Institution: The Open University		
Unit of Assessment: B10 Mathematical Sciences		
Title of case study: Timely detection of outbreaks of infectious diseases		
Period when the underpinning research was undertaken: 1 Jan 2009 - 31 Dec 2017		
Details of staff conducting the underpinning research from the submitting unit:		
Name(s):	**Role(s) (e.g. job title):**	**Period(s) employed by submitting HEI:**
Prof Paddy Farrington	Professor of Statistics	1998 – 2015
Prof Paul Garthwaite	Professor of Statistics	2000 – 2020
Dr Angela Noufaily	Research Fellow	2011 – 2017
Dr Doyo Gragn Enki	Research Fellow	2011 – 2017
Period when the claimed impact occurred: 1 August 2013 - 31 December 2020		
Is this case study continued from a case study submitted in 2014? N		

1. Summary of the impact

Statistical methodology developed at The Open University is embedded in routine outbreak detection systems including *LabBase* at Public Health England, the *Signale* system at the Robert Koch Institute in Germany, and the *Signal Detection* tool at the European Centre for Disease Prevention and Control. The methods are also being used by the Centers for Disease Control and Prevention in the USA and Statistics Canada to estimate excess deaths during the COVID pandemic. They are implemented in a robust quasi-Poisson regression model for early detection of aberrant disease counts. These methods contribute directly to the public health in several countries in Europe and in North America.

2. Underpinning research

As part of its public health remit, Public Health England (PHE) routinely collects information on infections identified at NHS laboratories in England, Wales and Northern Ireland (Scotland has a separate system). In a typical week, reports on several hundred different infection types may be obtained. These data form the basis of a regional and national surveillance network to identify outbreaks of infectious disease caused, for example, by contaminated food or water. An automated outbreak detection system automatically scans the data each week to check for aberrances in the numbers of reports for particular infectious organisms. This automated outbreak detection system is the focus of this impact case study.

The system was developed at the Open University (OU) as part of a research programme undertaken in 2009 – 2015, in collaboration with PHE (who supplied the data), and led by **Farrington**, who maintains an active link with the OU as Emeritus Professor.
The key feature of the system that distinguishes it from other outbreak detection algorithms is that it provides early warnings of potential outbreaks across all possible infections surveyed, the number of which may run into hundreds or thousands. To this end, the system operates automatically, requiring no manual input, and is sufficiently robust to cope with the widely diverse features of the many different organism types that may be reported each week, notably their different frequency, seasonality, trend structure and intrinsic variability.

The outbreak detection algorithm developed at the OU **[O2]** is based on a quasi-Poisson regression model, which compares the current weekly count for each organism type reported with its expected value, adjusted for seasonal variation, trends, and past outbreaks. The output includes a list of potential problem organisms, ranked by their degree of aberrance.

The statistical and operational assumptions underlying the algorithm have been thoroughly validated on real surveillance data **[O1, O5]**. Further methodological developments to explore aspects of the system and, potentially, extend its range of application have been undertaken. Thus, the time delay between collection of the specimen for analysis and its identification was investigated **[O3]** as were methods for adjusting for such delays **[O4]**. Also, the relationship between the mean, variance, and skewness of weekly counts for different organisms was studied both empirically **[O1]** and more theoretically **[O6]** across the range of organism types.

3. References to the research

O1. **Enki, D.G.**, **Noufaily, A.**, **Garthwaite, P.H**. et al (2013) Automated biosurveillance data from England and Wales, 1991-2011. *Emerg Infect Dis* 19(1): 35 – 42. https://doi.org/10.3201/eid1901.120493

O2. **Noufaily, A.**, **Enki, D.G.**, **Farrington, P**. et al (2013) An improved algorithm for outbreak detection in multiple surveillance systems. *Stat Med* 32, 1206-1222. https://doi.org/10.1002/sim.5595

O3. **Noufaily, A.**, Weldeselassie, Y.G., **Enki, D.G**. et al (2015) Modelling Reporting Delays for Outbreak Detection in Infectious Disease Data. *J R Statist Soc A* 178, 205-211. https://doi.org/10.1111/rssa.12055

O4. **Noufaily, A.**, **Farrington, P.**, **Garthwaite, P**. et al (2016) Detection of infectious disease outbreaks from laboratory data with reporting delays. *J Am Stat Assoc* 111 (514): 488-499. https://doi.org/10.1080/01621459.2015.1119047

O5. **Enki, D.G.**, **Garthwaite, P.H.**, **Farrington, P**. et al (2016) Comparison of statistical algorithms for the detection of infectious disease outbreaks in large multiple surveillance systems. *PLoS ONE* 11 (8): e0160759. https://doi.org/10.1371/journal.pone.0160759

O6. **Enki, D.G.**, **Noufaily, A.**, **Farrington, P**. et al (2017) Taylor's power law and the statistical modelling of infectious disease surveillance data. *J R Statist Soc A* 180(1), 45-72. https://doi.org/10.1111/rssa.12181

This research programme was supported by the following peer-reviewed grants and award:

Farrington (PI). *Methodological development of syndromic and laboratory statistical surveillance in England*. NIHR 2009-10, GBP28,000.

