

# Using Bayesian Networks to Assist Decision-Making in Syndromic Surveillance

Felipe J. Colón-González\*<sup>1</sup>, Iain Lake<sup>1</sup>, Gary Barker<sup>2</sup>, Gillian E. Smith<sup>3</sup>, Alex J. Elliot<sup>3</sup> and Roger Morbey<sup>3</sup>

<sup>1</sup>School of Environmental Sciences, University of East Anglia, Norwich, United Kingdom; <sup>2</sup>Institute of Food Research, Norwich, United Kingdom; <sup>3</sup>Public Health England, Birmingham, United Kingdom

## Objective

To analyse the use of Bayesian network structural learning to identify relations between syndromic indicators which could inform decision-making processes

## Introduction

Syndromic surveillance systems often produce large numbers of detections due to excess activity (alarms) in their indicators. Few alarms are classified as alerts (public health events that may require a response). Decision-making in syndromic surveillance as to whether an alarm requires a response (alert) is often entirely based on expert knowledge. These approaches (known as heuristics) may work well and produce faster results than automated processes (known as normative), but usually rely on the expertise of a small group of experts who hold much of their knowledge implicitly. The effectiveness of syndromic surveillance systems could be compromised in the absence of experts, which may significantly impact their response during a public health emergency. Also, there may be patterns and relations in the data not recognised by the experts. Structural learning provides a mechanism to identify relations between syndromic indicators and the relations between these indicators and alerts. Their outputs could be used to help decision makers determine more effectively which alarms are most likely to lead to alerts. A normative approach may reduce the reliance of the decision making process on key individuals

## Methods

We used Bayesian network structure learning to represent and quantify relations between indicators used by the English Real-time National Syndromic Surveillance System for classifying and investigating public health alerts. Summary data were obtained from four national syndromic surveillance systems coordinated by Public Health England. The network structure was initially learnt from the data using score-based algorithms. Temporal tiers were defined to separate variables on a timeline based on expert knowledge. Networks were constructed using a greedy search and iteratively adding, deleting, or reversing arcs, assessing the accuracy of the final network. The Bayesian Information Criterion was used to evaluate goodness of fit. Joint probability distribution for the data was estimated once the algorithm learnt the network structure. We evaluated how different algorithms and data features influence alert classification. The ability of the learnt networks for classifying alerts of public health importance was assessed retrospectively using 1000 bootstrap samples from the summary dataset for 2014

## Results

Provisional results show that our networks effectively evidenced the relations between variables and trends in the data known by the experts. The algorithm identified relations between variables that were unknown. The decision as to whether an alarm leads or not to an epidemiological alert was sensitive to several characteristics of the data stream and the period when such extra activity was detected. Different algorithms showed different degrees of specificity but most of them showed similar levels

## Conclusions

Decision-making methods may be difficult to maintain and replicate if they exclusively rely on expert knowledge. Bayesian networks may be useful to explicitly decompose and graphically describe the relationships among variables that are more useful for deciding whether extra activity in a syndromic indicator requires a response. When Bayesian networks are able to closely mimic the behaviour of syndromic surveillance variables, they may offer a useful explanation of how variables interact with each other which could assist syndromic surveillance teams for training, diagnostic and predictive purposes

## Keywords

Syndromic Surveillance; Bayesian Networks; Structural Learning

## Acknowledgments

We acknowledge support from Royal College of Emergency Medicine, EDs participating in the emergency department system (EDSSS), Ascribe Ltd and L2S2 Ltd; OOH providers submitting data to the GPOOHSS and Advanced Health & Care; TPP and participating SystemOne practices and University of Nottingham, ClinRisk, EMIS and EMIS practices submitting data to the QSurveillance database; and NHS 111 and HSCIC for assistance and support in providing anonymised call data the underpin the Remote Health Advice Syndromic Surveillance System. We thank the PHE Real-time Syndromic Surveillance Team for technical expertise. The authors received support from the National Institute for Health Research Health Protection Research Unit in Emergency Preparedness and Response. The views expressed in this abstract are those of the authors and not necessarily those of the NHS, the NIHR, the Department of Health or Public Health England

\*Felipe J. Colón-González

E-mail: f.colon@uea.ac.uk

