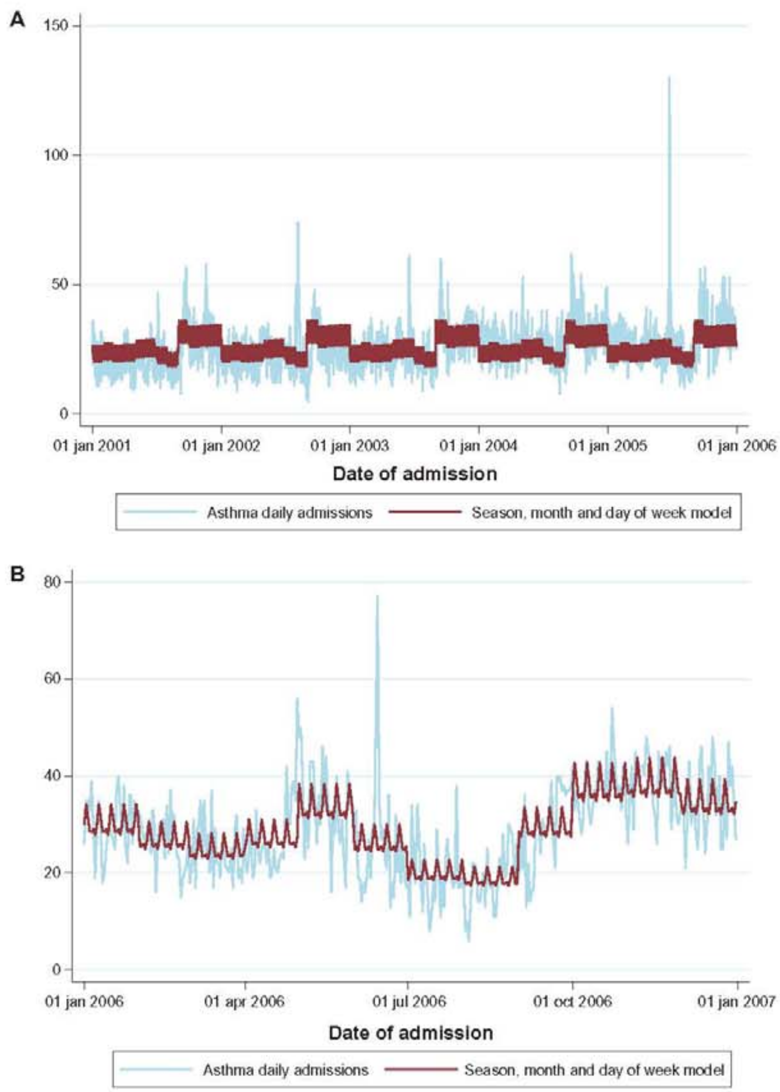


Figure 5 Asthma daily admissions and predictive model based on season month and week day. **(A)** Model development sample (hold-in dataset). **(B)** Model validation sample (hold-out dataset).

of the data from January 1, 2006 to December 31, 2006. Three bivariate and four multivariate predictive models were generated (Figures 4 and 5, Appendix 1) and compared based on their Akaike information criterion (Table 3). These models were cross-validated with the hold-out dataset. For instance, in Figure 4A, the light blue (or gray) time series plot shows total daily asthma admissions in the London area, and the red (or dark) plot shows the predicted multivariate model based on month and week day. However, Figure 4B shows the validation output. The predicted plot reasonably tracks the real distribution of asthma admissions, although it misses out on extreme variations. Even though the focus of the work is on predictive modeling and not causal modeling, the negative binomial regression outputs for these models are also presented (Tables 1 and 2).

Findings from the Akaike information criterion model fitness tests (Table 3) show that the "day of the week model" (III) was the least performing model, whilst the VI and VII multivariate models outperformed all the others, and by as much as 2.4% of III in the hold-in data and over 6.6% in the holdout dataset. Model V was slightly better than both II and IV in the hold-in dataset, and conversely less fit than the latter in the hold-out dataset. Temporal factors have been used to predict exacerbations of chronic respiratory diseases, such as asthma and chronic obstructive pulmonary disease.[83] In addition to temporal factors, other studies have used allergens, weather, and air quality factors in specific areas to predict asthma hospital admissions.[84–86] Hence our findings should be interpreted with caution because the dataset used represents a large and diverse population for London, and does not account for other important predictors of asthma, such as weather and air quality. Nonetheless, predicting asthma events can have policy implications for planning and executing health care delivery. This may then produce some indirect benefits pertaining to health budgets and resource allocation.

Forecasting extreme events with quantile regression models

This paper also proposes the use of quantile regression models in forecasting peak events as a more ideal approach compared with others. Quantile regression is an extension of the linear model, and is better equipped to characterize the relationship between a response distribution and explanatory variable(s) for selective quintiles.[87–89] It has been used more extensively in other areas of forecasting, but has yet to be fully tapped in health forecasting. Quantile regression models can be considered as fitting a linear model to a cross-section of the data/distribution within the anticipated range.

Using the asthma data as an example, for a peak number of daily admissions, the quantile regression model is presented in the form:

$$Y_i = \beta_0^{(p)} + \beta_1^{(p)} x_i + \varepsilon_i^{(p)} \tag{3}$$

where Y_i is asthma hospital admissions for a given day, $\beta_0^{(p)}$ is a constant term, $\beta_1^{(p)}$ is a coefficient of exposure term, x_i is the exposure term, $\varepsilon_i^{(p)}$ is the error term, and p is the quantile. Further illustration of the quantile regression model equation above is available elsewhere.[88,89]

The pseudo R^2 (comparable with the R^2 for least-squares procedures)[14] is the coefficient of determination for quantile regression, and it represents the goodness-of-fit statistic, which is most appropriate for comparing models of specific quantiles.[90,91] It is based on change in the deviance statistic, and ranges between 0 and 1. The pseudo R^2 is thus estimated as:

$$\text{Pseudo } R^2 = 1 - \frac{\text{Sum of deviations about the estimated quantile}}{\text{Sum of deviations about the raw quantile}} \tag{4}$$

Table 1 Bivariate temporal models of daily asthma admissions in London for 2001–2005

Temporal factors	Coefficient	95% CI	
Seasonal model			
Spring#			
Summer	−0.0458*	−0.0872	−0.0043
Autumn	0.2378***	0.1974	0.2783
Winter	0.0758***	0.0347	0.1169
Month of year model			
January#			
February	0.0358	−0.0352	0.1069
March	0.0065	−0.0631	0.0761
April	0.0168	−0.0533	0.0869
May	0.0733*	0.0042	0.1424
June	0.0865**	0.0169	0.1561
July	−0.0252	−0.0951	0.0447
August	−0.1061***	−0.1767	−0.0356
September	0.3112***	0.2431	0.3792
October	0.2451***	0.1771	0.3130
November	0.2553***	0.1868	0.3237
December	0.2628***	0.1950	0.3307
Day of week model			
Sunday#			
Monday	0.1376***	0.0831	0.1922
Tuesday	0.0488	−0.0062	0.1038
Wednesday	−0.0364	−0.0918	0.0190
Thursday	−0.0702**	−0.1258	−0.0146
Friday	−0.0705**	−0.1261	−0.0149
Saturday	−0.1275***	−0.1834	−0.0716

Notes: #Reference category; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.
Abbreviations: Coefficient, coefficient of the negative binomial regression; CI, confidence interval.

Table 2 Multivariable temporal models of asthma daily admissions in London, 2001–2005

Temporal factors	Season and month model			Season and day of week model			Month and day of week model			Season, month, day of week model		
	Coefficient	95% CI		Coefficient	95% CI		Coefficient	95% CI		Coefficient	95% CI	
Season												
Spring[#]												
Summer	−0.1794***	−0.2494	−0.1094	−0.0446*	−0.0847	−0.0045				−0.1781***	−0.2458	−0.1105
Autumn	0.1819***	0.1141	0.2498	0.2370***	0.1979	0.2760				0.1823***	0.1169	0.2477
Winter	−0.0733*	−0.1424	−0.0042	0.0758***	0.0360	0.1156				−0.0731*	−0.1399	−0.0063
Month												
January[#]												
February	0.0358	−0.0352	0.1069				0.0351	−0.0336	0.1037	0.0351	−0.0336	0.1037
March	−0.0668	−0.1359	0.0022				0.0077	−0.0596	0.0750	−0.0654	−0.1321	0.0013
April	−0.0565	−0.1261	0.0131				0.0159	−0.0519	0.0836	−0.0572	−0.1244	0.0099
May	(omitted)	–	–				0.0731*	0.0063	0.1399	(omitted)	–	–
June	0.1926***	0.1221	0.2630				0.0897**	0.0225	0.1569	0.1948***	0.1266	0.2629
July	0.0809*	0.0102	0.1517				−0.0261	−0.0937	0.0414	0.0789*	0.0105	0.1473
August	(omitted)	–	–				−0.1050**	−0.1733	−0.0368	(omitted)	–	–
September	0.0559	−0.0109	0.1227				0.3092***	0.2436	0.3749	0.0538	−0.0105	0.1181
October	−0.0102	−0.0769	0.0565				0.2445***	0.1789	0.3101	−0.0109	−0.0751	0.0533
November	(omitted)	–	–				0.2554***	0.1894	0.3214	(omitted)	–	–
December	0.2628***	0.1950	0.3307				0.2635***	0.1981	0.3290	0.2635***	0.1981	0.3290
Week day												
Sunday[#]												
Monday				0.1382***	0.0868	0.1895	0.1385***	0.0886	0.1884	0.1385***	0.0886	0.1884
Tuesday				0.0494	−0.0024	0.1012	0.0513*	0.0010	0.1016	0.0513*	0.0010	0.1016
Wednesday				−0.0339	−0.0862	0.0184	−0.0335	−0.0843	0.0173	−0.0335	−0.0843	0.0173
Thursday				−0.0685**	−0.1210	−0.0161	−0.0674**	−0.1184	−0.0164	−0.0674**	−0.1184	−0.0164
Friday				−0.0682**	−0.1207	−0.0157	−0.0669**	−0.1179	−0.0159	−0.0669**	−0.1179	−0.0159
Saturday				−0.1243***	−0.1771	−0.0715	−0.1259***	−0.1772	−0.0745	−0.1259***	−0.1772	−0.0745

Notes: [#]Reference category; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.
Abbreviations: Coefficient, coefficient of the negative binomial regression; CI, confidence interval; omitted, omitted because of collinearity.

Table 3 Comparison of model fitness using the Akaike information criterion

Number	Temporal models	Hold-in model		Hold-out model	
		AIC	%	AIC	%
I	Seasonal model	12647.92	0.8	2578.40	3.8
II	Month of year model	12560.37	1.5	2518.14	6.1
III	Day of week model	12746.41	0.0	2681.11	0.0
IV	Season and month model	12560.37	1.5	2518.14	6.1
V	Season and day of week model	12532.29	1.7	2568.72	4.2
VI	Month and day of week model	12436.21	2.4	2504.38	6.6
VII	Season, month, and day of week model	12436.21	2.4	2504.38	6.6

Note: %, percentage improvement of model fit over the least performing model (III).
Abbreviation: AIC, Akaike information criterion.

Variable selection approaches

The variable selection approach is critical in modeling because it determines the final functional form, which is subsequently used for forecasting. As discussed earlier, there is a wide range of environment-related variables known to influence the incidence and/or exacerbation of asthma and other respiratory illnesses. However, common limitations in using environmental measures to forecast a wide population health issue like asthma include a reliable data source and its quality, as well as the extent to which these measures can represent an individual's level of exposure. The inclusion of temporal components as independent factors in a predictive model (eg, day of the week, month, season), accounts for any temporal kinetics and also allows for the identification of lag times, which may improve predictions.[78,92]

There are equally unlimited approaches that can be adopted for the selection of variables in predictive modeling.[27] These approaches range from selection by convenience to computationally exhaustive searches for the best combination of predictors.[27,28] The selection of potential predictors can also be biased by our common understanding of the mechanisms by which environmental agents cause diseases, particularly for respiratory illnesses. However, the ultimate goal is that one obtains a reliable and parsimonious forecasting model.

Backward elimination is one of the commonest approaches utilized to select an appropriate model. This method involves a systematic and/or automatic procedure of reducing a base model, ie, a multivariable model consisting of all possible predictors that are either independently strongly correlated with the dependent variable, or are of already known importance (ie, removing variables that are not statistically significant), while ensuring that the fit (ie, with either AIC or pseudo-R^2) is improved or at least maintained.

Another strategy of variable selection in modeling is to conduct an exhaustive search of the best fitting model using all possible combinations of the available predictors. This approach is however computationally intensive and has a very high chance of over fitting the model. Nonetheless, such computationally exhaustive approaches are most adapted to the novel idea of semistructured black-box models.[28]

Conclusion

Asthma poses a great burden to populations. Discrete measures of the incidence of the disease can be used for forecasting. Though environmental factors have specific effects on the disease, and these effects often vary by location, they may still provide supplementary variables for developing a forecast.

Two methods for developing predictive models as well as variable selection strategies, which have potential roles in semistructured black-box forecasting models, have been discussed. Both negative binomial models and quantile regression models are applicable to integer value health indicators (eg, total daily hospital admissions records). The negative binomial models predict anticipated events, whilst quantile regression models are designed to predict peculiar events. These kinds of forecast models vary in their complexity and methods, depending on the specific health condition and population data. The idea of semistructured black-box predictive modeling may stimulate further research on asthma and, possibly, health forecasting.

Disclosure

The authors declare they have no competing interests in this work.

References

1. Global Initiative for Asthma. Global Strategy for Asthma Management and Prevention, 2008. Available from: http://www.ginasthma.com/Guidelineitem.asp??l1=2&l2=1&intId=60. Accessed February 1, 2009.

2. Kaur B, Anderson HR, Austin J, 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):118–124.
3. Pearce N, Ait-Khaled N, Beasley R, 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): 758–766.
4. Masoli M, Fabian D, Holt S, Beasley R. The global burden of asthma: executive summary of the GINA Dissemination Committee report. *Allergy*. 2004;59(5):469–478.
5. Subbarao P, Mandhane PJ, Sears MR. Asthma: epidemiology, etiology and risk factors. *CMAJ*. 2009;181(9):E181–E190.
6. Fanta CH. Asthma. *N Engl J Med*. 2009;360(10):1002–1014.
7. Dougherty RH, Fahy JV. Acute exacerbations of asthma: epidemiology, biology and the exacerbation-prone phenotype. *Clin Exp Allergy*. 2009;39(2):193–202.
8. Neville RG, McCowan C, Hoskins G, Thomas G. Cross-sectional observations on the natural history of asthma. *Br J Gen Pract*. 2001; 51(466):361–365.
9. Soyiri IN, Reidpath DD, Sarran C. Asthma length of stay in hospitals in London 2001–2006: demographic, diagnostic and temporal factors. *PLoS One*. 2011;6(11):e27184.
10. Editorial. A fresh perspective on asthma. *Nat Med*. 2012;18(5):631.
11. Holt PG, Sly PD. Viral infections and atopy in asthma pathogenesis: new rationales for asthma prevention and treatment. *Nat Med*. 2012;18(5): 726–735.
12. Holgate ST. Innate and adaptive immune responses in asthma. *Nat Med*. 2012;18(5):673–683.
13. Rodrigo GJ, Rodrigo C, Hall JB. Acute asthma in adults: a review. *Chest*. 2004;125(3):1081–1102.
14. Soyiri IN, Reidpath DD. Evolving forecasting classifications and applications in health forecasting. *Int J Gen Med*. 2012;5(1): 381–389.
15. Panettieri RA Jr, Covar R, Grant E, Hillyer EV, Bacharier L. Natural history of asthma: persistence versus progression – does the beginning predict the end? *J Allergy Clin Immunol*. 2008;121(3): 607–613.
16. Fishwick D, Curran AD. Variability in the diagnosis of occupational asthma and implications for clinical practice. *Curr Opin Allergy Clin Immunol*. 2008;8(2):140–144.
17. Holgate ST. The epidemic of asthma and allergy. *J R Soc Med*. 2004;97(3):103–110.
18. Galant SP, Crawford LJR, Morphew T, Jones CA, Bassin S. Predictive value of a cross-cultural asthma case-detection tool in an elementary school population. *Pediatrics*. 2004;114(3):e307–e316.
19. Gaffin JM, Phipatanakul W. The role of indoor allergens in the development of asthma. *Curr Opin Allergy Clin Immunol*. 2009;9(2): 128–135.
20. Darby AC, Waterhouse JC, Stevens V, et al. Chronic obstructive pulmonary disease among residents of an historically industrialised area. *Thorax*. 2012. Epub 2012 Jun 28.
21. Fishwick D, Barber CM, Darby AC. Chronic obstructive pulmonary disease and the workplace. *Chron Respir Dis*. 2010;7(2):113–122.
22. British Thoracic Society Scottish Intercollegiate Guidelines N. British guideline on the management of asthma. *Thorax*. 2008;63 Suppl 4: iv1–iv121.
23. Holgate ST. Lessons learnt from the epidemic of asthma. *QJM*. 2004;97(5):247–257.
24. Nowak D, Suppli Ulrik C, von Mutius E. Asthma and atopy: has peak prevalence been reached? *Eur Respir J*. 2004;23(3):359–360.
25. National Asthma Campaign. Out in the open: a true picture of asthma in the United Kingdom today. *Asthma Journal*. 2001;6(Suppl 3): 3–14.
26. Asthma UK. Wish you were here, 2008. Available from: http://www.asthma.org.uk/media/141047/wish_you_were_here__-_uk_report.pdf. Accessed July 29, 2012.
27. Chatfield C. *The Analysis of Time Series: An Introduction*. 6th ed. London, UK: CRC Press; 2004.
28. Breiman L. Statistical modeling: the two cultures. *Stat Sci*. 2001;16(3): 199–215.
29. Ma L, Shima M, Yoda Y, et al. Effects of airborne particulate matter on respiratory morbidity in asthmatic children. *J Epidemiol*. 2008;18(3): 97–110.
30. Parker JD, Akinbami LJ, Woodruff TJ. Air pollution and childhood respiratory allergies in the United States. *Environ Health Perspect*. 2009;117(1):140–147.
31. Diette GB, McCormack MC, Hansel NN, Breysse PN, Matsui EC. Environmental issues in managing asthma. *Respir Care*. 2008;53(5): 602–615.
32. Gent JF, Triche EW, Holford TR, et al. Association of low-level ozone and fine particles with respiratory symptoms in children with asthma. *JAMA*. 2003;290(14):1859–1867.
33. Department of Health, Health Protection Agency. Health effects of climate change in the UK 2008: An update of the Department of Health report 2001/2002. Department of Health, Health Protection Agency. December 29, 2008. Available from: http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_080702. Accessed July 29, 2012.
34. Met Office. Climate science explained, 2009. Pope V, Haywood J, climate scientists discussants. Available from: http://www.metoffice.gov.uk/climatechange/science/explained/index.html. Accessed February 1, 2009.
35. Lin S, Gomez MI, Hwang S-A, Munsie JP, Fitzgerald EF. Self-reported home environmental risk factors for childhood asthma: a cross-sectional study of children in Buffalo, New York. *J Asthma*. 2008;45(4): 325–332.
36. Becklake MR, Ernst P. Environmental factors. *Lancet*. 1997;350: S10–S13.
37. Zeka A, Eisen EA, Kriebel D, Gore R, Wegman DH. Risk of upper aerodigestive tract cancers in a case-cohort study of autoworkers exposed to metalworking fluids. *Occup Environ Med*. 2004;61(5): 426–431.
38. Fouillet A, Rey G, Wagner V, 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): 309–317.
39. Kovats RS, Hajat S. Heat stress and public health: a critical review. *Annu Rev Public Health*. 2008;29:41–55.
40. Koskela HO. Cold air-provoked respiratory symptoms: the mechanisms and management. *Int J Circumpolar Health*. 2007;66(2): 91–100.
41. Raphoz M, Goldberg MS, Garneau M, Héguy L, Valois M-F, Guay F. Associations between atmospheric concentrations of spores and emergency department visits for asthma among children living in Montreal. *Arch Environ Occup Health*. 2010;65(4):201–210.
42. Héguy L, Garneau M, Goldberg MS, Raphoz M, Guay F, Valois M-F. Associations between grass and weed pollen and emergency department visits for asthma among children in Montreal. *Environ Res*. 2008;106(2):203–211.
43. O'Shea SD, Taylor NF, Paratz JD. But watch out for the weather: factors affecting adherence to progressive resistance exercise for persons with COPD. *J Cardiopulm Rehabil Prev*. 2007;27(3): 166–174.
44. Menne B, Kovats S, Bell J. Climate change and health: international research agenda. European Preparatory Meeting for the Global Ministerial Forum on Research for Health. Copenhagen, Denmark, April 29–30, 2008.
45. Intergovernmental Panel on Climate Change, Andy Jorgensen, lead author. IPCC Fourth Assessment Report, Working Group II: Chapter 8. Encyclopedia of Earth. Cleveland CJ, editor. *Washington, DC: Environmental Information Coalition, National Council for Science and the Environment*. Washington, DC: Encyclopedia of Earth; 2010.
46. Sauerborn R. Climate change: an agenda for research and teaching in public health. *Scand J Public Health*. 2007;35(6):561–563.

47. Schonwetter RS, Clark LD, Leedy SA, Quinn MJ, Azer M, Kim S. Predicting emergency room visits and hospitalizations among hospice patients with cardiac disease. *J Palliat Med*. 2008;11(8):1142–1150.
48. Tobias A, Saez M, Galan I, Campbell MJ. Sensitivity analysis of common statistical models used to study the short-term effects of air pollution on health. *Int J Biometeorol*. 2003;47(4):227–229.
49. Intergovernmental Panel on Climate Change. *Assessing the Health Impacts of Climate Change*. Geneva, Switzerland: World Meteorological Organization/United Nations Environment Programme; 1995.
50. Storr J, Lenney W. School holidays and admissions with asthma. *Arch Dis Child*. 1989;64(1):103–107.
51. Rossi OV, Kinnula VL, Tienari J, Huhti E. Association of severe asthma attacks with weather, pollen, and air pollutants. *Thorax*. 1993;48(3):244–248.
52. Epton MJ, Martin IR, Graham P, et al. Climate and aeroallergen levels in asthma: a 12 month prospective study. *Thorax*. 1997;52(6): 528–534.
53. Hales S, Lewis S, Slater T, Crane J, Pearce N. Prevalence of adult asthma symptoms in relation to climate in New Zealand. *Environ Health Perspect*. 1998;106(9):607–610.
54. Weiland SK, Husing A, Strachan DP, Rzehak P, Pearce N. Climate and the prevalence of symptoms of asthma, allergic rhinitis, and atopic eczema in children. *Occup Environ Med*. 2004;61(7):609–615.
55. Chavarria JF. Short report: asthma admissions and weather conditions in Costa Rica. *Arch Dis Child*. 2001;84(6):514–515.
56. Ivey MA, Simeon DT, Monteil MA. 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): 1526–1530.
57. Khot A, Burn R, Evans N, Lenney W, Storr J. Biometeorological triggers in childhood asthma. *Clin Allergy*. 1988;18(4):351–358.
58. Johnston NW, Sears MR. Asthma exacerbations. 1: Epidemiology. *Thorax*. 2006;61(8):722–728.
59. Hajat S, Haines A. Associations of cold temperatures with GP consultations for respiratory and cardiovascular disease amongst the elderly in London. *Int J Epidemiol*. 2002;31(4):825–830.
60. Priftis KN, Paliatsos AG, Panagiotopoulou-Gartagani P, et al. Association of weather conditions with childhood admissions for wheezy bronchitis or asthma in Athens. *Respiration*. 2006;73(6):783–790.
61. Hospital Episode Statistics online. Asthma, Health and seasons. Health and Social Care Information Centre, 2009. Available from: http://www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=1937&category ID=972. Accessed January 1, 2009.
62. Yang C-Y, Chen C-C, Chen C-Y, Kuo H-W. Air pollution and hospital admissions for asthma in a subtropical city: Taipei, Taiwan. *J Toxicol Environ Health A*. 2007;70(2):111–117.
63. Kim J, Ha J-S, Jun S, Park T-S, Kim H. The weather watch/warning system for stroke and asthma in South Korea. *Int J Environ Health Res*. 2008;18(2):117–127.
64. Global Initiative for Asthma. Global Burden of Asthma, 2004. Available from: http://www.ginasthma.com/download.asp?intid=29. Accessed February 1, 2009.
65. Holgate ST, Commins BT, Anderson HR. *Asthma and Outdoor Air Pollution*. London, UK: Department of Health Committee on the Medical Effects of Air Pollutants; 1995. Available from: http://www.comeap.org.uk/documents/reports.html. Accessed August 13, 2012.
66. Jones SA, Joy MP, Pearson J. Forecasting demand of emergency care. *Health Care Manage Sci*. 2002;5(4):297–305.
67. Boyle J, Jessup M, Crilly J, et al. Predicting emergency department admissions. *Emerg Med J*. 2012;29(5):358–365.
68. Bibi H, Nutman A, Shoseyov D, et al. Prediction of emergency department visits for respiratory symptoms using an artificial neural network. *Chest*. 2002;122(5):1627–1632.
69. Blaisdell CJ, Weiss SR, Kimes DS, et al. Using seasonal variations in asthma hospitalizations in children to predict hospitalization frequency. *J Asthma*. 2002;39(7):567–575.
70. Schweigler LM, Desmond JS, McCarthy ML, Bukowski KJ, Ionides EL, Younger JG. Forecasting models of emergency department crowding. *Acad Emerg Med*. 2009;16(4):301–308.
71. Wiler JL, Griffey RT, Olsen T. Review of modeling approaches for emergency department patient flow and crowding research. *Acad Emerg Med*. 2011;18(12):1371–1379.
72. McCarthy ML, Zeger SL, Ding R, Aronsky D, Hoot NR, Kelen GD. The challenge of predicting demand for emergency department services. *Acad Emerg Med*. 2008;15(4):337–346.
73. Novikov I, Kalter-Leibovici O, Chetrit A, Stav N, Epstein Y. Weather conditions and visits to the medical wing of emergency rooms in a metropolitan area during the warm season in Israel: a predictive model. *Int J Biometeorol*. 2011;56(1):121–127.
74. Long JS, Freese J. Predicted probabilities for count models. *Stata J*. 2001;1(1):51–57.
75. Long JS, Freese J. *Regression Models for Categorical Dependent Variables Using Stata*. College Station, TX: Stata Press; 2003.
76. Stata SE. Release 11. *Statistical Software*. College Station, TX: StataCorp LP; 2009.
77. Hospital Episode Statistics. HES User Guide. Health and Social Care Information Centre 2008. Available from: http://www.hesonline.nhs.uk. Accessed July 29, 2009.
78. Soyiri IN, Reidpath DD. Predicting extreme asthma events in London using quantile regression models. *Emerg Health Threats J*. 2011;4(162): 39–40.
79. Hilbe JM. *Negative Binomial Regression*. 2nd ed. Cambridge, UK: Cambridge University Press; 2011.
80. Burg JG. Seasonal activity and spatial distribution of host-seeking adults of the tick Dermacentor variabilis. *Med Vet Entomol*. 2001;15(4): 413–421.
81. Gray B, Haro R, Rogala J. Addressing among-group variation in covariate effects using multilevel models. *Environ Ecol Stat*. 2009; 17(4):573–591.
82. Cameron AC, Trivedi PK. *Essentials of Count Data Regression*: Malden, MA: Blackwell Publishing Ltd; 2007.
83. Moineddin R, Nie J, Domb G, Leong A, Upshur R. Seasonality of primary care utilization for respiratory diseases in Ontario: A time-series analysis. *BMC Health Services Res*. 2008;8(1):160.
84. Sheppard L, Levy D, Norris G, Larson TV, Koenig JQ. Effects of ambient air pollution on nonelderly asthma hospital admissions in Seattle, Washington, 1987–1994. *Epidemiology*. 1999;10(1): 23–30.
85. Raphoz M, Goldberg MS, Garneau M, Héguy L, Valois M-F, Guay F. Associations between atmospheric concentrations of spores and emergency department visits for asthma among children living in Montreal. *Arch Environ Occup Health*. 2010;65(4):201–210.
86. Nastos PT. Weather, ambient air pollution and bronchial asthma in Athens, Greece. In: Thomson M, Herrera RG, Beniston M, editors. *Seasonal Forecasts, Climatic Change and Human Health. Health and Climate. Series: Advances in Global Change Research*. New York, NY: Springer; 2006.
87. Barbosa SM. Quantile trends in Baltic sea level. *Geophys Res Lett*. 2008;35:L22704.
88. Hao L, Naiman DQ, editors. Quantile regression. In: *Quantitative Applications in the Social Sciences*. Thousand Oaks, CA: Sage Publications Inc; 2007.
89. Koenker R. *Quantile Regression*. New York, NY: Cambridge Press; 2005.
90. Zietz J, Zietz E, Sirmans G. Determinants of house prices: a quantile regression approach. *J Real Estate Finan Econ*. 2008;37(4):317–333.
91. Barnes ML, Hughes AW. A quantile regression analysis of the cross section of stock market returns. In: *EconPapers No. 02-2, Working Papers*. Federal Reserve Bank of Boston; 2002 [updated June 6, 2012]. Available from: http://EconPapers.repec.org/RePEc:fip:fedbwp:02-2. Accessed November 1, 2010.
92. Racloz V, Ramsey R, Tong S, Hu W. Surveillance of dengue fever virus: a review of epidemiological models and early warning systems. *PLoS Negl Trop Dis*. 2012;6(5):e1648.

Chapter 5

5.0 Health forecasting

In addition to providing general background information on health forecasting and a discussion of the key issues, the overview of health forecasting (chapter 5), also presents a *schematic approach* to health forecasting (54). The importance of including the schematic approach was in part to address the basic potential needs of public health practitioners who may be interested in developing, testing and maintaining a simple but pragmatic health forecasting plan, which has been discussed (215, 216), but implemented less successfully. Chapter 5 further highlights the successes and challenges associated with the practical implementation of health forecasting using programmes in the UK as case examples.

5.1 An overview of health forecasting

5.1.1 Declarations for Thesis Chapter 5

Monash University

Declaration for Thesis Chapter 5

Declaration by candidate

In the case of Chapter 5: *An overview of health forecasting* (DOI: 10.1007/s12199-012-0294-6, *Env Health Prev Med*. 2013), the nature and extent of my contribution to the work was the following:

Nature of contribution	**Extent of contribution (%)**
Lead and corresponding author, conceptualized the idea, accessed and organized the literature, drafted initial manuscript, submitted manuscript to journal and then managed correspondence with editors/reviewers, editorial staff and publishers until the final publication	100

The following co-authors contributed to the work. Co-authors who are students at Monash University must also indicate the extent of their contribution in percentage terms:

Name	**Nature of contribution**	**Extent of contribution (%) for student co-authors only**
Daniel D. REIDPATH	Supervised all activities leading to the publication of the paper; critiqued conceptual ideas, analyses and all drafts of manuscript	N/A

Candidate's Signature	Ireneous N. Soyiri	**Date:** 22-11-2012

Declaration by co-author

The undersigned hereby certify that:

(1) the above declaration correctly reflects the nature and extent of the candidate's contribution to this work, and the nature of the contribution of each of the co-authors.

(2) they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;

(3) they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
(4) there are no other authors of the publication according to these criteria;
(5) potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and
(6) the original data are stored at the following location(s) and will be held for at least five years from the date indicated below:

Location(s)	**Global Public Health, School of Medicine & Health Sciences, Monash University Sunway campus**

Signature 1	Daniel D. Reidpath	**Date:** 22-11-2012

- -

Environ Health Prev Med
DOI 10.1007/s12199-012-0294-6

REVIEW

An overview of health forecasting

Ireneous N. Soyiri · Daniel D. Reidpath

Received: 14 March 2012 / Accepted: 25 June 2012

Abstract Health forecasting is a novel area of forecasting, and a valuable tool for predicting future health events or situations such as demands for health services and healthcare needs. It facilitates preventive medicine and health care intervention strategies, by pre-informing health service providers to take appropriate mitigating actions to minimize risks and manage demand. Health forecasting requires reliable data, information and appropriate analytical tools for the prediction of specific health conditions or situations. There is no single approach to health forecasting, and so various methods have often been adopted to forecast aggregate or specific health conditions. Meanwhile, there are no defined health forecasting horizons (time frames) to match the choices of health forecasting methods/approaches that are often applied. The key principles of health forecasting have not also been adequately described to guide the process. This paper provides a brief introduction and theoretical analysis of health forecasting. It describes the key issues that are important for health forecasting, including: definitions, principles of health forecasting, and the properties of health data, which influence the choices of health forecasting methods. Other matters related to the value of health forecasting, and the general challenges associated with developing and using health forecasting services are discussed. This overview is a stimulus for further discussions on standardizing health forecasting approaches and methods that will facilitate health care and health services delivery.

Keywords Health forecasting · Principles · Prediction · Forecasting horizon · Health services

Introduction

Forecasting is about predicting future events based on a foreknowledge acquired through a systematic process or intuition [1, 2]. Some of the earliest forms of health forecasting date back to the period of Hippocrates of Cos (460 BC–370 BC). Hippocrates studied the natural history of diseases and their major environmental sources (including food and water) [3], and believed that *prognosis* was an important part of medical treatment, because by forecasting disease outcome, the physician established his expertise for treating the patient [4]. He was able to develop and to forecast the occurrence of many diseases and conditions. One of the classical terms in medicine, '*Hippocratic facies*', describes the procedure for forecasting impending death based on the observation of distinctive signs and symptoms that he identified [5]. The birth of forecasting as a science, however, is associated with weather forecasting and, is credited to Francis Beaufort, who developed the popularly known scale for measuring wind force (the Beaufort scale) and Robert Fitzroy, who developed the Fitzroy barometer for measuring atmospheric pressure [6]. Forecasting has advanced over time and has increased in sophistication in many specialised areas, including the fields of health [7–10], economics and commerce [1, 11], sports [12, 13], environment (including meteorology) [14, 15], technology and politics [16–18].

Two approaches to forecasting, statistical and judgmental are widely discussed in the forecasting literature [19]. An integration of both approaches has been discussed by some as the best way to obtain a more reliable forecast [19–22].

I. N. Soyiri · D. D. Reidpath
Global Public Health, Jeffrey Cheah School of Medicine and Health Sciences, Monash University, 46150, Bandar Sunway, Selangor, Malaysia

I. N. Soyiri (✉)
School of Public Health, University of Ghana, Legon, P.O. Box LG 13, Accra, Ghana
e-mail: soyiriin@yahoo.com

Published online: 28 July 2012

Monitoring population health, which includes demographic and health surveillances and epidemiological studies on disease surveillance, can generate very useful data that can be used in health forecasting. A reliable health forecast is important for health service delivery, because it can: (1) enhance preventive health care/services; (2) create alerts for the management of patient overflows (in situations of peak demand for health care services); and (3) significantly reduce the associated costs in supplies and staff redundancy.

Health forecasting is a useful tool for health service provision, but very few reviews on the subject exist. Some previous studies on health forecasting focused on very specific conditions, like ischaemic heart disease [23], chronic obstructive pulmonary disease (COPD) [24], diabetes prevalence [25], or aggregate health situations, such as emergency department visits [26, 27]. These individual studies adapted environmental, climatic and other factors as predictors in forecasting health. They are very specific and do not give information on general approaches that could guide the development of other health forecasts. An overview published by Ioannidis discussed the limits of forecasting in personalised medicine and focused only on challenges associated with this form of health forecasting [28]. A systematic review conducted by Wargon et al. [26] on models for forecasting focused only on the number of emergency department visits. More recently, a similar study was conducted by Boyle et al. [27], in which they reviewed and predicted emergency department admissions. The above-mentioned reviews on health forecasting had a very specific focus on emergency attendance. However, health forecasting possesses potential applications across a wider range of health issues. There is dearth of information pertaining to the many possible applications of health forecasting in relation to health service delivery. There seem to be no reports that have gathered the basic principles and procedures for developing pragmatic health forecasting schemes.

This paper describes the key issues of health forecasting; including definitions, principles of health forecasting, and the properties of health data, which influence the choice of health forecasting methods. It also identifies the values of health forecasting in health service provision, and further presents the general challenges associated with developing and using health forecasting services.

In preparing this review, a search of the literature on health forecasting and statistical methods used in the analysis of health conditions/situations was conducted, using popular medical-related databases including PubMed (Medline) and then Google Scholar. Our search strategy included: "health" and "forecast*" and combinations of terms, including "principle*", "data", "predict*", "model*", "method*", "challenge*". Based on the titles and subsequently on the abstracts, articles unrelated to our objectives were excluded. Additional literature searches were done through citation mapping of key papers and the selected papers and documents were synthesized and summarized according to the set objective of this paper.

Defining health forecasting and related terminologies

Health forecasting is predicting health situations or disease episodes and forewarning future events. It is also a form of preventive medicine or preventive care that engages public health planning and is aimed at facilitating health care service provision in populations [8, 10, 29, 30]. Health forecasting has been commonly applied to emergency department visits, daily hospital attendance and admissions [27, 31–33].

There are important terms in forecasting that are worth noting because of the way in which they are used across various fields. The term *prediction* is mainly used across several fields of study to mean an opinion-based speculation with no explicit causal assumptions [34]. In the health forecasting literature, however, the terms *prediction* and *prognosis* could mean different things, even though they are sometimes used interchangeably and without clarity. The term *prognosis* refers to a forecasting of outcomes under no intervention, whilst *prediction* is used to mean forecasting health outcomes that are associated with some health-related intervention [28, 35]. *Syndromic surveillance* is another closely related concept that is well known in disease surveillance literature. This concept focuses on case detection and events that lead to/precede an outbreak, and involves detecting aberrations in the patterns of diseases and using this information to determine future outbreaks [36–38]. Syndromic surveillance was initially developed as an innovative electronic surveillance system and was aimed at improving early detection of outbreaks attributable to biologic terrorism or other causes [36].

Forecasting is a key component in the practice of medicine, with the main purpose of improving both health service provision and individual patient outcome [10, 24, 30, 39, 40]. For example, the United Kingdom Meteorological Office developed a health forecasting service for COPD, which provides health alerts to both individuals and health service providers through an automated call system [7, 24, 41]. This forecast combines a rule-based model that predicts risks based on environmental conditions, with an anticipatory care intervention to provide information, which is then communicated. The service enables patients and care providers to take precautionary actions to improve health service delivery and reduce COPD events [7, 10, 30, 42].

Principles of health forecasting

There are four main principles of health forecasting: (1) the measure of uncertainty and errors, (2) the focus, (3) the nature of data aggregation and how it affects accuracy, and (4) the horizon of health forecasting. These properties are not only hypothetically important, but also have applications that are exemplified in the literature, as discussed below.

Uncertainty and error of health forecasting

According to the definition of health forecasting, determining future health events or situations involves a degree of uncertainty, as it is virtually impossible to have a perfect (i.e. 100 % error free) prediction. We therefore describe the measurement of uncertainty and error of health forecasting as a principle in forecasting, because it is a basic requirement, and is also desirable for validation and determining the real value of a forecast. The data used is a major source of uncertainty and error, but this basic problem can partly be addressed methodologically, to obtain health forecasts with the least possible error [43].

The focus of health forecasting

The focus of a health forecast relates to the central targeted issue that is being forecast. This is with reference to the basic unit of the health outcome measure that is being forecast. One focus of health forecasting is to predict population health outcome in terms of the number of events occurring within a space of time; for example, the forecasting of life expectancy and health expectancies [44]. Another focus is to determine the course of an ailment for a particular individual, which is usually referred to as prognosis [28]. These two categories are related to how the data is aggregated in health forecasting.

Data aggregation and accuracy of health forecasting

Forecasting a health condition or situation for a population aggregate of a particular problem, or for groups of the same *family*, presents a lesser challenge than doing so for an individual case. This is because by pooling the variances of the population-related factors (which are usually broad and well known), the behavior of the aggregated data can have very stable characteristics, even when the individuals within exhibit high degrees of randomness [45]. It is therefore easier to obtain a higher degree of accuracy in forecasting specific health events when using pooled population data versus data for specific individuals.

Horizons of health forecasting

A health forecasting horizon refers to the range of the period the forecast is intended to cover. The demand for a health forecast determines the forecast horizon (range), and this could be in a short, medium or long term. There are no clearly defined boundaries to health forecast horizons in the literature. However, borrowing the common classifications from other disciplines such as finance, business or econometric forecasting, a short-range forecast horizon refers to a period of 1 day to a quarter of a year; a medium-range forecast horizon refers to a quarter of a year to a year; and long-range forecasts refer to a year to five or more years. These horizons are, however, not fixed for all situations, but rather may be defined in relation to the qualitative indicator being forecast (e.g. life expectancy), as well as its weighting over an extended time period. Major population health issues, such as life expectancy or future health expectancies [44], or the forecasting of some chronic disease prevalence (i.e. obesity and diabetes) in large populations [25, 46], are often forecast with a long range. Short-range and medium-range health forecasts are applicable to routine health service uptake (e.g. hospital visits), and some chronic disease exacerbations resulting from environmental exposures [7, 24]. The choice of a long-range, medium-range, or short-range forecast is critical in developing a forecast,as health forecasting horizons also have applications in the planning of health care service deliveries.

The discussions around short, medium, and long range health forecasting do not identify some of the fundamental differences in assumptions between the various forecasting horizons. Yet, these differences are important since forecasting future events is based on a strong assumption that the current drivers or predictors will also follow the trend over the future horizon. Hence, long-range forecasting models will be prone to having more “shocks” compared to short-term forecasts. The “shocks” herein refer to disruptions/disturbances of function of the distributions’ equilibrium, which is caused by a significant change in magnitude of the forecast model predictor(s). This may then lead to a shift in the trend. Shocks also have effects on forecast errors because their occurrence, which is between the time of the forecast and the realization of the outcome, determines the error of the forecast. However, research on the mechanisms by which health forecasting models are developed to accommodate shocks at various thresholds is not explicit.

The principles discussed above also serve as a guide to creating simple decision tools for health forecasting, based on: the type, amount and distribution of data (the kernel density) that is required by a quantitative predictive model; the forecasting horizon for which the health forecast is being created; and then the degree of accuracy or error that is acceptable—taking into account the need for a parsimonious model. The type of data, as described elsewhere [47], refers to the classification of the data as either continuous (ratio or interval scales) or categorical (ordinal, nominal, or dichotomous scales); the amount of data simply refers to the sample size or total number of the unit of reference of the primary variable and its corresponding independent variables/predictors. The section below exemplifies a hypothetical approach and framework for developing a health forecasting scheme with simple decision tools.

A schematic approach to health forecasting

A framework for health forecasting is an essential guide. It is, however, uncommon in the literature and so the following framework, which presents a summary of the key processes involved in developing a general health forecasting service, is illustrated below (Fig. 1). The steps help to identify and broadly define the needs and tools of health forecasting. Further, they state the key processes involved in developing and perfecting a health forecasting scheme over time. Thus, the framework demonstrates a dynamic process in which the forecast models created at any time would be continuously improved to meet the purpose of the forecast or the client's needs.

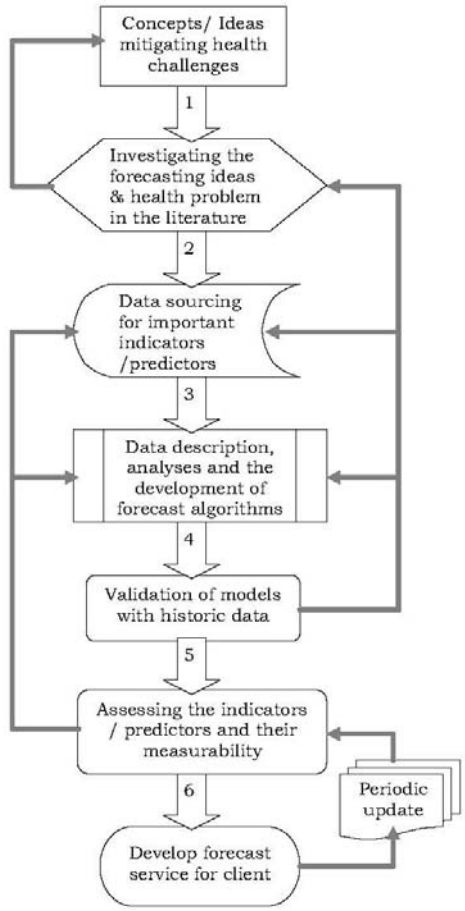


Fig. 1 A schematic approach to health forecasting

Step 1: Identify the concepts and ideas that address an important health condition of great burden and ones that significantly cost the health service. Provide a precise specification of the health outcome to be forecast and a clear definition of the forecasting horizon;

Step 2: Use the literature to identify causal or highly correlated variables that are associated with the identified health outcome measures in Step 1 (expert consult may be required in building this domain knowledge);

Step 3: Identify the data sources for both the health outcome measure (Step 1) and all of the potential predictors, and ascertain the availability and completeness (i.e. checking for gaps in the data series) of the data;

Step 4: Prepare the data sets for basic statistical analyses, including descriptive patterns and the development of forecast algorithms. Some preliminary activities include data cleaning and management, and the generation of supplementary variables for further analyses;

Step 5: Generate the predictive models and validate them using different sets of similar historical data;

Step 6: Evaluate and determine the final lists of indicators needed for good predictive model(s) based on the practical access to their measures (data);

Step 7: Develop very specific and tailor-made health forecast services for a specific purpose/client, and then periodically update the model(s).

Patterns of health data and applications in forecasting

In health forecasting, the pattern of distribution of previous health data over a period of time (i.e. in the form of time series) is important for determining the choice of an appropriate forecasting method. Time series plays an important role in many forecasting approaches, and has been extensively used in subject areas such as climate science, finance and econometrics. The patterns of health data in time series, which are of importance to health forecasting are trend, seasonality, cyclicality, and randomness [48, 49].

Time series and health forecasting

Time series is defined by Shumway and Stoffer [50] as "*a collection of random variables indexed according to the order they are obtained in time*". In the broader literature, time series is similarly defined as a collection of data points that are typically measured at successive and uniformly spaced time intervals. In relation to health forecasting, the importance of this second definition is the emphasis it places on the "uniformly spaced time intervals", which is important in the use of health data for health forecasting. Thus, time series provides statistical setting for describing seemingly random fluctuating health data and projecting the data series into the future [49, 50].

Trend is the long-term variation in a time series that is not influenced by irregular effects or seasonally related components in the data. For instance, in health data, an overall record of a progressively increasing incidence over a specified period would show an increasing trend, irrespective of any random or systematic fluctuations.

When the pattern of health data (e.g. containing the incidence of health events/situations) is influenced by some periodic (long-term/short-term) fluctuations that are associated with other characteristics, it is described as cyclical. *Cyclicality* therefore refers to the extent to which disease incident data points are influenced by overall disease patterns. *Seasonality* is also a cyclic phenomenon, but is related to annual events, and is described as the predictable and repetitive positions of data points around the trend line within a year. A major difference between cyclical and seasonal patterns is that the former varies in length and magnitude, as compared to the latter. Chatfield describes how seasonality and cyclicality can be estimated either in an additive or multiplicative form [49]. *Additive seasonality* is estimated as a function of the sums of the de-seasonalized mean (m), the seasonal factor (S) and an error term (ε) (i.e. additive seasonality $= m + S + \varepsilon$). *Multiplicative seasonality* is defined by two functions, either the product of m, S and ε (multiplicative seasonality $= m{\cdot}S{\cdot}\varepsilon$), or the product of m and S and sum of ε (i.e. multiplicative seasonality $= m{\cdot}S + \varepsilon$). In order to minimise the overall error, shorter cyclical effects that fall within the annual seasonal effect are best estimated with additive seasonality, whereas the effect of annual seasonality is best computed as "$m{\cdot}S{\cdot}\varepsilon$" [49].

