A BAYESIAN MIXTURE MODEL FOR DETECTING UNUSUAL TIME TRENDS IN SMALL AREA ESTIMATES: APPLICATION TO COPD MORTALITY IN ENGLAND

G. Li[1], N. Best[1,2], A. Hansell[1,2] and S. Richardson[1,2]

[1] Department of Epidemiology and Biostatistics and [2] MRC-HPA Centre for Environment and Health, Imperial College London

INTRODUCTION

- Space-time modeling of small area data is often used in epidemiology for mapping chronic disease rates and by government statistical agencies for producing local estimates of, for example, unemployment or crime rates. At any given time point, the spatial variations can be attributable to the differences in the distribution of predictors such as socio-demographic and environmental risk factors. Since most of these predictors are likely to remain stable over time, the temporal changes in most local areas tend to resemble each other closely.
- However, some may exhibit unexpected changes over time, suggesting, e.g., the emergence of localized predictors/risk factor(s) or impacts from a new policy. Detection of areas with "unusual" temporal patterns is therefore of importance for several reasons, including identifying possible predictors/risk factors or assessing the effectiveness of a policy.
- We have proposed a novel Bayesian mixture model for short time series of small area data that provides estimates of both the common temporal trend and the area-specific temporal trends. For each area, the posterior probability of belonging to the area-specific versus the common trend is used to classify the local time trend as "unusual" or not.

THE DETECTION FRAMEWORK

Given that the disease of interest is rare and non-contagious, we assume that the count of cases in area *i* at time *t* follows a Poisson distribution. That is,

$$y_{i,t} \sim Poisson(\theta_{i,t} \cdot E_{i,t})$$

where $\theta_{i,t}$ and $E_{i,t}$ are the relative risk (RR) and the expected number of cases, respectively. The log RR is decomposed into two components through a mixture model, one to describe a common trend pattern of all areas assuming space-time separability while the other to provide trend estimates for each spatial unit locally:

$$log(\theta_{i,t}) = z_i \cdot (\alpha + \eta_i + \nu_t) + (1 - z_i) \cdot (u_i + \xi_{i,t})$$

- The first component, $(\alpha + \eta_i + \nu_t)$, combines the effects of space, η_i, and time, ν_t, additively, an assumption that can over-smooth local trends that display unusual patterns;
- A second component, $(u_i + \xi_{i,t})$, is introduced so that the temporal trends are estimated locally, retaining the characteristics of the local trend patterns; u_i is the area-specific intercept and $\xi_{i,t}$ models the local trend patterns;
- The latent indicator, z_i, selects one of the above two components that is best described the observed data. Its posterior mean, p_i, interpreted as evidence of area *i* to follow a common trend pattern, is used to classify areas (see below);
- For the spatial random effects we assign the convolution BYM prior [1] while for the temporal random effects terms, a random walk of order 1 is used.

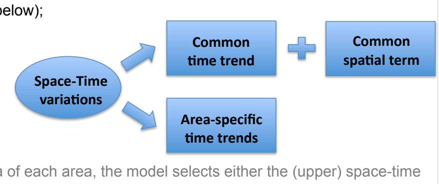
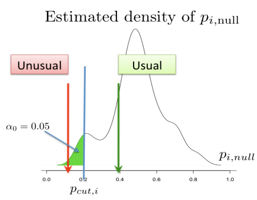


Figure 1. For fitting data of each area, the model selects either the (upper) space-time separable model with a common temporal trend or an area-specific trend model.

CLASSIFICATION WITH A CONTROL FOR FDR

- Given a cutoff, detection of unusual trends is achieved based on the rule outlined on the right.
- However conventional methods for estimating the cutoff based on some pre-defined false discovery rate (FDR) are not applicable since here the alternative hypothesis is specific to each area. Thus area-specific cutoffs are required.



