BaySTDetect: Detecting unusual temporal patterns in small area disease rates using Bayesian posterior model probabilities

Guangquan Li, Sylvia Richardson, Lea Fortunato, Ismaïl Ahmed, Anna Hansell, Mireille Toledano and Nicky Best

Department of Epidemiology and Biostatistics
Imperial College London

ISBA2012 Kyoto, June 29, 2012
Outline

Motivation

BaySTDetect: Bayesian model choice for detecting unusual temporal patterns in small area data

Simulation study

Application1: Policy assessment

Application2: Data mining cancer incidence

Conclusions
Outline

Motivation

BaySTDetect: Bayesian model choice for detecting unusual temporal patterns in small area data

Simulation study

Application1: Policy assessment

Application2: Data mining cancer incidence

Conclusions
Motivation

▶ For many areas of application such as small area estimates of income, unemployment, crime rates and rates of chronic diseases, there is typically a general time trend that affects most areas similarly.

▶ However, abrupt changes may occur in a particular area due to, for example,
  ▶ emergence of localized risk factor(s);
  ▶ local policy implementation (e.g., health awareness or screening campaigns);
  ▶ changes to health care provision or social structure of the local population;
  ▶ local variations in diagnostic or coding practice;
  ▶ ...

▶ Detection of areas with unusual temporal patterns is therefore important as a screening tool for further investigation.
Motivation: Two applications

1. COPD: Policy assessment

▶ Industrial Injuries Disablement Benefit was made available for miners developing COPD from 1992 onwards in the UK. ▶ There was a debate on whether this policy may have differentially increased the likelihood of a COPD diagnosis in mining areas as miners with other respiratory problems with similar symptoms (e.g., asthma) could potentially have benefited from this scheme.

2. TCR: Retrospective surveillance on cancer incidence

▶ to highlight areas with a potential need for further investigation and/or intervention
Motivation: Two applications

1. COPD: Policy assessment
   - Industrial Injuries Disablement Benefit was made available for miners developing COPD from 1992 onwards in the UK.
   - There was a debate on whether this policy may have differentially increased the likelihood of a COPD diagnosis in mining areas. As miners with other respiratory problems with similar symptoms (e.g., asthma) could potentially have benefited from this scheme.

2. TCR: Retrospective surveillance on cancer incidence
Motivation: Two applications

1. COPD: Policy assessment
   - Industrial Injuries Disablement Benefit was made available for miners developing COPD from 1992 onwards in the UK.
   - There was a debate on whether this policy may have differentially increased the likelihood of a COPD diagnosis in mining areas as miners with other respiratory problems with similar symptoms (e.g., asthma) could potentially have benefited from this scheme.

2. TCR: Retrospective surveillance on cancer incidence
   - to highlight areas with a potential need for further investigation and/or intervention
Problems in small area detection

1. Sparse data (small number of cases)
   - BaySTDetect employs the Bayesian multilevel modelling framework to allow appropriate information borrowing.
Problems in small area detection

1. Sparse data (small number of cases)
   - BaySTDetect employs the Bayesian multilevel modelling framework to allow appropriate information borrowing.

2. Multiple comparisons are made
   - A Bayesian procedure is used in BaySTDetect to derive decision rules which enable the control of the false discovery rate (FDR).
Outline

Motivation

BaySTDetect: Bayesian model choice for detecting unusual temporal patterns in small area data

Simulation study

Application1: Policy assessment

Application2: Data mining cancer incidence

Conclusions
BaySTDetect: Model specification

Data level

\[ y_{it} \sim \text{Poisson}(\mu_{it} \cdot E_{it}) \]

Modelling underlying risks

\( \log(\mu_{it}) \)

**Common trend**

**Common spatial pattern**

**Area-specific time trends**

**Model 1:** Time trend pattern is the same for all areas

**Model 2:** Time trends are estimated independently for each area
BaySTDetect: Model specification

Data level

\[ y_{it} \sim \text{Poisson}(\mu_{it} \cdot E_{it}) \]

Modelling underlying risks

Common trend + Common spatial pattern

Area-specific time trends

Model 1: Time trend pattern is the same for all areas

Model 2: Time trends are estimated independently for each area

Selection

A model indicator \( z_i \) indicates for each area whether Model 1 \((z_i = 1)\) or Model 2 \((z_i = 0)\) is supported by the data.

\[ \mu_{it} = z_i \cdot \mu_{it}^{(M1)} + (1 - z_i) \cdot \mu_{it}^{(M2)} \]
BaySTDetect: Model specification

\[ y_{it} \sim \text{Poisson}(E_{it} \cdot \mu_{it}) \]

\[ \log(\mu_{it}) = \begin{cases} 
\alpha_0 + \eta_i + \gamma_t & \text{Model 1 for all } i, t \\
u_i + \xi_{it} & \text{Model 2 for all } i, t. 
\end{cases} \]
BaySTDetect: Model specification

\[ y_{it} \sim \text{Poisson}(E_{it} \cdot \mu_{it}) \]

\[ \log(\mu_{it}) = \begin{cases} \alpha_0 + \eta_i + \gamma_t & \text{Model 1 for all } i, t \\ u_i + \xi_{it} & \text{Model 2 for all } i, t. \end{cases} \]

