Attribution of the French human Salmonellosis cases to the main food-sources according to the type of surveillance data.

J M David, P Sanders, N Bemrah, S A Granier, M Denis, François-Xavier Weill, D Guillemot, L Watier

Attribution of the French human Salmonellosis cases to the main food-sources according to the type of surveillance data

J.M. David[1,2,3,4], P. Sanders[1], N. Bemrah[5], S. Granier[6], M. Denis[7], F-X. Weill[8], D. Guillemot[2,3,4], L. Watier*[2,3,4]

[1] Anses, BP 90203 Fougères, F-35302, France

[2] INSERM, U 657, Paris, F-75015, France

[3] Institut Pasteur, Pharmacoépidémiologie et Maladies Infectieuses, Paris, F-75015, France

[4] Université Versailles Saint Quentin, EA4499, F-Garches 92380, France

[5] Anses, Maisons-Alfort, F-94701, France

[6] Anses, Maisons-Alfort, F-94700, France

[7] Anses, BP 53, Ploufragan, F-22440, France

[8] Institut Pasteur, Unité des Bactéries Pathogènes Entériques, Centre National de Référence des *Salmonella*, Paris, F-75015, France

* Inserm U657, UVSQ EA 4499, Unité de Santé Publique, Hôpital Raymond Poincaré, Bat Rabelais, 104 av R. Poincaré, F-92380 Garches, France. Tel: +33 1 47 10 79 08 poste 6132 ; Fax: +33 1 47 14 49 08. Email: laurence.watier@inserm.fr

Present address of J.M. David: Public Health Agency of Canada, Laboratory for Foodborne Zoonoses, 3200 rue Sicotte, CP5000, St Hyacinthe, Quebec, J2S 7C6

Abstract

Salmonella are the most common bacterial cause of food-borne infections in France and ubiquitous pathogens present in many animal productions. Assessing the relative contribution of the different food-animal sources to the burden of human cases is a key step toward the conception, priorisation and assessment of efficient control policy measures. For this purpose, we considered a Bayesian microbial subtyping attribution approach based on a previous published model (Hald et al., 2004). It requires quality integrated data on human cases and on the contamination of their food sources, per serotype and microbial subtype, which were retrieved from the French integrated surveillance system for *Salmonella*. The quality of the data available for such an approach is an issue for many countries in which the surveillance system has not been designed for this purpose. In France, the sources are monitored simultaneously by an active, regulation-based surveillance system that produces representative prevalence data (as ideally required for the approach) and a passive system relying on voluntary laboratories that produces data not meeting the standards set by Hald *et al.* (2004) but covering a broader range of sources. These data allowed us to study the impact of data quality on the attribution results, globally and focusing on specific features of the data (number of sources and contamination indicator). The microbial subtyping attribution model was run using an adapted parameterization previously proposed (David et al., 2012). A total of 9,076 domestic sporadic cases were included in the analyses as well as 9 sources among which 5 were common to the active and the passive datasets. The greatest impact on the attribution results was observed for the number of sources. Thus, especially in the absence of data on imported products, the attribution estimates presented here should be considered with caution. The results were comparable for both types of surveillance, leading to the conclusion that passive data constitute a potential cost-effective complement to active data collection, especially interesting because the former encompass a greater number of sources. The model appeared robust to the type of surveillance, and provided that some

methodological aspects of the model can be enhanced, could also serve as a risk-based guidance tool for active surveillance systems.

Keywords: source attribution, *Salmonella*, surveillance, public health, zoonosis

1. Introduction

Worldwide, *Salmonella* is a priority public health issue and one of the main causes of bacterial foodborne diseases (Sofos, 2008). This concern prompted France to start a nationwide intervention in 1999 that targeted the most frequent serotypes, Enteritidis and Typhimurium, in the main known sources of *Salmonella* at an early stage of the food chain – in the breeding flocks of laying hens and broilers. The *Gallus gallus* plan has been successful at mitigating the number of human cases (Poirier et al., 2008), as have other similar plans implemented in other European countries (Edel, 1994; Wegener et al., 2003). However, *Salmonella* still is one of the most common recorded bacterial causes of foodborne gastroenteritis, as well as the most common cause of hospitalizations and deaths (InVS, 2004).

