Epidemiol. Infect. (2016), **144**, 2165–2175.
doi:10.1017/S0950268816000182

Utility of algorithms for the analysis of integrated *Salmonella* surveillance data

L. VRBOVA[1]*, D. M. PATRICK[1,2], C. STEPHEN[3,4], C. ROBERTSON[5], M. KOEHOORN[1], E. J. PARMLEY[6], N. I. DE WITH[7] AND E. GALANIS[1,2]

[1] *School of Population and Public Health, University of British Columbia, Vancouver, British Columbia, Canada*
[2] *British Columbia Centre for Disease Control, Vancouver, British Columbia, Canada*
[3] *Canadian Wildlife Health Cooperative, University of Saskatchewan, Saskatoon, Saskatchewan Canada*
[4] *Western College of Veterinary Medicine, University of Saskatchewan, Saskatoon, Saskatchewan, Canada*
[5] *Geography and Environmental Studies, Wilfrid Laurier University, Waterloo, Ontario, Canada*
[6] *Laboratory for Foodborne Zoonoses, Public Health Agency of Canada, Guelph, Ontario, Canada*
[7] *British Columbia Ministry of Agriculture, Abbotsford, British Columbia, Canada*

Received 11 February 2015; Final revision 20 December 2015; Accepted 19 January 2016; first published online 4 March 2016

SUMMARY

The objective of this study was to assess the use of statistical algorithms in identifying significant clusters of *Salmonella* spp. across different sectors of the food chain within an integrated surveillance programme. Three years of weekly *Salmonella* serotype data from farm animals, meat, and humans were used to create baseline models (first two years) and identify weeks with counts higher than expected using surveillance algorithms in the third (test) year. During the test year, an expert working group identified events of interest reviewing descriptive analyses of same data. The algorithms did not identify *Salmonella* events presenting as gradual increases or seasonal patterns as identified by the working group. However, the algorithms did identify clusters for further investigation, suggesting they could be a valuable complementary tool within an integrated surveillance system.

Key words: Foodborne zoonoses, food safety, public health, *Salmonella*, surveillance.

INTRODUCTION

Salmonella is a major bacterial pathogen that continues to pose a health and economic burden in Canada [1]. The reservoirs of *Salmonella* include humans and a wide range of wild and domestic animals. Transmission occurs through ingestion of food derived from infected animals, food contaminated with animal or human faeces, or contact with infected animals, people, or their environment [2].

Analysis of human laboratory isolates of *Salmonella* for surveillance purposes can facilitate the early detection of outbreaks; when combined with epidemiological data (e.g. age, sex, risk factor, and exposure information), such analyses can also identify emerging sources of infection to aid in the development and implementation of control measures. Many surveillance efforts focus solely on human *Salmonella* isolates [3–5]. However, there are initiatives using isolate data from non-human (e.g. animal, food) sources for surveillance, such as the World Health Organization's Global Foodborne Infections Network (GFN) [6], FoodNet Canada [7], the Canadian Integrated Program for Antimicrobial Resistance Surveillance (CIPARS) [8], and the Danish Programme for surveillance of antimicrobial

* Author for correspondence: Dr L. Vrbova, School of Population and Public Health, University of British Columbia, Vancouver, British Columbia, Canada.
(Email: linda.vrbova@gmail.com)

consumption and resistance in bacteria from animals, food and humans (DANMAP) [9]. These preceding initiatives have largely focused on identifying trends, *vs.* outbreaks, in the non-human data to target and evaluate measures for reducing *Salmonella* incidence and/or antimicrobial resistance.

In British Columbia (BC), the Integrated Salmonella (IS) Surveillance programme compiles laboratory *Salmonella* data from three sectors – humans, animals and food. Data are analysed regularly by the IS epidemiology working group (IS WG), a subset of the larger British Columbia Integrated Surveillance of Foodborne Pathogens Working Group (see Appendix). The IS WG members consist of public health and animal health practitioners from provincial and federal levels. IS WG analyses consist of regular qualitative reviews of descriptive data from each sector and across sectors (e.g. monthly counts by animal species, commodity, serotype, subtype), as well as looking for matching strains, trends over time, and clusters (increased number of isolates higher than expected based on historical data) [10].

Methods to analyse laboratory surveillance data to detect statistically significant clusters for further investigation include time-series analysis, regression-based methods, scan statistics, and cumulative sum statistics [11, 12]. The use of statistical methods to identify clusters of *Salmonella* in people has become part of routine public health surveillance and the use of these methods has been suggested in animal surveillance [12]. Recent studies have successfully used statistical surveillance algorithms to detect clusters of *Salmonella* in animals [13, 14]. However, these approaches have not been used to identify clusters spanning different sectors (animals, food, and humans) as part of an integrated surveillance system for public health.

The purpose of our study was to (1) assess whether human and non-human *Salmonella* data may be analysed using statistical surveillance algorithms to identify clusters across three sectors (animal, food, human), (2) assess the impact of adding travel history information to human isolate data, and (3) validate whether the algorithms identified the same clusters as those identified by the IS WG using a qualitative review.

METHODS

Data sources

This study used *Salmonella* serovars isolated from samples collected in BC between January 2008 and December 2010 from three sectors: human, animal, and food (meat). The data extracted from the BC IS database came from the following sources: (1) human laboratory diagnostic data from the BC Centre for Disease Control Public Health and Microbiology Reference Laboratory (BCCDC PHMRL) and the Public Health Agency of Canada National Microbiology Laboratory, (2) animal laboratory diagnostic data from the BC Ministry of Agriculture Animal Health Centre, and (3) retail meat (food) data from CIPARS, an integrated surveillance programme coordinated by the Public Health Agency of Canada that aims to monitor trends in antimicrobial use and resistance [8]. Details on these source data and their limitations have been published previously [10].