Farrington (PI) and **Garthwaite**. *Statistical outbreak detection methods for large multiple surveillance systems*. MRC 2011-14, GBP621,000.

Farrington. *Statistical methods for pharmacoepidemiology and infectious disease epidemiology*. Royal Society Wolfson Research Merit Award 2011-15, GBP10,000 per annum salary uplift.

4. Details of the impact

The impacts described relate to use of OU research by five distinct public health or governmental agencies: Public Health England, the Robert Koch Institute in Germany, the European Centre for Disease Prevention and Control (an agency of the EU, operating across EU and EEA countries), the Centers for Disease Control and Prevention in the USA, and Statistics Canada.

Impact 1: The Public Health England outbreak detection system

In the UK, the control of infectious disease outbreaks is an important public health priority to which substantial resources are committed. For example, in 2018-19 Public Health England (PHE) devoted GBP86,900,000 and 2093 staff to this task through its national centres and regional network and responded to over 10,000 outbreaks **[C1** pp 11 and 14**]**.

The outbreak detection algorithm was developed at the OU in consultation with PHE statisticians and has been in use at PHE nationally and at its regional centres since 2013. A key function of the outbreak detection system is as a last line of defence, to identify outbreaks that have been missed at the local or regional level, through routine automated screening of all the available data. The system previously in use produced too many false alerts, resulting in an unnecessary burden of signal verification **[C2** p379 para 2**]**.

The new exceedance monitoring system forms an integral part of PHE's outbreak surveillance strategy **[C3** p2 para 1**]**. The algorithm runs automatically every week to flag up possible outbreaks, which are then reviewed by a team of epidemiologists. The new system has led to the detection of numerous infectious disease outbreaks. For example, an outbreak of dysentery among the English Orthodox Jewish community was first flagged by the outbreak detection system in North London in December 2013 **[C4** p2 para 3**]**. There were 52 cases; the outbreak was later linked to travel from an endemic area, indicating that public health advice to travellers may need to be reinforced to reduce imported infections **[C4** Abstract**]**.

Impact 2: The Robert Koch Institute outbreak detection system

The Robert Koch Institute (RKI), based in Berlin, is Germany's federal Public Health Institute. Among its tasks it lists in first place the "*Identification, surveillance and prevention of diseases, especially infectious diseases*". To support this key strategic aim, RKI have developed a comprehensive online surveillance system called *Signale.* The outbreak detection component of this system uses the algorithm developed at the OU **[C5** p3 para 2**]**. A senior infectious disease epidemiologist at RKI writes **[C6**; our emphasis**]**:

> "*One of RKI's core competencies is infectious disease epidemiology, and for this purpose we record and analyse data on the occurrence of numerous infectious diseases in Germany. A key component of this function is the Signale early warning system.*
>
> *The Signale system provides epidemiologists in Germany with early alerts of possible outbreaks of infectious diseases. Its aim is to flag up infectious diseases that register an unusually high number of reports, compared to the number expected.* ***One of the statistical algorithms used to perform the calculations is that published in 2013 by Noufaily A., Enki DG, Farrington P et al in the journal Statistics in Medicine (vol 32, pp 1206-1222). This algorithm was selected as it is appropriate for this application in view of its robustness and flexibility. [****Note: the reference is* ***O2]****.*
>
> *The Signale early warning system was developed in 2015 and runs weekly. Reports of possible alerts are provided to epidemiologists in Public Health agencies throughout Germany. A key function of the Signale system is to provide reassurance that major outbreaks have not been missed, notably outbreaks of food-borne infections. Also, in case of other available information, e.g. alerts about detection of special pathogens in food items, the Signale system is used to evaluate the epidemiological situation. It is especially useful for the analysis of the notification data for salmonellosis because we do not have the capacity to check all the combinations of different serotypes, locations and time periods manually*".

