Int J Biometeorol (2013) 57:569–578
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 / Published online: 12 August 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

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 μg/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

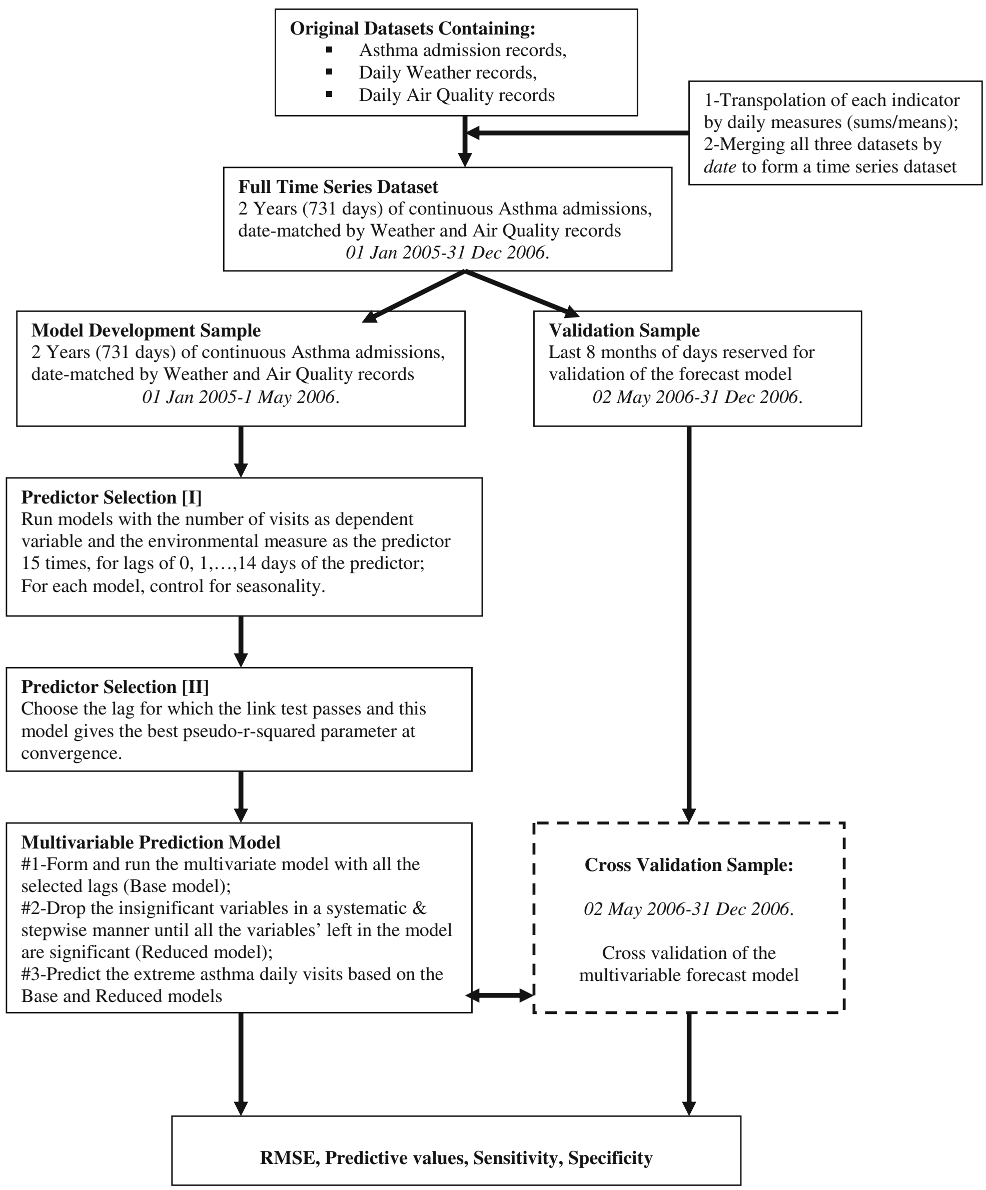

Fig. 1 Decision tree for developing quantile regression model (QRM) forecasting models for peak asthma events using temporal, weather and air quality factors

