# Evaluating multi-purpose syndromic surveillance systems – a complex problem

**Roger Morbey[1*], Gillian Smith[1], Isabel Oliver[2], Obaghe Edeghere[3], Iain Lake[4], Richard Pebody[5], Dan Todkill[3], Noel McCarthy[6], and Alex J. Elliot[1]**

[1] Real-time Syndromic Surveillance Team, Field Service, National Infection Service, Public Health England, Birmingham B2 4BH, United Kingdom; [2] Field Service, National Infection Service, Public Health England, Bristol BS1 6EH, United Kingdom; [3] Field Epidemiology West Midlands, Field Service, National Infection Service, Public Health England, Birmingham B2 4BH, United Kingdom; [4] School of Environmental Sciences, University of East Anglia, Norwich, NR4 7TJ, United Kingdom; [5] Influenza and Other Respiratory Virus Section, Immunisation and Countermeasures Division, National Infection Service, Public Health England, London NW9 5EQ, United Kingdom; [6] Warwick Medical School, Division of Health Sciences, University of Warwick, CV4 7AL, United Kingdom

## Abstract

Surveillance systems need to be evaluated to understand what the system can or cannot detect. The measures commonly used to quantify detection capabilities are sensitivity, positive predictive value and timeliness. However, the practical application of these measures to multi-purpose syndromic surveillance services is complex. Specifically, it is very difficult to link definitive lists of what the service is intended to detect and what was detected.

First, we discuss issues arising from a multi-purpose system, which is designed to detect a wide range of health threats, and where individual indicators, e.g. 'fever', are also multi-purpose. Secondly, we discuss different methods of defining what can be detected, including historical events and simulations. Finally, we consider the additional complexity of evaluating a service which incorporates human decision-making alongside an automated detection algorithm. Understanding the complexities involved in evaluating multi-purpose systems helps design appropriate methods to describe their detection capabilities.

**Keywords:** public health, epidemiology, surveillance, outbreaks

**Abbreviations:** positive predictive value (PPV)

* Correspondence: [roger.morbey@phe.gov.uk](mailto:roger.morbey@phe.gov.uk)

# Introduction

## Syndromic surveillance

Syndromic surveillance involves monitoring health care data on symptoms, signs and diagnoses to provide information for public health action [1]. Syndromic surveillance is often multi-purpose, using many different syndromes or clinical indicators to monitor different conditions and events of public health interest. Public health organisations may operate a syndromic surveillance 'service' that includes several 'systems', with each 'system' using data from one source, e.g. emergency departments, family doctors or ambulances. An on-going syndromic surveillance service is more than a series of data processing steps, it involves analysis, interpretation, reporting and enabling decision-making for appropriate action. It also requires a cycle of continuous improvement, with development of novel approaches and their subsequent application into the service.

When interpreting information from syndromic surveillance systems, public health practitioners, e.g. epidemiologists or incident directors, need to understand the capabilities of those systems to support decision making and choice of actions. Incident directors and other users want answers to apparently simple questions such as: "How many cases of cryptosporidiosis need to occur before your system detects an outbreak in this area?"; or "How much early warning can you provide of increases in seasonal influenza?"

## Evaluating syndromic surveillance - existing evidence base

The Centre for Disease Control and Prevention (CDC) in the United States of America created a framework for evaluating a syndromic surveillance service [2]. This framework has been widely adopted and used to evaluate both syndromic and traditional non-syndromic surveillance. The framework has been applied to evaluate services both quantitatively and qualitatively [3,4]. Furthermore, a wide range of statistical aberration detection algorithms have been applied to syndromic surveillance, to identify unusual exceedances that might indicate a threat to public health [5-7]. Consequently, much of the published research on quantifying the public health benefit of syndromic surveillance focuses on the use of the statistical algorithms. However, retrospectively identifying that an algorithm can detect outbreaks does not inform whether appropriate public health action was taken by the syndromic surveillance service or the impact on public health [8]. It is also important to evaluate the service's decision-making and operational processes [9]. Surveillance does not end with the generation of a statistical alarm. Following an alarm there will be decisions about the importance of the alarm, possibly further epidemiological investigations and analysis to summarise findings in key messages, and finally there will decisions about appropriate public health action. Therefore, further work is also needed to evaluate these later stages of syndromic surveillance as well as the detection algorithms.

