

Investigating trends in asthma and COPD through multiple data sources: a small area study

Areti Boulieri^{a, b}, Anna Hansell^{a, b} and Marta Blangiardo^{a, b}

^a*Department of Epidemiology & Biostatistics, Imperial College London, UK*

^b*MRC-PHE Centre for Environment & Health, Imperial College London, UK*

Abstract

This paper investigates trends in asthma and COPD by using multiple data sources to help understanding the relationships between disease prevalence, morbidity and mortality. GP drug prescriptions, hospital admissions, and deaths are analysed at clinical commissioning group (CCG) level in England from August 2010 to March 2011. A Bayesian hierarchical model is used for the analysis, which takes into account the complex space and time dependencies of asthma and COPD, while it is also able to detect unusual areas. Main findings show important discrepancies across the different data sources, reflecting the different groups of patients that are represented. In addition, the detection mechanism that is provided by the model, together with inference on the spatial, and temporal variation, provide a better picture of the respiratory health problem.

Keywords:

Space-time analysis, detection, asthma and COPD

1. Introduction

Asthma and chronic obstructive pulmonary disease (COPD) are the most common chronic respiratory conditions worldwide, contributing to heavy social and economic burden (World Health Organization, 2012).

The number of people suffering from asthma in 2014 was estimated to be 334 million around the world (World Health Organization, 2014) and this number is projected to rise to 400 million by 2025. Around 250 000 deaths per year are caused by the disease, with the majority of them considered to be preventable (Masoli et al., 2004). COPD has a lower prevalence of 64 million people but much higher mortality, with 3 million deaths annually, an estimated 6% of all deaths worldwide. COPD is predicted to become the third leading cause of death by 2030 (World Health Organization, 2014). In the UK, asthma affects 1 in 5 households, and COPD is the

fifth leading cause of death after cancer and cardiovascular disease (Masoli et al., 2004).

Asthma and COPD have similarities in symptoms and treatment and there may be considerable overlap between these conditions making them difficult to distinguish clinically (Drazen et al., 2015). Asthma commonly starts in childhood and is often allergic in origin, while a large proportion of COPD is caused by smoking and the condition starts in mid to later life. Common symptoms in both conditions are shortness of breath and wheeze, with worsening of symptoms with respiratory infections, with similarities in treatments of bronchodilators, steroids and antibiotics for infections. Triggers for exacerbations, the major determinants of admissions and possibly deaths are likely to be influenced by infectious disease trends and by common environmental factors with a spatio-temporal structure, such as air pollution (Eeftens et al., 2012).

A study of trends of these chronic respiratory diseases is important, as it allows a better understanding of the characteristics of the disease, to determine whether health policies or preventive measures are effective, and to identify high-risk populations that might require additional care and treatments. A challenge, however, is the choice of data to use for asthma and COPD analyses. Studies on mortality (Sin et al., 2006) investigate only the highest degree of severity and findings may differ from those for hospital admissions particularly for asthma (Hansell et al., 2003), which is a more heterogeneous condition. Most studies have used data from secondary and tertiary care, such as hospital admissions and emergency care data, but these will not capture milder cases seen in primary care. It is estimated that only 20% of asthmatic patients and less than half of COPD patients suffer from severe symptoms (Lindebeg et al., 2006).

A useful addition to asthma and COPD research is the use of General Practice (GP) drug prescription data which consist in the numbers of items that are prescribed in England by GPs and are dispensed anywhere in UK or Europe. GP drug prescriptions can be very relevant for asthma and COPD as these are long-term conditions that are controlled by regular medication. These capture patients of any severity of the disease, from mild to severe, and hence they can provide a general picture of the respiratory health of the population at small area level. Only a few authors have used GP drug prescriptions to investigate asthma and COPD trends (Hansell et al., 2003; Naureckas et al., 2005; Vegni et al., 2005; Laurent et al., 2009; Sofianopoulou et al., 2013).

In addition, the geographical trends of asthma and COPD have only been studied by a few authors. Hansell et al. (2003) found COPD mortality, hospital admissions and GP prescriptions for COPD were higher in urban areas and northern regions of England, but less clear patterns were seen for asthma in comparisons using age-sex standardised event ratios. Holt et al. (2011) analysed hospitalisations within a Bayesian hierarchical framework, while other examples include Joo et al. (2007); Centers for Disease Control and Prevention (2008); Lipton and Banerjee (2006) and

Nandram et al. (2000). Sofianopoulou et al. (2013) explored geographical patterns of GP drug prescriptions, considering the Newcastle and North Tyneside area in the UK as a study region. Studies (Hacking et al., 2011; Wells and Gordon, 2008; Hansell et al., 2003) suggest that there is a significant difference of morbidity and mortality within and between regions of the UK over the last 40 years. This needs to be taken into account in order to both provide reliable statistical estimates, as well as to help public health policy makers more clearly identify target areas with great needs and improve disease prevention and treatment.

The objective of this study is to investigate trends in asthma and COPD at the population level across England by using multiple data sources to help understanding the relationships between disease prevalence, morbidity and mortality. We explore spatial and temporal patterns of GP drug prescriptions, hospital admissions and deaths, and we evaluate if different behaviours can be seen for different data sources which underline different condition severity. We also focus on the detection of unusual areas, i.e characterised by a temporal trend which deviates from the general one, suggesting the presence of a policy or an emerged localised factor. In this analysis we combine information on asthma and COPD, given the similarities in these conditions, issues distinguishing between them (Drazen et al., 2015) and the fact that the GP prescription dataset used does not provide information on diagnosis.

The remainder of the paper is structured as follows. Section 2 describes the study design and the data sources used for the analysis, and Section 3 describes the statistical modelling framework. In Section 4 the results of the study are presented, followed by a discussion, and finally, Section 5 summarizes the main findings of the paper, and suggests recommendations for future research.