Impact 3: The ECDC outbreak detection system

The European Centre for Disease Prevention and Control (ECDC), based in Stockholm, is an agency of the EU whose aim is to strengthen Europe's defences against infectious diseases. In 2018, ECDC adopted the outbreak surveillance system developed at the OU and incorporated it into its routine surveillance tools. The Chief Scientific Officer at ECDC, writes **[C7**; our emphasis**]**:

> "*Our long-term strategy includes encouraging the development of strong, harmonised and efficient European surveillance systems to serve the member states, the European Commission and public health professionals. As part of this strategy, ECDC recently developed an Epidemiological Signal Detection package which is specifically designed to detect possible outbreaks at EU or country level, using infectious disease surveillance data*

retrieved from the European Surveillance System. Such outbreak detection systems are already in use at national level in some member states; they are applied in conjunction with other surveillance methods to provide early warnings of potential outbreaks. ***One of the two detection algorithms included in the Signal Detection tool at the heart of the package is that developed by Farrington and colleagues at The Open University and Public Health England, and published as Noufaily et al (2013) in Statistics in Medicine (vol 32, pp 1206-22).*** *We chose this algorithm because it is a validated and robust method that can detect the emergence of very rare infections, as well as outbreaks of more common infections, and has proved to be a popular choice with public health professionals […]. The package has been well received and will help to harmonise Europe's infectious disease surveillance systems in line with ECDC's strategic objectives*" **[**Note: the reference is **O2]**.

The Signal Detection Tool was developed at ECDC as an R package incorporating the OU algorithm, which is freely available online from the ECDC website. The purpose of the package is described in the following terms **[C8** p2 Background**]**:

"*The monitoring of routinely collected infectious disease surveillance data has the potential to ensure that relevant changes in disease incidence such as possible outbreaks are promptly recognised, investigated and that control measures are initiated in a timely manner. Due to the large amount of data being collected as part of such surveillance, there is a need to develop automatic detection algorithms, which can assist epidemiologists in this task*".

Impact 4: Monitoring excess mortality during the COVID pandemic

Since the start of the COVID-19 pandemic, the OU outbreak detection algorithm has been adapted to estimate excess mortality (the most reliable measure of deaths attributable to this coronavirus) in the USA and Canada.

The Centers for Disease Control and Prevention (CDC), the main public health agency responsible for surveillance of infectious diseases in the USA, describe the methodology underpinning their system for monitoring excess mortality in the following terms **[C9** Technical Notes para 1**]**:

"*Counts of deaths in the most recent weeks were compared with historical trends (from 2013 to present) to determine whether the number of deaths in recent weeks was significantly higher than expected, using* ***Farrington*** *surveillance algorithms (****Noufaily*** *2013)*" **[**Note: the reference is **O2]**.

Likewise, the OU algorithm was used by Statistics Canada, that country's national statistics agency, to monitor excess mortality during the COVID-19 pandemic. Their methods are briefly described as follows **[C10** p7 para 1**]**:

"The model used to estimate the expected number of deaths is adapted from an infectious disease detection algorithm developed by ***Farrington*** *et al. and modified by* ***Noufaily*** *et al. and Salmon et al."* **[**Note: these last two references are **O2** and **C5]**.

In conclusion

The statistical outbreak detection methodology developed at the OU has had wide-ranging impact on systems designed to protect the public health, both within the UK and in other countries. In addition to its intended use for outbreak detection, the methodology has proved to be versatile and of wider significance, as shown by its use to monitor excess mortality during the COVID-19 pandemic.

5. Sources to corroborate the impact

C1. PHE Annual Report and Accounts, 2018-2019.

C2. Freeman R, Charlett A, Hopkins S et al (2013). Evaluation of a national microbiological surveillance system to inform automated outbreak detection. *Journal of Infection* 67(5), 378-384. https://doi.org/10.1016/j.jinf.2013.07.021

C3. Chalmers RM, Robinson G, Elwin K and Elson R (2019). Analysis of the *Cryptosporidium* spp. and *gp60* subtypes linked to human outbreaks of cryptosporidiosis in England and Wales, 2009 to 2017. *Parasites & Vectors* 12: 95 https://doi.org/10.1186/s13071-019-3354-6

C4. Rew V, Mook P, Trienekens S et al (2018). Whole-genome sequencing revealed concurrent outbreaks of shigellosis in the English Orthodox Jewish Community caused by multiple importations of *Shigella sonnei* from Israel. *Microbial Genomics* 4(3). https://doi.org/10.1099/mgen.0.000170

C5. Salmon M, Schumaker D, Burrmann H et al (2016a). A system for automated outbreak detection of communicable diseases in Germany. *Euro Surveillance* 21(13): pii=30180. https://doi.org/10.2807/1560-7917.ES.2016.21.13.30180

C6. Supporting statement from the senior infectious disease epidemiologist at RKI confirming the use of the OU algorithm in designing *Signale* for the early detection of infectious disease outbreaks in Germany.

C7. Supporting statement from the Chief Scientific Officer at ECDC confirming the use of the OU algorithm in designing the *Signal Detection* tool for the early detection of outbreaks of infectious diseases in Europe.

C8. Merdrignac L, Gomes J, Kissling E et al (2018). The 'Epidemiological Signal Detection' package. https://ecdc.europa.eu/sites/portal/files/documents/EpiSignalDetection-Vignette.pdf

C9. Centers for Disease Control and Prevention. Excess Deaths Associated with COVID-19.

C10. Statistics Canada. Excess mortality in Canada during the COVID-19 Pandemic.