Human data

All diagnostic *Salmonella* isolates sent to the BCCDC PHMRL were included in this study. Subtyping by pulse-field gel electrophoresis (PFGE) was done for all human isolates and phage typing was done on isolates received in the first 15 days of each month for certain serotypes [10]. Travel history is not included in this laboratory-based dataset, and therefore is generally not available for prospective surveillance analyses (such as those in this study). However, since this study was conducted retrospectively, we were able to investigate the potential effect of including travel data history on a subset of *Salmonella* Enteritidis isolates. *S.* Enteritidis isolates were linked to information in the integrated Public Health Information System (iPHIS) to obtain travel exposure information (i.e. travel history, no travel history, and unknown travel history). *S.* Enteritidis cases with no travel history were classified as 'domestic' cases.

Animal data

Submissions from chicken, cattle, swine, and turkeys with a laboratory isolation of *Salmonella* were obtained from the provincial animal health diagnostic laboratory. Phage-type (PT) subtyping was done for certain serotypes [10]. Isolates were stratified by type of submission: diagnostic (disease cases and investigations), monitoring (routine animal health programmes, chicken only), and special project investigations. Data for this study is primarily from diagnostic submissions. Although diagnostic isolates are often from animal necropsies, and ill animals are

unlikely to enter the food chain, we assume that the farm and all animals on the farm are potentially contaminated/infected with *Salmonella* where there is a positive *Salmonella* test. In addition, since diagnostic samples can also include swabs, faeces, or other samples from the animal and/or farm, no animal may have been removed from the food chain.

Samples submitted for unknown reasons (i.e. not identified as diagnostic, monitoring, or project) were included in analyses. Data generated through targeted research projects were excluded from further analysis due to highly intermittent sampling. Monitoring data (available only for chicken) consisted of batched fluff samples from hatcheries; while sparse and intermittent, the data were included in analyses as a supplement to the constant diagnostic data.

For chicken, two time series were created for each serotype: 'all chicken' and 'diagnostic chicken', where 'all chicken' included diagnostic, monitoring, and samples submitted for unknown reasons. Since subclinical infection with *Salmonella* is common in many animals [15], we assumed that inclusion of asymptomatic animal isolates could correlate better with human risk. Therefore, for chicken, 'all chicken' time series were chosen over 'diagnostic chicken' time series for examination of cross-sectoral signals.

Food data

Data about *Salmonella* identified in retail chicken and pork meat samples purchased at the point of sale in BC were extracted from the CIPARS system. Subtyping (PT) was done for certain serotypes [10]. Retail sampling occurred approximately every 2 weeks, with eight samples of pork and eight samples of chicken collected in each sampling week [16]. Retail meat samples included both domestic and imported products. Data were stratified by meat type (chicken and pork).

Events identified by the IS WG

Information about events identified by the IS WG in 2010 was abstracted from IS WG meeting minutes and through supplemental interviews with the IS WG members. The IS WG met six times to review 2010 data: five times in 2010 [March (week 10), June (week 23), July (week 30), September (week 37), November (week 45)], and once in 2011 [March (week 11)]. The meeting minutes recorded information from the group's review of data from each sector that were then summarized in tables as monthly counts of all serotypes for each sector (and species) for the previous 3 months and previous years, matching serotypes (and PFGE/PT subtypes when available) spanning the sectors. Additionally, monthly graphs for the more common serotypes illustrating patterns (e.g. seasonality), increases above that expected, and trends across sectors were reviewed. No statistical tests were used by the IS WG to identify patterns, trends or clusters. Cross-sectoral events were defined as those where the group noted a pattern of interest (e.g. an increase) in a particular serotype in at least two of the three sectors.

Statistical signals across time series

Three algorithms were used in to identify a particular week of interest; this was done in order to increase the specificity of the signals in any particular week (i.e. all three algorithms had to agree a signal is present), while allowing for increased sensitivity in detecting smaller but more sustained increases (i.e. a signal identified in only one algorithm but in two consecutive weeks). The surveillance algorithms used to detect statistically significant signals in individual time series were: (1) the Farrington algorithm [17] (also used by Kosmider *et al.* [18] for detection of *Salmonella* clusters in animal data), (2) a Bayesian algorithm, and (3) the Robert Koch Institute (RKI) algorithm (all algorithms in Hohle [19]). These algorithms were chosen as they have been used previously to identify unusual agri-food chain contamination and the emergence of a serotype that correlated with an emergence of the same serotype in humans [14]. Weekly time series were created for each serotype for: humans, each animal species, and each meat type separately. An additional time series of 'domestic human' cases was created for *S.* Enteritidis. All time series with at least one submission in the 'test year' (2010) and two submissions in the 3-year time interval were included in the analyses. Since *Salmonella* infections are often found to be seasonal in both humans and animals, these algorithms obtain expected values in the form of 'moving windows' of pre-specified subsets of past counts (3-week windows were used in analyses) that form the baseline, thereby directly accounting for seasonality. The past counts in the years 2008 and 2009 were used to create the baseline, with statistical signals evaluated for weeks in 2010. Expected values consisted of counts from 6 weeks (three from each of the previous 2 years): the same week, the week before and week after were used.