Randomness is also a common feature of all time series data, and refers to unexpected distortions of existing or anticipated trends.

Lag refers to the lapse of time before an effect is manifested. Lags have proven useful in forecasting events globally, and are a feature of time series data that is widely exploited in many forecasting techniques, e.g. in auto regressive integrated moving averages (ARIMA) [27]. In developing health forecast models for a particular condition/situation, the key questions are: how many days back should one go back in history to identify appropriate predictors, and how many lags should be included.

The properties of time series mentioned above require specific treatment prior to any analysis, and they have been described more elaborately elsewhere [48–51]. However, the statistical forecasting models that involve time series analysis and are commonly used in health forecasting include moving average models, such as ARIMA, and smoothing techniques, e.g. the Holt-Winters methods. For instance, the Box–Jenkins ARIMA model, is commonly used in fitting forecasting models when dealing with a non-stationary time series, and this model has been used extensively in health forecasting [27, 33, 52–55]. *Stationarity* is a feature of trend in a time series, and refers to the level of variation in the statistical properties (such as the mean, variance, auto-correlation, etc.) over time. Smoothing models have also been used in health forecasting studies conducted by Medina et al. and Hyndman et al. [56–58]. In the study conducted by Champion et al. [33], the authors identified trend, seasonal variations and randomness/"noise" in the data distribution, but used a time series statistical package to automatically identify optimal models to forecast monthly emergency department presentations. After, the authors proceeded to compare forecasts, based on a simple seasonal exponential smoothing model to an ARIMA model. Similarly, the study conducted by Medina et al. also identified seasonal oscillations and trends in the time series data (of the diseases they analyzed). The harmonics in the data distributions were handled as level, and trend components by the multiplicative Holt-Winters forecasting method, which is also a smoothing technique in forecasting [56].

Probabilistic health forecasting methods for peak events

Health forecasting techniques generally rely on modelling expectancy of the mean, but this is not useful for looking at extreme events. Nonetheless, extreme events represent the greatest test of a health system, because they expose the weaknesses of the system whenever they occur. A reliable method of modelling and predicting extreme events is therefore important. Quantile regression models (QRMs) and fractional polynomial models (FPMs) are potential probabilistic techniques that could be adopted for predicting extreme health situations/conditions.

Quantile regressions are extensions of the linear-regression models, and do not assume normality of the dependent variable. They model the conditional quantiles as functions of predictors, specifying changes in any conditional quantile [59, 60]. Unlike linear-regression models, QRMs have the ability to characterize the relationship between the dependent variable and the independent variable(s), particularly in the extremes of the distribution. They have common applications in medical reference charts, and could be used in preliminary medical diagnosis to identify unusual subjects by providing robust regressions for estimating extreme values [61]. QRMs also have the potential of predicting and forecasting extreme chronic respiratory illnesses like asthma. For instance, a QRM could be used to estimate extreme variations in daily asthma hospital admissions resulting from the changing patterns of selected meteorological and air quality indicators that are known to exacerbate asthma in a given location/area [62].

Williams [63] also showed how fractional polynomials could be used in modelling specific categories of dependant variables within a linear distribution of data, and thus target specific groups more precisely. In this study, the author used various categories of age groups as regressors to model a dichotomous health care demand. Logistic regression outputs of two arbitrary age-categorized models were then compared to a fractional polynomial model. The polynomial method of categorizing had clear advantages because it allowed a fuller representation of non-linear relationships between the predictor and outcome variables. This approach can be extended to a wide range of health situations or conditions.

Both approaches (QRM and FPM) can be adapted to suit extreme health forecasting.

The value of health forecasting in health services provision

Health service(s) is (are) the most important component of any health system. The World Health Organisation (WHO), reports that effective health service delivery requires some key resources including information, finance, equipment, drugs and well motivated staff [64]. Given the ever-increasing demand for both the coverage and quality of health care services, health service delivery institutions and service providers struggle to tackle situations of excess demand particularly associated with peak events [65–67]. This is because front-line health delivery services and providers are not usually adequately informed and do not have adequate resources to meet the needs of a "higher than normal" demand for health care. Hence, improving the access, coverage and quality of health services depends on the ways these services are pre-informed, organised and managed. Even though there are still unanswered questions about how to improve the organisation and management of health service delivery in a manner that would help achieve a better and more equitable coverage and quality [64], there are equally untapped resources such as health forecasting, which can aid the process. Health forecasting services enable both individuals and service providers to anticipate situations, and hence take the necessary steps to manage peak or extreme events.

There are important features or health outcome measures that are considered to have a significant impact on the coping mechanisms of health service providers. These features include the total duration of the care/support being provided (also described in the literature as "length of stay" in the hospital or "spell duration"), and the periodic (daily) rates of attendance of patients to the general practitioner or emergency rooms. The length of stay provides some insights into the disease burden of a particular health condition. The length of stay, in combination with other related factors like demographic, diagnostic and temporal factors, can explain and forecast future events [68–70]. On the other hand, the daily rate of health events is a very useful indicator, which can be used in time series forecasting.

The challenges in developing and using health forecasts

The value of health forecasting has been mentioned in previous discussions, but there are a number of challenging issues to be noted and addressed in developing and using a health forecast. These include limitations in the scope and reliability of health data, the robustness of health forecasting tools and techniques, and the poor demand for health forecasting [11, 28]. In recent times, technological advances have enabled health indicators to be more easily and cheaply measured, and yet the record capture of important population health indicators is not very efficient and not easily accessible or validated [28]. In the practice of personalised medicine, for instance, there are slight prognostic effects attributable to a wide range of complex factors (including some unknown factors), and these factors usually intermingle (randomly) to generate clinical outcomes. Data limitation on these complex factors can pose a challenge in developing a reliable health forecast. Aside from the data and methodological limitations in developing reliable health forecast, it is difficult to convincingly demonstrate the performance of a health forecasting model in realistic settings [71].

Health forecasting-related researches have sometimes focused on methods or procedures for forecasting aggregate health conditions, or on situations like crowding at emergency departments and total admissions [72–74]. Even though these kinds of aggregate health forecasts are useful, health care providers would be better informed and prepared with condition-specific health forecasts. Therefore, health forecasts need to be more specific for particular health conditions. For example, the health forecast service provided by the United Kingdom Meteorological Office to some Primary Care Trusts (PCT) is very specific for conditions such as COPD [7, 8]. This kind of service is rare but useful.

Health forecasts are most valuable when they provide sufficient warning for timely, remedial action to be taken. Providers make critical decisions and resource allocations to meet the potential demand for health care services. Some of the complexities associated with these types of health care provider actions could range from providing basic social care for early symptoms, to using sophisticated staff and facilities and attending to extreme events [7, 24, 41]. Meanwhile, being able to meet the demand for a health forecast that provides ample time for preparatory activities often requires the use of a good forecasting technique and ample reliable data. It also comes with an additional compromise as to the precision and accuracy of the forecast [75]. Hence, finding a fine line between what is predictable vis-à-vis the demand for specific health forecast is a key challenge in health forecasting.

Another challenge in health forecasting relates to its practical use. A health forecast is usually developed to target the needs of susceptible individuals or institutions (health care providers). In any instance, there is a need for a technology with an intelligent early warning system that can communicate the forecast to the users. Automated telephone services, home visits/treatment, and direct health forecast (to individuals and service providers) are means through which some health forecast services have been delivered [76]. Although there have been some challenges and debates regarding the relevance of some of these existing health forecasting programmes, there are a couple of success stories which provide compelling evidence for their usage [7]. The case of the UK Meteorological Offices' COPD forecast, which was available to general practitioners in Bradford and Airendale, is an example. In 2009, Maheswaran et al. [77] evaluated this health forecasting alert service and failed to show that any change in admissions associated with the forecasting service was significant, and hence they challenged the effectiveness of the COPD forecast. Meanwhile, in cross-sectional study on the acceptability and utility of this same service in England, Scotland and Wales, Marno et al. [7] concluded that the service was both viable and useful. Further research to improve or develop new approaches or schemes in health forecasting is therefore important and will contribute to easing disease burden.

Conclusion

Health forecasting is a dynamic process and requires frequent updates. This can be done with novel techniques and data, taking into consideration the principles of health forecasting. The methodologies currently used involve time series analyses with smoothing or moving average models, and less probabilistic forecasting models like QRM, which offers a useful alternative for predicting and forecasting extreme health events. The horizons of health forecasting are important but not classified in the literature, and so the approaches used to forecasting various horizons have no common benchmarks to guide new health forecasts. The patterns of health data can be exploited in health forecasting, using time series analysis or other probabilistic techniques. Health forecasting is a valuable resource for enhancing and promoting health services provision; but it also has a number of drawbacks, which are related either to the data source, methodology or technology. This overview is presented to stimulate further discussions on standardizing health forecasting approaches and methods, so that it can be used as a tool to facilitate health care and health services delivery.

Conflict of interest The authors declare they have no competing interests.

References

1. Armstrong JS. Principles of forecasting: a handbook for researchers and practitioners. Norwell: Kluwer Academic Publishers; 2001.
2. Lewis JB, McGrath RJ, Seidel LF. Essentials of applied quantitative methods for health services managers. Sudbury: Jones and Bartlett Publishers, LLC.; 2011.
3. Antoniou SA, Antoniou GA, Granderath FA, Mavroforou A, Giannoukas AD, Antoniou AI. Reflections of the Hippocratic Oath in modern medicine. World J Surg. 2010;34(12):3075–9.
4. Chang A, Lad EM, Lad SP. Hippocrates' influence on the origins of neurosurgery. Neurosurg Focus. 2007;23(1):E9.
5. Chadwick J, Mann WN. Medicine. In: Lloyd GER, editor. Hippocratic writings. England: Penguin Group; 1978. p. 170–171.
6. Craft ED. An economic history of weather forecasting. 2007. http://eh.net/encyclopedia/article/craft.weather.forcasting.history. Accessed 02 Jan 2010.
7. Marno P, Chalder M, Laing-Morton T, Levy M, Sachon P, Halpin D. Can a health forecasting service offer COPD patients a novel way to manage their condition? J Health Serv Res Policy. 2010;15(3):150–5.
8. Met-Office. The Met Office health forecasting services, Exeter. 2009. http://www.metoffice.gov.uk/health/. Accessed 01 Jan 2009.
9. WHO. Using climate to predict infectious disease epidemics. In: Kuhn K, Campbell-Lendrum D, Haines A, Cox J, editors. World Health Organization. Geneva: World Health Organization; 2005.
10. Rogers DP, Shapiro MA, Brunet G, Cohen JC, Connor SJ, Diallo AA, et al. Health and climate—opportunities. Proc Environ Sci. 2010;1:37–54.

11. Makridakis S, Taleb N. Decision making and planning under low levels of predictability. Int J Forecast. 2009;25(4):716–33.
12. Vaughan Williams L, Stekler HO. Sports forecasting. Int J Forecast. 2010;26(3):445–7.
13. Vaughan Williams L. Information efficiency in financial and betting markets. Cambridge: Cambridge University Press; 2005. p. 1–123.
14. McMichael AJ, Campbell-Lendrum DH, Corvalán CF, Ebi KL, Githeko AK, Scheraga JD, et al., editors. Climate change and human health: risks and responses. Geneva: World Health Organization; 2003.
15. WMO. Guidelines on analysis of extremes in a changing climate in support of informed decisions for adaptation. In: Klein Tank AMG, Zwiers FW, Zhang X, editors. Climate data and monitoring. Geneva: World Meteorological Organization; 2009.
16. Manton KG, Soldo BJ, Vierck E. "Death and taxes": a contrary view. Popul Today. 1984;12(11):2, 8–9.
17. Rogge JR. Population trends and the status of population policy in Africa. J Geogr. 1982;81(5):164–74.
18. Fildes R. The forecasting journals and their contribution to forecasting research: citation analysis and expert opinion. Int J Forecast. 2006;22(3):415–32.
19. Armstrong JS, Green KC. Demand forecasting: evidence-based methods. In: Working Paper 24/05, Demandforecasting35-Monashdoc. Department of Econometrics and Business Statistics, Monash University, Australia; 2005.
20. Fildes R, Goodwin P, Lawrence M, Nikolopoulos K. Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. Int J Forecast. 2009;25(1):3–23.
21. Lawrence M, Goodwin P, O'Connor M, Önkal D. Judgmental forecasting: a review of progress over the last 25 years. Int J Forecast. 2006;22(3):493–518.
22. Wright G, Lawrence MJ, Collopy F. The role and validity of judgment in forecasting. Int J Forecast. 1996;12(1):1–8.
23. McGregor GR, Watkin HA, Cox M. Relationships between the seasonality of temperature and ischaemic heart disease mortality: implications for climate based health forecasting. Clim Res. 2004;25(3):253–63.
24. Marno P, Bryden C, Bird W, Watkin HA. How different measures of cold weather affect chronic obstructive pulmonary disease (COPD) hospital admissions in London. Eur Respir Rev. 2006;15(101):185–6.
25. Shi L, van Meijgaard J, Fielding J. Forecasting diabetes prevalence in California: a microsimulation. Prev Chronic Dis. 2011;8(4):A80. (Epub 2011 Jun 2015).
26. Wargon M, Guidet B, Hoang TD, Hejblum G. A systematic review of models for forecasting the number of emergency department visits. Emerg Med J. 2009;26(6):395–9.
27. Boyle J, Jessup M, Crilly J, Green D, Lind J, Wallis M, et al. Predicting emergency department admissions. Emerg Med J. 2011;. doi:10.1136/emj.2010.103531.
28. Ioannidis JPA. Limits to forecasting in personalized medicine: an overview. Int J Forecast. 2009;25(4):773–83.
29. van Meijgaard J, Fielding J, Shimkhada R, Eslami E, Cole B. Forecasting health outcomes related to air quality for placer county. Issue brief. Health Forecasting, UCLA School of Public Health; 2010.
30. Sekhri N, Chisholm R, Longhi A, Evans P, Rilling M, Wilson E et al. Principles for forecasting demand for global health products. In: Global health forecasting working group, background paper. The Center for Global Development; 2006.
31. Hoot NR, LeBlanc LJ, Jones I, Levin SR, Zhou C, Gadd CS, et al. Forecasting emergency department crowding: a discrete event simulation. Ann Emerg Med. 2008;52(2):116–25.
32. McCarthy ML, Zeger SL, Ding R, Aronsky D, Hoot NR, Kelen GD. The challenge of predicting demand for emergency department services. Acad Emerg Med. 2008;15(4):337–46.
33. Champion R, Kinsman LD, Lee GA, Masman KA, May EA, Mills TM, et al. Forecasting emergency department presentations. Aust Health Rev. 2007;31(1):83–90.
34. Gentry L, Calantone RJ, Cui SA. The forecasting classification grid: a typology for method selection. J Glob Bus Manag. 2006;2(1):48–60.
35. Poikolainen K. A comment diagnosis as a means of health forecasting. Soc Sci Med Part A Med Psychol Med Sociol. 1979;13:165–6.
36. Henning KJ. What is syndromic surveillance? MMWR Morb Mortal Wkly Rep. 2004;53(Suppl):5–11.
37. Unkel S, Farrington CP, Garthwaite PH, Robertson C, Andrews N. Statistical methods for the prospective detection of infectious disease outbreaks: a review. J R Stat Soc Ser A. 2011;175(1): 49–82.
38. Burkom HS, Murphy SP, Shmueli G. Automated time series forecasting for biosurveillance. Stat Med. 2007;26(22):4202–18.
39. Krishna S, Boren SA, Balas EA. Healthcare via cell phones: a systematic review. Telemed J E Health. 2009;15(3):231–40.
40. Jessup M, Wallis M, Boyle J, Crilly J, Lind J, Green D, et al. Implementing an emergency department patient admission predictive tool: insights from practice. J Health Organ Manag. 2010;24(3):306–18.
41. Hemming D, Colman A, James P, Kaye N, Marno P, McNeall D, et al. Framework for COPD forecasting in the UK using weather and climate change predictions. IOP Conf Ser Earth Environ Sci. 2009;6:142021.
42. Oh J, Kim B. Prediction model for demands of the health meteorological information using a decision tree method. Asian Nurs Res. 2010;4(3):151–62.
43. van Meijgaard J, Fielding JE, Kominski GF. Assessing and forecasting population health: integrating knowledge and beliefs in a comprehensive framework. Public Health Rep. 2009;124(6): 778–89.
44. Majer I. Modeling and forecasting health expectancy: theoretical framework and application. In: Netspar Discussion Papers: 01/ 2011-009. Network for Studies on Pensions, Aging and Retirement. 2011. http://arno.uvt.nl/show.cgi?fid=113977. Accessed 01 Nov 2011.
45. Sanders NR. Forecasting theory. In: Wiley encyclopedia of electrical and electronics engineering. New York: Wiley; 2001.
46. Zaninotto P, Wardle H, Stamatakis E, Mindell J, Head J. Forecasting obesity to 2010. In: DOH: Publications and Statistics; 2006.
47. Soyiri IN, Reidpath DD. Evolving forecasting classifications and applications in health forecasting. Int J Gen Med. 2012;5(1): 381–9.
48. Barnett AG, Dobson AJ. Analysing seasonal health data. Heidelberg: Springer; 2009.
49. Chatfield C. The analysis of time series: an introduction. 6th ed. London: CRC Press; 2004.
50. Shumway RH, Stoffer DS. Time series analysis and its applications with R examples. 2nd ed. New York: Springer; 2006.
51. Zhang GP, Qi M. Neural network forecasting for seasonal and trend time series. Eur J Oper Res. 2005;160(2):501–14.
52. Reis B, Mandl K. Time series modeling for syndromic surveillance. BMC Med Inform Decis Mak. 2003;3(1):2.
53. Sharma P, Chandra A, Kaushik SC. Forecasts using Box–Jenkins models for the ambient air quality data of Delhi City. Environ Monit Assess. 2008;157(1–4):105–12.
54. Wang XK, Lu WZ. Seasonal variation of air pollution index: Hong Kong case study. Chemosphere. 2006;63(8):1261–72.

55. Armstrong JS. Findings from evidence-based forecasting: methods for reducing forecast error. Int J Forecast. 2006;22(3):583–98.
56. Medina DC, Findley SE, Guindo B, Doumbia S. Forecasting non-stationary diarrhea, acute respiratory infection, and malaria time-series in Niono, Mali. PLoS One. 2007;2(11):e1181.
57. Mott JA, Mannino DM, Alverson CJ, Kiyu A, Hashim J, Lee T, et al. Cardiorespiratory hospitalizations associated with smoke exposure during the 1997, Southeast Asian forest fires. Int J Hyg Environ Health. 2005;208(1–2):75–85.
58. Hyndman RJ, Koehler AB, Snyder RD, Grose S. A state space framework for automatic forecasting using exponential smoothing methods. Int J Forecast. 2002;18(3):439–54.
59. Koenker R. Quantile regression. New York: Cambridge Press; 2005.
60. Hao L, Naiman DQ (eds) Quantile regression. USA: Sage Publications, Inc; 2007.
61. Yu K, Lu Z, Stander J. Quantile regression: applications and current research areas. J R Stat Soc Ser D. 2003;52(3):331–50.
62. Soyiri IN, Reidpath DD. Predicting extreme asthma events in London using quantile regression models. Emerg Health Threats J. 2011;4:s162. doi:10.3134/ehtj.10.162:39-40.
63. Williams JS. Assessing the suitability of fractional polynomial methods in health services research: a perspective on the categorization epidemic. J Health Serv Res Policy. 2011;16(3):147–52.
64. WHO. Health service delivery. In: WHO programmes and projects: health systems; 2010.
65. Asthma-UK. No asthma improvement for decade has cost NHS 100 m pounds—Asthma UK Scotland. Scotland: Asthma UK; 2010. http://www.medicalnewstoday.com/articles/203257.php. Accessed 07 Nov 2010.
66. Bradley VM. Placing emergency department crowding on the decision agenda. J Emerg Nurs. 2005;31(3):247–58.
67. Derlet RW. Overcrowding in emergency departments: increased demand and decreased capacity. Ann Emerg Med. 2002;39(4):430–2.
68. Morrison DS, McLoone P. Changing patterns of hospital admission for asthma, 1981–97. Thorax. 2001;56(9):687–90.
69. Lambe S, Washington DL, Fink A, Herbst K, Liu H, Fosse JS, et al. Trends in the use and capacity of California's emergency departments, 1990–1999. Ann Emerg Med. 2002;39(4):389–96.
70. Soyiri IN, Reidpath DD, Sarran C. Asthma length of stay in hospitals in London 2001–2006: demographic, diagnostic and temporal factors. PLoS One. 2011;6(11):e27184.
71. Buckeridge DL, Burkom H, Campbell M, Hogan WR, Moore AW. Algorithms for rapid outbreak detection: a research synthesis. J Biomed Inform. 2005;38(2):99–113.
72. Sun Y, Heng B, Seow Y, Seow E. Forecasting daily attendances at an emergency department to aid resource planning. BMC Emerg Med. 2009;9(1):1.
73. Schweigler LM, Desmond JS, McCarthy ML, Bukowski KJ, Ionides EL, Younger JG. Forecasting models of emergency department crowding. Acad Emerg Med. 2009;16(4):301–8.
74. Lisa MS, Jeffrey SD, Melissa LM, Kyle JB, Edward LI, John GY. Forecasting models of emergency department crowding. Acad Emerg Med. 2009;16(4):301–8.
75. Myers MF, Rogers DJ, Cox J, Flahault A, Hay SI. Forecasting disease risk for increased epidemic preparedness in public health. Adv Parasitol. 2000;47:309–30.
76. Cooper R, O'Hara R. Patients' and staffs' experiences of an automated telephone weather forecasting service. J Health Serv Res Policy. 2010;15(Suppl 2):41–6.
77. Maheswaran R, Pearson T, Hoysal N, Campbell MJ. Evaluation of the impact of a health forecast alert service on admissions for chronic obstructive pulmonary disease in Bradford and Airedale. J Public Health. 2009;32(1):97–102.

Chapter 6

6.0 Forecasting classifications and health forecasting

In the previous chapter we presented an overview of health forecasting, focussing on issues around definitions, principles of health forecasting and some of the factors that influence the choices of health forecasting methods. In this chapter, particular attention is paid to the evolution of forecasting approaches (typologies), and the forecasting methods which mark developments of these approaches. The chapter also includes a brief review of the methods and techniques used in measuring the accuracy and validation of forecasting models. Earlier classification typologies could afford to be concise, exclusive and exhaustive. However, as a result of innovations and the development there are overlaps and nuanced differences that no longer make such typologies possible or practicable. On the flipside, there is a greater flexibility in selecting/choosing forecasting methods, particularly when dealing with non-causal approaches. They are focussed on prediction.

The review also discusses the challenges associated with developing and using a health forecasting.

6.1 Evolving forecasting classifications and applications in health forecasting

6.1.1 Declarations for Thesis Chapter 6

Monash University

Declaration for Thesis Chapter 6

Declaration by candidate

In the publication: *Evolving forecasting classifications and applications in health forecasting.* in the *Int J Gen Med.* 2012. 5(1): p. 381-9 / Chapter 6, the nature and extent of my contribution to the work were the following:

Nature of contribution	Extent of contribution (%)
Lead and corresponding author, conceptualized the idea, accessed and organized the literature, drafted initial manuscript, submitted manuscript to journal and then managed correspondence with editors/reviewers, editorial staff and publishers until the final publication	100

The following co-authors contributed to the work. Co-authors who are students at Monash University must also indicate the extent of their contribution in percentage terms:

Name	Nature of contribution	Extent of contribution (%) for student co-authors only
Daniel D. REIDPATH	Supervised all activities leading to the publication of the paper; critiqued conceptual ideas, analyses and all drafts of manuscript	N/A

Candidate's Signature	Ireneous N. Soyiri	**Date:** 22-11-2012

Declaration by co-author

The undersigned hereby certify that:

(1) the above declaration correctly reflects the nature and extent of the candidate's contribution to this work, and the nature of the contribution of each of the co-authors.

(2) they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;

(3) they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
(4) there are no other authors of the publication according to these criteria;
(5) potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and
(6) the original data are stored at the following location(s) and will be held for at least five years from the date indicated below:

Location(s)	**Global Public Health, School of Medicine & Health Sciences, Monash University Sunway campus**

Signature 1	Daniel D. Reidpath	**Date:** 22-11-2012

- -

International Journal of General Medicine

Dovepress
open access to scientific and medical research

Open Access Full Text Article

REVIEW

Evolving forecasting classifications and applications in health forecasting

This article was published in the following Dove Press journal:
International Journal of General Medicine
3 May 2012

Ireneous N Soyiri[1,2]
Daniel D Reidpath[1]

[1]Global Public Health, JCSMHS, MONASH University, Selangor, Malaysia; [2]School of Public Health, University of Ghana, Legon, Accra, Ghana

Abstract: Health forecasting forewarns the health community about future health situations and disease episodes so that health systems can better allocate resources and manage demand. The tools used for developing and measuring the accuracy and validity of health forecasts commonly are not defined although they are usually adapted forms of statistical procedures. This review identifies previous typologies used in classifying the forecasting methods commonly used in forecasting health conditions or situations. It then discusses the strengths and weaknesses of these methods and presents the choices available for measuring the accuracy of health-forecasting models, including a note on the discrepancies in the modes of validation.

Keywords: health forecast, health data, electronic health records, accuracy, cross validation, method, strengths and limitations

Introduction

Forecasting is the process of predicting future events based on foreknowledge acquired through a systematic process or intuition.[1,2] It requires data, information, and advanced knowledge. Forecasting has evolved over the years and now has wide applications in many fields, including economics and commerce,[1,3] sports,[4,5] the environment (including meteorology),[6,7] technology and politics,[8–10] and health.[11–14]

Health forecasting predicts health situations or disease episodes and forewarns about future events. It is also a form of preventive medicine or preventive care engaged in public health planning, and it is aimed at facilitating health care service provision in populations.[12,14–16] One of the least developed branches of forecasting, health forecasting is a useful tool for decision making in health services provision. Health forecasting has been commonly applied to emergency department visits, daily hospital attendance, and admissions.[17–20]

Various methods and approaches have been applied in forecasting events, but some outstanding issues are yet to be addressed. Even though a comprehensive classification of all forecasting approaches and methods would serve as a useful guide to forecasters searching for appropriate forecasting methods, there have been limited discussions and debates around this need.[21] Health forecasting studies have often adapted statistical techniques used by other well-established areas of forecasting, such as econometrics and finance. However, little has been said about the strengths and weaknesses of these techniques when they are applied to health forecasting.[22] Another important issue that has not been explicitly presented in the literature relates to approaches used to determine the accuracy and validity of health-forecasting models. The applications

Correspondence: Ireneous N Soyiri
Global Public Health, JCSMHS, MONASH University, 46150 Bandar Sunway, Selangor DE Malaysia
Tel +603 5514 6300 Extn 61569
Fax +603 5514 6323
Email soyiriin@yahoo.com

http://dx.doi.org/10.2147/IJGM.S31079

available for measuring and determining the validity and accuracy of health forecasts have not been compared or presented as guides for health forecasting or even used to stimulate discussions that can contribute to improve health forecasting. This paper therefore aims at presenting a brief overview of the evolution of forecasting classifications and methods. It discusses the strengths and weaknesses of various health-forecasting techniques and methods, and then presents the choices available for validating and measuring the accuracy of health-forecasting models. Because of the new approach it brings to medical and health sciences, health forecasting is important for practices in these fields. Advances in health-forecasting research will facilitate the decision-making processes that are associated with healthcare planning and management.

In preparing this review, a search of the literature on health forecasting and statistical methods used in the analysis of health conditions was conducted in popular medical databases, such as PubMed (Medline) and Google Scholar. Additional literature searches were done through citation mapping of key papers. The selected papers and documents were synthesized and summarized according to the objectives of this paper.

An enumeration of forecasting typologies

Although several authors have made attempts to schematically classify the wide variety of forecasting techniques, many have not been adequately exclusive[a] or both concise and exhaustive[b] enough to meet the needs of forecasters across all fields.[21] A good classification system for forecasting methods can facilitate the process of choosing an appropriate method for forecasting, in addition to providing a better understanding and organization of the methods involved in designing a forecasting system. The enumeration of forecasting classifications presented below reveals the typologies and methods that have been involved. These classifications also justify the need for further research involving health-forecasting methodologies, since the latter have played a minimal role in shaping contemporary theory and methods in this area.

In 1971, Cetron and Ralph developed one of the earliest classifications of forecasting methods and approaches. It consisted of five categories, including intuitive methods, trend extrapolation, trend correlation, analogy, and dynamic predictive models.[23] Generally, intuitive methods in forecasting are based on individual opinion, whether structured or unstructured. Trend extrapolation is an approach that uses known existing trends, whereas trend correlation forecasts are based on the causal links between a dependent factor and another factor or factors. Cetron and Ralph also used the category, Analogy, to describe forecasting approaches that used similarity in patterns for forecasting. They also classified dynamic predictive models (also later known as structural models), which describe simulation procedures that involve high impact causal factors. Although Cetron and Ralph's classification is concise, it has been criticized for being neither exhaustive nor exclusive enough.[21]

Similarly, in 1972, in his classification of forecasting methods, Martino provided a five-category scheme consisting of the following: intuitive, consensus (ie, obtaining results from several experts), analogy, trend extrapolation, and structural models.[24] Although Martino's classification was concise and exclusive, it was not sufficiently exhaustive to meet the needs of forecasters. In 1978, another classification of forecasting methods by Bright considered as many as eight different categories, some of which were later thought contentious.[21,25] His classification included intuitive forecasting, trend extrapolation, dynamic modeling, morphological analysis, normative forecasting, monitoring, cross-impact analysis, and scenarios.[25] The key strength of Bright's classification was that it added an entirely new concept of scenarios, and could be viewed liberally as exhaustive. However, it was neither exclusive nor concise. Furthermore, as mentioned earlier, some categories, such as monitoring, have been challenged because they are inappropriate as forecasting methods or approaches.[21]

In 1985, Armstrong published his first "forecasting methodology tree," which was based on three assumptions or decisions. More recently, his *Methodology Tree for Forecasting* (2010) assumed that before arriving at an appropriate choice of analytical forecasting method, it is first necessary to decide on whether to use intuitive (judgmental) or objective (statistical) methods. Second, if the choice of approach is statistical, then a choice between causal and noncausal approaches is required. After a causal approach is chosen, the final decision is whether to select either linear or nonlinear (classification) statistical approaches.[26] Hence, Armstrong introduced five categories in his maiden classification: judgmental, bootstrapping, extrapolation, econometric, and segmentation. Armstrong's classification was concise and contributed new approaches to forecasting

[a]Exclusive means that anything belonging in one category should clearly not belong in another category.
[b]Exhaustive means that the classification system should cover every potential option.

(ie, naïve/causal continuum), as well as providing guidance on the selection of forecasting approaches, which previous classifications lacked. However, Armstrong's classification was neither exhaustive nor exclusive.

In 2001, Armstrong revised the classification of his forecasting methods and provided eleven categories of methods that could be derived from the *Methodology Tree* (including role playing, intentions, conjoint analysis, expert opinions, judgmental bootstrapping, analogies, extrapolation methods, rule-based forecasting, expert systems, econometric models and multivariate models).[1] His classification further illustrated the primary distinction between methods that rely on judgment[c] and those that estimate relationships using statistical[d] approaches or quantitative data.[1] The classification was however not concise because there were too many categories (eleven in all). It was also not exclusive because of the subclassification extrapolation, which has a much wider application in statistical forecasting.

In 2006, Gentry et al proposed an entirely new form of categorizing forecasting approaches and methods in the form of a grid. In *Forecasting Classification Grid*,[21] argue that two independent dimensions can determine forecasting approaches, which are on the continuums of Opinion and Empirical and Causal and Naïve. This classification helps to distinguish between opinions and ideas that can be empirically verified, and it is also simpler and more logical compared with earlier descriptions.[21] The *Grid* has just four classifications and is therefore a concise scheme. It also is exhaustive because it is designed to fit in as many forecasting methods as are available. Even though the authors believed that the classifications were exclusive, grey boundaries could exist between the continuums. The key challenge in this classification is that the grid appears to be an uncompleted template, so the exact relative positioning of a forecasting method could be the subject of debate.

In a 2010 revision of the *Methodology Tree for Forecasting*,[e] Armstrong and Green further extend the list of forecasting methods. The authors differentiated between structured and unstructured approaches related to judgmental forecasting and further classified the theory-based approaches of forecasting into the categories, linear and classification.[27] Armstrong and Green provided guidance on choosing suitable forecasting approaches and methods based on specific contexts. However, there are still flaws in their classification because some methods that had multiple applications were not explicitly illustrated in the *Methodology Tree*. For instance, both univariate and multivariate approaches can be applied in data mining or causal modeling.[28] Furthermore, noncausal (black box) approaches, such as those involving data mining and neural nets, are equally applicable to causal modeling and share similar methods and techniques (eg, regressions and segmentation).

All the classifications of forecasting approaches and methods discussed above have significantly contributed to the organization of forecasting. Even though most of these developments have taken place in nonhealth-related areas (eg, marketing, management and finance/econometrics), they have direct applications to health forecasting. For instance, health forecasting has used neural networks,[29–31] and many emergency department forecasts use one or more forms of regression analysis. It is therefore imperative that the lessons learned from previous forecasting topologies should inform any development of a typology of health-forecasting approaches and methods. Some related health-forecasting methods involved in the typologies listed above are exemplified in subsequent sections of this essay, which considers their strengths and limitations, their accuracy, and their validation procedures.

Health data and forecasting

Although data is vital in forecasting, what constitutes "health data" is poorly defined in the literature. Health data can be defined as: records of health conditions and situations that refer to individuals or populations and carry information about disease prevalence, incidence, diagnoses, treatments, prognosis, preventive strategies, and health systems. Moreover, these records are categorized by demographics and factors that directly affect health and are collected systematically or otherwise. For example, this definition could apply to hospital attendance or admission records that contain a variety of measures that are recorded in their respective units – age in complete years, for example.

In the practice of medicine, the diagnosis of disease is focused on determining the presence or absence of a condition so that the appropriate treatment can be given. However, the measures taken to facilitate this process are selected from a continuum. For example, diastolic blood pressure and pulse rate are taken to help determine whether a person is hypertensive or not (ie, more than or equal to 90 mmHg). Other factors that may have an effect on health status and whose levels

[c]Judgmental forecasting techniques include prediction markets, Delphi, structured analogies, game theory, decomposition, judgmental bootstrapping, and expert systems.
[d]Statistical forecasting approaches include causal models and segmentation.
[e]Also available at www.forecastingprinciples.com.

are measured to generate health data include environmental exposure (eg, weather and air quality). At the point of measure or use, health data could be classified as either continuous (ratio or interval scales) or categorical (ordinal, nominal, or dichotomous scales).[32] The definition and classification of health data determine how data are accumulated over time in addition to the method or methods of analyses that can be employed in analyzing this information.

An emerging form of health data – electronic health records (EHR) – refers to digital health data that is stored in secured repositories and shared only among authorized users.[33] Hayrinen and colleagues identified the following as components of electronic health records: daily charting, medication administration, physical assessment, admission nursing notes, nursing care plan, referral, present complaint (eg, symptoms), past medical history, life style, physical examination, diagnoses, tests, procedures, treatment, medication, discharge, history, diaries, problems, findings, and immunization.[34] These kinds of structured records have applications beyond health forecasting because they can be used to make predictions about the occurrence of future health events.

Strengths and limitations of health-forecasting techniques

Many reported studies on health forecasting adopted statistical techniques and methods, the theories of which are described in the standard literature. The choice of method depends on the purpose of forecasting and the nature of the data that are available. The strengths and limitations of these methods pertaining to health forecasting are discussed in the following paragraphs.

Linear regression methods are commonly used because they provide reasonably accurate results, are easy to interpret, and have wide applications in modeling trends and seasonality. However, like most regression methods, linear regression uses the method of ordinary least squares to derive estimates, which may wrongly assume that but for the dependent variable, the independent variables or regressors have no error.[35,36] Hence, to account for this problem in modeling, there is always a need to factor in an error component. Linear regressions also require large amounts of data on all variables for parameter estimations.[35]

Logistic regressions provide a means for analyzing binary or categorical dependent variables, but they are not useful for count data.[37] Logistic regressions can thus be applied to forecast the presence or absence of an event in a dichotomous (categorical) state. Poisson and negative binomial regressions are generally used for analyzing count data, and the latter is particularly suitable for analyzing count data that have a skewed distribution with a considerable number of zero entries.[38–42] For instance, Negative Binomial Models (NBMs) were used in previous work to investigate the determinants of asthma in the length of stay in hospitals, for which the dependent variable bore the aforementioned characteristics.[43,44] NBMs were also used to compare various statistical forecasting models for predicting the number of daily admissions of asthmatic patients in London (personal work yet to be published).

Moving average methods, which include autoregressive-integrated moving average (ARIMA), seasonal autoregressive-integrated moving average (SARIMA), and exponential smoothing (eg, Holt–Winters) are widely used forecasting approaches. They have the advantage of modeling trend and seasonal variations, as well as accommodating multivariable models.[45,46] The exponential smoothing methods used in health forecasting are effective with data that change over time.[22] However, the main challenge in using these complicated methods is that they require specialist knowledge and expertise.

Time series regressions generally have a much wider application and capabilities in forecasting than all the other nontime-series approaches mentioned here. Time series regressions provide easily interpretable outputs that can be more consistent than ordinary linear regressions.[45,46] The use of time series approaches in forecasting ideally requires sufficient data for not only the dependent variables, but also the matching independent variables.

Quantile regressions and fractional polynomials are rarely mentioned in the health-forecasting literature, but they provide a means for predicting and forecasting peculiar events.[47,48] For instance quantile regression models allow the modeling and forecasting of anticipated extreme events based on data distributions that are outside the normal range, which is more useful than linear regressions whose forecasts are based on the overall mean distribution. One major limitation of these approaches is that they deal with only the relevant or specific category of the data, and hence some information that could affect the accuracy and statistical power of the analysis is lost.[48,49]

Artificial neural network (ANN) is a black box modeling procedure known to provide more reliable results than the traditional causal approach.[50] ANN is capable of modeling complex and random systems by automatically controlling adjustments to the changes in time series based on the design of the experiment.[22,51] The major challenge of these models is that they are difficult to interpret and, unlike the other approaches described earlier, very few statistical software packages are available.[52–54]

Measuring forecast accuracy and validation

Forecasting is generally aimed at predicting future events in order to inform and guide precautionary measures. It is an art in as much as it is a science, and therefore the degree of certainty of every forecast is imperative. A number of techniques and approaches are used to determine the accuracy or validation of a forecast. The main purpose of measuring the accuracy of a forecast model is to assist in choosing the best model.[55] This can be done in several ways: traditional forecasting accuracy measures; model discrimination approaches like receiver-operating characteristic (ROC) curves; and the use of model fit statistics, eg, R-square, Akaike information criterion (AIC), Schwarz information criteria (SIC) and Bayesian information criteria, which are discussed below.

Forecast accuracy

Forecast accuracy is a quantitative measure of the efficiency of the forecasting process, and it is performed by comparing the forecast to the actual situation. Forecast accuracy measures and parameters are usually supplied alongside the forecast as constituent elements to aid in decision making.

The accuracy of any forecast depends on objective features of the environment, such as the nature of the variable being forecast, or the length of the forecast horizon. Accuracy also depends on attributes of the forecast relating to the theories involved.[56] Hence, the choice of accuracy measures may depend on the method of forecasting. However, there has been considerable discussion about appropriate measures of forecasting accuracy, which have a wide range.[55,57–61]

Measures of forecasting accuracy have three main categories: (a) scale-dependent measures (ie, accuracy measures whose scale depends on the scale of the data); (b) percentage error measures (ie, independent measures that can compare forecast performance across different datasets); and (c) relative error measures (ie, scaled errors based on error measured from a reference standard forecast), including the relative measures of each type of error measure.[59] Examples of these measures are listed in Table 1.

Subsequent discussion focuses on selected scale-dependent and percentage error measures (Table 2), which are commonly used in health-forecasting studies. Scale-dependent error measures have been recommended for the comparison of different methods that are applied to the same set of data and scales.[59] For example, the root-mean-square error (RMSE) has traditionally been widely used for forecasting evaluation[62] and specifically for comparing models of the same series.[61,63] Even though some scale-dependent error measures, such as the MSE and RMSE, have been theoretically more relevant in statistical modeling, they have also been found more sensitive in detecting outliers than the mean absolute error (MAE) or median absolute error (MdAE).[59] Mean absolute scaled error (MASE) is also another scaled error approach recommended for comparing forecast accuracy across series on different scales.[59,64] According to Hyndman and Koehler, MASE provides the

Table 1 List of forecast accuracy measures

B. Scale-dependent measures
   I. Mean square error (MSE)
   II. Root mean squared error (RMSE)
   III. Mean absolute error (MAE)
   IV. Median absolute error (MdAE)

C. Percentage error measures
   I. Mean absolute percentage error (MAPE)
   II. Median absolute percentage error (MdAPE)
   III. Root mean square percentage error (RMSPE)
   IV. Root median square percentage error (RMdSPE)

D. Relative error measures[a]
   I. Mean relative absolute error (MRAE)
   II. Median relative absolute error (MdRAE)
   III. Geometric mean relative absolute error (GMRAE)

Notes: [a]The relative error measures are obtained by dividing each forecast error by the error obtained using a benchmark procedure, such as the grand mean (ie, a reference or benchmark average, which could be determined by taking the average of all averages of several subsamples). The accuracy measures of GMRAE and MDRAE, for instance, were presented by Armstrong and Collopy (1992)[61] and Fildes (1992).[06] Even though both reports recommend the use of forecast accuracy measures based on relative errors, they express these measures in different and complicated forms. Hyndman and Koehler (2006)[59] have however noted that these relative error methods could have some deficiencies that are associated with the difficulty of dealing with extremely small benchmark forecast error measures, resulting in the relative error measures having infinite variances.

Table 2 A comparison of scale-dependent error measures

Scale-dependent measures	Definition	Error spread	Error weights
Mean square error (MSE)	$\mathrm{Mean}(O_t{-}F_t)^2$	Yes	Yes
Root mean squared error (RMSE)	$\sqrt{\mathrm{MSE}}$	Yes	Yes
Mean absolute error (MAE)	$\mathrm{Mean}\lvert(O_t{-}F_t)\rvert$	Yes	No
Median absolute error (MdAE)	$\mathrm{Median}\lvert(O_t{-}F_t)\rvert$	–	–
Mean absolute scaled error (MASE)	$\mathrm{Mean}\lvert(Q_t)\rvert$	Yes	Yes

Notes: Error spread refers to the ability of the measure to capture an error that is not localized and not widely distributed in the dataset. Error weights refers to the ability of the measure to differentiate the error at different points in history.
Abbreviations: t, at a time; O, observation; F, forecast; Q:A, scaled error independent of scale of data.[59]

most reliable approach because compared to others, it has a meaningful scale, is widely applicable, and is not subject to "degeneracy" problems.[59] Moreover, MASE is seen less sensitive to outliers and is more easily interpreted. It shows smaller variations, even with small samples, than other measures in the same category.[59,65]

Measures based on percentage errors are not dependent on the scales of data and hence can be used to compare forecast errors across different datasets. However, their results tend to be infinite or undefined if a given forecast result equates 0 at any given time or has an extremely skewed distribution when the forecast is close to 0.[59] A further challenge in this category of error measures, particularly for mean absolute percentage error (MAPE), is that they tend to over penalize positive errors compared to negative ones and thereby create an unbalanced symmetry in the measures.[66]

Model discrimination test (ROC curve)

The ROC curve is another measure of forecast error that is associated with discrimination and has been used in health related forecasting studies. ROC provides a means of measuring and comparing the accuracy of predictive models. It is a graphical plot of Sensitivity versus 1-Specificity in a binary classifier system, and it is constructed to assess the varying thresholds for discrimination of comparable statistical predictive models.[67–70] The accuracy of prediction is measured by comparing the true positives against false positives.[67,68] The ROC curve has very wide applications in many fields, and its use in forecasting has been described by many authors. It was for example used by Classen and Hokayem to compare various econometric models and to select a suitable model for forecasting "Childhood influences on youth obesity."[71]

Model fit statistics

Widely used statistical model fitness tests include R-square, adjusted R-square, AIC and SIC. These model parameters are defined and estimated as follows:

$$R^2 = 1 - \frac{RSS}{TSS} \tag{1}$$

$$\text{Adjusted}\, R^2 = 1 - \frac{RSS/_{n=k}}{TSS/_{n-1}} = 1 - (1 - R^2)\frac{n-1}{n-k} \tag{2}$$

$$AIC = e^{2k/_n}\frac{RSS}{n} \tag{3}$$

$$SIC = n^{k/_n}\frac{RSS}{n} \tag{4}$$

where RSS is the residual sum of squares; TSS is the total sum of squares; n is the sample size; k is the number of parameters in the fitted model.

The use of R^2 as a measure of fit or model variability in health-forecasting-related studies is very common in the literature.[20,72–75] A higher value of a model's R^2 could be interpreted as having a better fit, which tends to increase with the addition of every extra explanatory variable. However, using R^2 as a measure of fit can be unreliable in forecasting because the R^2 of a model can be high or equal to 1 and yet be consistently wrong.[57] Like the R^2, the adjusted R^2 also increases with every additional explanatory variable, but this test is more reliable because it tends to penalize the model for every additional explanatory variable as long as the new addition does not significantly reduce the RSS (Equation 2). The AIC is superior to the adjusted R^2 because it has a harsher penalty and is preferred in forecasting models as a measure of fit.[62,76,77] This technique is based on the maximum likelihood and the number of independently adjusted parameters within a predictive model.[76] Compared with AIC and R^2, SIC gives the best model diagnostic fit because it imposes the highest penalty on the model.[45] However, in forecasting, given the balance between the need for a predictive model that has a good fit and a high explanatory power, AIC is currently the most popular and recommended technique for model fit statistics,[78] and it is commonly used in model selection.[43,63,79–81]

Forecast validation

Forecast predictions are rarely perfect, so validation or cross-validation is an essential process that allows estimation of the extent to which a predictive model emulates the natural phenomenon that produces the data.[50,60,82] Validating a forecast requires appropriate techniques and reliable measures. In developing a health-forecasting predictive model, two types of validity can be examined: model validity and predictive validity. Both are important and can be used to generate a useful and reliable forecast. Model validity represents the extent to which the model fits the data that was used for the model development (ie, the fit of the model to the experimental sample). This type of validity test is also referred to in the literature as internal validity. The second type, predictive validity (also known as external validity), represents the extent to which the predicted forecast values fit the observed values (ie, the fit of the model to the test sample).[61] Predictive validity is usually carried out through a process described as cross validation.