In this context, interest in methods to understand and assess management measures in various commodities and at different points in the agro-food chain has been growing. Source attribution is one of these tools; it allows identification, prioritization and assessment of the impact of interventions at the farm level and subsequently (Batz et al., 2005). The microbial subtyping attribution approach has been already successfully applied for such purposes in Denmark (Hald et al., 2004). The corresponding model requires intensive data collection on microbial types repartition both in human and in food animal sources, as well as food consumption data (Hald et al., 2004; Mullner et al., 2009). The data originally used in the model were ideally designed for such an approach – representative; based on an efficient typing system (systematic sero- plus phage typing); covering a broad spectrum of sources' and with sources contamination measured as prevalence for each bacterial type. Because surveillance systems are usually restrained by budget considerations or structural aspects and may have been designed before source attribution became a potential objective, the data collected may not fully meet the ideal requirements cited above.

The French integrated surveillance system for *Salmonella* covers the whole food-chain from the breeding farms to the human cases. Many actors contribute to this “mosaic” surveillance system (David et al., 2011). Salmonellosis cases are monitored by the National Reference Centre for *Salmonella* (NRC, Institut Pasteur, a collection of human strains sent on a voluntary basis by private and public medical laboratories), and the National Public Health Institute (InVS, mandatory declaration of enteric disease outbreaks). Regarding food-animal reservoirs of *Salmonella*, two parallel systems gather data on the contamination level in the different commodities at the national level. The Food Directory of the French Agriculture Ministry (DGAL) coordinates sampling plans at the farm level within the framework of European and national regulation-based surveillance. Those plans are designed so that surveillance is representative of national production and constitute an active surveillance of food-animal sources. The *Salmonella* network (SN), hosted by the French Agency for Food, Environmental and Occupational Health and Safety (Anses), gathers non-human *Salmonella* strains sent on a voluntary basis by public and private veterinary laboratories spread all over the national territories, ensuring a passive surveillance of *Salmonella* in the food-animal sources (Danan et al., 2011; David et al., 2011).

We thus had the opportunity to compare two surveillance datasets of food-animal sources of *Salmonella*. In each dataset, the sources contamination data were linked to human cases data described above. The first dataset, referred to as the active dataset, comprised data that were similar to the data used in the original attribution approach (representative prevalence data). The second dataset, referred to as the passive dataset, had some weaknesses (unknown representativeness, no access to prevalences but only types repartition in a commodity) but covered supplementary food-animal sources.

In this paper, using our previous proposal of an adapted parameterization for the Bayesian microbial subtyping attribution model (Hald et al., 2004; David et al., 2012), we assessed the impact of the deviation from the required data quality on source attribution results globally

and more particularly the impact of the sources contamination indicator used (prevalence in the active dataset *versus* proportion, that is, repartition of *Salmonella* types in a given commodity, in the passive dataset) and the number of sources included.

2. Materials and methods

2.1 *Data*

Two databases were used in the analyses, associating the human data with either passive or active surveillance data for the contamination of the sources. In the active dataset, source contamination data meet the requirements of the original model and this dataset constitutes our reference for attribution results. In the passive dataset, source contamination data had unknown representativeness and were measured as types proportions and not prevalences, so that this dataset does not meet the data requirements. However, it covers a broader range of sources. The targeted study period was 2005, but some data on sources contamination in the active surveillance dataset were retrieved from 2004 to 2007 according to the commodity. In the following paper, we will use “type” to designate both serotype (non Enteritidis and Typhimurium *Salmonella*) and subtype (for *S*. Enteritidis and Typhimurium) depending on the situation.