All three algorithms assume that the past counts follow a Poisson distribution, a distribution used for modelling counts of (rare) events in a specific period of time. The Farrington method calculates an expected value by fitting a Poisson regression model to the past counts, creating 95% confidence intervals using a transformed Normal distribution. The Bayesian algorithm employs an empirical Bayes approach to Poisson-distributed reference values by defining a non-informative Gamma prior distribution on the Poisson rate parameter (i.e. observed counts), which leads to a negative binomial distribution for the posterior distribution (i.e. expected counts). The RKI algorithm sets the expected value to the mean of the reference distribution, calculates a 95% Poisson confidence interval around the mean, and compares whether the current value is within the interval. All analyses were performed in R [20], using the 'surveillance' package [12, 19].

The following assumptions were required for the analyses: *Salmonella* isolates are independent (i.e. one isolate does not affect the probability of another isolate), samples are submitted at a constant rate (hence denominators do not need to be included), and isolation counts follow Poisson distributions. The data from each of the three sectors used in this study (human, animal, food) may violate these assumptions for different reasons (see Table 1 for details).

Once algorithm analyses were conducted in each sector for each serotype, the weeks with important public health *signals* were those meeting the following criteria:

(1) Weeks with statistically significant results ($\alpha = 0{\cdot}01$) for unique serotypes identified by all three algorithms (for higher specificity).
(2) Weeks with statistically significant results ($\alpha = 0{\cdot}01$) for unique serotypes (using $\geqslant 1$ algorithm) persisting for $\geqslant 2$ consecutive weeks in humans or animals and either consecutive or alternating weeks for meat samples (due to bi-weekly sampling).

Serotypes with a public health signal in more than one sector (animal, food, human) for a week in 2010 were considered to be cross-sectoral signals, regardless of the time difference between significant weeks. For each serotype with a cross-sectoral signal, available subtyping (PT or PFGE patterns) as well as travel information for the human cases was extracted from the IS database and iPHIS, respectively.

RESULTS

Salmonella isolates and signals in individual time series

Over the 3-year study period there were a total of 3335 human isolates, with 150 serotypes observed overall. There were 1477 *S.* Enteritidis isolates, 68% (1007/1477) were linked to travel history data, and of these 69% (693/1007) were domestically exposed.

In 2010, the 'test year', 68% (67/99) of human serotype time series had at least one isolate and were therefore tested for signals. For the 67 time series, 66 had a signal using the Farrington algorithm, 63 using the Bayes, and seven using the RKI algorithm. There were 750 animal isolates representing 68 serotypes; 79% (595/750) of isolates were from chicken. In the 'test year' 2010, there were 31 time series created for serotypes with at least one positive isolate to test for signals: 15 'all chicken', nine 'diagnostic chicken', three cattle, two swine, and two turkey. For each of the 31 time series, there was at least one corresponding signal in both the Bayes and Farrington algorithms; there were signals in four time series using the RKI algorithm.

There were 169 meat isolates representing 20 serotypes; 96% (162/169) from chicken and 4% (7/169) from pork. Six serotypes has positives isolates in the 'test year' 2010, and all six had at least one signal using the Farrington algorithm, five had signals using the Bayes algorithm, and none had signals using the RKI algorithm.

Eighteen *Salmonella* serotypes were present in at least two sectors; six serotypes were present in all three sectors: *S.* Enteritidis, *S.* Hadar, *S.* Heidelberg, *S.* Kentucky, *S.* Schwarzengrund, and *S.* Typhimurium.

Comparison of events identified by the IS WG with statistically significant algorithm signals across sectors

The IS WG identified four serotypes for investigation, *S.* Enteritidis, *S.* Heidelberg, *S.* Typhimurium, and *S.* I 4,[5],12:i:-, while the algorithms identified three cross-sectoral alerts among *S.* Enteritidis, *S.* Hadar, and *S.* Kentucky (Table 2).

Identified by both methods: *S.* Enteritidis investigation by subtype and travel history

The IS WG and the statistical algorithms both identified *S.* Enteritidis; there were statistically significant

Table 1. *Data limitations: examination of assumptions for surveillance algorithms by sector*

Sector	Assumption	Assessment of assumption violations
Human	Independence	Unlikely violated: population-based data, isolates are diagnostic cases with each isolate likely representing one person; while repeat testing for an individual could violate the independence assumption, laboratory data are routinely checked for repeat isolates, and cases linked with iPHIS data (i.e. domestic cases) are case-based (rather than sample-based) since iPHIS is a case-based database.
	Constant rate	Unlikely violated: samples submitted at a constant rate
Animal	Independence	Likely violated: clustering likely present, since one isolate can represent a group of animals such as a flock or herd (or conversely, a number of samples from one group can be sent at one time), isolates coming from animals and their environment can be sent in at the same time (i.e. chicken monitoring samples), and isolates from the same flock/herd or animal can be sent sequentially until infection is no longer found in the samples.
	Constant rate	Likely violated: Agricultural animals are part of species-specific agricultural production systems (e.g. chicken for meat or egg-production, cattle for milk or meat production), differing widely in the reasons samples are sent to the laboratory, a biased fraction of samples are submitted the provincial laboratory, susceptible to differential testing due to seasonality and other changes within their particular production system affecting submission rates.
Food (meat)	Independence	Unlikely violated: samples independent by design: samples based on statistical and convenience sampling strategy
	Constant rate	Unlikely violated: sampling frequency determined by design, although small changes in sampling occurred throughout 2008–2009.

iPHIS, Integrated public health information system.