Similarly, published evaluations of syndromic systems often focus on just one disease or syndrome [10], whereas syndromic surveillance services are often multi-purpose [5]. Importantly, syndromic surveillance has the potential to detect future unknown hazards, for instance symptoms resulting from a newly emerging disease, such as COVID-19, for which laboratory tests may not yet be available [11]. Therefore, there is a gap in our understanding of the detection capabilities of multi-purpose syndromic surveillance services because services are usually only evaluated as if they have a single purpose and only in terms of the ability to generate statistical alarms.

2

**Quantifying the detection capabilities of a multi-purpose service - a complex problem**

Ideally, simple clear quantitative measures should be provided to describe a multi-purpose service's detection capabilities. However, published quantitative estimates for detection capabilities have usually been restricted to single diseases or to the automated part of a service. For example, it is much easier to deliver estimates structured as "the algorithm had a sensitivity of 98% and a specificity of 84% for simulated influenza outbreaks" rather than "this syndromic service resulted in appropriate action 85% of the time, with 20% of actions subsequently found to be unnecessary". This research focus may be because quantifying the detection capabilities of a multi-purpose syndromic service is not as straightforward as it might initially appear. In fact, this is not just a complicated problem but a complex one. A complicated problem might be large and require considerable resources but can be answered by a single rule-based process, whereas a complex problem requires a range of context-specific methods to obtain answers. Similar issues of complexity have been found in evaluating public health interventions [12]. Here, we provide a perspective paper on the complexities involved in providing meaningful answers for what can and cannot be detected by a multi-purpose syndromic surveillance service. Thus, we aim to suggest a way forward in tackling this complex problem, which can be adopted by other organisations and countries coordinating a multi-purpose syndromic surveillance service.

**Measures for quantifying detection – laboratory tests analogy**

Syndromic surveillance systems are often used alongside and complement traditional surveillance systems such as those based on laboratory testing. Therefore, we use laboratory tests as an example to describe how detection capabilities can be quantified. Then, by analogy we discuss what is required to quantify the detection capabilities of syndromic systems.

**Quantifying laboratory tests – a 'simple' example**

A laboratory test needs to be able to identify disease rapidly with few 'false alarms' [13]. Therefore, evaluation measures must include: a measure for how likely the test is to detect disease; a measure for how likely it is to create false alarms; and for how quickly it will detect disease. Firstly, sensitivity (also called recall) can be defined as the proportion of patients with disease correctly identified by a positive test. Secondly, false alarms can be quantified using, specificity or positive predictive value (PPV; also, called precision). Specificity can be defined as the proportion of tested patients without a disease with a negative test result, and PPV by the proportion of positive tests that come from patients with the disease. Finally, timeliness can be defined as the time between a sample being taken and the laboratory report being available.

Calculating these quantitative measures for a laboratory test requires a list of patients, with a variable for whether the disease or condition is present, and a linked list of samples, with a variable for whether the laboratory test was positive for the disease or condition (Figure 1).

| $\text{Sensitivity} = \dfrac{Correct\ detections}{Patients\ with\ disease}$ $\text{Specificity} = \dfrac{Correct\ reassurances}{Patients\ with\ no\ disease}$ $\text{PPV} = \dfrac{Correct\ detections}{Positive\ test\ results}$ | | Did the patient have the disease? | |
|---|---|---|---|
| | | **Yes** | **No** |
| Was the laboratory test positive? | Yes | Correct detection (true positive) | False warning (false positive) |
| | No | Fail to detect (false negative) | Correct reassurance (true negative) |

Figure 1. Results matrix for evaluating the sensitivity and specificity of a single laboratory test

**Quantifying syndromic surveillance – a 'complex' example**

By analogy, it should be possible to create the same quantitative measures i.e. the sensitivity, specificity, PPV and timeliness of a syndromic surveillance service (Figure 2). However, instead of comparing a list of patients and test results, we need a list of events we want to detect and a linked list of detections made by the service (throughout this paper, we will use the term 'event' to cover all the different public health threats a service aims to detect, including outbreaks with different aetiologies, public health incidents and the impact of environmental exposures etc., Figure 3).

| $\text{Sensitivity} = \dfrac{Correct\ detections}{Events\ occuring}$ $\text{Specificity} = \dfrac{Correct\ reassurances}{No\ events\ occurring}$ $\text{PPV} = \dfrac{Correct\ detections}{All\ detections\ reported}$ | | Did an event occur? | |
|---|---|---|---|
| | | **Yes** | **No** |
| Did the syndromic service report detection? | Yes | Correct detection (true positive) | False warning (false positive) |
| | No | Fail to detect (false negative) | Correct reassurance (true negative) |

Figure 2. Results matrix for evaluating a multi-purpose syndromic surveillance service.