2. Data sources

To gain a better understanding of asthma and COPD, we make use of three different data sources: *i*) General practice (GP) drug prescription data of treatments used for these conditions, which capture patients with mild to severe symptoms and will give a general picture of the disease prevalence across the study region; *ii*) Hospital Episode Statistics (HES) admissions with primary diagnosis of asthma or COPD; *iii*) mortality data with asthma and COPD disease as cause of death. The latter two data sources will inform on cases characterised by higher severity. We are going to describe each data source in the rest of this section.

2.1. GP drug prescription data

The Prescription Cost Analysis (PACT) data are accessed from the NHS Business Services Authority. These include the monthly prescriptions of all drugs from 8003 general practices across England from August 2010 onwards at a monthly temporal resolution. In this study we use the prescriptions on Salbutamol, Ventolin and Clenil Modulite, with corresponding British National Formulary (BNF) codes

1011R0AAAPAP, 0301011R0BEAIAP and 0302000C0BPABBF respectively. These account for more than 90% of the total prescription of short acting beta2-agonist (SABA), a class of drugs that relieves patients from bronchospasm which characteristically occurs in acute symptoms (Drazen et al., 2015). Every GP is part of a local clinical commissioning group (CCG), which is the authority responsible for local healthcare services including local hospitals and NHS services, according to the 2012 Health and Social Care Act. GPs of the same CCG collaborate to evaluate local needs, monitor services, set priorities and make area-specific decisions to promote healthcare services for local residents. This suggests that GPs within the same CCG should share similarities. Therefore, the available GP data are aggregated at CCG (211 in England) level to be used for the analysis. The PACT data also contain the number of patients registered within each GP, with information on age group and sex. These are also aggregated at CCG level and they are used for the calculation of the expected number of drugs which will be the offset for the analysis of GP drug prescriptions.

2.2. HES and mortality data

Health data for England from August 2010 to March 2011 were obtained from the Small Area Health Statistics Unit (SAHSU) at Imperial College London. Hospital Episode Statistics (HES) admission data, supplied by the Health and Social Care Information Centre contain the number of admissions with a primary diagnosis of asthma or COPD, and may also include readmissions of the same patient. Mortality data were also obtained from SAHSU, supplied by the Office for National Statistics (ONS), derived from the national mortality registrations. The number of deaths with underlying cause of death (UCD) of asthma or COPD was collected. International Classification of Disease coding version 10 (ICD-10) was used for admission and mortality coding throughout this time period and asthma and COPD were defined as ICD-10 codes 490–496. Linkage between HES and mortality data to identify individual patients was not possible. Population data were also obtained from the Office for National Statistics, with individual-level information on age and sex and these were used for the calculation of the expected number of cases. HES, mortality, and population data were all aggregated to CCG level.

We have considered the same time period of August 2010 to March 2011 across all three datasets.

3. Statistical Analysis

The analysis is conducted within a Bayesian hierarchical framework that takes into account the complex dependence patterns of asthma and COPD over space and time. Bayesian methods have been extensively applied in epidemiological studies, in order to summarise the spatial and temporal variations of the disease risk (Best et al., 2005). Approaches within the spatio-temporal setting have been suggested by

many authors (Bernardinelli et al., 1995; Waller et al., 1997; Knorr-Held and Besag, 1998; Knorr-Held, 2000; Abellan et al., 2008).

The spatio-temporal model that we use in this paper, known as BaySTDetect, is a recently developed method by Li et al. (2012) that is able to estimate spatial and temporal patterns, and to also detect areas whose temporal pattern deviates from the general one.

3.1. Model specification

The first level of the hierarchical model is given by

$$Y_{it} \sim \text{Poisson}(\mu_{it} E_i) \quad (1)$$

where Y_{it} and E_i are the observed and expected counts in CCG $i = 1, \dots, 211$ at time points $t = 1, \dots, 8$, corresponding to months August 2010, September 2010, ..., March 2011.

In the second level of the hierarchy, the rate μ_{it} follows a mixture of two components as follows:

$$\log(\mu_{it}) = z_i \log(\mu_{it}^{\text{C}}) + (1 - z_i) \log(\mu_{it}^{\text{AS}}) \quad (2)$$

where

$$\log(\mu_{it}^{\text{C}}) = \alpha_0 + h_i + \gamma_t \quad (\text{Common Model}) \quad (3)$$

and

$$\log(\mu_{it}^{\text{AS}}) = u_i + k_{it} \quad (\text{Area-Specific Model}) \quad (4)$$

The Common Model (3) consists of spatial and temporal effects, h_i and γ_t respectively, that are combined additively on the log scale, thus estimating the temporal pattern to be the same for all areas. An overall intercept α_0 is also included.

To incorporate unusual temporal patterns that may occur in any particular month, we allow for the selection of an alternative Area-Specific Model (4), which estimates the temporal effects k_{it} independently for each area. An area-specific intercept u_i is also included in the model.

In the third level of the hierarchy, priors are specified for all model parameters as follows:

$$\begin{aligned} \alpha_0 &\sim \text{U}(-\infty, +\infty) & u_i &\sim \text{N}(0, 1000) \\ h_i &\sim \text{N}(v_i, \sigma_h^2) \text{ and } v_i \sim \text{ICAR}(\text{W}, \sigma_v^2) & k_{i,t} &\sim \text{ICAR}(\text{Q}, \sigma_{ik}^2) \\ \gamma_t &\sim \text{ICAR}(\text{Q}, \sigma_\gamma^2) & \log(\sigma_{ik}^2) &\sim \text{N}(\alpha, \beta^2) \end{aligned}$$