Table 2. *Comparison of* Salmonella *serotypes identified by the BC Integrated Salmonella Surveillance Working Group (IS WG) and by surveillance algorithms in at least two of the three sectors (human, animal and food) in 2010*

Serotype	Identified by	Reason	Sector	Details
Enteritidis	IS WG	Serotype and subtypes in three sectors	H, A, F	PT51 increase in chicken (animal) and chicken (meat) PT13a and PT8 identified in three sectors PT 15a increase in humans and identified in chicken (animal)
	Algorithms	Signals in three sectors	H, A, F	Signals in humans, chicken (animal) and chicken (meat) with high proportions of PT8, PT13a, and PT13
Heidelberg	IS WG	Increase in humans, serotype in three sectors	H, A, F	PT19 identified in three sectors, increase in humans led to outbreak investigation: no common source found
Typhimurium	IS WG	Serotype in two sectors	H, A	Serotype in swine and cattle, seasonality in cattle
I 4,[5],12:i:-	IS WG	Serotype in two sectors	H, A	Seasonal trend in humans
Hadar	Algorithms	Signals in two sectors	H, F	Signal in humans and chicken (meat)
Kentucky	Algorithms	Signals in two sectors	H, A, F	Signal in humans, chicken (animal) and chicken (meat)

H, Human; A, animal; F, food.

signals in chicken (animal), chicken meat, humans (all), and 'domestic humans' (Fig. 1*a*, *b*).

However, the algorithms and IS WG identified events with different subtypes: both methods identified PT8 and PT13a, the statistical algorithms identified PT13, and the IS WG identified PT51 and PT15a (Table 3). PT51 was identified by the IS WG because of a gradual increase of the subtype, particularly in animals. There were not enough isolates of PT15a to generate a statistically significant signal in animals.

Figure 1(*a*, *b*) shows the chicken (animal) time series and chicken meat time series for 2010, as well as weeks with statistically significant signals and the PT composition of isolates within the signals (for the top PTs of interest). Figure 1*a* shows weekly counts

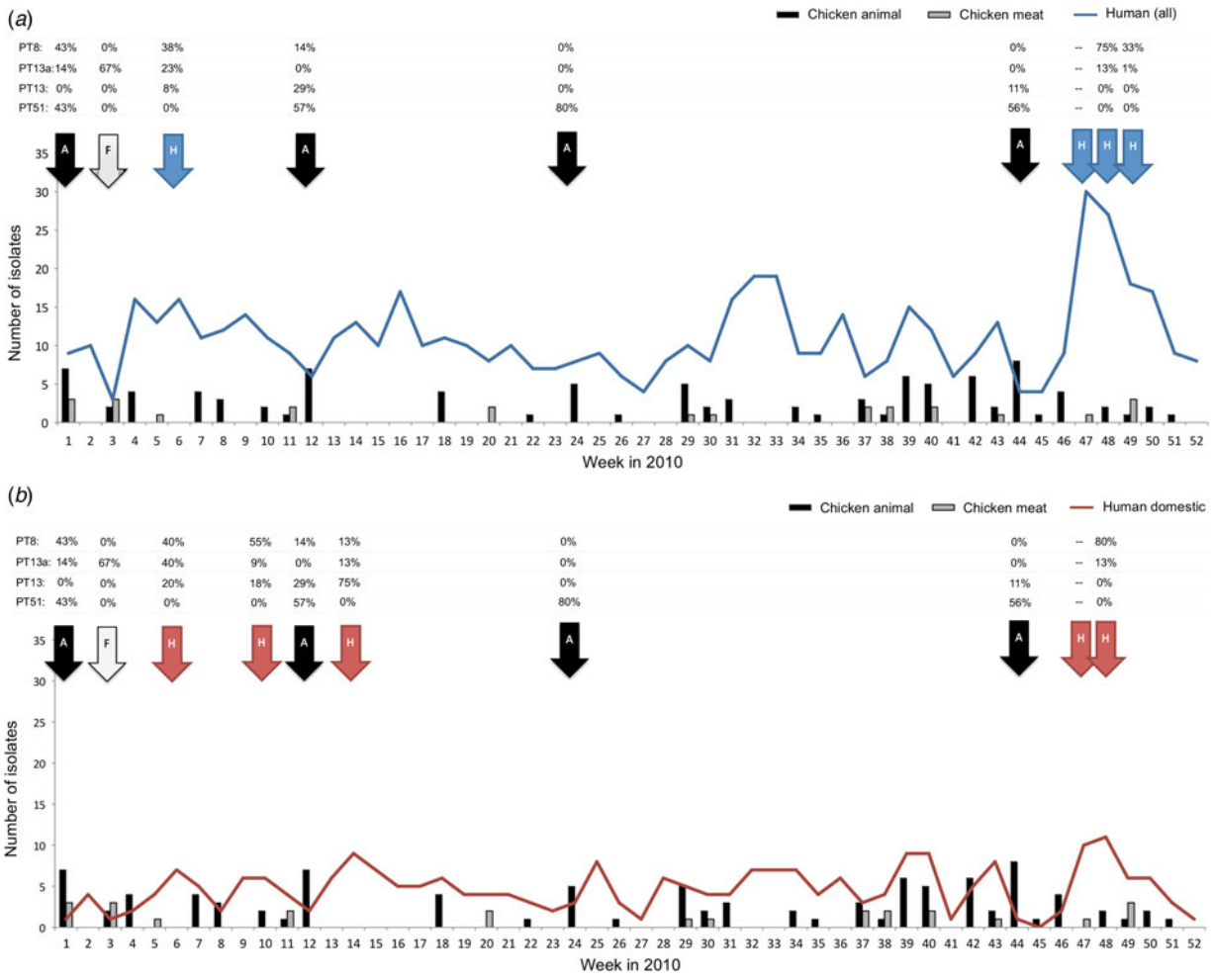

Fig. 1. (*a*) *Salmonella* Enteritidis (SE) isolates in 2010 from animals (chicken), food (chicken meat) and humans (all human cases). (*b*) SE isolates in 2010 from animals (chicken), food (chicken meat) and humans (domestic human cases). The blue line represents the weekly number of SE isolates from humans; the red line represents the weekly isolates of domestic SE cases. Black bars are the weekly number of live chicken (and their environment) SE isolates, grey bars are the weekly number of chicken meat SE isolates. Black 'A' arrows indicate statistically significant signals in animals (chicken), the grey 'F' arrow indicates a statistically significant signal in food (chicken meat), and red 'H' arrows indicate statistically significant signals in humans. Proportions of each particular phage type (PT) out of the total with a valid PT for the week for the sector with the statistically significant signal (indicated by arrows).

