Exploring the joint analysis of routine data and pathogen genomic datasets in the investigation of outbreaks of GI infection.

Funded under Food Standards Agency call FS301019: Exploring the potential of integrating next generation sequencing and other 'big data'

Draft report to FSA 16 August 2016

Final report post FSA review 5 April 2017

Lead institution

University of Warwick

Collaborating Institutions:

Public Health England

University of Oxford

University of Liverpool

Project team

Study investigators (applicants and researchers named on the bid in *italics* and those leading case studies marked with ***)**

University of Warwick

*Noel McCarthy** (Principal Investigator), *Hendramoorthy Maheswaran**, Pam Royle*, Eileen Taylor

Public Health England

Tim Dallman, Kathie Grant, Jeremy Hawker, Jonathan Green, Chantil Sinclair*, Eleanor Rees*, Jacquelyn McCormick*

University of Oxford

Martin Maiden

University of Liverpool

Sarah O'Brien

Contributors of outbreaks and in the development of case studies from these

Public Health England

Charles Beck, Emma Bennett, Paul Cleary, Richard Elson, Kirsty Foster ,Russell Gorton, Ian Hall, John Harris, Tom Inns, Chris Lane, Lesley Larkin, Noeleen McFarland, Piers Mook, Gordon Nichols, Deborah Wilson, Roberto Vivancos

European Centre for Disease Prevention and Control

Julien Beauté

Expert workshop participants are listed in the workshop report

Contents

Main lessons priorities and recommendations

Genomic clusters detection

1. Genomic clusters appear to detect foodborne outbreaks with good specificity and can allow earlier detection if sequencing is rapid and appropriate small clusters are investigated. Investigation of even small clusters may efficiently identify sources for small or early outbreaks and is recommended where resources allow.
2. Cluster investigation is substantially more efficient where complementary forms of data are available. Obtaining at least basic case epidemiology data on a consistent and accessible basis nationally is the first priority for complementary information. A proposal for this is outlined in Annex 2.
3. Widely distributed (in space and time) outbreaks, which were often missed are identified more commonly with the addition of pathogen genome sequencing (GS) to surveillance. Complementary case epidemiology data to support this outbreak type (e.g. shared brands or suppliers) should be considered in decisions on what data to collect for each case.

Epidemiological control data for comparison of e.g. exposures with cases

4. Routine sources of population exposure data may provide a valid alternative or complement to control data from outbreak studies and should be explored for use in outbreaks. This requires mapping of available data and establishing access so that it is available rapidly in outbreak settings.
5. Internet panel controls are increasingly used, very efficient, and supported by good work to develop a system to make access efficient. They are inevitably a biased population sample and protocols for use should consider approaches such as over sampling to allow better matching or propensity scoring to decrease bias.
6. Gaining timely access to good population exposure representative controls could be facilitated by PHE pre-arranging access to more representative controls such as could be accessed via the NHS. This would require agreements and a protocol analogous to the developments already made for internet panel controls.

Increased descriptive and analytical epidemiology accuracy and efficiency and GS data

7. Integrating GS data into case definitions supports more efficient and informative application of trawling questionnaires and food tracing studies.
8. Where appropriate increasing specificity of case definition by using GS data can allow more efficient and accurate case-control studies. Increased accuracy may support undertaking these in smaller outbreaks. This integration of GS into epidemiological investigation is recommended. Retrospective application to some completed investigations would allow increased experience and demonstration of the effects of this added dimension of data.

Linking cases to upstream sources in the food chain

9. Linking human cases to possible sources using GS may be very useful. However, especially in the absence of corroborative epidemiological information, linkage to a reservoir or source by GS alone could be incorrect if the outbreak clone is relatively stable and widespread. Pathogen population data including GS across possible sources of human infection, and an understanding of the distribution of pathogen populations these will allow more reliable inference.
10. Both structured samples of such sources (animal, food and environment) and the assembly from other potentially available data, such as data from testing done by industry for their internal quality control, are high priorities to support more robust inference from GS data in the investigation of foodborne disease.
11. Extensive epidemiological food chain investigation is likely to continue to increase and complement GS data. This motivates both the development of shared infrastructure and agreements to facilitate FSA-PHE data sharing, and the development of formal study methods. An outline protocol for a food venue case-control study is included in Annex 2.

More detailed consideration of lessons from the outbreak case studies are given on pages 12 to 17 and a summary of discussion of the main themes at the expert workshop on pages 21 and 22.

Background and aim

The application of pathogen genome sequencing (GS) to surveillance has many impacts including the increased detection of widely distributed outbreaks. In parallel with human surveillance GS data the collection of extensive data by industry, regulatory, health and other sectors now creates large datasets on food, and microbial populations in the food chain, environment, and human disease. These developing resources complement bespoke datasets generated during incident investigation and offer substantial potential to aid the detection and investigation of outbreaks. However data sharing across public and private sector organisations can be difficult to achieve during an acute incident and where the purpose is not clear. This motivates the development of: i) processes to support ad hoc collaboration including data sharing; ii) underlying data sharing infrastructure for agreed purposes; and iii) learning across incidents and other data to support interpretation of novel data and analyses.

This project used expert review of outbreak case studies to identify:

1) benefits achievable by the effective integration of genomics and other datasets in incident and outbreak investigation;

2) issues and barriers to this integtration;

3) learning points from the reviewed outbreaks;

4) examples that will support the motivation of wider partners to work together to establish data sharing priorities.

Some of the findings support implementation while others could guide further research into optimising the use of available and emerging genomic and big data technologies in the investigation of potentially foodborne disease incidents.

Methods

Summary of methods

Identification of outbreaks was by literature review, review of a national outbreak database and contact with experts. Selection of outbreaks used criteria agreed by the project group. Outbreaks for case studies were selected to cover a range of complexity, geographies, types of pathogen, outbreaks involving collaborative work with the FSA on food sources, and different data issues and opportunities (Table 1).

The original investigators of the outbreaks selected as case studies were contacted. We requested sharing of outbreak reports, publications and other available data; and their opinion on approaches and data identified as useful at the time or in retrospect. A data extraction form was used to capture a summary of each outbreak including documents reviewed, individuals contacted, and specific issues and learning identified. These are included in the Annex 1.

All case studies were reviewed by the project PI to collate lessons learned and present these to the expert workshop. Themes for prioritisation were selected and expanded at the expert workshop discussions.

Where appropriate and feasible additional work was undertaken to scope proposals or gaps in the evidence base arising from review of case studies and priorities identified at the expert workshop. These are presented in Annex 2.

Identifying outbreaks for case studies

We undertook a literature review of food-borne pathogen associated outbreaks in the UK in February 2016 published since 2010. Searches included the electronic databases (MEDLINE; Embase). The full search strategies are included in the appendix. These identified 202 entries in Embase and 186 in Medline with 60 and 57 retained after title screening respectively. Removal of duplicates left 71 peer-reviewed publications. After assessment of the abstracts by the study team, 14 publications relating to foodborne outbreaks were retained for consideration (listed in the appendix).

We reviewed outbreaks that occurred since 2010 and had been reported to the PHE electronic Food Outbreak Surveillance System (eFOSS). This is a national database to which PHE staff report outbreaks from across the country that are, or might be, foodborne. In total there were 395 foodborne outbreaks and 143 non-foodborne outbreaks. Of the 395 foodborne outbreaks on the database, 40 outbreaks were selected for more detailed assessment for inclusion by excluding those that occurred in institutional settings where it would be relatively easy to investigate outbreak (e.g. residential care home; schools), occurred at small events (e.g. weddings/parties), or were associated with a single restaurant. Of the 143 non-foodborne outbreaks reviewed from the database, eight were selected for detailed assessment having excluded those that could be easily traced to water sources (swimming pool) or farm visits, and those where it was evident that food would not have been considered a likely source during the investigation. To identify outbreaks that might not have been published or reported to eFOSS we emailed and made either telephone or face to face contact with members of each PHE Field Epidemiology Service team to brief them on the type of outbreaks that would be useful for the study and to request suggestions for suitable outbreaks. The study investigators involved in leading case studies and PHE colleagues in the PHE Gastrointestinal Infections Department also met to identify suitable outbreaks known to the national team.

Selecting outbreaks

Outbreaks identified were screened using our minimum inclusion criteria of:

1. Suspected or confirmed food poisoning outbreak, including outbreaks which were originally believed likely to be foodborne and subsequently identified as having a different vehicle.
2. Where whole genome sequencing or other specific typing methodology was applied.

Outbreaks remaining were reviewed and a shortlist selected that provided sufficient heterogeneity (Table 1) across criteria agreed with the study team including the FSA. Fifteen outbreaks were selected in case of delays in getting access to data or if some outbreaks proved uninformative. The outbreaks selected are listed in Table 2 and

mapped against themes addressed in Table 3. All fifteen outbreaks were considered informative and all were included as case studies rather than the planned twelve.

Table 1: Criteria for selection of outbreaks from the long list to ensure diversity.

Criteria	Examples
Varied geographical distribution of cases	Localised, National with local cluster, National, International
Timescale of outbreak	Short (<21 days), Medium (3-10 weeks), Persistent (>10 weeks)
Pathogen responsible for outbreak	Bacterial, Viral, Parasitic
Source and route identified	Unknown, Local supplier, Non-food, National retailer Restaurant chain, Business to business (e.g. wholesaler)
(Potential) use of a range of routine and bespoke data sources	Electronic customer records, Food chain data, External genetic databases.

Obtaining information for and conduct of case studies

Published papers and outbreak reports were obtained for outbreaks as the minimum basis for a case study. Additional documents and data were requested where available. Outbreak investigation leads and epidemiology leads were contacted to confirm interpretations being made in case studies where needed. Each case study was undertaken by one or two members of the project team. All case studies used a standard data extraction form that identified the sources of data used, contacts, information on the outbreak, and specific lessons from each outbreak (Annex 1).

Other aspects of case studies varied with some involving additional analysis by the project team or original outbreak leads, or supported by literature reviews of the area and some identifying and specifying additional studies that could be done with outline protocols for these. These are included within in each case study (Annex 1), or Annex 2 where more substantial extensions such as literature review or the development of protocols for additional studies were done.

Table 2. List of outbreaks and associated case study identification numbers.

Case study	Outbreak title
1	Outbreak of *E. coli* O157 infection associated with two butcher's shops in Wingate and Billingham in July 2015.
2	An outbreak of untypable *Salmonella* Typhimurium gastroenteritis linked to ham and gammon consumption in Thames Valley, July – September, 2013. [and related outbreaks]
3	Outbreak of *Salmonella* Agona phage type 40 associated with the Street Spice Festival, Newcastle upon Tyne, February / March 2013.
4	Feeder mouse associated outbreak of *Salmonella* Typhimurium DT191a in England and Wales, 2008-2009
5	Outbreak of listeriosis linked to crabmeat consumption in England.
6	*Cryptosporidium parvum* exceedance – November 2015. [and other outbreaks that used internet survey panels to obtain control data]
7	Related Hepatitis A outbreaks associated with semi-dried tomatoes.
8	Recurrent seasonal outbreak shiga-toxin producing *Escherichia coli* (STEC) O55:H7 in South West England, July 2014-September 2015.
9	National outbreak of *Salmonella enterica* Bovis-morbificans July 2015.
10	*E.coli* O157 outbreak linked to the consumption of raw cows' drinking milk from a farm dairy. [name removed]
11	VTEC *E.coli* O157 outbreak associated with a restaurant in Belfast.
12	Norovirus outbreaks in England linked to consumption of raw Oysters in 2010, with trace back to predominantly one producer in Ireland.
13	Outbreak of *Salmonella* Goldcoast, September-November 2013.
14	Prospective use of Whole Genome Sequencing (WGS) detected a multi-country outbreak of *Salmonella* Enteritidis.
15	A multi-country *Salmonella* Enteritidis phage type 14b outbreak associated with eggs from a German producer: 'near real-time' application of whole genome sequencing and food chain investigations, United Kingdom, May to September 2014.

Lessons identified across outbreaks

Table 3. Lessons identified grouped in main themes, and outbreaks supporting each of these indicated with an X.

Theme addressed	CS1	CS2	CS3	CS4	CS5	CS6	CS7	CS8	CS9	CS10	CS11	CS12	CS13	CS14	CS15
Alternative controls	X		X	X		X							X		
FSA role in outbreaks	X							X		X					X
Outbreak detection	X	X		X	X		X			X	X		X	X	
Descriptive epidemiology combining GS and other data	X				X		X	X	X	X				X	X
Inference from geographical patterning	X					X	X	X		X	X			X	
GS refined epidemiological studies	X												X		
Gaps in sampled population genetic reference data	X				X		X	X		X	X	X			X
Food chain investigation and uses		X	X	X	X		X	X		X				X	X
Sharing rich data rapidly		X			X		X	X							X
Extending studies (e.g. long term outcome)			X												
Novel linkage			X					X	X	X				X	X
GS data saving work													X		

Outbreak detection

Genomic cluster investigation would most evidently have speeded the detection of the outbreak in CS4 where development at about three cases per week for almost six months in a common *Salmonella* serotype delayed identification. Outbreak CS13 was identified by a temporospatial cluster of three cases of a rare serotype (*Salmonella* Goldcoast). Nationally exceedance monitoring would have triggered one week later. A genomic cluster of two was detectable nine weeks earlier (plus sequencing time) and of four one week before the temporospatial cluster, so that only action with a cluster of two would have allowed earlier detection if real time GS was in place compared to the far greater scope for earlier identification in CS4. Literature review supporting CS5 identified that many outbreaks of listeriosis outbreaks are detected when relatively small, in part due to highly discriminatory subtyping being common, and additionally this being coupled with further investigation of these small clusters. Outbreak CS14, occurring after real-time GS cluster identification was implemented in PHE was identified through this genomic surveillance. The outbreak strain showed substantial diversity of phage type which would have severely limited the investigation without GS. Overall GS clusters allowed or could have allowed earlier identification of outbreaks, in particular (i) for slow burn outbreaks, (ii) for those of common phenotype, and (iii) if it is possible to investigate small clusters.

Linking of microbiological characterisation to other data was critical to rapid confirmation of clusters as outbreaks and to support control. In CS1 a single VTEC case plus risk premises identification initiated action. In CS10 two VTEC cases with a shared risk factor initiated action including a product recall. In CS11 two local VTEC cases in context of a risk premises initiated substantial action. In CS7 two HAV cases in the context of external information. Common exposure information across two small local *Salmonella* clusters allowed outbreak detection in CS2. CS14 and CS15 each identify that Salmonella questionnaires vary across the UK and that this limited early access to comparable information. The literature review of listeriosis outbreaks in CS5 also identified the role of detailed questionnaires for each case, targeted food sampling, and in several cases reference to routine food samples to support both detection and investigation of outbreaks. CS7 showed how genetic sequencing could act as an indicator for relatedness between outbreaks with actions on just 2-5 cases in the UK and Netherlands in 2011 informed by the earlier outbreaks in Australia and elsewhere that had shared the outbreak genotype.

Overall the reviewed outbreaks showed that (i) triggering outbreak investigation in response to GS clustering alone can detect outbreaks earlier, (ii) GS data can link case clusters to other contextual information such as identical or highly similar isolates in particular sources, and (iii) combining different forms of epidemiological intelligence is usual in the early identification of outbreaks and their source. Barriers to accessing epidemiological information for cases in GS clusters are a source of inefficiency that may limit the effective use of the technology to support disease control.

Outbreak investigation: identifying sources of infection

Positive cultures at reservoirs and from cases are often accompanied by failure to detect infection along the transmission chain. Examples were CS10 where farm and cases shared GS (and multi-locus variable number tandem repeat) identity but no culture positive food (unpasteurised milk) could be identified in between despite sampling and CS14 where none of 1962 eggs sampled were positive despite a substantial outbreak linked to a source and transmission chain. In CS1 the certainty provided by the combination of shared GS across cases and temporospatial clustering initially, with later GS support for origin at a farm, allowed apparently conflicting exposures reported by cases to be interpreted as demonstrating cross contamination at implicated butchers premises rather than this lack of consistency in exposures being a problem. Detection of the source and relatively detailed mapping of aspects of the transmission route could therefore be achieved in a small outbreak (13 lab confirmed cases) by the synthesis of detailed data in several dimensions. The restricted geographical pattern of cases supported investigators in not following up major retailers as possible sources.

The geographical origin of several outbreaks was inferred from GS of case isolates combined with existing GS data from samples with known geographic origin. For example the Hepatitis A genotype 1B in CS7 is a strong predictor for Turkish origin, the CS10 VTEC where cases were due to infection from a farm in an area where distinct but related human clusters (5-25 SNPs) had been observed in the past, and CS11 where travel related cases suggested that the foodstuff producing these domestic cases may have had a Mediterranean origin.

Food chain tracing and GS allowed construction of the transmission network for most cases in the UK outbreak(s) in CS15 and this could be linked to an international context by Rapid Alert System for Food and Feed (RASFF) and shared MLVA and GS data. In CS14 attempted food chain tracing identified gaps in traceability of food despite EC Regulation 178/2002 under which this data should have been available.

These examples confirm the role of GS in demonstrating the relationship between cases and an upstream reservoir or suggesting the geographical origin that can provide a framework for more detailed identification of sources and transmission routes. However some of these inferences rest on an assumption of informative geographical structure in the pathogen population biology. Persistent and widespread bacterial clones could undermine the reliability of these assumptions. The case studies showed obvious gaps in the evidence base for this as outlined below as well as and gaps in compliance with food traceability regulations.

Several outbreaks highlighted limitations to inference and evidence gaps on the distribution of pathogen populations across possible sources of human infection. In CS1 the “same” strain as tested by GS was identified two months later on implicated farm. There was no data on how stable and widespread the identified clade is to support interpretation and inference. Literature review to support the listeriosis case study

(CS5) and summarised in the Annex 2 showed on the one hand many practical examples where food subtype matching case subtype identified or confirmed source. However there were also examples using PFGE and even GS of (almost) indistinguishable isolates in separate food production plants and over many years. In CS7 Hepatitis A partial sequencing suggested similarity to an Australian outbreak linked to dried tomatoes from Turkey. However a subsequently published study showed the very limited variation among clinical Hepatitis A isolates in Turkey limiting the ability to infer any particular food source from GS at the level performed. In CS12 the pattern of norovirus is considered to reflect human infection coming to oysters via sewage but not monitored in a way that allows any direct linkage of cases and source.

