

Data and text mining

The application of a novel ‘rising activity, multi-level mixed effects, indicator emphasis’ (RAMMIE) method for syndromic surveillance in England

Roger A. Morbey^{1,*}, Alex J. Elliot¹, Andre Charlett²,
Neville Q. Verlander², Nick Andrews² and Gillian E. Smith¹

¹Real-time Syndromic Surveillance Team, Public Health England, Birmingham B3 2PW, UK and ²Statistics and Modelling Economics Department, Public Health England, London, UK

*To whom correspondence should be addressed.

Associate Editor: Jonathan Wren

Received on February 19, 2015; revised on July 1, 2015; accepted on July 14, 2015

Abstract

Motivation: Syndromic surveillance is the real-time collection and interpretation of data to allow the early identification of public health threats and their impact, enabling public health action. The ‘rising activity, multi-level mixed effects, indicator emphasis’ method was developed to provide a single robust method enabling detection of unusual activity across a wide range of syndromes, nationally and locally.

Results: The method is shown here to have a high sensitivity (92%) and specificity (99%) compared to previous methods, whilst halving the time taken to detect increased activity to 1.3 days.

Availability and implementation: The method has been applied successfully to syndromic surveillance systems in England providing realistic models for baseline activity and utilizing prioritization rules to ensure a manageable number of ‘alarms’ each day.

Contact: roger.morbey@phe.gov.uk

1 Introduction

Syndromic surveillance is the near real-time collection and interpretation of data to allow the early identification of potential public health threats and their impact, enabling public health action (Triple, 2011). Syndromic surveillance fulfils part of Public Health England’s (PHE) role to give early warning of potential threats to public health due to infectious disease; and to quantify any impact (or reassure about the absence of impact) on public health during mass gathering or following major environmental incidents.

Within PHE, syndromic surveillance is co-ordinated by the Real-time Syndromic Surveillance Team (ReSST). ReSST monitors general practitioner (GP) consultations using an in-hours syndromic system (GPIHSS), currently covering about 61% of the registered GP patient

population in England, and a GP out of hours and unscheduled care system (GPOOHSS) (Harcourt *et al.*, 2012), covering about 76% of the provision in England. GPIHSS replaced a previous weekly system, HPA/QSurveillance in April 2013 (Harcourt *et al.*, 2011). A sentinel emergency department system (EDSSS) captures attendance data from 34 sites in England and Northern Ireland (Elliot *et al.*, 2012). The NHS 111 syndromic system analyses all calls to the national telephone health service, and replaced the existing NHS Direct syndromic system (Baker *et al.*, 2003) in September 2013.

Data are anonymized and then aggregated from diagnostic/symptom codes using code lists created by experienced epidemiologists. For most systems, the vast majority (over 80%) of patient consultations receive a useable diagnostic code; however coding rates can be very low (less than 5%) for a few GPOOHSS providers.

The following terms are used throughout this report:

- Signal—a measure of syndromic activity which has a specified system, syndrome and geography. Signals are measured on a daily basis e.g. ‘15 GP in hours’ consultations for diarrhoea in Hackney on 2nd January 2014.’
- (Syndromic) indicator—an aggregated group of diagnostic/symptom codes used to capture activity associated with one or more illnesses of public health importance e.g. acute respiratory infection, severe asthma, gastroenteritis, impact of cold weather.
- Baseline—the expected activity estimated using statistical methods for each signal.
- Thresholds—limits estimated using statistical methods with a known probability that a daily signal be within the thresholds.
- Alarm—an indication that a signal’s value on one particular day exceeds its upper threshold e.g. ‘GP In-hours asthma attendances for Birmingham have a baseline of 10 and an upper limit of 18, signal was 20, hence alarm.’

Each indicator has national, regional and local signals. Regional signals use the boundaries of the 15 PHE Centres. Local signals use the 152 English upper tier local authority boundaries, except for EDSSS, which is a sentinel system and thus uses the sentinel emergency department sites. The combination of syndromes and locations means that there are over 12 000 ‘signals’ to be reviewed daily by the team.