In theory, given a linked list of events to be detected and a list of detections reported by a syndromic service, we can quantify the detection capabilities of the service. However, in practice, creating definitive linked lists of events and detections is complex.

4

# What do we want to detect with syndromic surveillance?

**Multi-purpose surveillance**

Syndromic surveillance was originated to provide population-level surveillance for early warning for bioterrorism threats but it has subsequently been used for early warning of other events and is increasingly used for reassurance of the lack of adverse health impact in a specific context, or for situational awareness after a known exposure [1,14]. A multi-purpose syndromic surveillance service may have multiple objectives [2, 8, 10]:

- Early warning of unexpected events, e.g. bioterrorism, emerging new diseases, outbreaks;
- Early warning of aberrant trends by monitoring endemic or seasonal diseases, e.g. scarlet fever or seasonal influenza;
- Reassurance and monitoring during mass gatherings e.g. Olympic and Paralympic Games;
- Situational awareness during pre-identified outbreaks or environmental incidents, e.g. COVID-19, an influenza pandemic or heat wave;

Therefore, a multi-purpose syndromic surveillance service will need to detect a wide range of events, reflecting potential threats to public health, including infectious disease, environmental impacts and mass gatherings (Figure 3).

| Purpose | Objective | Event type |
|---|---|---|
| Comprehensive population surveillance | provide early-warning of unexpected threats to public health | epidemic of severe respiratory illness, e.g. SARS, COVID-19 |
| | | cryptosporidium outbreaks |
| | | norovirus outbreaks |
| | | food poisoning outbreaks |
| | | bioterrorism |
| | monitor trends to give early warning of atypical activity | seasonal influenza |
| | | seasonal respiratory syncytial virus |
| | | scarlet fever |
| | | "Back to school" asthma [1] |
| | | measles |
| | | mumps |
| | | rubella |

| | | pertussis |
|---|---|---|
| | | hay fever |
| | | insect bites |
| **Targeted sub-group surveillance** | Monitoring of specific context to provide reassurance or early warning of impact on health | vaccine impact |
| | | volcanic ash cloud |
| | | floods |
| | | large industrial fires |
| **Situational awareness** | Measuring impact of known exposure | out of season pandemic influenza |
| | | extreme cold weather |
| | | heat waves |
| | | "thunderstorm asthma" [2] |
| | | impact of air pollution |
| | | impact of water contamination |

Figure 3. Types of events that a multi-purpose syndromic surveillance service aims to detect.

Compiling a list of events to be detected through multi-purpose surveillance is complex because different types of events are defined in different ways. For example, point-source outbreaks might have a clear start and end date, whilst propagated or seasonal epidemics cannot be clearly defined in this way [8]. Similarly, how suspected events are validated will vary by type. For infectious diseases, laboratory reports provide a 'gold-standard' for incidence, however, independent data may not be available for other types of events, e.g. increase in hay fever reports. For some types of events, e.g. extreme weather or mass-gatherings, it may be easy to validate exposure but less obvious how to independently validate impact on the population's health. Consequently, we may be able to create a list of events which have been detected by other surveillance systems (but not those which haven't), but not be certain about the timing and size of any public health impacts that the syndromic service needs to detect.

**Obtaining historical examples**

It is important that syndromic services are evaluated across the full range of event types and different sizes of event [17, 18]. However, for some types of event there may be no historical data available or only a limited range of outbreak sizes, locations etc. [8]. Therefore, synthetic simulated data are often used to evaluate syndromic systems [19]. There are advantages and disadvantages for using real historical events or using synthetic events, historical events may be rare whilst synthetic events may be unrealistic [20]. The main disadvantage of using synthetic events is that they require modelling assumptions, for example, healthcare seeking behaviours for a range of diseases need to be estimated from other research, which is not straightforward [21]. A commonly used approach is to 'inject' synthetic simulations of events into 'real' historic syndromic data [5]. Furthermore, real scaled events can be injected to reduce modelling assumptions about the relationship between outbreak size and syndromic indicators [17, 22-24]. However, results will still depend upon assumptions about the lag between exposure, symptom onset and whether a person presents to health care.