The outbreak CS8 appears to originate from a local reservoir with relatively high diversity compared to usual VTEC outbreaks. Outbreak CS10 was investigated in the context of isolates with 5-25 SNP distances from the outbreak strain typically found in the same broad geographical area as the outbreak source. CS15 Identified sampling from a moderately diverse reservoir as the likely source of the observed phylogeny, and that this population was likely present substantially before the start of the outbreak. The extent to which the population is limited to the implicated farms is unclear. Only a small number of similar isolates in UK human surveillance in 2012 suggests that this was the first time it caused substantial disease in the UK. CS15 was also an outbreak where the pattern of GS relatedness of cases could be mapped onto the identified probable transmission chain and supported separate but related sources of infection. Other examples where more sophisticated analysis of GS than only identifying if isolates are highly similar or not include CS8 where inference on a nursery sub-cluster included monophyletic pattern as well as SNP distance, CS11 where phylogenetic tree topology showed relatedness but not descent of the second VTEC outbreak in a Belfast restaurant by comparison to isolates from the first outbreak. CS14, like CS15 mapped tree topology to food distribution systems to give informative results.

Review across these outbreaks identifies (i) that in most, but not all, we are sampling diversity in the pathogen population giving rise to the outbreak with little within outbreak evolution, and (ii) that we have very little empirical data to understand either diversity in these source populations or the extent to which stable “identical” clones can be widespread and persistent rather than specific to a source of an outbreak. Geographical and reservoir population biology needs to be quantified to support this inference. Building up both this empirical evidence base and approaches that use appropriate population genetic theory in our analyses will make inferences more robust.

Outbreak investigation: food chain studies

Food chain investigation is increasingly common in local as well as larger scale outbreak investigations. There was extensive FSA involvement in the geographically restricted outbreak CS8 and CS1 involved substantial food chain tracing in a local outbreak. In CS10 urgent recall was based on two cases with a shared high risk exposure and there was also substantial subsequent FSA involvement in this seven (identified primary)

case outbreak. In CS3 local authorities undertook extensive food chain investigation which was shared with the FSA and the FSA was involved in discussions on what this meant for guidance on the use of curry leaves. These food chain investigations included information from human cases as well as the food chain that needs to be synthesised. The literature review for CS5 identified that extensive food chain investigations in listeriosis are common in small outbreaks and have often identified sources of infection. The increasing local outbreak partnership of the FSA and PHE and tendency to include food chain investigations in these smaller outbreaks might be better supported by an agreed appropriate FSA role and representation on OCTs and identifying shared data areas to work on e.g. supporting joint inputs into food tracing. This may fit within agreed approaches to much large outbreaks (e.g. CS15) where there is longer experience of collaboration, although here also the nature and extent of food chain tracing may be substantially changing with GS data.

Industry collaboration in CS4 included the Reptile and Exotic Pet Trade Association (REPTA) using their data to trace the pet reptile food chain and support microbiological testing and CS10 where the supplier's internet orders identified supply to cases who did not recall raw milk exposure.

Several investigations showed ineffective implementation of recall such as in CS2 where presumed recall from main distributor had not been effectively cascaded and CS10 where GS identified ongoing cases after the FSA recommended recall via internet customer records was not implemented. Food chain networks identified in tracking can be used to reinforce recall implementation. Active oversight of recalls could use GS to support monitoring of the effectiveness of recall.

Most food chain investigations were descriptive without quantitative or statistical inference. Substantial investigation in CS14 identified that 43/87 cases could be linked to a food chain network. How can this move from description to quantitative inference on how supportive this is for an association between the food chain and the disease? One approach would be to undertake analytical study approaches to test this. An outline protocol for this is given in Annex 2. This would involve mapping out control food chains, unrelated to cases, to allow for comparison. Further forms of analyses that can support disease control are those comparing identified food chain structures and the phylogenetic pattern among cases mapped onto this food chain. CS15 demonstrated joint formal analysis across food chain and phylogeny with results that supported understanding of spread and control. The large scale of the outbreak and tracing opens a question on whether a sampled approach could be efficient. CS14 showed that in some outbreaks genomic diversity may map to the food chain giving support to this as the route of transmission and possibly helping to identify which exposures should be sought for cases linked initially through genome sequence to an outbreak. Most of the association was related to the part of the close to the point of consumption in this instance. Both the novel approaches to jointly analyse phylogeny and the possibility for formal analytical studies as proposed in Annex 2 identify the need to consider sampling and

specific study design rather than global descriptive approaches. Such formal studies may be more beneficial to improving outbreak investigation and guiding intervention.

Outbreak investigation: refining epidemiological studies by integrating GS

Although repeating the analytical study analysis undertaken in the investigation of CS13 but using GS to refine the case definition increased the strength of association by 20%, the original analytical study had identified the association with source effectively if less accurately. For smaller outbreaks the difference might be more critical motivating the theoretical estimation of the effect of reduced case misclassification possible by GS make a concrete difference. Considering an outbreak with eight true cases, 75% of whom recall eating the food actually associated with illness and 16 corresponding controls 25% of whom recall eating the food would produce an analysis as in Panel A below. Allowing a further four people to be incorrectly classified as cases (e.g. same phenotype in the absence of GS) would give an analysis as in Panel B. The impact of GS informed analytical epidemiology here moves the result from one of weak support for the source (Panel B) to strong evidence (Panel A).

Panel A
Odds Ratio = 9 [1 – 114], p = 0.03

	Case	Control
Ate	**6**	**4**
Did not	**2**	**12**

Panel B
Odds Ratio = 4 [0.7– 29], p = 0.12

	Case	Control
Ate	**7**	**4**
Did not	**5**	**12**

In CS1 a case-control study with just 10 well defined cases gave a practical example of how effective analytical epidemiology can be when true outbreak associated cases are not mixed with other cases even when numbers are small. GS can rule out background VTEC cases to support this, especially for common phage types or widespread outbreaks where the risk of non-outbreak cases being incorrectly included is higher.

In CS13 GS would also have ruled out two of the 10 people where trawling questionnaires were undertaken saving work and strengthening the association with whelk consumption from 5/10 cases (50%) to 5/8 (62.5%). The time saving here where an unusual food caused the outbreak is relatively lower than where a common food is involved that could lead to food chain investigations for several possible foods. In CS8 the

particular pattern of GS relatedness, identifying a monophyletic cluster with different transmission influenced the approach to the investigation.

Non-genomic lessons and record linkage

Studies used a wide variety of control data. Panel controls from internet panel surveys were used and there is now a PHE guide and established set of relationships to support this. However, the (PHE published) literature reviewed in preparing CS6 identified clear evidence of bias in these controls compared to PHE staff controls. In contrast controls identified via the NHS patient registers appear to be at substantially less risk of bias. However there is a lot of work in gaining access to these and no national approach has been identified and agreed between PHE and the NHS. This could potentially increase the efficiency of this approach and make it more feasible in an outbreak setting. One study assembled a cohort in large part using social media, which approach might be useful in future for some types of populations. In CS4 population exposure data could have been more easily and reliably obtained using estimates from an existing study, the pet food manufacturers pet population survey. The literature review around CS7 identified an example of ECDC using European Food Safety Authority estimates for berry consumption as their comparator. These forms of routine data may be increasingly available as estimates of population exposure levels.

Follow up to identify longer term burden was undertaken in one outbreak (CS3) where there were some concerns for persistent severe symptoms. This identified the feasibility of this for consideration as a more general approach to monitor disease burden and factors contributing to severe outcome.

Linkage of multiple datasets included a capture recapture study in CS3 that used a social media cohort, laboratory notification and reports to local public health authorities as data sources. This work to estimate overall outbreak size used publicly accessible tools to match individuals without unambiguous identification data.

Outbreak CS9 where a moderate sized distributed outbreak was traced back to a train station venue highlights the potential power of linking genomic and geographical data if it is possible to follow movements of cases in their exposure period. This motivates consideration of linking to mobile phone data with initial investigation of issues in this outlined in the Annex 2.

As well as individual case linkage other forms of linkage identified in examples included geographical analysis of bird migrations and farm animal movements in CS8 seeking to identify an animal reservoir to explain a recurrent outbreak of VTEC and using genomics databases to link

an outbreak strain to other individuals or settings where this was common such in GS14 where cases in the US were linked to a European outbreak.

Workshop report

FS301019 expert workshop, 28 June 2016, University of Warwick.

Participants

Paul Cook	Food Standards Agency
Giorgia Albieri	Food Standards Agency
Drazenka Tubin-Delic	Food Standards Agency
Tracy Bishop	Food Standards Agency
Andy Hill	BAE Systems
Pam Royle	University of Warwick
Noel McCarthy	University of Warwick
Hendy Maheswaran	University of Warwick
Laura Guzman Rincon	University of Warwick
Chantil Sinclair	Public Health England
Eleanor Rees	Public Health England
Tom Inns	Public Health England
Jacquelyn McCormick	Public Health England
Paul Cleary	Public Health England
Tim Dallman	Public Health England
Jeremy Hawker	Public Health England
Alison Waldram	Public Health England
Lesley Larkin	Public Health England
Martin Maiden	University of Oxford

Apologies

Public Health England: Jonathan Green, Richard Elson, Kirsty Foster, Kathie Grant, Deb Wilson, Russell Gorton, Bob Adak, Piers Mook.
University of Liverpool: Sarah O'Brien
Animal and Plant Health Agency: Charlotte Featherstone
Imperial College: Thibaut Jombart

The workshop comprised:

1. Presentation of themes emerging across outbreak case studies.
2. Presentation of another project funded by FSA under the same theme:

Towards an integrated food safety surveillance system: exploring the potential for combining genomic and epidemiological metadata. Andy Hill, BAE Systems.

3. Group participation in discussion of findings.
4. Group discussion of main themes and identification of priorities for implementation and further research.

Presentation and discussion of themes from the case studies:

Early detection was identified as a critical strength of genome sequence data but that this is (i) dependent on speed of turn-around of isolate sequencing and (ii) that the detection of many very small outbreaks may not be useful without other forms of complementary data since investigation of very large numbers of clusters is not feasible. Lower speed than some pre-genomic methods has been a problem in both detection and efficient investigation of outbreaks. Linking to possible sources where similar isolates have been identified in the food chain is increasingly supporting outbreak detection and investigation. Combination of genomic clustering with other data increases effectiveness and efficiency of outbreak detection but this is limited for most pathogens. Efficient capture of data from cases is a priority including using available technologies to support this.

Inference on source based on similarity of genomes alone may not be valid. There are large gaps in data on the extent to which clones persist in the environment and the food chain, and on how widespread clones with very limited diversity can become. Where available structured surveys (such as FSA funded VTEC work) are invaluable. Studies on *Listeria monocytogenes* in the food industry suggest that narrow clones can persist for several years in food premises and across sites (rapid literature review within case study 5). Hepatitis A (limited) sequencing could link outbreaks back to an origin in Turkey but inferences in the literature reviewed as part of CS7 that this identified a likely food source are less valid given the narrow range of diversity of hepatitis A in Turkey. In contrast confirmation by sequencing of a source identified by other means is more reliable. The added discrimination offered by genome sequencing over other typing can support mapping of transmission routes between reservoir and cases not possible without these points of relative certainty. Positive cultures in cases and reservoirs are usual but contamination levels and positivity of culture in the food chain in between is often low (e.g. 0 of 1962 eggs in case study 14). Case study 1 shows how knowing the main transmission pathway supports different interpretation of apparently discordant exposure information from cases: here showing that different reported foods likely demonstrates cross-contamination in meat preparation.

Patterns of genetic relatedness were informative as well as closeness, e.g. monophyletic subgroup in an outbreak mapping to human to human transmission compared to other cases sampling from an unidentified source (case study 8). Similarly analysis of food distribution network and outbreak phylogeny has helped to clarify transmission in a widely distributed egg associated outbreak (case study 15).

An expansion of food chain studies including FSA involvement in local as well as national outbreaks was identified, with this becoming an increasingly standard part of outbreaks not limited to a single venue. Industry collaboration supported investigation of internet based raw milk sales leading to illness in contrast to typically less organised data that is often not accessible electronically, especially for small retailers. Even in large supermarkets the capacity to identify which batches were sold to which customers appears to be limited. The lack of infrastructure and arrangements to share data during outbreaks between PHE and FSA was identified as a persistent problem of increasing importance. Genome sequencing may increase efficiency in food chain tracing by identifying outbreak associated cases more specifically.

Refinement of analytical epidemiological studies was tested on case study 13 where the measure of association between the exposure and outcome increase by 20% if only the cases confirmed by genome sequencing were included. In small outbreaks and those where the non-genomic typing is less discriminatory this may be even more critical. Sequencing will also support more valid case-case studies.

Different approaches were used in seeking control data to compare to the exposures reported by cases. These include internet panel surveys where PHE has developed protocols. However work by PHE has also confirmed substantial bias by these. Solutions to decrease this may be substantially oversampling to allow selection of a subset of more valid controls. Use of aggregate national data was identified as more efficient and valid in some cases (e.g. using population data for exotic pet ownership rather than undertaking new survey work and similarly to estimate berry consumption in a hepatitis A outbreak.). GP controls (case study 1) may provide particularly valid data but with difficulties in an outbreak setting in the absence of established agreements and processes with the NHS. Novel methods such as using social media to assemble a study cohort (case study 3) may be increasingly useful.

Areas where record linkage was applied or identified as potentially useful in case studies included linkage to mobile phone data (case study 9), electronic purchase record (case study 10), linking of social media and routine notification data to estimate outbreak size (case study 3) wild and farmed animal movement (case study 8) and linking via international genomics databases (case study 14). Sharing of narratives linked to genetic data were identified as useful in case study 7. The potential to link individual case history data to genomic data was repeatedly identified as an important gap for some pathogens (case studies 2, 15, 15).

Follow up to identify longer term outcomes of patients may become more feasible and has already being undertaken in routine outbreak investigation (Case study 3).

Parallel project under this call

Andy Hill presented “Towards an integrated food safety surveillance system: exploring the potential for combining genomic and epidemiological metadata.” This demonstrated the limits to predicting illness from food chain data, possible roles for

machine learning and that food chain data will need to be richly sampled to be of maximum value. It is fully reported as a separate Food Standards Agency project.

The main themes identified as priorities and for more detailed discussion were:

1. Food chain sampling, both structured and opportunistic, with genome sequencing of isolates and methods to access broad samples of data.
2. Public Health-Regulatory Authority-Industry-Academic data sharing
3. Analytical food chain investigations, such as case-control studies comparing the food supply chain of venues associated with cases of illness and that for venues not associated (outline protocol in Annex 2).
4. The value of rapid access to comparable basic epidemiological data for human cases of infection alongside rapid genome sequence data of their isolates (case questionnaire proposal in Annex 2).

1. & 2. Extensive food chain sampling by industry, with increasing interest in genome sequencing of isolates, was identified by FSA. Examples of both good practice in data sharing allowing identification of the source of outbreaks and opportunistic linking of food source and human samples were raised. Issues in non-sharing of samples from literature and local experience as were discussed and contrasted with some US systems to support data sharing including by industry. Inter-operable databases rather than monolithic visions for data integration or standardisation were the preferred model. Developing the culture of sharing and application programming interface (API) type solutions is needed to support this. Costs of prolonged outbreaks leading to substantial potential for savings to industry as well as public health benefit were identified as motivators to support data sharing and linkage. Examples from the literature review of listeria outbreaks as part of case study five identified a large business going bankrupt after one prolonged outbreak and another with total costs estimated at 234 million Canadian dollars.

Structured food chain sampling was strongly supported as a funding priority whether that be as a regulatory environmental health activity or through research projects. This would complement the benefits offered by accessing industry sampling. Collaborative approaches with industry on sampling and data sharing would maximise benefits to industry and public health from this data providing industry with quality control data as well as the shared interest of minimising foodborne disease by the earlier identification of sources of outbreaks.

3. Difficulties in food chain tracing due to traceability data being limited to “one-up and one down” and to variable quality and formats of records for this data were identified. Software to support collation of the data such as used in Germany would still be limited by data quality and resources to extract this. Although larger companies are generally more sophisticated there are limits even in major supermarkets where specific batches of product cannot be traced reliably from receipts and loyalty card data. Extremely positive examples of engagement by industry bodies such as REPTA in CS4 supporting tracing of sources in the supply chain were identified as well as difficulties. The lack of efficient sharing such as a shared data environment between FSA and PHE was also identified as a barrier. FSA

proposing a standard approach including data format to industry might be useful for industry as well as public authorities.
Along with extensive descriptive studies of the food chain the idea of case – control studies for food chains was discussed – with e.g. case venues which were associated with illness among those consuming foods from them compared to control venues that are broadly similar in type but without links to identified cases of illness. Comparison of the food supply for each could identify possible sources or test specific hypotheses. Identification of the purpose of food chain studies in an outbreak will guide whether detailed studies of the whole food chain for all cases is needed, or sampling , and whether there is value in assembling "control food chain" data to allow quantitative inference. Studies such as undertaken following the outbreak contributing to case study 15, showing that food chain distributions map to phylogeny, suggest that GS could support sampling of cases for whom the food chain would be followed if there is a phylogenetic structure.

A shared working space for PHE and FSA with use of similar tools would increase efficiency and effectiveness. The increasing frequency and complexity of this work makes this important. There is also scope to collate information being gathered on stable parts of the food chain. For eggs for example, although the primary supply can change substantially over short timescales the main networks of importers, wholesalers and retailers or caterers are more stable and information could be built up across incidents and from routine work.

4. Very large numbers of small clusters of *Salmonella* infection at PHE cannot be investigated efficiently and effectively. Having basic case epidemiology data for each would be a major improvement compared to needing to gather variable data from local questionnaires. This was also identified as an issue in outbreaks 2, 13 and 15. The type of widespread outbreak identified by genome sequence data also changes which questions are most important. For example, where food purchases are made is increasingly important as well as what was eaten. The increasing normality of people entering questionnaire data on mobile devices in everyday life supports the development of patient facing electronic surveys as already undertaken in many outbreak investigations. Even if questionnaires need to be posted due to limited electronic contact details an option for on-line data entry by patients as well as a shared system with local PHE teams and local authorities is seen as feasible and of potentially great value. It could support efficient detection of outbreaks from genomic cluster data and prioritisation among these for further investigation. Difficulties in gaining full coverage are likely given some local arrangements but there is increasing positivity among local authority colleagues as well as PHE toward a shared approach.