and signals for all human cases (domestic, travel, and unknown), while Figure 1*b* shows weekly counts and signals for domestic human cases only. Overlap of PTs (PT8, PT13a, PT13) within signals is present in the first third of the year (up to week 14), but not in the second two thirds of the year, where animal signals consisted of PT51 and PT13, while human signals consisted largely of PT8 and PT13a.

The analysis of all human isolates yielded one additional signal in week 49 (subsequent to signals in week 47 and 48) that was not identified in the domestic isolates (Fig. 1*a*). Conversely, analysis of domestic human isolates alone yielded two additional signals in the first third of the year (weeks 10 and 14) that the analysis with all cases combined did not identify (Fig. 1*b*).

The time difference from an initial animal signal and any subsequent human signal (with the same PT) ranged from 5 to 48 weeks (2–47 weeks for domestic cases); for food the time to the human signals with the same PT ranged from 3 to 46 weeks (3–45 weeks for domestic cases).

Identified by algorithms only: *S.* Hadar and *S.* Kentucky

The algorithms identified signals in *S.* Hadar in chicken meat and 36 weeks (over 8 months) later in

Table 3. *Comparison of* Salmonella *Enteritidis subtypes identified by the BC Integrated Salmonella Surveillance Working Group (IS WG) and by surveillance algorithms and identification of cross-sectoral clusters in at least two of the three sectors (human, animal and food) in 2010 by week of investigation*

Data review	Week*	Sector	2010 data available	PT subtypes identified by		Cross-sectoral clusters identified†			
				IS WG	Algorithms	PT8	PT13	PT13a	PT51
1	10	Human	Jan.–Feb.		8, 13, 13a	Algorithms	None	Algorithms	None
		Animal	Jan.–Feb.		8, 13a, 51				
		Food	None		Not applicable				
2	23	Human	Jan.–May		8, 13, 13a	Algorithms	Algorithms	Algorithms	None
		Animal	Jan.–May		8, 13, 51				
		Food	Jan.–May		13a				
3	30	Human	Apr.–June	8, 13a		IS WG	None	IS WG	None
		Animal	Apr.–June	8, 13a	51				
		Food	Apr.–May	8, 13a					
4	37	Human	Apr.–July	51		None	None	None	IS WG
		Animal	Apr.–July	51					
		Food	Apr.–July	51					
5	45	Human	June–Aug.	8, 13a, 51		IS WG	None	IS WG	IS WG
		Animal	June–Aug.	8, 13a, 51					
		Food	June–Aug.	8, 13a, 51					
6	11	Human	Sep.–Dec.	8	8, 13a	None	None	None	None
		Animal	Sep.–Dec.	51	13, 51				
		Food	Sep.–Dec.						

PT, Phage type.
* Week data review occurred in 2010 (for data reviews 1–5), and 2011 (for data review 6).
† Cross-sectoral cluster: PT identified in at least two of three sectors (human, animal and food).

humans. Of the five human isolates in the cluster, three had no travel information and two were travel related. The signals in *S.* Kentucky were in: 'all chicken' (weeks 35 and 43), chicken meat (weeks 12, 41, and 50), and humans (week 38). Of the five human isolates in the cluster in week 38, three had no travel information and two were travel related.

Identified by IS WG only: *S.* Heidelberg, *S.* Typhimurium and *S.* I 4,[5],12:i:-

An investigation into *S.* Heidelberg started at the beginning of 2010 based on observing similar PT/PFGE patterns in all three sectors (humans, chicken, and chicken meat), in late 2009 and early 2010. The IS WG observed another increase in *S.* Heidelberg in autumn 2010; however, they did not find common subtypes to warrant further investigation. *S.* Typhimurium was identified by the IS WG based on seasonality in animals and an increase in humans. While the increase in humans was also identified using the algorithms, there were no signals in the animal or meat sectors, hence no cross-sectoral signal was generated. The *S.* I 4,[5],12:i:- serotype was identified by the IS WG due to differences in seasonality patterns in humans and animals. While a significant signal in *S.* I 4,[5],12:i:- was identified in humans, no cross-sectoral signal for this serotype was generated because seasonal trends in animals were not identified as an alert by the algorithms (as seasonality is accounted for by the window-based sampling method of the algorithms).

DISCUSSION

Analysis of integrated *Salmonella* surveillance data from agricultural animals, meat, and humans using surveillance algorithms did not identify the same cross-sectoral clusters as the IS WG. This lack for agreement may be due to the IS WG using more diverse criteria to identify events of interest, animal data violating statistical assumptions required by the algorithms, and/or the lack of a defined time interval between statistically significant weeks across sectors.

The IS WG used descriptive analyses, such as graphs illustrating longer-term trends and seasonality, while the algorithms were designed to detect weeks with increases above expected based on baseline data.