Cross validation is a statistical technique commonly used in forecasting for estimating the performance of a predictive model. It is usually carried out using a similar, but separate, sample of the data that was used in developing the forecast model. The health-forecasting literature does not provide standard procedures for conducting cross validation. Hence, the proportion of an evaluating sample (compared to the test or model development sample of data) that is suitable and sufficient for validating a health-forecasting model remains unclear. A scan of the literature revealed a wide range of arbitrary choices. As illustrated in Table 3, the relative proportion of a data sample used for cross validation of health-forecasting models could range from 1:1 to 12:1.

Table 3 Varying ratios of period of training to period of evaluation of health forecasting models

Author	Ratio of period of training: evaluation	Analytical techniques used in forecasting and study purpose
Hoot et al 2008[17]	1:1	ARIMA; to predict ED operation conditions within 8 hours
McCarthy 2008[18]	1:1	Poisson regression; to predict hourly ED presentations
Boyle 2011[19]	1:1/2:1/3:1/4:1	ARIMA, regression, ESM; to predict ED presentation and admission
Hoot et al 2007[51]	2:1	Logistic regression and ANN; to predict ED overcrowding
Wargon et al 2010[83]	3:1	Regression model; to predict daily ED presentation
Reis and Mandl, 2003[84]	4:1/5:1	ARIMA models; to predict daily pediatric ED presentation
Schweigler et al 2009[85]	7:1/14:1	SARIMA, hourly historical averages; to predict hourly ED bed occupancy
Jones et al 2008[22]	8:1	SARIMA, regression, ESM, and ANN; to predict daily presentation
Batal et al 2001[75]	9:1	Stepwise linear regression; to predict daily presentation
Champion et al 2007[20]	12:1	ARIMA, ESM; to predict aggregate monthly ED presentations

Abbreviations: ANN, Artificial Neural Networks; ARIMA, autoregressive-integrated moving average; SARIMA, seasonal autoregressive-integrated moving average; ESM, exponential smoothing; ED, emergency department.

For example, studies conducted by McCarthy et al[18] and Hoot et al[17] on forecasting emergency department visits used similar proportions of data for a cross validation of their output. However, other researchers have done this differently. To develop and to test their forecasts, Wargon et al[83] and Rotstein et al[74] used three quarters of their data as a training sample and the other one quarter as an evaluating sample.

Currently, there are no common scales for validating health forecasts based on a particular forecasting horizon, and the information available suggests that any appropriate cross validation strategy should be considered case by case. Thus, further research is necessary to help define and streamline the process of validating health-forecasting models.

Conclusion

The review identifies a number of knowledge gaps in health forecasting, which presents a challenge for further studies. These gaps include the following:

1. Typologies that classify health-forecasting approaches and methods;
2. A clear definition for health data, which nonetheless is an important ingredient for health forecasting;
3. Discussions on the strengths and limitations of statistical methods that are applicable to health forecasting, particularly for extreme health events;
4. A classification and ranking of various accuracy measures applicable to health forecasting; and
5. A clearly defined approach to cross validation of health-forecasting models.

The classifications of forecasting approaches have evolved over time. Several researchers have attempted to classify forecasting methods into typologies that are concise, exclusive, and exhaustive for all purposes. Lessons learned from these attempts will serve as useful guides in developing health-forecasting classification topologies and schemes, which are currently nonexistent. Few statistical methods have been identified to forecast extreme health events. Compared with percentage and relative error measures, scale-dependent error measures are easier and more frequently used in health forecasting. Because no common guidelines are available for cross validation in health forecasting, the current practice is quite irregular. Therefore, detailed studies are needed to help define standard classifications and applications for health forecasting.

Disclosure

The authors declare they have no competing or conflicts of interest in this work.

Authors' contributions

The paper was conceived and drafted by INS. Both INS and DDR participated in revising the draft. The final manuscript was approved by both authors.

Acknowledgments

We would like to thank the peer reviewers of this article for the useful comments and Ginarthini D/O Pachlai for editorial expertise.

References

1. Armstrong JS. *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Norwell, MA: Kluwer Academic Publishers; 2001.
2. Lewis JB, McGrath RJ, Seidel LF. *Essentials of Applied Quantitative Methods for Health Services Managers*. Sudbury, MA: Jones and Bartlett Publishers, LLC; 2011.
3. Makridakis S, Taleb N. Decision making and planning under low levels of predictability. *Int J Forecast*. 2009;25(4):716–733.
4. Vaughan Williams L, Stekler HO. Sports forecasting. *Int J Forecast*. 2010;26(3):445–447.
5. Vaughan Williams L. *Information Efficiency in Financial and Betting Markets*. Cambridge, MA: Cambridge University Press; 2005.
6. McMichael AJ, Campbell-Lendrum DH, Corvalán CF, et al, editors. *Climate Change and Human Health: Risks and Responses*. Geneva, Switzerland: World Health Organization; 2003.
7. World Meteorological Organization. *Guidelines on Analysis of Extremes in a Changing Climate in Support of Informed Decisions for Adaptation*. WCDMP-No 72. Geneva, Switzerland: World Meteorological Organization; 2009.
8. Manton KG, Soldo BJ, Vierck E. Death and taxes: a contrary view. *Popul Today*. 1984;12(11):2, 8–9.
9. Rogge JR. Population trends and the status of population policy in Africa. *J Geog*. 1982;81(5):164–174.
10. Fildes R. The forecasting journals and their contribution to forecasting research: citation analysis and expert opinion. *Int J Forecast*. 2006;22(3):415–432.
11. Marno P, Chalder M, Laing-Morton T, Levy M, Sachon P, Halpin D. Can a health forecasting service offer COPD patients a novel way to manage their condition? *J Health Serv Res Policy*. 2010;15(3):150–155.
12. Met-Office. The Met Office health forecasting services. 2009. Available from: http://www.metoffice.gov.uk/health/. Accessed January 1, 2009.
13. World Health Organization. *Using Climate to Predict Infectious Disease Epidemics*. Geneva, Switzerland: World Health Organization; 2005.
14. Rogers DP, Shapiro MA, Brunet G, et al. Health and climate – opportunities. *Procedia Environmental Sciences*. 2010;1:37–54.
15. Van Meijgaard J, Fielding J, Shimkhada R, Eslami E, Cole B. *Forecasting Health Outcomes Related to Air Quality for Placer County. Issue Brief*. Los Angeles, CA: Health Forecasting, UCLA School of Public Health; 2010.
16. Sekhri N, Chisholm R, Longhi A, et al. Principles for forecasting demand for global health products. Global Health Forecasting Working Group Background Paper. 2006. Available from: http://www.cgdev.org/doc/DemandForecasting/Principles.pdf. Accessed February 2, 2009.
17. Hoot NR, LeBlanc LJ, Jones I, et al. Forecasting emergency department crowding: a discrete event simulation. *Ann Emerg Med*. 2008;52(2):116–125.
18. McCarthy ML, Zeger SL, Ding R, Aronsky D, Hoot NR, Kelen GD. The challenge of predicting demand for emergency department services. *Acad Emerg Med*. 2008;15(4):337–346.
19. Boyle J, Jessup M, Crilly J, et al. Predicting emergency department admissions. *Emerg Med J*. Jun 24, 2011; [Epub ahead of print.]
20. Champion R, Kinsman LD, Lee GA, et al. Forecasting emergency department presentations. *Aust Health Rev*. 2007;31(1):83–90.
21. Gentry L, Calantone RJ, Cui SA. The forecasting classification grid: a typology for method selection. *J Global Bus Manage*. 2006; 2(1):48–60.
22. Jones SS, Thomas A, Evans RS, Welch SJ, Haug PJ, Snow GL. Forecasting daily patient volumes in the emergency department. *Acad Emerg Med*. 2008;15(2):159–170.
23. Cetron MJ, Ralph CA. *Industrial Applications of Technological Forecasting: Its utilization in R&D management*. New York, NY: Wiley-Interscience; 1971.
24. Martino JP. *Forecasting the Progress of Technology*. 2nd ed. New York, NY: Gordon and Breach Science Publishers; 1972.
25. Bright JR. *Practical Technology Forecasting: Concepts and Exercises*. Austin, TX: Industrial Management Center, Inc; 1978.
26. Martino JP. *An Introduction to Technological Forecasting*. London, UK: Gordon and Breach; 1972.
27. Armstrong JS, Green KC. JSA-KCG methodology tree for forecasting – forecasting principles. *Forecasting Principles*. 2010. Available from: http://www.forecastingprinciples.com/. Accessed March 19, 2011.
28. Fildes R, Nikolopoulos K, Crone S, Syntetos A. Forecasting and operational research: a review. *J Oper Res Soc*. 2008;59(9):1150–1172.
29. Blaisdell CJ, Weiss SR, Kimes DS, et al. Using seasonal variations in asthma hospitalizations in children to predict hospitalization frequency. *J Asthma*. 2002;39(7):567–575.
30. Kassomenos P, Papaloukas C, Petrakis M, Karakitsios S. Assessment and prediction of short term hospital admissions: the case of Athens, Greece. *Atmos Environ*. 2008;42(30):7078–7086.
31. Moustris KP, Douros K, Nastos PT, et al. Seven-days-ahead forecasting of childhood asthma admissions using artificial neural networks in Athens, Greece. *Int J Environ Health Res*. 2012;22(2):93–104.
32. Bland JM. *An Introduction to Medical Statistics*. 3rd ed. Oxford, UK: Oxford University Press; 2006.
33. ISO/TR 20514. Health Informatics – Electronic Health Record – Definition, Scope, and Context: International Organization for Standards (ISO). 2005. Available from: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=39525. Accessed March 10, 2012.
34. Hayrinen K, Saranto K, Nykanen P. Definition, structure, content, use and impacts of electronic health records: A review of the research literature. *Int J Med Inform*. 2008;77(5):291–304.
35. Cook RD, Weisberg S. Criticism and influence analysis in regression. *Sociol Methodol*. 1982;13:313–361.
36. Armstrong J. *Illusions in Regression Analysis*. Available from: http://ssrn.com/abstract=1969740. Accessed on December 8, 2011.
37. Collett D. *Modelling Binary Data*. 2nd ed. London, UK: Chapman and Hall/CRC Press; 2003.
38. Zorn C. Evaluating zero-inflated and hurdle Poisson specifications. *The Society for Political Methodology ad indicia spectate* 1998; Working Papers. Available from: http://polmeth.wustl.edu/workingpapers.php?text=use+of+force&searchkeywords=T&order=dateposted. Accessed November 1, 2008.
39. Long JS. Predicted probabilities for count models. *Stata J*. 2001;1(1):51–57.
40. Anon. A brief overview of count data commands in Stata. 2008. Available from: http://www.sts.uzh.ch/past/hs08/count/stata_glossar.pdf. Accessed November 1, 2008.
41. Alt JE, King G, Signorino C. Aggregation among binary, count, and duration models: estimating the same quantities from different levels of data. *Political Analysis*. 1999;9:21–44.
42. Mitchell SM, Moore WH. 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; Working Papers. Available from: http://polmeth.wustl.edu/workingpapers.php?text=use+of+force&searchkeywords=T&order=dateposted. Accessed November 1, 2008.

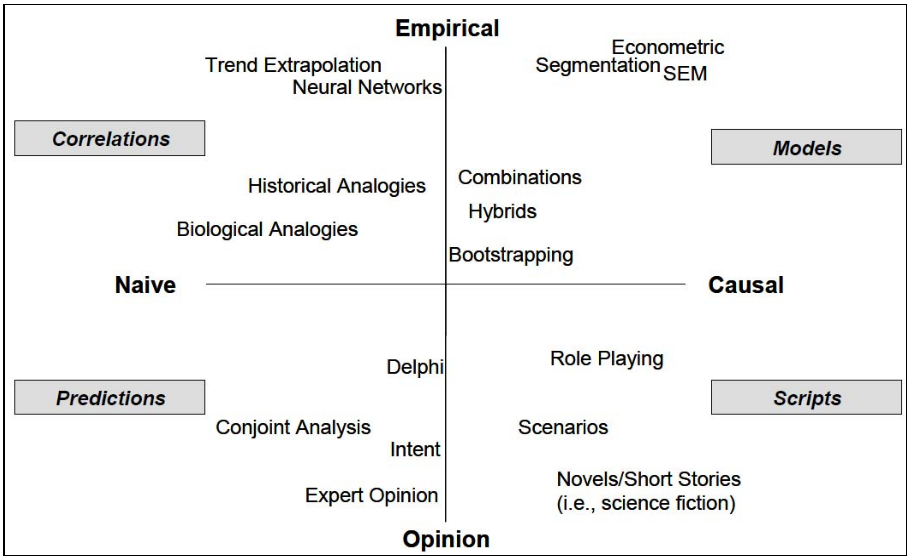
43. Soyiri I, Reidpath D, Sarran C. Determinants of asthma length of stay in London hospitals: individual versus area effects. *Emerg Health Threats J*. 2011;4(0):143–143.
44. Soyiri IN, Reidpath DD, Sarran C. Asthma length of stay in hospitals in London 2001–2006: demographic, diagnostic and temporal factors. *PLoS One*. 2011;6(11):e27184.
45. Shumway RH, Stoffer DS. *Time Series Analysis and its Applications with R Examples*. 2nd ed. New York, NY: Springer; 2006.
46. Chatfield C. *The Analysis of Time Series: An Introduction*. 6th ed. London, UK: CRC Press; 2004.
47. Koenker R. *Quantile Regression*. New York, NY: Cambridge Press; 2005.
48. Williams JS. Assessing the suitability of fractional polynomial methods in health services research: a perspective on the categorization epidemic. *J Health Serv Res Policy*. 2011;16(3):147–152.
49. Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *J R Stat Soc Ser C Appl Stat*. 1994;43(3):429–467.
50. Breiman L. Statistical modeling: the two cultures. *Stat Sci*. 2001;16(3):199–215.
51. Hoot NR, Zhou C, Jones I, Aronsky D. Measuring and forecasting emergency department crowding in real time. *Ann Emerg Med*. 2007;49(6):747–755.
52. Spencer SJ, Alun T, Evans RS, Shari JW, Peter JH, Gregory LS. Forecasting daily patient volumes in the emergency department. *Acad Emerg Med*. 2008;15(2):159–170.
53. Lu WZ, Wang WJ. Potential assessment of the "support vector machine" method in forecasting ambient air pollutant trends. *Chemosphere*. 2005;59(5):693–701.
54. McKendry IG. Evaluation of artificial neural networks for fine particulate pollution (PM10 and PM2.5) forecasting. *J Air Waste Manag Assoc*. 2002;52(9):1096–1101.
55. Mahmoud E. Accuracy in forecasting: a survey. *J Forecast*. 1984; 3(2):139–159.
56. Bathcelor R, Dua P. Forecaster ideology, forecasting technique, and the accuracy of economic forecasts. *Int J Forecast*. 1990;6(1):3–10.
57. Reidpath DD, Diamond MR, Hartel G, Glasziou P. Improving interpretability: as an alternative to R2 as a measure of effect size. *Stat Med*. 2000;19(10):1295–1302.
58. Lim C, McAleer M. Time series forecasts of international travel demand for Australia. *Tourism Management*. 2002;23(4):389–396.
59. Hyndman RJ, Koehler AB. Another look at measures of forecast accuracy. *Int J Forecast*. 2006;22(4):679–688.
60. Willmott CJ. Some comments on the evaluation of model performance. *Bulletin of the American Meteorological Society*. 1982;63(11):1309–1313.
61. Armstrong JS, Collopy F. Error measures for generalizing about forecasting methods: empirical comparisons. *Int J Forecast*. 1992;8(1):69–80.
62. Barnett AG, Dobson AJ. *Analyzing Seasonal Health Data*. London, UK: Springer; 2009.
63. Hu W, Tong S, Mengersen K, Connell D. Weather variability and the incidence of cryptosporidiosis: comparison of time series poisson regression and SARIMA models. *Ann Epidemiol*. 2007;17(9):679–688.
64. Smyth C, Mori J, Jiang C. Model ensembles for prediction of Wenchuan aftershock activity. *Bulletin of the Seismological Society of America*. 2010;100(5B):2532.
65. Koehler AB, Snyder RD, Keith Ord J, Beaumont A. A study of outliers in the exponential smoothing approach to forecasting. *Int J Forecast*. 2012;28(2):477–484.
66. Makridakis S. Accuracy measures: theoretical and practical concerns. *Int J Forecast*. 1993;9(4):527–529.
67. Pepe M, Longton G, Janes H. Estimation and comparison of receiver operating characteristic curves. *Stata J*. 2009;9(1):1.
68. Fawcett T. An introduction to ROC analysis. *Pattern Recognit Lett*. 2006;27(8):861–874.
69. Obuchowski NA. An ROC-type measure of diagnostic accuracy when the gold standard is continuous-scale. *Stat Med*. 2006;25(3):481–493.
70. Gorr WL. Forecast accuracy measures for exception reporting using receiver operating characteristic curves. *Int J Forecast*. 2009;25(1):48–61.
71. Classen T, Hokayem C. Childhood influences on youth obesity. *Econ Hum Biol*. 2005;3(2):165–187.
72. Tandberg D, Qualls C. Time series forecasts of emergency department patient volume, length of stay, and acuity. *Ann Emerg Med*. 1994;23(2):299–306.
73. Holleman DR, Bowling RL, Gathy A. Predicting daily visits to a walk-in clinic and emergency department using calendar and weather data. *J Gen Intern Med*. 1996;11(4):237–239.
74. Rotstein Z, Wilf-Miron R, Lavi B, Shahar A, Gabbay U, Noy S. The dynamics of patient visits to a public hospital ED: a statistical model. *Am J Emerg Med*. 1997;15(6):596–598.
75. Batal H, Tench J, McMillan S, Adams J, Mehler PS. Predicting patient visits to an urgent care clinic using calendar variables. *Acad Emerg Med*. 2001;8(1):48–53.
76. Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Contr*. 1974;19(6):716–723.
77. Bisgaard S, Kulahci M. Quality Quandaries*: Time series model selection and parsimony. *Qual Eng*. 2009;21(3):341–353.
78. Hyndman RJ, Koehler AB, Snyder RD, Grose S. A state space framework for automatic forecasting using exponential smoothing methods. *Int J Forecast*. 2002;18(3):439–454.
79. Beyerlein A, Fahrmeir L, Mansmann U, Toschke AM. Alternative regression models to assess increase in childhood BMI. *BMC Med Res Methodol*. 2008;8:59.
80. Lindsey JK, Jones B. Choosing among generalized linear models applied to medical data. *Stat Med*. 1998;17(1):59–68.
81. Tangri N, Stevens LA, Griffith J, et al. A predictive model for progression of chronic kidney disease to kidney failure. *JAMA*. 2011;305(15):1553–1559.
82. Fildes R, Kourentzes N. Validation and forecasting accuracy in models of climate change. *Int J Forecast*. 2011;27(4):968–995.
83. Wargon M, Casalino E, Guidet B. From model to forecasting: a multicenter study in emergency departments. *Acad Emerg Med*. 2010;17(9):970–978.
84. Reis B, Mandl K. Time series modeling for syndromic surveillance. *BMC Med Inform Decis Mak*. 2003;3(1):2.
85. Schweigler LM, Desmond JS, McCarthy ML, Bukowski KJ, Ionides EL, Younger JG. Forecasting models of emergency department crowding. *Acad Emerg Med*. 2009;16(4):301–308.
86. Fildes R. The evaluation of extrapolative forecasting methods. *Int J Forecast*. 1992;8(1):81–98.

6.1.2 Further comments on the classification and applications in health forecasting

The idea behind this chapter is to present the literature around forecasting methods, highlight the techniques applicable to health forecasting and illustrate some of the methods and approaches commonly used in evaluating forecasts. It was noted that there are a wide range of statistical techniques that could be adapted to suit the purpose of a health forecast, following either a causal or a non-causal (black box) approach in predictions. The forecasting classification /typology proposed by Gentry and colleagues for instance stand out from others, but also appears to be one of the most comprehensive. This is because it can accommodate all types of forecasting methods. For the lack of space/journal paper restrictions our enumeration of forecasting typologies did not include the forecasting grid by Gentry et al. (Fig. 6) or other typologies. However, the importance of this classification grid is because of its pragmatic description of forecasting methods which fits between the continuum of naïve (non-causal) and causal models. The authors describe "Naïve" (non-causal) methods as those which only use data on the variable of interest for predictions. Non-causal forecasting models have sparingly been applied in health forecasting. The thrust of such approaches form the basis of the health forecasting models described in subsequent sections, using asthma daily admissions in London.

In the following section, we begin to apply forecasting techniques to actual data.

Figure 6 Existing forecasting techniques and the Grid



Source: Gentry *et al*., 2006 [*The Forecasting Classification Grid: A Typology for Method Selection*]

SECTION IV

IV Developing health forecasting methods

In previous sections, we described the literature and some empirical analysis to highlight the importance of health forecasting to health services. We also exemplified approaches for establishing the disease burden and statistical methods which could be used in further empirical analyses. This section (IV) presents four empirical studies, which successively contribute to advancing the science in health forecasting methods. The studies include: Forecasting asthma related hospital admissions with negative binomial models; Predicting asthma daily admissions with lag models; and Forecasting peak asthma admissions in London, or peak respiratory related deaths in New York City using quantile regressions.

The first study was focused on testing the idea that asthma events could be better predicted using temporal, weather, and air quality factors. This study confirmed that selected lags of weather and air quality indicators can predict asthma daily admissions; however their combined effect was not better than predictive models which used temporal factors such as season alone.

To investigate the question about suitable indicators for forecasting asthma admissions besides the environmental predictors, a number of potential approaches exist. Among these possible approaches, we chose to investigate whether previous days' admissions could predict future admissions, using autoregressive analysis of daily asthma admissions. The idea behind this was that in a population, sensitive lungs act as sentinels for less sensitive lungs, so as environmental triggers build within a population, some one should be able to observe the effects. The variables for modelling included selected lags of up to 7-days with specific lag selected with reference to their partial autocorrelation function (PACF) plots. The results from this study showed among other things that the lag model prediction of peak admissions were often slightly out of synchronization with the actual data, but the days of greater admissions were better matched than the days of lower admissions.

Both the negative binomial and autoregressive techniques model the expected value of the daily admissions – i.e. the distributional mean. In many cases, particularly in an area such as health services planning, the mean value while useful, may not be as useful as expected peak in demand.

We developed multivariable QRMs to predict peak daily asthma admissions in London using selected weather and air quality factors. This study established that the associations between asthma and environmental factors including temperature, ozone & carbon monoxide can be exploited in predicting peak asthma admissions using a multistage variable selection criteria and QRMs. However, a major weakness in this study was paucity in data; considering the approach which limits prediction only to a specific percentile (90^{th} percentile).

Reviewers of the paper in chapter 9 in which we model the peaks in asthma admissions using the London data 2005-2006 expressed concern about the small size of the dataset.

In the absence of a larger admissions dataset we were compelled to cast our net more widely. There arose suitable dataset, which contained 70,830 respiratory related deaths in NYC, we investigated how to predict higher than expected respiratory deaths. This final study provided a further backing to the approach of predicting extreme/peak health events for alerting health services using QRMs.

Being an important decision making tool for health services delivery, the nature of demand for health forecast determines the approaches needed to forecast. In the case of asthma, forecasting immediate future events is important for service providers, but even more important is the forecasting of peak events, where extra resource mobilization is pertinent. Both approaches have been investigated in this section, and these analytical studies may have implications for health policy and programs.

Chapter 7

7.0 Introduction

In the research paper presented in this chapter, which is on forecasting asthma related hospital admissions with negative binomial models, a two pronged strategy was adopted in the selection of variables. When lags of data are available, and there are numerous measurements from which one has to choose, there is no single strategy that is endorsed. It is particularly uncertain when the goal is forecasting and not causal modelling. These approaches were designed to be exhaustive and sensitive enough to pick up the right combination of predictors. The first approach involved modelling with a 7-day average of each exposure variables, in order to account for the cumulative effect of various exposures. This followed a search of suitable predictor variables. The second approach to variable selection was a thorough and exhaustive search for the combinations of lagged variables that could best predict asthma daily admissions. This followed a non-explicit data mining approach (30). The NBMs obtained from these two approaches to variable selection were then contrasted with a base model (i.e. seasonal effects only). Issues of over-fitting are discussed and managed in the paper.

7.1 Forecasting asthma related hospital admissions with negative binomial models

7.1.1 Declarations for Thesis Chapter 7

Monash University

Declaration for Thesis Chapter 7

Declaration by candidate

In the paper submitted to the *Chronic Research Disease* Journal: *Forecasting asthma related hospital admissions in London using negative binomial models* /Chapter 7 of thesis, the nature and extent of my contribution to the work was the following:

Nature of contribution	Extent of contribution (%)
Lead and corresponding author, conceptualized the idea, accessed, organized and analyzed data, discussed preliminary results with stakeholders, drafted initial manuscript for circulation, submitted manuscript to journal and then managed correspondence with editors/reviewers, editorial staff and publishers until the final publication	100

The following co-authors contributed to the work. Co-authors who are students at Monash University must also indicate the extent of their contribution in percentage terms:

Name	Nature of contribution	Extent of contribution (%) for student co-authors only
Daniel D. REIDPATH	Supervised all activities leading to the publication of the paper; critiqued conceptual ideas, analyses and all drafts of manuscript	N/A
Christophe SARRAN	Assisted in accessing data, provided expert interpretation to the data and took part in reviewing and critiquing manuscript prior to submission	N/A

Candidate's Signature	Ireneous N. Soyiri	**Date:** 22-11-2012

Declaration by co-authors

The undersigned hereby certify that:

(1) the above declaration correctly reflects the nature and extent of the candidate's contribution to this work, and the nature of the contribution of each of the co-authors.
(2) they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
(3) they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
(4) there are no other authors of the publication according to these criteria;
(5) potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and
(6) the original data are stored at the following location(s) and will be held for at least five years from the date indicated below:

Location(s)	**Global Public Health, School of Medicine & Health Sciences, Monash University Sunway campus**

Signature 1	Daniel D. Reidpath	**Date:** 22-11-2012
Signature 2	Christophe Sarran	**Date:** 14-08-2012

- -

Article

Chronic Respiratory Disease

Chronic Respiratory Disease
10(2) 85–94

Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/1479972313482847
crd.sagepub.com
SAGE

Forecasting asthma-related hospital admissions in London using negative binomial models

Ireneous N Soyiri[1], Daniel D Reidpath[1] and Christophe Sarran[2]

Abstract

Health forecasting can improve health service provision and individual patient outcomes. Environmental factors are known to impact chronic respiratory conditions such as asthma, but little is known about the extent to which these factors can be used for forecasting. Using weather, air quality and hospital asthma admissions, in London (2005–2006), two related negative binomial models were developed and compared with a naive seasonal model. In the first approach, predictive forecasting models were fitted with 7-day averages of each potential predictor, and then a subsequent multivariable model is constructed. In the second strategy, an exhaustive search of the best fitting models between possible combinations of lags (0–14 days) of all the environmental effects on asthma admission was conducted. Three models were considered: a base model (seasonal effects), contrasted with a 7-day average model and a selected lags model (weather and air quality effects). Season is the best predictor of asthma admissions. The 7-day average and seasonal models were trivial to implement. The selected lags model was computationally intensive, but of no real value over much more easily implemented models. Seasonal factors can predict daily hospital asthma admissions in London, and there is a little evidence that additional weather and air quality information would add to forecast accuracy.

Keywords

Health forecasting, asthma, air quality, weather, population health, respiratory, model, statistics, London, hospital admission

The successful forecasting of the future health events can be used to improve health service provision and individual patient outcomes.[1–5] An example of the latter form of health forecasting was developed for the patients having chronic obstructive pulmonary disease (COPD) and was offered by the United Kingdom Meteorological Office (the Met Office). By understanding the relationship between weather, air quality and the onset of COPD crises,[6–9] the Met Office sought to alert COPD patients about changes in their personal risk of an adverse event.[2] Once alerted to an increased risk, individual COPD patients could then take steps to mitigate that risk. Nonetheless, there are reported challenges associated with realising the acclaimed benefits of a health forecasting scheme,[10,11] even though successes have equally been documented.[12]

Forecasting health events for health service provision, however, has the potential to have a more far reaching public health effect than simply mitigating adverse health events in individuals at known risk.[3] Health service providers can also be alerted to a likely increase in demand for services. Forewarned, hospitals can make rational decisions about resource allocation. Do extra beds need to be made available? Do extra staff or staff with particular skill sets need to be put on the roster? The obvious area in which health forecasting could play a role is in the prediction of adverse events that related to time-varying environmental exposures, such as those diseases affected by weather and air quality.

[1] Monash University, Bandar Sunway, Malaysia
[2] Met Office, Exeter, UK

Corresponding author:
Ireneous N Soyiri, Monash University, Jalan Lagoon Selatan, Bandar Sunway, Kuala Lumpur 46150, Malaysia.
Email: soyiriin@yahoo.com

Though there are a number of studies on the association between asthma and environmental factors,[13–16] relatively little research has been focussed on health forecasting.[17] In general, the approach taken to understanding the relationship between the occurrence of adverse health events and factors such as weather, air quality and season has relied on disentangling the causal relationships between the environmental factors and the health outcome. Time series analysis has been used to answer such questions as "is there a statistically significant relationship between an air quality factor (say $PM_{2.5}$) and cardiovascular disease, after controlling for other potential causes?"[18] In this way, researchers seek to understand the specific causes of adverse health; and by understanding the causes, it is hoped that long-term management strategies can be developed and government policy adjusted appropriately (such as emissions policies).

However, if the goal is to forecast the increase in the demand for hospital services, the causal relationships need not necessarily be understood. An indicator that is known not to have a causal relationship with an adverse health event may nonetheless be an exceptional predictor. As Breiman,[19] and Kostenko and Hyndman[20] have all observed, statistically significant causal models need not generate good predictions, and the measure of a forecasting model is its predictive performance.

Where there is a substantial literature looking at causal effects, there is surprisingly little literature that looks at forecasting the demand for health services based on environmental predictors (exceptions include studies by Bibi et al.,[21] Moustris et al.,[22] and Novikov et al.[23]). Using weather, air quality and hospital asthma admissions data from 2005 to 2006 in the London area (the region bounded by the M25 motorway), two related negative binomial models were developed and compared with a naive seasonal model. The modelling was constrained by need for a low cost, relatively easily implemented forecast.[24]

Methods

Ethics statement

An exemption from ethical review for the secondary analysis of hospital administrative data was obtained from the Monash University Human Research Ethics Committee (Number: 2011001092).

Forecasting models were developed using three date-linked datasets that included a daily record of hospital admissions, daily weather and daily air quality. The asthma admissions data are count data, and following earlier work, negative binomial models were developed with the key focus on forecasting.[25,26]

Data

Hospital (asthma) admissions data were sourced from the nationally recorded Hospital Episode Statistics maintained by the National Health Service, England. The data included an unidentified record of all asthma-related, emergency hospital admissions within London from 1 January 2005 to 31 December 2006 (i.e. 731 days of continuous data).

The operational definition for an asthma admission was any hospital emergency admission with a primary diagnosis of asthma (i.e. an ICD-10 code of "J-45"). A count of the asthma admissions across all the hospital emergency departments within London was recorded for each day of the study period, and this daily count was used as the primary dependent variable in the analyses.

Weather data were obtained from the UK Met Office database and were based on averaged daily results from the weather monitoring sites across London. The weather data contained 97% of complete daily records for: ambient air temperature, vapour pressure (HPa) and humidity (%). All temperature data were recorded in °C.

Air quality data were based on 24 h averages from air quality monitoring sites across London. The Numerical Atmospheric-dispersion Modelling Environment of Met Office was used to generate measures for all corresponding postcodes in the database. The indicators available with full daily records were carbon monoxide, formaldehyde, nitrogen dioxide, nitrogen oxide, ozone, particulate matter (specifically PM_{10}) and sulphur dioxide. All data were recorded in kilograms per cubic metre but converted to milligram per cubic metre for carbon monoxide and parts per million for the other pollutants. All the measured weather and air quality factors examined were identified in previous studies of respiratory or cardiac-related adverse health events, including asthma.[27–30] The incidence of respiratory illnesses are also known to be seasonally dependent[31] and so an additional temporal predictor (i.e. meteorological seasons) was generated to account for seasonality.

The final time series dataset aggregated the daily count of asthma admissions as the dependent variable, and potential predictors included the averaged 24h daily weather measures (including temperature and humidity) as well as the averaged 24 h daily air quality measures of ozone, nitrogen dioxide and nitrogen oxide for London. The complete date-linked dataset covered the period from 1 January 2005 to 31 December 2006. There were 24 days with missing temperature and humidity data. The missing data points were approximately uniformly distributed over the 2-year period.

Data analysis

Generally, in the analysis of time series data, particularly in causal modelling, it is important to base the analysis on a modified form of the time series known as a stationary time series. "Stationarity" refers to the idea that the probabilistic structure of the time series data is the same, no matter where in the series one begins to observe the data.[32] The value of a stationary time series in causal modelling arises because when two non-stationary time series are generated by independent processes, they can appear to be related simply in virtue of shared temporality and not from any true underlying relationship. In applied settings, *stationarity* is achieved by the removal of trend and periodicity elements from the data.[33] By contrast, in predictive modelling, a stationary time series is not critical because the test of the model is ultimately a predictive validity.[34] Furthermore, working with stationary time series for predictive modelling created a level of complexity not warranted given the implementation issues for asthma forecasting.

The data were analysed using negative binomial regression. Poisson regression is generally well suited to the modelling count data and is one of the most common techniques used for modelling asthma admissions.[35] However, it is not applicable for the cases in which the variance of the count data is substantially above the mean (i.e. over dispersion). The number of daily hospital admissions for asthma across London ranged from six admissions per day to 130 admissions per day. The distribution was observed to be slightly skewed with over dispersion of the variance, and it was for this reason, negative binomial regression was the preferred modelling technique.[26,36] The choice of negative binomial modelling technique for this dataset was further confirmed by a likelihood test suggested by Long and Freese.[37] This test is available as an application in the Stata statistical software and has been described earlier elsewhere.[38]

The negative binomial model for an expected number of daily admissions for asthma can be presented in the following form

$$\Pr(Y = y|\lambda, \alpha) = \left[\Gamma\left(y + \alpha^{-1}\right)/y!\Gamma\left(\alpha^{-1}\right)\right]\cdot \left[\alpha^{-1}/\left(\alpha^{-1} + \lambda\right)\right]^{\alpha^{-1}}\cdot\left[\lambda/\alpha^{-1} + \lambda\right]^{y}$$

where λ is the mean of the distribution, α is the over-dispersion parameter, y is the number of daily asthma admissions and Γ is the gamma function. Further interpretation to the negative binomial model has been described earlier.[38]

Data analysis followed a traditional approach taken in forecasting, which is to divide the dataset into a hold-in sample (which refers to the sample of the data used in developing forecast models) for model development, and then test the fit of the model against a hold-out sample (refers to the sample of the data used in validating forecast models) or cross-validation sample. Model development was conducted using 16 months of data, reserving the last 8 months of data for cross-validation. The selected split on the data allowed for more than a single annual cycle of data for the model development, unfortunately constrained by the fact that there were only two annual cycles in the full dataset. Limitations of this are discussed later. A simple seasonal model based on dummy predictors for autumn, winter and spring was developed for comparison purposes. This had a better fit than the equivalent Fourier series.[39] However, a similar model including dummy predictors of "day of the week" did not yield a better fit for the overall model, and so was dropped. Subsequent models involving weather and air quality variables included the seasonal predictors. A forecast "model" of the 2005 average daily admissions (i.e. 28.05) was also used as a point of comparison.

One of the challenges in time series analyses is the selection of appropriate lags. That is, how many days back should one go back in history to identify appropriate predictors, and how many lags should be included; that is, was it just yesterday's humidity level that was important or was it yesterday's and the day before's? (See. for instance, study by Peng and Dominici[33]). Most of the causal modelling research in asthma has limited the investigated lags to 14 days or fewer, but even with this constraint, there are 2^{14}-1 (i.e. 16,383) possible combinations of lags for each of the three weather variables and each of the seven air quality variables. Two strategies were adopted to reduce the variable selection space; both strategies used the Akaike Information Criteria (AIC) as a measure of the fit of the models for the purposes of model development[40] and as a mechanism for removing variables that did not contribute to the models' fit.[39]

The first strategy was to fit a model with the 7-day averages of each of the 10 predictors,[22] and use backward elimination to remove variables that were not statistically significant, while checking that the fit (AIC) improved. This was the 7-day average model. When data were missing in the 7-day period, the average of the remaining data was used. The second strategy was to conduct an exhaustive search of the fit (AIC) between the 2^{14}-1 possible combination of lags and asthma admission, for each weather and air quality variable in turn. For any one variable, the smallest model that had an AIC no greater than 2 and more than the best fitting model was selected.[39] An initial model was developed using the combination of best-fit lags for all 10 predictors. Backward elimination was then used to remove variables that were not statistically significant, while ensuring that the fit (AIC) improved. This was the selected lags model. Unfortunately, when data were missing for a particular day, the "missingness" was propagated across the lags. Data were missing for 24 of the 730 days for air temperature, humidity and vapour pressure. No data were missing for the air quality measures.

Validation and forecasting. There has been considerable discussion in the literature about appropriate measures of forecasting accuracy.[40–42] We used three measures: root mean squared error (RMSE), because it has traditionally been widely used in forecasting evaluation[42]; mean absolute percentage error (MAPE), because it is currently the most widely used measure of forecast accuracy[41]; and mean absolute scaled error (MASE), because it has desirable properties for comparing across models.[40,41]

The RMSE is an error measure of the squared difference between an observation (A) and a forecast (F) at a given time, t, and is usually presented as follows

$$\mathrm{RMSE} = \sqrt{(A_t - F_t)^2}$$

This approach for estimating the forecast error has the ability to capture an error that is not localised, but also not widely distributed in data. It also has the ability to differentiate error measures at different points in history.

The MAPE is an error measure based on generic percentages. MAPE is estimated as follows

$$\mathrm{MAPE} = \frac{100\%}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right|$$

where A_t is the true value, F_t is the forecast value, and n and t represent the number of individuals and time, respectively. MAPE is a useful measure because it is less cumbersome to report comparative forecast models. It is however worth noting that MAPE has a limitation in measuring or estimating forecast error when $n = 0$.

MASE is another approach for estimating forecasting error that compares forecast accuracy across a series on different scales. It is presented as follows

$$\mathrm{MASE} = \frac{1}{n} \sum_{t=0}^{n} \left(\left| \frac{e_t}{\frac{1}{n-1} \sum_{i-2}^{n} |A_i - A_{i-1}|} \right| \right)$$

where, A_t = actual value; F_t = forecast value; e_t = forecast error for a given period (t); and $e_t = A_t - F_t$. The MASE has been described in greater detail by Hyndman and Koehler.[41]

The analyses were conducted in the R statistical environment (R Foundation for Statistical Computing, R Development Core Team, Vienna, Austria) and Stata statistical software version 11 (Stata Corp LP., College Station, Texas, USA).

Results

The summary statistics of the key variables used in the analyses are presented in Table 1. During the hold-in period, the average daily number of admissions was 27.9 (SD = 9.3), and in the hold-out period, it was 29.7 (SD = 10.7). Because the hold-in and hold-out periods cover different periods of time in a year, it is unwise to try and interpret the difference. There appear, however, to be no radical differences in the means and ranges of the weather and air quality data for the hold-in and hold-out periods. Missing weather data are observable from the variations in sample size for both the hold-in and hold-out data.

A graph of asthma admissions from 1 January 2005 to 31 December 2006 is shown in Figure 1. The separation between the hold-in and hold-out periods is indicated by a vertical line on 25 April 2006. Even over the 2-year period, there appear to be cycles for the admissions. There are two notable admission peaks that occur in the series, one in the early summer of 2005 (130 admissions) and one in the early summer of 2006 (77 admissions).

The base model against which the other models were compared was the seasonal, negative binomial model that was developed using the hold-in data; and forecasts were made using the seasonal model against

Table 1. Summary statistics of variables used in analysis.

Variable	Observation	Mean	SD	Minimum	Maximum
Hold-in sample					
Asthma admissions	486	27.88	9.3	10	130
Air temperature	470	9.67	6.1	−2.54	26.48
Vapour pressure	470	9.85	3.7	3.425	20.68
Humidity	470	78.24	11.90698	35.2	99.5
Ozone	486	0.010959	0.0059467	0.000848	0.032167
Nitrogen dioxide	486	0.021878	0.0076514	0.009238	0.052433
Nitrogen oxide	486	0.017128	0.0115094	0.002534	0.06597
Carbon monoxide	486	0.000254	0.0000621	0.000155	0.000523
PM_{10}	486	0.011079	0.0088792	0.00166	0.059955
Sulphur dioxide	486	0.011972	0.0068856	0.002902	0.037904
Formaldehyde	486	0.006517	0.0032655	0.001703	0.018419
Hold-out sample					
Asthma admissions	244	29.69	10.7	6	77
Air temperature	236	14.35	5.6	−1.8	26.28
Vapour pressure	236	12.76	3.4	5.18	20.75
Humidity	236	77.1	14.2	35	99
Ozone	244	0.012231	0.0055678	0.001136	0.031115
Nitrogen dioxide	244	0.023201	0.0083452	0.009518	0.056262
Nitrogen oxide	244	0.017336	0.0113111	0.002157	0.073087
Carbon monoxide	244	0.000226	0.0000531	0.000137	0.000437
PM_{10}	244	0.011565	0.0092531	0.001461	0.045469
Sulphur dioxide	244	0.013756	0.0081773	0.003051	0.042769
Formaldehyde	244	0.006473	0.0032274	0.001667	0.018969

PM_{10}: particulate matter.

the hold-out data. The initial 7-day average model was developed on the hold-in data. The model included season and the 7-day averages for all 10 available weather and air quality variables (AIC = 3251). The reduced 7-day average model had an AIC of 3245, that is, an improvement in AIC with a reduction in the number of predictors. The reduced 7-day average model included: season, air temperature, vapour pressure, carbon monoxide, sulphur dioxide, nitrogen oxide and PM_{10}. Figure 2 shows the plot of the seasonal model and the reduced 7-day average model for the hold-in and hold-out (forecast) periods. In this illustration, the 7-day average model (solid black line) and the seasonal model (dashed black line) are shown over the time series of asthma admissions (grey line). Visually, the fit appears better for both the seasonal model and the 7-day average model for the hold-in period than the hold-out period. In the hold-out period, neither model seems to predict few enough admissions during the period of low admissions or a large enough number during the period of high admissions.

The selected lags models included season, and a total of 21 separate lags from 10 different weather and air quality variables. The AIC for the initial model fitted to the hold-in data was 2598. The reduced, selected lags model included season, three lags for air temperature (2, 6 and 9 days), three lags for humidity (2, 3 and 4 days), one lag for vapour pressure (14 day), two lags for ozone (7 and 14 days), one lag for nitrogen oxide (3 day) and one lag for formaldehyde (1 day). The AIC for the reduced model was 2585—a definite improvement with a reduction in model size. Figure 3 shows the selected lags model (solid black line) and the seasonal model (dashed black line) over the time series of asthma admissions (grey line). The gaps in the fitted line for the selected lags models indicate the dates with missing data. Because missing data are propagated across the dataset when they are lagged, the dates with no fitted data occur relatively frequently. The fit of the selected lags model in the hold-in period appears to track the actual admission data better than in the hold-out period, with the forecasts cutting through the trough (occurring around late July 2006) and the peak (around early November 2006).

A comparison of the performance of the models was made using RMSE, MAPE and MASE (Table 2).

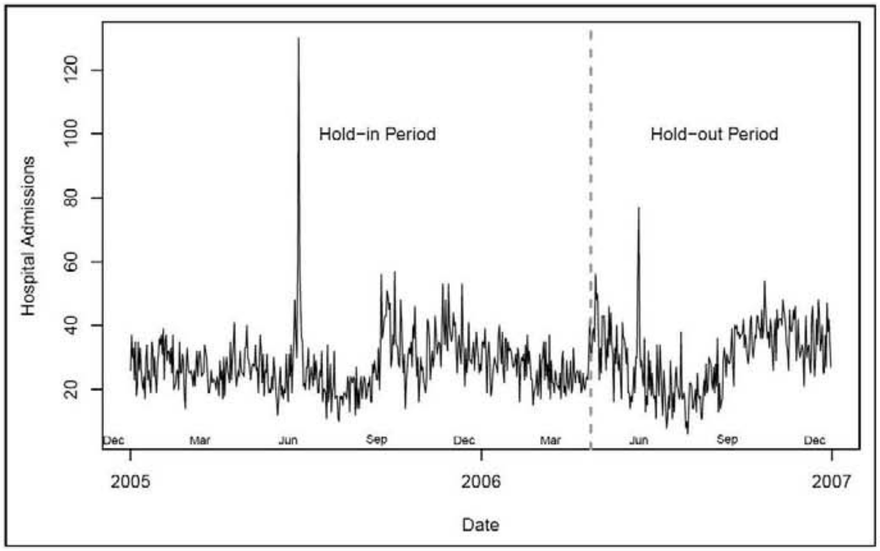


Figure 1. A graph of asthma admissions in London (1 January 2005 to 31 December 2006). Vertical dotted line indicates the separation between the hold-in and hold-out periods.

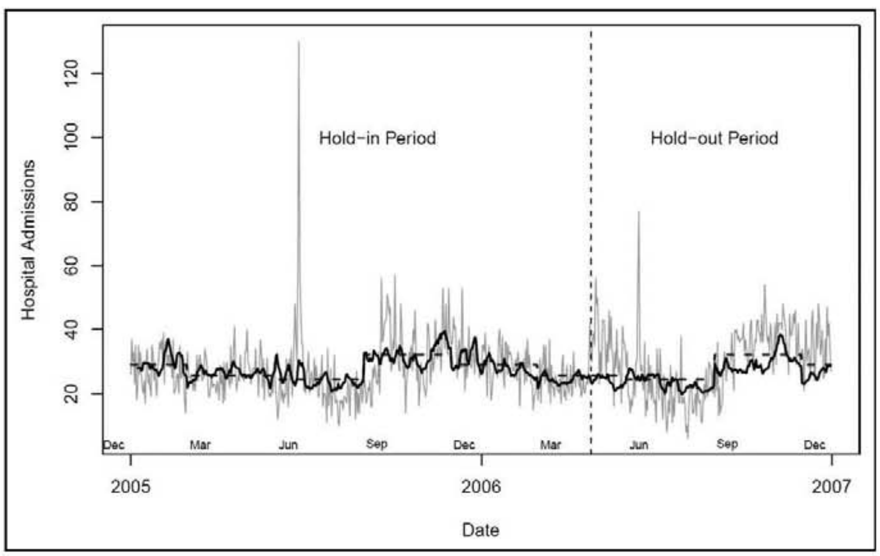


Figure 2. The plot of seasonal model and the reduced 7-day average model for the hold-in and hold-out (forecast) periods for asthma daily admissions in London (2005–2006). Solid black line indicates the 7-day average model. Dashed black line indicates the seasonal model. Grey line indicates the time series of asthma admissions.