For the Common Model, a spatial convolution prior based on the standard Besag-York-Mollie formulation (Besag et al., 1991) is assigned to the spatial random effects. This combines a spatially structured component that follows a conditional autoregressive prior (ICAR) (Besag, 1974), accounting for spatial correlation in the data, and a spatially unstructured component following a Gaussian prior, accounting for heterogeneity in the data. For the spatial ICAR prior, we specify the neighbourhood structure by defining an adjacency matrix W of size $N \times N$ such that the diagonal entries $w_{i,j} = 0$ and the off-diagonal entries $w_{i,j} = 1$ if areas i and j share a common boundary, and 0 otherwise. The temporal effects are assigned the temporal analogue of the ICAR prior, thus accounting for the temporal correlation in the data. Similar to the spatial ICAR prior, the temporal neighbourhood structure is defined through a matrix Q , where $q_{h,t} = 1$ if $|h - t| = 1$ and $q_{h,t} = 0$ otherwise, with h and t indexing units of time. For the Area-Specific Model, the same ICAR prior is assigned to the temporal component k_{it} . In addition, a prior is assigned to the area-specific variances σ_{ik}^2 as an extra hierarchical level for identifiability reasons. For the overall intercept α_0 and the area specific intercept u_i , vague priors are specified.

A weakly informative half Normal prior $N(0,1)$ is assigned to each of the parameters σ_h , σ_v and σ_γ (Gelman, 2006), while a $N(\alpha, \beta^2)$ prior is assigned to the $\log(\sigma_{ik}^2)$ with parameters α and β^2 following priors $N(0, 1000)$ and $N(0, 2.5^2)$, based on the specification by Li et al. (2012). As the latter prior is somewhat informative, sensitivity analysis is carried out to assess the robustness of the results.

The model indicator z_i follows a Bernoulli (0.95) prior, selecting estimates from either the Common Model ($z=1$) or the Area-Specific Model ($z=0$). The parameter 0.95 reflects our expectations that only a small number of areas express an unusual temporal pattern.

To classify areas as unusual, we use the posterior estimates of the indicator z_i representing how likely it is for area i to follow the Common Model, i.e. to exhibit a usual pattern in the risks, and we select the ones that satisfy the following condition: $r_j \leq 0.05$, where r_j is the j th ordered posterior z_i .

3.2. Applications

The model is fitted to all three datasets described in Section 2. In the prescription model, Y_{it} is the number of GP drug prescriptions in CCG i at month t and E_i are the expected counts. Appropriate covariates are also included in (2) to adjust for age and sex. As an indicator of age, we use the percentage of active population (age group 15 to 64), while as an indicator of sex, we use the male to female ratio.

In the admission model, Y_{it} are the observed cases of asthma and COPD admissions in CCG i at month t and E_i are the expected ones based on age and sex direct standardisation using the whole of England as standard population. Similarly, in the mortality model, Y_{it} is the observed number of deaths due to asthma or COPD in CCG i at month t and E_i is the corresponding expected number based on age and sex direct standardisation using the whole of England as standard population.

The models are implemented in OpenBUGS using Markov Chain Monte Carlo (MCMC) integration algorithms (Gilks, 2005). Three chains are run for each parameter per model with different initial values for 80 000 iterations, from which 20 000 are discarded as burn-in, and estimates are based on the remaining samples using only every 5th iteration to limit autocorrelation. The simulations took around 17 hours per model on an Intel Xeon processor 2.50GHz with 23.4 RAM. Convergence was assessed through trace plots, BGR statistic, and Monte Carlo error.

4. Results & Discussion

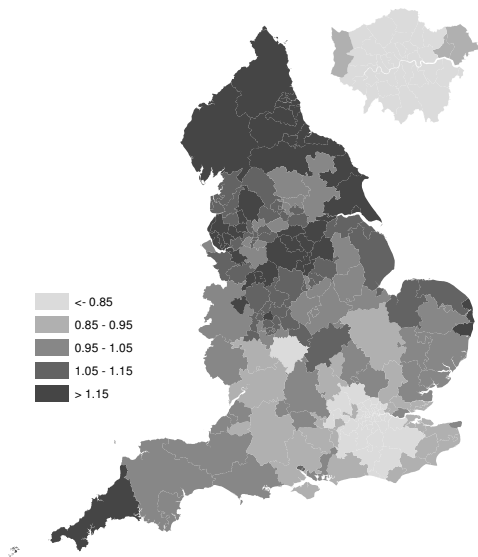
4.1. Spatial Patterns

We provide maps of posterior rates to investigate the geographical patterns of asthma and COPD across the three different data sources and evaluate whether these share any similarities or differences. Figure 1a shows the residual relative risk of the spatial component $\exp(h_i)$ of the Common Model (2) for the GP drug prescriptions across England during the period August 2010 to March 2011, while Figures 1b and 1c show the corresponding risk for HES admissions, and deaths respectively.

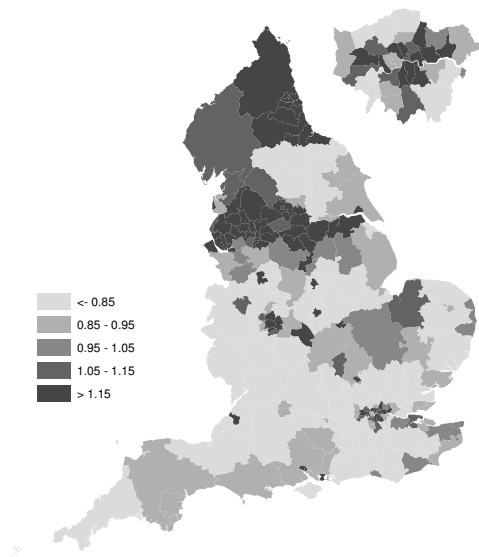
A clear pattern is observed across all datasets with a strong increasing effect from south to north. High risk is focused mainly on the region that includes Liverpool, Manchester, Leeds, and Sheffield, as well as on the region Newcastle and Durham, with the highest risks being observed in the north east of England. Potential shared risk factors may include deprivation, which is highest in urban areas of the north-east and north-west, environmental factors such as air pollution and smoking prevalence (Prescott and Vestbo, 1999; Lopez et al., 2006; Andersen et al., 2011). The north west is the areas with the highest prevalence of smokers (HSCIC, 2015).

