The practical application of statistical methods to improve the utility of syndromic surveillance in England

By

Roger Morbey

Submitted for consideration for the degree of

Doctor of Philosophy by published work – Health Sciences

Warwick Medical School

University of Warwick

February 2017

Table of Contents

Acknowledgements

Firstly I'd like to thank my wife, Chandra for supporting me in my studies when I decided to re-train as a statistician after more than 15 years working in the voluntary sector. I hope this thesis will confirm our belief that a person is most effective when applying themselves to a field that they both enjoy and uses their peculiar talents. Secondly I'd like to thank my colleagues, Gillian Smith, Alex Elliot, Nick Andrews, Andre Charlett, Sally Harcourt, Helen Hughes, Winnie Lee, Paul Loveridge, Sue Smith, Neville Verlander and other co-authors. I have tried in this submission to use pronouns to distinguish between my purely technical contributions from the collaborative team work. However, it would not be unreasonable to argue every 'I' should be a 'we' given my colleagues' wonderful support. I'd especially like to thank Gillian and Alex for guiding and encouraging me in this PhD and my academic writing. Finally I'd like to thank my supervisor, Noel McCarthy for his support and clarity in helping with the thesis structure, and Mihai Balanescu for his excellent course on academic writing. If, gentle reader, you find this submission sometimes too dry or difficult to read, rest assured it would have been much, much worse without their help.

Index of published work for consideration

Paper	Reference
1	Morbey, R. A., Elliott, A. J., Charlett, A., Andrews, N., Verlander, N. Q., Ibbotson, S., Smith, G. E. Development and refinement of new statistical methods for enhanced syndromic surveillance during the 2012 Olympic and Paralympic Games. *Health Informatics J* 2014; 21(2):159-169.
2	Morbey, R. A., Elliot, A. J., Charlett, A., Verlander, N. Q., Andrews, N., Smith, G. E. The application of a novel 'rising activity, multi-level mixed effects, indicator emphasis' (RAMMIE) method for syndromic surveillance in England. *Bioinformatics* 2015.
3	Morbey, R. A., Elliot, A. J., Charlett, A., Ibbotson, S., Verlander, N. Q., Leach, S., Hall, I., Barrass, I., Catchpole, M., McCloskey, B., Said, B., Walsh, A., Pebody, R., Smith, G. E. Using public health scenarios to predict the utility of a national syndromic surveillance programme during the 2012 London Olympic and Paralympic Games. *Epidemiol Infect* 2014; 142(5):984-993.
4	Elliot, A. J., Morbey, R. A., Hughes, H. E., Harcourt, S. E., Smith, S., Loveridge, P., Edeghere, O., Ibbotson, S., McCloskey, B., Catchpole, M., Smith, G. E. Syndromic surveillance - a public health legacy of the London 2012 Olympic and Paralympic Games. *Public health* 2013; 127(8):777-781.
5	Harcourt, S. E., Morbey, R. A., Loveridge, P., Carrilho, L., Baynham, D., Povey, E., Fox, P., Rutter, J., Moores, P., Tiffen, J., Bellerby, S., McIntosh, P., Large, S., McMenamin, J., Reynolds, A., Ibbotson, S., Smith, G. E., Elliot, A. J. Developing and validating a new national remote health advice syndromic surveillance system in England. *J Public Health* 2016: DOI: 10.1093/pubmed/fdw1013.
6	Todkill, D., Elliot, A. J., Morbey, R., Harris, J., Hawker, J., Edeghere, O., Smith, G. E. What is the utility of using syndromic surveillance systems during large subnational infectious gastrointestinal disease outbreaks? An observational study using case studies from the past 5 years in England. *Epidemiol Infect* 2016:1-10.
7	Hughes, H. E., Morbey, R., Hughes, T. C., Locker, T. E., Pebody, R., Green, H. K., Ellis, J., Smith, G. E., Elliot, A. J. Emergency department syndromic surveillance providing early warning of seasonal respiratory activity in England. *Epidemiol Infect* 2015:1-13.
8	Hughes, H. E., Morbey, R., Hughes, T. C., Locker, T. E., Shannon, T., Carmichael, C., Murray, V., Ibbotson, S., Catchpole, M., McCloskey, B., Smith, G., Elliot, A. J. Using an Emergency Department Syndromic Surveillance System to investigate the impact of extreme cold weather events. *Public health* 2014; 128(7):628-635.

9	Newitt, S., Elliot, A. J., Morbey, R., Durnall, H., Pietzsch, M. E., Medlock, J. M., Leach, S., Smith, G. E. The use of syndromic surveillance to monitor the incidence of arthropod bites requiring healthcare in England, 2000-2013: a retrospective ecological study. *Epidemiol Infect* 2016:1-9.
10	Bawa, Z., Elliot, A. J., Morbey, R. A., Ladhani, S., Cunliffe, N. A., O'Brien, S. J., Regan, M., Smith, G. E. Assessing the Likely Impact of a Rotavirus Vaccination Program in England: The Contribution of Syndromic Surveillance. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America* 2015; 61(1):77-85.

Statement of ethical considerations

There are no ethical considerations as this body of work was carried out within the context of anonymised aggregated public health data.

Submission declaration

I declare that the submitted material as a whole is not substantially the same as published or unpublished material that I have previously submitted, or am currently submitting, for a degree, diploma, or similar qualification at any university or similar institution. No parts of the works submitted have been submitted previously for any aforementioned qualification.

Glossary

- CDC – Centre of Disease Control and Prevention – Public Health authority in the United States of America (USA)
- ED – Emergency Department
- EDSSS – Emergency Department Syndromic Surveillance System
- GP – General Practitioner – term used for family doctors within the UK
- GPIHSS – GP In-Hours Syndromic Surveillance system
- GPOOHSS – GP Out-of-hours and unscheduled care Syndromic Surveillance system
- HPA – Health Protection Agency – National agency responsible for health protection in England between 2003 and 2013
- NHS – National Health Service in England
- NHS 111 – National telephone advice line for England
- NHS 24 – National telephone advice line for Scotland
- NHS Direct – National telephone advice line, replaced by NHS 111 in 2013
- PHE – Public Health England – National agency responsible for public health in England since 2013
- RAMMIE – Rising Activity Multi-level Mixed effects Indicator Emphasis method for aberration detection
- ReSST – Real-time Syndrome Surveillance Team
- RSV – Respiratory Syncytial Virus
- UK – United Kingdom of Great Britain (including England, Wales and Scotland) and Northern Ireland

Statement of candidate's contribution to the published work

Paper	Contribution of candidate	Lead co-author in agreement
4	Roger Morbey was a co-author on the peer reviewed paper 'Syndromic surveillance – a public health legacy of the London 2012 Olympic and Paralympic Games'. I was the lead author on this paper. Roger made a significant contribution to this paper through the description of the statistical alarm methodologies developed for the London 2012 Olympic Games, and provided guidance on the interpretation of alarms. Roger also contributed to the writing of the manuscript, providing comments on drafts and the final version.	Elliot, A.J.
5	Roger Morbey was a co-author on the peer reviewed paper 'Developing and validating a new national remote health advice syndromic surveillance system in England.' He advised on the statistical methods used to compare the new NHS 111 surveillance system with the previous NHS Direct system. Also, he applied the RAMMIE aberration detection method to the new system, incorporating variables to account for changes from the old system.	Harcourt, S.E.
6	Roger Morbey was a co-author on the peer reviewed paper 'What is the utility of using syndromic surveillance systems during large subnational infectious gastrointestinal disease outbreaks? An observational study using case studies from the past 5 years in England.' He advised on the surveillance coverage provided by different systems over time and its impact on the outbreaks studied. Furthermore, he provided details of the statistical alarms and action taken by ReSST during the outbreak periods.	Todkill, D.
7	Roger Morbey was a co-author for 'Emergency department syndromic surveillance providing early warning of seasonal respiratory activity in England.' He advised on the most appropriate statistical tests for determining estimates of the	Hughes, H.E.

	proportion of emergency department attendances which were attributable to respiratory pathogens.	
8	Roger Morbey was a co-author for 'Using an Emergency Department Syndromic Surveillance System to investigate the impact of extreme cold weather events.' He advised on and carried out the most appropriate statistical tests for determining significant increases in activity during periods of cold weather.	Hughes, H.E.
9	Roger Morbey was a co-author on the peer reviewed paper 'The use of syndromic surveillance to monitor the incidence of arthropod bites requiring healthcare in England, 2000-2013: a retrospective ecological study.' He advised on the availability and completeness of syndromic data from different systems relating to arthropod bites. Furthermore he consulted on the statistical methods used to identify the impact of temperature and the interpretation of the statistical test results.	Newitt, S.
10	Roger was a co-author for 'Assessing the Likely Impact of a Rotavirus Vaccination Program in England: The Contribution of Syndromic Surveillance.' He advised on the statistical methodology to be used in comparing seasons before and after the introduction of a Rotavirus vaccination and the interpretation of the results.	Bawa, Z.

1. Public health and syndromic surveillance

1.1 National public health surveillance systems

The World Health Organisation estimates infectious disease to be the second biggest cause of death worldwide, killing around 15 million people a year [1, 2]. Although developing countries suffer the biggest impact, all countries need to take precautions against the threat of emerging diseases and bioterrorism. This means providing the national infrastructure for health protection including an integrated surveillance system [3].

Syndromic surveillance has been defined as:
"A real-time (or near real-time) collection, analysis, interpretation, and dissemination of health-related data to enable the early identification of the impact (or absence of impact) of potential human or veterinary public health threats that require effective public health action; syndromic surveillance is based not on the laboratory-confirmed diagnosis of a disease but on non-specific health indicators including clinical signs, symptoms as well as proxy measures (e.g., absenteeism, drug sales, animal production collapse) that constitute a provisional diagnosis (or 'syndrome')" [4].

The Centre for Disease Control (CDC) in the USA pioneered the use of syndromic surveillance in the late 1990s [5]. Initially, they aimed to use syndromic surveillance to improve the early warning of infectious diseases and bioterrorism. Importantly, syndromic surveillance was able to report on increases in symptomatic patients earlier than laboratory reporting. Also, syndromic data might detect new threats to public health that they were not monitoring via laboratory tests. Many governments now include syndromic surveillance alongside traditional public health laboratory surveillance. However, over time as syndromic surveillance systems became more widespread, its role and aims have evolved.

The initial focus of syndromic surveillance on monitoring bioterrorism and infectious disease has widened to include "situational awareness" [6]. This is partly due to questions around the ability of syndromic surveillance to detect small bioterrorism incidents, but also because public health authorities have found it useful in new areas. Decision makers value the ability of syndromic surveillance to provide near real-time information during an incident with a public health impact. This includes not just surveillance of infectious

diseases like seasonal influenza but also non-infectious and environmental impacts. For instance, authorities have used syndromic surveillance to monitor the impact of heat waves, floods, forest fires, or contaminated water supplies. Furthermore, there is increasing interest in using syndromic surveillance to monitor longer term trends in chronic illness, injuries, drug abuse, or mental illness. Finally, in chapter 4.6, I consider a new role for syndromic surveillance to evaluate public health interventions.