2.1.1 *Human cases*

The human cases data were retrieved from the NRC database for the year 2005. The cases included in the study correspond to confirmed non-Typhi salmonellosis cases for which the strain, or an analysis report, has been sent on a voluntary basis to the NRC by a public or private microbiology laboratory, together with epidemiological information. Travel-related cases and outbreak-related cases were excluded, based on the travel information available in the NRC database, and on the mandatory declarations of outbreaks centralized by InVS. The cases registered in the overseas “départements” and territories were excluded as well. For each outbreak, one case was kept in the dataset to represent it in the model (Hald et al.,

2004) This avoids (i) overestimation of the role of the source of the outbreak type when this type is specific to that source and (ii) underestimation of the role of the source of the outbreak type when this type occurs in several other sources.

The types that caused only a few (or no) infections or were not observed in the animal reservoirs were grouped into the categories “others” for serotypes and “other Typhimurium” or “other Enteritidis” for the Typhimurium and Enteritidis subtypes (see paragraph on subtyping below). Subsequently, the types’ composition of the “other” categories can be different for the human data and for the different sources. The corresponding human cases were consequently considered as not attributable (they do not have the potential to be attributed, or linked, to any source because their type has not been observed in any of the sources). They were excluded from the attribution process and affected to an “unknown source” category.

2.1.2 *Food-animal sources*

2,1,2,1 *Active dataset*

In the frame of European regulations, national prevalence studies of one year in length have been conducted from 2004 to 2007 in the main animal commodities either at the farm (layers in 2004–2005, broilers in 2005–2006, turkeys in 2006–2007) or at the slaughterhouse (pigs in 2006–2007) level. For each food-animal species, a random sample of flocks or carcasses representative of the national production has been tested for *Salmonella*: 519 layer flocks (faeces and environmental samples), 371 broiler flocks (faeces samples), 331 turkey flocks (faeces samples) and 1,166 pig carcasses (ileo-caecal lymph nodes). A flock or carcass was considered positive if at least one sample of *Salmonella* was positive, and all isolates were serotyped. The estimated prevalences were adjusted for the flock size for layers, broilers and turkeys. The flock size was known for layers and we used the barn capacity as an indicator of the flock size for broilers and turkeys. The weights were calculated as follow: $\omega_i = \frac{F_i}{\sum F_i}$

where F_i is either the flock size (layers) or the barn capacity (broilers and turkeys). More details on the sampling and analyses methodologies can be found in the corresponding European Food Safety Agency reports (Anonymous, 2007a, 2007b, 2008a, 2008b). The 2005 data on cattle were gathered within the framework of the national surveillance plan of antimicrobial resistance in indicator and zoonotic bacteria in cattle, a national annual monitoring plan conducted by the French Ministry of Agriculture aimed at assessing the prevalence of resistant bacteria in this commodity. The plan was based on the sampling of caecal content of 334 cattle carcasses (veal, young beef cattle and cull cows from beef herds) randomly selected at the abattoirs of the nine “départements” where the production is greatest, stratifying by abattoir size.

2,1,2,2 *Passive dataset*

In the SN database, we extracted the 2005 serotyping data for the 5 main commodities common to the active surveillance and added sheep, ducks, other poultry and sea products. The “sea products” category gathered data on fish, shellfish and seafood, the “other poultry” category gathered data on geese, guinea fowls, pheasants, quails and pigeons. We only included strains collected at the farm, on production animals and their environment, and at the abattoir, on whole carcasses. Strains relative to breeding animals, to diseased animals or for which the origin was unclear were excluded. For cattle, an overrepresentation of pathogenic strains has been identified at the farm level. As a consequence, for this commodity, we only included the strains collected at the abattoir.

2.1.3 *Subtyping*

To optimize the attribution process, Enteritidis and Typhimurium strains that represented more than 30% of the human cases each, were further subtyped. The subtyping was based on antimicrobial resistance profiles (Berge et al., 2003). One hundred and two strains out of 3,138 for Enteritidis and 92 strains out of 3,536 for Typhimurium were subtyped, representing

about 4% of all Enteritidis and Typhimurium human strains. For the food-animal sources, all strains of Enteritidis and Typhimurium collected in the active surveillance were analyzed for antimicrobial resistance, and all Typhimurium and Enteritidis strains collected in the passive surveillance by the SN were tested after exclusion of duplicates.