Annex 1. Case study summaries

Outbreak Case Study CS1

Outbreak Title: Outbreak of E. coli O157 infection associated with two butcher's shops in Wingate and Billingham in July 2015

Organism: VTEC O157

Year(s): 2015

Source (if known): Bovine transmitted via various raw and cooked meats.

Reason for outbreak selection:

Detailed trace back of a local moderately complex food chain including FSA involvement

Strong inference on transmission route by synthesis of multiple forms of data including pathogen genomes.

Control selection methods (stratified random from GP lists).

Summary of outbreak

Fifteen cases of E. coli O157 infection with onset of symptoms from 4th to 21st July 2015. Ten hospitalised and seven cases developed HUS. A case-control study including ten cases found a strong statistically significant association between being a case and purchase and consumption of raw and ready-to-eat (RTE) food supplied by one butcher across multiple sites and supplied sites. GP controls used. GS on human, food and farm isolates showed links within the limits of discrimination of GS. Variation 0-3 between isolates implicated including after substantial time gap between cases and farm testing.

Laboratory methods used

GS

Available documents (and location)

Incident report. Redacted lessons learned. Phylogenetic tree – on study site

Additional Datasets

Reference data sets to describe on and between farm variation in VTEC would have supported more robust inference – not available.

People contacted

Deb Wilson (OCT chair) :

Russell Gorton (Epi lead) :

Learning points and follow up

GP controls effective and good science but difficult. Establishing pre-agreed methods to sample and access would be good.

Role of FSA in (local) outbreaks. Increasing frequency of involvement with food chain investigations for local as well as national outbreaks but role uncertain. Propose identification of agreed appropriate role and representation on OCTs.

Farm sampling timing and interpretation. Here thorough sampling done but two months after food exposures on two farms potentially implicated by trace back. Close match on GS but difficult to interpret given time and the lack of GS data sampled across farms and longitudinally. Here the combination of GS with strong corollary evidence through of a limited number of farms supplying the outbreak associated butchers, and reasonably close timing combined to identify this as a likely source but with some uncertainty and no capacity to use the GS for quantitative probabilistic inference.

GS similarity between cases (0-1 SNP), on some sampled food (0-2 SNP), and some sampled farm isolates (1-2 SNP) but how to interpret quantitatively?

Strength of GS linking of cases at the same time and in the same broad area supported persistence in identifying that an outlier butcher had received a part carcase from the main implicated butcher. The contrast of information from detailed exposure history on the one hand (showing different specific food exposures) with GS and descriptive epidemiology on the other (confirming a shared source) allowed to confirm that different foods from different sites shared the same source and gave

strong evidence for cross contamination at the butchers contributing to the outbreak.

Detection of the source and relatively detailed mapping of aspects of the transmission route was achieved in a small outbreak (13 lab confirmed cases and a case-control study with 10 cases) by the synthesis of detailed data in several dimensions.

Local practice to raise a single case with a possibly informative exposure (here consumption of product from a small local butcher). Extending this - two cases with strong GS similarity and sharing a plausible source on exposure histories would support substantial further investigation. Might this carry over to e.g. Salmonella in contrast to some considerations of only investigating e.g. clusters of at least five or at least 10 – to modify if there are shared high risk exposures among fewer cases at initial review. Maybe link this case study with the national Salmonella typhimurium where local (Thames Valley) outbreak and national signal detected in parallel.

Observed pattern of cases interpreted as meaning that a supermarket source is unlikely. This reasonable inference would be better supported if there is information on the extent to which supermarkets and other major retailers take in any local lines of food that might produce patterns of cases that might steer away identifying these as the source. This may be an area where FSA can gain background information in general or in specific outbreaks.

Control selection methods (stratified random from GP lists). This worked and may be more valid than many approaches to control selection – but was time consuming as regards both technical aspects and negotiation. Greater efficiency possible if there were national agreements in place.

Comments –other issues

Other linked cases from the same farm via other butchers not part of this outbreak but included on the tree.

Outbreak Case Study CS2

Outbreak Title: Local (Thames Valley) and national non-phage typable *Salmonella* Typhimurium outbreak

Organism: *Salmonella* Typhimurium MLVA profile 3-13-13-NA-0211.

Year(s): 2013

Source (if known): Ham

Reason for outbreak selection:

Differing overlapping detection methods of an outbreak in an instance where there were multiple localised clusters as well as a national picture.

Local and national traceback and recall.

Summary of outbreak

A national outbreak of *Salmonella* Typhimurium was detected in the UK from July-September 2013, while local outbreaks of the same strain of *S.* Typhimurium were detected in local areas in Thames Valley, Wales, Leeds, Cumbria and Stafford. Local investigations were undertaken in Thames Valley and Wales. These investigations both linked cases to suppliers of ham, which were ultimately both traced back to the same supplier in importing from Denmark.

Laboratory methods used

Phage-typing was attempted however the strain was not typable. MLVA typing was done and the strain profile was 3-13-13-NA-0211.

Available documents (and location)

Thames Valley outbreak report – available on Thames Valley and Colindale shared drives.

Wales outbreak report (provided by an OCT member).

National investigation meeting minutes / sitreps / traceback investigation – available on the Colindale shared drive.

Outbreak detection

Thames Valley

Given two cases temporospatially clustered cases of salmonellosis in Oxfordshire both were contacted to ascertain and compare exposure histories. One identified cooked ham from a local butchers shop PG as an exposure and the other having eaten at a cricket club event following which others were also ill. The cricket club lunch organiser had used the same local butcher PG. Follow up of the others with illness after the cricket club event identified another confirmed case and that this patient had a subtype matching an outbreak in Wales.

Traceback from PG identified a wholesaler TWL supplied from an importer and distributer DCUK. DCUK used a firm LBC to cook hams. Traceback investigations from Wales also identified supply from DCUK of ham cooked by LBC. TWL was not aware of the need to cascade the recall of its supply until advised by local environmental health.

Further cases were later identified, two of whom reported purchasing ham from butchers shops PB and CW which had also been supplied by TWL.

Wales

Three cases of suspected *Salmonella* Typhimurium that were temporally and geographically clustered in Gwynedd. A family link between one of these cases and two additional cases outside this area was identified, and an ICT was convened to examine possible exposures as food histories gathered suggested ham as a common exposure.

The two family members reported sharing a meal (ham sandwiches); the ham had been purchased from a local butchers shop. Further investigation revealed that the

majority of primary cases had consumed cooked meat from four separate butchers shops, where it was sliced for sale (not pre-packaged). No single type of meat was reported as consumed by all cases, however a supply chain investigation was able to link three of four butchers back to a distributor DCUK and manufacturer LBC where subsequent investigations identified hygiene problems.

National

A national OCT was convened after the investigations of the Welsh cases had been carried out and the manufacturer in the North West had been identified. At that point seven cases had been identified in the UK; four of these were part of the on-going Thames Valley investigation.

Learning points and follow-up

Local investigations prompted by very small case numbers (2 and 4) had identified sources before a national exceedance was tripped. Small numbers of temporospatially linked cases with exposure information can offer substantial insight into source, if all are part of the same outbreak. This may be similarly true for e.g. Temporo-Genomic clusters with relevant exposure information.

Triangulating the local food chain tracing from small local clusters confirmed a shared upstream source of infection.

Linking / coordinating information from local outbreaks or across genomic sub-clusters

Presumed recall from main supplier had not been effectively cascaded. Food chain networks identified in tracking can be used to reinforce recall implementation.

Outbreak Case Study CS3

Outbreak Title: Outbreak of Salmonella Agona phage type 40 associated with the Street Spice Festival, Newcastle upon Tyne, February / March 2013.

Organism: *Salmonella* Agona and PCR evidence for *Shigella* and Enteroaggregative *E. coli*

Year(s): 2013

Source (if known): Uncooked curry leaves.

Reason for outbreak selection:

Social media use for case finding and recruitment to a cohort study

Linkage of datasets allowing capture-recapture estimate of total outbreak size

PCR panel tests identifying Enteroaggregative *E. coli* (not usually tested) and *Shigella* (not identified on culture).

Follow up study conducted to assess outcome.

Summary of outbreak

- No more than 200 words

Twenty five lab confirmed Salmonella Agona and four cases of other individual *Salmonella* serotypes among 592 reported cases (cohort study, EH reports and lab test requests) and an estimated 926 (capture-recapture). PFGE indistinguishable Salmonella Agona from a food ingredient, curry leaves, in a food implicated by epidemiological association as the likely source of infection. Low rate of lab positivity led to multiplex PCR that identified other pathogens. Combination of food sampling and epidemiology identified curry leaf ingredient as plausible source of the outbreak (/many cases within the outbreak). Tracing via importer and follow up of other batches showed no evidence for ongoing risk other than a lack of labelling as not

ready to eat. A follow up study based on case reports of recurrent and long-lived symptoms confirmed relatively long symptom duration.

Laboratory methods used

Serophage typing and PFGE.

Available documents (and location)

Incident report. Study folders.

Additional Datasets

- Additional datasets used

Hospital records of laboratory samples (positive and negative) used to identify cases and support capture-recapture analysis.

People contacted

Kirsty Foster (OCT chair):

Russell Gorton (Epi lead):

Learning points and follow up

- Social media case finding and recruitment to a cohort study

Rapid large cohort study achieved with recruitment via twitter using festival organiser tweets as well as public health authority and case tweets and cascade. Tinyurl! Shortening of web addressed used on electronic communications and on posted letters.

- Linkage of datasets allowing capture-recapture estimate of total outbreak size

Three sources of data – reports to PH authorities, Cohort study, and laboratory test records used to conduct a capture –recapture study to scale the outbreak. Web

based software used to allow fuzzy logic matching of cases to identify those identified by one, two or three sources. Flagging to local GPs to clearly identify outbreak samples may have helped in leading to the outbreak being mentioned in laboratory sample clinical details, as may the cascading of information via social media to make people aware of the risk.

- PCR panel tests identifying Enteroaggregative *E. coli* (not usually tested) and *Shigella* (not identified on culture).

Substantial illness with well-timed but negative laboratory samples in many alongside clear identification of an outbreak strain in others led to testing available samples by multiplex PCR which identified both EAEC that would be missed in standard testing and *Shigella* that should have been identified with high sensitivity if viable. What should be the threshold for multiplex testing in outbreaks where it is not routine local clinical practice? Going forward which pathogens should be cultured if PCR positive if culture is not being undertaken. Interpretation of positive *Shigella* PCR and absence on culture,

- Follow up study conducted to assess outcome.

Some reports of recurrence and persisting symptoms led to a follow up study. Relatively high response rate (71%). Check if this was consistent across paper (with Tinyurl!) and electronic invitations. It appeared to be informative. Should it be a default part of large outbreaks?

- There was liaison between local authority and the FSA during the investigation. The outbreak provided good evidence for a particular ingredient as the source of infection raising questions on guidance for the use of curry leaves. Teleconferences were held with FSA when agreeing the final report and recommendations. Findings were also presented at the ACMSF.

Comments –other issues

Outbreak Case Study CS4

Outbreak Title: An outbreak of *Salmonella* Typhimurium DT 191a associated with reptile feeder mice in England and Wales, 2008-2009

Organism: Salmonella Typhimurium DT191a

Year(s): 2008 (but cases ongoing)

Source (if known): Reptile feeder mice

Reason for outbreak selection: Detailed trace back done to source. Source linked back to United States. Top down approach used.

Summary of outbreak

In 2008 a large number of cases of *Salmonella enterica* Typhimurium. Phage typing identified it to be a previously undefined strain, and was subsequently defined as DT191a [1]. *S.* Typhimurium is commonly associated with pork, beef and poultry. Cases accrued slowly at approximately 3 per week with an excess identified and investigated in December 2008 with 55 cases available for descriptive epidemiology [2]. As of April 2009 there had been 110 cases; 55% male, mean age 15 and median age 9. Cases were on going into 2015.

Standard food poisoning questionnaires were reviewed and a link with reptiles noted. A case-case study was carried out on laboratory confirmed cases using *S.* Enteritidis cases as controls. The study found that cases had 17 times the odds of having contact with reptiles than controls.

A list of pet shops who had supplied frozen mice was supplied to the Reptile and Exotic Pet Trade Association (REPTA) and the Veterinary Laboratory Agency (VLA). At the same time, REPTA supplied samples of mice representing all the major suppliers for testing. Through this testing, positive samples were traced back to a single supplier and then back to a producer in the USA.

Although this was reported as a reptile associated outbreak, it was associated with feeder mice and therefore some bird of prey handlers were affected as well. A second outbreak, this time of *Salmonella* Enteritidis P8 occurred in April 2015. GS and market research panel controls were used.

Laboratory methods used

Pulse-field gel electrophoresis (PFGE), variable-number tandem repeat (VNTR) typing and multilocus sequence typing (MLST) for phagetyping and genetypic identification.

Whole Genome Sequencing was used for the most recent outbreak.

Available documents (and location)

An investigation into the source of frozen mice used as reptile feed, associated with an outbreak of salmonellosis in the UK. – Unpublished Report (Personal correspondence, Chris Lane)

1. Peters T, Hopkins KL, Lane C, Nair S, Wain J, de Pinna E. Emergence and characterization of Salmonella enterica serovar Typhimurium phage type DT191a. J Clin Microbiol. 2010;48(9):3375-7
2. Harker KS, Lane C, De Pinna E, Adak GK. An outbreak of Salmonella Typhimurium DT191a associated with reptile feeder mice. *Epidemiology & Infection*. 2011;139(8):1254-61

National Outbreak of Salmonella Enteritidis PT8 (WGS t5: 1.5.159.280.280.280%), Final Outbreak Report, February 2016

Additional Datasets

Food chain tracing done by Reptile and Exotic Pet Trade Association (REPTA) using their databases that supported testing of mice across suppliers.

Controls used to identify population exposure to reptiles as pets. An alternative estimate could have been the Pet Food Manufacturers' Association pet population report estimating the proportion of households owning snakes, lizards and Tortoises at approximately 1% each from a survey sample of 2,000 – 4,

000 annually. Current year: http://www.pfma.org.uk/pet-population-2016, relevant past years http://www.pfma.org.uk/pet-population-2008-2012 .

People contacted

- Chris Lane PHE lead for *Salmonella* epidemiology
- Sanch Kanagarajah — Epi lead for 2015 reptile outbreak

Learning points and follow up

Development at 3 cases per week for almost 6 months shows how a small or slowly accumulating outbreak of a common serotype may have delayed identification. Genomic clustering would likely to foreshorten this.

Collaboration with the Reptile and Exotic Pet Trade Association (REPTA) in reptile food chain tracing using their data and contacts may be a model for future outbreaks.

Difficulties in obtaining case nominated controls undermined a planned case-control study. A case-case alternative used *S.* Enteritidis successfully. Case-case could be even more efficient if (i) GS and demographic information was used to select a set of control cases that would best represent the population generating the cases and (ii) exposure and contact data was available to the outbreak investigation team across these cases such as if there were a national standard questionnaire. Additionally information from the annual http://www.pfma.org.uk/pet-population-2016 might be considered as an alternative.

Outbreak Case Study CS5

Outbreak Title: Outbreak of listeriosis linked to crabmeat consumption in England

Organism: *Listeria monocytogenes*

Year(s): 2010-2013

Source (if known): Crabmeat

Reason for outbreak selection: Use of routine food microbiological sampling data

Summary of outbreak

In England, an outbreak of listeriosis, where crab meat was reported as the implicating food vehicle, was investigated in 2013. Three cases of listeriosis of a rare fluorescent amplified fragment length polymorphism (fAFLP) type were reported in London between 2011 and 2013. All cases had underlying medical conditions were aged between 55 and 79 years of age. Food and environmental samples from routine sampling of a crab meat producer in November 2012 revealed the same rare fALFP type. Food history of the cases were reviewed and revealed that two of the cases reported the consumption of fresh crab and dressed crab, bought from two outlets in London. The third case reported consumption of other type of seafood but not crab meat. Food traceability investigations confirmed that the producer supplied products regionally in the North West as well as selling at a Wholesale Market in London. This was supported by a shared fAFLP type accessible via routine sampling by local authorities of a crabmeat producer.

Laboratory methods used

- Molecular serogrouping
- Fluorescent Amplified Fragment Length Polymorphism (fAFLP)

Datasets used

- Human enhanced surveillance data (includes a food and clinical questionnaire). All cases isolates at this time were sent to the reference laboratories and typed using fAFLP.

- Food surveillance data. Local authority samples were sent to PHE for fALFP typing.

Available documents (and location)

Draft of the outbreak report – final version should be available soon (as of 24 June 2016).

Additional Datasets

Industry sampling. In this case, as in other outbreaks [1-3], identification of source was supported by identification of the same subtype in testing unrelated to the outbreak investigation. Routine typing of food chain isolates at the reference laboratory in PHE and similar arrangements in other such as Finland where all official test isolates are typed in a national laboratory [2] is usually restricted to regulatory samples and those undertaken as part of outbreak investigation. Including industry testing in comparable typing offers the possibility to detect outbreaks more quickly. Motivations from an industry perspective for this could be that:

(i) Large outbreaks detected late can be catastrophically expensive. For example one with an estimated Ca$234,000,000 in Canada [4] and permanent businesses closure after delayed identification [5]

(ii) As public health authority application of GS on all isolates supports more accurate epidemiology outbreaks are more likely to have an identified source eventually even without collaboration between authorities and industry.

(iii) As methods converge on GS such collaborations become much easier.

1. Gaulin C, Gravel G, Bekal S, Currie A, Ramsay D, Roy S. Challenges in listeriosis cluster and outbreak investigations, Province of Quebec, 1997-2011. Foodborne Pathogens & Disease. 2014;11(1):1-7.

2. Nakari UM, Rantala L, Pihlajasaari A, Toikkanen S, Johansson T, Hellsten C, et al. Investigation of increased listeriosis revealed two fishery production plants with persistent Listeria contamination in Finland in 2010. Epidemiology & Infection. 2014;142(11):2261-9.

3. Winter CH, Brockmann SO, Sonnentag SR, Schaupp T, Prager R, Hof H, et al. Prolonged hospital and community-based listeriosis outbreak caused by ready-to-eat scalded sausages. Journal of Hospital Infection. 2009;73(2):121-8.

4. Thomas MK, Vriezen R, Farber JM, Currie A, Schlech W, Fazil A. Economic Cost of a Listeria monocytogenes Outbreak in Canada, 2008. Foodborne Pathogens & Disease. 2015;12(12):966-71.