A wide range of statistical techniques are used internationally for syndromic surveillance (D’Errigo *et al.*, 2007; Tokars *et al.*, 2009; Robertson *et al.*, 2010; Corberan-Vallet and Lawson, 2011; Xing *et al.*, 2011; Lau *et al.*, 2012; Unkel *et al.*, 2012). Previously a number of different statistical methods were used by ReSST: the NHS Direct syndromic surveillance system used a regression model utilizing the historical data dating back to January 2003 (Baker *et al.*, 2003); HPA/QSurveillance used a standardized incidence ratio that compared sub-national with national incidence rates (Harcourt *et al.*, 2011); the EDSSS and GPOOH systems introduced prior to the 2012 Olympics used control chart methods (Morbey *et al.*, 2014).

The ‘rising activity, multi-level mixed effects, indicator emphasis’ (RAMMIE) method was developed to provide a single robust method for all systems, including any future new data sources. Having a single method improves interpretation of signals, with RAMMIE designed to combine the strengths of the existing methods.

The RAMMIE method incorporates the benefits of two types of previously used methods; using historical data and regression models to identify whether activity is unusual for the time of year, and a control chart type approach to identify significant recent rises in activity. The combination of these two types of alarm, ‘historical’ and ‘spike’ for each signal gives more information to interpret potential public health threats.

In application, the RAMMIE method needed to be robust and computationally fast, able to review over 12 000 signals daily and cope with the following data quality issues:

- The scale of signal counts varies greatly, a national signal may have tens of thousands of counts per day, whilst many local signals, for less common indicators like ‘meningitis’, will most days have a zero count.
- Many syndromic indicators, e.g. influenza-like illness, have a strong seasonal pattern, but peaks can vary unpredictably year on year.
- The volume of people accessing health care varies greatly by day of the week, due to the closure of GP in-hours surgeries during weekends and bank holidays.

- Syndromic systems use data that are collected for other purposes and are subject to unforeseen changes in the relevant health care system which can affect the volume of data or the coding. Many syndromic systems grow over time e.g. as new emergency department sites are added.
- Syndromic surveillance is required to identify a wide range of unforeseen events, varying from gradual changes in disease incidence (e.g. national pertussis outbreak in England during 2012; www.hpa.org.uk/NewsCentre) that may last over a year, to one-day ‘spikes’ in syndromes caused by e.g. air pollution.

Whilst primarily concerned with events of a national importance it is necessary to identify regional differences during an outbreak, and increasingly help is sought in identifying the impact of regional or local incidents (e.g. flooding). This paper outlines how the RAMMIE method has been developed and applied as a robust method for analysing and interpreting syndromic surveillance data, across a range of data sources, and describing initial validation work.

2 Methods

The RAMMIE method identified activity that was either unusually high for the time of year or had recently risen significantly, incorporating prioritization rules to ensure manageable numbers of statistical alarms.

2.1 Multi-level mixed effects modelling

One of the novel aspects of the RAMMIE method was to use multi-level modelling and benefit from the hierarchical structure of signals, with local authorities or emergency department sites grouped into English regions (PHE Centres) which are themselves subsets of national signals. This method enabled signals to be modelled at a local level despite sparse data by ‘borrowing strength’ from other areas.

Separate models were created for each syndromic indicator. A Poisson or a negative binominal model was used for the count data but, to allow for changes in data volume, an offset was used. For GPIHSS the offset was the registered patient population of GP practices, whilst for other systems the total daily activity was used. Separate models were used for national, regional and local signals, because low counts for the uncommon syndromes mean that some models did not converge at a local or regional level. The national signals used a negative binominal model to allow for over-dispersion. The more complicated multi-level sub-national models used a Poisson model that was computationally quicker to converge.

Independent variables included day of the week, whether or not the day was a bank holiday and month of the year. To avoid step changes at month end, which could lead to models with a ‘saw tooth’ appearance, each day was given a weighting based on two months: day 16 was treated as the middle of the month with each other day in the year contributing to two of the neighbouring month categories, with weighted values between 0.5 and 1, summing to 1 for each day.