**Model 1**

- \( \eta_i \sim \text{spatial BYM model} \) Common spatial pattern
- \( \gamma_t \sim \text{random walk } [\text{RW}(\sigma_\gamma^2)] \) Common temporal pattern
BaySTDetect: Model specification

\[ y_{it} \sim \text{Poisson}(E_{it} \cdot \mu_{it}) \]

\[
\log(\mu_{it}) = \begin{cases} 
\alpha_0 + \eta_i + \gamma_t & \text{Model 1 for all } i, t \\
\mu_{it} + \xi_{it} & \text{Model 2 for all } i, t.
\end{cases}
\]

**Model 1**

- \( \eta_i \sim \) spatial BYM model
- \( \gamma_t \sim \) random walk \([\text{RW}(\sigma^2_{\gamma})]\)

**Common spatial pattern**

**Model 2**

- \( u_i \sim \) \(N(0, 1000)\)
- \( \xi_{i,t} \sim \) random walk \([\text{RW}(\sigma^2_{\xi,i})]\)

**Area-specific temporal pattern**
BaySTDetect: Model specification

\[ y_{it} \sim \text{Poisson}(E_{it} \cdot \mu_{it}) \]

\[
\log(\mu_{it}) = \begin{cases} 
\alpha_0 + \eta_i + \gamma_t & \text{Model 1 for all } i, t \\
u_i + \xi_{it} & \text{Model 2 for all } i, t.
\end{cases}
\]

Model 1

\[ \eta_i \sim \text{spatial BYM model} \quad \text{Common spatial pattern} \]

\[ \gamma_t \sim \text{random walk } [\text{RW}(\sigma^2_\gamma)] \quad \text{Common temporal pattern} \]

Model 2

\[ u_i \sim \mathcal{N}(0, 1000) \]

\[ \xi_{i,t} \sim \text{random walk } [\text{RW}(\sigma^2_{\xi,i})] \quad \text{Area-specific temporal pattern} \]

Selection

\[ z_i \sim \text{Bern}(0.95) \]
A detection rule based on FDR

- Define $f_i = P(z_i = 1|\text{data})$ which is the probability that area $i$ belongs to the common trend model (Model 1)
  - A small $f_i$ suggests that area $i$ is unlikely to follow the common trend.

- We need to set a suitable cut-off value, $C$, such that areas with $f_i < C$ are declared to be unusual.
  - Put another way, if we declare area $i$ to be unusual, then $f_i$ can be thought of as the probability of false detection for that area.
  - We choose $C$ in such a way that we ensure that the average probability of false detection (i.e. the average value of $f_i$) amongst areas declared to be unusual is less than some pre-set level $\alpha$.
  - This procedure ensures that, on average, the number of false positives is no more than $(k \times \alpha)$, where $k$ is the number of declared unusual areas.
A detection rule based on FDR

- Define \( f_i = P(z_i = 1 | \text{data}) \) which is the probability that area \( i \) belongs to the common trend model (Model 1)
  - A small \( f_i \) suggests that area \( i \) is unlikely to follow the common trend.

- We need to set a suitable cut-off value, \( C \), such that areas with \( f_i < C \) are declared to be unusual.
A detection rule based on FDR

- Define $f_i = P(z_i = 1|\text{data})$ which is the probability that area $i$ belongs to the common trend model (Model 1)
  - A small $f_i$ suggests that area $i$ is unlikely to follow the common trend.

- We need to set a suitable cut-off value, $C$, such that areas with $f_i < C$ are declared to be unusual.

- Put another way, if we declare area $i$ to be unusual, then $f_i$ can be thought of as the probability of false detection for that area.

- We chose $C$ in such a way that we ensure that the average probability of false detection (i.e. the average value of $f_i$) amongst areas declared to be unusual is less than some pre-set level $\alpha$. 
A detection rule based on FDR

- Define $f_i = P(z_i = 1 | \text{data})$ which is the probability that area $i$ belongs to the common trend model (Model 1)
  - A small $f_i$ suggests that area $i$ is unlikely to follow the common trend.

- We need to set a suitable cut-off value, $C$, such that areas with $f_i < C$ are declared to be unusual.

- Put another way, if we declare area $i$ to be unusual, then $f_i$ can be thought of as the probability of false detection for that area.

- We chose $C$ in such a way that we ensure that the average probability of false detection (i.e. the average value of $f_i$) amongst areas declared to be unusual is less than some pre-set level $\alpha$.

