Disease surveillance on COPD using BaySTDetect

Areti Boulieri and Marta Blangiardo
Dept of Epidemiology and Biostatistics, Imperial College London

Introduction

Disease surveillance is an important public health practice, as it provides information which can be used to make successful interventions. Innovative surveillance systems are being developed to improve early detection and investigation of outbreaks, with the Bayesian models attracting a lot of interest recently. Outbreak detection requires a system that will be able to flag areas that are differentially expressed. Both test-based and model-based techniques exist in the literature. Within the Bayesian framework, spatio-temporal hierarchical models are able to give robust results due to their flexibility. Through the specification of spatially and/or temporally structured random effects information is shared between areas and/or time points, increasing the strength of the parameter estimates. These models are designed to provide estimates and describe risk patterns, however very limited research exists in models that are able to provide a detection mechanism. BaySTDetect is a recently developed method by Li et al. (2012) able to detect outbreaks, and also to control for multiple testing through the specification of the False Discovery Rate (FDR). The objective of this work is to analyse hospital admission data for chronic obstructive pulmonary disease (COPD) by using the BaySTDetect method. Important aspect of this work is to examine the association between hospital admissions for COPD and GP drug prescriptions for COPD.

Data

Hospital Episode Statistics (HES) data for COPD in England are obtained from the Small Area Health Statistics Unit (SAHSU) of Imperial College. Population data are also obtained from SAHSU. GP drug prescription data are released monthly by the English National Health Service (NHS) for all general practices in England and are publicly available. The datasets are aggregated at Clinical Commissioning Groups (CCG) (211 in England) and the temporal coverage is from August 2010 to March 2011.

Statistical Framework

First level of hierarchy:

$$Y_{it} \sim \text{Poisson}(\mu_{it}\, \text{E}_{it}),$$

for CCG $i = 1, 2, \ldots, 211$, and time point $t = 1, 2, \ldots, 8$. Y_{it} and E_{it} are the observed and expected numbers of COPD admissions, respectively, in CCG i at time t.

Second level of hierarchy:

$$log(\mu_{it}) = \begin{cases} \alpha_0 + \eta_i + \gamma_t + \beta * X_{it} & \text{Common model (C)} \\ u_i + \xi_{i,t} & \text{Area-specific model (AS)} \end{cases}$$

$$\mu_{it} \sim z_i \times \mu_{it}^{\text{C}} + (1 - z_i) \times \mu_{it}^{\text{AS}} \quad \text{Selection model}$$

where $z_i \sim \text{Bernoulli}(0.9)$
X_{it} is the drug prescription rate in CCG i at time t.

Third level of hierarchy:

Common model

$$\alpha_0 \sim \text{U}(-\infty, +\infty), \beta \sim N(0, 1000)$$
$$\eta_i \sim N(v_i, \sigma_\eta^2), v_{1:N} \sim \text{CAR}(\mathbf{W}, \sigma_v^2)$$
$$\gamma_{1:T} \sim \text{CAR}(\mathbf{Q}, \sigma_\gamma^2)$$

Area-specific model

$$u_i \sim N(0, 1000)$$
$$\xi_{i,1:T} \sim \text{CAR}(\mathbf{Q}, \sigma_{i,\xi}^2)$$
$$\log(\sigma_{i,\xi}^2) \sim N(a, b^2)$$

A weakly informative half Normal prior $N(0, 1)$ is assigned to each of the variances σ_η, σ_v, σ_γ (Gelman, 2006). An $N(0, 1000)$ prior is assigned to the hyperparameter a, and a moderately informative prior $N(0, 2.5^2)$ is assigned to the hyperparameter b.

Detection rules based on the Bayesian FDR

In oder to control for the FDR in our model, we use the posterior model probability f_i, which is the posterior mean of the model indicator z_i (Newton et al., 2004).
We classify areas based on the following criterion.

$$f_i < C = f_{(k)}, \text{where k is the maximum integer such that}$$

$$\frac{1}{k} \times \sum_{j=1}^{k} f_{(j)} < \alpha$$

with $f_{(j)}$ denoting the j^{th} ordered posterior model probability,

and α is some preset level, here taken to be 0.05.

Results

At the FDR level of 0.05, seven CCGs out of 211 were detected by BaySTDetect to follow an unusual temporal trend over the period August 2010 to March 2011. In addition, the effect of the drug prescription rate on the hospital admissions is shown to be significant ($\exp(\beta) = 1.015$).