For GPIHSS a variable for the day after a bank holiday was included in the models and found to be significant (with a p-value of less than 0.05): this was to account for the increased GP activity following the extended closure of GP surgeries over a bank holiday. For some systems binary coefficients were added to identify periods when a significant change occurred, e.g. moving from NHS Direct to NHS 111 data.

The formula for the three level models was expressed algebraically as follows:

$$\log(\mu_{ijk}) = \log(\text{total}_{ijk}) + \beta_0 + \beta_1 X_{ijk} + u_k + v_{jk}$$

For PHE Centre k , location (local authority or emergency department site) j on day i . Total is the offset (as described above) and $\beta_1 X_{ijk}$ represents a vector for all the independent variables and their coefficients, u_k represents the PHE Centre level specific random effect and v_{jk} represents the specific random error for each local area within a PHE Centre.

Coverage varied by signal and where data were very sparse models failed to converge. A pragmatic approach was applied, signals were only included where they had at least 365 days of non-zero counts, therefore some indicators were only modelled at a national or regional level.

All modelling and analysis described below was carried out using Stata version 13, (StataCorp LP; 2013).

2.2 Historical alarm

The independent variables for the subsequent weeks were known, being based on day of the week, month and holidays and so it was possible to predict activity. An average of the recent totals was used for the prediction which was then revised using the actual total on the day data were received.

'Historic thresholds' were calculated using either three standard deviations or three times the square root of the count above the modelled mean, whichever was greater. This approach provided a robust upper limit in order to generate 'historical' alarms.

2.3 Spike alarm

Occasionally a signal would remain significantly higher or lower than modelled for several weeks, e.g. if there was a change in community incidence of a disease (e.g. during the influenza season) or following the introduction of a new vaccine, or a change in recording practices. This could lead to either the historic alarm sounding every day or never sounding, therefore a second 'spike' threshold was created to identify recent increases in activity. The spike threshold was a scaled version of the historic threshold, based on how close to the model the signal had been over the past week (the data for the past week were weighted based on total consultations to prevent any single outlier days of missing data distorting the threshold).

When data were too sparse for a local model to be calculated the spike threshold was based on the national historic threshold, scaled to the recent difference between total local activity and national activity across the system.

2.4 Prioritization rules

Prioritization rules were created to prevent duplication of investigation and excessive numbers of alarms. Some indicators were more specific subsets of a more generalized indicator, for instance, diarrhoea was a subset of gastroenteritis. If an alarm for a specific indicator and its more general form sounded, then only the specific one was prioritized. When seasonal illnesses, for example seasonal influenza or norovirus, are widespread they generated a large number of alarms; if the increase in illness affected more than three local areas and also resulted in a PHE Centre alarm, then only the Centre alarm was prioritized. Similarly, if three or more Centres were affected and there was also a national alarm then only the national alarm was prioritized. If an increase in just one local location was sufficiently large to create a Centre or a national alarm, then only the local location was prioritized.

The season specific indicators, e.g. hypothermia were only prioritized during the appropriate season.

2.5 Validation of RAMMIE method

The RAMMIE method was tested in terms of specificity, sensitivity, positive predictive value and timeliness against known 'incidents' (Table 1). An incident was defined as a period of increased activity detected by ReSST using previous statistical methods and which had been additionally reported in the syndromic surveillance weekly bulletins (<https://www.gov.uk/government/collections/syndromic-surveillance-systems-and-analyses>) or internally reported within PHE. Thus the incidents contained increased syndromic surveillance activity between April 2012 and July 2013, all of which had been verified by reference to other surveillance systems within PHE e.g. laboratory reporting or by other sources of outbreak reports. In addition, we defined incidents as also including a number of widespread events that affected air quality during the same period.

The period selected for the validation covered the time for which historical data were available for all the current systems up to the introduction of RAMMIE as a replacement for previous methods. For this period, the RAMMIE method was applied retrospectively and the resulting alarms compared with previous methods on the same or comparable syndromic systems. The main difference between systems was for GPIHSS: prior to April 2013 the HPA/QSurveillance system had less coverage than GPIHSS and used a weekly data extract (except for a few months when daily data were used for enhanced surveillance during the 2012 London Olympic and Paralympic games).