By comparing the risk distribution across different data sources, we observe that for admissions and deaths, these are somewhat consistent (Figures 1b and 1c), while for prescriptions important differences are highlighted (Figure 1a). This observation is confirmed by the heatmap in Figure 1d which shows the correspondence across the three data sources using the spatial relative risk; the area in light grey represents the CCGs in England that share a low relative risk while the area in dark grey represents the ones with a high relative risk. It is clear that the proportion of areas with a low/high relative risk across all three data sources is relatively small compared to the proportion of areas with a low/high relative risk among HES admissions and mortality data. The correlation of the latter two was estimated to be 0.77, while the corresponding correlations of GP drugs and mortality, and GP drugs and HES admissions were close to 0.5.

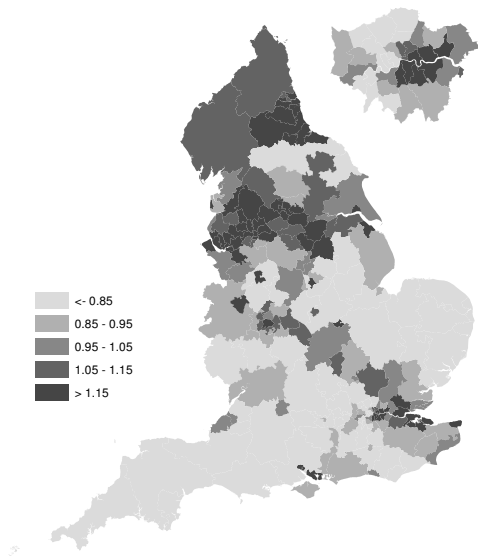
Low risk is observed for admissions and deaths (Figures 1b and 1c) in the south of England, and in the coastal areas. At the same time these areas exhibit high prescription rates (Figure 1a). This suggests a number of hypotheses for further investigation e.g. that higher prescription rates are a causal factor in lower morbidity and mortality, that there are more patients with milder disease in the south-east



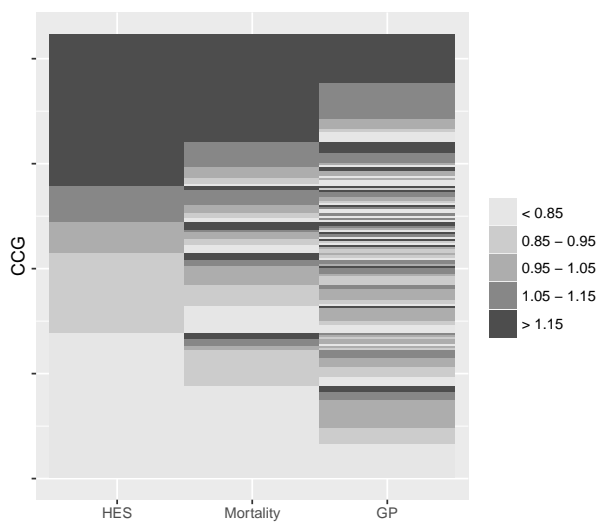
(a) GP drug prescription data



(b) HES admission data



(c) Mortality data



(d) Heatmap

Figure 1: Spatial patterns of chronic respiratory disease across England for GP drugs (a), admissions (b), and deaths (c); heatmap showing correspondence across the three data sources (d)

potentially related to lifestyle factors (lower smoking) and environmental factors (lower air pollution), that deprivation results in higher morbidity and mortality but lower use of primary care. Age-sex effects could also explain these differences; these are accounted for at an aggregate level for drugs, through the inclusion of appropriate covariates at the GP practice level in the model, since information was not available at an individual level, while direct standardisation was used for admissions and mortality.

Only a few CCGs in the south appear to have high admission and mortality rates (Figures 1b and 1c), including Portsmouth and Southampton CCGs. Increased smoking prevalence and high deprivation can be related to this. Unlike the rest of big cities in England, London appears to have low prescription rates (Figure 1a) potentially related to the low number of smokers in the area (HSCIC, 2015). Comparing between mortality and admission spatial patterns, a stronger south to north effect can be seen for mortality (Figure 1c), showing the south eastern and central west parts of England as the least risky for disease death. Another region which potentially represents a group of patients with mild symptoms is around the Yorkshire area, where national parks are, and this area clearly stands out in the admissions map (Figure 1b).

4.2. Temporal patterns

The general temporal patterns in England from August 2010 to March 2011 under the different datasets considered in the study can be seen in Figure 2, which plots the component $\exp(\gamma_t)$ of the Common Model (2).

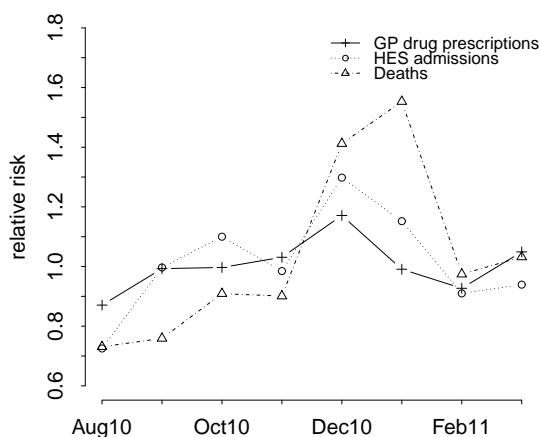


Figure 2: Temporal trends in chronic respiratory disease across different data sources

Generally, there was a seasonal pattern in the monthly risk for asthma and COPD across all data sources over the period August 2010 to March 2011 however the trends show important discrepancies. The highest risk was recorded around Christmas across

all data sources. This is likely to represent peaks in respiratory infections, especially influenza and respiratory syncytial virus (RSV) that peak around this year. These result in increased respiratory symptoms from mild to severe and cause exacerbations of disease that are occasionally fatal (Fleming et al., 2015). The lowest peak was recorded in August for all three data sources, a period when people are on holidays, away from their homes. The raise in the rates across all datasets around September coincides with the start of term at schools.