The seasonal model was used as the comparison model and provided the scaling factor for the calculation of MASE. The 2005 mean admissions was used as a naive model for comparison purposes. The selected lags model consistently underperformed the seasonal and the 7-day average models for the hold-in and the hold-out data and underperformed the mean model for a number of comparisons. The 7-day average model outperformed the seasonal model for both the hold-in and the hold-out data when the models were compared using MAPE. When compared using RMSE, the seasonal model outperformed the 7-day

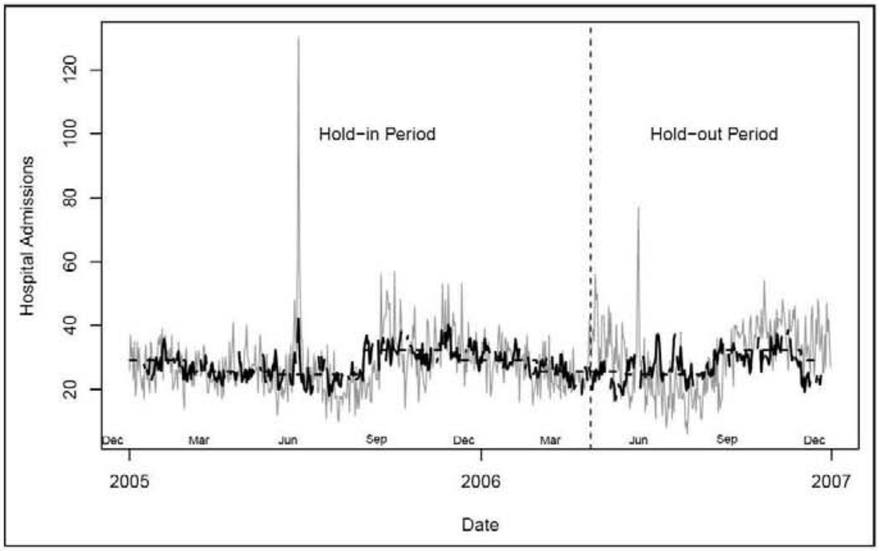


Figure 3. The plot of the selected lags model (solid black line) and the seasonal model (dashed black line) over the time series of asthma admissions (grey line) in London (2005–2006). Solid black line indicates the selected lag models. Dashed black line indicates the seasonal model. Grey line indicates the time series of asthma admissions.

Table 2. Performance of the seasonal (naive) model, the 7-day average model and the selected lags model in forecasting hospital asthma admissions in London (2005–2006).

Models	MAPE		RMSE		MASE	
	Hold-in	Hold-out	Hold-in	Hold-out	Hold-in	Hold-out
2005 Mean	25.59	37.97	9.24	10.98	1.11	1.54
Season only	22.65	31.1	8.8	9.84	1	1.32
7-day Average	22.02	30.50	8.47	10.4	0.98	1.40
Selected lags	23.04	37.54	8.67	11.15	1.01	1.57

MAPE: mean absolute percentage error; RMSE: root mean squared error; MASE: mean absolute scaled error.

average model for the hold-out data, but not the hold-in data. When compared using MASE, the seasonal model outperformed the 7-day average and selected lags models on the hold-in and hold-out data. The seasonal model and the 7-day average model consistently outperformed the mean model.

Discussion

In this study, we assessed the temporal variation in environmental exposures and asthma daily admissions in London over a 2-year period, using negative binomial models to iteratively model the effects of these exposures. We observed that seasonality was the main predictor of asthma daily admissions with little influence of additional meteorological data.

Health forecasting is important for health systems and services delivery and can provide additional decision-making tools/ways of doing more with existing resources and health data.[1,5] Given its potential far reaching public health benefit, health forecasting for chronic health conditions like asthma, can guide the planning process and also safeguard resource utilisation in health delivery.

Environmental—weather and air quality—factors are known to have a significant causal effect on respiratory events, including asthma; and this has a small but significant literature.[6,7] In contrast, forecasting future adverse respiratory events based on current and past weather and air quality factors is an inchoate field,[2,7,11] and within that field, there has been relatively little research looking at the forecasting of the demand for hospital services.[21–23] The task is important for ensuring the delivery of efficient and appropriate care according to the needs of the community.

Three models were considered here. A base model-seasonal effects only-contrasted with a 7-day average model and a selected lags model, which included weather and air quality predictors consistent with the literature.[9,13–16,27] The naive mean model was also included. The single most striking feature of the models was the importance of season as a predictor. For RMSE and MASE, season produced the best forecasting models, and the forecast accuracy deteriorated with the inclusion of weather and air quality data. Only for MAPE did the 7-day average model marginally outperform the seasonal model, and MAPE is known to be biased by analysing whether the forecast value is above or below the true value.[41] Even the mean model outperformed the selected lags model for the hold-out data for the RMSE and MASE measures.

The failure of weather and air quality models to outperform a seasonal model is surprisingly given the reported success of the Met Office COPD forecasting model[2] and other asthma forecasting models (Moustris et al.[22]). However, there are some important differences in the approach, which may shed some light on this. The COPD model[2] included seasonal effects in the model and did not contrast the seasonal model with the combined seasonal, weather and air quality factors. The COPD model also used R^2 as the measure of fit, which is quite unreliable in forecasting, because one can obtain an $R^2$of 1 and be consistently wrong. It is well known that for the last 40 years, air temperature exacerbates COPD[8] and air temperature is strongly seasonal. In a recent asthma forecasting model, season again was implicitly included in the model as dummy variables of month,[22] and again, there was no contrast model that just contained season. The study also used R^2 as a measure of fit as well as RMSE. The more widely accepted MAPE and recently proposed MASE were not used.

The 7-day average and season models were trivial to implement and do not rely on excessive numbers of weather or air quality factors. The selected lags model is computationally intensive, but appears to be of no real value over much more easily implemented models. There are, however, a number of limitations with the approach taken, and these need to be factored into decisions about future directions for research. The hold-out dataset on which the validation was conducted did not cover a full year. This means that predictions associated with certain times of the year were missed, and this would need to be considered in future research, utilising more than 2-years of data. The causal relationships between weather, air quality and asthma are not uniform across geographical locations.[27–30,35] That is, the findings cannot be uncritically generalised from one setting to another. A similarly cautious approach should be taken in the development of forecasting models. Where the approach may be used as a guide for future research, the specifics would almost certainly require “localisation”.

There is a clear need in the health forecasting area for researchers to adopt consistent approaches that allow a ready comparison of models. Season is a basic factor influencing hospital utilisation for respiratory diseases, and it is important to know if (for the sake of simplicity) it is enough just to take account of season, or whether additional factors would add significantly to forecasting accuracy. Future research would need to explore alternative modelling techniques, and forecasting peak admissions rather than average admissions may ultimately be of greater value to health service planners. In situations like these when multivariate time series predictions become limited to temporal factors, because of the lack of exposure-related phenomena that predicts asthma admissions, it is suggested that non-linear techniques be used to complement predictions of particularly extreme events. One of such approaches is the use of quantile regression models that help to predict unusual events.[43] Other approaches may involve a detailed examination of the temporal fluctuations of daily admissions, in order to identify if the pattern of behaviour follow a power law/function that can be used in prediction. The latter approach is yet to be investigated with our dataset.

The daily hospital asthma admissions in London may be predicted and forecast using seasonal factors, and there is a little evidence that additional weather and air quality information would add to forecast accuracy. It is not trivial to assemble data on all the known confounders. A weakness in our study was the lack of data on some commonly known effects like viral or influenza epidemics and pollen counts, which have a major influence in exacerbating respiratory diseases. Furthermore, obtaining representative population exposure measures for a wide and diverse area like London is difficult. This is because weather conditions and air pollutant levels vary widely even in small areas, and, more particularly, between indoors and outdoors. It is therefore difficult to know if this result is unexpected, because other forecasting studies seem to have included seasonal factors as a matter of course. The computationally intensive–exhaustive search for the best fitting lags results in a relatively poorly fitting model. There is real potential value for relatively simple models in forecasting demand for hospital services, and hence this article presents an opportunity for further analysis and forecasting of asthma daily admissions in London using any available current data.

Acknowledgements

The article was conceived by INS with input from DDR. INS reviewed the literature and conducted data analysis with input from DDR. The first draft was produced by INS, and INS, DDR and CS participated in the writing/editing of subsequent drafts. The final manuscript was approved by all authors.

Conflict of Interest

The authors declared no conflicts of interest.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

1. Krishna S, Balas EA, Boren SA and Maglaveras N. Patient acceptance of educational voice messages: a review of controlled clinical studies. *Methods Inf Med* 2002; 41(5): 360–369.
2. Marno P, Bryden C, Bird W and Watkin HA. How different measures of cold weather affect chronic obstructive pulmonary disease (COPD) hospital admissions in London. *Eur Respir Rev* 2006; 15(101): 185–186.
3. Rodo X, Ballester J, Cayan D, Melish ME, Nakamura Y and Uehara R. Association of Kawasaki disease with tropospheric wind patterns. *Sci Rep* 2011; 1: 152.
4. Bryden C, Bird W, Titley HA, Halpin DMG and Levy ML. Stratification of COPD patients by previous admission for targeting of preventative care. *Respir Med* 2009; 103(4): 558–565.
5. WHO. *Everybody's Business: Strengthening Health Systems to Improve Health Outcomes*. Geneva, Switzerland: World Health Organization Document Production Services, http://www.who.int/entity/healthsystems/strategy/everybodys_business.pdf (2007, accessed 9 February 2012).
6. Ostro BD, Feng WY, Broadwin R, Malig BJ, Green RS and Lipsett MJ. The impact of components of fine particulate matter on cardiovascular mortality in susceptible subpopulations. *Occup Environ Med* 2008; 65(11): 750–756.
7. Arbex MA, Martins LC, de Oliveira RC, de Oliveira RC, Pereira LAA, Arbex FF and Cancado JED. Air pollution from biomass burning and asthma hospital admissions in a sugar cane plantation area in Brazil. *J Epidemiol Community Health* 2007; 61(5): 395–400.
8. Hsieh YC, Frayser R and Ross JC. The effect of cold-air inhalation on ventilation in normal subjects and in patients with chronic obstructive pulmonary disease. *Am Rev Respir Dis* 1968; 98(4): 613–622.
9. Muenchberger H and Kendall E. Predictors of preventable hospitalization in chronic disease: priorities for change. *J Public Health Policy* 2010; 31(2): 150–163.
10. Bakerly ND, Roberts JA, Thomson AR and Dyer M. The effect of COPD health forecasting on hospitalisation and health care utilisation in patients with mild-to-moderate COPD. *Chron Respir Dis* 2011; 8(1): 5–9.
11. Maheswaran R, Pearson T, Hoysal N and Campbell MJ. Evaluation of the impact of a health forecast alert service on admissions for chronic obstructive pulmonary disease in Bradford and Airedale. *J Public Health* 2009; 32(1): 97–102.
12. Marno P, Chalder M, Laing-Morton T, Levy M, Sachon P and Halpin D. Can a health forecasting service offer COPD patients a novel way to manage their condition? *J Health Serv Res Policy* 2010; 15(3): 150–155.
13. Abe T, Tokuda Y, Ohde S, Ishimatsu S, Nakamura T and Birrer RB. The relationship of short-term air pollution and weather to ED visits for asthma in Japan. *Am J Emerg Med* 2009; 27(2): 153–159.
14. Li S, Batterman S, Wasilevich E, Wahl R, Wirth J and Su FC. Association of daily asthma emergency department visits and hospital admissions with ambient air pollutants among the pediatric Medicaid population in Detroit: time-series and time-stratified case-crossover analyses with threshold effects. *Environ Res* 2011; 111(8): 1137–1147.
15. Matyasovszky I, Makra L, Bálint B, Guba Z and Sümeghy Z. Multivariate analysis of respiratory problems and their connection with meteorological parameters and the main biological and chemical air pollutants. *Atmos Environ* 2011; 45(415): 2e4159.
16. Rage E, Siroux V, Kunzli N, Pin I and Kauffmann F. Air pollution and asthma severity in adults. *Occup Environ Med* 2009; 66(3): 182–188.
17. Ward M. Health forecasting and COPD. *Chron Respir Dis* 2011; 8(1): 3–4.
18. Dominici F, Peng RD, Bell ML, Pham L, McDermott A and Zeger S. Fine particulate air pollution and

hospital admission for cardiovascular and respiratory diseases. *JAMA* 2006; 295(10): 1127–1134.

19. Breiman L. Statistical modeling: the two cultures. *Stat Sci* 2001; 16(3): 199–215.
20. Kostenko AV and Hyndman RJ. A note on the categorization of demand patterns. *J Oper Res Soc* 2006; 57(10): 1256–1257.
21. Bibi H, Nutman A, Shoseyov D, Shalom M, Peled R and Kivity S. Prediction of emergency department visits for respiratory symptoms using an artificial neural network. *Chest* 2002; 122(5): 1627–1632.
22. Moustris KP, Douros K, Nastos PT, Larissi IK, Anthracopoulos MB and Paliatsos AG. Seven-days-ahead forecasting of childhood asthma admissions using artificial neural networks in Athens, Greece. *Int J Environ Health Res* 2011; 1–12.
23. Novikov I, Kalter-Leibovici O, Chetrit A, Stav N and Epstein Y. Weather conditions and visits to the medical wing of emergency rooms in a metropolitan area during the warm season in Israel: a predictive model. *Int J Biometeorol* 2011; 56(1): 121–127.
24. Gujarati DN. *Essentials of Econometrics*. 3rd ed. London: McGraw Hill Higher Education, 2005.
25. Elvik R. The stability of long-term trends in the number of traffic fatalities in a sample of highly motorised countries. *Accid Anal Prev* 2010; 42(1): 245–260.
26. Soyiri IN, Reidpath DD and Sarran C. Asthma length of stay in hospitals in London 2001-2006: demographic, diagnostic and temporal factors. *PLoS One* 2011; 6(11): e27184.
27. Babin S, Burkom H, Holtry R, Tabernero N, Davies-Cole J and Stokes L. Medicaid patient asthma-related acute care visits and their associations with ozone and particulates in Washington, DC, from 1994-2005. *Int J Environ Health Res* 2008; 18(3): 209–221.
28. Hajat S, Haines A, Goubet SA, Atkinson RW and Anderson HR. Association of air pollution with daily GP consultations for asthma and other lower respiratory conditions in London. *Thorax* 1999; 54(7): 597–605.
29. Hajat S, Kovats RS, Atkinson RW and Haines A. Impact of hot temperatures on death in London: a time series approach. *J Epidemiol Community Health* 2002; 56(5): 367–372.
30. Peng RD, Chang HH, Bell ML, McDermott A, Zeger SL and Samet JM. Coarse particulate matter air pollution and hospital admissions for cardiovascular and respiratory diseases among medicare patients. *JAMA* 2008; 299(18): 2172–2179.
31. Koster ES, Raaijmakers JAM, Vijverberg SJH, van der Ent CK and Maitland-van der Zee A-H. Asthma symptoms in pediatric patients: differences throughout the seasons. *J Asthma* 2011; 48(7): 694–700.
32. Diggle P. *Time Series: A Biostatistical Introduction*. Oxford, UK: Oxford University Press, 1990.
33. Peng RD and Dominici F. Statistical methods for environmental epidemiology with R: a case study in air pollution and health.USA: Springer-Verlag New York Inc., 2008.
34. Medina DC, Findley SE, Guindo B and Doumbia S. Forecasting non-stationary diarrhea, acute respiratory infection, and malaria time-series in Niono, Mali. *PLoS One* 2007; 2(11): e1181.
35. Erbas B, Kelly AM, Physick B, Code C and Edwards M. Air pollution and childhood asthma emergency hospital admissions: estimating intra-city regional variations. *Int J Environ Health Res* 2005; 15(1): 11–20.
36. Long JS. Predicted probabilities for count models. *Stata J* 2001; 1(1): 51–57.
37. Long JS and Freese J. *Regression Models for Categorical Dependent Variables Using Stata*. Texas, TX: Stata Press, 2003.
38. Soyiri IN and Reidpath DD. Semi-structured black-box prediction: proposed approach for asthma admissions in London. *Int J Gen Med* 2012; 5(1): 693–705.
39. Barnett AG and Dobson AJ. *Analysing Seasonal Health Data*. Heidelberg Dordrecht London New York: Springer, 2009.
40. Soyiri IN and Reidpath DD. Evolving forecasting classifications and applications in health forecasting. *Int J Gen Med* 2012; 5(1): 381–389.
41. Hyndman RJ and Koehler AB. Another look at measures of forecast accuracy. *Int J Forecast* 2006; 22(4): 679–688.
42. Armstrong JS and Collopy F. Error measures for generalizing about forecasting methods: empirical comparisons. *Int J Forecast* 1992; 8(1): 69–80.
43. Soyiri IN, Reidpath DD and Sarran C. Forecasting peak asthma admissions in London: an application of quantile regression models. *Int J Biometeorol*. Epub ahead of print 12 August 2012. DOI: 10.1007/s00484-012-0584-0.

Chapter 8

8.0 Introduction

In the previous chapter the seasonal model appears to be the strongest model. The paper (chapter 8) on using humans as animal sentinels for forecasting asthma events follows a similar approach in variable selection as the previous paper, but in this case uses only reported asthma daily admissions. Against one measure of error (MAPE) to model that included 7-day averages was slightly stronger. However it was not unusually stronger and it is more – although not much more - complex to implement.

The results are, however, somewhat surprising. We know that daily variations in weather and air quality are causally related to asthma events (151, 152). We are not however seeing the value of this information in the forecasting. Nonetheless, the paper examines the underlying idea that asthma sufferers with more sensitive lungs respond more quickly to changes in environmental exposure than those with less sensitive lungs, and hence serve as early warning signal for the latter group. This is tested using the lag models. The approach could have potential applications in other chronic disease conditions that are largely dependent on common environmental exposures.

8.1 Humans as animal sentinels for forecasting asthma events

8.1.1 Declarations for Thesis Chapter 8

Monash University

Declaration for Thesis Chapter 8

Declaration by candidate

In the paper: *Humans as animal sentinels for forecasting asthma events: helping health services become more responsive* /(Chapter 8) published in the *PLoS One* Journal (*PLoS One, 2012*. 7(10): e47823), the nature and extent of my contribution to the work was the following:

Nature of contribution	**Extent of contribution (%)**
Lead and corresponding author, conceptualized the idea, accessed, organized and analyzed data, discussed preliminary results with co-author, drafted initial manuscript, submitted the final manuscript to journal and then managed correspondence with editors/reviewers, editorial staff and publishers until the final publication	100

The following co-authors contributed to the work. Co-authors who are students at Monash University must also indicate the extent of their contribution in percentage terms:

Name	**Nature of contribution**	**Extent of contribution (%) for student co-authors only**
Daniel D. REIDPATH	Supervised all activities leading to the publication of the paper; critiqued conceptual ideas, analyses and all drafts of manuscript	N/A

Candidate's Signature	Ireneous N. Soyiri	**Date:** 22-11-2012

Declaration by co-authors

The undersigned hereby certify that:

(1) the above declaration correctly reflects the nature and extent of the candidate's contribution to this work, and the nature of the contribution of each of the co-authors.

(2) they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
(3) they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
(4) there are no other authors of the publication according to these criteria;
(5) potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and
(6) the original data are stored at the following location(s) and will be held for at least five years from the date indicated below:

Location(s)	**Global Public Health, School of Medicine & Health Sciences, Monash University Sunway campus**

Signature 1	Daniel D. Reidpath	**Date:** 22-11-2012

- -

Humans as Animal Sentinels for Forecasting Asthma Events: Helping Health Services Become More Responsive

Ireneous N. Soyiri[1,2]*, Daniel D. Reidpath[1]

1 SEACO, Monash University, Bandar Sunway, Selangor, Malaysia, 2 School of Public Health, University of Ghana, Accra, Ghana

Abstract

The concept of forecasting asthma using humans as animal sentinels is uncommon. This study explores the plausibility of predicting future asthma daily admissions using retrospective data in London (2005–2006). Negative binomial regressions were used in modeling; allowing the non-contiguous autoregressive components. Selected lags were based on partial autocorrelation function (PACF) plot with a maximum lag of 7 days. The model was contrasted with naïve historical and seasonal models. All models were cross validated. Mean daily asthma admission in 2005 was 27.9 and in 2006 it was 28.9. The lags 1, 2, 3, 6 and 7 were independently associated with daily asthma admissions based on their PACF plots. The lag model prediction of peak admissions were often slightly out of synchronization with the actual data, but the days of greater admissions were better matched than the days of lower admissions. A further investigation across various populations is necessary.

Citation: Soyiri IN, Reidpath DD (2012) Humans as Animal Sentinels for Forecasting Asthma Events: Helping Health Services Become More Responsive. PLoS ONE 7(10): e47823. doi:10.1371/journal.pone.0047823

Editor: Devendra Amre, University of Montreal, Canada

Received May 25, 2012; **Accepted** September 17, 2012; **Published** October 31, 2012

Funding: The authors have no support or funding to report.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: soyiriin@yahoo.com

Introduction

Asthma is a chronic respiratory illness of immense global proportions, and it affects over 300 million people. Recent reviews have reaffirmed the highly heterogeneous nature of the disease, which is influenced by complex genetic and environmental effects as well as an evolving knowledge-base of its key determinants [1]. Many of these reviews comprehensively addressed the key factors which contribute to the manifestation and progression of asthma in individuals and lab based experiments [2,3,4]. There was however less content on the forecasting of asthma events for the purposes of providing early warning systems to help manage the condition in larger populations. Meanwhile, an approach to develop a forecast for respiratory conditions that are dependent on environmental exposures (e.g. asthma), which is yet to be reported in the literature, is the use of humans as animal sentinels to forecast asthma.

The classical animal sentinel is the canary in the coal mine. Coal miners would carry a caged canary with them into mines knowing that the birds were more sensitive to the toxic gases found in the coal seams than were the miners [5,6,7]. If the canary died then the humans had early warning about the presence of toxic gases and could evacuate the mine.

Since those times animal sentinels have been widely used for monitoring changes in environmental exposures [3,8,9,10]. Although it is not usually discussed in these terms, there is also a potential for humans to act as animal sentinels for environmental exposures for other humans. The use of syndromic surveillance to detect non-infectious bioterrorism is an example of this [11,12]. Unlike animal sentinels, however, where specific identifiable animals are followed up over time, human sentinel surveillance follows fluctuations in health events over entire populations. The logic is that people who are more sensitive to environmental exposures or (because of geographic location) people who experience earlier exposure will present in hospital records sooner than the less sensitive. As the dose of an environmental exposure increases (or diffuses across the population), so more people will experience health events. Thus, temporal fluctuations in the numbers patients presenting to hospitals will be, in part, attributable to fluctuations in environmental exposure.

There is the potential to utilize human sentinels for predicting more routine variations in disease events to inform health service provision. For example, in the case of asthma events, those people with more sensitive lungs are likely to respond more quickly to changes in environmental exposure than those people with less sensitive lungs. In effect, the sensitive lung is "the canary in the coal mine" for the less sensitive lung. Without having to measure any particular environmental trigger or determine the causal relationships between environmental exposures and asthma events, the potential exists to use the frequency of asthma events today to predict the frequency of asthma events in the future and feed this into decision making about health services provision.

Previous studies have looked at the forecasting of asthma events, but have tended to focus on relationships between the environmental exposures which are known to trigger asthma events, such as weather conditions or Ozone and PM10 levels, as well as the extent to which these can be used to forecast asthma [13 18]. Other related studies, such as the recent study by Eisner and colleagues on the use of an assessment tool for measuring the "severity of asthma score" and using it to predicts clinical outcomes in patients with moderate to severe persistent asthma, have demonstrated the predictability of adverse clinical outcomes in specific group of patients (i.e. moderate to severe asthma) [19]. In contrast, it is the aim of the present study to ignore the specifics of any environmental exposure or demographic factor(s), and focus exclusively on the possibility of using sentinel humans living within the community to forecast asthma events. If asthma sufferers can be used as sentinels for other asthma sufferers, the possibility exists that by monitoring changes in the number of asthma events, health services would be able to respond more efficiently to the future demands. As a result individual asthma sufferers could be alerted to their personal increased risk. The plausibility however needs to be established first before the potential value to health issues can be explored.

The objective of this study was to examine the relative value of autoregressive models to forecast asthma admissions using data for two years of hospital admissions for asthma from London (2005 2006). Because the interest is forecasting performance, and there is no sense in which one can suggest that the lagged count of asthma admissions from some days ago caused the asthma admissions of today, reporting the parameter estimates for particular lags are likely to be of little value, or misleading [20]. We focus, therefore on the more relevant predictive performance of the models

Methods

This study involved the development of an asthma forecasting model based on a secondary analysis of hospital administrative data from London, England. The data covered 20,794 hospital admissions that occurred within the perimeter formed by the M25 Motorway (surrounding London) where the admissions had a primary diagnosis of asthma.

Data

Data were sourced from the nationally recorded Hospital Episode Statistics (HES) maintained by the National Health Service, England [21]. Asthma admissions were defined as any hospital admission with a primary diagnosis of asthma; i.e., an International Classification of Diseases (ICD-) 10 code of J45. The data covered all days between January 1st, 2005 and December 31st, 2006 with no missing data.

The outcome variable for the study was the daily count of admissions for asthma. The predictor variables were selected lags of previous days' admissions. The selection of lags is explained in the following section (Data Analysis). The data were divided into two annual sets: a model development data set from the 2005 admissions data and a cross validation data set from 2006 admissions data.

Based on the aggregate, anonymous and administrative nature of the data, an exemption from ethical review for the secondary analysis was obtained from the Monash University Human Research Ethics Committee (Number: 2011001092).

Data Analysis

The analysis of the data relied on a comparison of forecasting models of asthma daily admissions in which 2005 hospitals admissions data was used in the development of three negative binomial regression models, and 2006 data were used for cross-validation. The three models were:

A mean daily admissions (historical model). This model was a null model that included no predictor variables.

A seasonal model: The seasonal model included three dummy predictor variables to model the effects of the four seasons. Season was dummy coded, in keeping with earlier work using these data, because this fitted the data better than a smoothed seasonal model.

An autoregressive (lags) models: A lag represents the admissions count from a previous day. Thus a 1 day lag represents the admissions count from the day before the day being modelled, and a two day lag represents the admissions count from two days prior to the day being modelled. The lags model included the non-contiguous lagged data from the days prior to the modeled day as predictor variables. The lags were informed by a partial autocorrelation function (PACF) plot with a maximum lag of 7 days.

Negative binomial regression was chosen for the modelling because the asthma daily admissions counts were known to have issues with over dispersion, [22 27]. Following Hilbe, [28] the probability model can be conceptualised in the following way. P is the probability function of the negative binomial distribution:

$$P(Y=y_i|X_i)=\frac{\Gamma(y_i+1/\alpha)}{\Gamma(y_i+1)\Gamma(1/\alpha)}\cdot\left(\frac{1}{1+\alpha\mu}\right)^{1/\alpha}\cdot\left(\frac{\alpha\mu}{1+\alpha\mu}\right)^{y_i}$$

Where: y_i represents the number of admissions; $\mu=\exp(X_i\beta)$; β is the vector of coefficients; X_i is the vector of predictor variables (in this case "1" for the historical model, the dummy variables of three seasons for the seasonal model, and the admissions counts for the lagged days 1, 2, 3, 6 and 7 for the lags model); α is the overdispersion parameter; and Γ is the gamma function The predictor variable parameters (β) were estimated via maximum likelihood estimation.

A positive coefficient in the regression output indicates that a factor will increase the number of daily asthma admissions relative to its reference category and conversely a negative coefficient will decrease the number of daily asthma admissions relative to its reference category. The exponent of the coefficient can be interpreted, all other things being equal as the proportionate increase (for values greater than 1) or decrease (for values between 0 and 1) of number of daily asthma admissions associated with a one unit increase in the predictor variable [27,28]. The predictor variable(s) herein refers to the functional form of the lag term(s) constituting the NBM. As stated in the objective of this study, this univariate model does not account for other plausible indicators of asthma (e.g. pollution) other than lagged asthma events. We acknowledge that, accounting for multivariable factors is beyond the scope of this paper, even though they may be viewed as potentially confounding risk factors that are also time dependant. Hence for our analyses, specific potential covariates were selected nonlinear lags of 0 to 7 days of asthma admissions from the training dataset (i.e. 2005 asthma daily admissions in London). To the best of our knowledge, there is no standard reference in current literature for lag selection for this kind of study, as it has not been carried out before. Hence our choice of this range of lags was to satisfy the biological plausibility of our hypothesis and also develop a tool which relies on a "short memory". The selection of lag combinations for the models involved a computationally exhaustive process, selecting the best fit for all possible lags.

Model Formulation

Three models were developed for comparison purposes, using the 2005 data. The first model was the mean daily admissions (historical) model. The final model utilized non-contiguous autoregressive lags. Season was *dummy* coded, in keeping with earlier work using these data that indicated a better fit than with a smoothed seasonal model.

1. Mean daily admissions (historical model): This model was defined by a function of the average daily asthma admissions in London in 2005;
2. Seasonal model: Then seasonal model was defined by four meteorological seasons, categorized as dummy variable;
3. Lags models: The lag model was defined by a function of combinations of the 0 7day lags which yielded the best predictive model. The model comprised a multivariable 1, 2, 3, 6 and 7 day lags.

Error measures

Three measures of fit were used to evaluate modeled data for 2005 and the predictive forecast of the model on the cross-validation data from 2006. The measures of predictive performance were R-squared, root mean squared error (RMSE) and mean absolute scaled error (MASE) [29]. RMSE was included because it is well known and still popular in the literature although it has known problems [30]. R-squared, though flawed as a measure of predictive validity, [31] remains popular and was included purely for historical reasons. MASE is now regarded as one of the better measures of predictive validity, [32] but it requires a scaling factor against which to measure performance. The scaling factor was derived from the mean absolute error of the predictions based on the 2005 historical mean daily admissions. When interpreting the measures of error, it should be noted that with the exception of R-squared, smaller numbers indicate less error between the forecast and actual data. In contrast, larger R-squared values are indicative of a better fit between the forecast and actual data.

Analyses were conducted using the R (Version 2.14.1) statistical environment [33] and Stata (version 11.2) statistical package [34].

Results

The mean daily asthma admission in 2005 was 27.9 and in 2006 it was 28.9. The plot of the PACF indicated lags 1, 2, 3, 6 and 7 were independently associated with daily asthma admissions. These plots lie within reasonable confidence bounds (i.e. 95% Confidence Interval). The negative binomial regression model was developed using these lags.

Figure 1 shows a plot of the asthma admissions data (grey line), and the lag model (dashed black line), seasonal model (solid black line) and the historical model (straight dashed line). A solid vertical line (1 January 2006) shows the division between the data on which the models were developed and the data on which the models were cross-validated (i.e., the predictive forecasts were measured).

It appears from the figure that the lag model captures the daily variation in the admissions better than the seasonal model, which is certainly better than the historical model. Careful scrutiny of the figure however shows the peak admissions predicted by the lag model are often slightly out of synchronization with the actual data. It also appears that the days of greater admissions are somewhat better matched than the days of lower admissions.

Table 1 shows the measured fit of the lag model, the seasonal model and the historical model. The scaling factor for the MASE measure was derived from the historical model. As a consequence, the MASE for the historical model for 2005 is 1, and all comparisons of fit relate to the fit of the historical model.

Discussion

Using human sentinels to forecast asthma events in large concentrated populations is uncommon. Previous studies on animal sentinels have tended to use mammals, which occupy shared environments and/or exposures with humans [10]. This study makes an important contribution by using retrospective asthma admission records in London to demonstrate the plausible hypothesis.

The idea of forecasting asthma using human sentinels was based on the probable observation that asthma sufferers with more sensitive lungs, all things being equal, would react more to environmental changes or to the precursors of asthma exacerbations than their less sensitive counterparts. Where others have considered lagged effects of pollutants on asthma, and sometimes included autoregressive components in their analysis, these have not been used for forecasting [13 18]. Where research has been conducted on forecasting of asthma (and other respiratory conditions), this has not considered autoreggressive predictors [19].

There is no consensus on the approach to developing health forecasting models. There is also no agreed scale in determining what constitutes a good health forecast model, but for the fact that such a model predicts well. The modeling approach described in this study is quite flexible because it provided an opportunity to choose the most suitable predictors and guarding against over fitting of the model by limiting the range of lags (covariates) to be selected.

Partial autocorrelation function plots (and other model diagnostic tools like Plot of time series residuals, Normal quantile plot and Autocorrelation function) have been found to be useful guides in selecting covariates for modeling and prediction [35,36]. A key advantage of this model building approach is that it combines fast input selection with accurate but computationally demanding non-linear predictions [37]. Additionally, the complexity of the input variable selection process makes the approach viable for large scale population health challenges. Ultimately, it still provides a wide range of potential models for the best forecast model to be selected based on the chosen measures of fit and cross validation.

Forecasting and error measures

There is little difference in the R^2 for the lag model in 2005 or 2006. Both measures account for a little over 35% of the variation in asthma daily admissions. The seasonal model, surprisingly, accounts for a greater proportion of the variation of asthma daily admissions in the cross validation period.

The RMSE statistics show that the lag model consistently out performs the seasonal model, which in turn consistently out performs the historical model. For the modeled data (2005), the seasonal model has an RMSE around 8% smaller than the historical model and the lag model has and RMSE about 21% smaller than the historical model. In the cross validation period (2006), the forecast predictions of all the models are (as expected) worse than they were for the modeled data. The rank order however remains unchanged, with the lag model out performing either of the other models. With respect to MASE, the seasonal models performance is around 15% lower than the performance of the historical model, and the lag model is around 25% lower than the performance of the historical model

The preference of MASE over RMSE and R^2 as an error measure for forecasting has also been discussed by previous authors [29,32]. The MASE statistics are more easily interpreted, and potentially the most reliable and informative measure of accuracy in forecasting [29]. It is widely recommended for

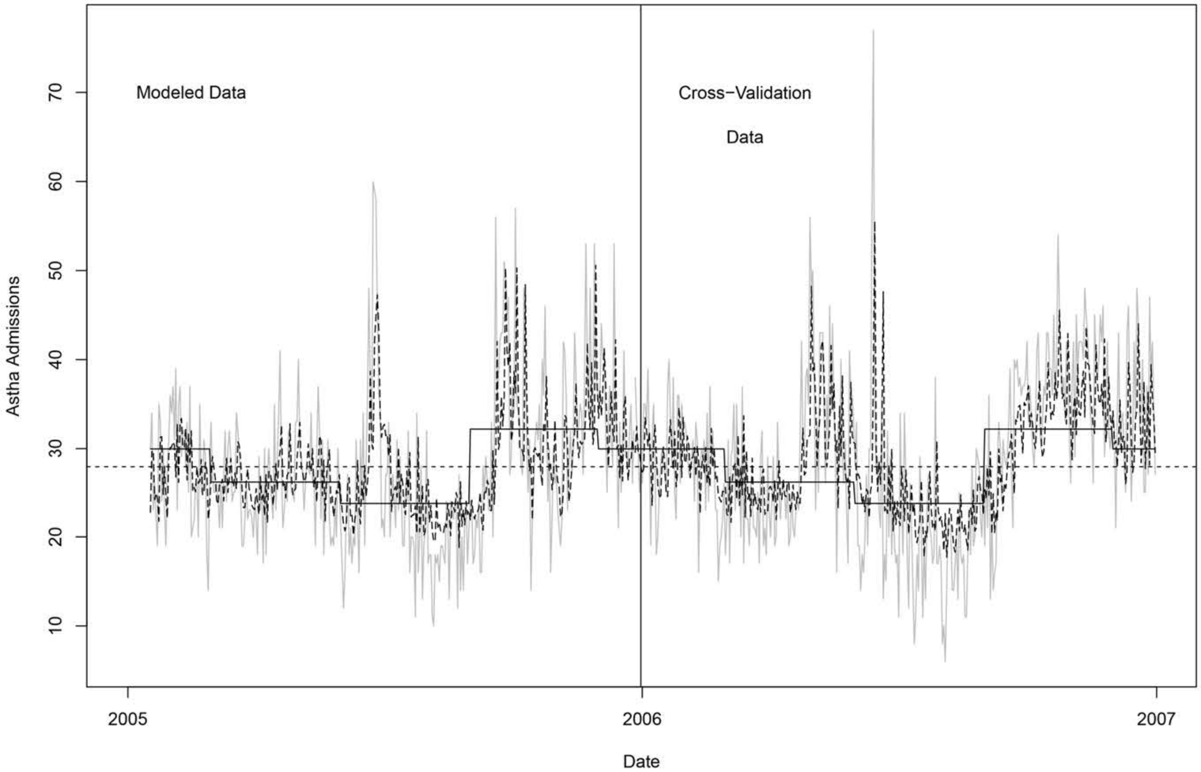


Figure 1. A plot of Asthma daily admissions in London (2005–2006). The grey line represents a plot of the actual asthma admissions data in London (2005–2006); The dashed black line shows the lag model of asthma daily admissions in London (2005–2006); The solid black line shows the seasonal model's plots; The straight dashed line represents the historical model; and The solid vertical line (1 January 2006) shows the division between the data on which the models were developed and the data on which the models were cross-validated.
doi:10.1371/journal.pone.0047823.g001

comparing forecast accuracy across series on different scales, because it is a scaled error measure. Hyndman and Koehler, (2006) have also reported that MASE provides the most reliable approach because of its meaningful scale, which is widely applicable and less prone to "degeneracy" problems [32].

Table 1. Measures of fit for the historical, seasonal, and lag models for asthma daily admissions in London, 2005 and 2006.

Error Measure	2005 (Model)	2006 (Forecast)
R^2 Historical	*	*
R^2 Seasonal	0.146	0.235
R^2 Lag	0.366	0.376
RMSE Historical	8.75	9.65
RMSE Seasonal	8.09	8.55
RMSE Lag	6.97	7.57
MASE Historical	1.000	1.150
MASE Seasonal	0.887	0.977
MASE Lag	0.784	0.857

*R^2 values cannot be computed for these models, because there is no variation in the predicted daily admissions.
doi:10.1371/journal.pone.0047823.t001

Furthermore, MASE shows smaller variations, even with small samples, than other measures in the same category and is also known to be less sensitive to outliers [32,38]. The use of MASE as a standard measure of accuracy may therefore enhance the utility of our lagged models in comparing the predictions of asthma daily admissions across various populations.

A comparison of the forecast models within the model development sample (i.e. *Modeled* data), and equally within the test sample (i.e. *Cross-Validation* data) shows various degrees of contrast between the three models we have presented. The observed contrasts between models that are within the same sample frame are useful for benchmarking and selecting the best model to be used in future predictions. These differences are attributed to the constituents (or covariates) of each specific model. On the other hand, it is expected that there are marked differences between the model parameters of the *Modeled* and *Cross-Validation* datasets because, their distributions vary as well. One important issue worth noting and also further investigation is the fact that the lag model predicts asthma daily admissions better during peak periods than moments of low admissions. Further analysis on the relationship between prediction and variations in admission rates is also recommended.

Limitations of study

A major limitation to this approach to forecasting asthma is the data sources and reliability. In this study we anticipated one major limitation could be from the inherent inaccuracies (reliability) of the original data/records. Generally it is assumed that everyone experiencing an asthma exacerbation would be recorded in the database, but conversely, some individuals may seek alternative care and hence go unnoticed. Also, issues of misdiagnoses could be a contributory factor to the data limitations.

In some regards, our choice of treating all cases as unique, including repeat admission cases in the dataset, may be seen as a limitation because of the unique characteristics of such individuals. Nevertheless, from a service provider's perspective, it may make no significant difference.

Implications of the study

This study aims at demonstrating a novel approach to developing an early warning system, which could then be used by health service providers. We however, do not anticipate that results of this current study would be used without circumspect, but hope that the procedure should be validated with larger population datasets and preferably across various populations. If this is done, we can be hopeful that health service providers, individual asthma sufferers and their care providers can be duly informed of when to expect peak and low asthma exacerbations. Such information, which comes as a guide, can enhance health policy decisions and resource allocation, health promotion via anticipatory care/management strategies for asthma and overall minimize the disease burden of the condition.

Conclusions

Uncertainty and chance is an inexorable element of any forecasting system or approach. Nonetheless this study highlights that, detailed and comprehensive retrospective records of asthma daily events can be used in forecasting future events. The study demonstrates that Lag models predict peak asthma admissions better than lower admissions.

All the three error measures (R^2, RMSE and MASE) were consistent in both the modeled data and cross-validation datasets.

The knowledge of the underlying relationships between asthma daily admissions and related lag events that precede the former has provided an underpinning prediction approach of future events. This approach to forecasting does not include other potential predictors that may be known as confounders, and thus minimizes the potential error in predictions associated with their measurement errors. However, important questions that remain unanswered include how such a proposed forecasting model will perform in different settings for different populations, and the precise mechanisms that will be most suitable for modifying the predictors of the respective population data.

Acknowledgments

We acknowledge the support by United Kingdom Meteorological Office's Health Forecast team, who provided the original data used in this analysis. The work here, however, does not reflect the views of the United Kingdom Meteorological Office or any of its staff.

Author Contributions

Conceived and designed the experiments: INS DDR. Performed the experiments: INS. Analyzed the data: INS. Contributed reagents/materials/analysis tools: INS DDR. Wrote the paper: INS DDR.

References

1. Editorial (2012) A fresh perspective on asthma. Nat Med 18: 631 631.
2. Holt PG, Sly PD (2012) Viral infections and atopy in asthma pathogenesis: new rationales for asthma prevention and treatment. Nat Med 18: 726 735.
3. Holgate ST (2012) Innate and adaptive immune responses in asthma. Nat Med 18: 673 683.
4. Rodrigo GJ, Rodrigo C, Hall JB (2004) Acute asthma in adults: a review. Chest 125: 1081 1102.
5. Johnson JA, Peltier-Glaze BM, Bailey-Perry D (2011) Entre Nous. A Critical Appraisal of Science Outcomes of Schooling for African American Children and Youth Enrolled in the Houston Independent School District. ECI Interdisciplinary Journal for Legal and Social Policy 1: 1 28.
6. de Graaff MB, Bröer C (2012) 'We are the canary in a coal mine': Establishing a disease category and a new health risk. Health, Risk & Society 14: 129 147.
7. Pollick M (2012) What Does It Mean to be a "Canary in a Coal Mine"? In: Foster N, editor. wiseGEEK Articles. Sparks, NV: wiseGEEK, Conjecture Corporation, U.S.A.
8. Kyriakis CS, Brown IH, Foni E, Kuntz-Simon G, Maldonado J, et al. (2011) Virological Surveillance and Preliminary Antigenic Characterization of Influenza Viruses in Pigs in Five European Countries from 2006 to 2008. Zoonoses and Public Health 58: 93 101.
9. Nitatpattana N, Le Flohic G, Thongchai P, Nakgoi K, Palaboodeewat S, et al. (2011) Elevated Japanese encephalitis virus activity monitored by domestic sentinel piglets in Thailand. Vector Borne Zoonotic Dis 11: 391 394. Epub 2011 Mar 2011.
10. O'Brien DJ, Kaneene JB, Poppenga RH (1993) The use of mammals as sentinels for human exposure to toxic contaminants in the environment. Environ Health Perspect 99: 351 368.
11. Henning KJ (2004) What is syndromic surveillance? MMWR Morb Mortal Wkly Rep 53: 5 11.
12. Unkel S, Farrington CP, Garthwaite PH, Robertson C, Andrews N (2011) Statistical methods for the prospective detection of infectious disease outbreaks: a review. J R Stat Soc Ser A 175: 49 82.
13. Erbas B, Hyndman RJ (2005) Sensitivity of the estimated air pollution-respiratory admissions relationship to statistical model choice. Int J Environ Health Res 15: 437 448.
14. Hajat S, Haines A, Atkinson RW, Bremner SA, Anderson HR, et al. (2001) Association between Air Pollution and Daily Consultations with General Practitioners for Allergic Rhinitis in London, United Kingdom. Am J Epidemiol 153: 704 714.
15. Hajat S, Haines A, Goubet SA, Atkinson RW, Anderson HR (1999) Association of air pollution with daily GP consultations for asthma and other lower respiratory conditions in London. Thorax 54: 597 605.
16. Peng RD, Dominici F (2008) Statistical Methods For Environmental Epidemiology With R: A Case Study In Air Pollution And Health: Springer-Verlag New York Inc. (United States).
17. Holgate ST, Commins BT, Anderson HR (1995) Asthma and outdoor air pollution. HMSO, London: Department of Health Committee on the Medical Effects of Air Pollutants. Available: http://www.comeap.org.uk/documents/reports.html. Accessed 2010 Nov 1.
18. Bibi H, Nutman A, Shoseyov D, Shalom M, Peled R, et al. (2002) Prediction of emergency department visits for respiratory symptoms using an artificial neural network. Chest 122: 1627 1632.
19. Eisner M, Yegin A, Trzaskoma B (2012) Severity of Asthma Score Predicts Clinical Outcomes in Patients With Moderate to Severe Persistent Asthma. Chest 141: 58 65.
20. Zhang G, Patuwo BE, Hu MY (1998) The state of the art. International Journal of Forecasting 14: 35 62.
21. HES (2008) HES User Guide. Health and Social Care Information Centre. Leeds: Health and Social Care Information Centre.
22. Zorn C (1998) Evaluating Zero-Inflated and Hurdle Poisson Specifications. The Society for Political Methodology ad indicia spectate. pp. Working Papers.
23. Long JS (2001) Predicted probabilities for count models. Stata Journal 1: 51 57.
24. Anonymous (2008) A brief overview of count data commands in Stata.
25. Alt JE, King G, Signorino C (1999) Aggregation Among Binary, Count, and Duration Models: Estimating the Same Quantities from Different Levels of Data. Political Analysis Vol. 9: Pp. 21 44.
26. Mitchell SM, Moore WH (2000) 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. pp. Working Papers.
27. Soyiri IN, Reidpath DD, Sarran C (2011) Asthma length of stay in hospitals in London 2001 2006: demographic, diagnostic and temporal factors. PLoS One 6: e27184.
28. Hilbe JM (2011) Negative binomial regression. Cambridge: Cambridge University Press.
29. Soyiri IN, Reidpath DD (2012) Evolving forecasting classifications and applications in health forecasting. Int J Gen Med 5: 381 389.
30. Barnett AG, Dobson AJ (2009) Analysing Seasonal Health Data. Heidelberg Dordrecht London New York: Springer. 168(139 140) p.
31. Reidpath DD, Diamond MR, Hartel G, Glasziou P (2000) Improving interpretability: γ as an alternative to R2 as a measure of effect size. Statistics in medicine 19: 1295 1302.
32. Hyndman RJ, Koehler AB (2006) Another look at measures of forecast accuracy. International Journal of Forecasting 22: 679 688.

33. R Development Core Team (2011) R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.
34. Stata SE (2009) Release 11. Statistical Software College Station, TX: StataCorp LP.
35. Craigmile P, Kim N, Fernandez S, Bonsu B (2007) Modeling and detection of respiratory-related outbreak signatures. BMC Medical Informatics and Decision Making 7: 28.
36. Singh J, Dutta P (2012) The Temporal Effect of Mobility on Path Length in MANET. International Journal of Wireless Information Networks 19: 38 48.
37. Tikka J, Hollmén J, Lendasse A (2005) Input selection for long-term prediction of time series. Computational Intelligence and Bioinspired Systems 3512/2005: 376 369.
38. Koehler AB, Snyder RD, Keith Ord J, Beaumont A (2012) A study of outliers in the exponential smoothing approach to forecasting. International Journal of Forecasting 28: 477 484.