- This procedure ensures that, on average, the number of false positives is no more than $(k \times \alpha)$, where $k$ is the number of declared unusual areas.
Outline

Motivation

BaySTDetect: Bayesian model choice for detecting unusual temporal patterns in small area data

Simulation study

Application 1: Policy assessment

Application 2: Data mining cancer incidence

Conclusions
Simulation: Setup

- Simulated data were based on the observed COPD mortality data (see Li et al. 2012).
- Three departure patterns were considered.
- When simulating the data, either the original set of expected counts from the COPD data or a reduced set (multiplying the original by 1/5) were used.
- 15 areas (approx. 4%) were chosen to have the unusual trend patterns.
  - areas were chosen to cover a wide range expected count values and overall spatial risks.
- Results were compared to those from the popular SaTScan space-time scan statistic.
Simulation: Unusual patterns

Figure: Illustration of the three departure patterns (red) with the common trend (black).

Pattern 1

Pattern 2

Pattern 3

Two departure magnitudes, $q = 1.5$ and 2, were considered.
Simulation: Sensitivity

Figure: Sensitivity of detecting the 15 truly unusual areas
Outline

Motivation

BaySTDetect: Bayesian model choice for detecting unusual temporal patterns in small area data

Simulation study

Application 1: Policy assessment

Application 2: Data mining cancer incidence

Conclusions
COPD application: Detected areas (FDR=0.05)
COPD application: Interpretation

- Results provide little support for hypothesis regarding the industrial injuries policy
  - only 3 out of 40 mining districts detected (Barnsley, Carmarthenshire and Rotherham);
  - unusual trend patterns in these areas are not consistent.
- Two unusual districts (Lewisham and Tower Hamlets) with an increasing trend (against a national decreasing trend) were identified in inner London.
- These areas are very deprived, with high in-migration and ethnic minorities → might expect different trends to rest of country.
- In fact, Tower Hamlets has been commissioning various local enhanced services to tackle high rates of COPD mortality since 2008.
- This rising trend could potentially have been recognised earlier in the 1990s through using BaySTDetect as a surveillance tool.
COPD application: Interpretation

- Results provide little support for hypothesis regarding the industrial injuries policy
  - only 3 out of 40 mining districts detected (Barnsley, Carmarthenshire and Rotherham);
  - unusual trend patterns in these areas are not consistent.
- Two unusual districts (Lewisham and Tower Hamlets) with an increasing trend (against a national decreasing trend) were identified in inner London.
  - These areas are very deprived, with high in-migration and ethnic minorities → might expect different trends to rest of country.
- In fact, Tower Hamlets has been commissioning various local enhanced services to tackle high rates of COPD mortality since 2008.
- This rising trend could potentially have been recognised earlier in the 1990s through using BaySTDetect as a surveillance tool.
Outline

Motivation

BaySTDetect: Bayesian model choice for detecting unusual temporal patterns in small area data

Simulation study

Application1: Policy assessment

Application2: Data mining cancer incidence

Conclusions
TCR application: Data

- The Thames Cancer Registry (TCR) collects data on newly diagnosed cases of cancer in the population of London and South East England.

- It is one of the largest cancer registries in Europe, covering a population of over 12 million, and holds nearly 3 million cancer registration records.

- We perform a retrospective surveillance of time trends for several cancer types using BaySTDetect
  - aim to provide screening tool to detect areas with unusual temporal patterns
  - automatically flag-up areas warranting further investigations
TCR application: results

Melanoma, FDR=0.05

<table>
<thead>
<tr>
<th>Period</th>
<th>Relative Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>81−84</td>
<td>85−88</td>
</tr>
<tr>
<td>89−92</td>
<td>93−96</td>
</tr>
<tr>
<td>97−00</td>
<td>01−04</td>
</tr>
<tr>
<td>05−08</td>
<td></td>
</tr>
</tbody>
</table>

Overall trend

Time period
Post-processing the detected trends

Cluster Dendrogram

1 cluster

2 clusters

Breast cancer
FDR=0.2

Black line = common trend
Coloured lines = average local trend in each cluster
Outline

Motivation

BaySTDDetect: Bayesian model choice for detecting unusual temporal patterns in small area data

Simulation study

Application1: Policy assessment

Application2: Data mining cancer incidence

Conclusions
Conclusions

▶ We have proposed a Bayesian space-time method for retrospective detection of unusual time trends;

▶ Simulation study has shown good performance of the model in detecting various realistic departures with relatively modest sample sizes

▶ We have demonstrated the use of BaySTDetect in policy assessment and in data mining;

▶ Implemented in R and WinBUGS, BaySTDetect enables real-time analysis of routinely collected data;

▶ Papers and WinBUGS codes for this model are available on www.bias-project.org.uk.
Acknowledgement

- This project is funded by the ESRC National Center for Research Methods through the BIAS II project.
- Thanks to the Thames Cancer Registry and the Small Area Health Statistics Unit (SAHSU) for providing the cancer incidence data.

Thank you!!
References