$p_i < p_{cut}$ → Area i is unusual

$p_i \geq p_{cut}$ → Area i is usual

Small p_i indicates unusual

- We have developed a simulation based approach which approximates the distribution of p_i under the null hypothesis, from which the area-specific cutoff is defined such that

$$P(p_{i,\text{under null}} < p_{cut,i}) = \alpha_0$$

where α_0 is the pre-defined FDR.

CASE STUDY: ASSESSING A POLICY ON COPD

- We have applied the proposed detection model to a retrospective study of chronic obstructive pulmonary disease (COPD) mortality data in England and Wales (a total of 376 local districts) to assess the impact of government policy to make COPD a "compensable" disease under the Industrial Injuries Disablement Benefit scheme in 1992, which would be expected to have greatest impact in industrial areas with exposures to dusts and fumes.
- With FDR=0.05, 41 districts are identified with unusual trend patterns. Further clustering of these trend patterns using a method by Heard et al. [2] shows that 21 of them display a trend pattern that goes above the common trend around 1993 (Figure 2B) while 20 have a sharper decreasing pattern (Figure 2C). The identified districts are colored correspondingly in Figure 2A.

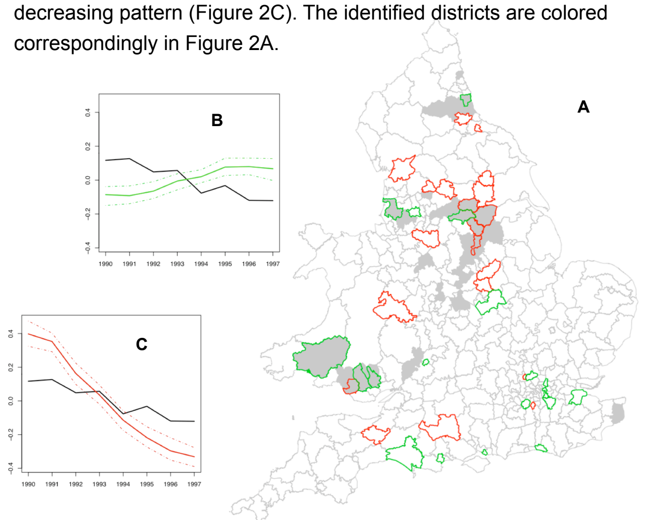


Figure 2. (A) Locations of the 41 unusual districts (mining districts, identified by high mortality of pneumoconiosis, a disease of miners, are colored in grey); (B) the cluster-specific trend pattern of the districts in green and (C) the trend pattern of those in red. In (B) and (C), the black curve represents the pattern of the common trend.

1. Under this compensable policy, we expect that general practitioners (GP) would be encouraged to diagnose aspiratory problems as COPD in miners so that the patients (or their family) could obtain compensation, a change that could lead to increase in COPD mortality.
2. The green districts in Wales and to the north of London either fall into the mining areas or are close to them. Others are seen along the south coast and along the Thames corridor to the east of London which are associated with heavy industry, shipping and port industry.
3. The districts in red, mainly located in Yorkshire and Nottinghamshire, Central England, are also associated with mining. However, there were wide scale closure of mining pits which could result in the decrease in COPD and COPD mortality.
4. The model has picked up changes in mining areas and those with heavy industry. The change cannot solely be linked to the compensable policy in question. Local factors such as the time of mining closure may also have important impacts. Further investigations are required to fully understand the differential of the trend patterns in different part of the county.

CONCLUSION AND FUTURE DIRECTIONS

- The strength of the proposed model over others is the ability to detect changes in trend patterns with distinctive departures such as elevation/ reduction of risks for some time points;
- The case study has demonstrated a use on policy assessment. Retrospective surveillance can be a potential field of application.

[1] Besag, York and Mollie. *Ann Inst Stat Math* 1991, 43, 1-59. [2] Heard, Holmes and Stephens (2006) *JASA* 2006, 101, 18-29.