5. Gaul LK, Farag NH, Shim T, Kingsley MA, Silk BJ, Hyytia-Trees E. Hospital-acquired listeriosis outbreak caused by contaminated diced celery--Texas, 2010. Clinical Infectious Diseases. 2013;56(1):20-6.

People contacted

- Richard Elson

Learning points and follow up

- The source was traced in this small outbreak. This phenomenon of definitive source identification even from diffuse small and moderate sized outbreaks is common in the literature (see review below).
- Food sampling and discriminatory comparable subtyping is often a critical part of these investigations.
- Introduction of Whole Genome Sequencing for both clinical samples and environmental/food sample may allow cluster detection and linking between food produces and clinical cases more reliably and quickly. This would be facilitated by shared sampling across sectors.
- EU member states typically use pulsed-field gel electrophoresis (PFGE) as a method of molecular typing, making routine comparison of strains difficult. Harmonisation of molecular typing methods, particularly utilising whole genome sequencing techniques, is needed to provide a pan European perspective.
- The extent to which WGS implementation across food and human surveillance can support disease control is in part dependent on how much outbreak and source identification can be speeded up compared to current outbreak duration. These in turn depend on outbreak durations and detectable persistent sources of infection in the food chain. Interpretation of similarity and difference requires empirical data on genetic relatedness of the population of isolates in the food industry, environment and human cases. A rapid literature review was conducted to summarise evidence for these issues.

Rapid literature review – this is in Annex 2.

Outbreak Case Study CS6

Outbreak Title: *Cryptosporidium parvum* exceedance – November 2015

Organism: *Cryptosporidium parvum*

Year(s):2015

Source (if known): Unknown

Reason for outbreak selection: Use of market research panels as a sampling frame for population controls

Summary of outbreak

An excess of *Cryptosporidium spp* throughout November 2015 was detected in the UK in December 2015. Many of these cases were further typed as *C. parvum* IIdA24G1, and were predominantely adult females. The unusual distribution of this combined with wide geographical spread led to concern that the source of infection may have been a widely distributed food or drink item and as such warranted further investigation. A case-control study was conducted based on the results of hypothesis-generating questionnaires. A market research panel was used to recruit two population controls per case, employing frequency-matching in a 3:1 ratio by country (England:Wales and Scotland). A multivariable model identified young age, visiting a particular coffee shop chain and consuming lettuce products from a variety of retailers as significantly associated with illness. A trace back exercise is being conducted to determine if there were common suppliers between different retailers in conjunction with the FSA.

Laboratory methods used

GP60 subtyping. Samples were sent for further typing to the *Cryptosporidium* Reference Units in Swansea and Glasgow, where *C. parvum* GP60 type IIdA24G1 was identified. This typing was employed in the case definition.

Available documents (and location)

There is a draft outbreak report, saved in shared drives for the GI (CIDSC) team and the FES North West, and circulated to OCT members. Food chain traceback is still being conducted by Gordon Nichols.

An SOP has been created for the recruitment of controls through market research panels. This was written by Piers Mook and is saved on a shared storage drive in both the FES SEaL and GI (CIDSC) teams, as well as on the FES Sharepoint. Contact information for market research panel companies is included. The SOP that has been compiled was not used to investigate the outbreak of *C. parvum*.

Additional Datasets

The market research panel company Bilendi was approached to supply population controls for the study. Frequency matching by country was employed, however this was not done by any other criteria (eg. age and sex) due to concern of overmatching. Bilendi (previously Maximiles) has been used in previous outbreak investigations by PHE:

- *Salmonella* Mikawasima (2013) – A market research panel used to recruit a control group for a national *S.* Mikawasima outbreak alongside the main OCT investigation where controls were obtained within PHE staff. Panel controls were selected through region, age and sex frequency matching to cases through Maximiles (since rebranded to Bilendi). The main results of the case-control study using staff controls were similar to those of the panel control studies – both found a variety of chicken exposures were significantly associated with illness, however the panel control CC study found coriander and salad garnishes were associated with illness while the staff controls did not, and the staff CC study found use of proton pump inhibitor medicine (PPIs) as associated with illness while the panel controls did not. Peer-reviewed papers have been published detailing both studies [1,2]. Direct comparison between the two control groups (using both papers) shows markedly different rates of use demonstrating that these groups do not represent the exposure experience of a common population. Cost and time savings were noted as benefits of using market research panels to recruit controls, however there were concerns regarding biases inherent in differing control groups given the difference in the results of some of the exposures in

the case-control study. The study team acknowledged all methods of control selection are biased, however the bias that is potentially introduced by using market research panels is not fully understood. The SOP written by Piers Mook was informed by this study. The team advised that the representativeness of market research panel participants requires further assessment for the use in case-control studies.

- *Salmonella* Enteritidis associated with feeder mice (2015) – Investigation of a *Salmonella* Enteritidis outbreak with a specific WGS profile used market research panels to provide evidence that feeder mice (for reptiles) were the cause of the outbreak. Bilendi were used for this study and frequency matched by age (including households with children) and sex. It was estimated that cost and time savings were made by using panel controls rather than other sources, however Bilendi could not place quotas on different age/sex groups, and extra manual effort was required to ensure individual surveys closed when the quotas were reached. Given the rarity of the exposure in the general population, bias in the control group chosen was unlikely to distort the association of contact with feeder mice/reptiles and illness. Consequently no concerns regarding representativeness or bias as a result of control selection were highlighted as relevant to the conclusions drawn during this investigation.
- VTEC (2014) – Investigation of outbreak of specific MLVA strain of VTEC associated with bagged salad. Maximiles was used to recruit controls, which was estimated to be more timely and cheaper than had another source of controls been used. Limitations included no matching of the controls against cases – the age and sex profiles of the controls was not representative of cases, which may have introduced selection bias.
- A second market research panel company (PureProfiles) was considered, as they are able to provide frequency matching as well as IMD score. Both companies are available on as PHE suppliers (on the FARM system). For the *C. parvum* outbreak, Bilendi was selected as it was the only one that had been used previously by PHE.

Other sources of data might have been considered. As attending a chain of coffee shops was identified in the trawling questionnaire as a potential source, asking cases to provide receipts if they were retained could have provided validation or additional information to the results of the trawling questionnaire.

Contacting the chain and asking for additional information and cooperation could provide further avenues for investigation. Additional details from cases (eg. Date, time and branch of shop and approximately how much they spent and what they

ate) could have potentially been provided to the chain's headquarters to check their electronic records and get full details on food items ordered by cases.

People contacted

OCT lead: Gordon Nichols

All contact in arranging recruitment of controls was arranged with one contact at Bilendi (see attached SOP for contact information). All contact was made by Sanch Kanagarajah in the Gastrointestinal Infections Department at CIDSC.

Learning points and follow up

Market research panels have been used as controls in several outbreak investigations and logistical and validity issues summarised in an SOP based on the *S*. Mikawasima study. No other learning points have been documented for the *Salmonella* Enteritidis feeder mice outbreak, the VTEC outbreak or the *C. parvum* outbreak reports, however those involved have provided feedback regarding their experiences informally to colleagues as below. Propose prospective lessons learned in using market research panel controls be collated and distributed across the GI department in CIDSC and FES along with current SOP.

- The further experience gained confirms the conclusions that this is relatively fast and inexpensive but with risk of bias. No data confidentiality issues or practical problems were identified in using this approach alongside case data.
- In dealing with bias a balance exists in the need for representative controls with concerns for overmatching. Controls exhibited differing age and sex distribution from cases, complicate analysis and interpretation of the analytical results and particular attention to bias and inference is needed.
- Substantial oversampling may allow better representation of the age and sex characteristics of cases and be feasible given the low cost per questionnaire by this method. If more widely adopted PHE may also want to consider using propensity score approaches [3] to further avoid the effects of selection bias.

References

1. Freeman et al. 2016. “Association between use of proton pump inhibitors and non-typhoidal salmonellosis identified following investigation into an outbreak of Salmonella Mikawasima in the UK, 2013”. *Epidemiol Infect*. 144(5):968-75.
2. Mook et al. 2016. “Selection of population controls for a Salmonella case-control study in the UK using a market research panel and web-survey provides time and resource savings”. *Epidemiol Infect*. 144(6):1220-30.
3. Austin, P. C. (2011). An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behavioral Research., 46:399–424.

Outbreak Case Study CS7

Outbreak Title: A possible outbreak of hepatitis A associated with semidried tomatoes, England, July–November 2011

Organism: Hepatitis A

Year(s): 2011 (2009-2011)

Source (if known): Semi-dried tomatoes. Turkish origin.

Reason for outbreak selection:

Example of hepatitis A potential outbreak investigation.

Definitions of outbreak

Linking international, foodborne disease, and travel.

Summary of outbreak

A simultaneous identification of small clusters of genetically related cases in the UK (1) and the Netherlands (2) in autumn 2011. This included two cases in the UK that (i) had a shared genotype with outbreaks in Australia and the Netherlands in 2009 and 2010 respectively both of which were linked to semi-dried tomatoes and (ii) both reported eating semi-dried tomatoes.

The 2009 Australian outbreak had 144 cases of hepatitis A genotype 1B which were identical on a partial genome sequencing (of the VP1-2A region) with peaks in April – May and September – November with the first peak widespread and the second mainly in one Victoria. Two case-control studies showed strong association with semi-dried tomatoes and extensive food chain tracing included testing for hepatitis A that demonstrated the presence of viral RNA in 22 samples. No specific supply could be identified and a range of general public health measures implemented such as pasteurisation (3). The Netherlands outbreak in the first two months of 2010 comprised 13 (4, 5) indistinguishable from those in the Australian outbreak. A case-control study showed strong association with semi-dried tomatoes but no specific type of brand was identified. No hepatitis A RNA was detected among 81 food samples tested. The isolates were different to other indigenous samples in the Netherlands but similar to the virus identified in travellers who became ill having been to Turkey. In France an outbreak occurred in three areas from November 2009 to February 2010 overlapping the Australian and Netherlands outbreaks (6, 7). Each of the three could be traced to a local sandwich shop and the supply chain identified

a single imported from Turkey. Genome sequences clustered with returned travellers from Turkey and were similar to those in Australia and Holland, but differing by 2/300 and 3/430 nucleotides respectively in the areas where overlapping sequence was available for comparison.

The 2011 Netherlands outbreak ultimately comprised four cases matching the previous Netherlands and Australian outbreaks and concurrent UK cases as well as 3 that were more similar to the French outbreak of 2009/10.

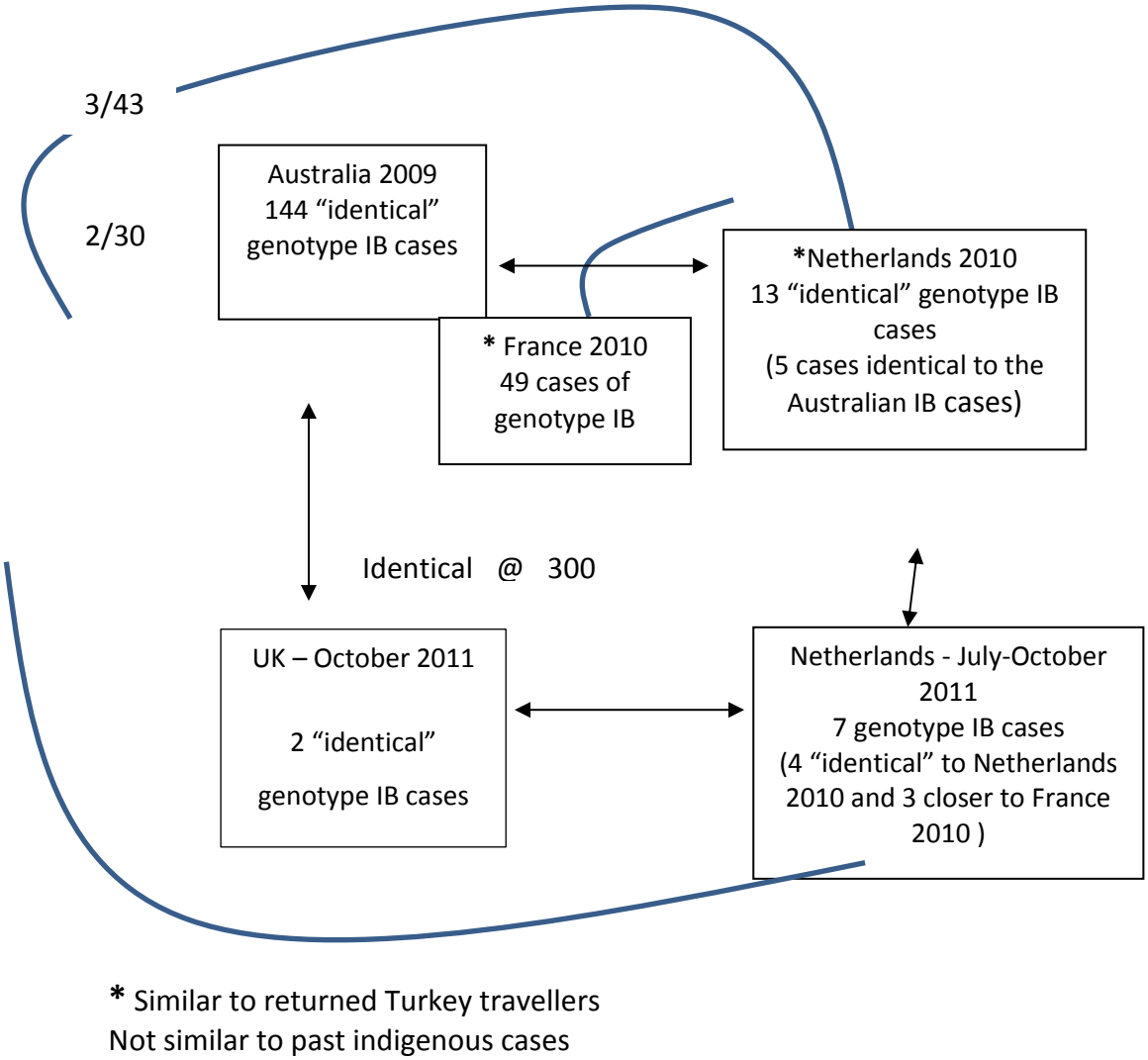


Figure. Relationship between semi-dried tomato outbreaks 2009-2011.

The later clusters were rapidly identified and extensively investigated but with no substantial outbreak developing in either the UK or the Netherlands.

This set of outbreaks and cases identify

(i) The difficulties in complex food chain tracing where even large indistinguishable strain outbreaks in Australia did not lead to the identification of a specific source.
(ii) The usefulness of traveller surveillance to indicate possible sources of infection.
(iii) The value of sharing complex incident data between countries.
(iv) Questions on relatedness. While the combination of genetic, temporal and spatial relatedness in Australia, supported by a strong exposure association in two case-control studies is strong evidence for epidemiological relatedness, and the French and earlier Netherlands outbreaks also formed epidemiologically coherent clusters the overall pattern of the same strain in Australia, the Netherlands and the UK over several years identifies the lack of specificity in the genotype as assayed.

A Turkish study on 76 isolates with information on an approximately 350 base fragment of the VP1-2A region showed that all were genotype 1B and highly similar to each other with little if any geographical structuring within the country. Only four polymorphic sites were found. The Turkish genotypes formed a defined cluster when compared to other 1B genotypes from outside Turkey (8). This would indicate that the diversity seen between the outbreaks, including between the French and other tomato associated outbreaks is similar to a sampling from Turkish isolates rather than there being any association of particular strains with tomatoes. A large US outbreak linked to pomegranates could be traced through mixed foods to identify this common ingredient and to an import from Turkey. Like the tomato associated outbreaks it showed strong clonality on the 315 base sequence studies with 60% of sharing an identical sequence. In contrast a study across the whole 7,500 base pair genome showed that in outbreaks with identical VP1 fragments there was a mean of approximately 7 varying nucleotides (9). These findings indicate the need for extended sequencing if outbreak structure is to be identified.

Laboratory methods used

Partial genome sequencing of VP1-2A region with precise target varying between countries but overlapping.

Available documents (and location)

Taken from published papers as per references.

Learning points and follow up

The difficulties in complex food chain tracing where even large indistinguishable strain outbreaks in Australia did not lead to the identification of a specific source.

This might be improved if outbreak substructure could be identified since the limited data available suggests that this genetic information is too coarse with 505 bases in the later UK outbreaks but only 300 to 420 overlapping comparable between France, Australian and Netherlands isolates.

The usefulness of traveller surveillance to indicate possible sources of infection.

The value of sharing complex incident data between countries.

References

1. Carvalho C, Thomas HL, Balogun K, Tedder R, Pebody R, Ramsay M, et al. A possible outbreak of hepatitis a associated with semidried tomatoes, England, July-November 2011. Eurosurveillance. 2012;17(6).
2. Fournet N, Baas D, van Pelt W, Swaan C, Ober H, Isken L, et al. Another possible food-borne outbreak of hepatitis a in the Netherlands indicated by two closely related molecular sequences, July to October 2011. Eurosurveillance. 2012;17(6).
3. Donnan EJ, Fielding JE, Gregory JE, Lalor K, Rowe S, Goldsmith P, et al. A multistate outbreak of hepatitis a associated with semidried tomatoes in Australia, 2009. Clinical Infectious Diseases. 2012;54(6):775-81.
4. Petrignani M, Verhoef L, van Hunen R, Swaan C, van Steenbergen J, Boxman I, et al. A possible foodborne outbreak of hepatitis A in the Netherlands, January-February 2010. Eurosurveillance. 2010;15(11):9-11.
5. Petrignani M, Harms M, Verhoef L, van Hunen R, Swaan C, van Steenbergen J, et al. Update: A food-borne outbreak of hepatitis a in the Netherlands related to semi-dried tomatoes in oil, January-February 2010. Eurosurveillance. 2010;15(20).
6. Gallot C, Grout L, Roque-Afonso AM, Couturier E, Carrillo-Santisteve P, Pouey J, et al. Hepatitis a associated with semidried tomatoes, France, 2010. Emerging Infectious Diseases. 2011;17(3):566-7.
7. Couturier E. Hepatitis A epidemics of dietary origin. On an epidemic linked to consumption of dried tomatoes, France, 2009-2010. [French]. Virologie. 2011;15(3):139-41.
8. Dinc B, Koyuncu D, Karatayli SC, Berk E, Karatayli E, Parlak M, et al. Molecular characterization of hepatitis A virus isolated from acute infections in Turkey. The Turkish journal of gastroenterology : the official journal of Turkish Society of Gastroenterology. 2012;23(6):714-9.
9. Vaughan G, Xia G, Forbi JC, Purdy MA, Rossi LM, Spradling PR, et al. Genetic relatedness among hepatitis A virus strains associated with food-borne outbreaks. PLoS One. 2013;8(11):e74546.