Tighter alignment of the objectives of surveillance and algorithm method selection may ensure a more prominent role for automated techniques [21] in supplementing working group investigations. For example, in order to detect changes in trends and seasonality, integrated surveillance studies could try to identify potential clusters using correlation between smoothed counts and statistically significant linear and nonlinear trends, such as those done previously in animals and humans separately that identified common patterns [22, 23], or use time series methods to detect gradual increases or correlations in seasonality [24]. Examination of longer-term changes in *Salmonella* PT distribution has been instrumental in generating hypotheses to explain increases in human cases and in identifying an emergence, such as that of PT 51 in animals, food, and humans [10]. Notably, however, the surveillance algorithms as defined and used in this study presented a new way to examine the same data, and provided targeted investigation opportunities for the IS WG, such as the identification of clusters of isolates to be examined in more detail using subtyping and/or epidemiological information.

Animal data likely violated the assumptions required for the algorithms (Table 1). Animal surveillance data (especially diagnostic data) are influenced by changing economic circumstances within the livestock industry, such as large outbreaks [18], with outbreaks potentially affecting independence as well as denominators (see Table 1). Further, while animal samples submitted for routine surveillance and monitoring to comply with various legislative or industry requirements (e.g. table egg production) could be a useful data source, the data are currently too sparse and intermittent (e.g. submitted to the laboratory in batches) for use in statistical algorithms. Future analyses could aggregate isolates into 'epidemiological units' and 'incidents' to address violations of the independence assumption in animal data streams whereby an 'incident' comprises the first isolation and all subsequent isolations of the same serotype (and PT) from the 'epidemiological unit' of animals (an individual animal or groups of animals in the same herd/flock) within a standard time period, such as 30 days [14]. Unfortunately, we did not have the necessary data (i.e. submitter information) to aggregate isolates into such units. Denominator data, such as the number of samples submitted for *Salmonella* testing [23], would help interpret animal surveillance data, allow for the assessment of sample submission rate, and justify constant rate assumptions.

The surveillance algorithm methodology enabled identification of high weekly counts as well as smaller sustained increases (i.e. over a few weeks); this was especially important for identification of signals in low-count data streams. With more years of data, larger time aggregation units (and associated higher counts in the time units) may lead to better performance of surveillance algorithms with more robust calculations of expected counts and confidence intervals. Kosmider *et al*. [14] successfully applied the Farrington algorithm to monthly *Salmonella* time series in animals over 10 years, and statistically significant temporal clusters of monthly counts have been found in *Salmonella* in both humans [22] and animals [23] using time scan statistics. Danan *et al*. combined all of the data from different species together, and were able to find statistically significant signals linked to contamination in the agri-food chain that were confirmed upon investigation [13].

The fact that we were limited to identifying statistically-significant weeks in *Salmonella* serotypes in more than one sector (animal, food, human), irrespective of the time difference between significant weeks (i.e. occurring within the same year) could be seen as a weakness of this study. However, as there was no defined time interval to determine the risk period to humans (or meat) following a signal in animals, we could at least investigate the possible correlations between these signals qualitatively.

At first sight, Figure 1 appears promising with signals in animals (and food in one instance) preceding human infections of *S*. Enteriditis. However, since we only had PT subtyping data to assess possible links between signals, the time lags between animal/food signals and human signals with the same PT ranged from 2 to 48 weeks. While the 48-week time lag may seem unreasonably long, a Canadian study found a 10-month temporal cluster in *S*. Typhimurium var. Copenhagen that ended in chickens 9 months prior to the start of a human temporal cluster, suggesting their results point to a possible association between human illness and exposure to chicken products [22].

A reasonable time interval between the two signals (animal and human, or animal and food, or food and human) is necessary to quantitatively test for correlation among the animal and human time series (i.e. positive predictive value of animal/food signals for human signals). Although the incubation period for salmonellosis is estimated to range between 12 and 36 h [2], due to the complexity of *Salmonella* dynamics

in the food chain, a reasonable time estimate requires either the results of previous outbreak investigations where human cases were linked to known food/animal contamination events; or, the collection of species-specific information that is currently lacking, such as the animal production type (e.g. egg *vs.* meat for chicken, milk *vs.* beef for cattle), and the age of the animal (e.g. to estimate time to slaughter). *Salmonella* contamination can also occur during transport, slaughter, or processing, whereby *Salmonella* is spread between infected and previously *Salmonella*-free animals resulting in *Salmonella*-contaminated meat cuts [25]. Additionally, salmonellosis has been associated with consumption of processed and frozen chicken products [26], adding to the potential shelf-life of such product. Finally, foods eaten in a jurisdiction are not necessarily produced in the same jurisdiction.

Our results suggest that a timely link between laboratory and epidemiological data would be beneficial for outbreak detection. There were differences in the signals generated by the time series for all *S.* Enteritidis human cases compared to the domestic human cases only. Both time series generated signals at the beginning and end of the test year. However, the domestic time series provided an extra two signals in the beginning of the year that the overall human time series did not. This suggests that inclusion of travel exposure information, especially if it is available for a substantial proportion of cases (68% of *S.* Enteritidis in this study), could result in generation of more meaningful alerts. The investigation of the two other cross-sectoral clusters identified by the algorithms (*S.* Hadar and *S.* Kentucky) support this, as both contained cases with travel history.

