

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

Ireneous N Soyiri^{1,2}
Daniel D Reidpath¹

¹Global Public Health, School of Medicine and Health Sciences, Monash University, Malaysia; ²School 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

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 Campaign²⁵ 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

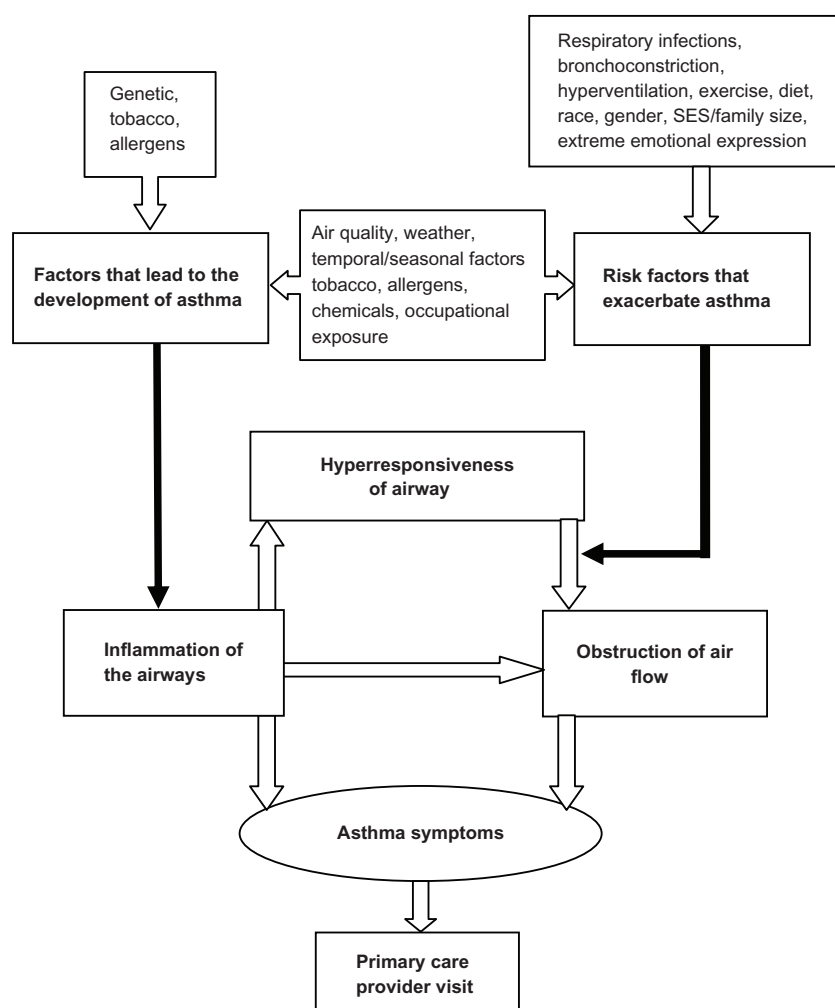


Figure 1 Factors involved in asthma manifestation.

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.

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

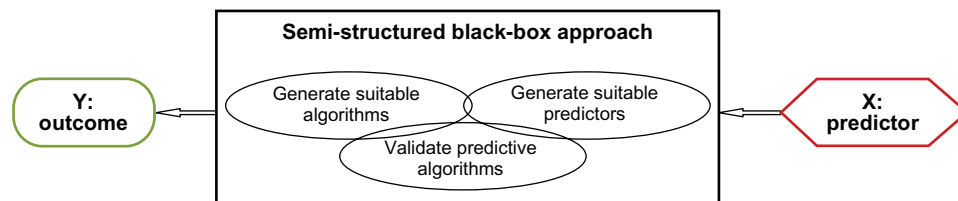


Figure 2 Schematic presentation of semistructured black-box modeling.

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.

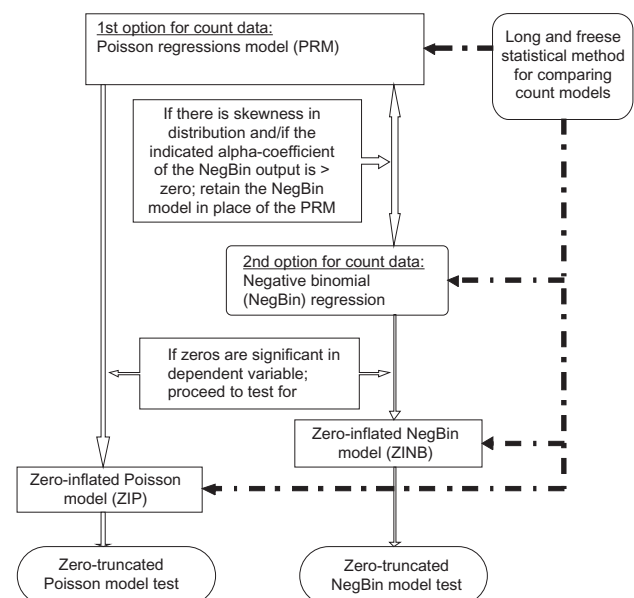


Figure 3 Decision tree for selecting an appropriate count model(s).