6

**Completeness of event lists**

To evaluate a syndromic service, the list of events to be detected must be comprehensive and exclusive (Figure 3). Furthermore, to estimate specificity or PPV, an identified period without such events is also needed. However, even for event types where numerous independently verifiable outbreaks are available, it may be impossible to guarantee that all events have been identified. It is perfectly plausible that syndromic data contain unverified events, for example, increases in respiratory illness have been observed in autumn that cannot be explained by comparison with laboratory data [20]. These unverified outbreaks within baseline syndromic data can result in lower specificity and PPV estimates [8, 14]. Figure 4 summarizes the complexities around defining what needs to be detected by syndromic surveillance, as discussed above.

| Reason definition is complex | Example |
|---|---|
| **Little or no historical data may be available** | Bioterrorism, newly emerging diseases |
| **Simulated data is sensitive to modelling assumptions** | Patients' health-seeking behaviour is difficult to predict |
| **Event may not be routinely monitored by non-syndromic systems** | Seasonal hay fever |
| **Exposure may be clearly defined but impact on public health is still uncertain** | Heat waves |
| **Laboratory 'gold-standard' for independent verification may not exist** | Newly emerging pathogen |
| **Precise start and end date of exposure might be uncertain** | Seasonal influenza |
| **Events causing similar symptoms may occur at the same time** | Air pollution and seasonal respiratory illness |
| **Control period without events may be unavailable** | Syndromic baseline data is rarely zero |

Figure 4. Reasons why defining 'events' to be detected by syndromic surveillance are complex.

## Defining detection with syndromic surveillance

Whilst it is relatively straightforward to define the detection parameters for statistical algorithms [25], it becomes more complex when we consider the whole syndromic surveillance service. Firstly, we need to consider how the service reports detection, which may depend on its 'surveillance objective'. Secondly, we need to decide how to link detection to events in the context of multi-purpose syndromic surveillance.

**Objectives for a syndromic surveillance service**

The objective that a syndromic service is fulfilling will affect both the definition of detection and its ability to detect events. For example, when acting as an early warning system a

7

syndromic service may define detection as alerting the appropriate authorities prior to any other surveillance system. Successful early warning depends on a service's routine surveillance practices and reporting arrangements. By contrast, when providing situational awareness during a known event, the multi-purpose service can focus on a geographical area and subset of syndromic indicators, which will increase the probability of detecting an impact. Also, when providing situational awareness, the service may define detection as identifying small changes in trends, which would not have triggered an early warning response to a hitherto unknown event. Similarly, a service that routinely monitors seasonal diseases (e.g. influenza) may have specifically developed thresholds that are more sensitive than those that warn of undefined new threats [26]. Finally, the objective of a syndromic service may change when an event becomes publicly known through media reports, e.g. COVID-19. Moreover, syndromic indicators may be affected by changes in patient health-seeking behaviour because of increased awareness after an event [8, 10, 27], or changes in government advice e.g. during a lock-down. In summary, creating a list of detections requires consideration of whether the event was expected and the service's objective at the time of detection.

**Multi-purpose syndromic indicators**

The ability to link what is detected by syndromic surveillance to specific events is further complicated because many syndromic indicators are multi-purpose. Whilst some syndromic indicators are very specific (e.g. bloody diarrhoea) others (e.g. gastrointestinal) are designed to have a high sensitivity but low specificity to maximise the chance of detecting events or to ensure that new emerging threats, such as COVID-19, are captured [3, 4, 28]. These broad syndromic indicators may detect a range of different types of events. For example, generic respiratory indicators (e.g. cough or difficulty breathing) have been found to be associated with changing trends in laboratory reports for several different respiratory pathogens [20, 29-30] as well as seasonal allergies [31]. Consequently, a syndromic service will often detect an increasing trend but not be able to link it to a specific event or individual organism, without further context. However, the ability to link detection to events may also depend on the objective of the surveillance system. For example, during a known laboratory-confirmed measles outbreak, a syndromic service may use a general indicator, e.g. rash, for situational awareness, which would not be considered as an effective early warning indicator for unknown measles outbreaks [10]. Furthermore, when laboratory data are not available to verify causal pathogens, syndromic indicators or combinations of symptoms may be used to suggest probable causes of outbreaks [3, 32], particularly for multi-system surveillance [20]. Finally, during a pandemic of an emerging disease like COVID-19, new processes or diagnostic codes may be introduced which have an impact on existing syndromic indicators.