Chapter 9

9.0 Introduction

This chapter (**chapter 9**) focuses on forecasting peak asthma admissions in London using quantile regression models. Though background literature and information on potential predictors of asthma events were used, the modelling approach was entirely predictive (involving a criterion for the selection of variables), which does not necessarily rely on their biological plausibility. Hence the paper is fundamentally about forecasting, and specifically forecasting the conditional 90th percentile of asthma admissions using quantile regression. We argue in this paper that, forecasting need not rely on good 'causal' models, because good correlation models may do just as well or indeed better. The proof of the forecasting model is its predictive capacity, not its conformance to a particular theory. This then means that, it is not strictly necessary to include any causal factors, as this approach (in this specific case) is data driven. We further acknowledge, with regards to this paper that, data driven approaches have sometimes created disagreements between 'causal modellers' and 'forecast modellers', but both approaches have their roles. And in the empirical forecasting and data mining areas, data driven approaches are generally regarded as superior for the purposes of forecasting and out-of-sample prediction.

Our choice of lags for modelling was therefore data driven, as we explain in the Methods section of the paper. Given this backdrop, emphasis was slightly more placed on the forecasting aspect of the paper, and the use of quantile regression. Even though for integer-valued data, it has been suggested that the data be transitioned from discrete to smoothed densities by “jittering” (52), we found insignificant differences without the *jittering* procedure.

9.1 Forecasting peak asthma admissions in London: an application of quantile regression models

9.1.1 Declarations for Thesis Chapter 9

Monash University

Declaration for Thesis Chapter 9

Declaration by candidate

In the published paper: *Forecasting peak asthma admissions in London: an application of quantile regression models* (In press DOI: 10.1007/s00484-012-0584-0; *Int J Biometeorol.*; Accepted: 26-07-2012) /Chapter 9 of thesis, the nature and extent of my contribution to the work was the following:

Nature of contribution	**Extent of contribution (%)**
Lead and corresponding author, conceptualized the idea, accessed, organized and analyzed data, discussed preliminary results with stakeholders (The Met Office, UK), drafted initial manuscript for circulation, submitted manuscript to journal and then managed correspondence with editors/reviewers, editorial staff and publishers until the final publication	100

The following co-authors contributed to the work. Co-authors who are students at Monash University must also indicate the extent of their contribution in percentage terms:

Name	**Nature of contribution**	**Extent of contribution (%) for student co-authors only**
Daniel D. REIDPATH	Supervised all activities leading to the publication of the paper; critiqued conceptual ideas, analyses and all drafts of manuscript	N/A
Christophe SARRAN	Assisted in accessing data, provided expert interpretation to the data and took part in reviewing and critiquing manuscript prior to submission	N/A

Candidate's Signature	Ireneous N. Soyiri	**Date:** 22-11-2012

Declaration by co-authors

The undersigned hereby certify that:

(1) the above declaration correctly reflects the nature and extent of the candidate's contribution to this work, and the nature of the contribution of each of the co-authors.

(2) they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;

(3) they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;

(4) there are no other authors of the publication according to these criteria;

(5) potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and

(6) the original data are stored at the following location(s) and will be held for at least five years from the date indicated below:

Location(s)	**Global Public Health, School of Medicine & Health Sciences, Monash University Sunway campus**

Signature 1	Daniel D. Reidpath	**Date:** 22-11-2012
Signature 2	Christophe Sarran	**Date:** 14-08-2012

Int J Biometeorol
DOI 10.1007/s00484-012-0584-0

ORIGINAL PAPER

Forecasting peak asthma admissions in London: an application of quantile regression models

Ireneous N. Soyiri · Daniel D. Reidpath · Christophe Sarran

Received: 31 December 2011 / Revised: 26 July 2012 / Accepted: 26 July 2012

Abstract Asthma is a chronic condition of great public health concern globally. The associated morbidity, mortality and healthcare utilisation place an enormous burden on healthcare infrastructure and services. This study demonstrates a multi-stage quantile regression approach to predicting excess demand for health care services in the form of asthma daily admissions in London, using retrospective data from the Hospital Episode Statistics, weather and air quality. Trivariate quantile regression models (QRM) of asthma daily admissions were fitted to a 14-day range of lags of environmental factors, accounting for seasonality in a hold-in sample of the data. Representative lags were pooled to form multivariate predictive models, selected through a systematic backward stepwise reduction approach. Models were cross-validated using a hold-out sample of the data, and their respective root mean square error measures, sensitivity, specificity and predictive values compared. Two of the predictive models were able to detect extreme number of daily asthma admissions at sensitivity levels of 76 % and 62 %, as well as specificities of 66 % and 76 %. Their positive predictive values were slightly higher for the hold-out sample (29 % and 28 %) than for the hold-in model development sample (16 % and 18 %). QRMs can be used in multistage to select suitable variables to forecast extreme asthma events. The associations between asthma and environmental factors, including temperature, ozone and carbon monoxide can be exploited in predicting future events using QRMs.

I. N. Soyiri (✉) · D. D. Reidpath
Monash University,
Kuala Lumpur, Selangor Darul Ehsan, Malaysia
e-mail: soyiriin@yahoo.com

C. Sarran
The Met Office,
Fitzroy Road,
Exeter, UK

Published online: 12 August 2012

Keywords Asthma · Emergency department · Health forecast · Hospital admission · Lag · Predictive model

Introduction

Research in health forecasting is gaining greater attention because of the potential role a reliable health forecast can play in enhancing health service delivery. Health care services are the most important component of any health system, and their functions are more efficient and useful when the related institutions are pre-informed of anticipated excess demand. The World Health Organisation (WHO) reports that effective health service delivery requires some key resources including information, finance, equipment, drugs and well motivated staff (WHO 2010). Given the ever increasing demand for both the coverage and quality of health care services, health service delivery institutions and service providers struggle to tackle situations of excess demand particularly those associated with peak events (Bradley 2005; Derlet 2002). This is because frontline health delivery services and providers are not usually adequately informed and resourced enough to meet the needs of a "higher than normal" demand for health care. Therefore, improving the access, coverage and quality of health services depends on the ways these services are pre-informed, organised and managed. Health forecasting services enable both individuals and service providers to anticipate situations, and hence take the necessary steps to manage peak or extreme events (Hoot et al. 2008, 2009; Jones et al. 2008; Bradley 2005; Soyiri and Reidpath 2012b).

Health forecasting can be conducted through causal (structured) modelling, semi-structured or unstructured (black-box) approaches. There is considerable literature on/related to health forecasting, which is focused on causal modelling (Dominici et al. 2006; Hajat et al. 1999; Babin et al. 2008; Peng et al. 2008; Pascual et al. 2008). Forecasting, however, need not rely on good causal models, because good correlational models will do just as well. The proof of the forecasting model is its predictive capacity, not its conformance to a particular theory. This means that it is not strictly necessary to include any causal factors, and the approach is usually data driven. Data-driven approaches have sometimes created disagreements between causal modelers and forecast modelers, but both approaches have a role to play, and in the areas of empirical forecasting and data mining, data-driven approaches are generally regarded as superior for the purposes of forecasting and out-of-sample prediction (Breiman 2001).

The analytical tools and techniques, like hospital attendance and admissions, that have been involved in predicting and forecasting health events are regression-based methods, which model the conditional mean (Hao and Naiman 2007). Many health forecasting studies that use these techniques fail to address specific health conditions in context, but rather focus on the broader issues such as total hospital attendance or admissions (Champion et al. 2007; Milner 1988, 1997; Sterk and Shryock 1987; Abdel-Aal and Mangoud 2003; Holleman et al. 1996; Farmer and Emami 1990), and quite often assume normality of the data involved. These procedures are, however, limited because: (1) they do not account for outliers in the data; (2) they are unsuitable for heavily skewed data, and (3) they cannot be relied on if there is a need to examine detailed properties of certain important strata of the data (Hao and Naiman 2007; Koenker 2005). Hence, looking beyond the modeling of the conditional mean is particularly useful and applicable to the case of hospital admissions where one might want to focus on unusually high or low numbers of events.

Quantile regression models (QRMs) are a better option for modeling and forecasting peak events, because they are better equipped to characterise the relationship between a response distribution and explanatory variables for selective quantiles (Barbosa 2008; Hao and Naiman 2007; Koenker 2005). Unlike the traditional ordinary least squares method, quantile regressions do not assume a constant effect of the explanatory variables over the entire distribution of the dependent variable.

QRMs have been used extensively in other areas such as econometrics and engineering, to predict extreme events such as price volatility and exchange rates in stock markets (Huang et al. 2011), or to examine the properties of materials that are suited for particular purposes (Young et al. 2008). They have also been applied in some health-related studies to estimate the relationship between socioeconomic determinants and BMI (Pieroni and Salmasi 2010), as well as how access to public infrastructure affects child malnutrition in a developing country setting (Bassolé 2007). However, these and similar studies involving quantile regressions have been focused mostly on explaining the relationship of explanatory factors with respect to quintiles, but not necessarily in the forecasting of peak health events or conditions (Soyiri and Reidpath 2012a). Hence, the aim of this study was to develop predictive QRMs for peak asthma admissions in London, and to further assess the accuracy of selected predictive models using classical forecasting error measures. This study has important implications for health care provision and policies that target conditional distribution of health care services and resources.

Methods

Data

Hospital (asthma) admissions data were sourced from the nationally recorded Hospital Episode Statistics (HES) maintained by the National Health Service, England (HES 2008). The data included an anonymised record of all asthma-related, emergency hospital admissions within London from 1 January 2005 to 31 December 2006 (i.e. 731 days of continuous data).

The operational definition for an asthma admission was any hospital emergency admission with a primary diagnosis of asthma (i.e. an ICD-10 code of J 45). A count of the asthma admissions across all the hospital Emergency Departments within London was recorded for each day of the study period, and this daily count was used as the primary dependent variable in the analyses.

A secondary, binary dependent variable was also created to represent days of peak demand. Usually, a peak event should be defined in collaboration with the relevant stakeholders, taking into consideration the factors that determine the risks of an event (Ebi and Schmier 2005). In the absence of such a known threshold for daily asthma admissions within the London area, a day of peak demand was defined on the basis of a 90th percentile threshold at which the dataset was partitioned naturally for quantile regression modelling (Azuaje 2010). Specifically, a day of peak demand was any day on which the daily admissions count was equal to or exceeded the 90th percentile of daily asthma admissions (i.e. 40 or more asthma admissions). We therefore use the notional definition of "peak events" to refer to the number of asthma admission in the top 90th percentile as explained above.

The corresponding weather data, obtained from the UK Met Office database, was based on averaged daily measurements from the weather monitoring sites across London (Met-Office 2009b). The weather data contained 97 % of complete daily records for the following parameters: ambient air temperature recorded (° C), barometric vapour pressure (hPa) and humidity (%).

Air quality data were based on 24-h averages from air quality monitoring sites across London. The Met Office's Numerical Atmospheric-dispersion Modelling Environment (NAME) was used to generate measures for all corresponding postcodes in the database (Met-Office 2009a). The asthma-associated indicators available with full daily records were carbon monoxide, formaldehyde, nitrogen dioxide, nitrogen oxide, ozone, and particulate matter (specifically PM_{10}). All data were recorded in kilograms per cubic metre but converted to mg/m^3 for carbon monoxide and $\mu g/m^3$ for the other pollutants. All the measured weather and air quality factors examined were identified in previous studies of respiratory- or cardiac-related adverse health events, including asthma (Priftis et al. 2006; Abe et al. 2009; Hajat et al. 1999, 2002; Babin et al. 2008; Peng et al. 2008).

Data analysis and model evaluation

A decision tree was developed and used to generate QRMs of daily asthma admissions based on the temporal, weather and air quality factors (Fig. 1). The predictive validity of the models was compared using the sensitivity and specificity measures for the prediction of peak events.

For the expected total daily asthma admissions, the QRM can be presented in the form below (further illustration of the equation below is also available elsewhere: Hao and Naiman 2007; Koenker 2005):

$$Y_i = \beta_0^{(p)} + \beta_1^{(p)} x_i + \varepsilon_i^{(p)}$$

Where:

Y_i	is asthma hospital admissions for a given day, i
$\beta_0^{(p)}$	is a constant
$\beta_1^{(p)}$	is a coefficient of exposure term
x_i	is the exposure term
$\varepsilon_i^{(p)}$	is the error term
p	is the quantile

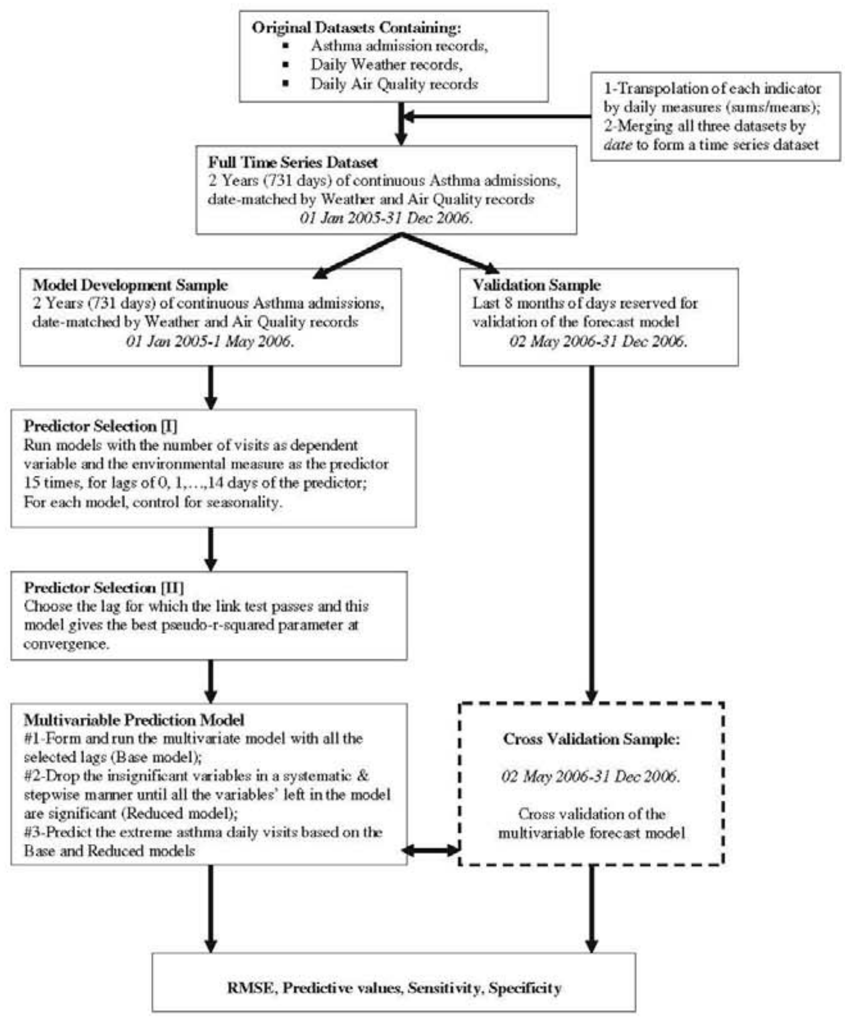


Fig. 1 Decision tree for developing quantile regression model (QRM) forecasting models for peak asthma events using temporal, weather and air quality factors

Trivariate QRMs were developed for the relationships between the daily asthma admissions count and each of the individual weather and air quality factors, controlling for seasonality (as the third variable). "Season" was modelled as a dummy variable with four categories: "spring, summer, autumn and winter". However, since the effect(s) of weather and air quality factors on respiratory health events are usually not instantaneous but rather delayed (Hajat et al. 1999; Sheppard et al. 1999; Arbex et al. 2007), the lagged properties of the predictors were also modelled. The procedure for the selection of suitable lags is described (Predictor selection I and II) in Fig. 1. Only significant lag predictors with the preferred pseudo-R^2 estimate were included in the multivariate models, and then only one lag for each predictor was selected. The range of lags (0, 1,…,14) were explored for a more suitable/optimal time frame for developing early warning messages. The lags found to be suitable were: 3-day lag air temperature; 4-day lag vapour pressure; 6-day lag humidity; 7-day lag ozone; 3-day lag carbon monoxide; 4-day lag nitrogen dioxide; 13-day lag nitrogen oxide; 4-day lag PM_{10}; and 13-day lag formaldehyde.

The pseudo R^2 (comparable to the R^2 for least squares procedures) is the coefficient of determination for QRs and it represents the goodness-of-fit statistic, which is most appropriate for comparing models of specific quantiles (Zietz et al. 2008; Barnes and Hughes 2002). Pseudo R^2 is based on change in the deviance statistic, and ranges between 0 and 1. The pseudo R^2 is thus estimated as:

$$1 - [\text{Sum of deviations about the estimated quantile}/ \\ \text{Sum of deviations about the raw quantile.}]$$

A backward stepwise reduction approach was then used to model weather and air quality effects on the predictive model. This approach involved the systematic elimination of statistically insignificant variables from the overall base model, until a reduced predictive model was achieved. The final reduced model included 3-day lag air temperature, 7-day lag ozone, and 3-day lag carbon monoxide. This multivariate model was used to predict the daily asthma admissions for peak events, and its outputs were then compared with the base model.

Validation and forecasting

Two types of validity were examined. The first was model validity. Model validity represents the extent to which the model fits the data with which the model was developed (i.e. the fit of the model to the hold-in sample). The second type of validity was predictive validity. Predictive validity represents the extent to which the predicted, forecast values fit the observed values (i.e. the fit of the model to the hold-out sample) (Armstrong and Collopy 1992).

Predictive values, sensitivity and specificity tests have been used extensively in many different ways to assess the accuracy of forecasts (Steyerberg et al. 2001; Galant et al. 2004; Sistek et al. 2001). In this study, sensitivity was estimated as a measure of the proportion of peak asthma events that were correctly identified; and specificity was estimated as a measure of the proportion of non-peak asthma events that were correctly identified.

Results

Summary statistics and distribution

The distribution of asthma daily admissions over the 2-year period of the data show two clear peaks, one in 2005 and the other in 2006 (Fig. 2). These peaks occur generally around the spring. Other minor peaks also occur, but these were distributed across all the seasons. Overall, daily asthma admissions in London over the 730 days (whole data sample) had a mean of 28.5±9.9 admissions per day ,and range from 6 to 130 in some peak situations. In the case of the hold-in and hold-out samples, this was, respectively, 27.9±9.4 (10–130) and 29.7±10.7 (6–77).

Summary statistics of the meteorological and air quality predictors used in the analyses are presented in Table 1. There are similarities between the summaries of the selected lags used for modeling (mean standard deviation and spread) and their original measures. Individual variables however, have a wide spread. Even though the summaries for nitrogen oxide, PM_{10} and formaldehyde are fairly similar across all their ranges, notable differences also occur between the whole, hold-in and hold-out data samples.

Table 2 summarizes the parameters of the trivariate analyses, which were used to select the individual environmental predictors for modeling. With the exception of a 4-day lag vapour pressure, a 6-day lag humidity and a 13-day lag formaldehyde, all the selected predictors had a statistically significant ($P<0.05$) association with asthma daily admissions.

Predictive QRM

Based on the design of the study, the predictive quantile regression base model was fitted with three weather related factors (i.e. 3-day lag air temperature, 4-day lag vapour pressure and 6-day lag humidity) and six air quality related factors (7-day lag ozone, 3-day lag carbon monoxide, 4-day lag nitrogen dioxide, 13-day lag nitrogen oxide, 4-day lag PM_{10} and 13-day lag formaldehyde), whilst controlling for the effects of the meteorological seasons. The second model, (i.e. reduced model), was developed from the base model through a systematic stepwise elimination of variables

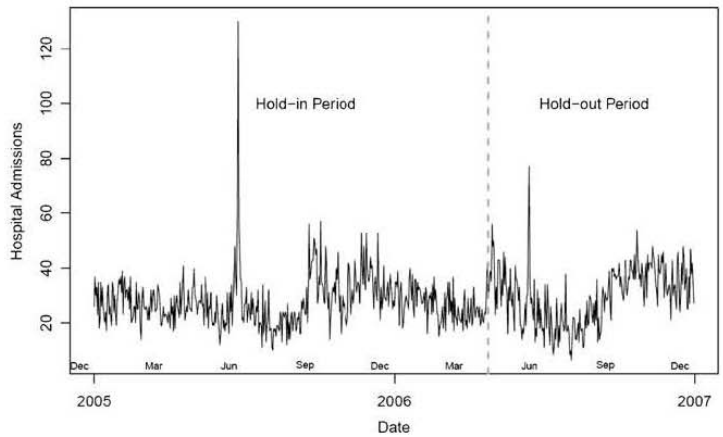


Fig. 2 Asthma admissions in London (2005–2006)

whose *P*-values were most insignificant, and at the same time deliberately controlling for seasonality. This reduced model was fitted with a 3-day lag air temperature; 7-day lag ozone and 3-day lag carbon monoxide. Both the base and reduced models passed the *Link test* model specification (i.e. *P*-value of the "*hatsq*" term >0.05) for the hold-in sample.

Figures 3 and 4 show the scatter plots of the actual daily asthma admissions and the solid green and orange lines showing the predicted asthma admissions for the hold-in and hold-out samples respectively, which are separated by the vertical arrow line (~2 May 2006). The horizontal lines illustrate the grand mean (solid grey line) and peak admissions at 40/day (dashed brown line). Since the model for predicted asthma admissions models the conditional 90th percentile, any datapoint that lies above this predicted line would have been rightly captured.

Sensitivity and specificity of the predictive models

Table 3 summarizes the predictive parameters of the base and reduced models for both hold-in and hold-out samples. The hold-in samples have fewer "true peak admissions" compared to the hold-out samples, and this is reflected in their respective sensitivity estimates of 76 %, 62 % versus 98 %, 96 %. The base model has a lower specificity (66 %) compared to the reduced model (76 %) for the hold-in sample, but a 1 % slightly greater specificity (45 %) in the hold-out sample. The positive predictive values were low; 16 % and 18 % for the hold-in sample and 28 % and 29 % for the hold-out samples.

Discussion

Predicting excess demand for health care services is useful to health care providers, because it enables them to adequately plan and appropriately allocate the resources that will enhance health service delivery (Hoot et al. 2008, 2009; Jones et al. 2008; Bradley 2005; Soyiri and Reidpath 2012a). In this study we designed a mechanism (Fig. 1) for developing predictive forecast models for peak number of asthma daily admissions in London. About 8 % of the daily admissions (between the 1 January 2005 and 2 May 2006) were classified as "peak", i.e. > 40 admissions/day. The base and reduced predictive models were able to detect these days at sensitivity levels of 76 % and 62 %; as well as specificities of 66 % and 76 % respectively. The positive

Table 1 Summary statistics of lags in the hold-in/model development sample. *SD* Standard deviation

Variable	Observed	Mean	SD	Minimum	Maximum
Asthma daily admissions	486	27.8786	9.3478	10	130
3-day lag air temperature (°C)	467	9.6621	6.1667	2.5400	26.4800
4-day lag vapour pressure (hPa)	466	9.8646	3.7482	3.4250	20.6800
6-day lag humidity (%)	464	78.3476	11.9055	35.2000	99.5000
7-day lag ozone (μg/m^3)	479	0.0109	0.0060	0.0008	0.0322
3-day lag carbon monoxide (mg/m^3)	483	0.2542	0.0622	0.1552	0.5227
4-day lag nitrogen dioxide (μg/m^3)	482	0.0219	0.0077	0.0092	0.0524
13-day lag nitrogen oxide (μg/m^3)	473	0.0171	0.0116	0.0025	0.0660
4-day lag PM_{10} (μg/m^3)	482	0.0111	0.0089	0.0017	0.0600
13-day lag formaldehyde (μg/m^3)	473	0.0065	0.0033	0.0017	0.0184

Table 2 Selected lags from a trivariate quantile (0.9) regression analysis of Asthma daily admissions, environmental predictors and seasons. *CI* Confidence interval

Selected individual lags	Coefficient[a]	95 %CI	
3-day lag air temperature (°C)	0.43*	0.0781	0.7825
4-day lag vapour pressure (hPa)	0.35	−0.131	0.8218
6-day lag humidity (%)	0.10	−0.024	0.2211
7-day lag ozone (μg/m³)	−270.28*	−542	1.4062
3-day lag carbon monoxide (mg/m³)	30.16**	6.0797	54.2365
4-day lag nitrogen dioxide (μg/m³)	228.24**	75.493	380.99
13-day lag nitrogen oxide (μg/m³)	116.81*	11.066	222.56
Four day lag PM_{10} (μg/m³)	137*	20.406	253.59
13-day lag formaldehyde (μg/m³)	344.71	−46.14	735.56

*$P<0.05$; **$P<0.01$; ***$P<0.001$

[a]For every one unit change in a predictor variable, the predicated value of asthma admissions will change by the coefficient

predictive values were slightly higher for the hold-out data sample (29 % and 28 %) than for the hold-in/model development sample (16 % and 18 %). Some of the reasons for these observed low predictions may be attributed to earlier observations made about the very wide variations and consistency in the distributions of the individual predictors. Whereas temperature variation over time appears to be consistent, there is less consistency in the variation of ozone and carbon monoxide. Furthermore, measurement of the latter is quite cumbersome, and obtaining area-specific estimates can only be an approximation (Nigam et al. 2010; Setton et al. 2011).

Among the nine variables selected for modeling, six were significantly ($P<0.05$) associated with asthma daily admissions, when controlling for seasonal effects. The variables found to be less significant in the multivariate base model have, however, been associated with asthma and other respiratory illnesses in earlier reports (Hajat et al. 1999, 2002; Babin et al. 2008; Peng et al. 2008). Our inability to use these variables as strong predictors of asthma daily admissions is partly because of the nature of interactions between them, as well as the consistency of their distributions within the dataset. For instance, humidity and barometric vapour pressure are linked independently to asthma exacerbations (Priftis et al. 2006; Abe et al. 2009), and again both are dependent on the seasons. Therefore, for a model that already accounts for meteorological seasons, the effects of humidity and barometric vapour pressure will

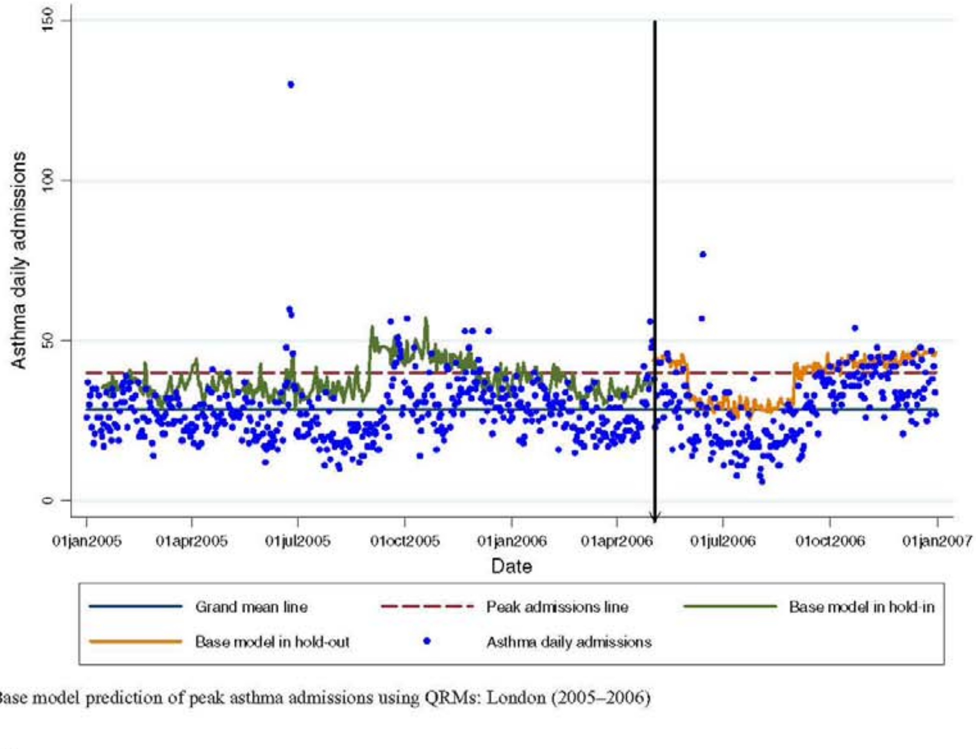


Fig. 3 Base model prediction of peak asthma admissions using QRMs: London (2005–2006)

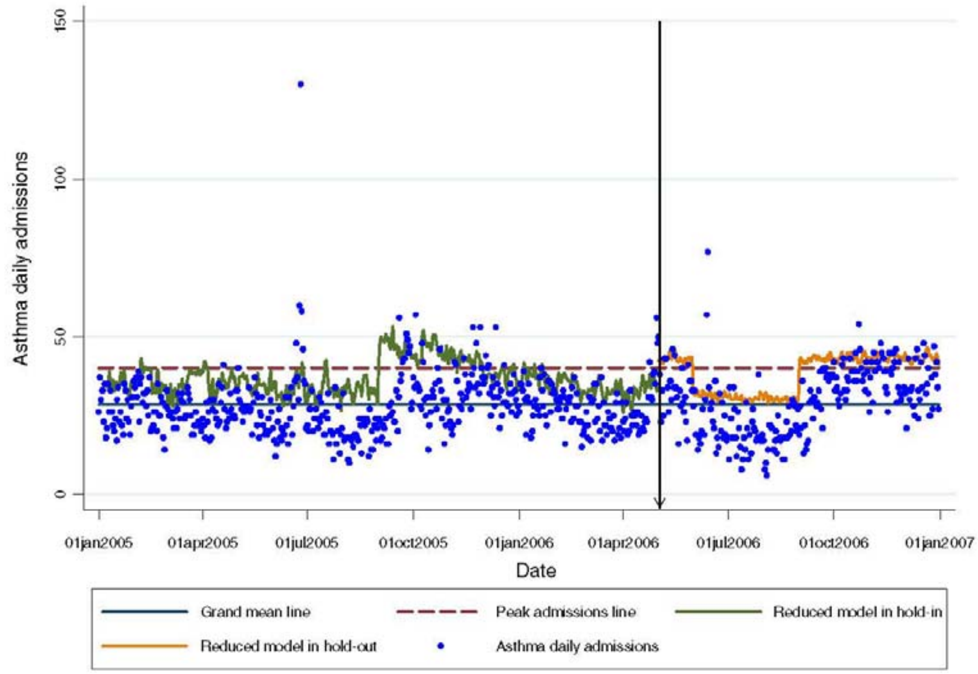


Fig. 4 Reduced model prediction of peak asthma admissions using QRMs: London (2005–2006)

be minimized or affected by collinearity. Similarly, the dominant effect of temperature on many air pollutants (Katsouyanni et al. 1993; Ren et al. 2008) may account for the lesser significance *P*-values observed for air quality measures such as nitrogen, PM_{10}, and formaldehyde.

In the multivariate modeling, the variability of asthma daily admissions could be explained with fewer variables, including a 3-day lag for air temperature; 7-day lag for ozone and a 3-day lag for carbon monoxide. Hence, in the reduced model, all other factors held constant, a one unit rise in, say, a 3-day lag carbon monoxide can result in 27 additional daily admissions ($P<0.05$; 95 % CI: 5.2259–48.8208) see Tables 4 and 5, below. These findings reaffirm the roles of temperature, ozone (Hajat et al. 1999) and carbon monoxide (Sheppard et al. 1999; Hajat et al. 1999) in exacerbating respiratory illnesses like asthma and further show how the same can be used in predicting future peak events.

The literature on forecasting suggests there is no single gold standard approach to forecasting any particular event, but rather recommends a complement of various approaches (Armstrong 2001; Fildes 2006; Armstrong and Collopy 1992). Quantile regressions present an option for predicting

Table 3 Sensitivity and specificity estimates of the base and reduced quantile regression models (QRMs) for the hold-in and hold-out samples

Estimated peak admissions	Base model		Reduced model	
	Hold-in	Hold-out	Hold-in	Hold-out
True non-peak admissions	297	89	341	87
False peak admissions	152	109	108	111
False non-peak admissions	9	1	14	2
True peak admissions	28	45	23	44
Total admissions	486	244	486	244
Prevalence	0.08	0.19	0.08	0.19
Sensitivity	0.76	0.98	0.62	0.96
Specificity	0.66	0.45	0.76	0.44
Positive predictive value	0.16	0.29	0.18	0.28
Negative predictive value	0.97	0.99	0.96	0.98

Table 4 Multivariate QRM base model for asthma daily admissions for hold-in and hold-out samples

Variables in the model	Hold-in sample			Hold-out sample		
	Coefficient[a]	95 %CI		Coefficient[a]	95 %CI	
3-day lag air temperature	0.67**	0.1912	1.1567	0.06	−0.6895	0.809822
4-day lag vapour pressure	−0.45	−1.1443	0.2449	−0.42	−1.14523	0.304593
6-day lag humidity	0.01	−0.1069	0.1270	0.1	−0.10335	0.310107
7-day lag ozone	−48.09	−270.0370	173.8541	8.52	−524.823	541.864
3-day lag carbon monoxide	19.28	−3.7678	42.3359	−5044.17	−40,002.9	29,914.54
4-day lag Nitrogen dioxide	132.14	−72.721	337.009	145.96	−158.75	450.6755
13-day lag nitrogen oxide	265.44	−90.3737	621.2513	41.87	−351.911	435.6489
4-day lag PM_{10}	18.28	−150.29	186.86	−18.19	−361.849	325.4737
13-day lag formaldehyde	−670.80	−2,000.73	659.14	−239.97	−1,587.22	1,107.273
Spring	1.00			1.00		
Summer	−2.84	−6.9188	1.2289	−12.64***	−19.3428	−5.94392
Autumn	9.92***	6.0282	13.8211	−1.75	−9.53913	6.033146
Winter	3.27	−0.2236	6.7542	−0.74	−10.6267	9.142975
The Link test: hatsq *P*-value	0.110			0.715		

$^{*}P<0.05$; $^{**}P<0.01$; $^{***}P<0.001$

[a] For every one unit change in a predictor variable, the predicated value of asthma admissions will change by the coefficient

peak hospital admissions for asthma (Soyiri and Reidpath 2012a, b). In this study, the peak number of daily asthma admissions was notionally defined with respect to the 90th percentile of the distribution. But, in a more practical setting, peak/extreme events would usually be defined by stakeholders, taking into consideration operational issues, as well as related population and demographic factors (Ebi and Schmier 2005). Nonetheless, our definition of a cut-off point allowed us to demonstrate a procedure that could be adapted for different conditions and situations in health forecasting.

Peak numbers of asthma daily admissions are often associated with variability in some environmental factors, which could impact the condition at different levels or thresholds. In this study, we identified a set of nine variables and constituted a multivariate predictive model with these variables (base model). However, to find a more efficient way of predicting the asthma events, the further analysis we conducted yielded a much more reduced predictive model consisting of three key variables. This reduced model, which is simpler, provides comparable and in some cases more competitive estimates to the base model.

The use of lags in predictive modeling presents both challenges and opportunities. Some of these challenges include the reduced sample size of the lag observations compared to the original corresponding data. Others relate to the complexity in choosing an appropriate lag for modeling.

Table 5 Multivariate QRM reduced model for Asthma daily admissions for hold-in and hold-out samples

Variables in the model	Hold-in sample			Hold-out sample		
	Coefficient[a]	95 %CI		Coefficient[a]	95 %CI	
3-day lag air temperature	0.58**	0.2342	0.9281	−0.08	−0.7984	0.6379
7-day lag ozone	−420.69***	−664.6763	−176.6986	167.51	−307.5432	642.5551
3-day lag carbon monoxide	27.02*	5.2260	48.8208	15.26	−18.0621	48.5737
Spring	1.00			1.00		
Summer	−4.08*	−8.1006	−0.0673	−10.83**	−18.0880	−3.5668
Autumn	7.24***	3.6532	10.8229	1.78	−5.2957	8.8487
Winter	1.22	−2.2223	4.6635	0.70	−8.4105	9.8118
The Link test: hatsq *P*-value	0.019			0.470		

$^{*}P<0.05$; $^{**}P<0.01$; $^{***}P<0.001$

[a] For every one unit change in a predictor variable, the predicated value of asthma admissions will change by the coefficient

However, the key advantage of using lagged models is that they detect and provide early warning signals of likely future events. For example, a 3-day lag temperature and carbon monoxide as well as a 7-day lag ozone, is able to predict, at least 3 days in advance, the daily asthma admissions with a positive predictive value of at least 28 %.

Even though our approach to forecasting is not entirely causal, but rather takes the form of a black box prediction (Breiman 2001), the predictors used in the model reaffirms the association between asthma and temperature, ozone, carbon monoxide and the seasons. This relationship is consistent with the literature discussed.

Study limitations

In this report, we acknowledge "asthma admission" itself as a limitation. Even though the study may draw quick attention to asthma in general, the definition of asthma admissions in our data only referred to code J45 of ICD10 that were recorded as primary diagnoses. This implies those admitted with co-infections or multiple conditions including asthma, but for which the latter was not the primary cause for admission, were not captured. It also misses out on the closely associated J46 diagnoses data that is classified as "Status Asmaticus". Another limitation was that we did not adjust for age differences (e.g. between children and adults) in the model; the manifestation of the disease is known to differ between various age groups, but that was not the focus of this study.

The study does not place much emphasis on the plausible association between asthma and the environmental factors used in the predictions. This is because the wide variations and spontaneous distributions of many environmental indicators makes it difficult to arrive at a single representative daily measure for a wide area like London. Hence the estimates used are approximations of the real situations.

The key interest of this paper was to predict peak daily admission of asthma, for which we notionally defined a cut-off point of the 90th percentile. For London as a whole, the study team had no further information that could be used to define a more realistic and applicable cut-off point for the condition. Since seasonality is known to have an influence on asthma exacerbations, it may be argued that the cut-off point should be specific for each season. Unfortunately, our methodology did not account for this, but makes recommendations for such an analysis in subsequent research. Inasmuch as the results of this study should be interpreted with caution, sections of the procedure also need amendment before adoption.

Conclusion

Excess demand for health care services is a great challenge to any health care service provider but the ability to forecast peak events is a promising resource. QRMs can be used as a multistage tool to select suitable variables for predictive modeling of peak daily asthma admissions using environmental factors. A base QRM was fitted with 3-day lag air temperature, 4-day lag vapour pressure; 6-day lag humidity, 7-day lag ozone, 3-day lag carbon monoxide, 4-day lag nitrogen dioxide, 13-day lag nitrogen oxide, 4-day lag PM_{10} and 13-day lag formaldehyde. Also, a second reduced model consisted of 3-day lag air temperature; 7-day lag ozone and 3-day lag carbon monoxide. Both the base and reduced predictive QRMs were able to detect peak number of daily asthma admissions at sensitivity levels of 76 % and 62 %; as well as specificities of 66 % and 76 % respectively. The positive predictive values of the base and reduced models were slightly higher for the hold-out sample (29 % and 28 %) than for the hold-in model development sample (16 % and 18 %). The findings also reaffirm the known associations between asthma and temperature, ozone and carbon monoxide levels.

References

Abdel-Aal R, Mangoud AM (2003) Modeling and forecasting monthly patient volume at a primary health care clinic using univariate time-series analysis. Comput Methods Programs Biomed 56 (3):235–247

Abe T, Tokuda Y, Ohde S, Ishimatsu S, Nakamura T, Birrer RB (2009) The relationship of short-term air pollution and weather to ED visits for asthma in Japan. Am J Emerg Med 27(2):153–159

Arbex MA, Martins LC, de Oliveira RC, Pereira LAA, Arbex FF, Cancado JED, Saldiva PHN, Braga ALF (2007) Air pollution from biomass burning and asthma hospital admissions in a sugar cane plantation area in Brazil. J Epidemiol Community Health 61 (5):395–400. doi:10.1136/jech.2005.044743

Armstrong JS (2001) Principles of forecasting: a handbook for researchers and practitioners. International series in operations research and management science. Kluwer, Dordrecht

Armstrong JS, Collopy F (1992) Error measures for generalizing about forecasting methods: empirical comparisons. Int J Forecast 8 (1):69–80. doi:10.1016/0169-2070(92)90008-w

Azuaje F (2010) Challenges and research directions in bioinformatics and biomarker discovery. In: Azuaje F (ed) Bioinformatics and Biomarker Discovery. Wiley, New York, pp 173–191. doi:10.1002/9780470686423.ch10

Babin S, Burkom H, Holtry R, Tabernero N, Davies-Cole J, Stokes L, DeHaan K, Lee D (2008) Medicaid patient asthma-related acute care visits and their associations with ozone and particulates in Washington, DC, from 1994–2005. Int J Environ Health Res 18 (3):209–221. doi:10.1080/09603120701694091

Barbosa SM (2008) Quantile trends in Baltic sea level. Geophys Res Lett 35(22):L22704. doi:10.1029/2008GL035182

Barnes ML, Hughes AW (2002) A quantile regression analysis of the cross section of stock market returns. EconPapers: Federal Reserve Bank of Boston Working Papers (02-2)

Bassolé L (2007) Child malnutrition in Senegal: does access to public infrastructure really matter? A quantile regression analysis. Paper presented at the African Economic Conference: Opportunities and Challenges of Development for Africa in the Global Arena, Addis Ababa, Ethiopia, Nov. 15–17

Bradley VM (2005) Placing emergency department crowding on the decision agenda. J Emerg Nurs 31(3):247–258

Breiman L (2001) Statistical modeling: the two cultures. Stat Sci 16 (3):199–215

Champion R, Kinsman LD, Lee GA, Masman KA, May EA, Mills TM, Taylor MD, Thomas PR, Williams RJ (2007) Forecasting emergency department presentations. Aust Health Rev 31(1):83–90. doi:10.1071/AH070083

Derlet RW (2002) Overcrowding in emergency departments: increased demand and decreased capacity. Ann Emerg Med 39(4):430–432

Dominici F, Peng RD, Bell ML, Pham L, McDermott A, Zeger SL, Samet JM (2006) Fine Particulate Air Pollution and Hospital Admission for Cardiovascular and Respiratory Diseases. JAMA 295(10):1127–1134. doi:10.1001/jama.295.10.1127

Ebi KL, Schmier JK (2005) A stitch in time: improving public health early warning systems for extreme weather events. Epidemiol Rev 27:115–121

Farmer RDT, Emami J (1990) Models for forecasting bed requirements in the acute sector. J Epidemiol Community Health 44:307–302

Fildes R (2006) The forecasting journals and their contribution to forecasting research: citation analysis and expert opinion. Int J Forecast 22(3):415–432

Galant SP, Crawford LJR, Morphew T, Jones CA, Bassin S (2004) Predictive value of a cross-cultural asthma case-detection tool in an elementary school population. Pediatrics 114(3):e307–e316. doi:10.1542/peds.2003-0575-F

Hajat S, Haines A, Goubet SA, Atkinson RW, Anderson HR (1999) Association of air pollution with daily GP consultations for asthma and other lower respiratory conditions in London. Thorax 54 (7):597–605

Hajat S, Kovats RS, Atkinson RW, Haines A (2002) Impact of hot temperatures on death in London: a time series approach. J Epidemiol Community Health 56(5):367–372

Hao L, Naiman DQ (eds) (2007) Quantile Regression, vol 149. Quantitative Applications in the Social Sciences, vol 149. Sage, Thousand Oaks, 91320

HES (2008) HES user guide. Health and Social Care Information Centre. http://www.hesonline.nhs.uk Accessed 1 January 2009

Holleman DR, Bowling RL, Gathy A (1996) Predicting daily visits to a walk-in clinic and emergency department using calendar and weather data. J Gen Intern Med 11(4):237–239

Hoot NR, LeBlanc LJ, Jones I, Levin SR, Zhou C, Gadd CS, Aronsky D (2008) Forecasting emergency department crowding: a discrete event simulation. Ann Emerg Med 52(2):116–125

Hoot NR, LeBlanc LJ, Jones I, Levin SR, Zhou C, Gadd CS, Aronsky D (2009) Forecasting emergency department crowding: a prospective, real-time evaluation. J Am Med Inform Assoc 16(3):338–345

Huang AY, Peng S-P, Li F, Ke C-J (2011) Volatility forecasting of exchange rate by quantile regression. Int Rev Econ Finance 20 (4):591–606

Jones SS, Thomas A, Evans RS, Welch SJ, Haug PJ, Snow GL (2008) Forecasting daily patient volumes in the emergency department. Acad Emerg Med 15(2):159–170

Katsouyanni K, Pantazopoulou A, Touloumi G, Tselepidaki I, Moustris K, Asimakopoulos D, Poulopoulou G, Trichopoulos D (1993) Evidence for interaction between air pollution and high temperature in the causation of excess mortality. Arch Environ Heal 48 (4):235–242

Koenker R (2005) Quantile regression. Cambridge Press, New York

Met-Office (2009a) NAME III dispersion model. http://www.metoffice.gov.uk/environment/name_iii.html. Accessed 1 May 2009

Met-Office (2009b) Synoptic and climate stations. http://www.metoffice.gov.uk/climate/uk/networks/index.html. Accessed 1 March 2009

Milner PC (1988) Forecasting the demand on accident and emergency departments in health districts in the Trent region. Stat Med 7:1061–1072

Milner PC (1997) Ten-year follow-up of ARIMA forecasts for attendances at accident and emergency department in the Trent region. Stat Med 16:2117–2125

Nigam S, Nigam R, Kulshrestha M, Mittal SK (2010) Carbon monoxide modeling studies: a review. Environ Rev 18:137–158. doi:10.1139/A10-005

Pascual M, Cazelles B, Bouma MJ, Chaves LF, Koelle K (2008) Shifting patterns: malaria dynamics and rainfall variability in an African highland. Proc Biol Sci 275(1631):123–132

Peng RD, Chang HH, Bell ML, McDermott A, Zeger SL, Samet JM, Dominici F (2008) Coarse particulate matter air pollution and hospital admissions for cardiovascular and respiratory diseases among Medicare patients. JAMA: J Am Med Assoc 299 (18):2172–2179

Pieroni L, Salmasi L (2010) Body weight and socio-economic determinants: quantile estimations from the British Household Panel Survey. Munich Personal RePEc Archive—unpublished. Munich University Library, Munich

Priftis KN, Paliatsos AG, Panagiotopoulou-Gartagani P, Tapratzi-Potamianou P, Zachariadi-Xypolita A, Nicolaidou P, Saxoni-Papageorgiou P (2006) Association of weather conditions with childhood admissions for wheezy bronchitis or asthma in Athens. Respiration 73(6):783–790

Ren C, Williams GM, Morawska L, Mengersen K, Tong S (2008) Ozone modifies associations between temperature and cardiovascular mortality: analysis of the NMMAPS data. Occup Environ Med 65(4):255–260

Setton EM, Allen R, Hystad P, Keller CP, Maantay JA, McLafferty S (2011) Outdoor air pollution and health—a review of the contributions of geotechnologies to exposure assessment geospatial analysis of environmental health. In: Gatrell JD, Jensen RR (eds) Geotechnologies and the environment, vol 4. Springer, Netherlands, pp 67–91

Sheppard L, Levy D, Norris G, Larson TV, Koenig JQ (1999) Effects of ambient air pollution on nonelderly asthma hospital admissions in Seattle, Washington, 1987–1994. Epidemiology 10(1):23–30

Sistek D, Tschopp JM, Schindler C, Brutsche M, Ackermann-Liebrich U, Perruchoud AP, Leuenberger P (2001) Clinical diagnosis of current asthma: predictive value of respiratory symptoms in the SAPALDIA study. Eur Respir J 17(2):214–219

Soyiri IN, Reidpath DD (2012a) Evolving forecasting classifications and applications in health forecasting. Int J Gen Med 5(1):381–389. doi:10.2147/IJGMS31079

Soyiri IN, Reidpath DD (2012b) An overview of health forecasting. Environ Health Prev Med. doi:10.1007/s12199-012-0294-6

Sterk WE, Shryock EG (1987) Modern methods improve hospital forecasting. Healthc Financ Manage 41:96–98

Steyerberg EW, Harrell FE, Borsboom GJJM, Eijkemans MJC, Vergouwe Y, Habbema JDF (2001) Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J Clin Epidemiol 54(8):774–781

WHO (2010) Health service delivery. WHO Programmes and Projects: Health Systems. WHO, Geneva

Young TM, Shaffer LB, Guess FM, Bensmail H, Leon RV (2008) A comparison of multiple linear regression and quantile regression for modeling the internal bond of medium density fiberboard. Forest Prod J 58(4):39–48

Zietz J, Zietz E, Sirmans G (2008) Determinants of house prices: a quantile regression approach. J Real Estate Financ Econ 37 (4):317–333

Chapter 10

10.0 Introduction

The study on NYC respiratory related deaths extends the application of QRMs, which we described in the previous chapter (Chapter 10). Both chapters bear similarity in analysis. Therefore the additional motivation for conducting this study for the NYC dataset was to further test the hypothesis of forecasting peak health events using QRMs. It also afforded us the opportunity of using a much larger dataset to replicate the analytical technique we proposed earlier.