Antimicrobial susceptibility was determined by disk diffusion on Mueller-Hinton agar according to the guidelines of the Antibiogram Committee of the French Society for Microbiology (Soussy et al., 2000). All the routine tested antimicrobials common to the human and animal databases were included: amoxicillin, chloramphenicol, ceftazidime, gentamicin, kanamycin, nalidixic acid, streptomycin, sulfonamides, sulfamethoxazole-trimethoprim and tetracycline. For each antimicrobial, the strains were classified either as susceptible or as not susceptible when intermediate or resistant. The subtypes were defined through Multiple Correspondence Analysis (MCA) and mixed classification (Berge et al., 2003) applied on the antimicrobial profiles obtained.

2.1.4 *Consumption*

The consumption data are individual consumption data extrapolated to the French population. To calculate the consumption data, we referred to the INCA study (Etude Individuelle Nationale sur les Consommations Alimentaires), which is a national survey of individual food consumption conducted by Anses, in 1999 on 3,003 representative subjects above 3 years of age (Volatier, 2000). The INCA study results were used by pooling the consumption of all animal products related to a commodity over the study population and then extrapolated to the national population. Because the consumption of animal products varies substantially from year to year (following long term trends and food crises such as BSE), those results were then updated for 2005 by the mean of an index based on the annual data on the amount of the different meat types, eggs and sea products available on the market for consumption, published by the French Livestock Institute and the French

Interprofessional National Board for Sea and Fish-Farming Products. This allowed us to estimate as close as possible to the actual consumption of the respective food products for the French population.

2.2 *Bayesian microbial subtyping attribution model*

In the model, the expected number of human cases due to a given type *i* in a given source *j* (λ_{ij}) depends on the prevalence of the type *i* in the source *j* (p_{ij}), on the consumption of the source *j* in the general population (M_j) and on a source-dependant and a type-dependant parameters (a_j and q_i respectively). The expected number of observed human cases (o_i) is assumed to be Poisson distributed.

$$o_i \sim \text{Poisson}\,(\Sigma_j\, \lambda_{ij}) \text{ with } \lambda_{ij} = M_j\, p_{ij}\, q_i\, a_j \qquad (1)$$

where :

i = sero- or sub-type index;

j = source index;

o_i = observed number of human cases for sero- or sub-type i;

λ_{ij} = expected number of human cases due to type i in source j;

M_j = observed consumption in the general population of the source j in tons;

p_{ij} = observed prevalence of type i in source j;

q_i = unknown type-dependant parameter of type i;

a_j = unknown source-dependant parameter of source j.

As defined by Hald *et al*., (2004) the type-dependant factor (q_i) summarizes the characteristics of the type (such as survivability, virulence and pathogenicity) which determine its capacity to cause an infection and the source-dependant factor (a_j) summarizes the characteristics of the source (such as physical properties, preparation methods and processing procedures) which determine its capacity to act as a vehicle for *Salmonella* (Hald et al., 2004).

Since the model is overparameterized, it is necessary to introduce information on some parameters, and we used the adapted parameterization recommended in David et al. (2012). The type-dependant parameters relative to specific types (types only present in a unique food-animal source) other than Enteritidis and Typhimurium were set to the following data-based value:

$$q_i = \frac{o_i}{\sum_i o_i} \times \frac{1}{\sum_j p_{ij}} \quad (2)$$

where o_i is the observed number of cases due to type *i* and p_{jj} is the prevalence in the sources. It should be noted that $\sum_j p_{ij} = p_j = p_i$ because the type *i* is only present in its unique source j. We thus used the percentage of human cases per point of prevalence in the source, as an indicator of the capacity of that particular type to cause an infection. The prior distributions for the other type-dependant parameters and the source-dependant parameters are uniform distributions with 0 as the lower value and an upper value defined to encompass all possible values for the parameter. This parameterization requires the availability of at least as many specific types not corresponding to Enteritidis and Typhimurium as there are different sources, which was the case for both datasets.