Governmental public health surveillance, including syndromic surveillance, monitors a range of different potential threats. Their surveillance often covers many geographical sub-regions, and is not limited to one time period but part of ongoing surveillance. Syndromic surveillance systems are an example of the practical use of what has come to be called 'Big Data [7].' In other words, they use large datasets that grow daily and have considerable structural variation and variable data quality. Therefore, investigators need statistical methods in order to automatically detect aberrations and distil relevant information from the sea of data.

1.2 Syndromic surveillance systems in England

Within the United Kingdom, Public Health England (PHE) is a national government agency whose first function is to protect the public's health from infectious diseases and other public health hazards. Public Health England was formed in April 2013, incorporating the Health Protection Agency (HPA), established in 2003. Both the HPA and then PHE included syndromic surveillance in their integrated surveillance systems, alongside traditional test-based laboratory surveillance. The Real-time syndromic surveillance team (ReSST), has been responsible for HPA and PHE syndromic surveillance systems.

ReSST has developed and maintained a number of different syndromic surveillance systems that are described in detail elsewhere [8-12]. Figure 1 summarises the main characteristics of each system.

Figure 1: Syndromic surveillance systems maintained by ReSST

System		Period in use	Data source	Coverage
Remote health advice systems	NHS Direct	Oct 2001 – Sept 2013	Calls to telephone helpline 'NHS Direct'	100% of calls in England and Wales
	NHS 111	Sept 2013 - present	Calls to '111'	100% of calls in England
	NHS 24	Apr 2004 – Apr 2015	Calls to 'NHS 24'	100% of calls in Scotland
Family doctor or 'General practitioner' (GP) systems	HPA/QSurveillance	2003 – Mar 2013	GP database	45% of patients registered with a GP in England
	GP in-hours syndromic surveillance (GPIHSS)	Apr 2013 - present	Two providers of GP computerised records	66% of patients registered with a GP in England
	GP out-of-hours and unscheduled care syndromic surveillance (GPOOHSS)	Nov 2009 - present	Forty providers of out-of-hours services	Data received from 70% of English local authority areas
Emergency department syndromic surveillance system (EDSSS)		July 2010 - present	Up to 36 sites across England and Northern Ireland	Approx. 16% of English ED attendances

Data scientists use the terms 'volume', 'velocity' and 'variety' when defining the characteristics of 'Big Data.' These are also the characteristics of ReSST's data that make analysis a big challenge.

ReSST has had to develop information systems that can cope with the volume of data received and statistical detection methods that can cope with sudden changes in volume, many of which are unpredictable. On a typical week day during the past 12 months, ReSST received information on:

- 38,000 GP in-hours consultations mapped to a syndromic indicator
- 24,000 NHS 111 telephone calls
- 19,000 GP out of hours Read-coded consultations
- 7,000 Emergency department attendances

There is large structural variation within this data volume. For instance, during weekends and public holidays GP in-hours consultations drop to near zero whilst GP out of hours consultations and NHS 111 calls roughly double [13]. There is also seasonal variation, driven mainly by increased incidence of respiratory pathogens during winter months [14]. Also, potentially more challenging for analysis, are changes to provider coverage. Over time coverage has gradually increased for most systems, for instance ReSST piloted EDSSS with just four providers, gradually rising over a number of years until 36 sites were reporting daily. However, all systems are subject to occasional days with missing data due to local technical issues. Hence data volume can vary unpredictably on a daily basis.

The high 'data velocity' of syndromic surveillance requires fast analytical methods. ReSST receives data daily overnight on the previous day's health care activity. Then, during each working day ReSST must assess the new data to identify changes that are indicative of potential public health threats. Therefore, there is no time for the exhaustive data cleaning or validating data entry that would normally accompany retrospective research, outbreak investigations or the production of annual public health statistics.

ReSST's policy is to be opportunistic in identifying data sources, and none of the data used is collected specifically for syndromic surveillance purposes. Therefore, the data streams are subject to change, sometimes without any prior notice and ReSST has limited control over the data format. Hence, the third characteristic of ReSST's syndromic surveillance systems is variety (figure 1).

All ReSST's data providers anonymise their data but they aggregate it in different ways. Some provide data as individual records for each diagnostic code, others as one record per patient, whilst others aggregate it into geographical regions. Also, coding practices and the level of detail available varies across local providers within syndromic systems. Finally the information available on total activity for coverage denominators varies by system. For instance, for GP in hours we know the registered population of participating practices but not the total number of consultations, whilst for GP out of hours services the reverse is true.

In this thesis I describe how I have developed statistical methods to enable the surveillance of high volume high velocity data. Furthermore, how I have applied these methods to cope with the challenged of considerable structural variation and unpredictable data volume.

Throughout this thesis I use the following terminology that ReSST uses for its internal use and external reporting:

- **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: Diarrhoea; Hackney on 2nd Jan 2014.'
- **Baseline** – the expected level of 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.'
- **Alert** – an external notification by ReSST to outside its team that unusual activity has been seen that may constitute a threat to public health. E.g. a key message in a weekly bulletin of 'fever consultations in children aged 1 to 4 years old are above seasonally expected levels.'

1.3 The use of statistical methods within syndromic surveillance in England

Since 2010 I have been employed as the Statistical Project Lead for ReSST, working as an 'embedded statistician' within a multi-disciplinary surveillance team. In this thesis I will demonstrate how I have used statistical methods to enhance syndromic surveillance in England. Importantly, this will include approaches that improved ongoing processes and practice. Consequently, I have used statistical methods for initial development and as part of the cycle of continuous improvement; where research into past events informed future developments (figure 2).

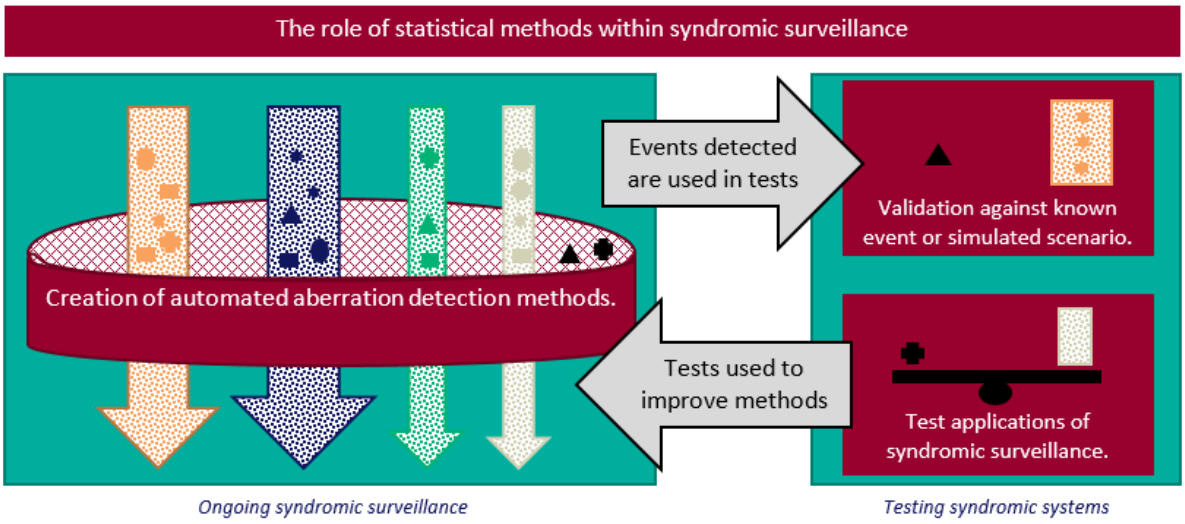

Figure 2 (Polygons epitomize events detected within the streams of syndromic data, different colours for different systems).

In the next section (1.4) I review the published literature of statistical methods applied to public health surveillance worldwide. Specifically, I distinguish between research that is

based on the potential use of a new statistical technique with evidence of practical use within an ongoing syndromic surveillance system. Also, I identify whether researchers have used statistical methods to detect one specific incident or a range of different events.

I have structured this thesis around three practical uses of statistical methods, described as 'creation, validation and application'. Firstly, in chapter 2, I describe how I have **created** new aberration detection methods and refined them through routine use. Secondly, in chapter 3, I demonstrate how we used statistical methods to **validate** syndromic surveillance systems. Finally, in chapter 4, I describe how we **applied** syndromic surveillance to detect different types of events and interventions, and used statistical methods to assess the results.

1.4 Statistics and surveillance, a literature review[1]

1.4.1 Aberration detection methods created for public health surveillance

Epidemiologists and analysts have created statistical tests called 'aberration detection methods' for public health surveillance. These methods are also referred to as 'outbreak detection methods' or 'biosurveillance'. Different reviewers have used different approaches to categorise the large number of diverse published aberration detection methods[15-20].Some reviewers have categorised methods by data type; spatial, temporal, or a combination of both. Some have characterised methods as either retrospective or prospective, depending on whether researchers had compared data with previous years. Alternatively, other reviewers have categorised methods by statistical methodology. I will use this last approach to describe four categories; statistical process control charts, model-based methods, time series analysis and cluster analysis. I have chosen this form of categorisation to highlight the range of statistical methods used for syndromic surveillance.

Manufacturing industries first developed statistical process control charts to test whether a process was 'in control'. Control charts are prospective methods, where investigators monitor a measure over time to identify any significant changes. Methods applied to syndromic surveillance include Shewhart range charts, cumulative sum (CUSUM) and exponentially weighted moving averages[21, 22]. Specifically, the widely-applied Early Aberration Reporting System (EARS) developed by the Centre of Disease Control and Prevention (CDC) incorporated a number of control chart methods[15, 23-25].

[1] It is beyond the scope of this research to provide a 'systematic' literature review, instead I provide an overview of the key research undertaken in this field.

Researchers have used model-based methods in order to determine baselines for surveillance data in the absence of outbreaks. Their methods can be classed as either regression or state change models. By using regression models, researchers have explicitly modelled seasonality. For instance, the Stroup or Farrington regression method uses only past data from a comparable time of year[26-29]. Alternatively, modellers included seasonality variables when using Generalised linear mixed models (GLMM) and Serfling models[30]. Researchers have also created and applied specialised regression techniques to public health surveillance, including Poisson regression[31], and wavelet algorithms[32, 33]. By contrast, researchers have used state change models to determine whether current data best fits a model with or without an outbreak present. For instance, developers of the open source R statistical software package 'surveillance' included Hidden Markov models and semiparametric methods[34].