This paper presents a relatively novel statistical application of quantile regression to respiratory deaths in New York City observed over a thirteen-year period. We used a hold-in sample (i.e. during the first half of the period) for model development and then applied to the hold-out sample (constituting the second half of the period). The 90th percentile was used in estimating the distribution of deaths instead of estimating the mean. The final model identifies seasonal variables, temperature, CO, and NO2, but excluded O3 and SO2. Data for particulate matter was spotty and hence excluded from modelling. This study is unique for respiratory related deaths and the results suggest there is a potential value of this approach, even when the model is no more sophisticated than a seasonal/temporal model. Health policy and programs may benefit from this study.

10.1 The use of quantile regression to forecast higher than expected respiratory deaths in a daily time series: a study of New York City data 1987-2000

10.1.1 Declarations for Thesis Chapter 10

Monash University

Declaration for Thesis Chapter 10

Declaration by candidate

In this paper designated as Chapter 10, *the use of quantile regression to forecast higher than expected respiratory deaths in a daily time series: a study of New York City data 1987-2000* and currently submitted to the *PLoS One* Journal and, the nature and extent of my contribution to the work was the following:

Nature of contribution	**Extent of contribution (%)**
Lead author, conceptualized the idea, organized and analyzed data, and drafted initial manuscript and contributed to the subsequent writing of the manuscript.	100

The following co-authors contributed to the work. Co-authors who are students at Monash University must also indicate the extent of their contribution in percentage terms:

Name	**Nature of contribution**	**Extent of contribution (%) for student co-authors only**
Daniel D. REIDPATH	Corresponding author, contributed to conceptualizing the idea, accessing the data and writing the paper.	N/A

Candidate's Signature	Ireneous N. Soyiri	**Date:** 22-11-2012

Declaration by co-author

The undersigned hereby certify that:

(1) the above declaration correctly reflects the nature and extent of the candidate's contribution to this work, and the nature of the contribution of each of the co-authors.

(2) they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;

(3) they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
(4) there are no other authors of the publication according to these criteria;
(5) potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and
(6) the original data are stored at the following location(s) and will be held for at least five years from the date indicated below:

Location(s)	**Global Public Health, School of Medicine & Health Sciences, Monash University Sunway campus**

Signature 1	Daniel D. Reidpath	**Date:** 22-11-2012

- -

PLOS ONE

The use of quantile regression to forecast higher than expected respiratory deaths in a daily time series: a study of New York City data 1987-2000

--Manuscript Draft--

Manuscript Number:	
Article Type:	Research Article
Full Title:	The use of quantile regression to forecast higher than expected respiratory deaths in a daily time series: a study of New York City data 1987-2000
Short Title:	Forecasting extreme respiratory deaths in NYC
Corresponding Author:	Ireneous N SOYIRI MONASH University Kuala Lumpur, Selangor D.E. MALAYSIA
Keywords:	Environmental exposure/statistics & numerical data; Epidemiologic Studies; Respiratory Tract Diseases/mortality; Time Factors; Analysis; Regression; Quantile regression; Forecasting; Predict; Respiratory Deaths; New York City; Risk Factors; Time Series; Extreme Events
Abstract:	Forecasting higher than expected numbers of health events provides potentially valuable insights in its own right, and may contribute to health services management and syndromic surveillance. This study investigates the use of quantile regression to predict higher than expected respiratory deaths. Data from 70,830 deaths occurring in New York were used. Temporal, weather and air quality measures were fitted using quantile regression at the 90th-percentile with half the data (in-sample). Four QR models were fitted: an unconditional model predicting the 90th-percentile of deaths (Model 1), a seasonal / temporal (Model 2), a seasonal, temporal plus lags of weather and air quality (Model 3), and a seasonal, temporal model with 7-day moving averages of weather and air quality. Models were cross-validated with the out of sample data. Performance was measured as proportionate reduction in weighted sum of absolute deviations by a conditional, over unconditional models; i.e., the coefficient of determination (R1). The coefficient of determination showed an improvement over the unconditional model between 0.16 and 0.19. The greatest improvement in predictive and forecasting accuracy of daily mortality was associated with the inclusion of seasonal and temporal predictors (Model 2). No gains were made in the predictive models with the addition of weather and air quality predictors (Models 3 and 4). However, forecasting models that included weather and air quality predictors performed slightly better than the seasonal and temporal model alone (i.e., Model 3 > Model 4 > Model 2) This study provided a new approach to predict higher than expected numbers of respiratory related-deaths. The approach while promising has limitations and should be treated at this stage as a proof of concept.
Order of Authors:	Ireneous N SOYIRI
	Daniel D. Reidpath
Suggested Reviewers:	Lane F Burgette, PhD Associate Statistician, RAND Corporation burgette@rand.org
	Sun Yan, PhD National Healthcare Group, Singapore yan_sun@nhg.com.sg
Opposed Reviewers:	

The use of quantile regression to forecast higher than expected respiratory deaths in a daily time series: a study of New York City data 1987-2000.

Ireneous N Soyiri [1,2] Daniel D Reidpath [1,2*]
[1]South East Asia Community Observatory (SEACO), School of Medicine and Health Sciences, Monash University, Malaysia; [2]Global Public Health, School of Medicine and Health Sciences, Monash University, Malaysia

*Corresponding Author:

Prof Daniel D Reidpath, PhD
School of Medicine and Health Sciences
Monash University
Jalan Lagoon Selatan
Bandar Sunway, Selangor
MALAYSIA

*Email: daniel.reidpath@monash.edu

Introduction

There is an increasing body of literature looking at the causal relationship between weather, air quality factors, and health outcomes [1,2,3,4]. Forecasting health outcomes has attracted less attention, but it too has a developing base in the scientific literature [5,6,7,8,9]. Traditionally, both the causal modelling and the forecast research has focused on the central tendencies of the distribution of data; i.e., the expected and conditional expected value. For instance, a typical generalised linear model of daily COPD events will model the expected number of COPD cases each day conditioned on a series of weather, air quality, and perhaps individual factors [10,11].

Although the expected outcomes can be important, the central portion of the conditional distribution is only one part of the story, and other parts of the conditional distribution can give quite different insights – particularly when the distributions are skewed. Studies of birth weight for example have shown quite different relationships between the explanatory variables and birth weight when modelling the conditional mean than they have when modelling low birth weight, such as birth weights in the lowest decile [12]. There is no requirement for the factors explaining low birth weight to be the same factors that explain average birth weight or for the explanation to be of the same form as for the central part of the conditional distribution.

Similarly, in modelling daily respiratory events (morbidity or mortality) and their relationship to air quality or weather, there is no strong requirement for the relationships that model average events to be the same as the relationships that model days with unusually high or unusually low numbers of events. By extension, forecasting the numbers of respiratory events on the outer arms of a conditional distribution need not rely on the same predictors that would be useful in forecasting the expected number of respiratory events.

To our knowledge, and not withstanding its potential value, there has only been one study looking at the forecasting of the number of respiratory events on the outer arm of a conditional distribution (such as the 90^{th} percentile)[13]. There is the simple scientific interest in our capacity to make such forecasts, and what insights it might provide into the data; but there is also potential value for forecasting likely resource needs, as well as in areas such as syndromic surveillance, where the number of events exceeding a threshold is used to trigger a health systems response. Quantile regression remains a relatively unusual modelling technique in health research, which can be used to model conditional responses at any quantile of interest; and – although it has been used (rarely) for forecasting [14,15] – to our knowledge has never been used to forecast mortality.

Methods

We investigate the use of quantile regression to forecast the 90^{th} percentile of daily, respiratory related deaths for New York City, in the period 1 January 1987 to 31 December 2000. The choice of the 90^{th} percentile was somewhat arbitrary but in keeping with the idea of understanding the general capacity that a health system might need to maintain to meet typical demand. The data were drawn from the National Morbidity, Mortality, and Air Pollution Study (NMMAPS)[16], which are publicly available data through the Health and Air Pollution Surveillance System website (http://www.ihapss.jhsph.edu), and, in our case, accessed using the NMMAPS package in the R statistical environment [17]. The daily count of respiratory deaths was the outcome measure of interest. The data included 70,830 respiratory deaths over 5,114 days of surveillance.

The data set also included a range of daily weather and air quality measures which were used as predictors in the modelling. The predictors included daily mean air temperature, dew point, ozone

(O_3), sulphur dioxide (SO_2), nitrogen dioxide (NO_2), and carbon monoxide (CO). Measures of particulate matter were not included because of the levels of missing data within the dataset. In addition to the measures of weather and air-quality, cosinor values representing a yearly and a half yearly cycle[18,19], and dummy variables representing the days of the week were also used as predictors.

The data were sub-divided into two equal sized sets, from 1 January 1987 to 31 December 1993 for model development (*in-sample*), and from 1 January 1994 to 31 December 2000 for cross validation (*out-sample*). The size of the *in-sample* data was subsequently reduced to 2405 days (94.0% of total days) because of the use of lagged data, and a small amount of pre-existing missing data. The *out-sample* data were almost complete with a loss of only 8 days of data (i.e., 2548 days).

The details of quantile regression have been described elsewhere[20,21], as has its application to health problems[22,23]. The use of quantile regression with count data is unusual and its application to health forecasting remains novel [24,25].

A common challenge in modelling outcomes related to environmental exposures is the lagged effect between weather and air quality exposures and the health outcome of interest [26]. To identify an appropriate lag to represent the exposure to each of the weather and air quality measures, a series of quantile regression count models using the *in-sample* data were constructed testing the fit for each lag in turn, from a 1-day lag through to a 7-day lag; and was similar to approaches used elsewhere [9]. The fit of each lag, for each weather and air quality measure was assessed using a function based on the asymmetric Laplace distribution commonly used in quantile regression. Best fit was determined by the lag that had the lowest value for:

$$\left[(1-p).\sum_{y_i<q}|y_i-q|\right]+\left[p.\sum_{y_i>q}|y_i-q|\right], \qquad (1)$$

where the absolute deviations below quantile q are weighted by $1-p$ if the actual values lies below q, and p if the actual value lies above q. The lags that were identified for inclusion were: CO, NO_2, O_3, dew point (1-day), temperature and SO_2 (3-days). A 7-day moving average value for each of the weather and air quality factors was also included in one model as a point of contrast.

Four separate models were subsequently developed, three of which used quantile regression with either the selected lags or 7-day moving averages as predictors with the *in-sample* data. Model 1 was the intercept only model, an unconditional model predicting the value of the 90^{th} percentile of daily respiratory deaths to be constant across the data. Model 2 was a conditional model in which the value of 90^{th} percentile of daily respiratory deaths varied, conditioned on seasonal (cosinor values) and temporal (day of the week) predictors. Model 3 was a conditional model in which the value of 90^{th} percentile of daily respiratory deaths varied conditioned on seasonal/temporal predictors and the selected lags of weather and air quality predictors. Model 4 was the same as Model 3 except that the 7-day moving averages of weather and air quality predictors were used instead of selected lags.

The parameter estimates and standard errors from the quantile regression are not reported, because they are essentially meaningless given the process for developing the forecasting model. Previous experience suggests that when they are provided, attention is inappropriately placed on that, rather than the predictive and forecasting capacity of the models – "the proof of the pudding is in the eating" not in knowing the ingredients.

The measure of fit used to establish the predictive validity (*in-sample* fit) and the forecasting accuracy (*out-sample* fit) of the quantile regression models was the coefficient of determination (R1)[20]. R1 measures the proportionate reduction in the weighted sum of the absolute deviations (WSADs) achieved by a conditional model over the unconditional model; where the weighted sum of the absolute deviations is given by eqn. 1. In this context, R1 is analogous to the mean absolute scaled error suggested by Hyndman and Koehler [27]. R1 was estimated for Models 2, 3, and 4, using the weighted sum of absolute deviations from Model 1 as the denominator.

Results

The time-series graph of the daily, respiratory deaths shows the familiar annual cycle with the peak deaths occurring in the winter months and the valleys occurring in the summer (Figure 1). The dashed vertical line in the middle of the figure shows the separation between the *in-sample* used to develop the quantile regression models and the *out-sample*, used to cross-validate the forecast models. The dotted horizontal line shows the *in-sample*, unconditional, 90^{th} percentile number of daily deaths (20.2 per day). It is clear that a conditional distribution at the 90^{th} percentile which included seasonal/temporal predictors would be quite different from the straight line of the unconditional quantile.

The top half of Table 1 shows the predictive capacity of the three conditional models (Models 2, 3, and 4) relative to the unconditional model (Model 1). The coefficient of determination (R1) showed improvements between .165 (i.e., a 16.5% improvement in the fit) and .191 over the unconditional model. The seasonal model (Model 2) slightly outperformed both the selected lags model and the 7-day moving average model.

The models developed using the *in-sample* data were cross-validated using the *out-sample* data.

The lower half of Table 1 shows the forecasting capacity of the three conditional models (Models 2, 3, and 4) relative to the unconditional model (Model 1). As anticipated, the forecasting performance of Model 2 (seasonal/temporal predictors) was slightly worse than the predictive performance (R1=.161); i.e., cross-validation against the *out-sample* was slightly weaker than predictions based on the model development data (*in-sample*). Model 3 (seasonal/temporal predictors and selected lags for weather and air quality) performed slightly better in forecasting than it had in prediction (R1=.194). The performance of the 7-day moving average model (Model 4) also improved slightly (R1=.187).

As an illustration, Figure 2 shows the predictions (*in-sample*) and forecasts (*out-sample*) based on Model 3 (seasonal, temporal and selected lags as predictors) against the unconditional Model 1. Each point (small dot, larger dot, and solid triangle) represents the number of respiratory deaths recorded for a particular day. The conditional quantile regression model for the 90th percentile (shown in grey) lies, as expected, well above the central portion of the data – where a model of the conditional means or median would lie – and shows clear seasonal variation. The dotted horizontal line shows the unconditional 90th percentile (Model 1). A few matters are worthy of note. All the points that lie above the dashed horizontal line would be regarded, under the unconditional model, as reflecting more unusual numbers of respiratory deaths. The points shown as black triangles reflect those days with numbers of deaths identified as more unusual under the unconditional model, but more typical under the more complex, conditional model. Conversely, all the points that lie below the dashed line would be regarded as more typical under the unconditional. The points shown as larger dots reflect those days with numbers of deaths identified as more typical under the unconditional model, but more unusual under the more complex, conditional model. Visually, the predictive and forecast performance of the model appears to be reasonably consistent (Figure 2).

Discussion

The notion of modelling and forecasting the expected number of daily deaths is well described in the literature [2,28,29,30]. Forecasting any health outcome on the outer arms of a conditional distribution, however, is unusual [15], and appears not to have been done in the analysis of daily time series data related to mortality. This is unfortunate, because there are things to be learned from forecasts made at, for instance, the 90^{th} percentile that could not be learned from forecasting the expected number of daily deaths.

For example, forecasts of the expected number of deaths will underestimate the kinds of resources that need to be available much of the time, particularly in an environment with the kind of variability shown in the daily respiratory deaths data. There is cyclical variation in the data, but even within the data at any one part of the annual cycle, there is substantial daily variation. Forecasts of likely numbers of deaths (i.e., occurring 90% of the time), can also feed into a mechanism for identifying when there is a concerning deviation in the number of deaths. Sustained numbers of days with deaths above the forecast can inform a health system about the occurrence of a likely environmental exposure or emerging disease.

Furthermore, because the forecasts are conditional, relatively low absolute numbers of deaths occurring in the summer, can still trigger a response when those numbers (although low) fall consistently above the conditional 90^{th} percentile. They also forecast when resources and capacity may be reduced.

The analysis presented here showed some forecasting benefit associated with the inclusion of selected lags of daily weather or air quality data (i.e., a difference in R1 between .161 and .194 – a 3% improvement over the unconditional model). A trade-off arises, however, between developing more complex conditional models over models including only temporal and seasonal predictors.

There are important limitations with the approach taken here, and these can be used to highlight future pathways for analysis. The first limitation is with the use of the 90th percentile. One can potentially analyse the data at any conditional quantile, and for different purposes (such as surveillance or resource allocation) analyses at different quantiles – or multiple quantiles – may be more useful. The utility is driven by the application, and as we were seeking a proof of concept, the 90th percentile seemed to be appropriate level. Using cosinor values of yearly and half yearly cycles may not capture important seasonal information that could be built into the forecasts, and is certainly worthy of future investigation. There is a balance to be made in forecasting between the gain in accuracy and the cost of implementation. Sinusoidal functions capturing seasonal and temporal variation are trivial to develop and implement, and provide around a 15-20% improvement in accuracy over using an annual figure for the 90th percentile. More complicated conditional models appear to add a 2-3% improvement. The utility of the gain for the effort is uncertain. The final limitation we consider here is a theoretical one. There is often concern expressed with forecasting models that do not take a more traditional causal modelling approach [31]. We would take two distinct lines of argument in response. The first line of response is that the purpose of forecasting is not about determining cause and effect, and therefore forecasting models should be judged according to their forecasting accuracy, not for their inadequacy at providing causal explanations. The second line of response is that if a causal model out-performs and non-causal model in forecast accuracy, then the causal model should absolutely replace the non-causal model. The causal model was not developed here, but there is some reason to believe that it may not perform as well as a "dust bowl" empirical approach [31].

Conclusion

This study reports for the first time, a statistical approach for forecasting respiratory related deaths at the 90th percentile using quantile regressions. The results suggest there is potential value in this, even when the model is no more sophisticated than a seasonal/temporal model. The study should, however, be treated as a proof of concept, rather than definitive.

References

1. Cheng Y, Kan H (2012) Effect of the Interaction Between Outdoor Air Pollution and Extreme Temperature on Daily Mortality in Shanghai, China. Journal of Epidemiology 22: 28-36.
2. Metzger KB, Ito K, Matte TD (2009) Summer heat and mortality in New York city: How hot is too hot? Environ Health Perspect 118.
3. Rajarathnam U, Sehgal M, Nairy S, Patnayak RC, Chhabra SK, et al. (2011) Time-series study on air pollution and mortality in Delhi. Res Rep Health Eff Inst 157: 47-74.
4. Zibners LM, Bonsu BK, Hayes JR, Cohen DM (2006) Local weather effects on emergency department visits: a time series and regression analysis. Pediatr Emerg Care 22: 104-106.
5. Chatterjee C, Sarkar RR (2009) Multi-step polynomial regression method to model and forecast malaria incidence. PLoS ONE 4: 6.
6. Gharbi M, Quenel P, Gustave J, Cassadou S, Ruche G, et al. (2011) Time series analysis of dengue incidence in Guadeloupe, French West Indies: Forecasting models using climate variables as predictors. BMC Infectious Diseases 11: 166.
7. Kam HJ, Sung JO, Park RW (2010) Prediction of Daily Patient Numbers for a Regional Emergency Medical Center using Time Series Analysis. Healthc Inform Res 16: 158-165.
8. Soyiri IN, Reidpath DD (2012) Semi-structured black-box prediction: proposed approach for asthma admissions in London. *Int J Gen Med* 5: 693-705.
9. Soyiri IN, Reidpath DD (2012) Humans as Animal Sentinels for Forecasting Asthma Events: Helping Health Services Become More Responsive. PLoS ONE 7: e47823.
10. Dominici F, Peng RD, Bell ML, Pham L, McDermott A, et al. (2006) Fine Particulate Air Pollution and Hospital Admission for Cardiovascular and Respiratory Diseases. JAMA: Journal of the American Medical Association 295: 1127-1134.
11. Stieb D, Szyszkowicz M, Rowe B, Leech J (2009) Air pollution and emergency department visits for cardiac and respiratory conditions: a multi-city time-series analysis. Environmental Health 8: 25.
12. Chen Y, Yang Q, Krewski D, Shi Y, Burnett RT, et al. (2004) Influence of relatively low level of particulate ar pollution on hospitalization for COPD in elderly people. Inhal Toxicol 16: 21-25.
13. Soyiri IN, Reidpath DD, Sarran C (2012) Forecasting peak asthma admissions in London: an application of quantile regression models. *Int J Biometeorol*: 1-10.
14. Burgette LF, Reiter JP, Miranda ML (2011) Exploratory quantile regression with many covariates: an application to adverse birth outcomes. Epidemiology 22: 859-866.
15. Sun Y, Teow KL, Heng BH, Ooi CK, Tay SY (2012) Real-Time Prediction of Waiting Time in the Emergency Department, Using Quantile Regression. Annals of emergency medicine 60: 299-308.
16. Dominici F, McDermott A, Daniels M, Zeger SL, Samet JM (2005) Revised analyses of the National Morbidity, Mortality, and Air Pollution Study: mortality among residents of 90 cities. J Toxicol Environ Health A 68: 1071-1092.
17. Peng RD, Welty LJ (2004) The "NMMAPSdata" Package. R News 4: 10-14.
18. Barnett AG, Dobson AJ (2010) Analysing Seasonal Health Data. Berlin: Springer-Verlag. 168(139-140) p.
19. Stolwijk AM, Straatman H, Zielhuis GA (1999) Studying seasonality by using sine and cosine functions in regression analysis. J Epidemiol Community Health 53: 235-238.
20. Hao L, Naiman DQ, editors (2007) Quantile Regression. Thousand Oaks, California 91320: Sage Publications, Inc. 139 p.
21. Koenker R, Xiao Z (2006) Quantile Autoregression. Journal of the American Statistical Association 101: 980-990.
22. Beyerlein A, Fahrmeir L, Mansmann U, Toschke AM (2008) Alternative regression models to

assess increase in childhood BMI. BMC Med Res Methodol 8: 59.
23. Beyerlein A, Nehring I, Rzehak P, Heinrich J, Müller MJ, et al. (2012) Gestational Weight Gain and Body Mass Index in Children: Results from Three German Cohort Studies. PLoS ONE 7: e33205.
24. Machado JAF, Silva JMCS (2005) Quantiles for Counts. Journal of the American Statistical Association 100: 1226-1237.
25. Winkelmann R (2006) Reforming health care: evidence from quantile regressions for counts. J Health Econ 25: 131-145.
26. Zeger SL, Irizarry R, Peng RD (2006) On Time Series Analysis of Public Health and Biomedical Data. Annual Review of Public Health 27: 57-79.
27. Hyndman RJ, Koehler AB (2006) Another look at measures of forecast accuracy. International Journal of Forecasting 22: 679-688.
28. Peng RD, Dominici F, Pastor-Barriuso R, Zeger SL, Samet JM (2005) Seasonal Analyses of Air Pollution and Mortality in 100 US Cities. American Journal of Epidemiology 161: 585-594.
29. Tao Y, Huang W, Huang X, Zhong L, Lu SE, et al. (2012) Estimated acute effects of ambient ozone and nitrogen dioxide on mortality in the Pearl River Delta of southern China. Environ Health Perspect 120: 393-398.
30. Yang C, Yang H, Guo S, Wang Z, Xu X, et al. (2012) Alternative ozone metrics and daily mortality in Suzhou: The China Air Pollution and Health Effects Study (CAPES). Science of The Total Environment 426: 83-89.
31. Breiman L (2001) Statistical Modeling: The Two Cultures. Statistical Science 16: 199-215.

Table 1: A comparison of the predictive and forecasting capacity of the models

	Parameter	Model 1	Model 2	Model 3	Model 4
Prediction					
In sample	WSAD	2328.2	1882.5	1943.3	1927.5
	R1	0	.191	.165	.172
Forecasting					
Out sample	WSAD	2529.5	2121.3	2039.7	2055.3
	R1	0	.161	.194	.187

Perdition was based on the In Sample (days=2445) and forecasting was based on cross-validation of the Out Sample (days=2548): Model 1, intercept only; Model 2, temporal/seasonal model, Model 3, temporal/seasonal model with selected lags of weather and air quality; and Model 4 temporal/seasonal model with a 7-day moving average of weather and air quality. The weighted sum of the absolute deviation (WSAD) and the coefficient of determination (R1) are used to compare the models.

Figure Legends

Figure 1: Time series of respiratory related deaths 1987—2000

The vertical dashed line indicates the separation between the in-sample and out-of sample data.

Figure 2: The quantile regression (90^{th} percentile) model of respiratory related deaths

The small dots indicate daily deaths

The dotted horizontal line shows the unconditional 90^{th} percentile

Figure
Click here to download Figure: Figure1-QRM forecast_insddr.doc

Figure 1: Time series of respiratory related deaths 1987—2000

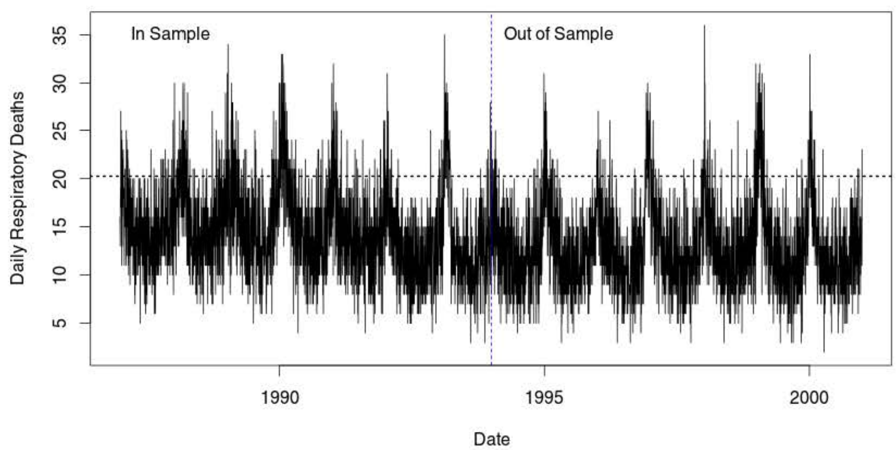
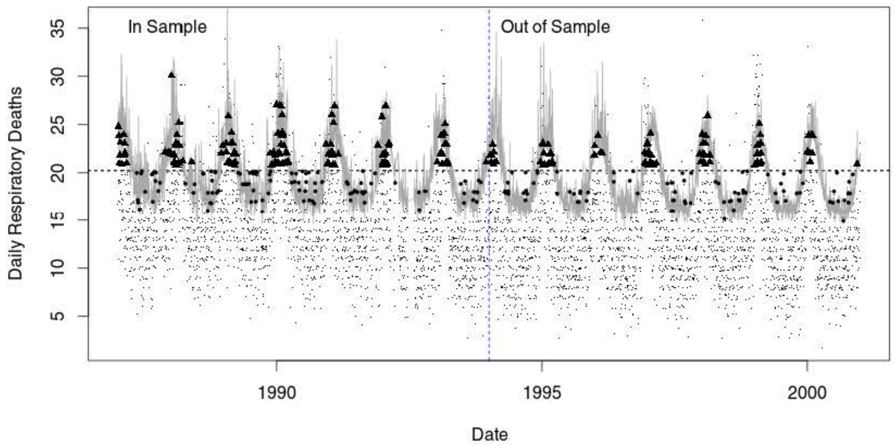


Figure
Click here to download Figure: Figure2-QRM forecast_insddr.doc

Figure 2: The quantile regression (90th percentile) model of respiratory related deaths



SECTION V

Chapter 11

11.0 General Discussion

Health forecasting predicts health situations or disease episodes and forewarns about future events (39). It also forms part of the activities of preventive medicine or preventive care engaged in public health planning, and it is aimed at facilitating health care service provision in populations (39, 54, 215, 217-219). Health forecasting is an age-old concept, practiced in various forms of health delivery over the years (220-222); but is also a relatively new and re-emerging terminology that is finding extended modern applications in the field of health. It has recently been championed by a number of organizations including the United Kingdom Meteorological Office (217) in collaboration with Medixine (COPD Health Forecasting) (223) and the Health Forecasting Project of UCLA School of Public Health (224) for the purposes of providing health forecasting services to health service providers like the NHS as well as to subscribed individuals. Health forecasting derives most of its current analytic approaches and techniques from other areas including econometrics, marketing, and meteorology. In this research we reviewed the literature on health forecasting and chronic respiratory illness (asthma), and identified statistical methods /techniques used in predictions. We also explored weather, air quality and hospital administrative datasets with the goal of understanding the data and its distributions, trends and nature of the relationships between the variables reported therein. It was hoped that this exercise would facilitate the design and development of methods for health forecasting.

Mainly structural (causal) models have been used to model and predict various health situations /conditions, but rather only few reports on the unstructured models. We also noted that the complexity of reported health forecasting models vary from individual /aggregated health situations or disease type being forecast, to the strategies involved in selection of variables/predictors, and through to the nature of the model diagnostic tools.

There are many reports on asthma and its impacts on various population demographics worldwide, and the disease burden has been described in many ways, as we reported in some of our reviews (Chapters 2, 3 & 4). In Chapter 3, we specifically investigated an aspect of the disease burden (LOS) in order to justify the importance of forecasting asthma, but also to show a unique statistical approach to investigating LOS. The relationships between asthma admissions and factors which partly define the burden of disease related to LOS (i.e. demographic, diagnostic and temporal factors (105)) were examined. Our approach, which used a fixed effect model (comparing individual versus area effects) was useful in categorizing the determinants and burden of LOS. Previous studies on LOS and even more recent ones, which reported on the factors associated with LOS, have failed to account for the fixed effects such as individual, area of residence or other constant factors that may influence the LOS in a unique manner (225, 226). This pragmatic approach to investigating LOS may therefore be of importance to health services provision.

Having tested a tool for estimating the disease burden, we focused our attention on the health forecasting literature in order to examine the key information that could support the development of health forecasting models. Reports on health forecasting often used forecasting terminologies and principles, which have been created for forecasting in other areas (227). These terms and principles have however not been well described in perspective of health forecasting. Since the interpretation of health data and information may vary extensively compared to say econometric or marketing data, streamlining health forecasting terminologies and principles is important. Even though we made attempts to define /describe some of the terms in this thesis (Chapters 4, 5 & 6), it is not sufficient and further research may need to focus on defining the 'fine boundaries' of terms and principles mainly applicable to health forecasting. Related paper in Chapters 5 & 6 for instance may stimulate further discourse and research on health forecasting issues which relate to how we define and handle uncertainty/risks, error, accuracy, the focus of a health forecast, data aggregation and how to define horizons for health forecasting. It would be interesting to see how future research on health forecasting address and discuss some of these issues identified.

Another important contribution of Chapter 5 is the proposed schematic approach to health forecasting. The scheme/outline also lists potential hurdles to overcome in health forecasting. Even though the proposed approach may not be entirely flawless, it was designed to capture the knowledge and information that is plausible and potentially relevant in health forecasting. We have not yet noted any similar proposed scheme in the health forecasting literature, and so it may be early to want to make comparisons. Nonetheless, it is intended to draw the appeal of public health practitioners who may be interested in developing a pragmatic tool for health forecasting.

A major component of our study was investigating the potential utility of a number of forecasting models and techniques on asthma daily admissions in London, and also respiratory related deaths in NYC. As a result, we developed statistical models (negative binomial regressions or quantile regression models), which were best suited for the datasets accessible to us. In the analyses, we examined the extent to which temporal, weather and air quality factors could help predict anticipated asthma daily admissions in London (chapters 6, 7 & 8), as well as extreme/peak health events in general (chapters 9 &10). The health forecasting methods and techniques discussed in this thesis gave reasonably good predictions based on the relative measures of model diagnostic parameters (measure of degree of error). The systematic approaches we described (in chapters 6 - 10) may be useful for standardizing model parameters in further research on health forecasting. They may also be useful in prioritizing indicators for future health and non-health data collection.

Generally, we observed a consistency between the original distribution of the dependant variable (asthma daily admissions) and that of its predicted form over time. However, as one would expect, there were significant differences between the hold-in and hold-out data sets, which was primarily attributable to the differences in their original distribution and characteristics. As we pointed out in Chapter 6, there is not as yet any clear criteria for deciding on exactly what proportion of data, in relation to the hold-in sample should be used as hold-out for model validation (39). The substantial variations of approaches used in validating various health forecasting models calls for further investigation to help define appropriate approaches for validating health forecasts

Earlier health forecasting models of asthma focused on the premise that increased exposure to the vagaries of the weather and to the toxicity effects of other environmental factors would cause exacerbations. Hence models were based on the well known causal factors like ozone, atmospheric temperature, just to mention a few, which are consistent with the literature (228, 229). Even though this is a reasonable assumption to base the development of a forecasting model on, there could be some problems with the exact representation of individual's exposure parameters, which may then subsequently affect the forecast model's reliability. However, as we argued in our analytical papers (Chapters 6 - 10), in the area of forecasting unlike causal modelling, data-driven approaches have proven to be better and more reliable in predicting future events (230).

Our analyses focused on how to best develop a reliable and parsimonious predictive model using the needed predictors – a data-driven approach. It is important to reemphasize that not all the plausible predictors were necessary for inclusion in the best predictive model. An air pollutant such as carbon monoxide may be well known for its role in exacerbating asthma, however, in some context of population health analyses involving a wide area; this indicator may be redundant in a multivariable predictive model. This is not unexpected, since population level exposures for wide areas like

London is difficult to generalize for individual exposures, and more particularly for asthma sufferers whose outdoor activities are largely unrecorded.

Since individuals would usually spend some time indoors and unspecified periods outdoors, where there is a wide variation in weather elements and air quality factors, it becomes imperative that potential health forecasting models for chronic respiratory related diseases such as asthma to consider other forms of data and novel forecasting approaches which will provide a better forecast. Hence our preference for non-causal or semi-structured forecasting models as discussed in chapters 4 & 6.

We were able to show how temporal and environmental factors could be used to develop models for forecasting both routine/anticipated health events and also extreme/peak events. Our inquiry into health forecasting also shed light on the extent to which predictions could vary based on differences in underlying models and their component predictors. Overall, the investigations suggest that a range of environmental factors and statistical tools can be used to develop health forecast for chronic respiratory conditions like asthma. But, there is no clear evidence that data on environmental factors can yield more effective predictions than ordinary temporal factors or univariate lagged models, particularly in targeting vulnerable individuals in large heterogeneous urban dwellings such as the city of London. Nonetheless the epidemiological evidence base for health conditions that are forecast needs to be regularly updated to supplement the strategies used by data-driven approaches to health forecasting. On the basis of our series of investigations, there may be relevant implications for further research on health forecasting, health care services and for policy. A few are mentioned here for reemphasis.

Clinical implications of health forecasting

Identifying the number of individuals who fall sick of asthma within a defined area or jurisdiction over a time period could help predict future events and hence help clinicians to make better decisions. It could also help health managers to make better decisions and resource allocations. Our study identified that, temporal factors such as day of the week and seasons as well as the lag asthma admissions such as the previous day(s) events, were strong predictors of future events. This therefore implies that information on the accurate diagnosis of a case, as at, and when it happens is useful for forecasting future events.

Though asthma is a well known chronic respiratory condition which is largely affected by the vagaries of environment (i.e. extreme weather conditions and poor air quality), obtaining representative human exposure records or estimates is hard to come by (231). This is because environmental conditions vary very widely across large populations and also between indoors and outdoors. Hence clinical records of reported individuals, over a period of time may be the most useful in forecasting.

Related to data /sustainable data management and public health surveillance

Lessons have been learnt in trying to use surveillance health records and other data to guide health care services delivery. The value of health forecasting in facilitating health care delivery and reducing overall health care expenditure, is however not well argued out in a compelling manner to ensure that it is widely adopted /used. Further research on the cost-saving benefits of health forecasting with examples on its return on investment may be a useful point to draw the attention and support of policy makers in health forecasting research

Health forecasting depends on the availability and access to reliable longitudinal data, which can be easily provided if national health systems set up and maintain sustainable public health surveillance systems

Related to methods and techniques

In the analytic sections of our research, we have used a number of statistical methods and techniques to develop health forecasting models, which can be validated with other population data and adapted to specific cases of interest.

Related work force development

Health forecasting can be realized with an investment in additional workforce development or outsourcing to specialized organizations that are professionally engaged in the activity. The UK Met Office Health Forecasting team is one of the examples which have been mentioned already in the thesis.

Chapter 12

12.0 Conclusions, Recommendations and Advances

12.1 Conclusions

Asthma poses a significant burden to populations, and health forecasting may provide an alternative solution to help in managing the condition at various levels (individual, care providers, service providers, and policy). In defining and establishing the disease burden of asthma, we found that the individuals' fixed effect negative binomial model was more robust in explaining the determinants of LOS, compared to the area effects model. We leave open the question; however whether the combination would be even more successful.

In our reviews we established that environmental factors have causal links to chronic respiratory diseases like asthma, and these effects often vary by location. However, our empirical studies showed that weather and air quality factors were not very useful in predicting asthma daily admissions. Nonetheless, we establish approaches that these environmental factors could be used in future health forecasting involving different datasets. Environmental factors may still be useful predictors for developing pragmatic forecasting tools for chronic respiratory illness.

Lagged predictors were also more relevant in model development compared to real measures, because of the potential delayed effects of exposures. Human sentinels are also useful in predicting future events, possibly because they succinctly capture the weather and air quality exposure effects.

Excess demand for health care services is a great challenge to any health care service provider but the ability to forecast peak events is a promising resource. Quantile regressions can be used in modelling peak daily asthma admissions.

12.2 Recommendations

Health forecasting should be viewed as a dynamic activity with an equally continuous update of whatever framework that is being employed. This is because of the changing circumstances of individuals /populations, and their local environments, which mostly determine the prognosis of diseases. Increasing health forecasting activity will also help perfect or sharpen its approaches.

Further research on health forecasting needs to focus on defining or classifying the tools *per* perspective. This includes the terms, principles and methodological typologies that are mainly applicable to health forecasting; so as to help guide the process of forecasting health events. There is a need to further examine the approaches for validating health forecasts and to propose suitable ones, which could then be adopted in health forecasting.

In predicting specific chronic respiratory diseases such as asthma, it is recommended that future work incorporate historical events in model development process and also provide notes for the interpretation of models based on peculiar antecedents.

Finally, further research on the cost-saving benefits of health forecasting with examples on its return on investment may be a useful point to draw the attention and support of policy makers in health forecasting research

12.3 Advances

The thesis makes a number of contributions to knowledge in health forecasting. First, it identifies the theoretical environmental and social settings (framework) which explains some of the peculiar aspects of the disease burden of asthma in the UK. The pragmatic approaches to modelling LOS with random effects models of negative binomial regressions was a modest but novel contribution to the area.

Secondly, the thesis advances the discourse of health forecasting literature by providing additional descriptions to key principles of health forecasting, which have not been explained elsewhere. Some of these include *Uncertainty and error*, *Focus* of health forecasting, *data aggregation and accuracy*, and *horizons* of health forecasting.

The third significant contribution of the thesis to the literature on health forecasting is encompassed in Chapter 6, which was focused on "classifications". A unique output of this paper is that it identifies and presents the choices available for measuring the accuracy of health forecasting models, and also unearths the lack of a common approach to /discrepancies in the modes of validation of a health forecast.

Another contribution of this work is its demonstration of forecasting asthma related hospital admissions in London using negative binomial models. This process reveals the importance of complementing variable selection approaches, particularly for the specification of environmental (weather and air quality) and temporal factors.

A further contribution which forms the thrust of Chapter 8, is the development of statistical models to forecast asthma related admissions in London using the concept of human sentinels. This technique theorizes that individual asthma sufferers who have a more sensitive lungs and more prone to exacerbations could be used as indicators to forewarn others. Hence the paper develops univariate lagged models which predicts future events.

Also, the thesis makes another significant contribution to health forecasting by developing approaches for forecasting extreme events using Quantile regression models (QRM). These models were developed and tested with asthma admissions data in London (Chapter 9) and also for respiratory related deaths in New York City (Chapter 10). Both papers provide compelling evidence that there are robust approaches for predicting situations of excess demand for health care services or resources.

The analysis in the contributions, which have been described above show that temporal and environmental factors (including weather and air quality measures) can be useful predictors of asthma daily admissions, but are mediated by lags.

The policy implications of this thesis are narrow in respect of the immediate beneficiary of the ideas it presents, but again could have a very wide population health impact depending on how the application is cascaded through health delivery systems. It is for instance anticipated that the results of this thesis will draw the interest and attention of strategic stakeholders involved in health forecasting such as the UK Met Office, who have a mandate to reach out to health services providers with health forecasting services – health alerts, *inter alia*.

List of references

1. Soyiri IN, Reidpath DD. An overview of health forecasting. Environ Health Prev Med. 2013;18(1):1-9.
2. Soyiri IN, Reidpath DD. Evolving forecasting classifications and applications in health forecasting. Int J Gen Med. 2012;5:381-9.
3. Erbas B, Hyndman RJ. Sensitivity of the estimated air pollution-respiratory admissions relationship to statistical model choice. Int J Environ Health Res. 2005;15(6):437-48.
4. Hemming D, Colman A, James P, Kaye N, Marno P, McNeall D, et al. Framework for COPD forecasting in the UK using weather and climate change predictions. IOP Conf Series: Earth and Environmental Science 2009;6 (142021).
5. Marno P, Bryden C, Bird W, Watkin HA. How different measures of cold weather affect chronic obstructive pulmonary disease (COPD) hospital admissions in London. Eur Respir Rev. 2006;15(101):185-6.
6. Schonwetter RS, Clark LD, Leedy SA, Quinn MJ, Azer M, Kim S. Predicting emergency room visits and hospitalizations among hospice patients with cardiac disease. J Palliat Med. 2008;11(8):1142-50.
7. Bakerly ND, Roberts JA, Thomson AR, Dyer M. The effect of COPD health forecasting on hospitalisation and health care utilisation in patients with mild-to-moderate COPD. Chronic respiratory disease. 2011;8(1):5-9.
8. Wilkinson P, Chalabi Z, Raine R, Stevens M. Evaluation of A Pilot Health Forecasting System for England. Epidemiology. 2005;16(5):S93.
9. Hoot NR, LeBlanc LJ, Jones I, Levin SR, Zhou C, Gadd CS, et al. Forecasting Emergency Department Crowding: A Discrete Event Simulation. Annals of emergency medicine. 2008;52(2):116-25.
10. Hoot NR, LeBlanc LJ, Jones I, Levin SR, Zhou C, Gadd CS, et al. Forecasting Emergency Department Crowding: A Prospective, Real-time Evaluation. Journal of the American Medical Informatics Association. 2009;16(3):338-45.
11. Boyle J, Jessup M, Crilly J, Green D, Lind J, Wallis M, et al. Predicting emergency department admissions. Emergency Medicine Journal. 2011;doi:10.1136/emj.2010.103531.
12. 2011 Census. Population and Household Estimates for England and Wales, March 2011. London: Office for National Statistics, 2012.
13. HES. HES User Guide. Leeds: Health and Social Care Information Centre; 2008 [01 Jan]; Available from: http://www.hesonline.nhs.uk
14. Hansell A, Bottle A, Shurlock L, Aylin P. Accessing and using hospital activity data. J Public Health. 2001;23(1):51-6.
15. Peled JU, Sagher O, Morrow JB, Dobbie AE. Do Electronic Health Records Help or Hinder Medical Education? PLoS Med. 2009;6(5):e1000069.
16. HESonline. Hospital Episode Statistics HESonline. Leeds: Health and Social Care Information Centre; 2009 [cited 2009 01 Jan]; Available from: http://www.hesonline.nhs.uk
17. Jha AK, DesRoches CM, Campbell EG, Donelan K, Rao SR, Ferris TG, et al. Use of Electronic Health Records in U.S. Hospitals. N Engl J Med. 2009;360(16):1628-38.

18. Westaby S, Archer N, Manning N, Adwani S, Grebenik C, Ormerod O, et al. Comparison of hospital episode statistics and central cardiac audit database in public reporting of congenital heart surgery mortality. BMJ. 2007;335(7623):759-.
19. Davies S. Summary of other responses. BMJ. 2007;335(7625):839-b-40.
20. Pullen I, Loudon J. Improving standards in clinical record-keeping. Adv Psychiatr Treat. 2006;12(4):280-6.
21. Met-Office. Synoptic and climate stations. 2009 [01/03/09]; Available from: http://www.metoffice.gov.uk/climate/uk/networks/index.html.
22. Met-Office. Methods and analysis. 2009 [01/03/2009]; Available from: http://www.metoffice.gov.uk/climate/uk/about/methods.html.
23. Eisenberg D. The mixed effects of precipitation on traffic crashes. Accident Analysis & Prevention. 2004;36(4):637-47.
24. Met-Office. NAME III dispersion model. 2009 [01/05/2009]; Available from: http://www.metoffice.gov.uk/environment/name_iii.html.
25. AEA. Monitoring Site Environment Type. 2009 [25/03/2009]; Available from: http://www.airquality.co.uk/archive/site_type.php.
26. Met-Office. Changing seasons. 2009 [26/04/2009]; Available from: http://www.metoffice.gov.uk/climatechange/guide/effects/seasons.html.
27. Dominici F, McDermott A, Daniels M, Zeger SL, Samet JM. Revised analyses of the National Morbidity, Mortality, and Air Pollution Study: mortality among residents of 90 cities. J Toxicol Environ Health A. 2005;68(13-14):1071-92.
28. Peng RD, Welty LJ. The "NMMAPSdata" Package. R News. 2004;4(2):10-4.
29. Stolwijk AM, Straatman H, Zielhuis GA. Studying seasonality by using sine and cosine functions in regression analysis. J Epidemiol Community Health. 1999;53(4):235-8.
30. Barnett AG, Dobson AJ. Analysing Seasonal Health Data. Heidelberg Dordrecht London New York: Springer; 2009. 168(39-40) p.
31. Peacock J, Kerry S. Presenting Medical Statistics from Proposal to Publication: A Step-by-Step Guide. Oxford: Oxford University Press; 2007.
32. Bland JM. An Introduction to Medical Statistics 3rd Edition ed. Oxford: Oxford University Press; 2006.
33. Soyiri IN, Reidpath DD, Sarran C. Asthma length of stay in hospitals in London 2001-2006: demographic, diagnostic and temporal factors. PLoS One. 2011;6(11):e27184. Epub 2011/11/11.
34. Collett D. Modelling Binary Data. Second Edition ed. London: Chapman & Hall / CRC Press; 2003. 387 p.
35. Hao L, Naiman DQ, editors. Quantile Regression. Thousand Oaks, California 91320: Sage Publications, Inc; 2007.
36. Koenker R, Hallock KF. Quantile Regression. The Journal of Economic Perspectives. 2001;15(4):143-56.
37. Gujarati DN. Essentials of Econometrics. Third ed: McGraw Hill Higher Education; 2005.
38. Armstrong JS, Collopy F. 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):150-.