## Discussion

Much of the published research evaluating syndromic surveillance focuses either on just one type of event or on the detection capabilities of statistical algorithms. We have reflected on and highlighted the complexities of evaluating and quantifying the detection capability of a multi-purpose syndromic service, which may explain the lack of published evidence on this subject. However, to address questions from users of syndromic surveillance about detection capabilities, we need to avoid over-simplifications and provide descriptions which directly address the complexities and wide-ranging utility of these services.

Therefore, we argue that syndromic surveillance service evaluations need to measure separately different types of event that the service aims to detect and to consider all surveillance

8

stages. Whilst the authors support the use of the CDCs framework for evaluation of surveillance systems [2], we also believe the complexity of multi-purpose systems needs to be considered in such frameworks. Firstly, separate answers are needed for different types of event both to address users' specific questions and because different types of events will require different methods for evaluation. Crucially, these separate evaluations should be done in the context of a multi-purpose service where other types of events can affect detection capabilities and the ability to identify causes is also addressed. Secondly, syndromic services should be evaluated beyond the generation of statistical alarms to provide results that inform public health action. Service evaluations should include consideration of the routine surveillance messages and the impact of public health actions for different event types.

To quantify the detection capabilities of syndromic surveillance it is important to compare events that the system aims to detect with what was detected. However, in this commentary we have shown that for a multi-purpose service, defining and linking these events is complex. The complexities arise from the wide range of events covered by a multi-purpose service and the need to assess not just the performance of statistical algorithms but the whole service process.

**Measure each event type separately**

When considering a multi-purpose syndromic surveillance service, no single measure can helpfully describe its detection capabilities across all the different types of events it aims to detect. Therefore, it is important to consider all the different type of events to be detected and measure detection capabilities separately for each.

Measuring each type of event separately means that a different approach can be used for different event types, for instance how events are defined or the user questions to be addressed. Involving key internal and external stakeholders (including users of the service) in the evaluation is very important to ensure relevance [17]. For example, stakeholders can steer how narrowly the event types are defined and to address issues such as whether it is sufficient to estimate detection for all gastrointestinal outbreaks or do users require separate estimates for specific pathogens e.g. cryptosporidium or rotavirus.

When measuring each event type separately there is still a need to consider how other types might affect detection capabilities. For example, does the ability to detect the health impact of air pollution change during an influenza epidemic? Also, where there are multi-purpose indicators, correct detection of one type of event could be considered as a false alarm for detecting another type of event. Importantly, evaluating a multi-purpose service by measuring different event types separately is not the same as performing a series of parallel evaluations in each of which the service is treated as if it had only one purpose.

Clearly, it requires much more work to tackle each event type separately, particularly if a range of different approaches are needed. However, this will provide a much richer understanding of the service's capabilities and enhance users' interpretation and confidence in the service outputs.

**Evaluating each stage in the surveillance process**

The automated statistical detection algorithm is just one stage in a syndromic service's many processes [33]. The stages can be characterized as: data collection, storage and extraction; aggregation to syndromic indicators; application of detection algorithms; and interpretation,

9

reporting and taking action. It is important to evaluate the service as a whole, so detection involves not just automated alarms but their interpretation, prioritization, reporting and public health impact [34]. However, evaluating each stage in the process separately can provide useful insights into which factors affect the service's ability to detect events [35].