Figure 1 presents the locations of these areas, and their corresponding temporal trends, compared to the common temporal trend.

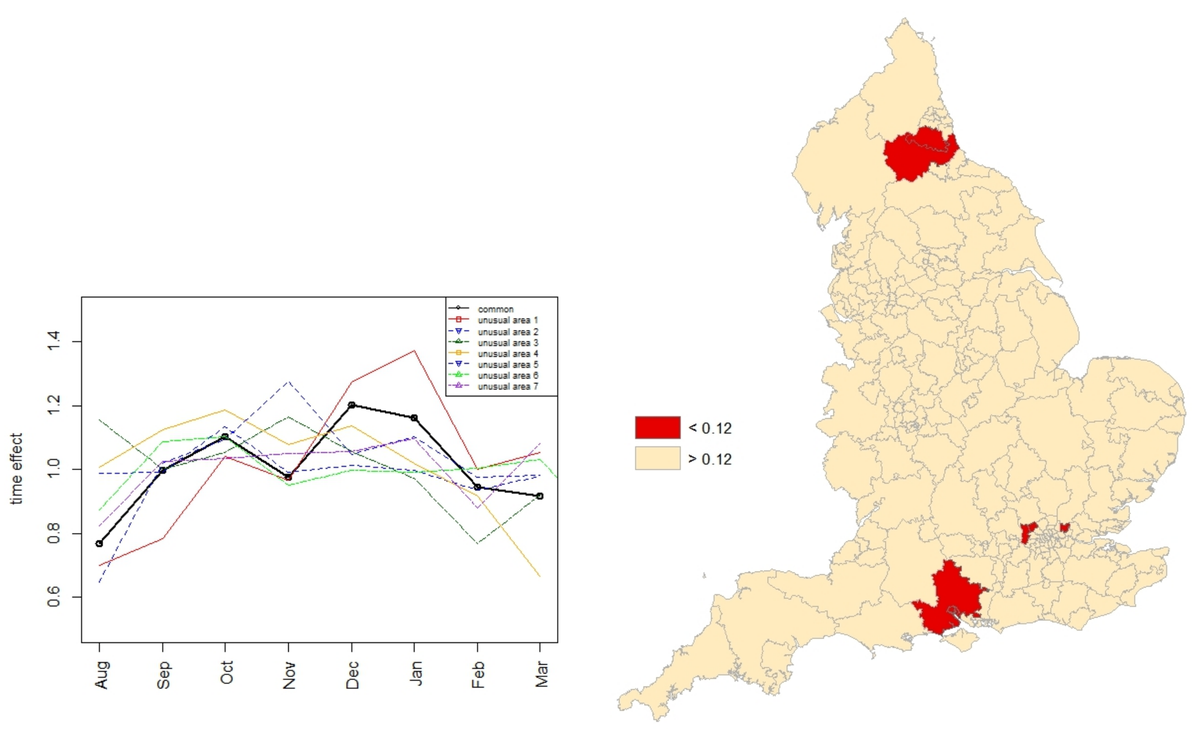

Figure 1: Comparison between the common and unusual temporal patterns (left) and the locations of the unusual CCGs in England (right).

Discussion

Among the CCGs that were detected by BaySTDetect, three of them are within the region of London (Harrow, Hillingdon, Redbridge). Other declared CCGs include Southampton and West Hampshire in South London, and Durham, Dales and Easington and North Durham in North London. From the perspective of policy making, several factors could have contributed to the unusual temporal trends of these areas, such as environmental factors, and further research is required to investigate this.
The next step of our analysis is to include additional time information in the model, and modify the BaySTDetect method to make it suitable for longer time series, by allowing the model indicator z_i to change across time $(z_{i,t})$.

References

[1] Besag, Julian. Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 36. 2 (1974): 192–236.

[2] Gelman, Andrew. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis* 1. 3 (2006): 515–534.

[3] Li, Guangquan, Nicky Best, Anna L. Hansell, Ismail Ahmed, and Sylvia Richardson. BaySTDetect: detecting unusual temporal patterns in small area data via Bayesian model choice *Biostatistics* 13. 4 (2012): 695–710.

[4] Newton, Michael A. and Noueiry, Amine and Sarkar, Deepayan and Ahlquist, Paul. Detecting differential gene expression with a semiparametric hierarchical mixture method. *Biostatistics* 5. 2 (2004): 155–176.