Researchers have used time series analysis to model and predict activity utilising the structure of health datasets, in particular autocorrelation. Time series methods could either be retrospective, or prospective. Examples that researchers have applied to syndromic surveillance include; autoregressive integrated moving average methods, integer-valued autoregressive models and the method of analogues where current data is matched to the most similar previous year[35].

The fourth type of statistical method researchers used was cluster methods, including scan statistics. Statisticians used cluster methods to identify non-random distributions of cases across time or spatially. Although these methods could be applied to just one location or time period they were more usually applied to both, for example using the Kulldorff statistic[36]. Two cluster methods created specifically for syndromic surveillance were 'What is strange about recent events' and the 'Population wide anomaly detection and assessment' (PANDA) method[15].

The four broad categories described above are not exhaustive and some methods straddled categories. For example, statisticians have applied Bayesian techniques to modelling and cluster methods to improve detection rates or computing speed[19, 37]. Furthermore, some surveillance programmes developed for syndromic systems involved using a suite of different methods.

1.4.2 Validation of detection methods

The utility of an aberration detection method depends both on its ability to detect events and its usability when applied to ongoing surveillance. Whilst there is considerable published literature validating different detection methods, the majority does not include evidence of use within ongoing surveillance. First I will describe how methods were validated before looking at evidence of ongoing use.

Usually researchers validated methods either by simulations or through retrospective application to a known event. The simulation approach had two advantages; researchers quickly tested a wide range of possible scenarios and they did not need an historical dataset[22, 23, 25, 32, 38-42]. Historical datasets may be difficult to obtain, particularly if researchers want to test a scenario that has not happened yet. The main limitation of simulations was dependence on the scenario assumptions. For a modelled scenario there will always be some uncertainty as to whether the model accurately reflects real events and patient presenting behaviour, and therefore whether we would detect real events. By contrast, evaluation against a specific historic example resulted in more certain conclusions[8, 30, 31, 33-35, 43-53].

Published applications of aberration detection methods are often specific, with many focussing on the influenza A (H1N1) 2009 pandemic[8, 20, 31, 34-37, 45, 46, 49, 52-55], using all the method types discussed above. Other published examples included as diverse threats as food poisoning[47], bioterrorism[38, 39], and meningitis[51]. Often researchers evaluated a method in a specific setting with the stated aim that it could be applied in a more general context[37].

Many researchers published new methods to argue the case for their adoption, or to show they could have detected a recent event. In these cases it is not surprising that there was no evidence of existing use within an ongoing surveillance system. However, evidence of use within ongoing surveillance systems can sometimes be inferred when methods were modified or compared with existing practice. For example, researchers have published modifications to the Early Aberration Reporting System (EARS) which describe how it is applied in the USA surveillance system called BioSense.[24]

One non-syndromic area where there is good evidence for practical use of statistical detection methods to ongoing systems is laboratory surveillance and the notification of infectious diseases[26-28, 56-58]. For example, epidemiologists monitoring notifiable diseases in China have used three different methods, providing evidence on the number of statistical

alarms and verified outbreaks over two years [56]. In general, laboratory surveillance systems have been established longer than syndromic surveillance systems and this may account for the better published evidence of practical use.

1.4.3 Evaluation of syndromic surveillance systems

In the United States the CDC pioneered the use of syndromic surveillance systems and drafted the first guidelines for how systems should be evaluated[5, 59]. These guidelines provided a comprehensive list of evaluation measures for each stage of the surveillance process. Specifically, data quality measures included representativeness and reliability; detection measures included sensitivity and timeliness; whilst 'experience' measures included usefulness, acceptability by stakeholders and cost. Several different countries have used the guidelines to evaluate their systems[60-63].

Some evaluators have focused on just one of the CDC evaluation measures. For instance, the Royal College of General Practitioners sentinel network in England has recently been evaluated for representativeness[64]. By contrast, researchers from North Carolina, USA, used stakeholder questionnaires to measure utility in terms of what public health actions had resulted from syndromic surveillance reports[65].

Many investigators have validated new syndromic surveillance systems by testing for correlations with existing surveillance data sources: for example, using known outbreaks[50, 66], laboratory surveillance reports[67-70], or an established syndromic system[30, 71, 72]. Recently, social media and the internet have been used as a new source of surveillance data. However, these sources are more sensitive to changes caused by media reporting, therefore researchers have analysed correlations with both disease incidence and with media reports[73, 74]. Analysts have used a variety of statistical methods to test for correlations, including: simple linear regression models, [66, 68] Generalised Additive Models, [69] Support Vector Regression[71], and the non-Parametric Spearman's rank correlation methods[60, 70].

Analysts also estimated improvements in timeliness when validating syndromic systems. Specifically, they incorporated lags in their models, used cross-correlation and cross-correlograms[60], or compared the timing of outbreak peaks[72].

Reviewers have used statistical methods to evaluate and compare multiple existing syndromic systems. Examples include; comparing data sources for timely influenza surveillance[67], comparing systems using drug sales data[75], and using Qualitative

Comparative Analysis[76]. In the last example, researchers combined literature searches with site visits and interviews.

Overall, as outlined above there are numerous published examples of different statistical methods used in a public health context. However, there is not yet an emerging consensus as to which methods work best for surveillance. Also, the majority of the literature does not include details of practical use within an ongoing surveillance system. This thesis provides a case study of the practical use of statistical methods to enhance syndromic surveillance systems. I describe how I have created aberration detection methods and refined them based on their use in ongoing daily surveillance. Furthermore, I have used statistical methods to validate not only the detection methods but the wider performance of syndromic systems and not just for a single incident but across a wide range of applications and for continuous use.

2. Creation and refinement of aberration detection methods

The wide range of different aberration detection methods in the published literature is a sign that no single method had been accepted as a universal standard for syndromic surveillance data. Moreover, researchers have tested several different methods concurrently within one surveillance system. This approach accepted that methods have different strengths and weaknesses, and that overlapping the methods weaves a tighter detection net than using just one method. Accordingly, I have used a range of methods for syndromic surveillance in England. Initially, I created a combination of retrospective regression models and prospective control charts, depending on availability of historical data (chapter 2.2). Subsequently, I have combined the strengths of both methods, creating a single approach across all the English surveillance systems (chapter 2.4).

I have used the aberration detection methods created in ongoing surveillance systems. Importantly, I was able to develop these methods whilst working as part of the team using them for daily surveillance. As a result, I could test, modify and validate methods as part of a continuous cycle of improvement.

2.1 Enhanced surveillance during the 2012 London Olympic and Paralympic games

The first syndromic surveillance system set up by ReSST in England collected data from the NHS Direct telephone helpline. Initially, we based our statistical alarms on fixed influenza

thresholds and thresholds created by Poisson regression models[77]. Subsequently, a GP in-hours syndromic surveillance system (HPA/QSurveillance) was created in 2004[9]. In this case, statistical alarms were spatial clusters, based on the ratio of local activity to the current national average.

The 2012 London Olympic and Paralympic games was the largest ever mass gathering held in the UK. Furthermore, a large number of international visitors plus media attention made it a priority for health protection surveillance. Therefore, the Health Protection Agency enhanced the resource put into syndromic surveillance prior to the games. Specifically, ReSST created two new surveillance systems with data from GP out-of-hours and unscheduled care services including walk-in centres[11] and from emergency departments.[10]

The enhancements to syndromic surveillance in England for the Games were fourfold. Firstly, emergency department data enabled ReSST to monitor patients with the most severe illnesses. Secondly, the inclusion of data from walk-in centres and emergency departments improved health surveillance of international visitors. Thirdly, these new data sources improved weekend surveillance. Finally, the HPA/QSurveillance system was upgraded from a weekly to a daily data feed in line with other systems. Consequently, ReSST was able to provide daily reports, 7 days a week, during the Games period. Therefore, we needed to create new aberration detection methods for the enhanced surveillance systems. ReSST created my post of Statistical Project Lead in December 2010 to create and maintain the statistical methods required.

2.2 Development and refinement of new statistical methods for enhanced syndromic surveillance during the 2012 Olympic and Paralympic games (paper 1)

I created new aberration detection methods for the enhanced HPA/QSurveillance GP in-hours system and for the two new syndromic surveillance systems. The choice of methods used was primarily dictated by the data available. Importantly, the new systems only had limited comparable historical data, therefore I used a prospective method, Shewhart control charts. By contrast, I used the retrospective Stroup/Farrington method for the older HPA/QSurveillance system. The Stroup/Farrington method uses weekly data from previous years at the same time of year to calculate baselines and alarm thresholds [26].

I chose a baseline for the Shewhart control charts that was the mean of the previous two weeks for GPOOHSS and three weeks for EDSSS. I did not need to model seasonal effects

because baselines only used recent data. However, there were very large 'day of the week' effects for GPOOHSS; typically there was twice as much daily activity at weekends and during public holidays. Therefore, I created separate baselines for holidays and working days for this system. Also, daily coverage in these new systems could vary so I excluded from the baselines any sites that had not provided data on every day.

For the Stroup/Farrington method I used a baseline that was a five-week rolling average from the previous three years. In other words, the baseline for week 30 2012 was the average of weeks 28-32 for 2009-2011. (Data prior to 2009 was not comparable due to changes in the geographical reporting boundaries.) The HPA/QSurveillance system had better coverage and a larger volume of data than the new systems, therefore it was possible to have local signals as well as regional and national ones. However, this meant we needed to check 3,500 signals daily. Therefore I used the Benjamini and Hochberg method for multiple testing to limit the number of alarms generated. This method accounts for excessive false-positives associated with multiple testing by adjusting the thresholds of what is considered a significant p-value [78]. Finally, I dealt with day of the week effects by using a 7 day moving average adjusted for public holidays to compare with the historical weekly data.

ReSST tested the methods prior to the Games period as part of a rehearsal of working practices during the games. In this way, we resolved teething problems, including IT issues and computing speed, prior to the games. However, refinements were still necessary up to and during the games to ensure that the volume of statistical alarms generated were manageable. For example, a slight rise in upper respiratory tract infections early in the Games period led to a large increase in alarm volume. Consequently, I was able to revise the thresholds, having identified the cause which was a problem with the criteria for excluding unusual historical data. Thereafter, the number of alarms during the Games were at manageable levels.

During the Games, ReSST was able to provide reassurance that no major incident had occurred and to identify minor examples of increased incidence. These examples included, increased levels of pertussis, unusually high numbers of severe asthma consultations in June and heat-related consultations in July. Additionally, ReSST were able to evaluate the methods in terms of the operational effectiveness of syndromic surveillance. During the Games we used the methods 347,754 times and generated nearly 4,000 statistical alarms over 73 days. (These alarms do not all represent distinct events. For example, the raised

incidence of pertussis caused multiple alarms, over many days in many different locations.) Subsequently, I assessed the alarm rates to identify any bias in alarm volume, whether by day of week, location or syndrome. As a result, we gained evidence and experience that would enable future improvements to the detection methods.