Firstly, evaluating data collection will reveal what proportion of the target population is covered by the service and whether there are any delays in receiving information. For example, a sentinel service will be unable to detect local outbreaks in locations not covered by the system [36]. Secondly, the underlying codes, diagnoses or free text included in syndromic indicators will determine their sensitivity and specificity [28], for example, a multi-purpose indicator may be able to detect different diseases with varying success due to different disease characteristics [7]. Evaluating detection algorithms enables users to choose the most appropriate method for their service, which may vary by event type. Finally, evaluating the interpretation and reporting stage usually involves assessing which automated statistical alarms require further action, therefore this stage should improve PPV and specificity but with a cost for timeliness and possibly sensitivity [6]. Considering each stage separately should enable service users to identify areas where a system can be improved, for example, what are the main causes of delays? or is more data being collected than can be analyzed? Figure 5 summarizes how each stage can impact on sensitivity, PPV and timeliness as discussed above. Each additional stage may introduce delays to timeliness and a drop in sensitivity but should increase the PPV.

| Surveillance stage | Potential problems causing… | | |
|---|---|---|---|
| | **Failure to detect** | **False alarms** | **Delays** |
| data collection, storage and extraction | Sentinel system does not cover location of 'event' | Data quality, duplicates, test data etc. | Delay between exposure and presenting to health care |
| aggregation to syndromic indicators | Symptoms not covered by existing indicators | Similar symptoms caused by other reasons | Data processing |
| application of detection algorithms | Alarm thresholds set too high (no alarm) or too low (more alarms than can be analysed) | Alarm thresholds set too low | Computational complexity also alarm volume impacts on next stage |
| interpretation, reporting and taking action | Failure to take appropriate public health action following alarm | Failure to distinguish between false alarm and potential health threat | Staff time, waiting for 'repeat' alarms to provide confirmation, decision-making processes |

Figure 5. Impact on detection capabilities of different stages in syndromic surveillance.

**Future work**

We have focused on the complexities surrounding evaluation of a multi-purpose syndromic service, therefore we have not considered other important issues such as cost-effectiveness or

the added value of additional data sources. However, understanding evaluation complexities will be useful for future studies into cost-effectiveness etc. Evaluation of a multi-purpose syndromic surveillance service should not be a one-off process, it should be periodic creating a positive feedback loop. Information about a service's detection capabilities should be updated as new evidence comes to light, or in response to major incidents such as the current COVID-19 pandemic. Also, the most valuable information for assessing a service will come from its on-going performance. Therefore, a syndromic service should have clear objectives and maintain a database of past events of different types and detections to enable on-going validation [37]. The process of identifying the different types of event that the users want a multi-purpose syndromic service to detect should help identify gaps in our knowledge about service detection capabilities, and in turn, this should help guide research priorities.

## Acknowledgements

## Financial Disclosure

No Financial Disclosures.

## Competing Interests

No Competing Interests.

## References

1. Triple S. 2011. Project. Assessment of syndromic surveillance in Europe. *Lancet*. 378, 1833-34. PubMed https://doi.org/10.1016/S0140-6736(11)60834-9

2. Sosin DM. 2003. Draft framework for evaluating syndromic surveillance systems. *J Urban Health*. 80, i8-13. PubMed

3. Jefferson H, Dupuy B, Chaudet H, Texier G, Green A, et al. 2008. Evaluation of a syndromic surveillance for the early detection of outbreaks among military personnel in a tropical country. *J Public Health (Oxf)*. 30, 375-83. PubMed https://doi.org/10.1093/pubmed/fdn026

4. Yih WK, Deshpande S, Fuller C, Heisey-Grove D, Hsu J, et al. 2010. Evaluating real-time syndromic surveillance signals from ambulatory care data in four states. *Public Health Rep*. 125, 111-20. PubMed https://doi.org/10.1177/003335491012500115

5.  Buckeridge DL, Burkom H, Campbell M, Hogan WR, Moore AW. 2005. Algorithms for rapid outbreak detection: a research synthesis. *J Biomed Inform*. 38, 99-113. PubMed https://doi.org/10.1016/j.jbi.2004.11.007

6.  Faverjon C, Berezowski J. 2018. Choosing the best algorithm for event detection based on the intend application: a conceptual framework for syndromic surveillance. *J Biomed Inform*. 85, 126-35. PubMed https://doi.org/10.1016/j.jbi.2018.08.001

7.  Yuan M, Boston-Fisher N, Luo Y, Verma A, Buckeridge DL. 2019. A systematic review of aberration detection algorithms used in public health surveillance. *J Biomed Inform*. 94, 103181. PubMed https://doi.org/10.1016/j.jbi.2019.103181