39. Soyiri IN, Reidpath DD. Evolving forecasting classifications and applications in health forecasting. *Int J Gen Med*. 2012;5(1):381-9. Epub 03 May 2012.
40. Galant SP, Crawford LJR, Morphew T, Jones CA, Bassin S. Predictive Value of a Cross-Cultural Asthma Case-Detection Tool in an Elementary School Population. Pediatrics. 2004;114(3):e307-16.
41. Sistek D, Tschopp JM, Schindler C, Brutsche M, Ackermann-Liebrich U, Perruchoud AP, et al. Clinical diagnosis of current asthma: predictive value of respiratory symptoms in the SAPALDIA study. Eur Respir J. 2001;17(2):214-9.
42. Sheppard L, Levy D, Norris G, Larson TV, Koenig JQ. Effects of ambient air pollution on nonelderly asthma hospital admissions in Seattle, Washington, 1987-1994. Epidemiology. 1999;10(1):23-30.
43. Hajat S, Anderson HR, Atkinson RW, Haines A. Effects of air pollution on general practitioner consultations for upper respiratory diseases in London. Occup Environ Med. 2002;59(5):294-9.
44. Erbas B, Kelly AM, Physick B, Code C, Edwards M. Air pollution and childhood asthma emergency hospital admissions: estimating intra-city regional variations. Int J Environ Health Res. 2005;15(1):11-20.
45. Soyiri IN, Reidpath DD. Semi-structured black-box prediction: proposed approach for asthma admissions in London. *Int J Gen Med*. 2012;5(1):693-705. Epub 20 August 2012.
46. Long JS. Predicted probabilities for count models. Stata Journal. 2001;1(1):51-7.
47. Zorn C. Evaluating Zero-Inflated and Hurdle Poisson Specifications. 1998 [01/11/2008]; Working Papers]. Available from: http://polmeth.wustl.edu/workingpapers.php?text=use+of+force&searchkeywords=T&order=dateposted.
48. Anon. A brief overview of count data commands in Stata. 2008 [01/11/2008]; Available from: http://www.sts.uzh.ch/past/hs08/count/stata_glossar.pdf.
49. Alt JE, King G, Signorino C. 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)):Pp. 21-44.
50. Mitchell SM, Moore WH. A New Look at Cold War Presidents' Use of Force: Aggregation Bias, Truncation, and Temporal Dynamic Issues. 2000 [01/11/2008]; Working Papers]. Available from: http://polmeth.wustl.edu/workingpapers.php?text=use+of+force&searchkeywords=T&order=dateposted.
51. Pregibon D. Goodness of Link Tests for Generalized Linear Models. Journal of the Royal Statistical Society Series C (Applied Statistics). 1980;29(1):15-4.
52. Koenker R. Quantile regression. New York: Cambridge Press; 2005.
53. Soyiri IN, Reidpath DD. Predicting extreme asthma events in London using quantile regression models. *Emerg Health Threats J*. 2011;4:s162:39-40.
54. Soyiri IN, Reidpath DD. An overview of health forecasting. *Environ Health Prev Med*. 2012;Accepted: 25-06-2012(In press). Epub 25 June 2012.
55. Taylor JW, Bunn DW. Investigating improvements in the accuracy of prediction intervals for combinations of forecasts: A simulation study. International Journal of Forecasting. 1999;15(3):325-39.

56. Baur D, Saisana M, Schulze N. Modelling the effects of meteorological variables on ozone concentration--a quantile regression approach. Atmospheric Environment. 2004;38(28):4689-99.
57. Barbosa SM. Quantile trends in Baltic sea level. Geophys Res Lett. 2008;35.
58. Vaz S, Martin CS, Eastwood PD, Ernande B, Carpentier A, Meaden GJ, et al. Modelling species distributions using regression quantiles. Journal of Applied Ecology. 2008;45(1):204-17.
59. Cade BS, Dong Q. A quantile count model of water depth constraints on Cape Sable seaside sparrows. J Anim Ecol. 2008;77(1):47-56.
60. Cade BS, Noon BR. A gentle introduction to quantile regression for ecologists. Frontiers in Ecology and the Environment. 2003;1(8):412-20.
61. Eide E, Showalter MH. The effect of school quality on student performance: A quantile regression approach. Economics Letters. 1998;58(3):345-50.
62. Beyerlein A, Fahrmeir L, Mansmann U, Toschke AM. Alternative regression models to assess increase in childhood BMI. BMC Med Res Methodol. 2008;8:59.
63. WHO. Chronic respiratory diseases - Asthma. Geneva: World Health Organization; 2010 [updated Oct. 2010]; No 307:[Fact Sheet]. Available from: http://www.who.int/respiratory/asthma/en/.
64. GINA. Global Strategy for Asthma Management and Prevention, Global Initiative for Asthma (GINA). 2008 [01/02/2009]; Electronic]. Available from: http://www.ginasthma.com/Guidelineitem.asp??l1=2&l2=1&intId=60.
65. Kaur B, Anderson HR, Austin J, Burr M, Harkins LS, Strachan DP, 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):118-24.
66. Asthma-UK. 'Where Do We Stand?' Asthma in the UK today Electronic. 2004 01/02/2009. Report No.
67. Pearce N, Ait-Khaled N, Beasley R, Mallol J, Keil U, Mitchell E, 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):758-66.
68. Masoli M, Fabian D, Holt S, Beasley R. The global burden of asthma: executive summary of the GINA Dissemination Committee report. Allergy. 2004;59(5):469-78.
69. Subbarao P, Mandhane PJ, Sears MR. Asthma: epidemiology, etiology and risk factors. CMAJ. 2009;181(9):E181-90.
70. Fanta CH. Asthma. N Engl J Med. 2009;360(10):1002-14.
71. Dougherty RH, Fahy JV. Acute exacerbations of asthma: epidemiology, biology and the exacerbation-prone phenotype. Clin Exp Allergy. 2009;39(2):193-202.
72. Neville RG, McCowan C, Hoskins G, Thomas G. Cross-sectional observations on the natural history of asthma. Br J Gen Pract. 2001;51(466):361-5.
73. Dosanjh A. Childhood asthma and anthropogenic CO(2) emissions. J Asthma Allergy. 2011;4:103-5.
74. Apologia. An introduction to the history and culture of Pharaonic Egypt. 2000 [cited 2010 December]; Available from: http://www.reshafim.org.il/ad/egypt/timelines/topics/medicine.htm.

75. Opolski M, Wilson I. Asthma and depression: a pragmatic review of the literature and recommendations for future research. Clinical Practice and Epidemiology in Mental Health. 2005;1(1):18.
76. Crosta P. Asthma History - Through The Ages. Medical News Today. 2010.
77. Marketos SG, Ballas CN. Bronchial asthma in the medical literature of Greek antiquity. J Asthma. 1982;19(4):263-9.
78. Cosman MP. A Feast for Aesculapius: Historical Diets for Asthma and Sexual Pleasure. Annual Review of Nutrition. 1983;3(1):1-34.
79. Holgate ST, Commins BT, Anderson HR. Asthma and outdoor air pollution. HMSO, London: Department of Health Committee on the Medical Effects of Air Pollutants. Available: http://www.comeap.org.uk/documents/reports.html. Accessed 1 November 2010., 1995.
80. Panettieri Jr RA, Covar R, Grant E, Hillyer EV, Bacharier L. Natural history of asthma: Persistence versus progression--does the beginning predict the end? Journal of Allergy and Clinical Immunology. 2008;121(3):607-13.
81. Fishwick D, Curran AD. Variability in the diagnosis of occupational asthma and implications for clinical practice. Curr Opin Allergy Clin Immunol. 2008;8(2):140-4 10.1097/ACI.0b013e3282f60f75.
82. Holgate ST. The epidemic of asthma and allergy. J R Soc Med. 2004;97(3):103-10.
83. Gaffin JM, Phipatanakul W. The role of indoor allergens in the development of asthma. Curr Opin Allergy Clin Immunol. 2009;9(2):128-35 10.1097/ACI.0b013e32832678b0.
84. British Thoracic Society Scottish Intercollegiate Guidelines N. British Guideline on the Management of Asthma. Thorax. 2008;63(Suppl_4):iv1-121.
85. Holgate ST. Lessons learnt from the epidemic of asthma. Qjm. 2004;97(5):247-57.
86. Nowak D, Suppli Ulrik C, von Mutius E. Asthma and atopy: has peak prevalence been reached? Eur Respir J. 2004;23(3):359-60.
87. 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):3-14.
88. HESonline. Asthma and allergies. Leeds: Health and Social Care Information Centre; 2009 [cited 2009 01 Jan]; Available from: http://www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=1937&categoryID=888.
89. Asthma-UK. ‘Wish you were here?’ Electronic. Asthma UK, 2008 01/02/2009. Report No.
90. Partridge MR. Examining the unmet need in adults with severe asthma. EUROPEAN RESPIRATORY REVIEW. 2007;16(104):67-72.
91. CDC. Adult asthma data: prevalence tables and maps. 2007 [19 June 2009]; Available from: http://www.cdc.gov/asthma/brfss/07/current/tableC1.htm.
92. Peters SP, Ferguson G, Deniz Y, Reisner C. Uncontrolled asthma: A review of the prevalence, disease burden and options for treatment. Respir Med. 2006;100(7):1139-51.
93. Asthma-UK. Living on a Knife Edge. 2004 Tue 20 Dec 2005. Report No.: Contract No.: 01/05/2009.

94. Demoly P, Paggiaro P, Plaza V, Bolge SC, Kannan H, Sohier B, et al. Prevalence of asthma control among adults in France, Germany, Italy, Spain and the UK. EUROPEAN RESPIRATORY REVIEW. 2009;18(112):105-12.
95. OHE. Compendium of Health Statistics. In: Economics OoH, editor. 15th edition ed2004.
96. Asthma-UK. No Asthma Improvement For Decade Has Cost NHS 100m Pounds - Asthma UK Scotland. Asthma UK Scotland; 2010 [07 Nov 2010]; Available from: http://www.medicalnewstoday.com/articles/203257.php.
97. Hoskins G, McCowan C, Neville RG, Thomas GE, Smith B, Silverman S. Risk factors and costs associated with an asthma attack. Thorax. 2000;55(1):19-24.
98. Xiao J, Lee AH, Vemuri SR. Mixture distribution analysis of length of hospital stay for efficient funding. Socio-Economic Planning Sciences. 1999;33:39-59.
99. Graves N, Weinhold D, Tong E, Birrell F, Doidge S, Ramritu P, et al. Effect of healthcare-acquired infection on length of hospital stay and cost. Infect Control Hosp Epidemiol. 2007;28(3):280-92.
100. Paradis AR, Stewart VT, Bayley KB, Brown A, Bennett AJ. Excess cost and length of stay associated with voluntary patient safety event reports in hospitals. Am J Med Qual. 2009;24(1):53-60.
101. Atienza N, García-Heras J, Muñoz-Pichardo JM, Villa R. An application of mixture distributions in modelization of length of hospital stay. Stat Med. 2008;27(9):1403-20.
102. Mulla ZD, Gibbs SG, Aronoff DM. Correlates of length of stay, cost of care, and mortality among patients hospitalized for necrotizing fasciitis. Epidemiol Infect. 2007;135(5):868-76.
103. Sobota A, Graham DA, Heeney MM, Neufeld EJ. Corticosteroids for acute chest syndrome in children with sickle cell disease: Variation in use and association with length of stay and readmission. American Journal of Hematology. 2009;85(1):24-8.
104. Morrison DS, McLoone P. Changing patterns of hospital admission for asthma, 1981-97. Thorax. 2001;56(9):687-90.
105. Soyiri I, Reidpath D, Sarran C. Determinants of asthma length of stay in London hospitals: individual versus area effects. *Emerg Health Threats J*. 2011;4(0):143-.
106. Lyratzopoulos G, Havely D, Gemmell I, Cook GA. Factors influencing emergency medical readmission risk in a UK district general hospital: A prospective study. BMC Emerg Med. 2005;5(1):1.
107. Callery P, Kyle RG, Campbell M, Banks M, Kirk S, Powell P. Readmission in children's emergency care: an analysis of hospital episode statistics. Archives of Disease in Childhood. 2009;95(5):341-6.
108. Watson L, Turk F, Rabe KF. Burden of asthma in the hospital setting: an Australian analysis. Int J Clin Pract. 2007;61(11):1884-8.
109. Arnold DH, Gebretsadik T, Minton PA, Higgins S, Hartert TV. Assessment of severity measures for acute asthma outcomes: a first step in developing an asthma clinical prediction rule. Am J Emerg Med. 2008;26(4):473-9.
110. Arnold DH, Gebretsadik T, Abramo TJ, Sheller JR, Resha DJ, Hartert TV. The Acute Asthma Severity Assessment Protocol (AASAP) study: objectives and methods of a study to develop an acute asthma clinical prediction rule. Emergency Medicine Journal. 2011.

111. Schenkel S. Promoting Patient Safety and Preventing Medical Error in Emergency Departments. Academic Emergency Medicine. 2000;7(11):1204-22.
112. Kuan V, Quantrill SJ. S7 Annual trends in asthma emergency attendance, hospital admission and readmission. Thorax. 2010;65(Suppl 4):A6-A7.
113. Rowe BH, Villa-Roel C, Abu-Laban RB, Stenstrom R, Mackey D, Stiell IG, et al. Admissions to Canadian hospitals for acute asthma: A prospective, multicentre study. Oakville, ON, CANADA: Pulsus; 2010. 6 p.
114. Scott IA. Public hospital bed crisis: too few or too misused? Australian Health Review. 2010;34(3):317-24.
115. Adams RJ, Smith BJ, Ruffin RE. Factors associated with hospital admissions and repeat emergency department visits for adults with asthma. Thorax. 2000;55(7):566-73.
116. Downing A, Rudge G, Cheng Y, Tu YK, Keen J, Gilthorpe MS. Do the UK government's new Quality and Outcomes Framework (QOF) scores adequately measure primary care performance? A cross-sectional survey of routine healthcare data. BMC Health Serv Res. 2007;7:166.
117. Gupta RS, Zhang X, Sharp LK, Shannon JJ, Weiss KB. The protective effect of community factors on childhood asthma. Journal of Allergy and Clinical Immunology. 2009;123(6):1297-304.e2.
118. Cakmak S, Dales RE, Judek S, Coates F. Does Socio-demographic Status Influence the Effect of Pollens and Molds on Hospitalization for Asthma? Results from a Time-series Study in 10 Canadian Cities. Annals of Epidemiology. 2005;15(3):214-8.
119. Rizza P, Bianco A, Pavia M, Angelillo IF. Preventable hospitalization and access to primary health care in an area of Southern Italy. BMC Health Serv Res. 2007;7:134.
120. Qian ZM, Lin HM, Chinchilli V, Lehman E, Stewart W, Shah N, et al. Associations Between Air Pollution and Peak Expiratory Flow Among Patients with Persistent Asthma. Journal of Toxicology and Environmental Health-Part a-Current Issues. 2009;72(1):39-46.
121. Rodrigo GJ, Plaza V, Forns SB, Tordera MP, Salas J. Factors associated with mortality in patients hospitalized in Spain and Latin America for acute severe asthma in 1994, 1999, and 2004. J Bras Pneumol. 2008;34(8):546-51.
122. Lobo FA, Almada-Lobo B. Quality of Life in Asthmatic Outpatients. Journal of Asthma. 2008;45(1):27 - 32.
123. Louise W, Florian T, Philip J, Stephen TH. Factors associated with mortality after an asthma admission: A national United Kingdom database analysis. Respir Med. 2007;101(8):1659-64.
124. Jeffery PK. Structural and inflammatory changes in COPD: a comparison with asthma. Thorax. 1998;53(2):129-36.
125. NIH. National Asthma Guidelines Updated. August 29, 2007 ed. U.S. Department of Health and Human Services: National Institute of Health News; 2007.
126. Busse WW, Lemanske Jr RF. Expert Panel Report 3: Moving forward to improve asthma care. Journal of Allergy and Clinical Immunology. 2007;120(5):1012-4.
127. Ho WC, Hartley WR, Myers L, Lin MH, Lin YS, Lien CH, et al. Air pollution, weather, and associated risk factors related to asthma prevalence and attack rate. Environ Res. 2007;104(3):402-9.
128. Diette GB, McCormack MC, Hansel NN, Breysse PN, Matsui EC. Environmental issues in managing asthma. Respir Care. 2008;53(5):602-15; discussion 16-7.

129. Mortimer KM, Neas LM, Dockery DW, Redline S, Tager IB. The effect of air pollution on inner-city children with asthma. Eur Respir J. 2002;19(4):699-705.
130. Benayoun L, Druilhe A, Dombret M-C, Aubier M, Pretolani M. Airway Structural Alterations Selectively Associated with Severe Asthma. Am J Respir Crit Care Med. 2003;167(10):1360-8.
131. Boulet LP, Sterk PJ. Airway remodelling: the future. Eur Respir J. 2007;30(5):831-4.
132. James AL, Wenzel S. Clinical relevance of airway remodelling in airway diseases. Eur Respir J. 2007;30(1):134-55.
133. Beasley R, Page C, Lichtenstein L. Airway remodelling in asthma. Clinical & Experimental Allergy Reviews. 2002;2(4):109-16.
134. Holgate ST. The epidemic of allergy and asthma. Nature. 1999;402(6760 Suppl):B2-4.
135. Chung KF, Adcock IM. Pathophysiological mechanisms of asthma - Application of cell and molecular biology techniques. Molecular Biotechnology. 2001;Volume 18(Number 3):213-32.
136. Sears MR. Descriptive epidemiology of asthma. The Lancet. 1997;350(Supplement 2):S1-S4.
137. Fleming DM, Cross KW, Sunderland R, Ross AM. Comparison of the seasonal patterns of asthma identified in general practitioner episodes, hospital admissions, and deaths. Thorax. 2000;55(8):662-5.
138. Campbell MJ, Cogman GR, Holgate ST, Johnston SL. Age specific trends in asthma mortality in England and Wales, 1983-95: results of an observational study. BMJ. 1997;314(7092):1439-41.
139. Campbell MJ, Holgate ST, Johnston SL. 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):1012.
140. Asthma-UK. Key facts & statistics London: Asthma UK; 2010 [updated 18 May 201010 June 2010]; Available from: http://www.asthma.org.uk/news_media/media_resources/for_journalists_key.html.
141. Patino CM, Martinez FD. Interactions between genes and environment in the development of asthma. Allergy. 2001;56(4):279-86.
142. Johnston NW, Sears MR. Asthma exacerbations ? 1: Epidemiology. Thorax. 2006;61(8):722-8.
143. Ma L, Shima M, Yoda Y, Yamamoto H, Nakai S, Tamura K, et al. Effects of airborne particulate matter on respiratory morbidity in asthmatic children. Journal of Epidemiology. 2008;18(3):97-110.
144. Parker JD, Akinbami LJ, Woodruff TJ. Air Pollution and Childhood Respiratory Allergies in the United States. Environmental Health Perspectives. 2009;117(1):140-7.
145. Walters S, Phupinyokul M, Ayres J. Hospital admission rates for asthma and respiratory disease in the West Midlands: their relationship to air pollution levels. Thorax. 1995;50(9):948-54.
146. Lin M, Chen Y, Burnett RT, Villeneuve PJ, Krewski D. 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):50-5.

147. Walters S, Griffiths RK, Ayres JG. Temporal association between hospital admissions for asthma in Birmingham and ambient levels of sulphur dioxide and smoke. Thorax. 1994;49(2):133-40.
148. Gent JF, Triche EW, Holford TR, Belanger K, Bracken MB, Beckett WS, 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):1859-67.
149. Repetto R. Population, resources, environment: an uncertain future. Popul Bull. 1987;42(2):3-43.
150. Sepulveda J, Lopez-Cervantes M, Frenk J, Gomez de Leon J, Lezana-Fernandez MA, Santos-Burgoa C. Key issues in public health surveillance for the 1990s. MMWR Morb Mortal Wkly Rep. 1992;41 Suppl:61-76.
151. IPCC. Assessing the health impacts of climate change. Geneva (Switzerland): WMO/UNEP, 1995.
152. DH/HPA. Health effects of climate change in the UK 2008: An update of the Department of Health report 2001/2002. Commissioned report Department of Health, Health Protection Agency 2008 29 Dec 2008 Report No.
153. Met-Office. Climate science explained. Exeter: The Met Office; 2009 [01/02/2009]; Available from: http://www.metoffice.gov.uk/climatechange/science/explained/index.html.
154. Lin S, Gomez MI, Hwang S-A, Munsie JP, Fitzgerald EF. 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):325 - 32.
155. Becklake MR, Ernst P. Environmental factors. The Lancet. 1997;350:S10-S3.
156. Menne B, Kovats S, Bell J. Climate change and health: international research agenda. Copenhagen, Denmark: European Preparatory Meeting for the Global Ministerial Forum on Research for Health, 2008 29-30 April. Report No.
157. Intergovernmental Panel on Climate Change (Lead Author) Andy Jorgensen. IPCC Fourth Assessment Report, Working Group II: Chapter 8. Encyclopedia of Earth Eds Cutler J Cleveland (Washington, DC: Environmental Information Coalition, National Council for Science and the Environment). Washington, D.C.: Encyclopedia of Earth; 2010.
158. Sauerborn R. Climate change: an agenda for research and teaching in public health. Scand J Public Health. 2007;35(6):561-3.
159. Storr J, Lenney W. School holidays and admissions with asthma. Arch Dis Child. 1989;64(1):103-7.
160. Rossi OV, Kinnula VL, Tienari J, Huhti E. Association of severe asthma attacks with weather, pollen, and air pollutants. Thorax. 1993;48(3):244-8.
161. Epton MJ, Martin IR, Graham P, Healy PE, Smith H, Balasubramaniam R, et al. Climate and aeroallergen levels in asthma: a 12 month prospective study. Thorax. 1997;52(6):528-34.
162. Hales S, Lewis S, Slater T, Crane J, Pearce N. Prevalence of adult asthma symptoms in relation to climate in New Zealand. Environ Health Perspect. 1998;106(9):607-10.

163. Weiland SK, Husing A, Strachan DP, Rzehak P, Pearce N. Climate and the prevalence of symptoms of asthma, allergic rhinitis, and atopic eczema in children. Occup Environ Med. 2004;61(7):609-15.
164. Chavarria JF. Short report: Asthma admissions and weather conditions in Costa Rica. Arch Dis Child. 2001;84(6):514-5.
165. Ivey MA, Simeon DT, Monteil MA. 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):1526-30.
166. Khot A, Burn R, Evans N, Lenney W, Storr J. Biometeorological triggers in childhood asthma. Clin Allergy. 1988;18(4):351-8.
167. Zeka A, Eisen EA, Kriebel D, Gore R, Wegman DH. Risk of upper aerodigestive tract cancers in a case-cohort study of autoworkers exposed to metalworking fluids. Occup Environ Med. 2004;61(5):426-31.
168. Fouillet A, Rey G, Wagner V, Laaidi K, Empereur-Bissonnet P, Le Tertre 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):309-17.
169. Kovats RS, Hajat S. Heat stress and public health: a critical review. Annu Rev Public Health. 2008;29:41-55.
170. Koskela HO. Cold air-provoked respiratory symptoms: the mechanisms and management. Int J Circumpolar Health. 2007;66(2):91-100.
171. Raphoz M, Goldberg MS, Garneau M, Héguy L, Valois M-F, Guay F. Associations Between Atmospheric Concentrations of Spores and Emergency Department Visits for Asthma Among Children Living in Montreal. Archives of Environmental & Occupational Health. 2010;65(4):201-10.
172. Héguy L, Garneau M, Goldberg MS, Raphoz M, Guay F, Valois M-F. Associations between grass and weed pollen and emergency department visits for asthma among children in Montreal. Environmental Research. 2008;106(2):203-11.
173. O'Shea SD, Taylor NF, Paratz JD. ...But watch out for the weather: factors affecting adherence to progressive resistance exercise for persons with COPD. J Cardiopulm Rehabil Prev. 2007;27(3):166-74; quiz 75-6.
174. Hajat S, Haines A. Associations of cold temperatures with GP consultations for respiratory and cardiovascular disease amongst the elderly in London. Int J Epidemiol. 2002;31(4):825-30.
175. Priftis KN, Paliatsos AG, Panagiotopoulou-Gartagani P, Tapratzi-Potamianou P, Zachariadi-Xypolita A, Nicolaidou P, et al. Association of weather conditions with childhood admissions for wheezy bronchitis or asthma in Athens. Respiration. 2006;73(6):783-90.
176. Nastos PT. Weather, Ambient Air Pollution and Bronchial Asthma in Athens, Greece. Seasonal Forecasts, Climatic Change and Human Health2008. p. 173-88.
177. Xirasagar S, Lin H-C, Liu T-C. Seasonality in pediatric asthma admissions: The role of climate and environmental factors. European Journal of Pediatrics. 2006;165(11):747-52.
178. Lecha Estela LB. Biometeorological classification of daily weather types for the humid tropics. Int J Biometeorol. 1998;42(2):77-83.

179. Schwartz J, Slater D, Larson TV, Pierson WE, Koenig JQ. Particulate air pollution and hospital emergency room visits for asthma in Seattle. Am Rev Respir Dis. 1993;147(4):826-31.
180. HESonline. Asthma, Health and seasons. Leeds: Health and Social Care Information Centre; 2009 [cited 2009 01 Jan]; Available from: http://www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=1937&categoryID=972.
181. Yang C-Y, Chen C-C, Chen C-Y, Kuo H-W. Air Pollution and Hospital Admissions for Asthma in a Subtropical City: Taipei, Taiwan. Journal of Toxicology and Environmental Health Part A. 2007;70:111-7.
182. Newson R, Strachan D, Corden J, Millington W. Fungal and other spore counts as predictors of admissions for asthma in the Trent region. Occup Environ Med. 2000;57(11):786-92.
183. Newson R, Strachan D, Archibald E, Emberlin J, Hardaker P, Collier C. Effect of thunderstorms and airborne grass pollen on the incidence of acute asthma in England, 1990-94. Thorax. 1997;52(8):680-5.
184. Newson R, Strachan D, Archibald E, Emberlin J, Hardaker P, Collier C. Acute asthma epidemics, weather and pollen in England, 1987-1994. Eur Respir J. 1998;11(3):694-701.
185. D'Amato G, Liccardi G, Frenguelli G. Thunderstorm-asthma and pollen allergy. Allergy. 2007;62(1):11-6.
186. Wark PA, Simpson J, Hensley MJ, Gibson PG. Airway inflammation in thunderstorm asthma. Clin Exp Allergy. 2002;32(12):1750-6.
187. Anderson W, Prescott GJ, Packham S, Mullins J, Brookes M, Seaton A. Asthma admissions and thunderstorms: a study of pollen, fungal spores, rainfall, and ozone. Qjm. 2001;94(8):429-33.
188. Higham J, Venables K, Kupek E, Bajekal M, Kopek E. Asthma and thunderstorms: description of an epidemic in general practice in Britain using data from a doctors' deputising service in the UK. J Epidemiol Community Health. 1997;51(3):233-8.
189. King A. Air pollution reduces EPC levels. Nat Rev Cardiol. 2010;7(10):540-.
190. Sarnat JA, Holguin F. Asthma and air quality. Curr Opin Pulm Med. 2007;13(1):63-6.
191. Babin S, Burkom H, Holtry R, Tabernero N, Davies-Cole J, Stokes L, 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):209-21.
192. Lin M, Chen Y, Villeneuve PJ, Burnett RT, Lemyre L, Hertzman C, 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):294-303.
193. Bartra J, Mullol J, del Cuvillo A, Davila I, Ferrer M, Jauregui I, et al. Air pollution and allergens. J Investig Allergol Clin Immunol. 2007;17 Suppl 2:3-8.
194. D'Amato G, Liccardi G, D'Amato M, Holgate S. Environmental risk factors and allergic bronchial asthma. Clin Exp Allergy. 2005;35(9):1113-24.
195. von Mutius E. The environmental predictors of allergic disease. Journal of Allergy and Clinical Immunology. 2000;105(1, Part 1):9-19.
196. Saxon A, Diaz-Sanchez D. Air pollution and allergy: you are what you breathe. Nat Immunol. 2005;6(3):223-6.

197. Katsouyanni K. Ambient air pollution and health. Br Med Bull. 2003;68(1):143-56.
198. Nicolai T, Carr D, Weiland SK, Duhme H, von Ehrenstein O, Wagner C, 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):956-63.
199. Ramadour M, Burel C, Lanteaume A, Vervloet D, Charpin D, Brisse F, et al. Prevalence of asthma and rhinitis in relation to long-term exposure to gaseous air pollutants. Allergy. 2000;55(12):1163-9.
200. Wang HC, Yousef E. Air Quality and Pediatric Asthma-Related Emergencies. Journal of Asthma. 2007;44(10):839-41.
201. Annesi-Maesano I, Moreau D, Caillaud D, Lavaud F, Le Moullec Y, Taytard A, et al. Residential proximity fine particles related to allergic sensitisation and asthma in primary school children. Respir Med. 2007;101(8):1721-9.
202. Hajat S, Haines A, Atkinson RW, Bremner SA, Anderson HR, Emberlin J. Association between Air Pollution and Daily Consultations with General Practitioners for Allergic Rhinitis in London, United Kingdom. Am J Epidemiol. 2001;153(7):704-14.
203. Hwang BF, Jaakkola JJ, Lee YL, Lin YC, Guo YL. Relation between air pollution and allergic rhinitis in Taiwanese schoolchildren. Respir Res. 2006;7:23.
204. Janssen NA, Brunekreef B, van Vliet P, Aarts F, Meliefste K, Harssema H, 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):1512-8.
205. Kramer U, Koch T, Ranft U, Ring J, Behrendt H. Traffic-related air pollution is associated with atopy in children living in urban areas. Epidemiology. 2000;11(1):64-70.
206. Lee YL, Shaw CK, Su HJ, Lai JS, Ko YC, Huang SL, et al. Climate, traffic-related air pollutants and allergic rhinitis prevalence in middle-school children in Taiwan. Eur Respir J. 2003;21(6):964-70.
207. Morgenstern V, Zutavern A, Cyrys J, Brockow I, Gehring U, Koletzko S, 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):8-16.
208. Penard-Morand C, Charpin D, Raherison C, Kopferschmitt C, Caillaud D, Lavaud F, et al. Long-term exposure to background air pollution related to respiratory and allergic health in schoolchildren. Clin Exp Allergy. 2005;35(10):1279-87.
209. Wong TW, Tam WS, Yu TS, Wong AH. Associations between daily mortalities from respiratory and cardiovascular diseases and air pollution in Hong Kong, China. Occup Environ Med. 2002;59(1):30-5.
210. Yu JH, Lue KH, Lu KH, Sun HL, Lin YH, Chou MC. 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):123-6.
211. Watson JP, Cowen P, Lewis RA. The relationship between asthma admission rates, routes of admission, and socioeconomic deprivation. Eur Respir J. 1996;9(10):2087-93.
212. Steenbergen MR, Jones BS. Modeling multilevel data structures. american Journal of political Science. 2002:218-37.
213. Stata SE. Release 11. Statistical Software College Station, TX: StataCorp LP. 2009.

214. Lewis JB, McGrath RJ, Seidel LF. Essentials of Applied Quantitative Methods for Health Services Managers. Sudbury MA, Boston: Jones and Bartlett Publishers, LLC; 2011.
215. Sekhri N, Chisholm R, Longhi A, Evans P, Rilling M, Wilson E, et al. Principles for Forecasting Demand for Global Health Products. Global Health Forecasting Working Group, Background Paper [Internet]. 2006; (01/02/2009). Available from: http://www.cgdev.org/doc/DemandForecasting/Principles.pdf.
216. Frank Wharam J, Weiner JP. The Promise and Peril of Healthcare Forecasting. Am J Manag Care. 2012;18(3):e82-e5.
217. Met-Office. The Met Office health forecasting services Exeter2009 [01 Jan 2009]; Available from: http://www.metoffice.gov.uk/health/.
218. Rogers DP, Shapiro MA, Brunet G, Cohen JC, Connor SJ, Diallo AA, et al. Health and climate - opportunities. Procedia Environmental Sciences. 2010;1:37-54.
219. van Meijgaard J, Fielding J, Shimkhada R, Eslami E, Cole B. Forecasting Health Outcomes Related to Air Quality for Placer County. Issue Brief 2010;Health Forecasting(UCLA School of Public Health).
220. Antoniou SA, Antoniou GA, Granderath FA, Mavroforou A, Giannoukas AD, Antoniou AI. Reflections of the Hippocratic Oath in Modern Medicine. World J Surg. 2010;34(12):3075-9.
221. Chang A, Lad EM, Lad SP. Hippocrates' influence on the origins of neurosurgery. Neurosurg Focus. 2007;23(1):E9.
222. Chadwick J, Mann WN. MEDICINE. In: Lloyd GER, editor. Hippocratic writings. England: Penguin Group; 1978. p. 170-1.
223. Medixine. Medixine Preventive E-care. Finland2009 [Accessed: 01 Jan 2009]; Available from: http://www.medixine.fi/.
224. UCLA HealthForecasting. Tools for Improving Population Health. Los Angeles2009 [Accessed: 01 Jan 2009]; Available from: http://www.health-forecasting.org/.
225. Malyuk RE, Wong C, Buree B, Kang A, Kang N. The interplay of infections, function and length of stay (LOS) in newly admitted geriatric psychiatry patients. Archives of gerontology and geriatrics. 2012;54(1):251-5.
226. Kudyba S, Gregorio T. Identifying factors that impact patient length of stay metrics for healthcare providers with advanced analytics. Health Informatics Journal. 2010;16(4):235-45.
227. Armstrong JS. Principles of Forecasting: A Handbook for Researchers and Practitioners. Hillier FS, editor. Dordrecht: Kluwer Academic Publishers; 2001.
228. Díaz J, Alberdi JC, Pajares MS, López C, López R, Lage MB, et al. A model for forecasting emergency hospital admissions: effect of environmental variables. J Environ Health. 2001;64(3):9-15.
229. Peng RD, Dominici F. Statistical Methods For Environmental Epidemiology With R: A Case Study In Air Pollution And Health: Springer-Verlag New York Inc. (United States); 2008.
230. Breiman L. Statistical Modeling: The Two Cultures. Statistical Science. 2001;16(3):199-215.
231. The PME. The Air That We Breathe: Addressing the Risks of Global Urbanization on Health. PLoS Med. 2012;9(8):e1001301.

Appendices

Appendix A

A1.0 Related peer reviewed conference papers

Environmental Health Conference 2011

International Society for Disease Surveillance 10th Annual Conference 2011

International Society for Disease Surveillance 9th Annual Conference 2010

A1.1 Environmental Health Conference 2011

[O16]

The role of weather and air quality factors in forecasting asthma admissions in London

I.N. Soyiri*[1], D.D. Reidpath[2]

[1]*University of Ghana, Ghana,* [2]*Monash University, Malaysia*

The subject of health forecasting is not entirely new, but the literature is limited, particularly in the area of forecasting of adverse respiratory health events like asthma. It is an unclear field; and there has been almost no research looking at the forecasting of the demand for asthma hospital admissions. Health forecast enables individuals to take precautionary measures, and service providers to manage demand for health care. The ability to produce reliable health forecast is often limited by insufficient data and information as well as verifiable methodologies that have a good predictive power; are simple to implement and relatively inexpensive.

The concept of health forecasting is described in relation to a proposed asthma forecast model for London. It illustrates the procedure for developing asthma forecast models using hospital admission records (total daily asthma admissions) as dependant factor and environmental data (weather and air quality factors) as predictors. A further exploration of the relative effects of different exposure durations to asthma admissions in London is carried out.

This paper demonstrates how to predict both normal and extreme asthma events that lead to hospital admissions in London using lagged exposures of weather (temperature and humidity) and air quality (nitrogen oxide and ozone) factors. Negative binomial regression count models and the Quantile regression models were used in the predictions. Error measures used to compare suitable models were the Root mean square error and Receiver operating characteristic curves /coefficients. The paper adds an explanation to the concept of health forecasting and proposes some frameworks for forecasting asthma admissions in London.

Keywords: Health forecast, Asthma Hospital Admissions, Weather, Air Quality

Soyiri, I.N. and D.D. Reidpath, The role of weather and air quality factors in forecasting asthma admissions in London, in Environmental Health 2011: Resetting our Priorities 2011, Elsevier: Salvador, Brazil. (Available: http://elsevier.conference-services.net/programme.asp?conferenceID=2205&action=prog_categories)

A1.2 International Society for Disease Surveillance 10th Annual Conference 2011

Soyiri, I., D. Reidpath, and C. Sarran, Determinants of asthma length of stay in London hospitals: individual versus area effects (Published: *Emerg Health Threats J.*, 2011. 4(0): p. 143-143).

CoAction

Determinants of asthma length of stay in London hospitals: individual versus area effects

Ireneous Soyiri[1,2]*, Daniel Reidpath[2] and Christophe Sarran[3]

[1]University of Ghana, Legon, Accra, Ghana; [2]MONASH University, Sunway Campus, Malaysia; [3]Met Office, Exeter, UK

Objective

To identify predictors of length of stay (LOS) of asthma admissions in London and to model their area and individual effects.

Introduction

Asthma is a chronic condition of public health concern associated with morbidity, mortality and healthcare utilisation. It disproportionately affects certain ethnic and demographic groups.

Methods

Asthma admission records in London (2001 2006) were used. Negative binomial regression was used to model the effect of demographic (*sex, age & ethnic group*), diagnostic (*primary & secondary diagnosis, method of admission*) and temporal (*day of the week, meteorological season & year of admission*) factors on the LOS, accounting for the random effects of each patient's attendance, as model 'I' and again for area of residence, model 'A'. Akaike information criterion (AIC) was used to compare the two models.

Table 1. Summary statistics of asthma-related hospital admissions in London, 2001–2006

Characteristics	N (%)
Age (years)	
< 5	12,420 (12.4)
5 14	10,700 (10.7)
15 44	16,612 (16.6)
45 59	7,029 (7.0)
60 74	5,698 (5.7)
≥ 75	4,309 (4.3)
Ethnic Group	
White	26,230 (46.2)
Black	6,604 (11.6)
Asian	6,382 (11.2)
Mixed/Other	5,780 (10.2)
Not stated	11,782 (20.8)
Secondary Diagnosis	
Other diseases of URT	25,053 (44.1)
Influenza and Pneumonia	692 (1.2)
Other acute lower respiratory infections	6,256 (11.0)
Acute upper respiratory infections	70 (0.1)
Chronic lower respiratory infections	1,207 (2.1)
Lung diseases due to external agents	1,519 (2.7)
Other diseases of respiratory system	378 (0.7)
Other Non-respiratory system diseases	15,227 (26.8)
Missing Values	6,376 (11.2)
Method of Admission	
Accident and emergency services	52,074 (91.7)
General Practitioner (GP)	2,602 (4.6)
Bed bureau	41 (0.1)
Consultants out patient clinic	577 (1.0)
Other means	1,484 (2.6)
Meteorological Season	
Summer	12,340 (21.7)
Spring	13,453 (23.7)
Autumn	16,800 (29.6)
Winter	14,185 (25.0)

Results

The median and mean asthma LOS over the period of study were 2 and 3 days, respectively. Admissions increased over the years from 8308 (2001) to 10,554 (2006), but LOS declined within the same period. Fewer males (48%) than females (52%) were admitted and, the latter had longer LOS compared to males. Only 5% were primarily diagnosed as *predominantly allergic*, whilst >94% were classified as '*asthma, unspecified*'. Younger people were more likely to be admitted than elderly, but the latter had higher LOS ($p < 0.001$). The secondary diagnosis and method of admission were important diagnostic determinants of length of stay, with very marginal differences between the two statistical models ('I' & 'A'). Again, all the temporal factors were significant determinants of LOS. Overall the patient cluster model (AIC=239394.8) outperformed the area model (AIC=247899.9).

Conclusions

Asthma LOS is best predicted by demographic, diagnostic and temporal factors with individual patients as a random effect.

Keywords

Asthma; length of stay; spell duration; risk factors; hospital admission

***I. Soyiri**
E-mail: soyiriin@yahoo.com

 Citation: Emerging Health Threats Journal 2011, 4: 11179 - DOI: 10.3402/ehtj.v4i0.11179

A1.3 International Society for Disease Surveillance 9th Annual Conference 2010

Soyiri, I.N. and D.D. Reidpath, Predicting extreme asthma events in London using quantile regression models. (Published: *Emerg Health Threats J.*, 2011. 4:s162: p. 39-40).

ISDS

Emerging Health Threats Journal 2011, 4:s162. doi: 10.3134/ehtj.10.162

www.eht-journal.org

ABSTRACT

Predicting extreme asthma events in London using quantile regression models

IN Soyiri[1,2], and DD Reidpath[2]

[1]University of Ghana, School of Public Health, Legon Accra, Ghana; and [2]Monash University, School of Medicine and Health Sciences, Kuala Lumpur, Malaysia
E-mail: soyiriin@yahoo.com

Objective

This paper describes a framework for creating a time series data set with daily asthma admissions, weather and air quality factors; and then generating suitable lags for predictive multivariate quantile regression models (QRMs). It also demonstrates the use of root mean square error (RMSE) and receiver operating characteristic (ROC) error measures in selecting suitable predictive models.

Introduction

The burden of asthma is a major public health issue, and of a wider interest particularly to public health practitioners, health care providers and policy makers, as well as researchers. The literature on forecasting of adverse respiratory health events like asthma attacks is limited. It is an unclear field; and there is a need for more research on the forecasting of the demand for hospital respiratory services.

Methods

Asthma hospital admissions from the hospital episode statistics (HES) database in the UK, weather and air quality measures obtained via the UK Met Office databases were used in the analysis (2005–2006). The key variables in the data were examined for their distribution and properties. Suitable time series lags were generated and converted into non-time series for bivariate quantile regression analysis. Multivariate QRMs were developed to predict extreme asthma events defined as > the 90th percentile. RMSE and ROC curves were used to compare error measures of each predictive model.

Results

All the potential predictors were independently significant with asthma daily admissions ($P<0.01$), however most proved to be insignificant in multivariate analysis when controlling for the other factors. Three significant predictive models were constructed: Model-I involved an 18-day lag barometric vapour pressure and a 3-day lag Nitrogen dioxide. Model-II included barometric pressure (18-day lag), Nitrogen dioxide (3-day lag) and air temperature (21-day lag), and then Model-III had barometric pressure (18-day lag), Nitrogen dioxide (3-day lag), air temperature (21-day lag), humidity (4-day lag) and formaldehyde (3-day lag). But for humidity and formaldehyde concentrations, all the variables were at least statistically significant at $P<0.01$ (see Table 1). Model-II had better predictive values for both normal and extreme asthma events compared with the other two

Table 1 QRM predictions of extreme (>90th percentile) asthma events in London

Asthma predictive models	*I*		*II*		*III*	
	s.e.	*Coef*[Ψ]	*s.e.*	*Coef*[Ψ]	*s.e.*	*Coef*[Ψ]
Barometric pressure HPa (18-day lag)	−0.18	2.00***	−0.3	1.44***	−0.24	0.85***
Nitrogen dioxide[§] (3-day lag)	−7.20e+07	1.8e+08*	−7.50e+07	4.4e+08***	−8.90e+07	2.9e+08**
Air temperature °C (21-day lag)			−0.18	−0.84***	−0.16	−1.11***
Humidity % (4-day lag)					−0.05	−0.04
Formal dehyde[§] (3-day lag)					−2.30e+08	3.70e+08
Specification test: (Linktest) *hatsq* *P*-value	0.155		0.922		0.136	

[Ψ]Expected change in log count for a one-unit increase in variable and degree of significance: *$P<0.1$, **$P<0.01$, ***$P<0.001$; [§]kgm ^{3}; standard error (s.e.).

models, and again had lower error measures compared with Model-I and Model-III.

Conclusions

Asthma daily admissions can be predicted from a combination of weather and air quality indicators including average daily measures of barometric pressure, Nitrogen dioxide, air temperature, humidity, and formaldehyde using QRMs. Barometric pressure, nitrogen dioxide, and air temperature were the best predictors of asthma daily admissions.

Acknowledgements

This paper was presented as an oral presentation at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA, on 1–2 December 2010.

Appendix B

B1.1 Synoptic and Climate Stations in London Area

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

B1.2 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

C1.0 Data visualisation and summary of preliminary analysis

C1.1 Daily Asthma Hospital Admissions in London (2005-2006)

The total daily asthma admission in London is illustrated in the time series plot in Figure 1 and 2. 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, 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.

C1.2 Asthma Admissions and Weather factors (London, 2005-2006)

The mean air temperature distribution is presented alongside the asthma daily hospital admissions for London in Figure 9. The higher temperatures were obviously 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 10-13).

The other indicators whose distribution patterns have been plotted with that of the Asthma daily admissions are presented in Figures 14 to Figure 16. These include

Humidity, Barometric vapour pressure and Wind speed. Both Humidity and Pressure show some seasonal patterns whilst Wind speed does not.

C1.3 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 19), do not show any regular seasonal or monthly effect. This is same for the other pollutants (Nitrogen dioxide, Nitrogen oxide, Sulphur dioxide Formaldehyde and PM_{10}) except ozone. Ozone shows some seasonal trends with notable peaks in the summer period and low records in the winter (See Figure 17).