By comparing temporal trends across different data sources, we observe that there is a high temporal variation for mortality, whereas this is lower for admissions and prescriptions. Interestingly, a time lag is apparent between admissions and mortality. The highest peak for admissions is December, whereas for mortality is January and this might reflect the group of admissions which were then followed by death.

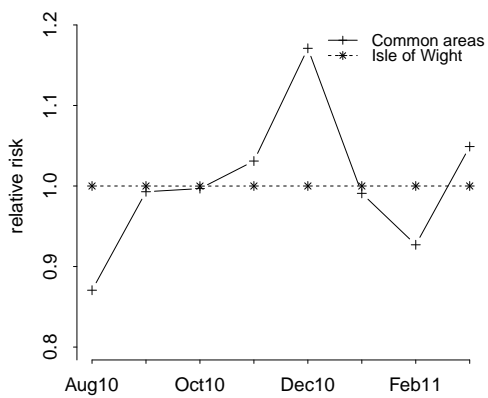
4.3. Detection of unusual areas

Finally, we obtain the posterior estimates of the indicator z_i for each model and we classify areas as unusual based on the rule described in Section 3.1.

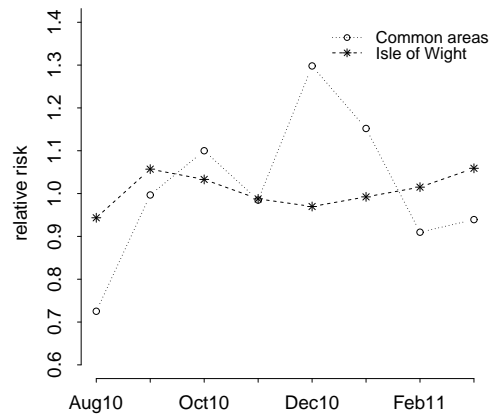
For GP drug prescription and mortality data, no areas were detected as unusual, meaning that all posterior estimates of the parameter z_i were above our *a priori* threshold value of 0.05. When ranking areas by probability, the Isle of Wight was the area with the smallest probability of following either the common prescriptions temporal pattern (with a probability of 0.13), or the common mortality temporal pattern (with a probability of 0.11). Looking at the temporal patterns for this area in Figure 3, we see that for prescriptions, a flat pattern is apparent, indicating low rates with no peaks throughout the whole time period (Figure 3a), while for mortality, a near exponential increase is observed, which after December 2010 exceeds the national death rates importantly (Figure 3c). The corresponding plot for the admissions shows a stable pattern (Figure 3b), similar to the one for prescriptions (Figure 3a).

After investigating the CCG of Isle of Wight, we found that an intervention was implemented in order to reduce the high prevalence of long respiratory diseases in the area, which was combined with an excess expenditure on respiratory medication. The project entitled ‘Isle of Wight Respiratory Inhaler Project’, led by the National Institute for health and Clinical Excellence (NICE), involved training of healthcare professionals in the use of the inhaler, patient training, and assessment of the inhaler technique. According to the results reported in 2009, the costs on selective beta-agonists fell by 22.7%, the number of prescriptions fell by 25.2%, the emergency admissions due to asthma were reduced by 50%, and associated deaths by 75% (NICE, 2015; The Pharmaceutical Journal, 2011).

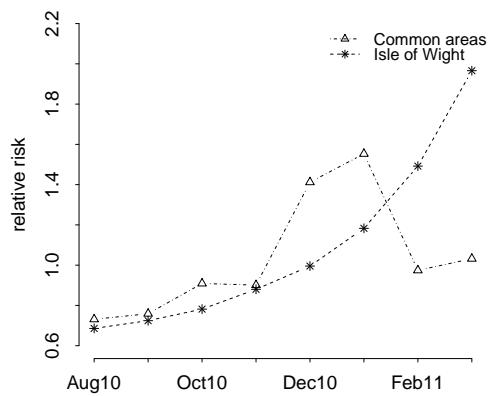
Given the above, the flat pattern in prescriptions seen in Figure 3a is well supported, as well as the one for admissions (Figure 3b). However, the increasing temporal trend in deaths in Figure 3c which peaks in the last month of the study period generates questions. Since the findings reported by NICE consider only asthmatic patients, meaning that COPD patients that use the same medication are



(a) GP drug prescription data



(b) HES admission data



(c) Mortality data

Figure 3: Temporal trend across different data sources for Isle of Wight

not represented, further analysis is required for the CCG of Isle of Wight separately on asthma and COPD in order to understand the underlying causes of the mortality trend. Additional data need to be also analysed in order to see the progression of the rates long after the intervention was implemented. Moreover, COPD is often misdiagnosed as asthma due to the similar symptoms that these two diseases share, and also depending on the availability of spirometry (Walker et al., 2006). This calls for further investigation into the misdiagnosed cases of COPD as asthma in the CCG of Isle of Wight which might be an adverse effect of the project.

On the other hand, four areas were detected as unusual using a 0.05 threshold level for hospital admissions model: Harrow, Hillingdon, Redbridge and Southampton. The time plots in Figure 4 show the temporal trends for each unusual area, compared to the general temporal trend.