Outbreak Case Study CS8

Outbreak Title: Recurrent seasonal outbreak shiga-toxin producing *Escherichia coli* (STEC) O55:H7 in South West England, July 2014-September 2015.

Organism: non-O157 VTEC

Year(s): 2014-5

Source (if known): Probable local animal reservoir. Some person to person transmission.

Reason for outbreak selection:

Some substructure in WGS investigation.

Extensive searches for food distribution to match outbreak and environmental sources including environmental sampling.

Network analysis for data collation and visualisation.

Summary of outbreak

Thirty one infections of whom 21 symptomatic cases during the summer and autumn of 2015 (July to November) and 2015 (May to August). Eighteen of the cases had cultures and three based on O55 serology. Six asymptomatic contacts were culture positive and four by serology. Eleven of fifteen symptomatic children had HUS.

Twenty-eight of the infections were in six epidemiological clusters; five household community clusters of between two and five cases each and one nursery cluster of 12 cases. The remaining three primary cases, with onset July 2014, October 2014 and July 2015, had no identified epidemiological links to any other cases. Locations scattered across South Dorset. Cases were related by whole genome sequence analysis to earlier cases in Ireland. Ten of 11 (co-)primary cases had contact with pets especially dogs (6) and / or cats (5).

Detailed investigation included environmental and pet sampling, work with FSA and retailers to identify any potentially locally restricted food, and use of datasets to identify how case locations fitted with bird migrations and with farm animal movement between Ireland and Dorset. One positive pet cat and one positive sample from cat faeces near the home of a case. No other positive environmental samples despite extensive sampling including bootsocks etc.

Laboratory methods used

WGS

Serology for culture negative.

Environmental sampling.

Available documents (and location)

Draft paper (SE England PHE Centre)

Additional Datasets

- Additional datasets used

British Trust for Ornithology mapping tool – using both English and Irish cases – [Atkinson PW, Robinson, R.A., Clark, J.A., Miyar, T., Downie, I.S., du Feu, C.R., Fiedler, W., Fransson, T., Grantham, M.J., Gschweng, M., Spina, F., Crick, H.Q.P. . Migratory movements of waterfowl: a web-based mapping tool. EURING report to the EU Commission. (http://blx1.bto.org/ai-eu/). 2007.]

Animal movements data (APHA) for cattle sales between Ireland and Dorset.

Rainfall and river height data.

- Additional datasets considered & reasons why they weren't used

Reference data sets to describe variation in different sources of VTEC – not currently available.

People contacted

Noëleen McFarland (OCT chair):

Learning points and follow up

- Contribution of WGS to undertaking more extensive investigation

Here a rare (O55) subtype and severity drove extensive investigation. Tight clustering in WGS further contributed to this clarity that cases were part of a linked outbreak. It may also have contributed to decisions for environmental sampling: wide sampling generally has a risk of finding similar pathogens with difficulty in

inference on true epidemiological linkage but a highly discriminatory isolate characterisation system such as WGS can decrease this difficulty, as does the context of a rare outbreak strain.

- Environmental sampling and geographical linking of this outbreak to environmental datasets

Boot sock work, bird migration data, comparison of case location with farm animal movement, and extensive novel in a local outbreak setting.

- Broad epidemiological trawl

Work with retailers to identify local food lines was unusual and the use of network software was very useful and although neither of these were novel there was a substantial learning curve supporting an increase in skills and methods for PHE staff to exploit network technology, approaches to gain rapid local food chain data in a usable form technology, and the role of FSA in local outbreaks.

- Technical aspects of data linkage

Linking to Irish data possible as Ireland use PHE Colindale within a culture of collaboration. No formal linkage needed.

- Other learning

1. WGS similarity between nursery associated cases vs (0-1 SNP) other Dorset cases vs Irish cases up to a year earlier. Demonstrate size of population reservoir being sampled and bottlenecks driving observed diversity rather than recent evolution.
2. As well as SNP distance evidence for monophyly in the nursery sub-outbreak supported interpretation of this as a transmission system.
3. Information to control for e.g. pet contact and outdoor recreation not available to support inference on the extent to which cases had more of these characteristics than usual or not.

Outbreak Case Study CS9

Outbreak Title: National outbreak of *Salmonella enterica* Bovis-morbificans July 2015.

Organism: *Salmonella enterica* Bovis-morbificans

Year(s): 2015

Source (if known): Probable train station food outlet.

Reason for outbreak selection:

National distribution of cases from a local source allowing an extreme example of the cases moving to get their exposure rather than the exposure moving to explore possible links with mobile phone or app data.

Summary of outbreak

A moderate exceedance of *Salmonella Enterica* Bovis-morbificans was identified with 27 cases by 13 August 2015 compared to 16 over the same period one year earlier. Fifteen of the 27 had laboratory sample dates of 6 July or later and were later shown to be part of an outbreak compatible genomic cluster on WGS. Ultimately 19 cases in this cluster were identified.

Review of 17 questionnaires completed on single cases before the cluster was considered and further cases afterwards identified travel through London, with 8 cases reporting eating from a specific branch of a Mexican-style restaurant and a further 4 eating Mexican style food in the same area.

A coincidental environmental health inspection of the implicated premises after the onset dates of illness among cases but before they had been diagnosed identified some hygiene and temperature control issues that were rectified.

The study provided strong evidence for a shared source of infection at a premises but not a confirmed food.

Laboratory methods used

WGS

Environmental sampling (negative).

Available documents (and location)

Draft outbreak report in the study folders.

ECDC protocol for using mobile phone data in a legionella outbreak – in study folders as part of a proposed collaboration.

PHE conference poster from Emma Bennett – in study folders.

Additional Datasets

- Potential data sources which could have been used – mobile phone data

Mobile phone data

People contacted

John Harris (Incident lead at PHE national GI team)

Ian Hall and Emma Bennett working in PHE Porton Down on a research protocol to use mobile phone data in emergencies.

Julien Beaute ECDC working on project to use mobile phone data in legionellosis outbreaks.

Andy Tatem (University of Southampton) working with Flowminder project that uses mobile phone data in disaster response in collaboration with companies.

Hypothesis to test

Cases were clustered in time, or a rare serotype and subsequently demonstrated to be closely related on WGS but were apparently widely dispersed geographically. This

is consistent with either widely distributed exposure, cases travelling to a geographically restricted exposure, or a combination of the two.

Identifying which of these scenarios was occurring could focus the investigation allowing more defined exposure histories and other investigations into e.g. food sources. The current outbreak gives an opportunity to assess one extreme of this spectrum, where cases travelled substantial distances and likely overlap mainly around the point of shared exposure.

The testable hypothesis is that a shared location can be accurately identified from mobile phone records. This could be a descriptive identification e.g. based on anywhere shared within e.g. 500 metres during the two weeks before illness or could be in the form of an analytical study comparing to other individuals or self-controlled using times outside the incubation period. Limitations to self-controlled analysis are the risk of habitual travel patterns in some cases.

Outbreak Case Study CS10

Outbreak Title: *E.coli* O157 outbreak linked to the consumption of raw cows' drinking milk from a farm dairy. [Paper title: Whole genome sequencing improved case ascertainment in an outbreak of Shiga toxin-producing *Escherichia coli* O157 associated with raw drinking milk.]

Organism: E. Coli O157

Year(s): 2014-2015

Source (if known): Raw milk

Reason for outbreak selection: Geography of related cases along with distributed foodborne source; Early detection and intervention; Internet based distribution of food.

Summary of outbreak

In September-October 2014, an outbreak of *Escherichia coli* O157 (*E.coli* O157) including 7 primary and 2 secondary cases, was linked microbiologically and epidemiologically to the consumption of raw cows' drinking milk from a farm dairy in North Devon. The farm sells bottled raw cows' drinking milk both locally and through online sales, delivered by courier, as well as cheese and other dairy products. Five cases were identified from routine investigations. The additional cases were identified by whole genome sequencing and subsequently epidemiological links were made for the majority of these cases, three of them firm and one probable. The outbreak control team was able to use internet sales records from the farm to link two cases epidemiologically to the farm as having received delivered milk where no history of consumption could be obtained even on direct questioning after WGS linked the cases to the outbreak.

Due to internet sales the outbreak was geographically dispersed and making it potentially difficult to identify the cause. Nonetheless two cases in Wessex and the West Midlands reported on 29 and 30 September as having VTEC and a reported

exposure to the PHE team covering the producer led to an FSA led recall on 2 October. The food business operator chose not to inform all direct customers of the risk and recall and primary cases occurred as late as 19 October 2014.

An identical type of *E. coli* O157 (identical to some of the human cases and within 2 SNPs of all outbreak cases) was identified in three isolates from animal faeces collected from the farmyard at the farm's premises but not cultured from environmental or from milk samples from the premises or retrieved from customers or retail. One milk sample from retail was positive for the O157 and a shiga-toxin gene by PCR but culture negative.

Laboratory methods used

- Whole genome sequencing (WGS)
- Multiple-Locus Variable number tandem repeat Analysis (MLVA)

Datasets used

Public Health England Gastrointestinal Diseases Data Warehouse [giving access to laboratory results from the reference laboratory]

Enhanced surveillance questionnaires were administered to all cases [with additional interviews to explore possible causal exposures in more detail]

The farm online sales records were used to link cases epidemiologically to the farm. These were also used to identify bulk supply lines to avoid further distribution of the product.

Data from FSA and APHA databases collected as part of this investigation including dairy hygiene inspection results and animal faecal microbiology results.

Spatial analysis methods

Available documents (and location)

Publication[1] and an Outbreak Control Team Report[2]

Additional Datasets

People contacted

Charles Beck Field Epidemiology Service lead for the outbreak

Learning points and follow-up

Identification and intervention while a very small outbreak due to just two linked cases with a shared high risk exposure.

Internet based delivery a challenge in creating a dispersed outbreak but it also supported identification of cases as customers in collaboration with the business and could have provided targeted recall if the business had also collaborated on this.

Interplay between WGS cluster identification, detailed exposure history, and internet commerce records improved case ascertainment to ascertain the full extent of the lab confirmed outbreak.

Cases in the same year outside the outbreak cluster but with similar (5-25 SNP differences) isolates showed geographical links to SW England where the source of the outbreak was identified. This may motivate further work on mapping the geography of VTEC populations in human cases and possible sources.

References

1. Butcher H, Elson R, Chattaway MA, Featherstone CA, Willis C, Jorgensen F, *et al.* Whole genome sequencing improved case ascertainment in an outbreak of Shiga toxin-producing Escherichia coli O157 associated with raw drinking milk. *Epidemiology & Infection*. 2016;1-12.
2. Outbreak Control Team Report: E.coli O157 outbreak linked to the consumption of raw cows' drinking milk from a farm dairy. [Title amended to remove farm name] https://www.food.gov.uk/sites/default/files/phebartonfarmoctreport.pdf

Outbreak Case Study CS11

Outbreak Title: VTEC E.coli O157 outbreak associated with a restaurant in Belfast

Organism: VTEC O157 PT54

Year(s): 2012

Source (if known): Circumstantial evidence for chopped parsley + food handler

Reason for outbreak selection: Geographical – Similar strain of cases found in travel associated cases

Summary of outbreak

Four cases were identified through routine surveillance in August 2012 and the restaurant was followed up through environmental health. Four further cases were identified between 9 and 11 October, the restaurant closed (11th), and an outbreak investigation carried out during which confirmed cases rose to 141, 118 of which were primary or co-primary and had eaten at the restaurant between 2 and 11 October. Cases from August were PT8 (common), while those from October were PT54 (unusual in the UK) or PT31 (common). 68% of cases were female with a mean age of 31.3 years. Approximately one third attended hospital but there were no cases of HUS or deaths.

Circumstantial evidence for parsley as source of infection included association with dishes likely to have parsley in a case-control study, testing showing faecal organism contamination of parsley, processes showing inadequate washing of parsley, and a probably origin in the Mediterranean from where travellers had acquired similar strains of VTEC. Some association of cases with meals (likely to have been) prepared by one food handler with confirmed infection was identified on case-control analysis and may indicate some transmission by this route.

WGS demonstrated that August and October cases were closely related consistent with a shared reservoir but not with October cases as a subpopulation of the August cases as well as supporting the geographical association with travel associated cases.

Laboratory methods used

Phagetyping, Variable Number Tandem Repeat (VNTR) and Whole Genome Sequencing (WGS) microbiological methods were used. Multiple phagetypes and a high degree of VNTR variation were identified making WGS vital in understanding this outbreak.

Available documents (and location)

Outbreak report - http://www.publichealth.hscni.net/publications/report-outbreak-control-team-investigations-outbreak-e-coli-o157-associated-flicks-rest

Additional Datasets

- Travel associated PHE WGS data and Republic of Ireland PFGE data

People contacted

- Tim Dallman

Learning points and follow up

WGS was used to

(i) identify the cases most likely associated with the outbreak despite conflicting phage type information,
(ii) describe the pattern of relatedness among isolates consistent with a shared reservoir but not with the October isolates as descended (as a monophyletic subgroup) from the August outbreak and
(iii) identify possible geography of the reservoir.

Case study (see also CS8) identified monophyly as an epidemiologically relevant population genetic concept.

Identification of just two cases in the recurrent outbreak led to substantial intervention. Even with this a further 116 occurred who had already been exposed but this number might have been substantially higher with delay since those exposed up to this time appeared to be at substantial risk with no evidence for it having stopped before intervention.

Outbreak case study CS12

Outbreak title: Norovirus outbreaks in England linked to Oysters

Organism: Norovirus

Year: 2010

Source: Oysters

Reason for outbreak selection: To explore utility of routine sampling data and environmental data

Summary of outbreak(s)

In early 2010 several clusters of acute foodborne gastroenteritis were found. Outbreak investigation found Norovirus to be the causative organism, and oysters the source. Although most oysters were from a single producer in Ireland, other producers across England and Scotland were implicated. One of these oyster-related norovirus outbreaks occurred amongst diners at a restaurant in England in February 2010. Initially five dinners from a group of six became unwell with diarrhea and vomiting, with all those symptomatic consuming oysters. The restaurant identified further cases and the health protection unit undertook an analytical study. A case-control study undertaken amongst a sample of dinners (11 cases and 15 controls) confirmed the association with oyster consumption. Stool samples were obtained from all staff, and from two of the cases. Restaurant inspection and sampling of food preparation surfaces in the kitchen did not find anything significant. Oysters were also submitted for microbiological analysis. Diners had different genogroup infections oysters were heavily contaminated with bother genogroup I and genogroup II isolates of several different genotypes. This outbreak (1) was in the context of linked outbreaks across Europe at the time (2).

1. Baker K, Morris J, McCarthy N, Saldana L, Lowther J, Collinson A, et al. An outbreak of norovirus infection linked to oyster consumption at a UK restaurant, February 2010. Journal of Public Health. 2011;33(2):205-11.

2. Westrell T, Dusch V, Ethelberg S, Harris J, Hjertqvist M, Jourdan-da Silva N, et al. Norovirus outbreaks linked to oyster consumption in the United Kingdom, Norway, France, Sweden and Denmark, 2010. Euro Surveillance: Bulletin Europeen sur les Maladies Transmissibles = European Communicable Disease Bulletin. 2010;15(12):25.

Laboratory methods used

RT-PCR detection of norovirus.

Quantitative testing on oyster samples.

Genogrouping and genotyping

People / organisations contacted

Centre for Environment, Fisheries and Aquaculture Science (CEFAS)

Surfers against sewage (via Piers Mook)

Additional datasets

1) Surfers against sewage: provides real-time water quality data with alerts for when there are combined sewer overflows into the sea. This data is provided across 326 beaches in England and Wales.

2) CEFAS: Routine monthly sampling of pre-determined monitoring sites, including Shellfish harvesting farms for E-Coli. They previously did some work for the FSA that suggested the use of E-coli sampling data and water temperature data could be used to determine sites with higher likelihood of Norovirus contamination.

Background and interpretation of utility of data for predictive purposes

Oysters are grown in coastal waters and feed by filtering large amounts of water through their gills. Oysters can concentrate viruses up to 99 times compared to the surrounding water, and therefore in water contaminated with norovirus, the virus accumulates within the flesh and gut of oysters. Oyster beds may become contaminated with sewage discharges from sewer overflows (e.g. after floods). Sampling of oysters finds the same seasonal trend as that observed in humans (October to February in Europe) for norovirus. Oyster consumption is often higher between January and April, reflecting the season when oysters are considered to be of the best quality.

The presence of mixed strains of norovirus in oysters has previously been found to be associated with sewage contamination. Sewage contamination of Oyster beds can occur when there combined sewer overflows (CSOs). However, CSOs play an important function in preventing flooding inland after periods of heavy rains. The rainwater drains into the sewers resulting in them becoming full. The CSOs are needed to prevent the full sewers backing up and flowing back out. With a combined sewer overflow in place, the rain water, mixed with sewage, will rise inside the sewer and eventually enter a separate pipe which runs off the main sewer and flows into a river or sea. Under strict conditions, water companies are allowed to spill wastewater into the sea because it is accepted there is a finite capacity inside sewer pipes.

These variations in factors that might be associated with risk are frequent but the history of detected outbreaks are much more sporadic. This suggests that they are unlikely to have useful predictive value for oyster associated outbreaks of norovirus.

Lessons

Although several forms of sewage and faecal contamination monitoring data exist there is no evidence to support their use to predict risk.

No routine norovirus monitoring in oysters was identified that might be linked to outbreaks or sporadic cases.

Outbreak Case Study CS13

Outbreak Title: Outbreak of *Salmonella* Goldcoast, September-November 2013

Organism: *Salmonella* Goldcoast

Year(s): 2013

Source (if known): Whelks

Reason for outbreak selection:

Testing the impact of WGS on

Outbreak suspicion via WGS cluster

Refining trawling and food chain investigation using WGS

Refining analytical study using WGS

Controls by sequential digit dialling (20 of 27 female)

Possible large exported exposure

Summary of outbreak

An outbreak with approximately 37 cases based on WGS (6 of original 43 in non-WGS case definition excluded). First identified by three cases of unusual serotype in a small area in a short time. Followed by identification of more widespread cases in the following weeks although still with a concentration in the East of England and around Sept 2013. Early identification of whelks as a suspect food on trawling questionnaires. Initial food tracing and environmental investigations informed by this led to early recall. Microbiology from product and plant matched cases. Subsequent case-control study showed strong association with infection and whelk consumption.