Further investigation into appropriate analytical methods for integrated surveillance programmes is needed. A review found that studies attempting an analytic linkage between animal and human health data (i.e. where animal data was used to quantitatively predict human risk), were limited largely to diseases that had a clear spatio-temporal component such as West Nile virus [27]. A pathogen such as *Salmonella* that is transmitted through a complex food production and distribution chain (as well as through other means, including direct contact and water) may only show a spatiotemporal association between human cases and the final products in the marketplace (e.g. meat on grocery store shelves), not throughout its entire production path. In general, a signal in meat should have a higher specificity than a signal in animals in terms of predicting risk of disease in humans, since meat is much closer to humans in the food chain than animals. Although limited, our results appear to support this: the only cross-sectoral signals with matching subtypes in *S.* Enteritidis were during the time interval that included a signal in meat (Fig. 1). An association with meat in our analyses is further supported by the knowledge that fresh chicken meat consumed in BC is largely produced locally within the province [10]. It is important to stress that appropriate farm and/or abattoir monitoring programmes could also detect relevant contamination events and would enable a timelier response (e.g. prior to meat distribution). The difficulty in identifying clear actionable alerts is likely due to the complexity of *Salmonella* dynamics within the food chain, specifically quantifying the risk to human health from a live animal with *Salmonella* and the lack of specificity and quantity of data in certain sectors. Without additional contextual information, especially relevant epidemiological data, animal and food surveillance data of this type may be more amenable to generating hypotheses in epidemiological investigations and in helping evaluate programmes by examining longer-term trends. For example, an international investigation of a human outbreak of *S.* Typhimurium used national animal and meat *Salmonella* surveillance databases to supplement human surveillance data and identify pork meat as the source of infection [28]. Moreover, animal surveillance data has been successfully linked with human data in source attribution studies, where various modeling approaches are used to make inferences about the relative importance of different food/animal sources for disease in the population [29].

While salmonellosis is generally considered to be a foodborne zoonosis acquired through the consumption of foods of animal origin, it can also be acquired thought the consumption of fruits and vegetables, as well as direct transmission from infected persons and animals [30]. Further, in previous investigations, *Salmonella* serotype distribution (number of specific serotypes and their frequencies) was different between animals at slaughter (chicken, cattle, and swine) and humans [31], and the distribution of *S.* Typhimurium DT104 (based on varying antimicrobial resistance patterns) was also found to be different between human and animal populations (primarily cattle) [32]. Such findings call into question links between agricultural animal and human isolates, especially causal links, even for the same subtype. As whole genome sequencing becomes more widely

available and affordable, novel genotyping methods based on sequence-based identification will result in better resolution of subtypes [33]. This may allow for future cluster/outbreak investigations to better specify the sources of initial infection isolates in human and animal populations, despite issues with categorizing this vast amount of new information into epidemiologically meaningful categories.

Our results indicate if surveillance algorithms were used prospectively they could complement IS WG manual reviews by identifying different cross-sectoral clusters. Ongoing cluster and outbreak investigations, as well as better epidemiological information associated with surveillance data, are needed to evaluate a relevant lag time between related statistically significant animal, meat, and human signals, as well as the predictive value of cross-sectoral clusters identified using surveillance algorithms for integrated surveillance systems using animal, human and food data.

APPENDIX

British Columbia Integrated Surveillance of Foodborne Pathogens Working Group (in alphabetical order, affiliations at time of analysis)

Sean Byrne[1], Nancy De With[1], Joe Fung[2], Eleni Galanis[3,4], Linda Hoang[2,3], Mira Leslie[1], Min Li[4], Sophie Li[4], Laura MacDougall[4], Lorraine McIntyre[4], Muhammad Morshed[2,3], Ana Paccagnella[2], E. Jane Parmley[5], Jane Pritchard[1], Brian Radke[4], Craig Stephen[6,7] Jason Stone[8], Marsha Taylor[4], Julie Wong[2], Erin Zabek[1]

[1]*British Columbia Ministry of Agriculture, Abbotsford, BC, Canada*
[2]*British Columbia Centre for Disease Control Public Health Microbiology & Reference Laboratory, Vancouver, BC, Canada*
[3]*University of British Columbia, Vancouver, BC, Canada*
[4]*British Columbia Centre for Disease Control, Vancouver, BC, Canada*
[5]*Public Health Agency of Canada, Guelph, ON, Canada*
[6]*Centre for Coastal Health, Nanaimo, BC, Canada*
[7]*University of Calgary, NW, Calgary, AB, Canada*
[8]*Fraser Health Authority, Surrey, BC, Canada*

ACKNOWLEDGEMENTS

The authors acknowledge David Graham of the Canadian Food Inspection Agency, Merv Wetzstein and Jaime Osei-Appiah of British Columbia (BC) Ministry of Agriculture, Loretta Janz of BC Centre for Disease Control (BCCDC) Public Health and Reference Microbiology Laboratory (PHMRL), BCCDC Environmental Microbiology PHMRL staff, BC Ministry of Agriculture Animal Health Centre staff, Public Health Agency of Canada (PHAC) Laboratory for Foodborne Zoonoses staff, PHAC National Microbiology Laboratory staff, and the Canadian Integrated Program for Antimicrobial Resistance Surveillance (CIPARS) BC field staff. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

DECLARATION OF INTEREST

None.

