Connected Health Cities – End of Project Report:

Workforce Development:

Spatio-temporal modelling of incidence in COPD emergency admissions in an area of North West England from 2012 to 2018

Contents:

20th April 2020

Abstract:

Chronic Obstructive Pulmonary Disease (COPD) is one of the leading causes of mortality worldwide with an estimated three million deaths in 2015, corresponding to 5% of all deaths globally.

Its exacerbation is a major contributor to the number of emergency admission and hospitalization in the UK.

COPD is the second most common cause (after a heart attack) of admission to a medical ward in the UK - ie. it's a huge cost burden and there is a belief that many cases could be prevented, hence the interest in predictions.

In this study, we pursue two objectives:

1) to assess the relative contribution of socio-economic and environmental variables for forecasting COPD emergency admissions

2) to develop a reliable early warning system that triggers an alarm whenever COPD emergency admissions signal the likely exceedance of predefined incidence thresholds

We developed a predictive model using a class of generalised linear mixed model. We selected the best predictors using the root mean square error (RMSE). We developed an early warning system based on exceedance probabilities.

The resulting predictors from our model selection are: minimum temperature; PM10; income deprivation; proportion of males; and proportion of the population aged above 75 years.

We found that, overall, the selected predictor variables explain about 22% per cent of the variability in the residual random effects. Among these variables, income deprivation attained the largest relative variance reduction of about 14%.

Our results demonstrate how to develop a predictive model as well as an early warning system for COPD emergency admission.

Our model has the potential to predict correctly in most areas with high sensitivity and specificity. The early warning system can potentially help to: identify and notify areas of high incidence of COPD emergency admission; and inform resource allocation for the healthcare system.

Keywords: *COPD; emergency admission; early warning system; variogram; geostatistics; generalised linear mixed model; prediction model.*

Introduction:

COPD is one of the leading causes of mortality worldwide[1] with an estimated three million deaths in 2015, corresponding to 5% of all deaths globally[2].

Its exacerbation is a major contributor to the number of emergency admission and hospitalisation, especially during the winter months as a result of the increase in respiratory viral infections.

Predictive models have been developed in several studies to identify patients at high risk of COPD exacerbations[3] which add significant cost to the patients care.

Hence, being able to accurately predict their occurrence can be especially useful to reduce avoidable COPD emergency admissions by targeting patients in most need.

In order to develop a reliable and scalable predictive model for COPD emergency admissions, the availability of comprehensive health records of patients is essential.

Predictive power can be further enhanced by incorporating risk factors concerning lifestyle behaviour (eg. smoking status), income, exposure to pollutants and other individual traits.

However, such detailed information may not be readily available to researchers due to confidentiality issues or because it has not been collected.

Notwithstanding, statistical modelling provides solutions that can be used to alleviate this issue. For example, generalised linear mixed models (GLMMs) are an extension of the classical generalised linear modelling framework that allows accounting for the unavailability of risk factors through the use of so-called random effects.

However, the full potential offered by this modelling framework has not been fully exploited in the analysis of COPD data and, in this work, we aim to fill this gap.

We argue further that instead of predicting the mean risk, statistical modelling should aim to predict the exceedance of clinically relevant thresholds beyond which COPD risk is of public health concern.

In our analysis of COPD admissions, we pursue two specific objectives:

1) to assess the relative contribution of socio-economic and environmental variables for forecasting COPD emergency admissions

2) to develop a reliable surveillance system that triggers an alarm whenever COPD emergency admissions signal the likely exceedance of predefined incidence thresholds. To the best of our knowledge, this is the first study that attempts to achieve these objectives using state-of-the-art spatio-temporal statistical methods for the analysis of data on COPD emergency admissions

Methods:

2.1 COPD admission data

Using the International Classification of Diseases (ICD) code (10th revision), J44 for COPD, we extracted monthly counts of COPD emergency admissions for patients above 18 years old living in the LA postcode area, covering parts of South Cumbria and North Lancashire in England.

The total population of the study region was 272,520 based on the 2011 census. The data cover the period from 1 April 2012 to 30 March 2018.

To protect confidentiality and anonymity of the patients, spatial information on their place of residence was provided at the Lower Super Output Area (LSOA).

From the same database, we also obtain the proportion of people older than 75 years and the proportion of male patients admitted, for each at LSOA-level.

2.1.1 Environmental and Socio-economic variables

In Table 1, we present the set of predictors used in our analysis, including weather, pollution and deprivation data.

Some of these variables are available on 1 × 1 km2 resolution, and for the analysis, we computed the population-weighted average of the variables over the LSOA.

Table 1. The table showing the set of predictors available for this study.