8.  Andersson T, Bjelkmar P, Hulth A, Lindh J, Stenmark S, et al. 2014. 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*. 142, 303-13. PubMed https://doi.org/10.1017/S0950268813001088

9.  Smith GE, Elliot AJ, Lake I, Edeghere O, Morbey R, et al. 2019. Public Health England Real-time Syndromic Surveillance T. Syndromic surveillance: two decades experience of sustainable systems - its people not just data! *Epidemiol Infect*. 147, e101. PubMed https://doi.org/10.1017/S0950268819000074

10. Thomas MJ, Yoon PW, Collins JM, Davidson AJ, Mac Kenzie WR. 2018. Evaluation of Syndromic Surveillance Systems in 6 US State and Local Health Departments. *J Public Health Manag Pract*. 24, 235-40. PubMed https://doi.org/10.1097/PHH.0000000000000679

11. Yoon PW, Ising AI, Gunn JE. 2017. Using Syndromic Surveillance for All-Hazards Public Health Surveillance: Successes, Challenges, and the Future. *Public Health Rep*. 132, 3S-6S. PubMed https://doi.org/10.1177/0033354917708995

12. Connelly JB. 2007. Evaluating complex public health interventions: theory, methods and scope of realist enquiry. *J Eval Clin Pract*. 13, 935-41. PubMed https://doi.org/10.1111/j.1365-2753.2006.00790.x

13. Lalkhen AG, McCluskey A. 2008. Clinical tests: sensitivity and specificity. *BJA Educ*. 8, 221-23.

14. Mathes RW, Lall R, Levin-Rector A, Sell J, Paladini M, et al. 2017. Evaluating and implementing temporal, spatial, and spatio-temporal methods for outbreak detection in a local syndromic surveillance system. *PLoS One*. 12, e0184419. PubMed https://doi.org/10.1371/journal.pone.0184419

15. Bundle N, Verlander NQ, Morbey R, Edeghere O, Balasegaram S, et al. 2019. Monitoring epidemiological trends in back to school asthma among preschool and school-aged children using real-time syndromic surveillance in England, 2012–2016. *J Epidemiol Community Health*. 73, 825-31. PubMed https://doi.org/10.1136/jech-2018-211936

16. Thien F. 2018. Melbourne epidemic thunderstorm asthma event 2016: Lessons learnt from the perfect storm. *Respirology*. 23, 976-77. PubMed https://doi.org/10.1111/resp.13410

17. Wallstrom GL, Wagner M, Hogan W. 2005. High-fidelity injection detectability experiments: a tool for evaluating syndromic surveillance systems. *MMWR Suppl*. 54, 85-91. PubMed

18. Todkill D, Elliot AJ, Morbey R, Harris J, Hawker J, et al. 2016. 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*. 144, 2241-50. PubMed https://doi.org/10.1017/S0950268816000480

19. Madouasse A, Marceau A, Lehebel A, Brouwer-Middelesch H, van Schaik G, et al. 2013. Evaluation of a continuous indicator for syndromic surveillance through simulation. application to vector borne disease emergence detection in cattle using milk yield. *PLoS One*. 8, e73726. PubMed https://doi.org/10.1371/journal.pone.0073726

20. van den Wijngaard C, van Asten L, van Pelt W, Nagelkerke NJ, Verheij R, et al. 2008. Validation of syndromic surveillance for respiratory pathogen activity. *Emerg Infect Dis*. 14, 917-25. PubMed https://doi.org/10.3201/eid1406.071467

21. Fan Y, Wang Y, Jiang H, Yang W, Yu M, et al. 2014. Evaluation of outbreak detection performance using multi-stream syndromic surveillance for influenza-like illness in rural Hubei Province, China: a temporal simulation model based on healthcare-seeking behaviors. *PLoS One*. 9, e112255. PubMed https://doi.org/10.1371/journal.pone.0112255

22. Noufaily A, Enki DG, Farrington P, Garthwaite P, Andrews N, et al. 2013. An improved algorithm for outbreak detection in multiple surveillance systems. *Stat Med*. 32, 1206-22. PubMed https://doi.org/10.1002/sim.5595

23. Morbey RA, Elliot AJ, Charlett A, Ibbotson S, Verlander NQ, et al. 2014. Using public health scenarios to predict the utility of a national syndromic surveillance programme during the 2012 London Olympic and Paralympic Games. *Epidemiol Infect*. 142, 984-93. PubMed https://doi.org/10.1017/S095026881300188X