2.3 Post-games, the Olympic legacy

London won their bid to host the 2012 Games by including a promise to create a positive legacy. For ReSST that legacy meant continuing with the enhanced surveillance systems after the games. Also, continuing with improved surveillance practices developed for the Games, including team rotation and a formal risk assessment system.[79]

Central to the enhanced surveillance systems were the new aberration detection methods. Without these methods, surveillance would have been restricted to reviewing a limited number of national signals, and reliant on the availability of individual expertise. I cover the legacy of the Games, including the validation of detection methods in more detail in chapter 3.3.

Introducing new surveillance systems and developing new detection methods during the Games period led to two new problems for ReSST, alarm volume and alarm interpretation. Firstly, increased volumes of data and the ability to apply methods to local areas led to a big increase in alarms. Furthermore these problems increased as we expanded the methods to a more general context, moving from detailed local surveillance of London during the Games, to year-round surveillance of the whole of England. Secondly, using different aberration detection methods for different systems, led to confusion over perceived conflicting alarms across systems.

Seasonal influenza surveillance provides an example of the confusion over alarm interpretation. Each of ReSST's four English surveillance had signals designed to detect changes in community incidence of influenza. Typically, these signals would rise gradually during September and then grow exponentially when a seasonal outbreak occurs, with peak incidence varying considerably between seasons. Control chart methods, like those originally used for GPOOHSS and EDSSS would produce alarms as soon as activity started increasing. However, the other system's methods using historical data would not alarm because it was not unusual to see increases at this time of year. Then, later in the season after the signals had peaked the GPOOHSS and EDSSS signals would stop alarming because activity was no longer rising. However, the other systems would continue to alarm until

activity had dropped down below thresholds based on past seasons. Whilst these differences in aberration detection methods were explainable it was not intuitive to the users.

I created a new detection method to cope with the increasing demands and requirements of the surveillance systems. Furthermore, more historical data was now available for the newest systems, enabling retrospective methods. Importantly, we applied the same method to all systems to ensure consistency of interpretation and we were able to add new systems with minimum extra work. Also, the new method included prioritisation rules to ensure that the number of alarms generated were manageable.

2.4 The application of a novel 'rising activity, multi-level mixed effects, indicator emphasis' (RAMMIE) method for syndromic surveillance in England (paper 2)

I created and implemented the new 'rising activity, multi-level mixed effects, indicator emphasis' (RAMMIE) method. The method's name described three of its properties. Firstly, the method detected unusual rises in activity. Secondly, the method used multi-level mixed effects regression to model baseline activity. Finally, the method included prioritisation rules to limit the number of alarms and ensure that the most important syndromic indicators were emphasised.

I designed RAMMIE to incorporate the strengths of both the prospective and retrospective methods developed for the Games (chapter 2.2). First of all, I modelled baseline activity using regression models to capture seasonality and day of the week effects from historic data. Then I created a 'historic' alarm threshold to warn of activity that was significantly higher than expected for the time of year. Additionally, I created a 'spike' alarm threshold based on the ratio of recent activity to the modelled baseline: Thereby warning of activity that had recently risen significantly, independently of whether it was high or low for the time of year. The availability of these two different alarm thresholds for each syndromic indicator provided consistent useful information for the surveillance team. Primarily, we were only interested public health threats leading to in increases in activity, hence thresholds based on upper confidence intervals. However lower thresholds could be created in a similar manner, e.g. for data quality monitoring.

I applied RAMMIE to all syndromic indicators at a local, regional and national level. Therefore the method needed to be able to model activity with a range of different

seasonal and weekly cycles, and with a very wide range of scales. For instance, the model for upper respiratory tract infection consultations to GP in-hours across England involved tens of thousands of daily cases. By contrast, local signals for rarer syndromes like measles, included a majority of days with zero cases. In order to model these low numbers I used a multi-level mixed effects model with the natural hierarchy of national signals sub-divided into geographical regions and then further divided into local signals. In this way, the local models could 'borrow power' from the national datasets, creating models that would otherwise not be possible using standard computing regression techniques. Additionally, this method was more powerful than the Stroup/Farrington method when only a couple of years' historical data were available. This is a novel multi-level approach I have not seen described in published literature on syndromic surveillance.

I used Poisson and negative binomial regression models for the count data, but using total consultations or population as an offset to account for daily fluctuations in system coverage. I included independent variables for public holidays, day of the week and month of the year in all the models along with system specific variables. For example, the day after a public holiday was included as a binary variable in the HPA/QSurveillance models.

In using parametric models I had to accept various assumptions about our data: specifically, the independence of daily counts and homoscedasticity across different locations. Furthermore, due to the limitations of computing speed, many hierarchical models used a Poisson rather than a negative binomial distribution. Therefore, these models assumed variance was equal to the mean.

Another limitation resulting from applying regression models across thousands of different signals is that I could not individually examine each model. Normal practice, after fitting a regression model is to explore residuals plots for any signs of non-random activity. Also, some of the independent variables may not have had a significant contribution to every individual model. However, any major problems with individual models would become apparent during daily surveillance, allowing me to make adjustments as needed.

I set the historical alarm threshold to be the maximum of either; three standard deviations above the mean, three times the square root of the model above the mean (i.e. a Poisson distribution), or a count of three cases. In this way, the alarm rate was set at around 1%, and we avoided excess alarms from signals with low baseline counts.

In addition to being an aberration detection method, I included in RAMMIE prioritisation rules to automatically select the most important alarms for further manual investigation. For instance, some syndromes are seasonal so were only prioritised during the summer or winter, e.g. sunstroke or hypothermia. Also, to prevent duplication of effort, when one event caused multiple alarms I only prioritised the most useful. For example, during seasonal influenza epidemics, only the national influenza-like illness alarms were prioritised, not the more general respiratory syndromes or multiple local alarms.

I initially validated the RAMMIE method by testing how well it would have detected events during the period April 2012 to July 2013. I chose events that had independent verification from non-syndromic sources, such as laboratory reporting or air quality monitoring. RAMMIE performed well, with a high sensitivity and specificity as well as being quicker to detect increases than the previously used methods.

The RAMMIE method enabled automated aberration detection of all syndromic surveillance systems at a local level for the first time. Between September 2013 and March 2014, we used RAMMIE over 2 million times, generating 25,000 prioritised alarms within its first seven months.

2.5 The routine use of detection methods in daily surveillance

PHE's automated statistical aberration detection methods enable daily surveillance of far more syndromic time series than would be possible manually. Also, I have used the detection methods to validate systems[12] and to retrospectively test applications of syndromic surveillance[80, 81]. Furthermore, they were used in the validations discussed in chapters 3.2, 3.3, and the gastrointestinal applications in chapter 4.2.

I developed the RAMMIE method to be flexible enough to enable modifications. For instance, priority rules have been modified and new independent variables added. Specifically, I introduced Christmas (December 25th) as a new variable because we found its impact was different to other public holidays. Also, I added variables to allow for the impact of a new rotavirus vaccine on gastrointestinal indicators (chapter 4.5).

We have applied RAMMIE to Ambulance dispatch data to test a new pilot English syndromic surveillance system (accepted for publication in Prehospital and Disaster Medicine 2017). Also, we have applied it to French emergency department data as part of a collaboration to compare similar events in Paris and London. Future developments will

include the creation of age-specific signals to improve the sensitivity and timeliness of detection.

3. Validation of syndromic surveillance systems

3.1 The need to validate evolving systems

The syndromic surveillance systems used by ReSST in England have evolved over time. We have developed new indicators, coverage has increased and data providers have changed. Consequently, I have (re-)validated systems and describe three examples in this chapter.

The first validation was to test if systems were ready for the 2012 Olympic and Paralympic Games. It was important not to provide false reassurance during the games, so we developed scenarios to determine what scale of events we could detect.

After the Games, we carried out a second validation to evaluate how the syndromic systems had functioned. We described the detected events and the new surveillance routines developed for the Games.

In the third example, I describe our re-validation of our Remote Health Surveillance System. This followed a major change in the way telehealth services were provided in England during 2013. Although ReSST provided continuity of indicator names and reporting, we also needed to validate the system to ensure it was still able to detect outbreaks.

3.2 Using public health scenarios to predict the utility of a national syndromic surveillance programme during the 2012 London Olympic and Paralympic games (paper 3)

During the 2012 London Olympic and Paralympic Games, ReSST provided daily reports from syndromic surveillance systems for the officials in charge of public health at the Games. Together, ReSST and its stakeholders had agreed the format of these reports, to ensure key information was included and messages were clear and unambiguous. To this end, prior to the Games I evaluated the systems using agreed scenarios. Specifically, it was important that all stakeholders would be aware of what scale of incident the syndromic systems could detect; then they would know how much reassurance we could provide when we reported a daily message of ‘nothing detected’.

The stakeholders and ReSST agreed on five scenarios that would cover a range of different types of events that syndromic systems should be able to detect:

- Contamination of a local water supply by *Cryptosporidium* oocysts.
- A localised food poisoning incident involving scombrotoxin.
- An outbreak of a new variant of influenza, arriving with Games overseas visitors.
- An intentional release of botulism into the food chain at a Games venue.
- An intentional release of anthrax via aerosol dispersion at a transport interchange.

I had to calculate, for each scenario, how many extra cases would be recorded by our syndromic surveillance systems. In order to do this, I sub-divided the scenarios into a number of stages. First, I created an epidemic curve to estimate the number of people symptomatic each day after exposure. Second, I estimated how many people with symptoms would access different types of health care; teleheath, GPs or emergency departments. Then, I considered what proportion of patients would be covered by services participating in our syndromic systems. Finally, I modelled the different symptoms that patients might present with and how these would be aggregated to our syndromic indicators. The resulting scenario datasets included explicit model assumptions that I investigated in a sensitivity analysis.

I used the scenario data in simulations to determine what could be detected and how quickly. I combined the extra scenario cases with expected baselines for syndromic data, based on historical data, including simulated random background variation. Then, I applied to the simulated combined datasets the aberration detection methods developed for the Games. Finally, I used the resulting alarms from the detection methods to calculate the probability of detection for each scenario.

I was able to vary the scenario model estimates to see how they impacted on detection rates. For instance, I could vary the number of symptomatic patients to represent different scales of events. Therefore, I could calculate the smallest size of incident that we would reliably detect.

I presented results in terms of timeliness as well as scale of incident detectable. Also, I could compare results across syndromic systems in order to find which system was the most sensitive. For instance, the smallest number of symptomatic patients that could be detected in the majority of simulations was; 1,100 within one London primary care trust (PCT) under the scenario of *Cryptosporidium* polluting the water supply, 43 within one

London PCT due to a scombrotoxin food-poisoning scenario, 65 for the scenario of botulism poisoning at a Games food outlet, and 510 for a scenario of anthrax exposure at a transport interchange. Furthermore, the GP out-of-hours system was the most sensitive for the *Cryptosporidium* scenario whilst the Emergency departments system was the most sensitive for the other three more severe illnesses. Under the influenza scenario, cases would grow exponentially and therefore the key question was how quickly it would be detected, not what size of event would be detected. The first system to detect influenza under the scenarios was GP out-of-hours, 15 days after the arrival of new cases.