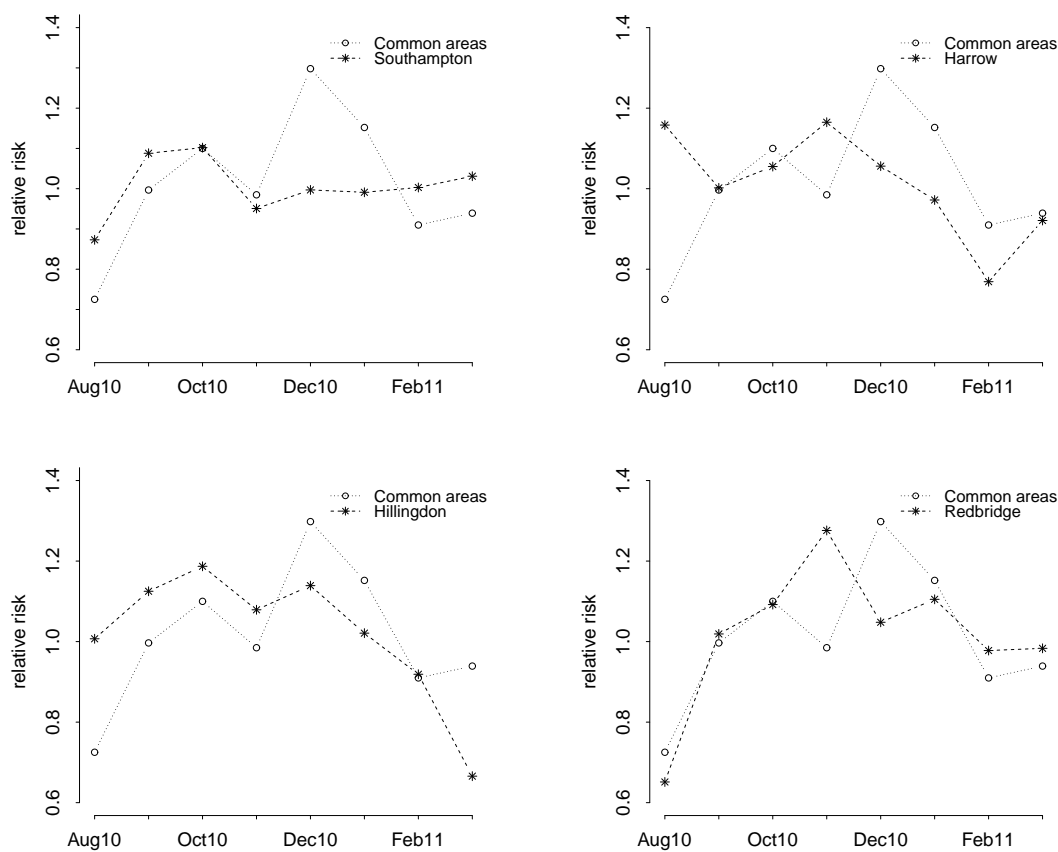


Figure 4: Unusual temporal trends under HES admissions data

As it can be seen, the corresponding areas do not follow particularly extreme patterns, but there are discrepancies from the common one that the model identifies as important. For instance, although Harrow in general exhibits lower than the

average national admission rates due to asthma or COPD, an unusual high peak is observed in August, as well as one in November, which months are considered safe for the rest of England. Hillingdon, on the other hand appears to have an extreme decrease in March 2011. Redbridge shows a different pattern in prescriptions from the other areas in that prescriptions increase in early November. London is a city known for its diversity, and for the population movements within it, hence it is not easy to make conclusions as to what has happened in those CCGs. Factors such as differences in local population characteristics, or differences in the quality of hospital care and the support people receive to manage their condition across time could be responsible.

The fourth area that was detected as unusual in terms of admission trends is Southampton. This is one of the few areas on the South Coast that show an increased risk for severe symptoms and deaths due to asthma and COPD (Figure 1). Interestingly, it seems that the low risk observed in Figure 4 from December onwards is the outcome of an intervention that was implemented in order to address the excess number of admissions in the CCG of Southampton. Wilkinson et al. (2014) found that 34 patients were responsible for 22% of the total COPD admissions over a 3-year period. The authors present an admission avoidance strategy that was constructed for this group of patients, and as a result the readmission rate fell from 13.4 to 1.9%.

5. Conclusions & further recommendations

In this paper we have investigated the spatial and temporal patterns of asthma and COPD, with a special focus on the detection of areas that follow an unusual temporal pattern compared to the general one, by using effective high quality datasets, aggregated by CCG which is meaningful for healthcare practices. There are two main conclusions we have drawn from our study.

First, we have shown that multiple data sources representing different degrees of disease severity give a more comprehensive picture of the respiratory health problem with potential implications for healthcare. Some similarities across the different datasets both in the spatial and temporal variation are indicated, however also important discrepancies are apparent reflecting the different groups of patients that are represented.

Second, the detection mechanism that is provided by the model we used, together with inference on the spatial, and temporal variation, can aid health care professionals and public health practitioners identify target areas, assess a policy impact or the quality provided by hospitals, and hence develop effective prevention programs to improve population health.

A strong aspect of the modelling approach is that it identifies unusual behaviour not only in terms of increased risk, but also in terms of any risk pattern that deviates from the expected one. Additionally, it only detects areas when a certain criterion is met, and the specification of this depends on how conservative we want to be. In the

original paper the model was applied on annual data by Li et al. (2012), while here we used monthly data as we were interested in the seasonal pattern of asthma and COPD. A great advantage of the additive model specification is that it allows for such flexibility, through the temporal random effects which account for unmeasured covariates that vary across time. Li et al. (2012) chose a Bayesian False Discovery Rate (FDR) to adjust for multiple testing, following Newton et al. (2004). Alternative FDR approaches have been suggested by other authors (Storey, 2003; Muller et al., 2006; Whittemore, 2007; Catelan et al., 2010). Here, we chose to adopt a standard classification rule, as we believe that the above FDR rule does not apply in our case, given the small number of positives we have. Besides, it is argued whether there is a need to perform multiple testing correction within a Bayesian framework (Gelman et al., 2012).