Since not all *S.* Typhimurium and *S.* Enteritidis isolates from human cases were subtyped, an allocation step was necessary to extrapolate from the observed subtypes' distribution among the subtyped strains to the strains that were not subtyped. Gamma distributions were used to reflect the uncertainty distributions for the proportions of cases per subtype. The proportions used for the allocation within a serotype are as follows:

$$\frac{Gamma(o_i;1)}{\sum_i Gamma(o_i;1)} \quad (3)$$

where o_i is the observed number of cases of subtype *i* (Hald et al., 2004).

The directed acyclic graph (DAG) relative to the Bayesian model is presented in Figure 1. It is divided into three parts to take into account the differences between the types. A general part represents the model common to all types that are neither SE or ST subtypes, nor specific types, and the two other parts represent the specificities linked to the specific types (parameterization) and to SE and ST subtypes (allocation).

When considering the passive dataset, since prevalences were not available, proportions were used instead.

2.3 *Analyses*

We first focused on the impact of two data characteristics that are part of the differences between the two types of surveillance on the attribution results: the indicator used to measure the sources contamination (proportions *versus* prevalences) and the number of sources included. We then studied the impact of the type of surveillance globally on the attribution results.

To study the impact of the indicator, we used the active dataset for which both prevalences and proportions were available to measure the sources contamination. The model was applied using both indicators alternatively. To measure a potential effect of the indicator, we compared first the sources ranking and second, for each source, the posterior means of the attribution estimates and their 95% credibility intervals (95% CIs).

To study the impact of the number of sources included, we used the passive dataset, in which 4 additional sources were available in addition to the five main sources. We then estimated the model, considering alternatively the entire dataset (9 sources) and only the five main sources. The impact of the number of sources was compared based on the total

number of attributable cases, the sources ranking and for each source, the posterior mean of the source attribution estimates and the corresponding 95% CI.

To study the global impact of the type of surveillance (sources contamination) on the attribution results, we applied the model alternatively to each of the datasets. The results were compared based on the number of attributable cases, on the ranking of the sources, on the posterior means and 95% CIs of the source attribution estimates for each source, on the source- and type-dependant parameters estimates and on the number of cases with unknown source.

Analyses were performed using Winbugs 1.4.3® 2007 (Lunn et al., 2000).

3. Results

3.1 *Description of the datasets*

A total of 9,076 domestic sporadic confirmed human cases were included in the study. Enteritidis and Typhimurium were the most frequent serotypes, representing respectively 35% and 39% of the cases. The subtyping resulted in 9 subtypes both for Enteritidis and Typhimurium. Some of those subtypes were just human or food-animal sources related. The main characteristics of the subtypes are summarized in Tables 1 and 2.

3.1.1 *Active dataset*

Among the 9,076 included cases, 5,938 were attributable to the five considered food sources, spread between 28 serotypes, 5 Typhimurium subtypes and 3 Enteritidis subtypes. The overall observed prevalences in the sources were 30.8% for layers, 18.2% for pigs, 13.8% for turkeys, 8.2% for broilers and 2.4% for cattle. The types proportions among the human cases and in the sources are presented in Table 3a. Enteritidis and Typhimurium represented over 75% (47.1 % and 30.4%, respectively) of the attributable human cases.

The human cases were mostly associated to SE-multiS (35.2%), ST1 (17.5%), SE1 (10.4%), ST4 (6.5%) and Infantis (2.8%). Only 2 types were observed in cattle, Indiana and Mbandaka, the latter representing 62.5% of all cattle strains. Derby was the most frequently observed type both in turkeys and pigs. Enteritidis was mainly observed in the layers source (7.9%) and Typhimurium, if ubiquitous, was mostly associated with pigs (38.7%) and to a lesser extent layers (14.3%) and broilers (13.4%).

This dataset comprised 15 specific types (Table 4), including one Typhimurium subtype and one Enteritidis subtype. All sources but cattle were represented among the specific types, pigs and layers being the sources that comprised the highest number of specific types.