24. Colon-Gonzalez FJ, Lake IR, Morbey RA, Elliot AJ, Pebody R, et al. 2018. A methodological framework for the evaluation of syndromic surveillance systems: a case study of England. *BMC Public Health*. 18, 544. PubMed https://doi.org/10.1186/s12889-018-5422-9

25. Noufaily A, Morbey RA, Colon-Gonzalez FJ, Elliot AJ, Smith GE, et al. 2019. Comparison of Statistical Algorithms for Daily Syndromic Surveillance Aberration Detection. *Bioinformatics*. 35, 3110-18. PubMed https://doi.org/10.1093/bioinformatics/bty997

26. Vega T, Lozano JE, Meerhoff T, Snacken R, Mott J, et al. 2013. Influenza surveillance in Europe: establishing epidemic thresholds by the moving epidemic method. *Influenza Other Respir Viruses*. 7, 546-58. PubMed https://doi.org/10.1111/j.1750-2659.2012.00422.x

27. Elliot AJ, Hughes HE, Astbury J, Nixon G, Brierley K, et al. 2016. The potential impact of media reporting in syndromic surveillance: an example using a possible

Cryptosporidium exposure in North West England, August to September 2015. *Euro Surveill*., 21. PubMed https://doi.org/10.2807/1560-7917.ES.2016.21.41.30368

28. Betancourt JA, Hakre S, Polyak CS, Pavlin JA. 2007. Evaluation of ICD-9 codes for syndromic surveillance in the electronic surveillance system for the early notification of community-based epidemics. *Mil Med*. 172, 346-52. PubMed https://doi.org/10.7205/MILMED.172.4.346

29. Bourgeois FT, Olson KL, Brownstein JS, McAdam AJ, Mandl KD. Validation of syndromic surveillance for respiratory infections. Ann Emerg Med. 2006;47:265 e1.

30. Morbey RA, Elliot AJ, Harcourt S, Smith S, de Lusignan S, et al. 2018. Estimating the burden on general practitioner services in England from increases in respiratory disease associated with seasonal respiratory pathogen activity. *Epidemiol Infect*. 146, 1389-96. PubMed https://doi.org/10.1017/S0950268818000262

31. Wallstrom GL, Hogan WR. 2007. Unsupervised clustering of over-the-counter healthcare products into product categories. *J Biomed Inform*. 40, 642-48. PubMed https://doi.org/10.1016/j.jbi.2007.03.008

32. Paterson BJ, Kool JL, Durrheim DN, Pavlin B. 2012. Sustaining surveillance: evaluating syndromic surveillance in the Pacific. *Glob Public Health*. 7, 682-94. PubMed https://doi.org/10.1080/17441692.2012.699713

33. Morbey RA, Elliot AJ, Charlett A, Verlander NQ, Andrews N, et al. 2015. The application of a novel 'rising activity, multi-level mixed effects, indicator emphasis' (RAMMIE) method for syndromic surveillance in England. *Bioinformatics*. 31, 3660-65. PubMed https://doi.org/10.1093/bioinformatics/btv418

34. Smith GE, Elliot AJ, Ibbotson S, Morbey R, Edeghere O, et al. 2017. Novel public health risk assessment process developed to support syndromic surveillance for the 2012 Olympic and Paralympic Games. *J Public Health (Oxf)*. 39, e111-7. PubMed

35. Buckeridge DL. 2007. Outbreak detection through automated surveillance: a review of the determinants of detection. *J Biomed Inform*. 40, 370-79. PubMed https://doi.org/10.1016/j.jbi.2006.09.003

36. Morbey R, Hughes H, Smith G, Challen K, Hughes TC, et al. 2019. Potential added value of the new emergency care dataset to ED-based public health surveillance in England: an initial concept analysis. *Emerg Med J*. 36, 459-64. PubMed https://doi.org/10.1136/emermed-2018-208323

37. Craig AT, Kama M, Samo M, Vaai S, Matanaicake J, et al. 2016. Early warning epidemic surveillance in the Pacific island nations: an evaluation of the Pacific syndromic surveillance system. *Trop Med Int Health*. 21, 917-27. PubMed https://doi.org/10.1111/tmi.12711