The scenario validations provided Games stakeholders with the information they required about the syndromic surveillance systems. Additionally, we learnt a number of key points from this research that would have wider implications. Firstly, the research supported our hope that emergency departments would prove to be more sensitive for the most severe illnesses. Secondly, we found that the systems involving more patient consultations, particularly GP in-hours, were more sensitive for small local events. Also, unsurprisingly, we found that it was much easier to detect events when they were clustered into a small geographical area or over a short period of time. We could improve mass gathering surveillance if systems could distinguish between patients who had attended and others. Finally, we discovered that there was considerable uncertainty over how patients would present to different health care providers, and more research was needed in this area.

3.3 Syndromic surveillance - a public health legacy of the London 2012 Olympic and Paralympic Games (paper 4)

A mass gathering, like the 2012 Games, provided a unique opportunity to validate syndromic surveillance systems in England. Normally, calculating the sensitivity and specificity of multi-purpose syndromic surveillance systems is complicated by the lack of a definitive list of what incidents did or did not occur that should have been detected. However, during the 2012 Games intense media and public health scrutiny meant it was less likely that substantial incidents would have been unrecorded, and risk assessments provided detail on what type of incidents we needed to detect. Furthermore, there were a number of increased risks during the games making incidents more likely; for instance, increased disease transmission due to travel and social mixing, increased risk of bioterrorism and greater pressures on infrastructure.

The retrospective validation of English syndromic surveillance systems involved detection methods that we used at the time. Specifically, we asked if the aberration detection methods generated alarms and were these important enough to alert the public health authorities, By using this approach we were able to validate the whole surveillance process and not just in theory but based on actual key messages recorded at the time.

No major public health incidents occurred during the games, nor were any detected by non-syndromic surveillance, and ReSST did not issue any false alerts suggesting they had occurred. We noted two minor increases in activity in the period immediately preceding the games that led to ReSST alerts. Firstly, an unusual increase in asthma and difficulty breathing symptoms led to an alert. The exact underlying cause is still unknown, but investigations have ruled out some of the common aetiologies, e.g. respiratory pathogens and environmental pollutants. A brief period of hot weather caused the second alert, during which ReSST was able to report on the resulting impact on health services, and to reassure that these were within seasonal expectations.

Our validation of systems after the 2012 Games showed the value of its legacy to syndromic surveillance. In addition to statistical aberration detection methods, our development of a formal risk assessment process became a key component of post-games daily surveillance [79]. Furthermore, the surveillance routines developed have generated considerable interest in other countries for routine surveillance and surveillance of mass-gatherings.

3.4 Developing and validating a new national remote health advice syndromic surveillance system in England (paper 5)

In 2001, ReSST developed a remote health advice syndromic surveillance system based on telehealth calls to the national NHS Direct service. During 2012-13 NHS Direct was gradually decommissioned and replaced with a new service, NHS 111. The new service used a free three-digit number accessible to all patients and was designed to be the gateway to out-of-hours urgent care provision. ReSST worked to ensure continuity between the two surveillance systems by creating syndromic indicators that were as similar as possible to those from NHS Direct. However, the new syndromic surveillance system required validation to ensure that the indicators detected the same events as the previous system.

We compared the data collected for NHS 111 with that from NHS Direct and also a Scottish system, NHS 24 which was a telephone helpline service similar to NHS 111. In particular, we compared day of week effects, age distributions and the proportion of calls aggregated to the different syndromic indicators. I applied the RAMMIE method (chapter 2.4) to the NHS 111 system, incorporating historical data from NHS Direct. In the process, I included specific independent variables to account for the change in services and a period during 2013 when there was a rapid decline in coverage of the NHS Direct service. Finally we completed the validation of the new system using two examples of its application, flooding in winter 2013 and air pollution events in 2014.

We found an increased volume of calls for the new NHS 111 system and a change in the age distribution compared to the previous system. Specifically, a higher proportion of elderly patients used NHS 111 compared to NHS Direct, probably due to its additional role as a gateway to other urgent care services. This change in age distribution was reflected in changed proportions of calls between different syndromic indicators. Also, there was a decrease in the proportion of calls that related to fever, due to a different coding practice, whereby patients were only assigned a fever diagnosis if they had no other symptoms.

We recorded statistical alarms for respiratory increases using RAMMIE during the winter of 2013-14, which coincided with RSV and influenza activity in the community seen in other systems. We also used the RAMMIE method to monitor diarrhoea and vomiting indicators during the winter floods (we found no significant outbreaks). Finally, we detected alarms between 3rd and 8th April 2014 for calls related to cough and difficulty breathing which coincided with a period of poor air quality.

The validation of NHS 111 showed that although there were significant differences compared to the former NHS Direct system, it was still an effective public health surveillance system. Furthermore it showed how, under certain conditions, the RAMMIE method could utilise historical data from one system to bolster a new system with limited historical data.

3.5 Statistical evidence for the validity of syndromic surveillance systems

Statistical methods were important in quantifying the detection capabilities of syndromic surveillance systems during validation. Specifically, published research used three types of methods; correlations with other data sources, scenarios and simulations, and validation of

aberration detection methods (see chapter 1.4). I used all these different approaches within the validations of English syndromic surveillance systems.

I used scenarios and simulations to validate our syndromic systems prior to the Olympic Games. In this way, I quantified what size of event we could detect and how quickly we were likely to detect it. This was very important to our stakeholders, enabling them to plan and assess syndromic surveillance's role alongside other surveillance systems. We were able to explicitly state model assumptions which revealed the key factors affecting our ability to detect events and the main sources of model uncertainty. Furthermore, we could use knowledge gained about system capabilities and factors affecting them in a more general context.

After the 2012 Games we validated our systems by considering the alarms produced by our aberration detection methods and the small number of alerts. Using the enhanced scrutiny provided by the Games we could show that no major incident went undetected. Furthermore, we validated the systems as whole, not just the automated alarms but also the process of investigation and interpretation. Consequently, we could show that we issued no false alerts to public health authorities. This validation established the legacy of the Games for syndromic surveillance, we continued with enhanced surveillance extending it to the whole of England.

We used our aberration detection methods again to validate a new syndromic surveillance system using calls from NHS 111. However, we also validated the new system by correlating with other data sources. Firstly, we compared the system with the previous NHS Direct system and discovered differences in age distributions and the proportion of calls aggregated to different syndromes. Secondly, we compared the timing of peak activity and statistical alarms for the new systems with known events. Specifically, seasonal RSV and influenza activity and air pollution incidents identified through other syndromic and non-syndromic systems.

I have used statistical methods to validate syndromic surveillance systems and the aberration detection methods developed for them. Importantly, these methods provide evidence of associations beyond those suspected by investigation through descriptive epidemiology. Moreover, we were able to quantify the detection capabilities of our systems, which would not have been possible without statistical methods.

4. Applications of syndromic surveillance

4.1 Choosing the appropriate statistical method for each context.

Although primarily developed for surveillance of infectious diseases, PHE has also applied its syndromic systems in a range of areas, including environmental incidents and assessing public health interventions. For each application we needed to apply statistical tests to quantify detection capabilities and test the evidence for syndromic surveillance's utility. Moreover, each application had a unique context which determined which statistical tests were the most appropriate to use. In this chapter I describe five different applications to illustrate the range of applications and statistical methods I have used in collaborative studies.

4.2 What is the utility of using syndromic surveillance systems during large subnational infectious gastrointestinal disease outbreaks? An observational study using case studies from the past 5 years in England (paper 6)

ReSST uses syndromic surveillance systems for surveillance of gastrointestinal disease, including seasonal rises in norovirus and rotavirus activity, at a national level. However, we did not know how effective the systems would be at detecting subnational outbreaks. Therefore, we identified examples of subnational outbreaks and investigated whether we had detected them with syndromic surveillance systems. Specifically, we choose eight outbreaks from a database of over 100 investigated outbreaks, four selected randomly and four purposively sampled. We selected the outbreaks based on size (over 75 cases) and duration (less than 3 weeks), hypothesised to be outbreaks most likely to be detected by syndromic surveillance.

In this application we were primarily concerned with the utility of our syndromic surveillance systems. Therefore, we studied each stage of the investigation process to identify what factors would determine our ability to detect an outbreak. Firstly, we considered coverage, which varies by syndromic system and has changed over time. Secondly, we used descriptive epidemiology to examine the local syndromic data during outbreak periods and compared it with preceding and subsequent periods. Thirdly, we retrospectively examined our aberration detection methods to review the statistical alarms during outbreak periods. Finally, we looked at our reporting records for the outbreak periods to identify whether we had issued any alerts for the outbreaks as they occurred.

System coverage was an important limiting factor during the outbreaks, but one that would not be obvious unless the researcher was familiar with the systems. For instance, during the period studied there was only very limited data available from our remote health system due to the changeover from NHS Direct to NHS 111, therefore we excluded this system. Also, the emergency department surveillance system did not have coverage in all the areas studied so we only included it for the relevant outbreaks.

Contemporaneously, we would normally only examine national surveillance graphs daily, looking at local ones only when there was a statistical alarm or other intelligence to suggest that specific local ones should also be reviewed. However, here we retrospectively examined all the local data around the outbreak periods. In this way, we could identify if there were any clear increases seen during outbreaks that did not result in alarms, suggesting the aberration detection methods required improvement.

We analysed the aberration detection methods to see retrospectively what alarms they generated locally during the outbreaks. Also, to provide context, we considered the periods before and after the outbreaks and neighbouring locations to compare alarm rates.

Finally, we reviewed the weekly bulletins we'd produced during the outbreak periods and our logs of any alerts issued. As a result, we were able to not just test whether the systems had detected any changes but whether this had resulted in any public health action.

Overall, we found that syndromic surveillance was unable to provide early warning of the local gastrointestinal outbreaks studied. Significantly, ReSST had issued no alerts at the time relating to these events. In some of the outbreaks we did find related alarms or visible increases in activity but ReSST had not considered these sufficiently unusual to require any further action at the time.

The inability of our syndromic system to detect these events could be related to low numbers of cases, partial coverage of health systems in the areas studied, or because the majority of people affected are self-treating and not seeking health care; in which case syndromic surveillance may not be best suited to this purpose. Therefore, we are planning a larger study of gastrointestinal outbreaks covering a longer period. These studies will aid our understanding, so that investigators will have more confidence in the ability of syndromic surveillance to detect specific types of outbreaks and therefore to take the appropriate action when statistical alarms occur.