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:

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

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}

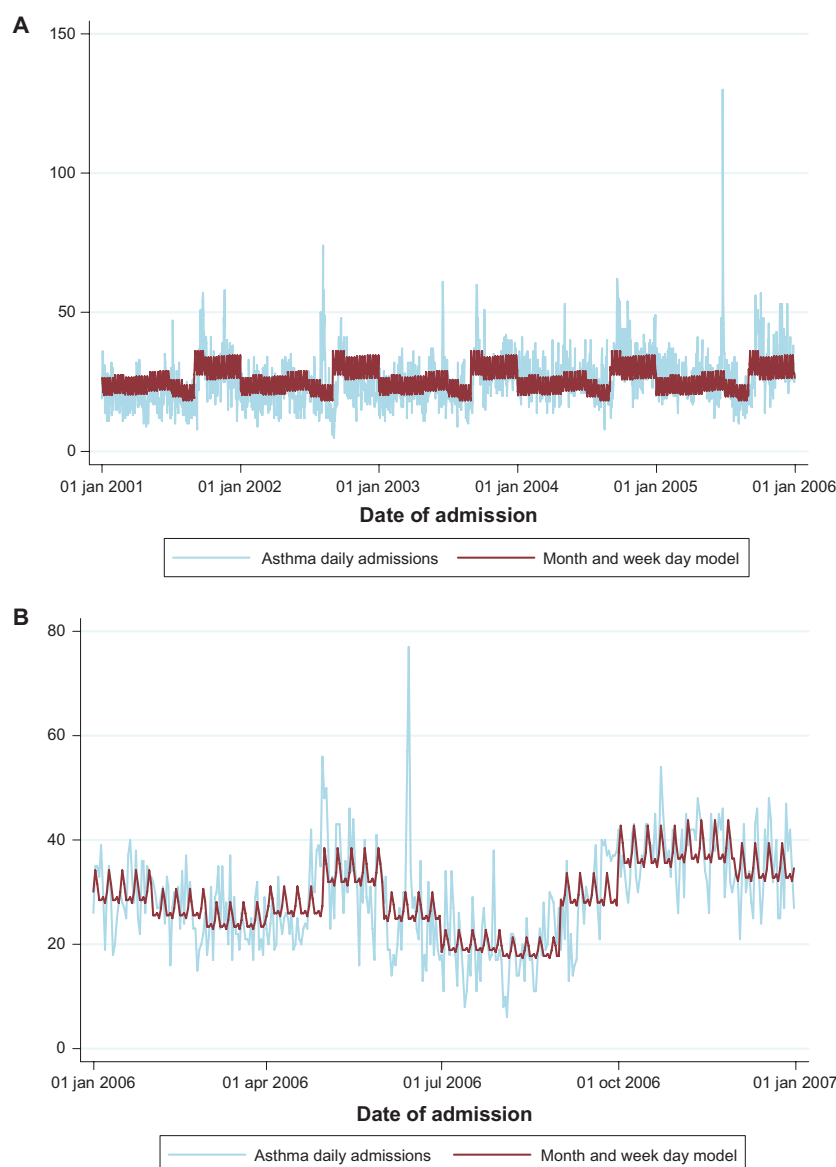


Figure 4 Asthma daily admissions and predictive model based on month and week day. (A) Model development sample (hold-in dataset). (B) Model validation sample (hold-out dataset).