3.1.2 *Passive dataset*

In total 6,527 human cases were attributable to the 9 included sources, spread between 37 serotypes, 5 Typhimurium subtypes and 4 Enteritidis subtypes. The number of strains included in the study was 617 for layers, 1,273 for broilers, 796 for turkeys, 213 for pigs, 71 for cattle, 104 for sheep, 40 for the sea products, 3,113 for ducks and 722 for other poultry. The distribution of types among the human cases and in the sources is presented in Table 3b.Serotypes Enteritidis and Typhimurium represented over 75% (47.6% and 27.7%, respectively) of the attributable cases. The five main types were almost the same as in the active dataset – SE-multiS (32.1%), ST1 (15.9%), SE1 (9.4%), ST4 (5.9%) and SE3 (4.7%). Poultry commodities were characterised by a high number of subtypes. Sheep was the source with the smallest number of types (6 types out of the other categories) and was strongly associated with the type S 61:k:1,5,7 of subspecies *diarizonae*, which represented 57.7% of the strains. Serotype Senftenberg was mainly associated with layers (19.8%), broilers (13.0%) and turkeys (14.7%). SE-multiS was particularly present in layers and broilers (16.9% and 16.6%). Serotype Indiana was associated with ducks (23.1%) and other poultry (31.0%). ST1 (22.5%), Derby (19.7%) and Montevideo (14.1%) were associated with cattle and Weltevreden with sea products (25.0%). Globally, serotype Enteritidis was mostly

found in layers (24.3%) and broilers (16.7%), whereas serotype Typhimurium was preferentially associated with other poultry (30.7%), pigs (23.5%), cattle (22.5%) and sea products (15.0%).

This dataset comprised 14 specific types (Table 4), including 2 Enteritidis subtypes. As for the active dataset, all sources except cattle were represented among the specific types and pigs and layers were the most frequently concerned, according to the number of specific types and the number of associated human cases. There were 4 specific types in common with the active dataset: SE2, Oranienburg and Havana, linked to layers; and Goldcoast, linked to pigs.

Finally, as expected, the types observed differed according to the dataset leading to a difference in the number of attributable cases, which was 10% higher in the passive dataset. When considering the 5 common food-animal species, the types' distribution differed between the two datasets for a given species, which was confirmed by Fisher exact tests. However, across the datasets Derby was always the most frequent type for pigs and also for Typhimurium globally. SE-multiS was the most frequent type in layers. Specific types were identified for all food sources except cattle (Table 4). However, there were few types in common between the datasets, and these were only observed in layers and pigs. Thus the two datasets gave divergent results of the food-sources contamination.

3.1.3 *Consumption data*

On the basis of the updated results of the INCA study, the consumption estimates for 2005 were 82,301 tons for layers (eggs), 84,842 tons for broilers, 18,967 tons for turkeys, 161,971 tons for pigs, 144,504 tons for cattle, 20,232 tons for sheep, 111,764 tons for sea products, 5,309 tons for ducks and 4,580 tons for other poultry (Table 3).

3.2 Impact of the data characteristics on the attribution estimates

Results presented correspond to runs of 100,000 iterations of the Gibbs sampler with a thin of 25 for 5 independent chains. Convergence diagnostics were satisfactory (Cowles and Carlin, 1996; Brooks and Gelman, 1998; Brooks and Roberts, 1998b). From these runs, parameters estimates (posterior means, and posterior 95% CIs) were computed from the last 50,000 iterations.

3.2.1 *Source contamination indicator: proportion vs prevalence*

Based on the active dataset, using proportions instead of prevalence as indicator for sources contamination gave the same ranking of the sources (Figure 2), layers being ranked number one source, and cattle ranked last. The numbers of cases attributed to each source differed significantly only for broilers. Using prevalence *versus* proportion led to a significant decrease in the attribution posterior means of 58.3% (427 cases attributed *versus* 178) for broilers and to a non-significant decrease of 24.9% (3,060 cases attributed *versus* 2,297) for layers. For the three other food-sources, a non-significant increase of 26.9% for pigs, 12.0% for cattle and 57% for turkeys was observed.