4.3 Emergency department syndromic surveillance providing early warning of seasonal respiratory activity in England (paper 7)

PHE uses syndromic systems as part of their surveillance of respiratory illnesses caused by seasonal respiratory pathogens. Early detection and identification of increases and peaks in pathogen incidence is important in planning public health action. Although syndromic surveillance cannot be used to explicitly identify causative pathogens, we can make inferences by understanding the contributions of different pathogens to different syndromes and variation in age distribution. Therefore, we studied the associations between pathogen incidence and syndromic respiratory indicators, specifically within the new emergency departments system.

We examined the correlation between syndromic indicators and pathogen incidence using confirmed laboratory reports. We used weekly data based on emergency department attendance or laboratory sample date, thereby avoiding any confounding due to day of the week effects. The period studied was the first three winter seasons that EDSSS was operational, although changes in site coverage meant that models were restricted to two years with the same sites included in both.

We chose to model the correlations using multiple linear regression models. This approach had already been used for other syndromic systems.[82] Moreover, it enabled us to model which pathogens were associated with each syndromic indicator, eliminating those with no association through backwards stepwise regression. Subsequently, we were able to model lags to determine if emergency departments gave early warning of increases in pathogen incidence. We allowed for autocorrelation within the time series by differencing, using the week on week changes in attendances and laboratory reports. Finally, we stratified by age band so we could separately model differences between infants, children and adults.

Our results showed that 25% of attendances for respiratory disease were associated with seasonal rises in respiratory pathogens. Furthermore, this rose to 40% for specific respiratory complaints like bronchiolitis. We found that RSV and influenza B were associated with the greatest burden of respiratory attendances, reflecting their greater seasonal variation compared with other pathogens. Conversely, we did not find a clear association with influenza A, although this may have been due to the relatively low incidence of this pathogen during the seasons studied.

We found that correlations were strongest when the laboratory reports lagged a week behind the syndromic data. Also, we found that indicators in the oldest age groups lagged behind the youngest. Therefore, we showed that emergency department syndromic data is able to provide early warning of RSV and influenza B compared to traditional laboratory surveillance. Furthermore, monitoring attendances amongst young children could provide several weeks warning of rises in attendances amongst the elderly. This is important because hospitals are more likely to admit the elderly who often require additional care.

4.4 Using an Emergency Department Syndromic Surveillance System to investigate the impact of extreme cold weather events (paper 8)

ReSST used syndromic surveillance to monitor the public health impact of extreme weather events. This is important for health care planning and mitigating adverse impacts. In particular, PHE uses syndromic data as part of the Cold Weather Plan for England.

With the introduction of a new emergency departments' syndromic surveillance system in England, we assessed its potential for cold weather surveillance after the first two years of operation. In addition to testing for correlations between extreme cold weather and emergency attendances we needed to ascertain what types of attendances cold weather affects. In this way, we would be able to develop sensitive cold weather indicators for monitoring future events.

We used the pilot data we had from two sites within one city which gave us the longest period of consistent emergency department data. We compared attendances with local meteorological data, including daily minimum temperature and snowfall. Our initial examination of the data included using the Dickey-Fuller test to confirm that we did not need transformations to remove autocorrelation. (The Dickey-Fuller test checks the null hypothesis that a time series is autoregressive with a unit root [83]).

We examined the diagnostic codes used for emergency attendances, identifying codes that directly related to the impact of cold, e.g. hypothermia or frostbite. Additionally, we examined fractures, including forearm/wrist/hip/femur that could be associated with slippery conditions experienced during cold weather. We also, considered respiratory and cardiac attendances that are reported to be affected by extreme reductions in

temperature. Consequently, we were able to develop and test indicators to find those that were the most sensitive to periods of cold weather.

Because we were only interested in periods of extreme weather, we did not attempt to model a general relationship between attendances and temperature or volume of snow. Instead we used the weather data to characterise each day as extreme or not and compared the periods of extreme weather with other days. We choose to use the Wilcoxon-Mann-Whitney non-parametric test, which meant we did not have to assume our data conformed to any particular statistical distribution. We incorporated lags of up to two days in our analysis, in case there was a delay in the impact of the extreme weather. Also, we stratified our analysis by age band and gender.

We found that the most sensitive syndromic indicators for extreme cold weather included attendances for specific fractures, but not for cardiac or respiratory attendances. Interestingly, we found that the strongest association with temperature was for fractures in females but for snowfall it was with fractures in both sexes. Also, we found associations were strongest with a lag of at least one day, or two days in the case of heavy snow (over 5cm).

Our analysis enabled us to develop new syndromic indicators for extreme weather events that were as sensitive as possible. Subsequently, ReSST incorporated these indicators into routine reporting on the impact of weather events.

4.5 The use of syndromic surveillance to monitor the incidence of arthropod bites requiring healthcare in England, 2000-2013: a retrospective ecological study (paper 9)

Arthropods bites and stings (primarily from insects, spiders and ticks) result in a range of medical conditions, from minor irritation to allergic shock and severe disease. Moreover, climate change may result in increased incidence and risk of arthropods being a vector of disease transmission in England. Therefore, it is important to develop monitoring systems as part of public health surveillance. Consequently, we decided to develop new syndromic indicators for arthropod bites and stings.

Within each of England's syndromic surveillance systems it was possible to identify diagnoses relating to bites and stings and thereby create syndromic indicators. However, there were no existing surveillance systems or obvious sources of independent data against which to validate the indicators, other than the potential confounder of temperature. Therefore, we validated the indicators through descriptive epidemiology, supplemented by statistical tests to identify significant features within the data. Specifically, we stratified the data by sex, age and region and calculated incidence rate ratios with confidence intervals. In addition, we analysed the relationship with temperature using negative binomial regression.

We extrapolated from our results to provide estimates of burden on different health care systems across England, including GP services, emergency departments and NHS Direct calls. We found differences between systems; with significantly more female patients with bites or stings visiting GPs but more males calling NHS Direct. Similarly, we found the highest identified burden was for adults aged 45-64 years for emergency departments and GP out-of-hours services, but amongst children aged 5-14 years for GP in-hours and NHS Direct services. Regionally, incidence rates tended to be lower in the north but this was not always the case. Finally, we found that for each degree increase in temperature (Celsius) rates increased by between 3 and 14% depending on system.

We were able to develop syndromic indicators for arthropod bites and stings that have become part of routine surveillance during summer months. Moreover, statistical methods enhanced our understanding of these new indicators, quantifying differences between systems that would have not been possible by descriptive epidemiology alone.

4.6 Assessing the Likely Impact of a Rotavirus Vaccination Program in England: The Contribution of Syndromic Surveillance (paper 10)

Rotavirus is a major source of gastrointestinal illness, particularly in young children. For this reason, PHE introduced a new rotavirus vaccine in July 2013 for children aged 2-3 months. Subsequently, we assessed the impact of the vaccine on our syndromic indicators for diarrhoea, vomiting and gastroenteritis.

Prior to the vaccine introduction there was a clear rotavirus season, visible in laboratory surveillance data. We therefore defined weeks 1 to 22 as the rotavirus season and

compared syndromic data for seasons pre and post the introduction of the vaccine. We calculated the mean weekly rates for each gastrointestinal indicator across our syndromic systems, stratifying the data by age band. We used the Taylor method to calculate confidence intervals for our incidence rate ratios and used unpaired student t-tests to compare between periods.

We found that there was a significant reduction in the amount of gastroenteritis, diarrhoea and vomiting after the introduction of the vaccine. Moreover, the biggest change was amongst children aged under 5 where rates were 23-33% lower, depending on the syndromic system, during rotavirus seasons after PHE introduced the vaccine. Furthermore, we found significant but smaller reductions in older age groups.

Although, PHE will usually measure vaccine effectiveness more directly through laboratory surveillance, syndromic surveillance provides an alternative with the advantage of data being available much sooner for analysis. Also, syndromic data can provide a wider context. The information about impact on other ages could assist future decisions about which age groups should be targeted for vaccination. Furthermore, we will use information about the impact on syndromic surveillance systems of vaccination to modify our aberration detection methods.

4.7 The role of the embedded statistician

The examples above illustrate the wide range of applications for syndromic surveillance and how they were tested using statistical methods. In order to choose which method was appropriate for each situation, I needed not just statistical expertise but also a deep understanding of the local context.

When testing the application to gastrointestinal disease outbreaks (chapter 4.2), it was necessary to understand not just the statistical aberration detection methods used by ReSST but also the local coverage of individual systems during the outbreak periods and the investigation processes used by ReSST prior to public health action. In this first application we compared syndromic data with recorded gastrointestinal outbreaks, whereas for seasonal respiratory activity (chapter 4.3) we used laboratory reports for respiratory pathogens, and for cold weather events (chapter 4.4) records of extreme daily temperatures. By contrast, the other applications to arthropod bites (chapter 4.5) and the rotavirus vaccine (chapter 4.6), did not focus on comparisons with external data. Therefore,

an understanding of the appropriateness of different statistical methods to different types of comparison was important

All these applications benefit from the establishment of my role as an 'embedded statistician'. This role combines a technical knowledge of statistical techniques and their application with a deep understanding of the context in which we apply them. Without statistical experience, researchers would have been limited to purely descriptive epidemiology or have risked misapplying techniques. Alternatively, using external statistical expertise would require considerable extra time for the statistician to understand the complexities of the syndromic data.

5. Conclusion

By using statistical methods we have been able to introduce practical improvements that have enhanced syndromic surveillance in England. Firstly, I have created and refined new aberration detection methods used routinely by ReSST. Secondly, I have validated syndromic systems by using statistical methods to quantify detection capabilities. Finally, I have used statistical methods to test the application of syndromic systems for monitoring specific events and public health interventions.

The creation of aberration detection methods has enhanced surveillance by enabling us to monitor more syndromes daily and at greater geographical granularity. Whereas, validation of systems has demonstrated new syndromic systems' utility and quantified for our stakeholders what scale of events were detectable. Also, I have used statistical methods to add to the evidence for the impact of various public health events or interventions on syndromic surveillance systems.

Statisticians often measure the effectiveness of detection systems in terms of sensitivity and specificity. However, applying these terms to syndromic surveillance reveals complexities that mean simple measures may be unhelpful. For instance, we create syndromic surveillance systems to detect a wide range of threats, and one system's ability to measure the impact of air pollution may be very different from its ability to detect a measles epidemic, therefore separate sensitivity measures are required for each. Moreover it is often very difficult to independently verify an outbreak or the extent of the impact of an event. What appears as an unverified false alarm in syndromic data could be a real event that has not been detected by any other method. Interestingly, whilst researchers

often use laboratory surveillance as a 'gold standard' to verify syndromic surveillance, syndromic data is used as a standard against which to measure new surveillance systems based on social media or web searches. Increasingly, there is a spectrum of data sources based on how confident we are that the data accurately measures the number of people actually affected by an event. Within this spectrum, syndromic data is less reliable than confirmed laboratory reports but more reliable than tweets about how people feel.