REFERENCES

1. **Thomas MK, *et al.*** Estimates of the burden of foodborne illness in Canada for 30 specified pathogens and unspecified agents, circa 2006. *Foodborne Pathogens and Disease* 2013; **10**: 639–648.
2. **Heymann D (ed.).** *Control of Communicable Diseases Manual*, 19th edn. Washington, DC: American Public Health Association, 2008.
3. **Public Health Agency of Canada.** National Enteric Surveillance Program (NESP) (www.nml-lnm.gc.ca/NESP-PNSME/index-eng.htm). Accessed 19 November 2014.
4. **Centers for Disease Control and Prevention.** National Salmonella Surveillance (http://www.cdc.gov/nationalsurveillance/salmonella-surveillance.html). Accessed 19 November 2014.
5. **European Centre for Disease Prevention and Control.** Food- and Waterborne Diseases and Zoonoses Programme (http://ecdc.europa.eu/en/activities/diseaseprogrammes/fwd/Pages/index.aspx). Accessed 19 November 2014.
6. **World Health Organization.** Global Foodborne Infections Network (GFN) (http://www.who.int/gfn/en/). Accessed 19 November 2014.
7. **Public Health Agency of Canada.** FoodNet Canada (formerly known as C-EnterNet) (http://www.phac-aspc.gc.ca/foodnetcanada/index-eng.php). Accessed 19 November 2014.
8. **Public Health Agency of Canada.** Canadian Integrated Program for Antimicrobial Resistance Surveillance (CIPARS) (http://www.phac-aspc.gc.ca/cipars-picra/index-eng.php). Accessed 19 November 2014.
9. **DANMAP.** DANMAP: the Danish programme for surveillance of antimicrobial consumption and resistance in bacteria from animals, food and humans (http://www.danmap.org). Accessed 19 November 2014.
10. **Galanis E, *et al.*** Integrated surveillance of salmonella along the food chain using existing data and resources

in British Columbia, Canada. *Food Research International* 2012; **45**.
11. **Sonesson C, Bock D.** A review and discussion of prospective statistical surveillance in public health. Journal *of the Royal Statistical Society Series A (Statistics in Society)* 2003; **166**: 5–21.
12. **Hohle M, Paul M, Held L.** Statistical approaches to the monitoring and surveillance of infectious diseases for veterinary public health. *Preventive Veterinary Medicine* 2009; **19**: 2–10.
13. **Danan C, *et al.*** Automated early warning system for the surveillance of salmonella isolated in the agro-food chain in France. *Epidemiology and Infection* 2011; **139**: 736–741.
14. **Kosmider R, *et al.*** A statistical system for detecting *Salmonella* outbreaks in British livestock. *Epidemiology and Infection* 2006; **134**: 952–960.
15. **Humphrey T, Threlfall E, Cruikshank J.** Salmonellosis. In: S. R. Palmer, Lord Soulsby, D. I. H. Simpson, eds. Zoonoses: Biology, Clinical Practice, and Public Health Control. New York, NY: Oxford University Press, 1998, pp. 191–206.
16. **Public Health Agency of Canada.** Canadian Integrated Program for Antimicrobial Resistance Surveillance (CIPARS) Annual Report 2008 (http://www.phac-aspc.gc.ca/cipars-picra/2008/index-eng.php). Accessed 20 January 2015.
17. **Farrington CP, *et al.*** A statistical algorithm for the early detection of outbreaks of infectious disease. *Journal of the Royal Statistical Society Series A* (*Statistics in Society*) 1996; **159**: 547–563.
18. **Kosmider RD, *et al.*** Detecting new and emerging diseases on livestock farms using an early detection system. *Epidemiology and Infection* 2011; **139**: 1476–1485.
19. **Hohle M.** Inside R: Surveillance (www.inside-r.org/packages/cran/surveillance). Accessed 7 October 2015.
20. **WU Wien Institute for Statistics and Mathematics.** The R Project for Statistical Computing (www.r-project-org). Accessed 9 October 2012.
21. **Robertson C, *et al.*** Review of methods for space-time disease surveillance. *Spatial and Spatio-temporal Epidemiology* 2010; **1**: 105–116.
22. **Guerin MT, Martin SW, Darlington GA.** Temporal clusters of salmonella serovars in humans in Alberta, 1990–2001. *Canadian Journal of Public Health* 2005; **96**: 390–395.
23. **Guerin MT, *et al.*** A temporal study of salmonella serovars in animals in Alberta between 1990 and 2001. *Canadian Journal of Veterinary Research* 2005; **69**: 88–99.
24. **Allard R.** Use of time-series analysis in infectious disease surveillance. *Bulletin of the World Health Organization* 1998; **76**: 327–333.
25. **Heyndrickx M, *et al.*** Routes for *Salmonella* contamination of poultry meat: epidemiological study from hatchery to slaughterhouse. *Epidemiology and Infection* 2002; **129**: 253–265.
26. **Currie A, *et al.*** Frozen chicken nuggets and strips and eggs are leading risk factors for *Salmonella* Heidelberg infections in Canada. *Epidemiology and Infection* 2005; **133**: 809–816.
27. **Scotch M, Odofin L, Rabinowitz P.** Linkages between animal and human health sentinel data. *BMC Veterinary Research* 2009; **5**: 15.
28. **Bruun T, *et al.*** An outbreak of *Salmonella* Typhimurium infections in Denmark, Norway, and Sweden, 2008. *Eurosurveillance* 2009; **14**: 1–6.
29. **Pires SM.** Assessing the applicability of currently available methods for attributing foodborne disease to sources, including food and food commodities. *Foodborne Pathogens & Disease* 2013; **10**: 206–213.
30. **Harker KS, *et al.*** National outbreaks of *Salmonella* infection in the UK, 2000–2011. *Epidemiology and Infection*. Published online: 31 May 2013. doi: 10.1017/S0950268813001210.
31. **Sarwari AR, *et al.*** Serotype distribution of *Salmonella* isolates from food animals after slaughter differs from that of isolates found in humans. *Journal of Infectious Diseases* 2001; **183**: 1295–1299.
32. **Mather AE, *et al.*** An ecological approach to assessing the epidemiology of antimicrobial resistance in animal and human populations. *Proceedings of the Royal Society of London, Series B: Biological Sciences* 2012; **279**: 1630–1639.
33. **Fournier PE, Drancourt M, Raoult D.** Bacterial genome sequencing and its use in infectious diseases. *Lancet Infectious Diseases* 2007; **7**: 711–723.