3.2.2 *Increasing sources' number*

Based on the passive dataset, when including 4 supplementary sources to the 5 initial sources, the total number of attributable cases increased slightly, leading to 1.1% newly attributed cases. Moreover, even though the ranking of the 5 common sources remained unchanged, 25.5% of the cases initially attributed to one of the 5 common sources were reattributed to one of the 4 additional sources (Figure 3). Thus, the number of attributed cases decreased by 46.6% for cattle, 39.9% for pigs, 24.3% for broilers and 19.0% for layers, turkeys being the only exception with an increase of 2.8%. This difference was significant only for pigs (1,728 attributed cases; 95%CI 1,345–2,094 *versus* 1,039; 95%CI 7,84.9–1,310).

3.3 Impact of the type of surveillance on the attribution results

The model was applied separately to the active and passive datasets, with prevalence and proportion as source contamination indicator and 5 and 9 sources included, respectively. To assess the adequacy of the model, estimated numbers of attributed cases per type and observed numbers of cases were compared. As can be seen in Figure 4, the model fit was satisfactory for both datasets. To be able to compare the estimated number of cases, the observed numbers of cases per Enteritidis and Typhimurium subtype were replaced by deterministically allocated numbers (the subtypes' distribution observed among the subtyped cases was deterministically applied to allocate the cases with unknown subtype). As expected, the estimated total number of cases was higher with the passive dataset than the passive dataset: 7,122 attributable cases (95% CI: 6,719–7,528) *versus* 5,746 (95% CI: 5,307–6,172), respectively. However, for the passive dataset, the estimated number was overestimated with regard to the 6,527 attributable cases.

With the active dataset (Figure 5), the top ranked source of infection was layers (eggs) to which 53.3% (95% CI, 46.0–60.0) of the estimated cases were attributed corresponding to 3,060 attributed cases, followed by pigs, to which 25.7% (95% CI, 20.7–31.1) of the cases were attributed (1,477 attributed cases). The third ranked sources were broilers and turkeys, which did not differ significantly, representing 7.4% and 12.4% of the cases respectively (95% CI, 5.6–9.5 and 6.9–19.2) corresponding to 427 and 714 attributed cases, respectively. Cattle was a minor source, comprising only 1.2% of the cases (69 attributed cases). The cases whose origin remained unknown, were as numerous as the expected number of cases attributed to layers.

Using the passive dataset led, in term of attributed cases, to comparable results concerning layers (2,328 *versus* 3,060 for active dataset), pigs (1,039 *versus* 1,477), turkeys (336 *versus* 714) and cattle (133 *versus* 69). However, for broilers the number of attributed cases was significantly higher (1,410 *versus* 427) so that this source was comparable to pigs (1,410;

95% CI, 1,026–1,821 *versus* 1,039; 95% CI, 785–1,310) in importance. The ranking of the sources indicated layers as the most important source of infection, broilers and pigs as second most important sources, all the other sources (turkeys, cattle and the four additional) following without distinction (as the expected numbers of attributed cases were not significantly different). As for the results obtained with the active dataset, the cases of unknown origin were as numerous as the cases attributed to layers.

The source-dependant parameter estimates for the five common sources were quite similar between the two datasets, except for cattle (Table 5). Ranking of the sources based on their ability as vectors of *Salmonella* could not be concluded from the active dataset because of the wide posterior credibility interval related to cattle (first ranking). However, with the passive dataset, the highest source-dependant parameter posterior means were relative to ducks (1.37) and other poultry (1.06).