The RAMMIE method was applied to local data for NHS Direct which previously had only regional and national signals.

Some incidents, e.g. major industrial fires, had a clear start date however others, e.g. a seasonal outbreak of norovirus were often less clear. Where the start date was unknown, the start was defined as one week prior to the first alarm by the previous methods, or two weeks for outbreaks lasting over four weeks. The end of an incident was defined as either one week after the final alarm created by previous methods or when levels were reported as returning to normal, whichever was sooner.

A signal was defined as being 'linked' to an incident if it could theoretically be expected to detect the increased activity, i.e. it was in the same geographical location and was for a syndromic indicator related to the incident.

Sensitivity was defined here as the percentage of incidents for which an alarm is generated; when measuring sensitivity by syndromic system, incidents were only included if they had at least one potentially 'linked' signal.

Specificity was defined here as the percentage of daily signals not linked to an incident that did not generate an alarm.

Timeliness was defined here as the mean number of days between the start of an incident and the first alarm.

Table 1. Number of public health incidents used in study to validate the RAMMIE method; incident presented by public health indicator category and geographical scale

| Type of incident | Local | Regional | National | Total |
|------------------|-------|----------|----------|-------|
| Respiratory | 5 | 4 | 7 | 16 |
| Gastrointestinal | 3 | 0 | 5 | 8 |
| Other | 5 | 1 | 8 | 14 |
| Total | 13 | 5 | 20 | 38 |

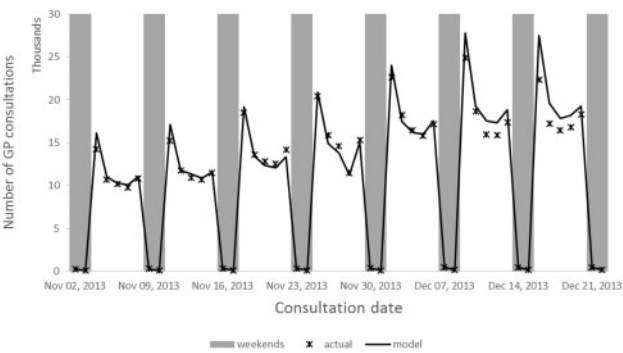


Fig. 1. Example signal with actual and modelled counts for GP in-hours consultations for upper respiratory tract infection across all participating practices in England between November and December 2013

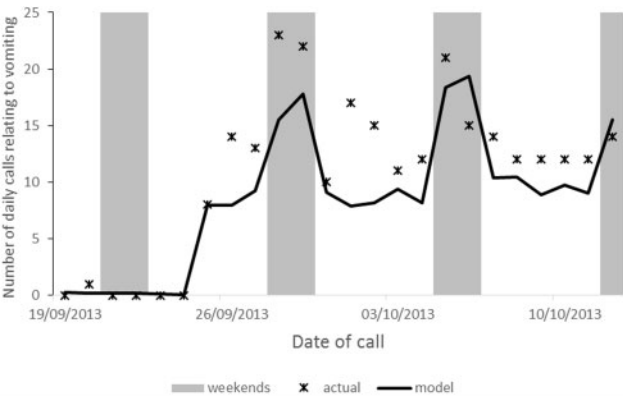


Fig. 2. Example signal with actual and modelled counts for telephone calls related to vomiting between September and October 2013. Prior to September 25, 2013 calls were to NHS Direct which was being phased out, subsequently to NHS 111

3 Results

3.1 Multi-level mixed effects modelling

The models created proved to be robust across the wide range of syndromes and locations tested, enabling public health surveillance to distinguish between usual seasonal factors and changes in data provision and potential outbreaks.

The model mirrored the weekly cycle and the seasonal winter increase in upper respiratory tract infections (Fig. 1). Offsetting by practice population enabled the model to adjust for occasional dips in activity due to technical issues.

The daily number of NHS 111 vomiting calls in one example PHE centre was again mirrored by the model (Fig. 2). The RAMMIE method was able to create models even when numbers were very low and was able to cope with sudden changes in coverage.