Asthma and COPD are jointly considered in the analysis, as the indication of the disease was not possible for the prescriptions, and one of our objectives was to compare between the different data sets. A next step is to separately explore asthma and COPD trends for all ages and for children, adults and elderly to see if our findings hold for these conditions individually, and to better understand the temporal patterns of the areas found to be unusual. Further analysis should formally investigate possible confounders, effect modifiers and causal factors such as deprivation, smoking, seasonal respiratory virus activity and service access, and to also adjust for socio-economic status, as this is likely to affect the spatial patterns. It is relatively easy to introduce covariates into the BaySTDetect model to start to explore some of the hypotheses as to why some CCGs differ from national trends. As an example we introduced a variable to represent deprivation - the Index of Multiple Deprivation (IMD) score, which was initially available at LSOA level for England. This was aggregated at CCG level, and then converted into a categorical variable of 5 levels (quintiles) which was introduced into the model through dummy variables. The results showed no important effect; the exponential of the posterior estimates of the coefficients of the dummy variables varied from 0.99 to 1.02 with all credible intervals including 1. The inclusion of deprivation in the model did not affect the posterior estimates of the other parameters, nor the areas that were detected as unusual.

In addition, a multivariate modelling approach will be considered, which could provide a more comprehensive picture of the trends and potential outbreaks of the disease, by borrowing information across different data sources. Finally, it would be useful to know the exact temporal point at which an unusual observation occurs, and the model should be modified to adjust for this.

Acknowledgments

The work of the UK Small Area Health Statistics Unit is funded by Public Health England as part of the MRC-PHE Centre for Environment and Health, funded also by the UK Medical Research Council.

SAHSU holds approvals from the National Research Ethics Service - reference 12/LO/0566 and 12/LO/0567 - and from the Health Research Authority Confidentially Advisory Group (HRA-CAG) for Section 251 support (HRA - 14/CAG/1039).

Hospital Episode Statistics data are copyright 2014, re-used with the permission of the Health and Social Care Information Centre. All rights reserved.

References

- Abellan, J., S. Richardson, and N. Best (2008). Use of space-time models to investigate the stability of patterns of disease. *Environmental Health Perspectives* 116(8), 1111–1119.
- Andersen, Z. J., M. Hvidberg, S. S. Jensen, M. Ketzel, S. Loft, M. Sørensen, A. Tjønneland, K. Overvad, and O. Raaschou-Nielsen (2011). Chronic obstructive pulmonary disease and long-term exposure to traffic-related air pollution: a cohort study. *American journal of respiratory and critical care medicine* 183(4), 455–461.
- Bernardinelli, L., D. Clayton, C. Pascutto, C. Montomoli, M. Ghislandi, and M. Songini (1995). Bayesian analysis of space-time variation in disease risk. *Statistics in Medicine* 14(21-22), 2433–2443.
- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 36(2), 192–236.
- Besag, J., J. York, and A. Mollié (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the institute of statistical mathematics* 43(1), 1–20.
- Best, N., S. Richardson, and A. Thomson (2005). A comparison of Bayesian spatial models for disease mapping. *Statistical methods in medical research* 14(1), 35–59.
- Catelan, D., C. Lagazio, and A. Biggeri (2010). A hierarchical Bayesian approach to multiple testing in disease mapping. *Biometrical Journal* 52(6), 784–797.
- Centers for Disease Control and Prevention (2008). Deaths from chronic obstructive pulmonary disease—United States, 2000–2005. *MMWR. Morbidity and mortality weekly report* 57(45), 1229.
- Drazen, J. M., D. S. Postma, and K. F. Rabe (2015). The asthma–COPD overlap syndrome. *New England Journal of Medicine* 373(13), 1241–1249.
- Eeftens, M., M.-Y. Tsai, C. Ampe, B. Anwander, R. Beelen, T. Bellander, G. Cesaroni, M. Cirach, J. Cyrys, K. d. Hoogh, A. D. Nazelle, F. d. Vocht, C. Declercq, A. Ddel, K. Eriksen, C. Galassi, R. Grauleviien, G. Grivas, J. Heinrich, B. Hoffmann, M. Iakovides, A. Ineichen, K. Katsouyanni, M. Korek, U. Krmer, T. Kuhlbusch, T. Lanki, C. Madsen, K. Meliefste, A. Mlter, G. Mosler, M. Nieuwenhuijsen, M. Oldenwening, A. Pennanen, N. Probst-Hensch, U. Quass, O. Raaschou-Nielsen, A. Ranzi, E. Stephanou, D. Sugiri, O. Udvardy, v. Vaskvi, G. Weinmayr, B. Brunekreef, and G. Hoek (2012). Spatial variation of PM_{2.5}, PM₁₀, PM_{2.5} absorbance and PM_{coarse} concentrations between and within 20 European study areas and the