Laboratory methods used

Serotyping, WGS.

Available documents (and location)

Outbreak report: study folders

Additional Datasets

- Food retailer and wholesaler dataset constructed by investigation.
- If international WGS it would have been possible to asses risk from exported product.

People contacted

- Tom Inns
- Epidemiological lead for outbreak with active participation in this case study including the additional analysis for learning points.

Learning points and follow up

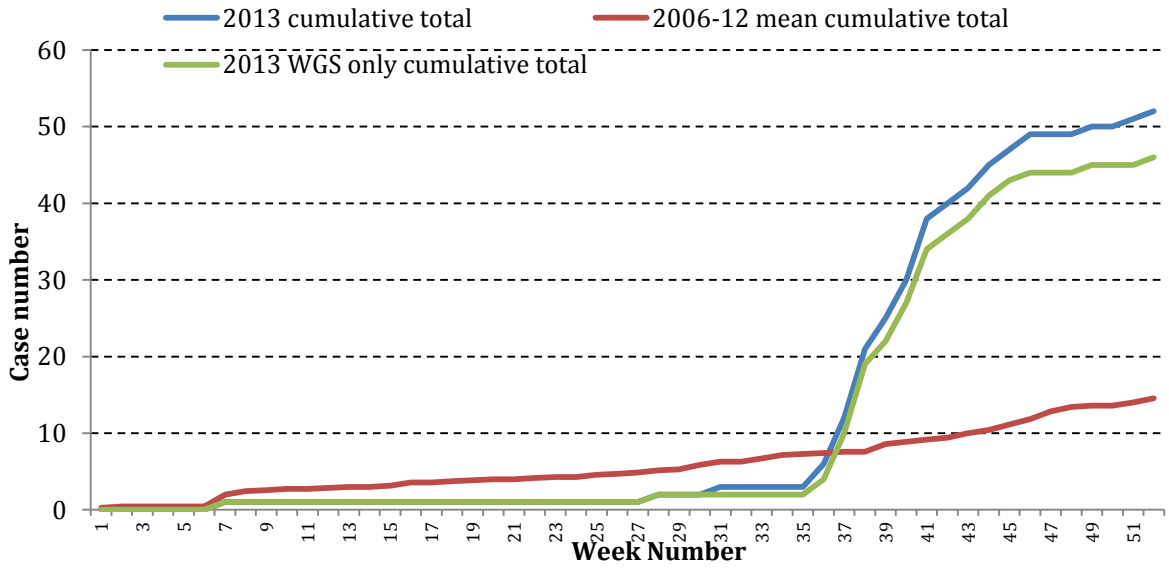
1. **Possible role for WGS in early outbreak detection**

On 12 September 2013, Anglia Health Protection Team (HPT) noted a cluster of 3 *Salmonella* Goldcoast cases with similar sample dates and resident within a 10km radius. During September 2013, 17 cases of *Salmonella enterica* serotype Goldcoast were reported in England. This number was greater than the annually expected number. Based on the previous isolations of S. Goldcoast from 2006-2012 it is possible to calculate a mean cumulative incidence based on date of report. On week 37 of 2013 the cumulative total of S. Goldcoast reports (12) exceeded the mean from 2006-12 (7.57). If non-WGS cluster cases are excluded, the cumulative total from 2013 (10) still exceeded the mean in the same week. Week 37 of 2013 is September 9 2013 to September 15, corresponding to the date that the first cluster was originally identified.

If we can assume that a highly related monophyletic cluster implies an outbreak then the WGS signal at week 28 (plus sequencing time) has 2 cases, at week 36 has 4 and

then the rise crossed the threshold on 37 and strongly by 38. In parallel week at week 37 identification of the 3 temporospatially linked cases prompted investigation.

Figure 1: Comparison of cumulative totals of *S*. Goldcoast laboratory reports, by date of report



2. Trawling questionnaires

To generate a hypothesis as to the cause of this outbreak, a detailed trawling questionnaire (used for Salmonella outbreaks) was undertaken for the first 10 cases. The food which most stood out from these questionnaires was whelks; 5 (50%) cases reported whelk consumption, which was higher than would be expected. Two of the ten cases who completed the trawl were outside of the WGS cluster. Neither of these cases reported whelk consumption. Therefore if the trawl included only those in the WGS cluster, 5/8 (62.5%) this would have been a stronger signal of an association with whelks. There were not sufficient cases excluded by WGS to allow a case-case contrast between WGS related and unrelated.

3. Possible role for WGS in refining analytical epidemiology

Conducting the case-control study with an altered case definition to exclude cases outside the WGS cluster increased the strength of association by approx. 20%. Given the strong association in this outbreak this did not change inference and would be more relevant in settings where inference is less robust on the case-control study with outcome misclassification. The full numbers for this analysis are given below.

The objective of the analytical study was to test the hypothesis that whelk consumption was associated with this outbreak of S. Goldcoast infection. To test this hypothesis, we undertook an unmatched case-control study, recruiting controls on a prospective basis, using a systematic digit dialling method. Cases and controls were interviewed by telephone. A structured and pre-tested questionnaire was used to collect information.

In total, 22 cases were eligible for the case–control study: of these, 20 cases were included (two cases declined to participate). A total of 27 controls were included. In a univariable analysis, cases were significantly more likely than controls to have consumed whelks, cockles, lettuce, fish and peppers (Table). No effect modification was detected. In the final multivariable model, when adjusted for sex, cases were significantly more likely to have consumed whelks (OR: 109; 95% CI: 7.7–1,539).

Food exposure	Case n=20[b]	Control n=27	Univariable analysis			Multivariable analysis		
	Number exposed	Number exposed	OR	95% CI	p value	OR	95% CI	p value
Whelks	16	1	104	9.54–4,517	<0.001	109	7.7–1,539	0.001
Cockles	5	0	NC	2.1–∞	0.010	–	–	–
Lettuce	10	22	0.2	0.05–0.99	0.030	–	–	–
Fish	7	18	0.3	0.07–1.06	0.420	–	–	–
Peppers	5	15	0.3	0.06–1.09	0.044	–	–	–

Only one of the 20 cases included in the analytical study would have been excluded based on WGS data. When excluding this case and re-running the analysis, the strength of association between case status and whelks is increased. The univariable association increases from an Odds Ratio of 104 (95% CI = 9.5-4517) to an Odds Ratio

of 138 (95% CI = 11.4-6019). The association adjusted for sex increases from OR 109 (95% CI=7.7-1539) to OR 126 (95% CI = 9.2-1745).

4. Possible impact of WGS on food chain tracing.

There were 43 cases, 37 had a food history. Of these, 24 had a report of eating whelks, the whelk supplier could be traced for 22. In total, 21 cases report consumption of whelks which were supplied by one factory.

The WGS data would have decreased the need for some extensive case interviews to support food chain tracing. None of the 6 cases outside of the WGS cluster were known to have been exposed to whelks. However having interviewed it would not have saved further work since no non-WGS case reported eating whelks. In outbreaks with a less unusual exposure WGS would be expected to both decrease workload and refine inference. It would have altered the strength of inference with 21 of 37 cases (57%) linked to one supplier as stronger evidence than 21 of 43 cases (48%).

5. Identifying controls

Systematic digit dialling from case land line exchange numbers recruited a sufficient number of controls in the desired timeframe and with the available resources. The age profile of controls was equivalent to cases, both being older, but there was a female bias in controls (20 of 27), compared to 67% of outbreak cases being male.

Outbreak Case Study CS14

Outbreak Title: Prospective use of Whole Genome Sequencing (WGS) detected a multi-country outbreak of *Salmonella* Enteritidis

Organism: *Salmonella* Enteritidis

Year(s): 2015 (2014)

Source (if known): Chicken eggs

Reason for outbreak selection:

Genomic investigation linked to food chain via extensive food chain investigation

"Blind" genomic search for related cases on large database

Detection by WGS

Summary of outbreak

Genomic cluster detected May 2015 with 29 cases (23 as PT59) and 136 cases ultimately identified and a further 18 in Spain. 17 of 21 UK cases reported travel to the Iberian peninsula, mainly Spain. 92 of 136 UK cases were PT59 and 44 not. Contemporaneous cases (28 WGS) clustered with the UK cases. WGS database searching identified UK (17) and US (3) cases in 2014 that clustered with these 2015 cases. Travel histories for two of the US cases were identified and reported travel to Europe.

Of 87 cases with food histories 67 reported recent eating out at 41 venues. Food tracing focussed on eggs (based on past similar outbreaks) from identified points of sale. Egg supply available for 31 of which 23 could be linked to a supply network that could account for 43 of the 87 cases with food histories. Isolates closely related on the food chain were more similar genomically – supported by formal testing. An outbreak cluster isolate was isolated from one egg containing food (omelette) in Spain but not from other egg (1962 sampled) and environmental samples.

Laboratory methods used

Phage typing, WGS

Available documents (and location)

Additional Datasets

Genome Trakr at NCBI allowed identification of related US cases.

US case exposure data via US teams

Investigators identify the need for comparable food industry and veterinary data on isolates.

People contacted

- Tom Inns, PHE

Learning points and follow up

High proportion not clustered on phenotype – non-specificity would slow detection and limit investigation.

Genomically diverse outbreak with food chain clustering mapped to WGS tree – especially as regards tips of the food chain network and with cases over a year apart. Travel histories support shared exposure for cases from US and UK in 2014 suggesting specificity of WGS.

Apparent specificity of WGS signal here supports investigators conclusions that industry and veterinary monitoring might have supported fuller linking to source.

Variable questionnaires across the UK.

Incomplete food chain despite extensive investigation due to food history gaps and imperfect recall and only 31 of 41 premises being able to provide traceability despite EC Regulation 178/2002.

Almost half (43/87) of cases with information could be associated with an egg supply network. Study designs to test this as a form of quantitative evidence are possible but time consuming. This prompts inclusion of a n outline framework for such a study in the Annex 2. This is adapted from work in the then HPA when considering how the UK would respond to an outbreak equivalent to the German *E. coli* O104.

Outbreak Case Study CS15

Outbreak Title: Multicountry *Salmonella* Enteritidis linked to eggs from German and Czech plants of an egg producer.

Organism: *Salmonella* Enteritidis

Year(s): 2014

Source (if known): Eggs

Reason for outbreak selection:

International scope

Extensive food chain traceback locally, nationally and internationally

Joint analysis of food chain and pathogen phylogeny

Summary of outbreak

An outbreak with 287 UK identified cases comprising several restaurant associated and one hospital associated outbreaks along with sporadic cases occurred in the context of over 350 cases in total across the UK, Austria, France, Germany and Luxembourg and communication between countries on the Rapid Alert System for Food and Feed (RASFF) and the Epidemic Intelligence Information System (EPIS). Extensive food traceback identified a single company supplying through 6 wholesalers could be linked to 198 of the UK cases for whom a food history was available to support food chain tracing.

Sampling from source plants supported the food chain history even thought there were no culture positive eggs identified within the UK. Differential subpopulations at the company supply plants could also be mapped onto different parts of the outbreak including showing a relationship with time. This was consistent with known interventions to control the outbreak. Variation in the phylogeny could be

substantially and statistically explained by the clade a clade structure and distance along the supply chain measured by numbers of nodes. Phylogenetic analysis identified that the variation was not compatible with evolution during the outbreak but was with sampling from existing diversity.

Comparison to reference populations identifies some similar isolates 2 years earlier without exposure history available to explain the relationship.

Laboratory methods used

MLVA and WGS.

Available documents (and location)

From published paper (1) and in press manuscript (2)

1. Inns T, Lane C, Peters T, Dallman T, Chatt C, McFarland N, Crook P, Bishop T, Edge J, Hawker J, Elson R, Neal K, Adak GK, Cleary P; Outbreak Control Team.A multi-country *Salmonella* Enteritidis phage type 14b outbreak associated with eggs from a German producer: 'near real-time' application of whole genome sequencing and food chain investigations, United Kingdom, May to September 2014.Euro Surveill. 2015 Apr 23;20(16). pii: 21098.
2. Dallman T et al Phylogenetic structure of European Salmonella Enteritidis outbreak correlated with national and international egg distribution network.

Additional Datasets

- Additional datasets used
- Additional datasets considers & reasons why they weren't used
- Potential data sources which could have been used
- Are these feasible?

People contacted

Tom Inns, PHE.

Learning points and follow up

Extensive food chain tracing was required. Strategic sampling of food chain tracing might be a consideration in very large outbreaks.

The pattern of relationship of isolates (tree topology) as well as closeness of isolates was informative in mapping out the transmission of the outbreak and monitoring the impact of control measures, with this investigation linking these datasets in formal analysis.

Phylogeny suggests that the source held this population of strains for some several years but only a small number of UK 2012 cases were identified. The phylogeny supports the transmission system as one of sampling from one or more somewhat diverse reservoirs rather than diversification during the outbreak.

EPIS and RASSF as ways of sharing rich data internationally. Collaboration through a national incident supported sharing across UK incidents. Should a system be in place to harmonise and archive data from local and national investigations during and after outbreaks. Similarly for the different aspects of food chain investigation.

Annex 2. Additional work

Literature review for Case study 5

Rapid review: Evidence on duration of listeriosis outbreaks and the pattern and persistence of *Listeria monocytogenes* in food plants.

The application of WGS to detect outbreaks and linking human surveillance data to regulatory and industry databases will be better supported if

1. Long lived outbreaks occur where early detection has the opportunity to give public health benefit.

2. Persistent sources exist in the food chain.

A single rapid review of English language peer reviewed literature was undertaken to give evidence on these two areas since they are not independent with long lived outbreaks a potential source of evidence for persistent sources.

The search terms used were as below.

1	▶ "listeri*".m_titl.	12060	Advanced
2	▶ food*.mp.	483092	Advanced
3	▶ 1 and 2	4217	Advanced
4	▶ outbreak*.tw.	66294	Advanced
5	▶ Food-Processing Industry/	4490	Advanced
6	▶ (Process* or produc* or manufactur* or industr* or plant*).m_titl.	656106	Advanced
7	▶ 3 and 4	466	Advanced
8	▶ limit 7 to english language	435	Advanced
9	▶ 5 or 6	658586	Advanced
10	▶ 1 and 2 and 9	744	Advanced
11	▶ limit 10 to english language	721	Advanced
12	▶ 11 not 8	655	Advanced

These produced 435 papers on outbreaks (#8) and 655 on *Listeria* in food processing (#12). Title and abstract scans were followed by full text reviews of papers (49 retrieved to 24 June 2016).

Results

Persistence is summarised in the table which identifies whether a study identified i) diversity in a single processing plant, ii) persistence of a single type in a plant and iii) sharing of subytpyes between plants in studies on two or more plants. Twenty-eight papers reported information on persistence and sharing of subtypes with all twenty-six reporting on persistence finding evidence for persistent subtypes, mostly using discriminatory methods, in particular PFGE.

Table. Evidence for diversity (two or more types in a plant), persistence (repeated isolation of a single subtype), and sharing of subtypes between plants (+ yes, - no, and ? if not reported)

Reference	Type of plants	Diversity	Persistence	Sharing
Rivoal et al 2013	Egg	+	+	+
Meloni et al 2014	Sausage	+	+	-
Malley et al 2013	Smoked fish	+	+	-
Ortiz et al 2010	Pig (one plant)	+	+	N/A
Cruz et al 2014	Seafood	?	+	?
Chen et al 2014	Mushrooms	+	+	+
Lambertz et al 2013	Meat and Fish	?	+	?
Nakari et al 2014	Fish	+	+	?
Bolocan et al 2014	Meat (one plant)	+	+	N/A
Camargo et al 2015	Beef (one plant)	+	+	N/A
CDC 2011	Pig (one plant)	?	+	N/A
Currie et al 2015	Meat (one plant)	+	+	N/A
Di Ciccio et al 2012	Smoked fish	+	+	N/A
Da Silva et al 2016	Butchers shop	+	?	?
Fagerlund et al 2016	Salmon and Chicken plants	+	+	+
Ferreira et al 2011	Sausage	+	+	+
Fox et al 2011	Cheese	+	+	-
Fox et al 2015	Chicken	+	+	N/A
Gaulin et al 2012	Cheese	?	+	?
Holch et al 2013	Fish	?	?	+
Jackson et al 2011	Cheese	?	+	?
Lamden et al 2013	Ox tongue	?	+	N/A
Leong et al 2014	48 plants	+	+	+
Magalhaes et al 2015	Cheese	?	+	N/A
Morganti 2016	Parma ham (genomic)	+	+	+
Schroder et al 2013	Cheese	?	+	?
Stessl et al 2014	Cheese	+	+	?
Vongkamjan et al 2013	Smoked fish	+	+	N/A

In five egg plants in France studied using PFGE across 4 seasons in a year showed (i) diversity within L. monocytogenes populations in each plant at individual time points and over time, (ii) substantial within plant persistence of PFGE types and (iii) some sharing of types across plants (1). A study in 10 sausage processing plants found positive results in 5 with (i) more than one PFGE type in each, (ii) multiple same type isolates within all, and (iii) no sharing across plants (2). In a study of two smoked fish processing plants using ribotyping there was (i) more than one type in each, (ii) persistence over time in both particularly on non-food surface samples, and (iii) no sharing across plants (3). A single pig processing plan study over 3 years with 541 isolates identified 29 PFGE types with 1 comprising 45% of isolates and 3 comprising 73%. The dominant type was mainly identified in the manufacturing area and manufactured products consistent with persistence in the plant (4). Persistence of the same PFGE type has been identified in seafood processing environments across two years (5). L. monocytogenes present in 3 of 4 mushroom plants was present across the plant with higher rates of positive in product and positive scraping machinery than samples at earlier stages of production. There was (i) more than one RAPD type in two plants, (ii) multiple same type isolates in all three, and (iii) sharing across two plants (6). A Swedish study across human, food and processing plan isolates inferred persistence in plants from ongoing presence in food (7). Finnish investigations into two outbreaks identified (i) within plant diversity and (ii) persistence in cured fish from two plants each sharing the one of the outbreak PFGE types (8). A yearlong follow up in a new plant meat processing (beef, pork and poultry) plant showed (i) diversity (4) and (ii) persistence of PFGE strains with one dominant strain across sampling 2,3, and 4 of the four times across the year (9). A 13 visit study in a beef processing plant showed both (i) diversity and (ii) persistence of PFGE subtypes. Persistence was inferred in a pork plant linked to outbreak cases from Feb to June and where product sampled on August was positive for the same strain (10). Persistence in a cooked meat plan was inferred from a lengthy outbreak and diversity identified on sampling (11). A six year follow up of one salmon smoking plant identified (i) substantial diversity over time and within year and (ii) persistent types interpreting the pattern as evidence for final product contamination being more related to persistent types than raw salmon contamination (12). A butcher's shop study showed diversity within shop but did not use a discriminatory typing methods to assess persistence or sharing across the three shops studied (13). Application of whole genome sequencing to strains of ST-8 isolates demonstrated only 3 SNPs between two isolates 10 years apart from a Norwegian salmon processing plant, and a Danish salmon processing plant isolate from 5 and 15 years earlier than these was separated by 8 and 9 SNPs respectively. Two isolates from different poultry plants sampled 2 years apart were separated by 2 SNPs showing

persistence and sharing of very similar types (14). A Danish study similarly found two highly similar isolates on WGS in separate fish processing plants six years apart with one analysis identifying no SNPs (15). A study across Portuguese sausage plants identified diversity and persistence within plan and sharing between (16). A cheese making plants study identified diversity and persistence within plants but no sharing among plants (17). A detailed PFGE study in a single poultry meat processing plant identified both diversity and persistence (18). Diced celery at a plant implicated in a 7 month long outbreak was positive at inspection for the outbreak strain demonstrating persistent presence (19). Concurrent and retrospective typing at a plant making sliced ox-tongue linked to a single case demonstrated persistent presence of the disease causing strain (20). A study with 2-monthly sampling in 48 varied plants in Ireland showed extensive diversity within plant, persistence, and sharing across plants of PFGE types (21). An extensive study with WGS in Parma ham manufacturing demonstrated persistent clones, within plant diversity and clones shared across plants. WGS was able to split PFGE clusters including one across two producers (22). A 34 case outbreak over 9 months was evidence for ongoing transmission from the source plant producing Quargel cheese (23). A survey across Austrian, Czech and Irish cheese plants using two enzyme PFGE identified diversity within plant, substantial persistence with some evidence for persistence of 8 and 11 years in plants with historical samples. Similar isolates were shared across plants but it is unclear if identical ones were identified across plants (24). Several persistent strains, identical on ribotyping and PFGE, some across several years, are described in a sample from a smoked fish plant used for evaluating the impact of listeriaphages on persistent strains (25).