Domain	**Variables**
Weather	Minimum temperature; relative humidity; and number of days of ground frost.
Pollution	PM10 SO2; and NO2 (micrograms per cubic metre, gm-3)
Deprivation	Income deprivation; employment deprivation; health deprivation and disability; education skills and training deprivation; barriers to housing and services; living environment deprivation; and crime deprivation.

2.2 Statistical modelling

Let Y_{it} denote the monthly COPD emergency admission count at LSOA i and month t. We then assume that the Y_{it}, conditionally on a random effect Z_{it}, follow a Poisson distribution with mean $m_{it}\lambda_{it}$, where it denotes the population at LSOA i and month t and it represents the monthly incidence of COPD emergency admission at given LSOA.

We define the log-linear model for the incidence it as $\lambda_{it} = exp\{d_{it}\beta + Z_{it}\}$ where d_{it} is a vector of covariates with associated regression coefficients β.

Finally, we assume that the Z_{it} are independent and identically distributed Gaussian variables with mean zero and variance σ^2.

In order to build our regression model, we select predictors within three domains that are known to affect COPD admissions: weather, pollution and deprivation.

The variables that we consider within each of these domains are listed in Table 1. As the variables within each group are highly collinear, our goal is to select the best predictor from each group. In addition to the variables of Table 1, we include proportion of males and proportion of the population aged above 75 years as background predictors at LSOA-level of the incidence of COPD emergency admission.

To carry out the selection of the best predictors, we split the dataset into training and test sets, with the former covering the months from April 2012 to March 2017 and the latter from April 2017 to March 2018.

We then fit 63 models obtained by combining one predictor from each domain of Table 1 and, for each of those, we compute the root-mean-square-error (RMSE) for the predicted COPD admissions incidence using the test set.

Results:

3.1 Spatio-temporal Analysis

By applying the variables selection procedure described in Section 2.2, our final set of predictors consists of minimum temperature, PM10, income deprivation.

Table 1 shows the relative variance reduction (RVR) of each predictor in the model.

We find that, overall, the selected predictor variables explain about 22% per cent of the variability in the residual random effects. Among these variables, income deprivation attained the largest RVR of about 14%.

We then predict the incidence of COPD emergency admission for April 2017 – March 2018 and classify each LSOA as being above or below an incidence threshold which we set to 12 per 100,000.

For this threshold, we found that the value of EP that maximizes the sensitivity and specificity of the early warning system was p=0.85, yielding a 72% sensitivity and 70% specificity.

We also found that the area under the curve of the final model was about 78% which indicates a satisfactory predictive performance.

Figure 1 shows the LSOA that were correctly and incorrectly classified based on our modelling approach.

Whilst it is evident that our model can potentially predict correctly in most LSOAs, there exist a very few LSOAs with incorrect prediction.

Table 1: The table showing the relative variance reduced by the predictors.

Predictors	**RVR (%)**
Minimum temperature	1.23
PM10	0.88
Income deprivation	14.33
Proportion over age 75	3.35
Proportion of male	0.05
All predictors	21.58