C1.4 Codebook and data summary (2005-2006)

Table 1 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

C1.5 Asthma Admissions (London, 2005-2006)

Figure 7 Daily Asthma hospital admissions in London (2005-2006)

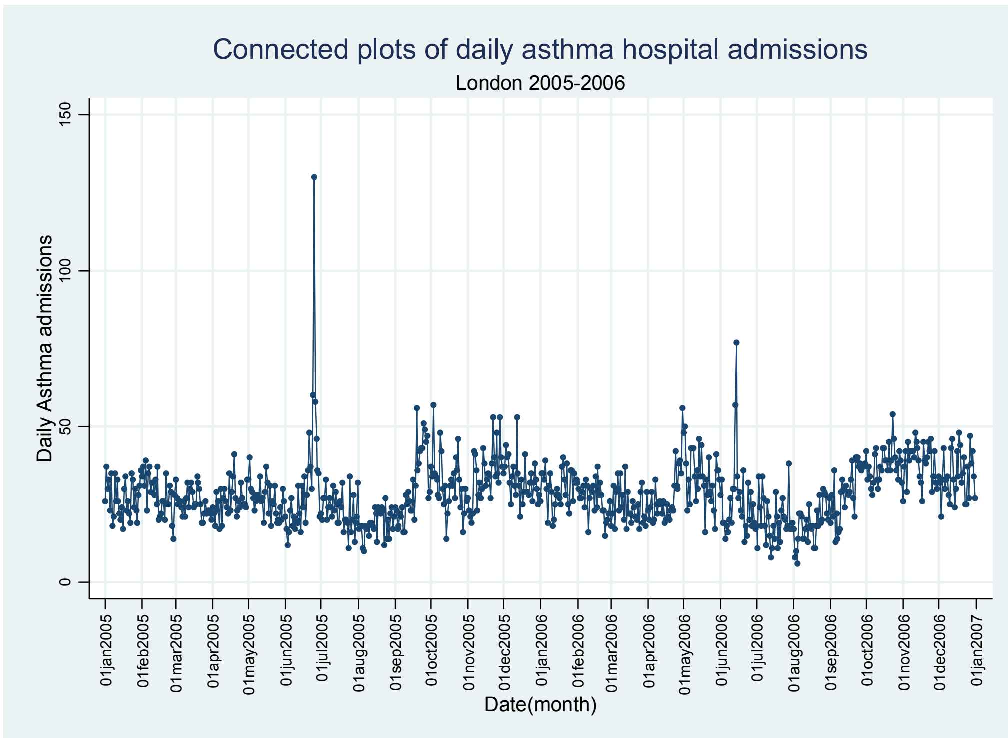
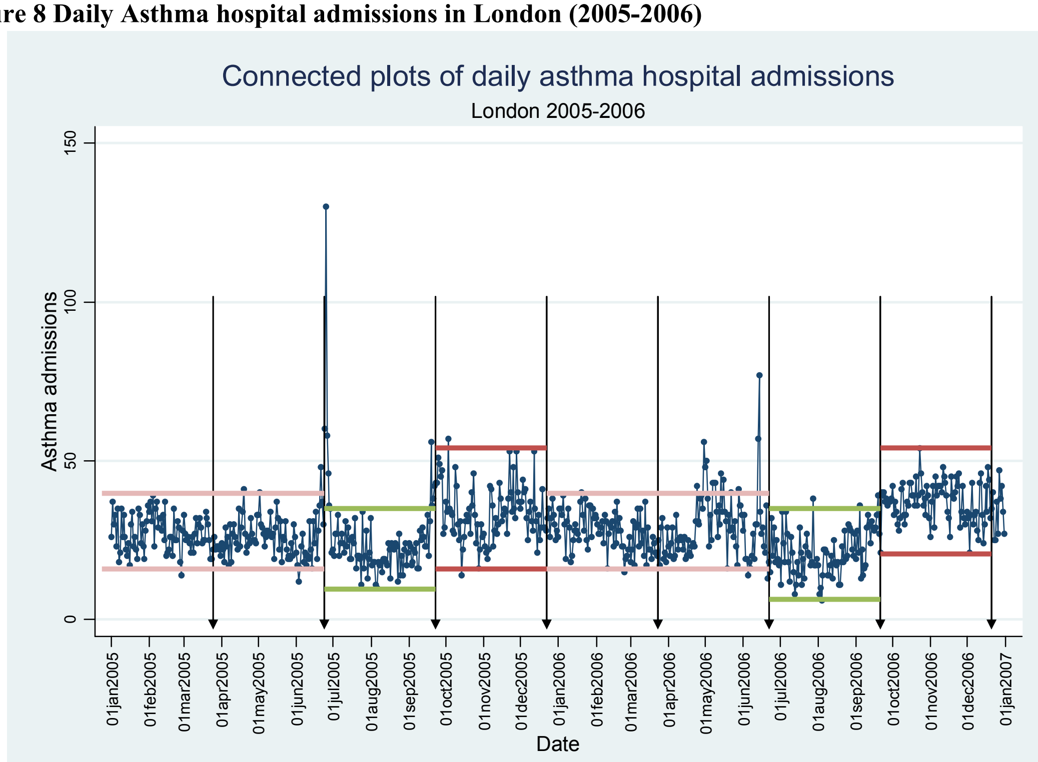


Figure 8 Daily Asthma hospital admissions in London (2005-2006)



Arrow marks /divisions represent the seasons in a year

C1.6 Asthma Admissions and Temperature (London, 2005-2006)

Figure 9 Asthma Admissions and Mean Air Temperature Distribution (London, 2005-2006)

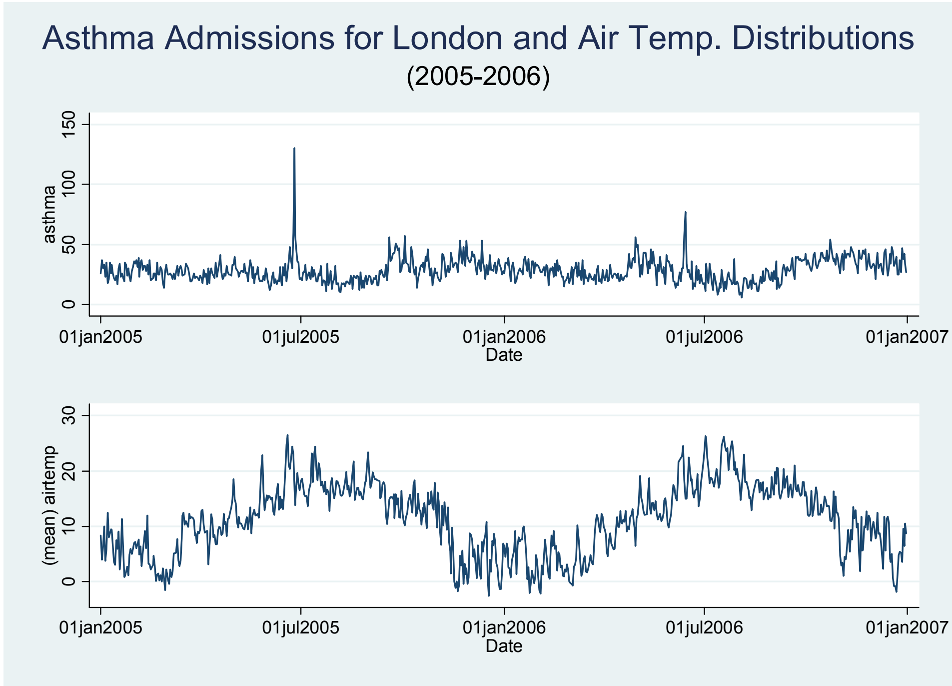


Figure 10 Asthma Admissions and Mean Daily Minimum/ Daily Maximum Temperature Distributions (London, 2005-2006)

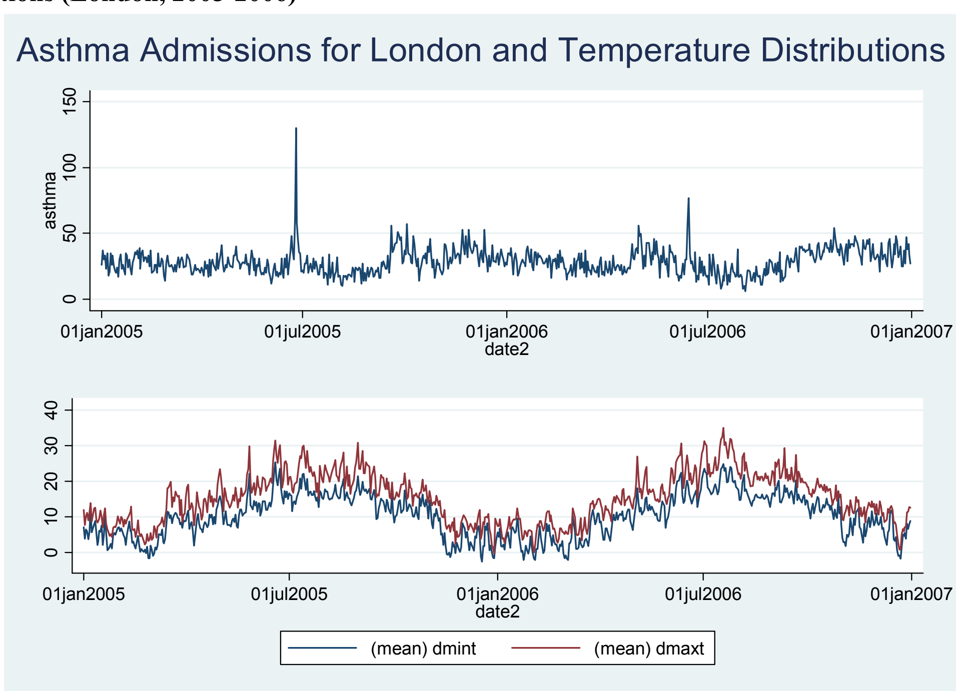


Figure 11 Asthma Admissions and Mean Night Minimum/ Night Maximum Temperature Distributions (London, 2005-2006)

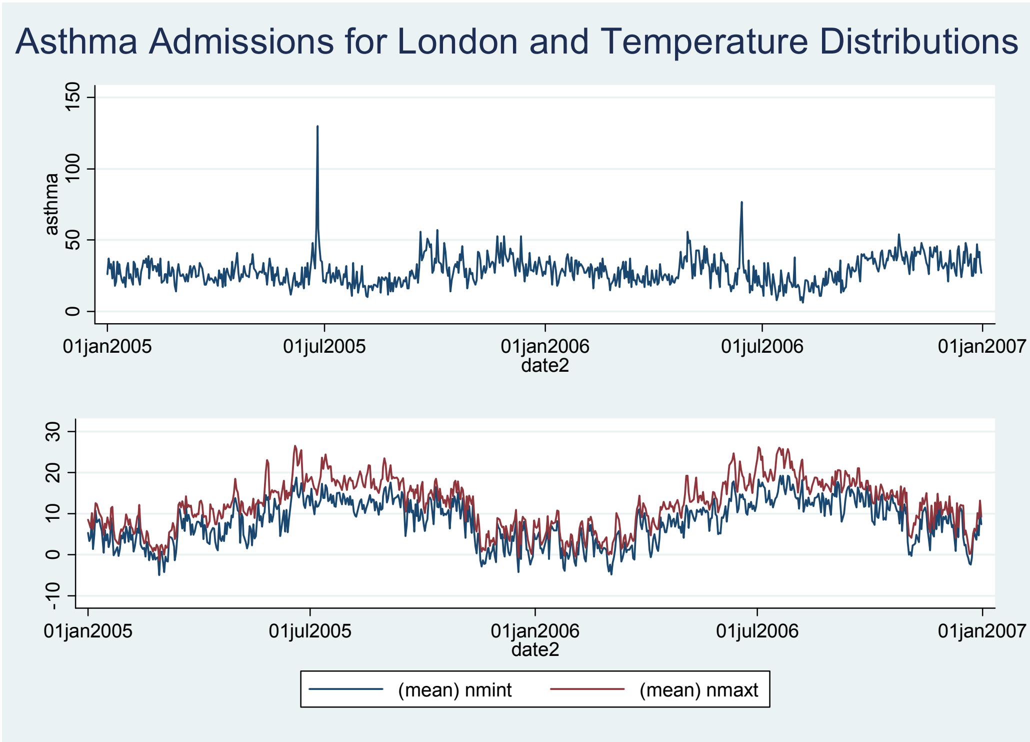


Figure 12 Asthma Admissions and Mean Dew point/ Wet bulb Temperature Distributions (London, 2005-2006)

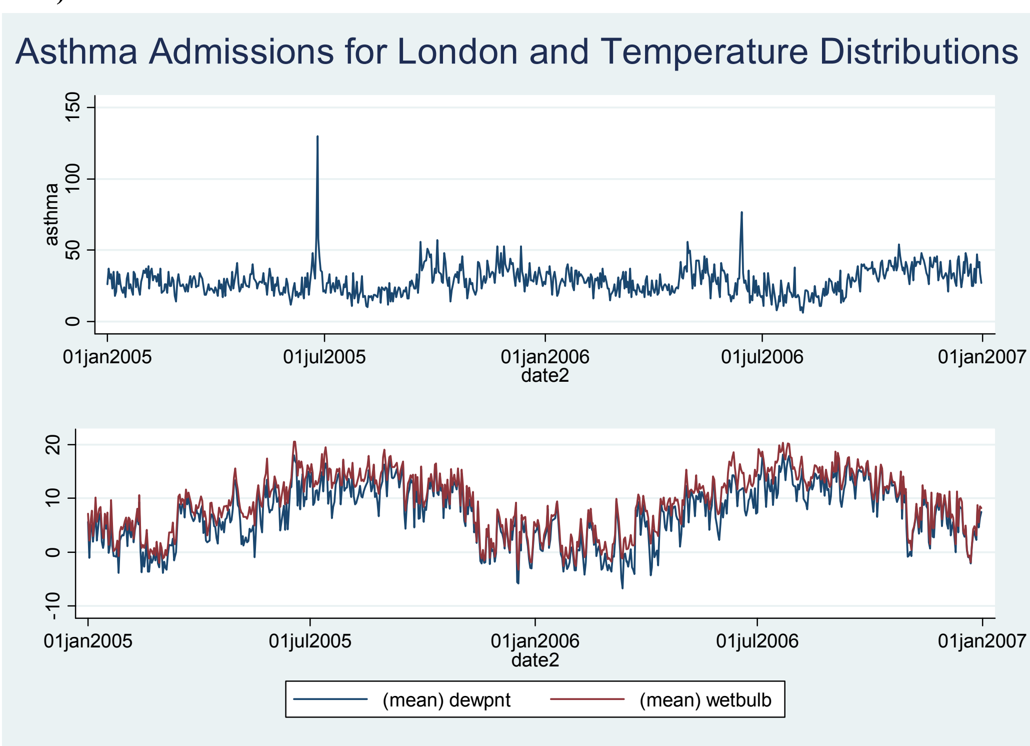


Figure 13 Asthma Admissions and Mean Night Minimum/ Night Maximum and Dew point/ Wet bulb Temperature Distributions (London, 2005-2006)

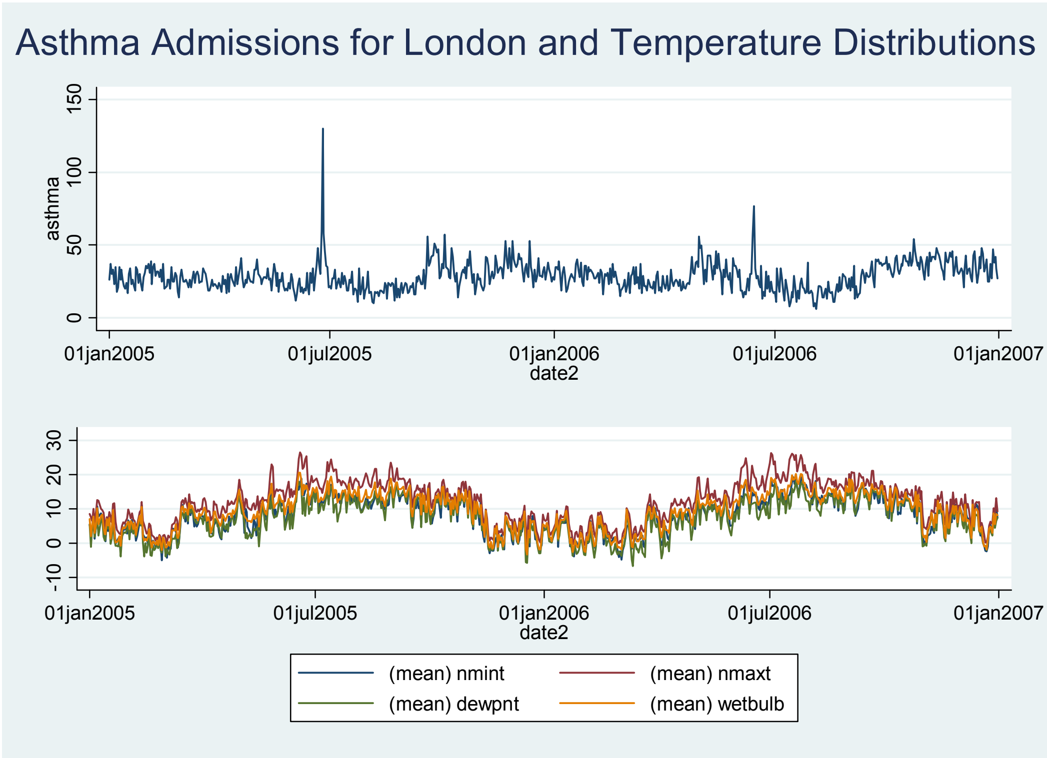


Figure 14 Asthma Admissions and Humidity Distributions (London, 2005-2006)

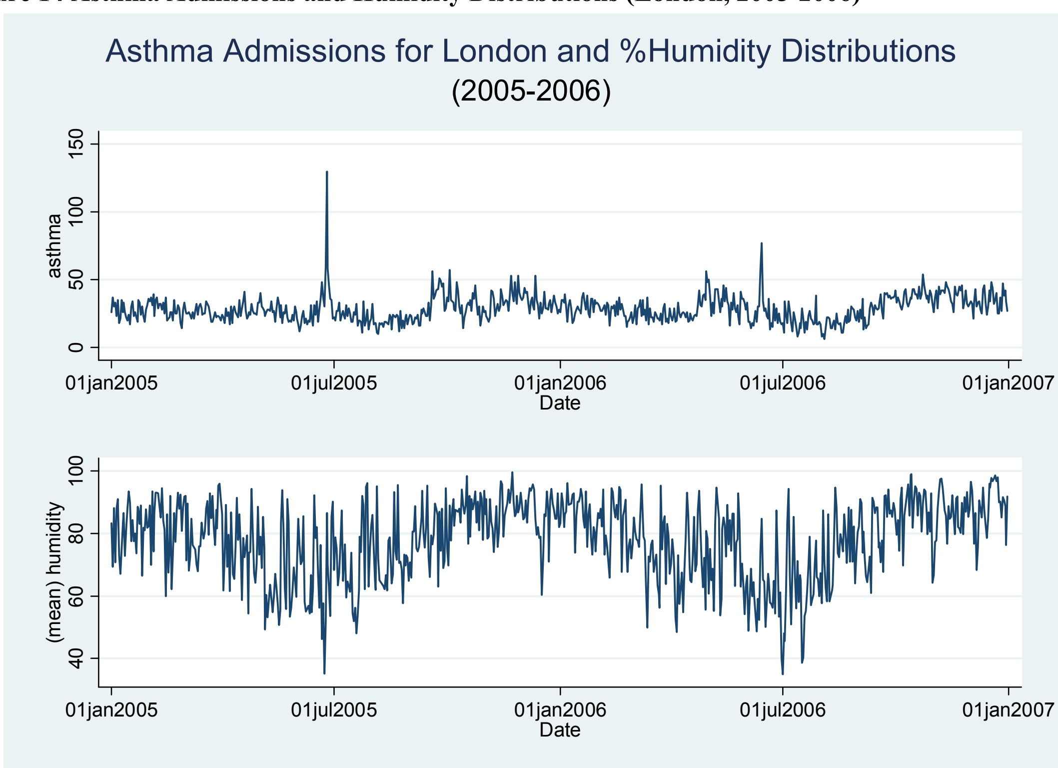


Figure 15 Asthma Admissions and Mean Wind speed Distributions (London, 2005-2006)

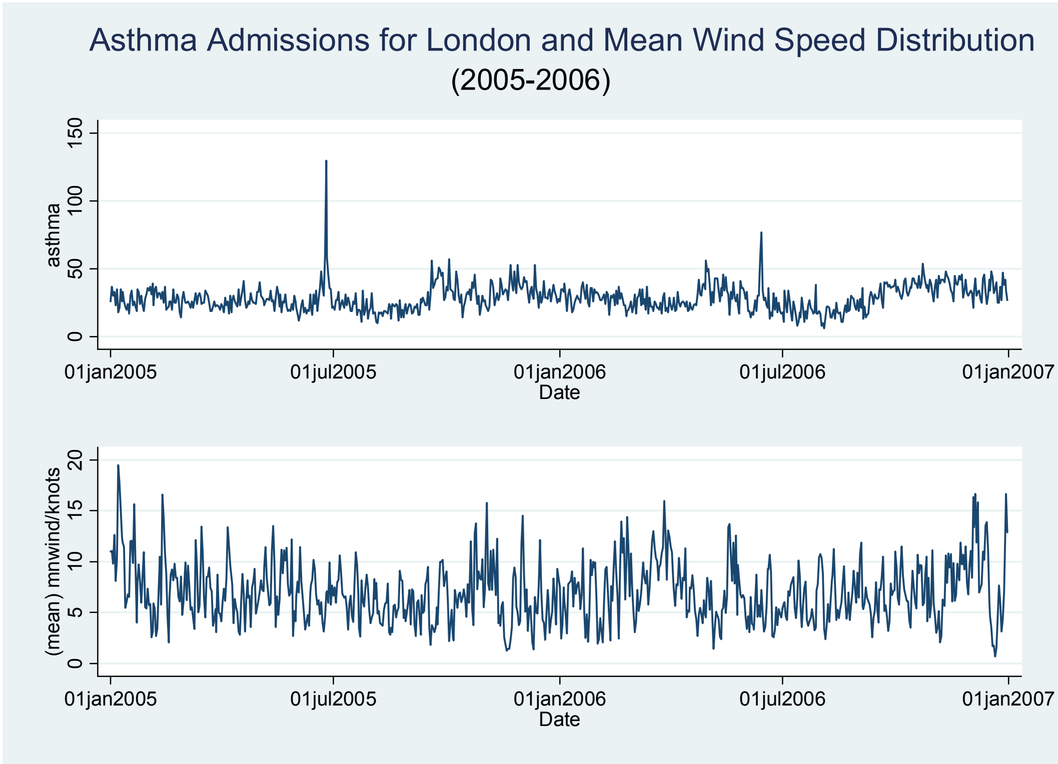
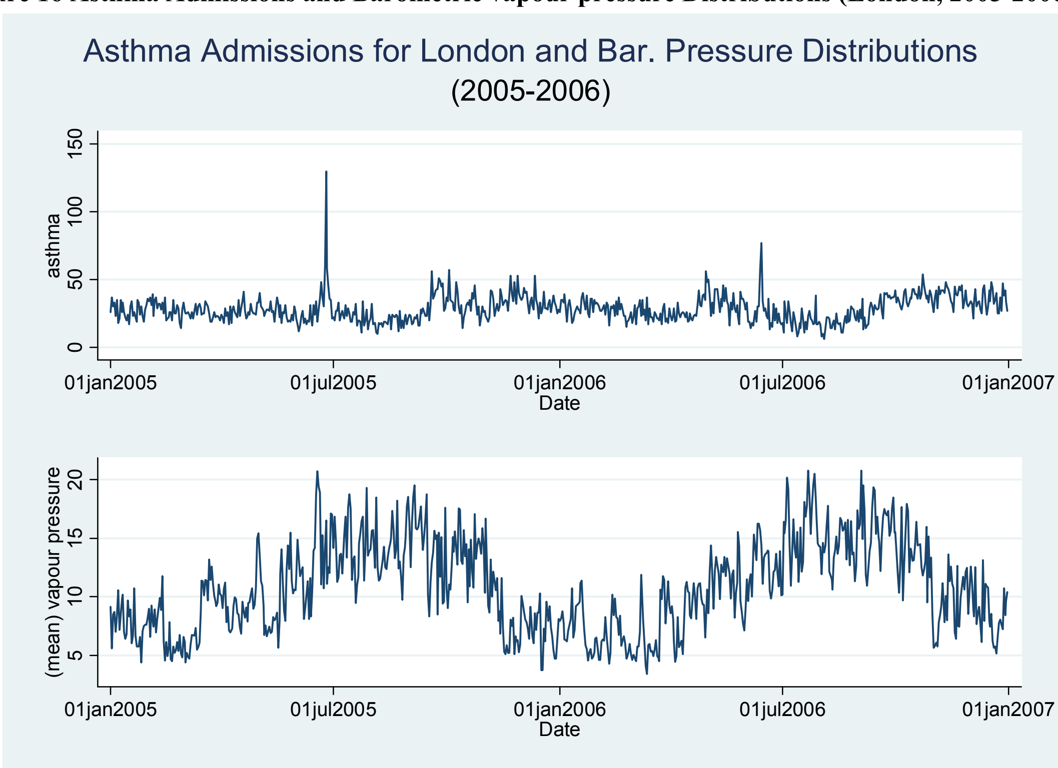


Figure 16 Asthma Admissions and Barometric vapour pressure Distributions (London, 2005-2006)



C1.7 Asthma Admissions and Air Quality (London, 2005-2006)

Figure 17 Asthma Admissions and Ozone Distribution (London, 2005-2006)

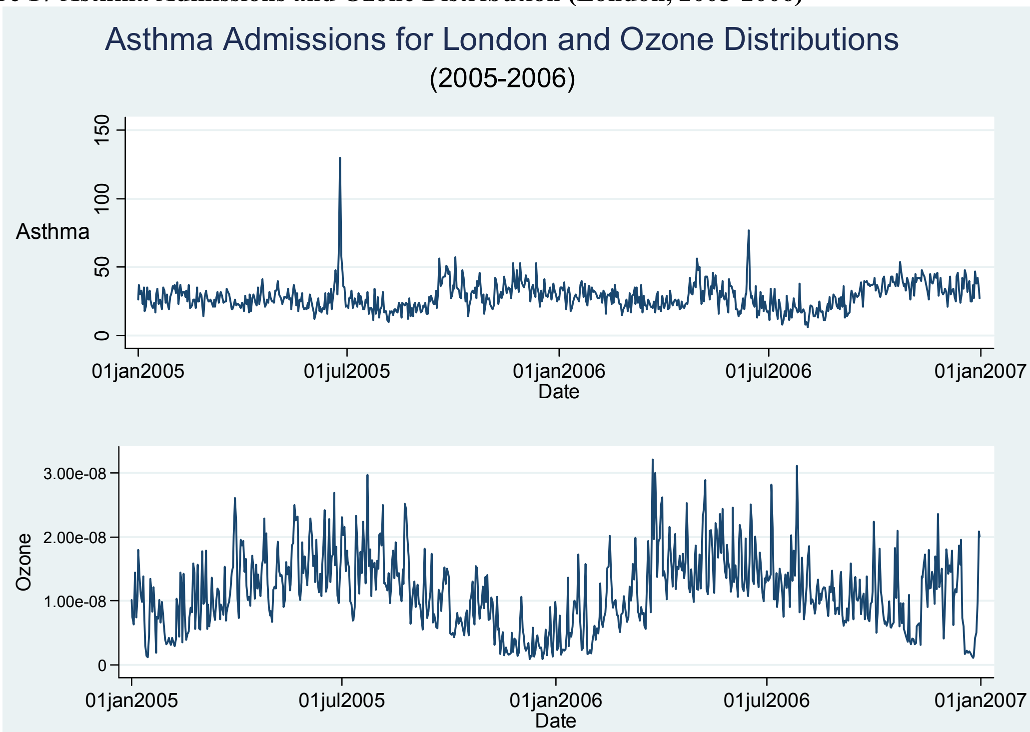


Figure 18 Asthma Admissions and Carbon monoxide Distribution (London, 2005-2006)

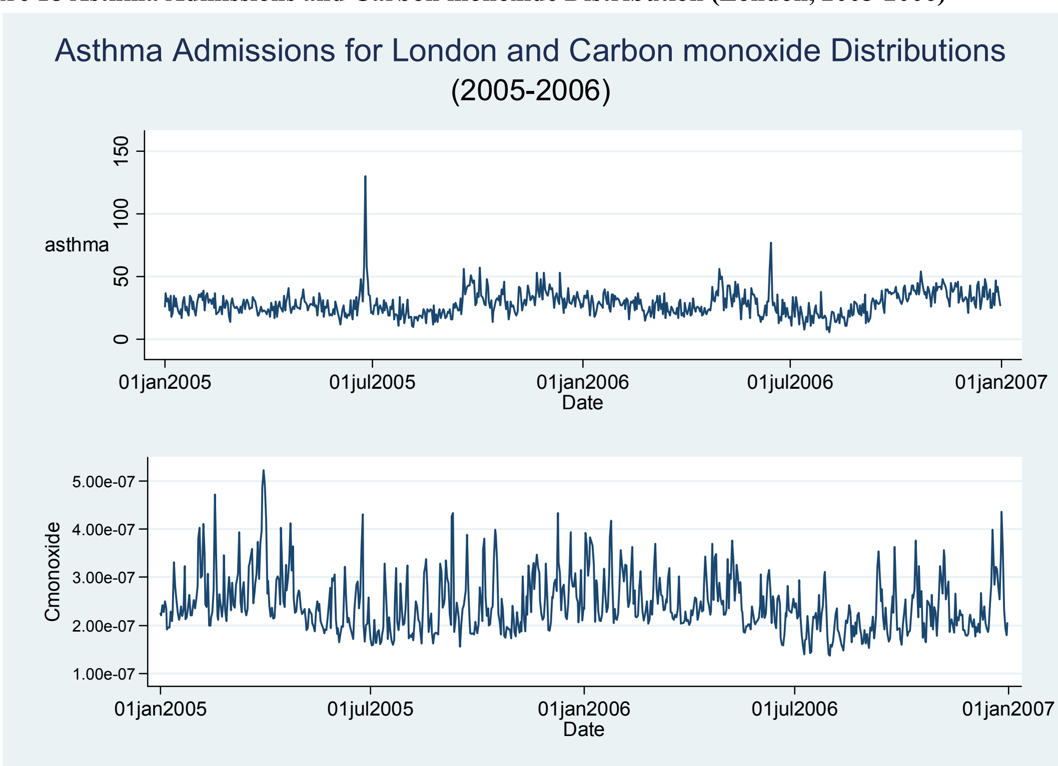
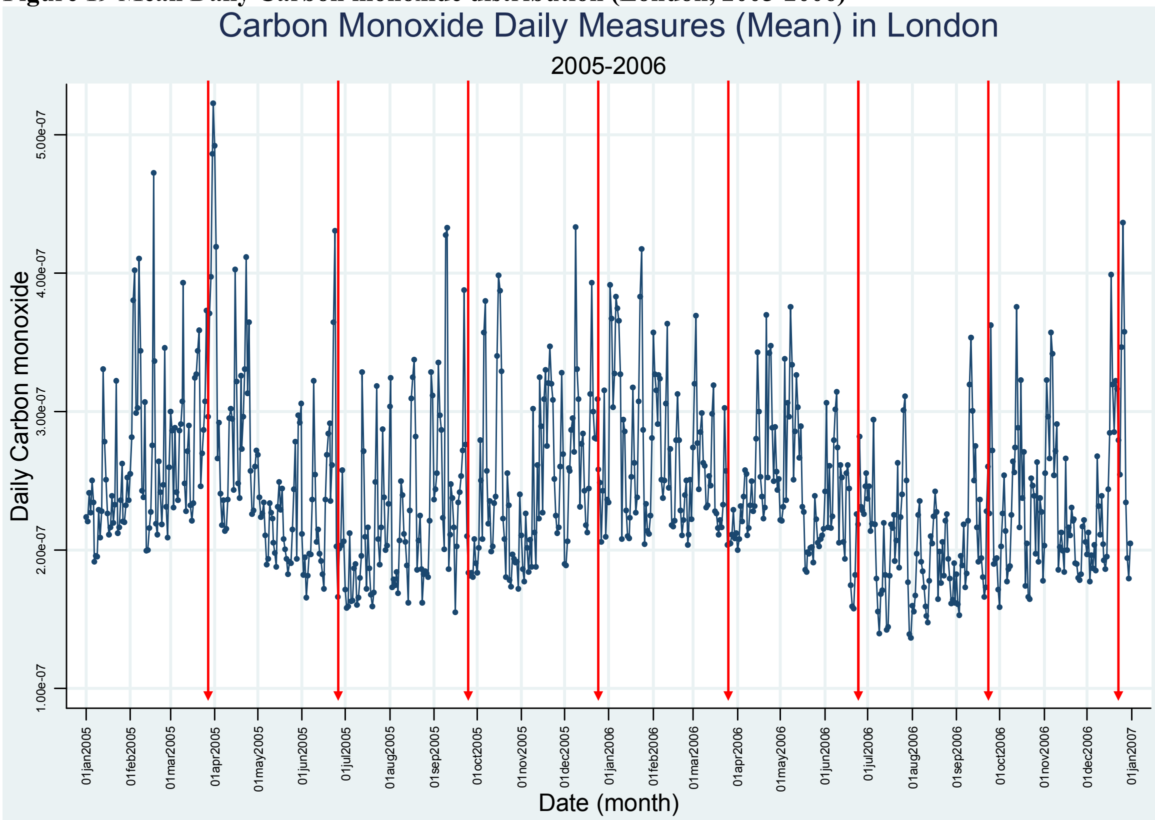


Figure 19 Mean Daily Carbon monoxide distribution (London, 2005-2006)



Arrow marks /divisions represent the seasons in a year

Figure 20 Asthma Admissions and NO Distribution (London, 2005-2006)

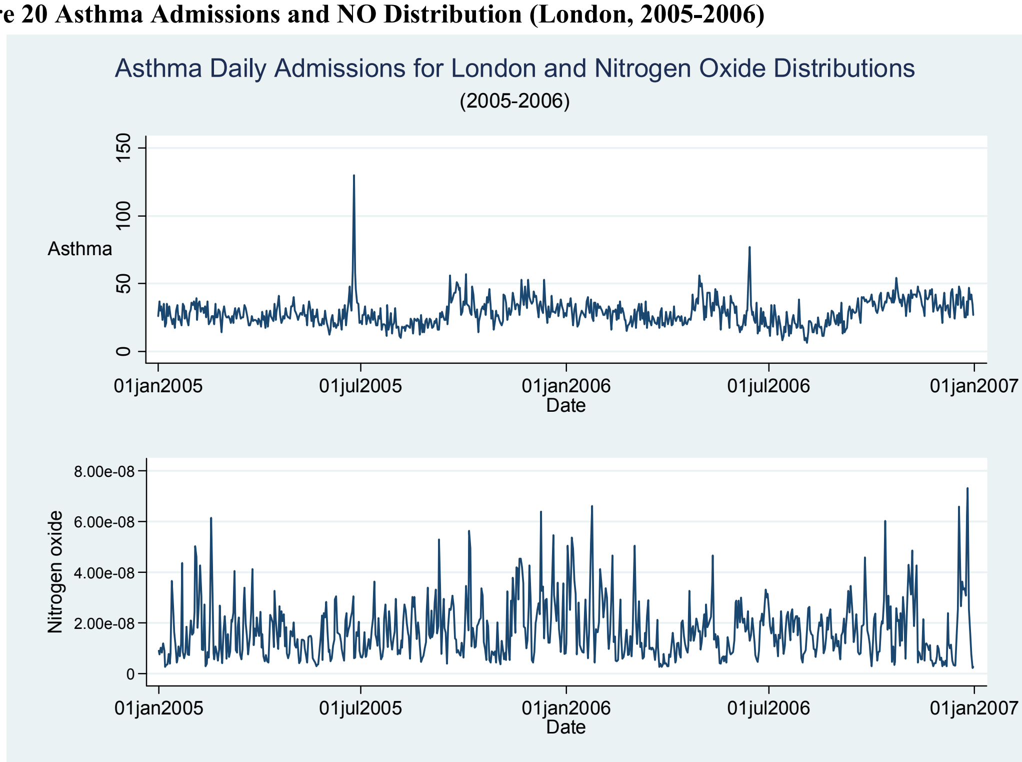


Figure 21 Asthma Admissions and Nitrogen dioxide Distribution (London, 2005-2006)

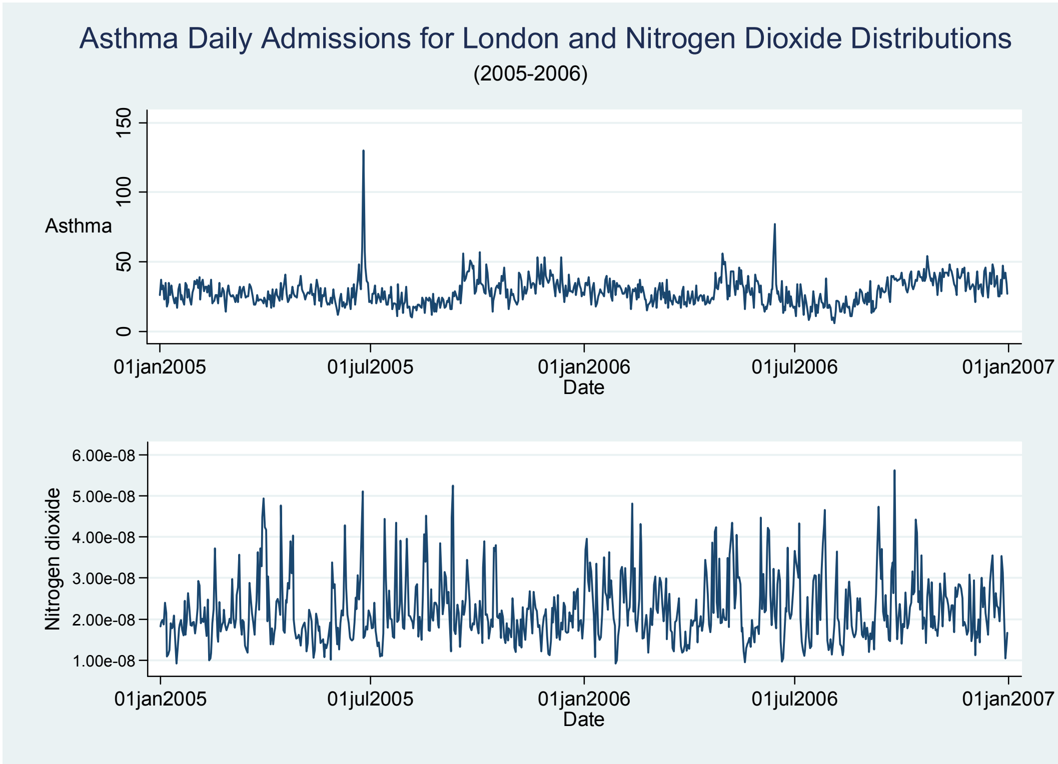


Figure 22 Asthma Admissions and Particulate matter Distribution (London, 2005-2006)

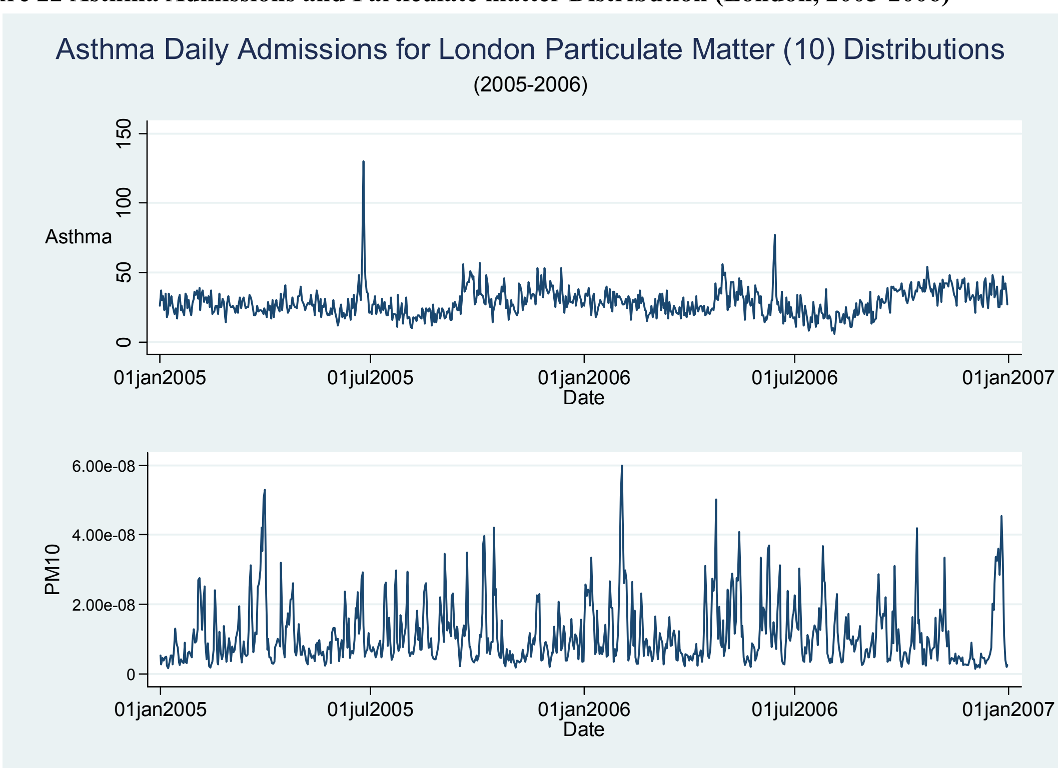


Figure 23 Asthma Admissions and Sulphur dioxide Distribution (London, 2005-2006)

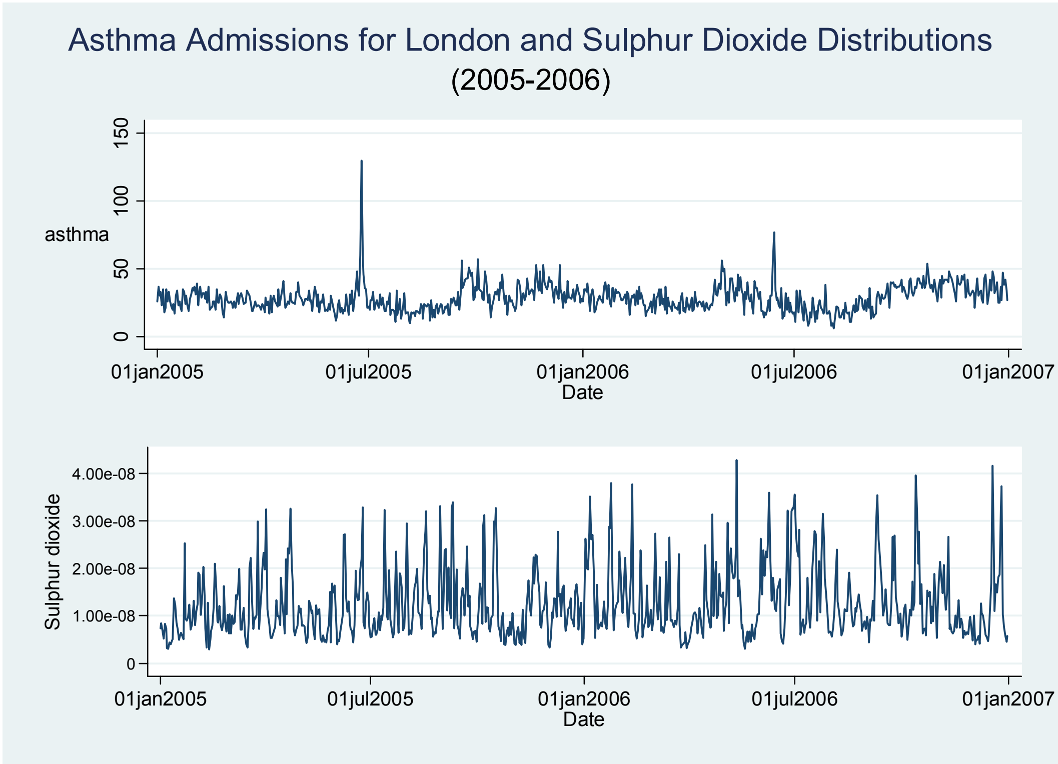
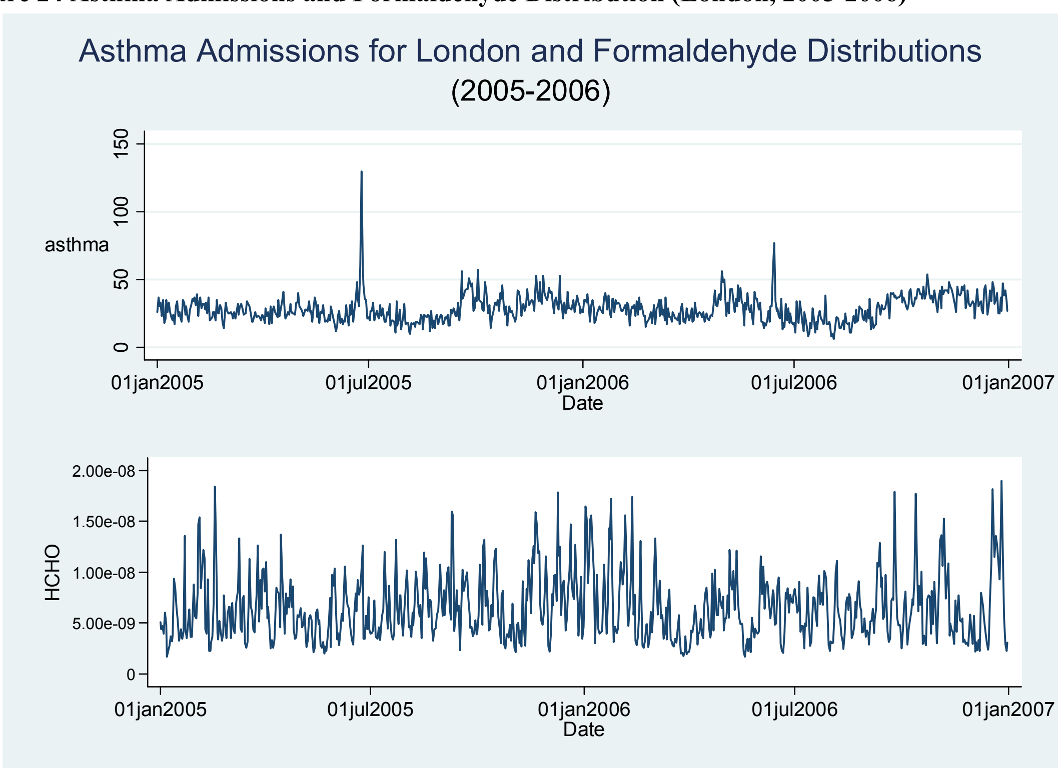


Figure 24 Asthma Admissions and Formaldehyde Distribution (London, 2005-2006)



C2.0 Notes on Bivariate Analysis

C2.1 Exploring Lag Days for the Explanatory Factors

In the preliminary data exploration, we examined lags from 1 to 21 days for each key independent variable through a bivariate test with asthma hospital admission as a dependent variable. The best fit bivariate lag-day was selected for further exploration in modelling collective effect(s) of the independent variables. The following were selected for consideration (Tab 2):

C3.0 Notes on Multivariable Analysis

C3.1 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 environmental factors that have an impact on health conditions like asthma are not frequently instantaneous, but rather cumulative in nature. Hence we 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

C3.2 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.

We then 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 PM_{10}, and a 21-day lag PM_{10}. 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 PM_{10} ($p=0.024$), and a 21-day lag PM_{10} ($p=0.03$) 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 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	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

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

Variable	Log change$^{\Psi}$	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

$^{\Psi}$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$^{\Psi}$	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

$^{\Psi}$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: -2035.88; **Chi2**: 86.03.

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.

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.

C4.0 Negative binomial regression Asthma Models

Table 7 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 8 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 9 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 10 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 11 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 12 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.