One review on mechanisms of persistence argues for environmental rather than pathogen factors as the driver for persistence (26) while another identifies that strain resistance to quaternary ammonium compounds and capacity to form biofilms may contribute to persistence (27). Airborne levels in ruminant and pig slaughter plants are low and identified as unlikely sources of spread (28) and reported as absent in mushroom plants (6). The minimal variation in two WGS studies across time and distance raises questions on the amount what can be inferred from identical or highly similar genomes since some types appear persistent and possibly widespread (14, 15). One study on the microbiome in drains showed some species that were associated with the presence or absence of *L. monocytogenes* with in vitro tests confirming biological effects (18).

Three papers reviewing more than one outbreak for other reasons than duration, and therefore potentially unbiased as an estimate of durations, reported a total of 21 outbreak durations with 8 of these of 20 weeks or longer and seven four week or shorter. The distribution of all 29 outbreaks with identified durations is given in Figure 1.

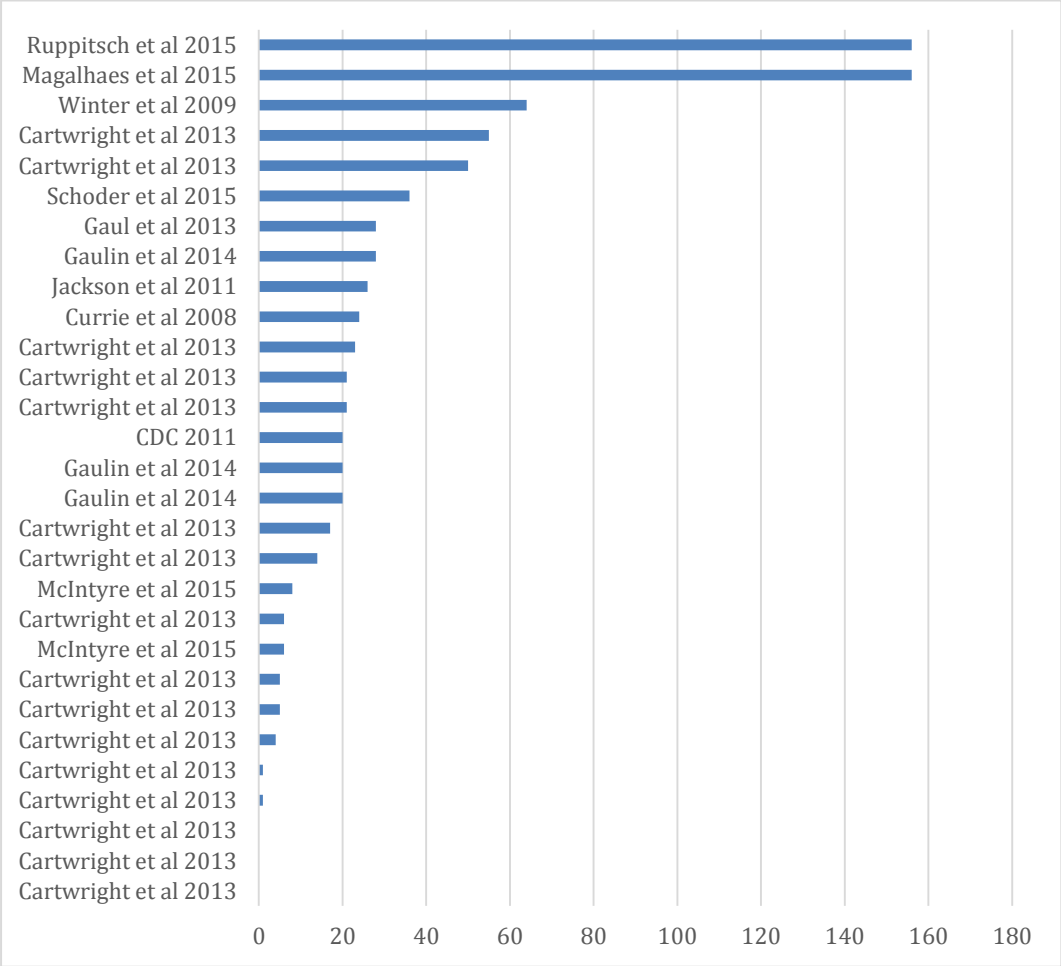


Figure 1. Outbreak duration in weeks among published outbreaks identified that reported duration.

Two Canadian cheese associated outbreaks comprising 48 and 86 cases had durations of 7 and 8 weeks with identification of source and product recall in weeks 4 and 6 of the outbreak respectively (29). In a review of US outbreaks from 1998-2008 median duration was 42 days with a range 1-389 and 7 persisting for 100 days or longer among the 16 of 26 for which data was available. The seven long duration outbreaks were associated with cheese (2 outbreaks), delicatessen meat (2 outbreaks), milk, frankfurter sausage, and sprouts (30). A German outbreak identified in October 2007 with two cases in the same hospital with identical PFGE types was identified to have started with cases in July and August 2006, to comprise

9 confirmed cases with identical PFGE and source identification to a sausage plant. A routine sample sausage sample had also tested positive in July 2006 (31). One of two outbreaks identified in Finland persisted (timing not precise) as a PFGE subtype but without a confirmed shared source (8). Detected cases associated with a pork jelly product sold through two shops occurred over 5 months with contaminated product identified on month 7 of the outbreak but with only 8 confirmed cases in total three of which had occurred by month 3 (10). In a Canadian cooked meat outbreak 43 cases occurred in the 12 weeks before recall was started and 14 in the 12 weeks after. Four cases occurred in the first 2 weeks. All isolates were clustered on one or other of two PFGE enzymes but there were minority populations with a one band difference on one enzyme for both enzymes and the shared source supports epidemiological and evolutionary relatedness (11). The authors identify that more rapid investigation was identified as a lesson. This might be supported by more robust genomic clustering supporting investigation at an earlier stage. Cost estimates were Ca$234 million (32). A seven month outbreak linked to diced celery was identified with just four cases in month 2 but limited case accrual contributed to difficulties in detecting the source which was supported by an identical PFGE pattern on the produce (19). A review of Canadian outbreaks identified 6 with investigations (33) including one described above (11). One lasting 5 months with 17 cases was identified following unrelated testing at a cheese facility showing the outbreak strain. Re-interview established recalled exposure among 9 of the 17 cases. One lasted 7 months with 24 cases and stopped without intervention. The otherwise unusual outbreak strain was identified in a range of products in four super markets on unrelated testing but no association to these products among the cases even on re-interview. A 38 case outbreak over 5 months was partly controlled by recall at weeks 9 and 11 (34). A further outbreak of unspecified duration had a source identified by testing at a cheese making facility in response to customer complaints. Details for other outbreaks are incomplete and the tables and text appear to mix up outbreaks I and J (33). A low level outbreak with 8 cases over 6 months stopped after a case-control study identified a form of cheese as associated and testing identified the outbreak strain (35). A three year outbreak with 30 cases was traced to several possible foods by case questionnaires and confirmed to cheese production by traceback and testing (36). A German outbreak with 68 PFGE cases, 38 of which have been confirmed as clustered on WGS within 5 SNPs of a central genotype on a core genome comparison, with 9 isolates in the central group and 18 within 1 SNP, has persisted for 3 years without identification of a source despite extensive epidemiological investigation and food testing (37). An outbreak across Austria, Germany and the Czech republic from Quargel cheese comprised 34 cases over approximately 9 months (23).

References

1. Rivoal K, Fablet A, Courtillon C, Bougeard S, Chemaly M, Protais J. Detection of Listeria spp. in liquid egg products and in the egg breaking plants environment and tracking of Listeria monocytogenes by PFGE. International Journal of Food Microbiology. 2013;166(1):109-16.
2. Meloni D, Consolati SG, Mazza R, Mureddu A, Fois F, Piras F, et al. Presence and molecular characterization of the major serovars of Listeria monocytogenes in ten Sardinian fermented sausage processing plants. Meat Science. 2014;97(4):443-50.
3. Malley TJ, Stasiewicz MJ, Grohn YT, Roof S, Warchocki S, Nightingale K, et al. Implementation of statistical tools to support identification and management of persistent Listeria monocytogenes contamination in smoked fish processing plants. Journal of Food Protection. 2013;76(5):796-811.
4. Ortiz S, Lopez V, Villatoro D, Lopez P, Davila JC, Martinez-Suarez JN. A 3-year surveillance of the genetic diversity and persistence of Listeria monocytogenes in an Iberian pig slaughterhouse and processing plant. Foodborne Pathogens & Disease. 2010;7(10):1177-84.
5. Cruz CD, Pitman AR, Harrow SA, Fletcher GC. Listeria monocytogenes associated with New Zealand seafood production and clinical cases: unique sequence types, truncated InlA, and attenuated invasiveness. Applied & Environmental Microbiology. 2014;80(4):1489-97.
6. Chen M, Wu Q, Zhang J, Guo W, Wu S, Yang X. Prevalence and contamination patterns of Listeria monocytogenes in Flammulina velutipes plants. Foodborne Pathogens & Disease. 2014;11(8):620-7.
7. Lambertz ST, Ivarsson S, Lopez-Valladares G, Sidstedt M, Lindqvist R. Subtyping of Listeria monocytogenes isolates recovered from retail ready-to-eat foods, processing plants and listeriosis patients in Sweden 2010. International Journal of Food Microbiology. 2013;166(1):186-92.
8. Nakari UM, Rantala L, Pihlajasaari A, Toikkanen S, Johansson T, Hellsten C, et al. Investigation of increased listeriosis revealed two fishery production plants with persistent Listeria contamination in Finland in 2010. Epidemiology & Infection. 2014;142(11):2261-9.
9. Bolocan AS, Nicolau AI, Alvarez-Ordonez A, Borda D, Oniciuc EA, Stessl B, et al. Dynamics of Listeria monocytogenes colonisation in a newly-opened meat processing facility. Meat Science. 2016;113:26-34.
10. Centers for Disease C, Prevention. Outbreak of invasive listeriosis associated with the consumption of hog head cheese--Louisiana, 2010. MMWR - Morbidity & Mortality Weekly Report. 2011;60(13):401-5.
11. Currie A, Farber JM, Nadon C, Sharma D, Whitfield Y, Gaulin C, et al. Multi-Province Listeriosis Outbreak Linked to Contaminated Deli Meat Consumed Primarily in Institutional Settings, Canada, 2008. Foodborne Pathogens & Disease. 2015;12(8):645-52.
12. Di Ciccio P, Meloni D, Festino AR, Conter M, Zanardi E, Ghidini S, et al. Longitudinal study on the sources of Listeria monocytogenes contamination in cold-smoked salmon and its processing environment in Italy. International Journal of Food Microbiology. 2012;158(1):79-84.
13. da Silva DA, Dias MR, Cossi MV, de Castilho NP, Camargo AC, Nero LA. Hygiene and Safety in the Meat Processing Environment from Butcher Shops: Microbiological Contamination and Listeria monocytogenes. Journal of Food Protection. 2016;79(4):628-34.
14. Fagerlund A, Langsrud S, Schirmer BC, Moretro T, Heir E. Genome Analysis of Listeria monocytogenes Sequence Type 8 Strains Persisting in Salmon and Poultry Processing Environments and Comparison with Related Strains. PLoS ONE [Electronic Resource]. 2016;11(3):e0151117.

15. Holch A, Webb K, Lukjancenko O, Ussery D, Rosenthal BM, Gram L. Genome sequencing identifies two nearly unchanged strains of persistent Listeria monocytogenes isolated at two different fish processing plants sampled 6 years apart. Applied & Environmental Microbiology. 2013;79(9):2944-51.
16. Ferreira V, Barbosa J, Stasiewicz M, Vongkamjan K, Moreno Switt A, Hogg T, et al. Diverse geno- and phenotypes of persistent Listeria monocytogenes isolates from fermented meat sausage production facilities in Portugal. Applied & Environmental Microbiology. 2011;77(8):2701-15.
17. Fox E, Hunt K, O'Brien M, Jordan K. Listeria monocytogenes in Irish Farmhouse cheese processing environments. International Journal of Food Microbiology. 2011;145 Suppl 1:S39-45.
18. Fox EM, Wall PG, Fanning S. Control of Listeria species food safety at a poultry food production facility. Food microbiology. 2015;51:81-6.
19. Gaul LK, Farag NH, Shim T, Kingsley MA, Silk BJ, Hyytia-Trees E. Hospital-acquired listeriosis outbreak caused by contaminated diced celery--Texas, 2010. Clinical Infectious Diseases. 2013;56(1):20-6.
20. Lamden KH, Fox AJ, Amar CF, Little CL. A case of foodborne listeriosis linked to a contaminated food-production process. Journal of Medical Microbiology. 2013;62(Pt 10):1614-6.
21. Leong D, Alvarez-Ordonez A, Jordan K. Monitoring occurrence and persistence of Listeria monocytogenes in foods and food processing environments in the Republic of Ireland. Frontiers in Microbiology. 2014;5:436.
22. Morganti M, Scaltriti E, Cozzolino P, Bolzoni L, Casadei G, Pierantoni M, et al. Processing-Dependent and Clonal Contamination Patterns of Listeria monocytogenes in the Cured Ham Food Chain Revealed by Genetic Analysis. Applied & Environmental Microbiology. 2016;82(3):822-31.
23. Schoder D, Skandamis P, Wagner M. Assessing in-house monitoring efficiency by tracing contamination rates in cheese lots recalled during an outbreak of listeriosis in Austria. International Journal of Food Microbiology. 2013;167(3):353-8.
24. Stessl B, Fricker M, Fox E, Karpiskova R, Demnerova K, Jordan K, et al. Collaborative survey on the colonization of different types of cheese-processing facilities with Listeria monocytogenes. Foodborne Pathogens & Disease. 2014;11(1):8-14.
25. Vongkamjan K, Roof S, Stasiewicz MJ, Wiedmann M. Persistent Listeria monocytogenes subtypes isolated from a smoked fish processing facility included both phage susceptible and resistant isolates. Food microbiology. 2013;35(1):38-48.
26. Carpentier B, Cerf O. Review--Persistence of Listeria monocytogenes in food industry equipment and premises. International Journal of Food Microbiology. 2011;145(1):1-8.
27. Martinez-Suarez JV, Ortiz S, Lopez-Alonso V. Potential Impact of the Resistance to Quaternary Ammonium Disinfectants on the Persistence of Listeria monocytogenes in Food Processing Environments. Frontiers in Microbiology. 2016;7:638.
28. Okraszewska-Lasica W, Bolton DJ, Sheridan JJ, McDowell DA. Airborne Salmonella and Listeria associated with Irish commercial beef, sheep and pig plants. Meat Science. 2014;97(2):255-61.
29. McIntyre L, Wilcott L, Naus M. Listeriosis outbreaks in British Columbia, Canada, caused by soft ripened cheese contaminated from environmental sources. BioMed Research International. 2015;2015:131623.
30. Cartwright EJ, Jackson KA, Johnson SD, Graves LM, Silk BJ, Mahon BE. Listeriosis outbreaks and associated food vehicles, United States, 1998-2008. Emerging Infectious Diseases. 2013;19(1):1-9; quiz 184.