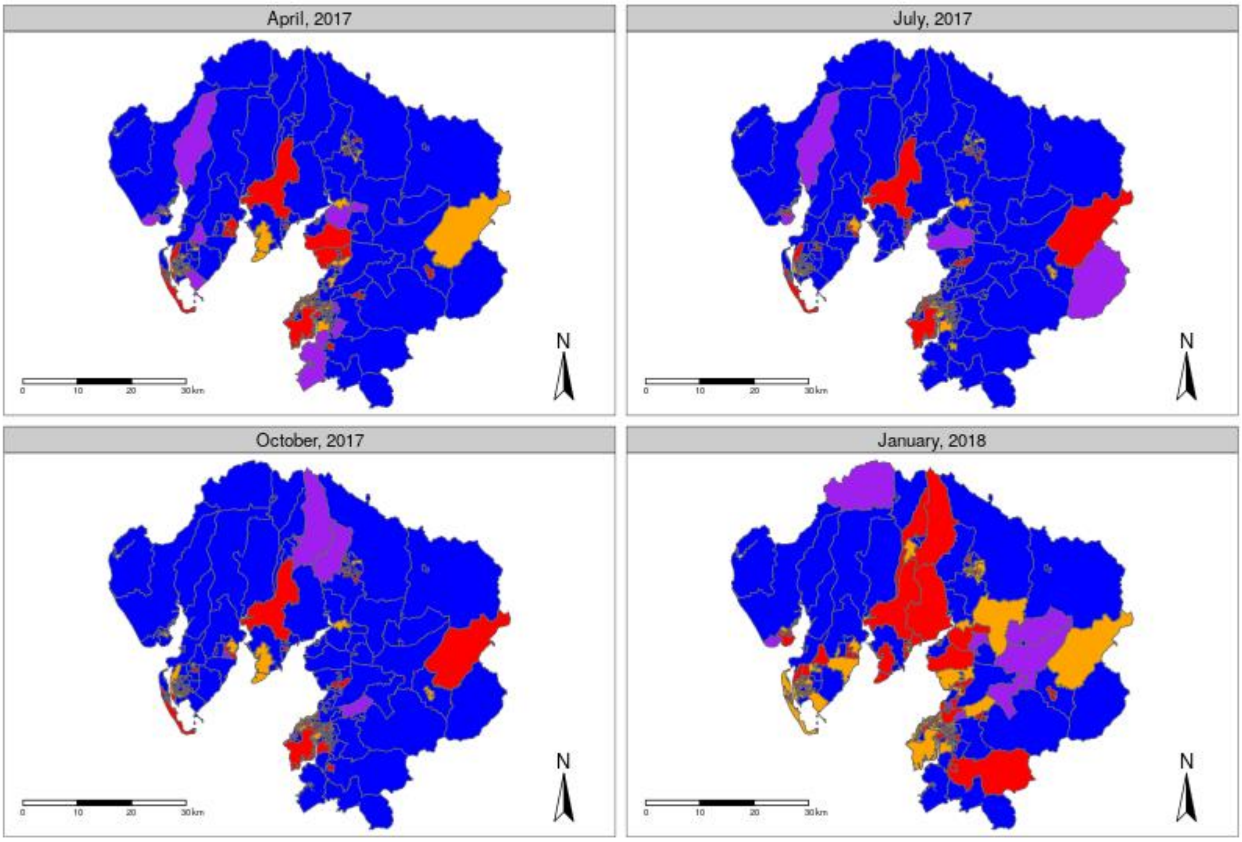

Figure 1: The monthly-predicted surveillance maps comparing the predicted and the true alarm for each LSOA. Colour blue indicates an LSOA that is correctly predicted to be below the threshold; orange indicates an LSOA that is correctly predicted to be above the threshold; purple indicates an LSOA that is incorrectly predicted to be below the threshold; and red indicates an LSOA that is incorrectly predicted to be above the threshold. The incidence threshold used is 12 per 100,000.

Conclusion/Discussion:

We have developed a predictive statistical model for the incidence of COPD in South Cumbria and North Lancashire district (North West England).

Our predictive model uses a combination of environmental and socio-economic variables as predictors. We also demonstrated that instead of predicting the incidence, a more meaningful prediction would be to predict the exceedance of clinically relevant threshold beyond which COPD risk is of public health concern.

Also, we found that income deprivation reduced the highest proportion of variance in the monthly incidence of COPD emergency admission. It has been shown in other studies that people who live in deprivation are more likely to be admitted[29]. This suggests that a good predictive model for incidence of COPD should take into account socio-economic status.

The model performs fairly well at predicting LSOA-level incidence of COPD emergency admission in the test set, however, there is clear room for improving the predictive accuracy. The value of the true positive rate is quite interesting which suggest that our model can potentially identify 71% of the high incidence LSOAs.

A good warning system model needs to achieve a balance between the sensitivity and specificity to avoid the waste of resources and identifying "real" high incidence LSOA. A warning system with high specificity is capable to detect LSOAs with "real" high incidence, but suffer losses from incurring additional resources due to low specificity.

Similarly, a warning system with high specificity benefits from a significant reduction in the consumption of resources but has a decreased capacity to detect "real" high incidence LSOA due to low sensitivity. However, our model has high sensitivity and high specificity suggesting a good balance.

Our results demonstrate how to develop a predictive model as well as a surveillance system for COPD emergency admission. The result of our analysis can potentially help NHS Morecambe Bay Clinical Commissioning Group stakeholders to define areas to target early intervention as well as inform resource allocation for healthcare system so that its limited resources can be used to maximum effect. Future studies will improve the model by accounting for some other known and unknown risk factors that are not captured in the study.

References:

1. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med*. 2006;3(11):2011-2030. doi:10.1371/journal.pmed.0030442
2. WHO. Chronic Obstructive Pulmonary Disease (COPD). World Health Organisation, fact sheet. http://www.who.int/mediacentre/factsheets/fs315/en/. Published 2016. Accessed February 6, 2018.
3. Billings J, Dixon J, Mijanovich T, Wennberg D. Case finding for patients at risk of readmission to hospital: Development of algorithm to identify high risk patients. *Br Med J*. 2006;333(7563):327-330. doi:10.1136/bmj.38870.657917.AE
4. Calderón-Larrañaga A, Carney L, Soljak M, et al. Association of population and primary healthcare factors with hospital admission rates for chronic obstructive pulmonary disease in England: National cross-sectional study. *Thorax*. 2011;66(3):191-196. doi:10.1136/thx.2010.147058

Authors:

Olatunji Johnson, Mike Pearson, Tim Gatheral, Peter Diggle, Jo Knight and Emanuele Giorgi

Lancaster University

20th April 2020