3.2 Spike alarms

The combination of spike and historical alarms provided extra information to aid public health interpretation. Figure 3 shows the number of emergency department attendances for bronchitis, along with alarms generated by the RAMMIE method.

The first historical alarm shown in Figure 3 came after a period when activity had been lower than expected and thus also generated a spike alarm, the second historical alarm came after a period when activity was already higher than expected so no spike alarm occurred.

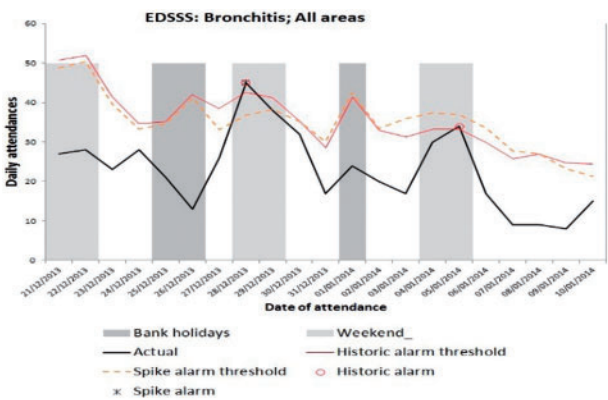


Fig. 3. Attendances between December 21, 2013 and January 10, 2014 to participating emergency departments that related to bronchitis; actual counts are shown along with statistical alarm thresholds and alarms where these are exceeded

Table 2. Number of alarms removed by using prioritization rules across all systems during study period (April 2012–July 2013)

| | Average number of alarms per day | Percentage of all alarms |
|---|--|--------------------------|
| Before rules applied | 209.1 | 100% |
| Prioritization rule | Average number of alarms removed by rule per day | |
| Don't prioritize seasonal alarms out of season | 4.6 | 2.2% |
| Only prioritize most specific indicator | 27.6 | 13.2% |
| Don't prioritize local alarms if outbreak is widespread | 58.3 | 27.9% |
| Only prioritize local alarms if increase is just in one local area. | 22.1 | 10.6% |
| All rules combined | 94.8 | 45.4% |

3.3 Prioritization rules

When applied retrospectively, the prioritization rules reduced the number of RAMMIE alarms from a mean of 209 to 114 per day (Table 2). Most of the reduction resulted from the rule that prioritized only national alarms when increased activity led to alarms in many local areas. These rules also reduced the peaks seen at busy times. Removing the prioritization rules would have resulted in days with over a thousand alarms, the busiest day having 1830: with the rules, the busiest day had 187 alarms.

Table 3 summarizes the sensitivity, specificity, positive predictive value (PPV) and timeliness of the different methods against incidents detected by the previous statistical methods, by syndromic system.

3.4 Sensitivity

One of the air quality incidents (a large plastics recycling fire in the West Midlands between June 30, 2013 and July 3, 2013) had not generated any alarms under previous statistical methods. Using the RAMMIE method, alarms would have been generated in the two GP surveillance systems.

There were three incidents involving local spikes in activity that generated alarms under the old HPA/QSurveillance system but no associated alarms using the RAMMIE method. All the other incidents detected by the previous methods would have been detected using the RAMMIE method. The RAMMIE method increased the

Table 3. Comparison of sensitivity and specificity of different methods by syndromic surveillance system

| System Coverage | | | Sensitivity | Specificity | Positive predictive value | Timeliness |
|---|---|----|--|---|---|---|
| Number of public health incidents with a 'linked' signal available (i.e. relevant syndromic indicator in same geographical location). | | | Percentage of linked incidents that have at least one statistical alarm. | Percentage of signals not linked to an incident that did not alarm. | Percentage of alarms that were linked to an incident. | Average number of days between start of incident and first alarm. |
| EDSSS | P | 25 | 92.0 | 99.5 | 23.7 | 1.5 |
| | R | 25 | 96.0 | 98.6 | 14.7 | 1.6 |
| GPIHSS | P | 36 | 72.2 | 99.0 | 29.1 | 5.5 |
| | O | 9 | 77.8 | 99.5 | 24.9 | 4.9 |
| | R | 36 | 75.0 | 99.2 | 27.9 | 2.6 |
| GPOOHSS | P | 37 | 67.6 | 99.9 | 28.9 | 2.8 |
| | R | 37 | 75.7 | 99.2 | 20.8 | 0.9 |
| NHS Direct | P | 19 | 63.2 | 99.4 | 38.9 | 6.2 |
| | R | 28 | 75.0 | 99.4 | 27.1 | 1.4 |
| All systems | P | 37 | 97.4 | 99.7 | 27.0 | 2.6 |
| | R | 37 | 92.1 | 99.1 | 22.2 | 1.3 |