31. Winter CH, Brockmann SO, Sonnentag SR, Schaupp T, Prager R, Hof H, et al. Prolonged hospital and community-based listeriosis outbreak caused by ready-to-eat scalded sausages. Journal of Hospital Infection. 2009;73(2):121-8.
32. Thomas MK, Vriezen R, Farber JM, Currie A, Schlech W, Fazil A. Economic Cost of a Listeria monocytogenes Outbreak in Canada, 2008. Foodborne Pathogens & Disease. 2015;12(12):966-71.
33. Gaulin C, Gravel G, Bekal S, Currie A, Ramsay D, Roy S. Challenges in listeriosis cluster and outbreak investigations, Province of Quebec, 1997-2011. Foodborne Pathogens & Disease. 2014;11(1):1-7.
34. Gaulin C, Ramsay D, Bekal S. Widespread listeriosis outbreak attributable to pasteurized cheese, which led to extensive cross-contamination affecting cheese retailers, Quebec, Canada, 2008. Journal of Food Protection. 2012;75(1):71-8.
35. Jackson KA, Biggerstaff M, Tobin-D'Angelo M, Sweat D, Klos R, Nosari J, et al. Multistate outbreak of Listeria monocytogenes associated with Mexican-style cheese made from pasteurized milk among pregnant, Hispanic women. Journal of Food Protection. 2011;74(6):949-53.
36. Magalhaes R, Almeida G, Ferreira V, Santos I, Silva J, Mendes MM, et al. Cheese-related listeriosis outbreak, Portugal, March 2009 to February 2012. Euro Surveillance: Bulletin Europeen sur les Maladies Transmissibles = European Communicable Disease Bulletin. 2015;20(17).
37. Ruppitsch W, Prager R, Halbedel S, Hyden P, Pietzka A, Huhulescu S, et al. Ongoing outbreak of invasive listeriosis, Germany, 2012 to 2015. Euro Surveillance: Bulletin Europeen sur les Maladies Transmissibles = European Communicable Disease Bulletin. 2015;20(50).

Template protocol for an analytical trace back study in a foodborne outbreak

This protocol is a framework to apply case-control methods in trace back for foodborne outbreaks comparing venues associated with case with control venues of a similar type but not associated with outbreak cases. Inclusion here is motivated by Case study 14.

1. Title

Trace back study to test association of foods and (food) supply chain factors with a food venue giving rise to cases with the XX outbreak.

2. Background

Literature review of trace back studies in template protocol as general background to be supplemented with summary of the specific incident being studied.

3. Aim

To contribute quantitative evidence to the identification of the source and path of infection.

4. Objectives

To identify a set of "Case" venues or food supply chain nodes, being those where people who were later ill report to have eaten or obtained food.

To identify comparison venues or nodes with no reported illness among their customers.

To compare supply chain factors between these two sets of venues or nodes and identify any supply chain factor associated with customer illness.

Hypothesis: These studies should be hypothesis driven where possible. If there are hypotheses for the source of the infection based on other information (e.g. a particular food type is suspected) trace back may seek data on the source of this food rather than being more general. A general approach comparing the range of suppliers (and foods) to each venue is very labour intensive but may be necessary if the outcome of infection is severe requiring an urgent response and bringing equivalent resources.

5. Methods

5.1 Study design

Case-control

5.2 Definition of cases

A case is any venue or supplier selling or serving food with one or more clinical case exposed to the venue. In a particular outbreak it may be restricted to e.g. restaurants.
Controls are then selected from equivalent venues without associated cases. In geographically structured outbreaks controls may be frequency matched by area.

5.3 Study population
The population of food suppliers (e.g. restaurants) some of which have been identified as exposures by cases in the outbreak.

5.4 Study period
To be defined relating to when cases occurred, when they were exposed to the venues under study, and the rate of turnover of the types of product being traced.

5.5 Exposure variables
These will vary with the type of premises and hypotheses being tested. Examples are:
Suppliers of any (food) product with full lists of suppliers from each premises.
All types of food served at the premises.
Specific batches of food that can be identified (batch numbers in records etc)

5.8 Data management
- 5.8.1 Data collection
- 5.8.2 Data cleaning

5.9 Data analysis plan
Analysis using odds ratios and associated statistical measures of confidence in line with usual case-control analysis and modified for any particular issues of the sampling frame or data in the study.

6. Role of team members
Each major agency needs to sign up to this as joint work, identify the shared aim and objectives, and agree roles, relationships and communications. An agreed template study protocol between PHE and FSA with roles and resources flowing from that protocol is recommended.

8. Infrastructure
Databases of types of venues and pre-existing food chain databases would make these types of studies more efficient. A single office for a multiagency team or linked offices between PHE, FSA and lead Environmental Health Department would be likely to optimize efficiency.

9. Variant designs

Case-crossover designs within case-linked premises would be more efficient as not involving separate control premises, and effective where the hypothesis being tested is of a food that is intermittently ordered and used. It might be informative for regularly used foods if batch numbers were accurately recorded for times of receipt and consumption but otherwise would be limited for these foods.

Proposal for routine on-line data capture from patients

Inclusion of this rationale for a standard online surveillance questionnaire for patients with laboratory confirmed GI infections, with content varied by pathogen, is motivated by case studies 4 and 14.

1. **Aim**

To make available a nationally coordinated on-line GI case questionnaire database supporting:

direct input by cases or professionals of GI surveillance questionnaire data;
linkage of this with isolate characterisation data including genome sequences;
national access for PHE epidemiological analysis and outbreak detection including by automated searching for related cases;
local access for PHE health protection teams and for local authorities;
compatibility with other PHE surveillance and case management software avoiding duplicate data entry.

2. **Background and current position**

Local authority environmental health and Public Health England health protection teams gather data from some individual patients with gastroenteritis, typically using structured questionnaires. Nationally standardised questionnaires are largely used for some such as shiga-toxin producing *E. coli* (STEC) but for most pathogens there is variation as to whether a questionnaire is undertaken, questionnaire content, and how the data is collated and reviewed. Local questionnaires are stored on paper only in most local authorities while in others questionnaires are entered onto databases held at local authority or PHE centre level or loaded onto PHE case management software (HPZone). In the national STEC system data is typically collected on paper, then entered onto the PHE case management system and then retyped onto the national surveillance system database by the Colindale GI team. This national STEC database is accessible to health protection teams to query.

The use of questionnaire data to detect outbreaks varies, including local review of paper questionnaires and data on databases in ad hoc or systematic ways. Local and national teams review data from STEC cases to identify links. Standardised or machine learning based approaches to identify links are not implemented at present. Questionnaire data can be sought and queried when an excess of isolates with a shared subtype is found, such as by exceedances identified on the national algorithm to identify an excess of a pathogen subtype over a short time period. This may involve gathering available local paper questionnaires from local authorities and health protection teams and administration of

questionnaires when none was done initially. Given this workload and the associated delays it is undertaken sparingly.

3. Motivation and drivers for standardised online data capture of case data now

Genomics

With genome sequencing becoming standard for *Salmonellae*, small clusters of highly related isolates are increasingly recognised, often across geographically dispersed populations. In retrospective analysis geographically restricted clusters of shared genotypes were typically from already detected outbreaks while geographically dispersed and travel related clusters had generally not been detected as an outbreak unless very large. Pathogen genome sequencing is therefore identifying gaps due to the absence of collation of intelligence across the country. An accessible shared database would facilitate both review of completed standard questionnaires and further contact with these cases where appropriate. In the absence of this the large workload prohibits following up many of these outbreaks with associated risk of delay for those that are the early stage of large outbreaks.

Technological developments

Online data capture tools and population access to appropriate IT and skills for on line data entry, have increased rapidly.

Expectations and comparisons

Other countries are beginning to deploy on-line survey software from national public health centres to collect case information (e.g. Belgium and Switzerland currently using local implementations of the open-access software LimeSurvey and the Robert Koch Institute in Germany has identified this as their preferred technology). Local health protection teams increasingly use on line data capture for studies in identified outbreaks of communicable disease. Efficient data capture, collation, and analysis are increasingly standard in other areas of society. Societal expectations are likely to include that lessons of problems arising from the failure to share intelligence across police forces would also be taken on board in other public services.

Environmental health officer attitudes

In contrast to a history of many local authorities favouring very local arrangements for designing questionnaires the Chartered Institute of Environmental Health (CIEH) has increasingly identified calls from local environmental health towards more standardised approaches and has confirmed support for this from a national perspective (Personal communication – Jenny Morris CIEH to Noel McCarthy)

Efficiency

Some environmental health departments are stopping the collection of information from individual cases due to resource restraints showing the need for a streamlined process at reduced cost.

4. Barriers and caveats

Where local surveillance and public health case management is working well locally adoption of a national system will be resisted unless the national system supports what is already in place.

Voluntary uptake is likely to be partial, at least initially, even with a system that supports current local practices.

Although potentially more efficient overall establishing an effective system requires central resource initially and widespread collaboration.

Patient confidentiality and data protection must be robust even where this may limit functionality or efficiency.

Since at present patient postal details are available but not electronic means of contact almost all questionnaires will likely need to be posted as a paper version with a simple link identified to an online option and sign in instructions allowing access to the correct record.

5. This work in the context of GI epidemiology and information systems in PHE

Work to develop a system supporting online data capture for GI infections needs to coordinate both with overall PHE information strategies, including developments in case management software, and with a broader project to oversee the development of standard national questionnaire content. This latter aspect is being developed in collaboration with the Colindale GI team and FES and PHE Centre GI leads and has identified four work areas as below in addition to the online data capture. The current proposal is focusing only on the development or implementation of software in PHE since this is the area that needs resource outside the specialist GI and field epidemiology teams and has interactions with wider information strategies.

Other work areas being progressed in parallel with this proposal:

1. Engage stakeholders in (the need for) a standard and nationally coordinated approach
2. Develop agreed content for core and additional modules for GI questionnaires
3. Develop routine standard analysis of collected data
4. Make targeted information accessible to patients completing questionnaires

6. Infrastructure and ongoing capacity requirements

Based in part on the Robert Koch Institute options appraisal for an on-line questionnaire the following characteristics are identified as important for the system supporting online

administration of questionnaires and feedback of information to national and local public health teams.

Easy to use for the patient / local health professional online.

Functional on a wide range of devices including mobile phones, online and off line.

Save and resume available and easy.

Print / PDF-Export function can generate a printable version suitable for use.

Full range of question types, branching structure, plausibility checks etc.

Supports modular construction of questionnaires including conditional modules.

Links to relevant on-line information pages possible.

High standard of data protection and security (encryption, local implementation on PHE servers, can be anonymised even for individual invitations).

Free open licence software.

Capable of being pre-filled with available data (e.g. patient details from PHE lab data or HPZone).

Back-end that supports extraction, linkage to other PHE data (e.g. isolate characterisation), and serving of data to national and local users or transfer into case management software.

Secure access to data for national and local teams including local authority staff.

Automated posting or capacity to email questionnaires for posting e.g. PDFs to relevant local authority with their local authority branding included and a personalised link or link with log in details to support identification of the correct record.

Limited need for ongoing maintenance other than content editing.

7. Rationale for initial postal contact to patients

At present NHS Patient Demographic Service has approximately 0.5% coverage for emails from GP systems and these are only accessible by individual patient look up. Mobile phone coverage is higher but % and accuracy uncertain as well as confidentiality issues in linking by text to PDS derived numbers. Although in the longer term centrally delivered electronic contact may be possible any initial roll out will therefore require the local posting of questionnaires, in line with usual current practice, but with a steer on the cover letter toward the online version and details allowing log into identify the correct case.

The PHE system should therefore allow local authority teams to generate letters for each case in their area with a questionnaire link. There may be efficiency in centralising this on behalf of local authorities and emailing a pdf for each case to the local authority to send on.

Linking case clusters identified on GS with mobile telephony

The value of identifying whether dispersed cases share travel to a particular location is identified in CS9. Mobile telephone data is likely to be very effective in identifying such links especially if "ping data" tracing location is used and not only during calls, or if other tracking data such as from Google applications has high coverage among cases. If this data proved very accurate it could be extended to identifying more subtle shared exposures such as presence in particular chains of restaurants or retailers motivating further exploration.

The testable hypothesis is that a shared location can be accurately identified from mobile phone records. This could be a descriptive identification e.g. based on anywhere shared within e.g. 500 metres during the 2 weeks before illness or could be in the form of an analytical study comparing to other individuals or self-controlled using times outside the incubation period. Limitations to self-controlled analysis are the risk of habitual travel patterns in some cases.

Necessary conditions and steps to using phone based data

Sufficient cases have a phone with them when travelling

Mobile phone companies (or an alternative source such as Google) hold data of sufficient detail and for long enough to allow tracing after cases have accrued and been identified as clustered by WGS (e.g. 2 months).

Access to phone data can be negotiated with the phone companies for individual cases.

An agreed consent procedure can be put in place for cases.

Exploration of these issues to date:

1. An approved protocol has been identified that was developed in relation to a Scottish legionellosis outbreak. This protocol development led by ECDC staff including gaining ethical approval but with no available mobile phone data by the time that this was implemented. This study still had public health implications when the protocol was being developed. Emergency response

studies are also being developed in PHE. Compared to these any research study to pilot this on an example of a concluded outbreak such as the current case study would need additional ethical review. The potential uses in PHE can be divided into three related categories each with different ethical and other approval issues:

   a. The acquisition and handling of individual mobile phone records for a research study (such as if possible to use CS9 as proof of principle).
   b. The acquisition and handling of individual mobile phone records in an emergency response.
   c. The acquisition and handling of mobile phone records in routine public health investigation (such as if a further instance arises with infrastructure in place to allow implementation to support public health investigation).

2. Mobile phone companies are able to provide “ping data” on where a phone was even if not in use but (i) apparently only retain this for a matter of weeks and (ii) treat this as a commercial opportunity with costs often handled by intermediaries. Coverage of Google app tracking not yet explored.
3. Approach to PHE Caldicott Guardian has given an opinion that standard clinical or research consent is needed as appropriate.
4. If this works consider other examples e.g. known food distribution related outbreaks to test whether these show lack of geographical overlap or clusters of geographical overlap e.g. from location specific sub-outbreaks within the overall outbreak and for some where exposure included a mix of travel to the main site of exposure and some distribution of product to assess accuracy in identifying these two types of exposure patterns for each case.

Appendix. Search strategy to identify outbreaks

Search History (8 searches) *(close)*

	# ▲	Searches	Results
☐	1	exp United Kingdom/	377300
☐	2	(United kingdom or UK or Britain or England or Scotland or Wales).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]	640174
☐	3	1 or 2	640174
☐	4	outbreak*.mp.	78075
☐	5	3 and 4	4280
☐	6	(food* or gastro*).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]	1352512
☐	7	5 and 6	921
☐	**8**	**limit 7 to yr="2010 -Current"**	**202**

Remove Selected | Save Selected | Combine selections with: And Or

Basic Search | Find Citation | Search Tools | Search Fields | **Advanced Search** | Multi-Field Searc

1 Resource selected | Hide | Change

Embase Classic+Embase 1947 to 2016 Week 08

Downloaded **60** records into Endnote on title screening.

# ▲	Searches		Results
1	exp United Kingdom/	►	318951
2	(United kingdom or UK or Britain or England or Scotland or Wales).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]	►	387960
3	1 or 2	►	404953
4	outbreak*.mp.	►	99387
5	3 and 4	►	6914
6	(food* or gastro*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]	►	870807
7	5 and 6	►	1148
8	**limit 7 to yr="2010 -Current"**	►	**186**

Re-ran same search in Ovid MEDLINE(R) 1946 to February Week 3 2016, Ovid MEDLINE(R) Daily Update February 24, 2016, Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations February 24, 2016 – retrieved 186 **–**

Downloaded **57** records into Endnote on title screening.

388 retrieved in Embase and Medline and 117 downloaded

From the combined **117** records from both databases, **71** remained after removing duplicates

Table. Outbreaks selected following abstract review.

Adak, G. K., et al. (2015). "A multi-country Salmonella Enteritidis phage type 14b outbreak associated with eggs from a German producer: 'Near real-time' application of whole genome sequencing and food chain investigations, United Kingdom, may to september 2014." Eurosurveillance 20(16).
Baker, K., et al. (2011). "An outbreak of norovirus infection linked to oyster consumption at a UK restaurant, February 2010." Journal of Public Health 33(2): 205-211.
Butcher, H., et al. (2016). "Whole genome sequencing improved case ascertainment in an outbreak of Shiga toxin-producing Escherichia coli O157 associated with raw drinking milk." Epidemiology & Infection FirstView: 1-12.
Carvalho, C., et al. (2012). "A possible outbreak of hepatitis a associated with semidried tomatoes, England, July-November 2011." Eurosurveillance 17(6).
Cleary, P., et al. (2010). "A foodborne outbreak of Salmonella Bareilly in the United Kingdom, 2010." Euro Surveillance: Bulletin Europeen sur les Maladies Transmissibles = European Communicable Disease Bulletin 15(48): 2.
Gajraj, R., et al. (2012). "Multiple outbreaks of Salmonella braenderup associated with consumption of iceberg lettuce." International Journal of Environmental Health Research 22(2): 150-155.
Harker, K. S., et al. (2011). "An outbreak of Salmonella Typhimurium DT191a associated with reptile feeder mice." Epidemiology & Infection 139(8): 1254-1261.
Inns, T., et al. (2013). "Outbreak of salmonella enterica goldcoast infection associated with whelk consumption, England, June to October 2013." Eurosurveillance 18(49).
Inns, T., et al. (2015). "A multi-country Salmonella Enteritidis phage type 14b outbreak associated with eggs from a German producer: 'near real-time' application of whole genome sequencing and food chain investigations, United Kingdom, May to September 2014." Euro Surveillance: Bulletin Europeen sur les Maladies Transmissibles. European Communicable Disease Bulletin 20(16).
Jenkins, C., et al. (2015). "Public Health Investigation of Two Outbreaks of Shiga Toxin-Producing Escherichia coli O157 Associated with Consumption of Watercress." Applied & Environmental Microbiology 81(12): 3946-3952.
Launders, N., et al. (2013). "Outbreak of Shiga toxin-producing E. coli O157 associated with consumption of watercress, United Kingdom, August to September 2013." Euro Surveillance: Bulletin Europeen sur les Maladies Transmissibles = European Communicable Disease Bulletin 18(44).
McCartney, G., et al. (2010). "The use of a new virtual cohort study design to investigate an outbreak of E. coli O157 linked to a supermarket delicatessen." Epidemiol Infect 138(10): 1439-1442.
Mook, P., et al. (2015). "Selection of population controls for a Salmonella case-control study in the UK using a market research panel and web-survey provides time and resource savings." Epidemiol Infect: 1-11
Westrell, T., et al. (2010). "Norovirus outbreaks linked to oyster consumption in the United Kingdom, Norway, France, Sweden and Denmark, 2010." Euro Surveillance: Bulletin Europeen sur les Maladies Transmissibles. European Communicable Disease Bulletin 15(12): 25.