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

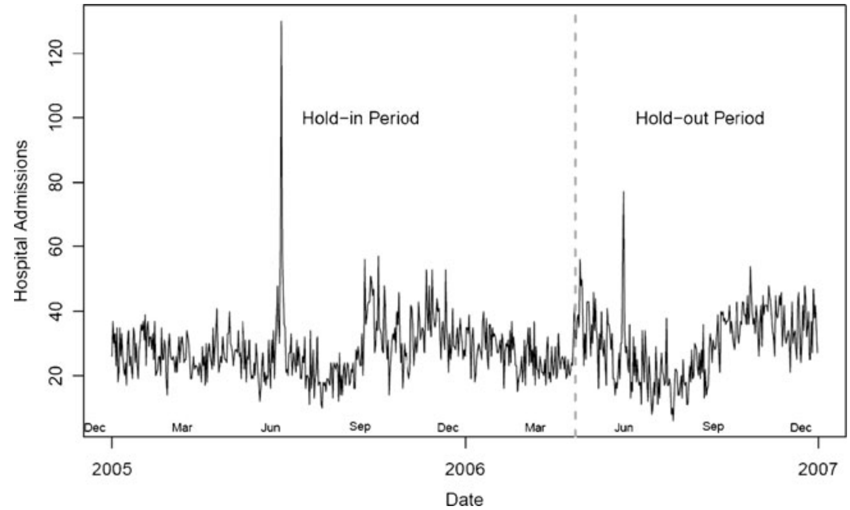

Fig. 2 Asthma admissions in London (2005–2006)

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/m^3)	−270.28*	−542	1.4062
3-day lag carbon monoxide (mg/m^3)	30.16**	6.0797	54.2365
4-day lag nitrogen dioxide (μg/m^3)	228.24**	75.493	380.99
13-day lag nitrogen oxide (μg/m^3)	116.81*	11.066	222.56
Four day lag PM_{10} (μg/m^3)	137*	20.406	253.59
13-day lag formaldehyde (μg/m^3)	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