Where researchers have applied measures of sensitivity and specificity to syndromic surveillance it usually just covers the statistical aberration detection methods. However, the detection capabilities of syndromic systems also depend on other parts of the process. Firstly, whether or not we detect an event will depend on the presenting behaviour of those affected. If people do not choose to consult health care we will not know they are ill. Secondly, the coverage of our systems will affect our ability to detect events. Finally, if a statistical alarm does not result in ReSST issuing an alert than no public health action will result. Therefore, we should consider the sensitivity and specificity for all parts of the process and the systems as a whole not just the aberration detection methods.

I have used statistical methods to improve our understanding of the syndromic surveillance systems, which then led to practical improvements. Indeed, the Olympic legacy of continued enhanced surveillance in England after 2012 demonstrated how stakeholders had valued the improvements we introduced before the Games. Further examples include how we developed new syndromic indicators following research into the impact of cold weather (chapter 4.4) and insect bites (chapter 4.5). Also, since we identified the impact of the new rotavirus vaccine (chapter 4.6) I have included it as a new confounding variable in the RAMMIE models used for daily surveillance.

Our improved understanding of syndromic surveillance systems and their limitations directs future research. For instance, we've planned more work to determine whether syndromic systems can detect small localised gastrointestinal outbreaks. Also, future work will consider whether we can enhance surveillance by focussing on changes in age distributions within the syndromic data.

How statistical methods are applied to syndromic surveillance systems depends on the role of statisticians within a public health organisation. Sometimes statisticians are only involved in research or development and not operations. Consequently, statisticians may develop detection methods as 'black boxes' where the users have to interpret alarms

without any input into the methodology creating them. Statisticians may work together in teams, acting as consultants to individual projects. This has the advantage of pooling expertise, and the team leader will allocate people to projects based on the statistical skills needed. Alternatively a statistician may be embedded in a multi-disciplinary team so that they have a deep understanding of the context in which they are working.

My role as an embedded statistician within a multi-disciplinary surveillance team has been very important to my effectiveness. As a member of the surveillance team I was immediately aware of any issues with the aberration detection methods. For example, I was able to promptly implement changes when respiratory thresholds were too low as described in chapter 2.2. Furthermore, I witnessed the problems of interpretation caused by using different detection methods and therefore I created a new unified approach for all systems (chapter 2.4). Importantly, close team-working has enabled us to develop prioritisation rules and risk assessment processes that are based on clinical expertise and experience rather than just using numerical multiple-testing algorithms [79]. Finally, my involvement in the use and development of new syndromic systems (e.g. chapter 3.4) has been invaluable when advising on statistical analyses to use for different applications; for example, in understanding system coverage, as discussed in chapters 4.2 and 4.3.

I have also been able to bring the statistician's viewpoint to wider discussions about the development of syndromic surveillance within PHE. For instance, to make recommendations about what new data sources would be most useful or how proposed change within the health service would impact on our detection capabilities. Sometimes these may be counterintuitive, for instance receiving extra more detailed information may be unhelpful if we can no longer compare it with historical data. Also, receiving extra information at a later date may improve data accuracy but make surveillance less timely. Although ReSST's policy is to obtain data opportunistically it has been able to provide input into discussions around national minimum standards for data sets, where this could improve our surveillance [84].

Syndromic surveillance is now a key component of public health surveillance in England. The enhancements we have introduced will therefore help protect the health of the nation. My work involving statistical methods has also had wider application. Thus, public health authorities in other countries regularly invite ReSST to present our findings. For example, ReSST shared their experience of the 2012 London Olympic and Paralympic Games (chapters 2.2, 3.2 & 3.3) with Brazil prior to the 2016 Rio Olympics and Paralympics and

with Canada prior to the 2015 Pan American Games. Also, current research is applying the RAMMIE method to both EDSSS and the French syndromic surveillance system, OSCOUR®[61], in joint collaborations on air pollution and mass gatherings.

We have improved the likelihood of early detection of a bioterrorism incident or emerging new infectious disease. Also, we have demonstrated the ability and the limitations of syndromic systems to provide situational awareness and reassurance to emergency services during events like extreme weather or air pollution. Statistical methods have been accused of reducing lives to mere numbers, however with syndromic surveillance these methods can use 'mere numbers' to save lives.

6. Bibliography

1. Abat C, Chaudet H, Rolain JM, Colson P, Raoult D. Traditional and syndromic surveillance of infectious diseases and pathogens. *Int J Infect Dis* 2016;**48:**22-8.
2. Morens DM, Folkers GK, Fauci AS. The challenge of emerging and re-emerging infectious diseases. *Nature* 2004;**430:**242-9.
3. Ellison J. Public Health England Strategic Remit & Priorities. In: Health Do, editor.2015.
4. Triple S. Assessment of syndromic surveillance in Europe. *Lancet* 2011;**378:**1833-4.
5. Buehler JW, Hopkins RS, Overhage JM, Sosin DM, Tong V, Group CDCW. Framework for evaluating public health surveillance systems for early detection of outbreaks: recommendations from the CDC Working Group. *MMWR Recommendations and reports : Morbidity and mortality weekly report Recommendations and reports / Centers for Disease Control* 2004;**53:**1-11.
6. Velsko S, Bates T. A Conceptual Architecture for National Biosurveillance: Moving Beyond Situational Awareness to Enable Digital Detection of Emerging Threats. *Health security* 2016;**14:**189-201.
7. Bansal S, Chowell G, Simonsen L, Vespignani A, Viboud C. Big Data for Infectious Disease Surveillance and Modeling. *J Infect Dis* 2016;**214 (suppl_4):**S375-S9.
8. Cooper DL, Verlander NQ, Elliot AJ, Joseph CA, Smith GE. Can syndromic thresholds provide early warning of national influenza outbreaks? *J Public Health* 2009;**31:**17-25.
9. Harcourt SE, Smith GE, Elliot AJ, Pebody R, Charlett A, Ibbotson S, et al. 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* 2011;**140:**100-5.
10. Elliot AJ, Hughes HE, Hughes TC, Locker TE, Shannon T, Heyworth J, et al. Establishing an emergency department syndromic surveillance system to support the London 2012 Olympic and Paralympic Games. *Emergency Medicine Journal* 2012;**29:**954-60.
11. Harcourt SE, Fletcher J, Loveridge P, Bains A, Morbey R, Yeates A, et al. Developing a new syndromic surveillance system for the London 2012 Olympic and Paralympic Games. *Epidemiol Infect* 2012;**140:**2152-6.
12. Harcourt SE, Morbey RA, Loveridge P, Carrilho L, Baynham D, Povey E, et al. Developing and validating a new national remote health advice syndromic surveillance system in England. *J Public Health* 2016;**39:**184-92.
13. Buckingham-Jeffery E, Morbey R, House T, Elliot AJ, Harcourt S, Smith GE. Correcting for day of the week and public holiday effects: improving a national daily syndromic surveillance service for detecting public health threats. *BMC Public Health* 2017;**17:**477.
14. Morbey RA, Harcourt S, Pebody R, Zambon M, Hutchison J, Rutter J, et al. The burden of seasonal respiratory infections on a national telehealth service in England. *Epidemiol Infect* 2017**:**1-11.
15. Yan P, Chen H, Zeng D. Syndromic surveillance systems. *Annual Review of Information Science and Technology* 2008;**42:**425-95.
16. Robertson C, Nelson TA, MacNab YC, Lawson AB. Review of methods for space-time disease surveillance. *Spatial and spatio-temporal epidemiology* 2010;**1:**105-16.
17. Unkel S, Farrington CP, Garthwaite H, Robertson C, Andrews N. Statistical methods for the prospective detection of infectious disease outbreaks: a review. *J R Stat Soc Ser A Stat Soc* 2012;**175:**49-82.

18. Rodriguez-Prieto V, Vicente-Rubiano M, Sanchez-Matamoros A, Rubio-Guerri C, Melero M, Martinez-Lopez B, et al. Systematic review of surveillance systems and methods for early detection of exotic, new and re-emerging diseases in animal populations. *Epidemiol Infect* 2014**:**1-25.
19. Jafarpour N, Izadi M, Precup D, Buckeridge DL. Quantifying the determinants of outbreak detection performance through simulation and machine learning. *J Biomed Inform* 2015;**53:**180-7.
20. Spreco A, Timpka T. Algorithms for detecting and predicting influenza outbreaks: metanarrative review of prospective evaluations. *BMJ Open* 2016;**6:**e010683.
21. Rogerson PA. Surveillance systems for monitoring the development of spatial patterns. *Stat Med* 1997;**16:**2081-93.
22. Hutwagner LC, Thompson WW, Seeman GM, Treadwell T. A simulation model for assessing aberration detection methods used in public health surveillance for systems with limited baselines. *Stat Med* 2005;**24:**543-50.
23. Hutwagner L, Browne T, Seeman GM, Fleischauer AT. Comparing aberration detection methods with simulated data. *Emerg Infect Dis* 2005;**11:**314-6.
24. Tokars JI, Burkom H, Xing J, English R, Bloom S, Cox K, et al. Enhancing time-series detection algorithms for automated biosurveillance. *Emerg Infect Dis* 2009;**15:**533-9.
25. Alencar AP, Lee Ho L, Albarracin OY. CUSUM control charts to monitor series of Negative Binomial count data. *Stat Methods Med Res* 2015.
26. Farrington CPA, N.J. Beale, A.D. Catchpole, M.A. A statistical algorithm for the early detection of outbreaks of infectious disease. *J R Stat Soc Ser A Stat Soc* 1996;**159:**547-63.
27. Freeman R, Charlett A, Hopkins S, O'Connell AM, Andrews N, Freed J, et al. Evaluation of a national microbiological surveillance system to inform automated outbreak detection. *The Journal of infection* 2013;**67:**378-84.
28. Noufaily A, Enki DG, Farrington P, Garthwaite P, Andrews N, Charlett A. An improved algorithm for outbreak detection in multiple surveillance systems. *Stat Med* 2013;**32:**1206-22.
29. Noufaily A, Ghebremicheal-Weldeselassie Y, Enki DG, Garthwaite P. Modelling reporting delays for outbreak detection in infectious disease data. *J R Stat Soc Ser A Stat Soc* 2013.
30. Pivette M, Mueller JE, Crepey P, Bar-Hen A. Surveillance of gastrointestinal disease in France using drug sales data. *Epidemics* 2014;**8:**1-8.
31. Pervaiz F, Pervaiz M, Abdur Rehman N, Saif U. FluBreaks: early epidemic detection from Google flu trends. *J Med Internet Res* 2012;**14:**e125.
32. Zhang J, Tsui F, Wagner MM, Hogan WR. Detection of outbreaks from time series data using wavelet transform. *AMIA Annual Symposium proceedings / AMIA Symposium AMIA Symposium* 2003**:**748-52.
33. Shmueli G. Wavelet-Based Monitoring for Biosurveillance. *Axioms* 2013**:**345-70.
34. Martinez-Beneito MA, Conesa D, Lopez-Quilez A, Lopez-Maside A. Bayesian Markov switching models for the early detection of influenza epidemics. *Stat Med* 2008;**27:**4455-68.
35. Cowling BJ, Wong IO, Ho LM, Riley S, Leung GM. Methods for monitoring influenza surveillance data. *Int J Epidemiol* 2006;**35:**1314-21.
36. Costa MA, Kulldorff M. Maximum linkage space-time permutation scan statistics for disease outbreak detection. *Int J Health Geogr* 2014;**13:**20.