- relationship with NO₂ - results of the ESCAPE project. *Atmospheric Environment* 62(0), 303 – 317.
- Fleming, D. M., R. J. Taylor, R. L. Lustig, C. Schuck-Paim, F. Haguinet, D. J. Webb, J. Logie, G. Matias, and S. Taylor (2015). Modelling estimates of the burden of respiratory syncytial virus infection in adults and the elderly in the united kingdom. *BMC infectious diseases* 15(1), 443.
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis* 1(3), 515–534.
- Gelman, A., J. Hill, and M. Yajima (2012). Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness* 5(2), 189–211.
- Gilks, W. R. (2005). *Markov chain Monte Carlo*. Wiley Online Library.
- Hacking, J. M., S. Muller, I. E. Buchan, et al. (2011). Trends in mortality from 1965 to 2008 across the English north-south divide: comparative observational study. *BMJ* 342.
- Hansell, A., J. Hollowell, R. McNiece, T. Nichols, and D. Strachan (2003). Validity and interpretation of mortality, health service and survey data on COPD and asthma in England. *European respiratory journal* 21(2), 279–286.
- Holt, J. B., X. Zhang, L. Presley-Cantrell, and J. B. Croft (2011). Geographic disparities in chronic obstructive pulmonary disease (COPD) hospitalization among medicare beneficiaries in the United States. *International journal of chronic obstructive pulmonary disease* 6, 321.
- HSCIC (cited October 2015). Statistics on Smoking: England, 2011. The Health and Social Care Information Centre. <http://www.hscic.gov.uk/catalogue/PUB00693/smok-eng-2011-rep.pdf>.
- Joo, M. J., T. A. Lee, and K. B. Weiss (2007). Geographic variation in chronic obstructive pulmonary disease exacerbation rates. *Journal of general internal medicine* 22(11), 1560–1565.
- Knorr-Held, L. (2000). Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine* 19(17-18), 2555–2567.
- Knorr-Held, L. and J. Besag (1998). Modelling risk from a disease in time and space. *Statistics in Medicine* 17(18), 2045–2060.
- Laurent, O., G. Pedrono, L. Filleul, C. Segala, A. Lefranc, C. Schillinger, E. Rivière, and D. Bard (2009). Influence of socioeconomic deprivation on the relation between air pollution and β -agonist sales for asthma. *CHEST Journal* 135(3), 717–723.

- Li, G., N. Best, A. L. Hansell, I. Ahmed, and S. Richardson (2012). BaySTDetect: detecting unusual temporal patterns in small area data via Bayesian model choice. *Biostatistics* 13(4), 695–710.
- Lindebeg, A., A. Bjerg-Backlund, E. Ranmark, et al. (2006). Prevalence and under-diagnosis of COPD by disease severity and the attributable fraction of smoking: report from the Obstructive Lung Disease in Northern Sweden Studies. *Respir Med* 100, 264–72.
- Lipton, R. and R. Banerjee (2006). The geography of chronic obstructive pulmonary disease in california, 1993 & 1999: Using a novel method to transform zip code level data to a uniform grid structure. In *American Journal of Epidemiology*, Volume 163, pp. S28–S28.
- Lopez, A., K. Shibuya, C. Rao, C. Mathers, A. Hansell, L. Held, V. Schmid, and S. Buist (2006). Chronic obstructive pulmonary disease: current burden and future projections. *European Respiratory Journal* 27(2), 397–412.
- Masoli, M., D. Fabian, S. Holt, and R. Beasley (2004). The global burden of asthma: executive summary of the GINA Dissemination Committee report. *Allergy* 59(5), 469–478.
- Muller, P., G. Parmigiani, and K. Rice (2006). FDR and Bayesian multiple comparisons rules. In *Proceedings Valencia*.
- Nandram, B., J. Sedransk, and L. W. Pickle (2000). Bayesian analysis and mapping of mortality rates for chronic obstructive pulmonary disease. *Journal of the American Statistical Association* 95(452), 1110–1118.
- Naureckas, E. T., V. Dukic, X. Bao, and P. Rathouz (2005). Short-acting β -agonist prescription fills as a marker for asthma morbidity. *CHEST Journal* 128(2), 602–608.
- Newton, M. A., A. Noueir, D. Sarkar, and P. Ahlquist (2004). Detecting differential gene expression with a semiparametric hierarchical mixture method. *Biostatistics* 5(2), 155–176.
- NICE (cited October 2015). National Institute for Health and Clinical Excellence, Kings Fund, Paul Jerram (IOW PCT). <http://www.nice.org.uk/proxy/?sourceurl=http://www.nice.org.uk/usingguidance/sharedlearningimplementingniceguidance/examplesofimplementation/eximresults.jsp?o=461>.
- Prescott, E. and J. Vestbo (1999). Socioeconomic status and chronic obstructive pulmonary disease. *Thorax* 54(8), 737–741.

- Sin, D. D., N. R. Anthonisen, J. B. Soriano, and A. Agusti (2006). Mortality in COPD: role of comorbidities. *European Respiratory Journal* 28(6), 1245–1257.
- Sofianopoulou, E., S. P. Rushton, P. J. Diggle, and T. Pless-Mulloli (2013). Association between respiratory prescribing, air pollution and deprivation, in primary health care. *Journal of Public Health* 35(4), 502–509.
- Storey, J. D. (2003). The positive false discovery rate: A Bayesian interpretation and the q-value. *The Annals of Statistics* 31(6), 2013–2035.
- The Pharmaceutical Journal (2011). *URI: 10047732 282*, 99.
- Vegni, F. E., B. Castelli, F. Auxilia, and P. Wilkinson (2005). Air pollution and respiratory drug use in the city of como, italy. *European journal of epidemiology* 20(4), 351–358.
- Walker, P., P. Mitchell, F. Diamantea, C. Warburton, and L. Davies (2006). Effect of primary-care spirometry on the diagnosis and management of COPD. *European Respiratory Journal* 28(5), 945–952.
- Waller, L. A., B. P. Carlin, H. Xia, and A. E. Gelfand (1997). Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association* 92(438), 607–617.
- Wells, C. and E. Gordon (2008). Geographical variations in premature mortality in England and Wales, 1981–2006. *Health statistics quarterly* 38, 6–18.
- Whittemore, A. S. (2007). A Bayesian false discovery rate for multiple testing. *Journal of Applied Statistics* 34(1), 1–9.
- Wilkinson, T., M. North, and S. C. Bourne (2014). Reducing hospital admissions and improving the diagnosis of COPD in Southampton city: methods and results of a 12-month service improvement project. *NPJ primary care respiratory medicine* 24.
- World Health Organization (2012). Prevention and control of noncommunicable diseases: Guidelines for primary health care in low resource settings.
- World Health Organization (2014). Global status report on alcohol and health.