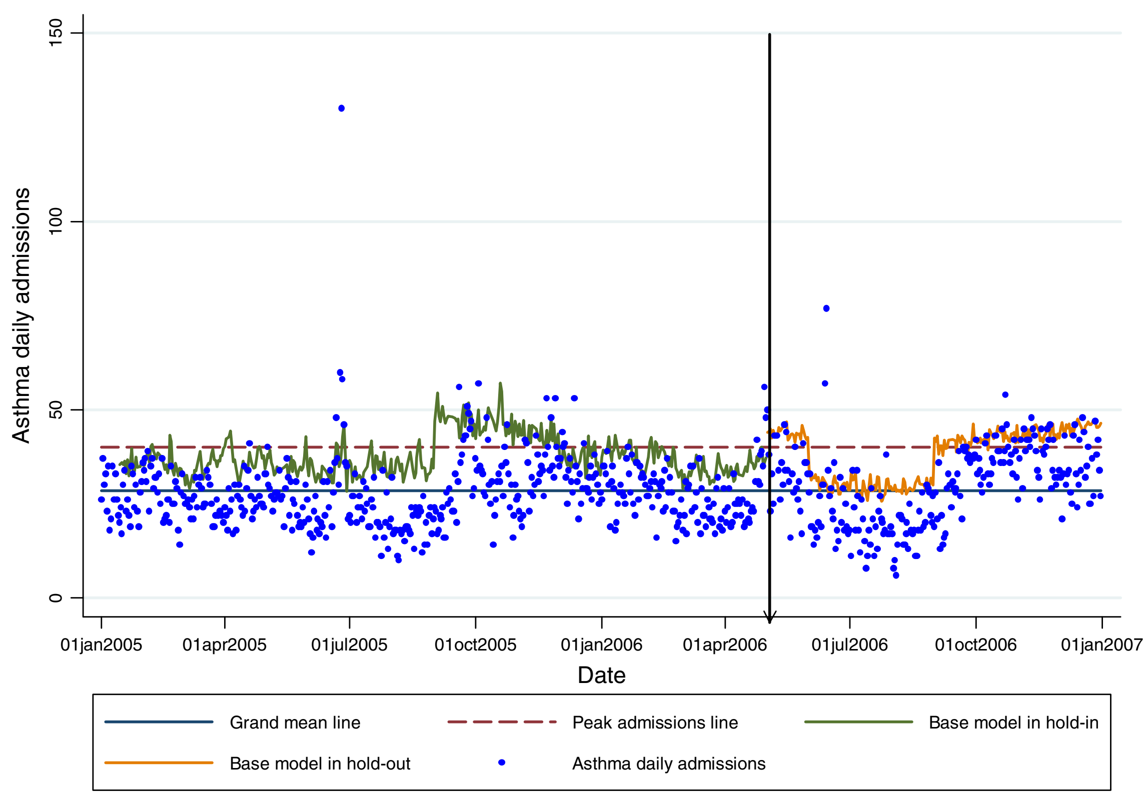

Fig. 3 Base model prediction of peak asthma admissions using QRMs: London (2005–2006)

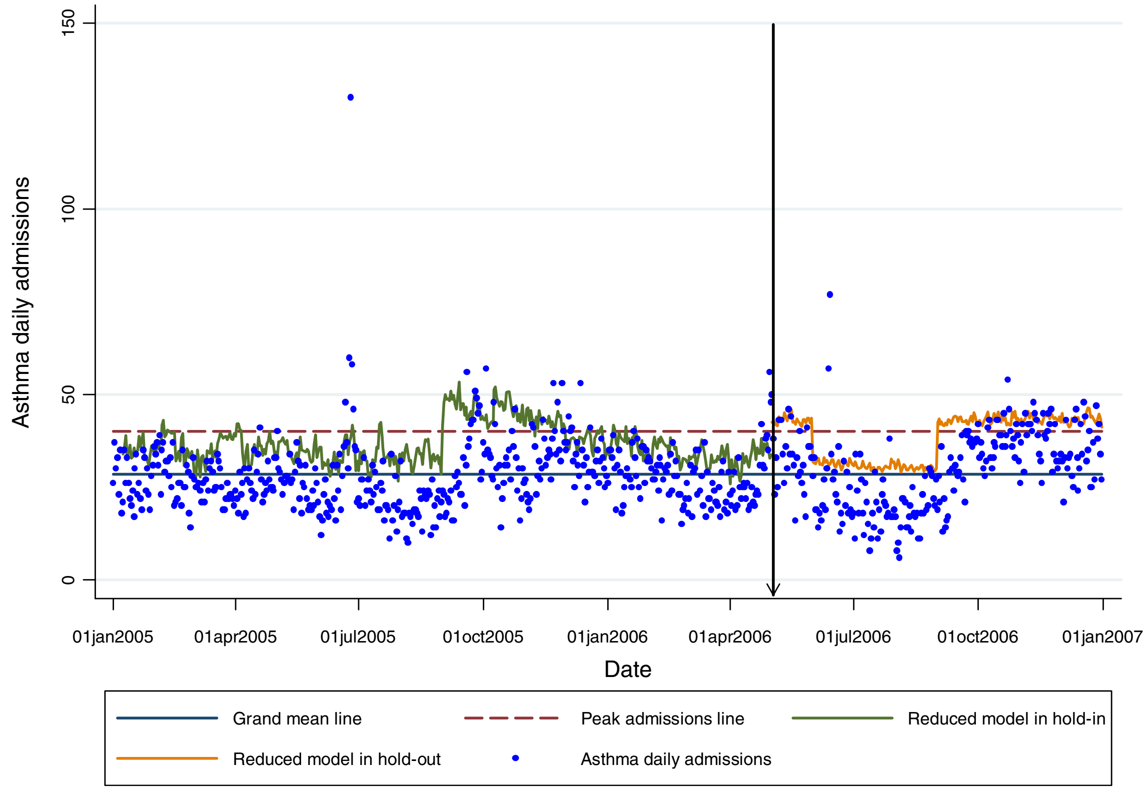

Fig. 4 Reduced model prediction of peak asthma admissions using QRMs: London (2005–2006)

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