37. Wagner M, Tsui F, Cooper G, Espino JU, Harkema H, Levander J, et al. Probabilistic, Decision-theoretic Disease Surveillance and Control. *Online journal of public health informatics* 2011;**3**.
38. Buckeridge DL, Switzer P, Owens D, Siegrist D, Pavlin J, Musen M. An evaluation model for syndromic surveillance: assessing the performance of a temporal algorithm. *MMWR Morbidity and mortality weekly report* 2005;**54 Suppl:**109-15.
39. Buckeridge DL, Owens DK, Switzer P, Frank J, Musen MA. Evaluating detection of an inhalational anthrax outbreak. *Emerg Infect Dis* 2006;**12:**1942-9.
40. Peter W, Najmi AH, Burkom HS. Reducing false alarms in syndromic surveillance. *Stat Med* 2011;**30:**1665-77.
41. Jafarpour N, Precup D, Izadi M, Buckeridge D. Using hierarchical mixture of experts model for fusion of outbreak detection methods. *AMIA Annu Symp Proc* 2013;**2013:**663-9.
42. Zhou HBHWCDAAU. Practical comparison of aberration detection algorithms for biosurveillance systems. *J Biomed Inform* 2015.
43. Rogerson PA, Yamada I. Approaches to syndromic surveillance when data consist of small regional counts. *MMWR supplements* 2004;**53:**79-85.
44. Reis BY, Kohane IS, Mandl KD. An epidemiological network model for disease outbreak detection. *PLoS Med* 2007;**4:**e210.
45. Chan TC, King CC, Yen MY, Chiang PH, Huang CS, Hsiao CK. Probabilistic daily ILI syndromic surveillance with a spatio-temporal Bayesian hierarchical model. *PloS one* 2010;**5:**e11626.
46. Boyle JR, Sparks RS, Keijzers GB, Crilly JL, Lind JF, Ryan LM. Prediction and surveillance of influenza epidemics. *The Medical journal of Australia* 2011;**194:**S28-S33.
47. Kaimi I, Diggle PJ. A hierarchical model for real-time monitoring of variation in risk of non-specific gastrointestinal infections. *Epidemiol Infect* 2011;**139:**1854-62.
48. Xing J, Burkom H, Tokars J. Method selection and adaptation for distributed monitoring of infectious diseases for syndromic surveillance. *J Biomed Inform* 2011;**44:**1093-101.
49. Vega T, Lozano JE, Meerhoff T, Snacken R, Mott J, Ortiz de Lejarazu R, et al. Influenza surveillance in Europe: establishing epidemic thresholds by the moving epidemic method. *Influenza Other Respir Viruses* 2013;**7:**546-58.
50. Andersson T, Bjelkmar P, Hulth A, Lindh J, Stenmark S, Widerstrom M. Syndromic surveillance for local outbreak detection and awareness: evaluating outbreak signals of acute gastroenteritis in telephone triage, web-based queries and over-the-counter pharmacy sales. *Epidemiol Infect* 2014;**142:**303-13.
51. Faryadres M, Karami M, Moghimbeigi A, Esmailnasab N, Pazhouhi K. Levels of Alarm Thresholds of Meningitis Outbreaks in Hamadan Province, west of Iran. *Journal of research in health sciences* 2015;**15:**62-5.
52. Reich NG, Cummings DA, Lauer SA, Zorn M, Robinson C, Nyquist AC, et al. Triggering interventions for influenza: the ALERT algorithm. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2015;**60:**499-504.
53. Buczak AL, Baugher B, Guven E, Moniz L, Babin SM, Chretien JP. Prediction of Peaks of Seasonal Influenza in Military Health-Care Data. *Biomedical engineering and computational biology* 2016;**7:**15-26.
54. Singh BK, Savill NJ, Ferguson NM, Robertson C, Woolhouse ME. Rapid detection of pandemic influenza in the presence of seasonal influenza. *BMC Public Health* 2010;**10:**726.

55. Rao Y, McCabe B. Real-time surveillance for abnormal events: the case of influenza outbreaks. *Stat Med* 2016.
56. Yang W, Li Z, Lan Y, Wang J, Ma J, Jin L, et al. A nationwide web-based automated system for outbreak early detection and rapid response in China. *Western Pac Surveill Response J* 2011;**2:**10-5.
57. Salmon M, Schumacher D, Hohle M. Monitoring Count Time Series in R: Aberration Detection in Public Health Surveillance. *Journal of Statistical Software* 2014.
58. Enki DG, Garthwaite PH, Farrington CP, Noufaily A, Andrews NJ, Charlett A. Comparison of Statistical Algorithms for the Detection of Infectious Disease Outbreaks in Large Multiple Surveillance Systems. *PloS one* 2016;**11:**e0160759.
59. Sosin DM. Draft framework for evaluating syndromic surveillance systems. *J Urban Health* 2003;**80:**i8-13.
60. Doroshenko A, Cooper D, Smith G, Gerard E, Chinemana F, Verlander N, et al. Evaluation of syndromic surveillance based on National Health Service Direct derived data--England and Wales. *Morbidity and mortality weekly report* 2005;**54 Suppl:**117-22.
61. Josseran L, Fouillet A, Caillere N, Brun-Ney D, Ilef D, Brucker G, et al. Assessment of a syndromic surveillance system based on morbidity data: results from the Oscour network during a heat wave. *PloS one* 2010;**5:**e11984.
62. Paterson BJ, Kool JL, Durrheim DN, Pavlin B. Sustaining surveillance: evaluating syndromic surveillance in the Pacific. *Global public health* 2012;**7:**682-94.
63. Rosewell A, Ropa B, Randall H, Dagina R, Hurim S, Bieb S, et al. Mobile phone-based syndromic surveillance system, Papua New Guinea. *Emerg Infect Dis* 2013;**19:**1811-8.
64. Correa A, Hinton W, McGovern A, van Vlymen J, Yonova I, Jones S, et al. Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC) sentinel network: a cohort profile. *BMJ Open* 2016;**6:**e011092.
65. Samoff E, Waller A, Fleischauer A, Ising A, Davis MK, Park M, et al. Integration of syndromic surveillance data into public health practice at state and local levels in North Carolina. *Public Health Rep* 2012;**127:**310-7.
66. Rha B, Burrer S, Park S, Trivedi T, Parashar UD, Lopman BA. Emergency department visit data for rapid detection and monitoring of norovirus activity, United States. *Emerg Infect Dis* 2013;**19:**1214-21.
67. Dailey L, Watkins RE, Plant AJ. Timeliness of data sources used for influenza surveillance. *J Am Med Inform Assoc* 2007;**14:**626-31.
68. van den Wijngaard C, van Asten L, van Pelt W, Nagelkerke NJ, Verheij R, de Neeling AJ, et al. Validation of syndromic surveillance for respiratory pathogen activity. *Emerg Infect Dis* 2008;**14:**917-25.
69. Schindeler SK, Muscatello DJ, Ferson MJ, Rogers KD, Grant P, Churches T. Evaluation of alternative respiratory syndromes for specific syndromic surveillance of influenza and respiratory syncytial virus: a time series analysis. *BMC Infect Dis* 2009;**9:**190.
70. Hall G, Krahn T, Van Dijk A, Evans G, Moore K, Maier A, et al. Emergency department surveillance as a proxy for the prediction of circulating respiratory viral disease in Eastern Ontario. *Can J Infect Dis Med Microbiol* 2013;**24:**150-4.
71. Signorini A, Segre AM, Polgreen PM. The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic. *PloS one* 2011;**6:**e19467.

72. Daughton AR, Velappan N, Abeyta E, Priedhorsky R, Deshpande A. Novel Use of Flu Surveillance Data: Evaluating Potential of Sentinel Populations for Early Detection of Influenza Outbreaks. *PloS one* 2016;**11:**e0158330.
73. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. *PloS one* 2010;**5:**e14118.
74. Zhang Y, Arab A, Cowling BJ, Stoto MA. Characterizing Influenza surveillance systems performance: application of a Bayesian hierarchical statistical model to Hong Kong surveillance data. *BMC Public Health* 2014;**14:**850.
75. Pivette M, Mueller JE, Crepey P, Bar-Hen A. Drug sales data analysis for outbreak detection of infectious diseases: a systematic literature review. *BMC Infect Dis* 2014;**14:**604.
76. Ziemann A, Fouillet A, Brand H, Krafft T. Success Factors of European Syndromic Surveillance Systems: A Worked Example of Applying Qualitative Comparative Analysis. *PloS one* 2016;**11:**e0155535.
77. Baker M, Smith GE, Cooper D, Verlander NQ, Chinemana F, Cotterill S, et al. Early warning and NHS Direct: a role in community surveillance? *J Public Health Med* 2003;**25:**362-8.
78. Benjamini Y, Hockberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol* 1995;**57 (1):**289-300.
79. Smith GE, Elliot AJ, Ibbotson S, Morbey R, Edeghere O, Hawker J, et al. Novel public health risk assessment process developed to support syndromic surveillance for the 2012 Olympic and Paralympic Games. *J Public Health* 2016.
80. Elliot AJ, Bone A, Morbey R, Hughes HE, Harcourt S, Smith S, et al. Using real-time syndromic surveillance to assess the health impact of the 2013 heatwave in England. *Enviromental Research* 2014;**135:**31-6.
81. Smith GE, Bawa Z, Macklin Y, Morbey R, Dobney A, Vardoulakis S, et al. Using real-time syndromic surveillance systems to help explore the acute impact of the air pollution incident of March/April 2014 in England. *Enviromental Research* 2015;**136:**500-4.
82. Cooper DL, Smith GE, Edmunds WJ, Joseph C, Gerard E, George RC. The contribution of respiratory pathogens to the seasonality of NHS Direct calls. *J Infect* 2007;**55:**240-8.
83. Dickey DA, Fuller WA. Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *Journal of the American Statistical Association* 1979;**74:**427-31.
84. Hughes HE, Hughes TC, Haile A, Smith GE, McCloskey B, Elliot AJ. Syndromic Surveillance Revolution? Public Health Benefits of Modernizing the Emergency Care Patient Health Record in England. *Public Health Rep* 2017.