P—Pre-RAMMIE methods for all systems (including weekly data for GP in hours surveillance); O—Pre-RAMMIE methods for GP in hours using daily data during the 2012 Olympic period; R—RAMMIE method for all systems; EDSSS—emergency department syndromic surveillance system; GPIHSS—GP in hours syndromic surveillance system; GPOOHSS—GP out of hours syndromic surveillance system. Shading used to delineate separate syndromic systems.

number of events detected and sensitivity for each system individually, across the study period.

3.5 Specificity

Specificity was very high, being over or approaching 99% for every system in England (Table 3).

3.6 Positive predictive value

Positive predictive value, (the proportion of alarms that were linked to an incident), was not high, being 22.2% overall for the RAMMIE method, slightly lower than for earlier methods.

3.7 Timeliness

Timeliness improved overall, halving from a mean of 2.6 days (between the start of an incident and the first alarm) for previous methods to 1.3 days for RAMMIE, with the majority of systems improving; by 3 days for GPIHSS, 2 days for GPOOHSS and 5 days for NHS Direct.

4 Application

In the seven months from September 2013 to March 2014, the RAMMIE method was used over 2 million times to assess signals, creating nearly 25 000 prioritized alarms. By April 2014 nearly 12 000 different signals were being assessed each weekday, generating on average 134 alarms a day. In October 2013 the RAMMIE method was applied to the new remote advice service, NHS 111, incorporating local signals for the first time to telephone call data in England. The alarms were used to supplement the interpretation of surveillance data published in weekly syndromic surveillance reports to highlight trends and any findings of public health significance.

RAMMIE alarms have highlighted a number of increases in activity that have led to alerting a potential public health impact to colleagues within PHE, including:

- A rise in asthma indicators in September 2013.
- Increases in respiratory indicators at the beginning of April 2014 that coincided with national reports of poor air quality associated with pollution and dust from the Sahara.

5 Conclusions

The RAMMIE method has proved to be a reliable and effective method for generating automated alarms for syndromic surveillance. The multi-level models have enabled local models to be created for the first time across all systems. The method is able to cope with the wide range of data volume and seasonal variation in activity. The prioritization rules reduce duplication by emphasizing the most important signals and help to keep alarms to manageable levels. New surveillance systems incorporating new data sources have been incorporated into RAMMIE with a minimum of extra development work, as could future potential data sources for syndromic surveillance.

The ability to automatically analyse thousands of signals and identify and prioritize those with increased activity improves the surveillance team's ability to detect outbreaks and potential rises in syndromes of public health concern. Having a single method for all syndromic systems and a combination of 'historic' and 'spike' alarms also aid interpretation of the surveillance.

One potential limitation of this approach is that all automatically generated alarms could be seen as equally important, whereas a national alarm is likely to be of more concern than a local alarm, which may involve a small number of patients. It should be noted that the generation and prioritization of statistical alarms is only the first stage in a public health risk assessment process used by ReSST. Further work is planned to publish the other stages of this risk assessment.

By using incidents defined as those statistical increases identified previously, there may be a bias in favour of the previous methods. Despite this, RAMMIE was able to maintain high specificity whilst improving timeliness and system sensitivity. It is inevitable that the positive predictive value should be considerably less than 100%, due to the large number of signals covering a wide range of syndromes across many locations. Further work is planned to strengthen the prioritization rules with an emphasis on prioritizing signals based on a clinical assessment of the potential risk to public health; this will reduce the number of false alarms and increase the overall positive predictive value.