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

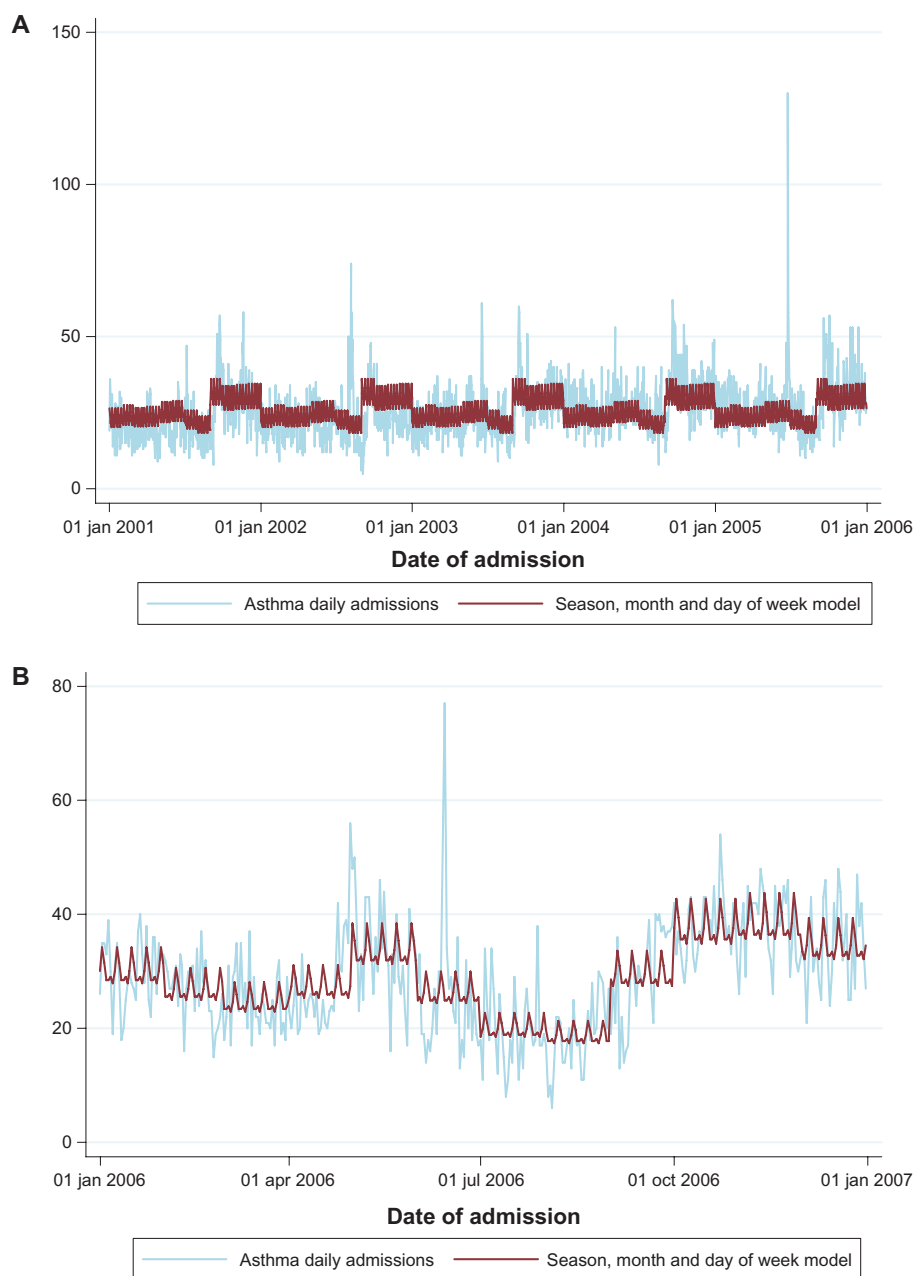


Figure 5 Asthma daily admissions and predictive model based on season month and week day. (A) Model development sample (hold-in dataset). (B) Model validation sample (hold-out dataset).

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:

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

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:

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

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
Autumn	0.1819***	0.1141	0.2498	0.2370***	0.1979	0.2760	0.1823***	0.1169
Winter	−0.0733*	−0.1424	−0.0042	0.0758***	0.0360	0.1156	−0.0731*	−0.1399
Month								
January [#]								
February	0.0358	−0.0352	0.1069				0.0351	0.1037
March	−0.0668	−0.1359	0.0022				−0.0654	−0.1321
April	−0.0565	−0.1261	0.0131				−0.0572	−0.1244
May	(omitted)	–	–				(omitted)	–
June	0.1926***	0.1221	0.2630				0.1948***	0.1266
July	0.0809*	0.0102	0.1517				0.0789*	0.0105
August	(omitted)	–	–				(omitted)	–
September	0.0559	−0.0109	0.1227				0.0538	−0.0105
October	−0.0102	−0.0769	0.0565				−0.0109	−0.0751
November	(omitted)	–	–				(omitted)	–
December	0.2628***	0.1950	0.3307				0.2635***	0.1981
Week day								
Sunday [#]								
Monday							0.1385***	0.0886
Tuesday							0.0513*	0.0010
Wednesday							−0.0335	−0.0843
Thursday							−0.0674**	−0.1184
Friday							−0.0669**	−0.1179
Saturday							−0.1259***	−0.1772

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?i1=2&i2=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, Mandhara 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, Morpew 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, Breyse 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, Héguay 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. Héguay 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, Paliatatos 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&categoryID=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. Raclou 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.

International Journal of General Medicine**Dovepress****Publish your work in this journal**

The International Journal of General Medicine is an international, peer-reviewed open-access journal that focuses on general and internal medicine, pathogenesis, epidemiology, diagnosis, monitoring and treatment protocols. The journal is characterized by the rapid reporting of reviews, original research and clinical studies across all disease areas.

Submit your manuscript here: <http://www.dovepress.com/international-journal-of-general-medicine-journal>

A key focus is the elucidation of disease processes and management protocols resulting in improved outcomes for the patient. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.