There is no definitive list of every real incident that should have been detected by syndromic surveillance and it is possible that some

of RAMMIE's false alarms were real events that were not detected by any other means (indeed the increase described associated with air pollution may be one example). Further work is planned to address such limitations, including identifying, through a variety of sources, a comprehensive list of local gastrointestinal outbreaks and assessing whether the syndromic surveillance systems were able to detect these incidents.

The timeliness measure used for verification, although simple could be biased in giving more weight to longer lasting incidents which can skew the results. Alternative measures are available (Jafarpour *et al.*, 2013) which could be used in future validation work.

A high performing statistical method can result in improved public health surveillance, with an increased likelihood of detecting key incidents early and less time wasted investigating false alarms.

Acknowledgements

The authors would like to acknowledge support from: NHS Direct, NHS 111 and HSCIC; Royal College of Emergency Medicine, EDs participating in the EDSSS, Ascribe Ltd and L2S2 Ltd; out of hours providers submitting data to the GPOOHSS and Advanced Health & Care; the contribution of EMIS and EMIS and the SystmOne GP practices to the GPIHSS, University of Nottingham and Clinrisk, TPP, and ReasearchOne. Also the help of all the ReSST team, including Sue Smith, Sally Harcourt, Paul Loveridge, Helen Hughes and Lee Carrilho who use the RAMMIE method on a daily basis and have helped to refine and improve it over the past year.

Conflict of Interest: none declared.

References

Baker, M. *et al.* (2003) Early warning and NHS Direct: a role in community surveillance? *J. Public Health Med.*, **25**, 362–368.

- Corberan-Vallet, A. and Lawson, A.B. (2011) Conditional predictive inference for online surveillance of spatial disease incidence. *Stat. Med.*, **30**, 3095–3116.
- D'Errigo, P. *et al.* (2007) Use of hierarchical models to evaluate performance of cardiac surgery centres in the Italian CABG outcome study. *BMC Med. Res. Methodol.*, **7**, 29.
- Elliot, A.J. *et al.* (2012) Establishing an emergency department syndromic surveillance system to support the London 2012 Olympic and Paralympic Games. *Emergency Med. J.*, **29**, 954–960.
- Harcourt, S.E. *et al.* (2012) Developing a new syndromic surveillance system for the London 2012 Olympic and Paralympic Games. *Epidemiol. Infect.*, **140**, 2152–2156.
- Harcourt, S.E. *et al.* (2011) Use of a large general practice syndromic surveillance system to monitor the progress of the influenza A(H1N1) pandemic 2009 in the UK. *Epidemiol. Infect.*, **140**, 100–105.
- Jafarpour, N. *et al.* (2013) Using hierarchical mixture of experts model for fusion of outbreak detection methods. *AMIA Annu. Symp. Proc.*, **2013**, 663–669.
- Lau, E.H. *et al.* (2012) Situational awareness of influenza activity based on multiple streams of surveillance data using multivariate dynamic linear model. *PLoS one*, **7**, e38346.
- Morbey, R.A. *et al.* (2014) Development and refinement of new statistical methods for enhanced syndromic surveillance during the 2012 Olympic and Paralympic Games. *Health Inf. J.*, **21**, 159–169.
- Robertson, C. *et al.* (2010) Review of methods for space-time disease surveillance. *Spatial Spatio-temporal Epidemiol.*, **1**, 105–116.
- Tokars, J.I. *et al.* (2009) Enhancing time-series detection algorithms for automated biosurveillance. *Emerg. Infect. Dis.*, **15**, 533–539.
- Triple, S. (2011) Assessment of syndromic surveillance in Europe. *Lancet*, **378**, 1833–1834.
- Unkel, S. *et al.* (2012) Statistical methods for the prospective detection of infectious disease outbreaks: a review. *J. R. Stat. Soc. Ser. A*, **175**, 49–82.
- Xing, J. *et al.* (2011) Method selection and adaptation for distributed monitoring of infectious diseases for syndromic surveillance. *J. Biomed. Inform.*, **44**, 1093–1101.