Unlike the source-dependant parameters, the type-dependant parameters estimates were systematically different, mostly significantly, between the two datasets; estimates for the passive dataset were lower. The only exception was serotype Stourbridge for which the estimates were coherent for both datasets. Another global feature was that the 95% CIs were narrower for Enteritidis and Typhimurium subtypes in the passive dataset. When considering the 10 top ranked type-dependant parameter estimates for each dataset (Table 6), 6 types appeared in both lists: SE1, SE-multiS, Stourbridge, Havana, Oranienburg and Napoli. However, only the parameter for Stourbridge had the same value, the calculated value in the active dataset (5.40) being in accordance with the estimate in the passive dataset (5.41; 95% CI, 2.42–10.72). For serotype Napoli, if the values were different, the ranking was the same in both cases (10^{th}). Heidelberg, ST2 and SE2 ranked high in the active dataset but low in the passive dataset. Finally, all Enteritidis subtypes but one were in the top ten for each dataset, whereas for Typhimurium only ST2, a subtype that gathers multi-resistant strains resistant to nalidixic acid, was present in the top ten of the active dataset.

4. Discussion

The model estimates appeared to be robust to a deviation from the required quality of the data, due to a passive design of the sources data collection (but with a national coverage ensured), and to the consequent use of proportions instead of prevalences for the contamination indicator, but sensitive to the number of sources included. The results confirmed layers as the top source for *Salmonella* and were globally coherent for both datasets except for broilers. This was probably linked to the higher frequency in this source of Typhimurium and Enteritidis subtypes in the passive dataset (27.4% of the strains globally *versus* 5.1% in the active dataset). In both configurations, the number of cases with unknown source represented about a third of all cases and with the passive dataset, the 4 additional sources contributed to a non-negligible proportion of all cases (around a quarter of all cases). Because source- and type-dependant parameters measured different quantities according to the dataset and because of the wide posterior credibility intervals, those parameters were difficult to interpret. However, it can be extrapolated from the few convergent results that birds appear to be a good vehicle for *Salmonella* and that Enteritidis has a high capacity for infection, whichever the subtype, as do some other serotypes such as Stourbridge, Havana, Oranienburg and Napoli.

When considering the active dataset complemented by the passive dataset for the 4 additional sources, relative importance of the sources deduced from the posterior estimations were in accordance with those from Denmark (Hald et al., 2004; Havelaar et al., 2007): layers were the most important source, followed by pigs and broilers, turkeys, ducks and cattle, which were not significantly different.

The significant difference observed in broilers between the two datasets could be related to the higher frequency of Typhimurium and Enteritidis among the strains in the passive dataset, the source-dependant factor being similar between both datasets. Another possibility is that some samples corresponding to abattoir data for broilers may correspond to spent hens, despite our efforts to discriminate these (many strains have been excluded because of incomplete or unclear information). Such a difference in the results obtained for *Salmonella* with active and passive data, especially for broilers (compared to pigs), has been recently shown for antimicrobial resistance (Mather et al., 2009). It would be interesting to further explore the reasons for this difference, which seems to be specific to broilers.

The number of cases with unknown origin was as high as the number of expected cases attributed to the main source, layers. A non-negligible part of these "unknown" cases belonged to Enteritidis and Typhimurium "other" categories. They correspond to subtypes not observed in the animal sources considered in this study. The restricted number of subtypes common to the cases and the sources could be linked either to a problem of detection due to the small number of human strains subtyped, or to the absence of some potential sources. Moreover, the subtyping based on antimicrobial resistance profiles, a phenotypic method, is not ideal. Namely, the antimicrobial resistance profile of a strain could evolve along the food chain, because of, for example, selection by disinfectants (cross-resistance) (Braoudaki and Hilton, 2004; Condell et al., 2012) or the impact of stress encountered at different stages of the food chain (Poole, 2012). Furthermore, a study by Mather *et al.* (2012) raises the possibility that the resistance diversity in human cases may not be exclusively of animal origin, based on observations of Enteritidis DT104 strains. Besides, some resistance traits are known to be linked to virulence factors, especially in *S.* Typhimurium (Martinez and Baquero, 2002; Mølbak, 2004; Foley and Lynne, 2008), what can influence the type-parameter. This makes it essential to be able to distinguish that parameter between subtypes within a serotype. Having based the subtyping on antimicrobial resistance in this study makes it impossible to assess the attribution of cases according to their resistance status.

For future studies a new genotypic subtyping method, based on CRISPR polymorphisms (Fabre et al., 2012), is now available. It will be more appropriate to assess strain linkage (rather than using a phenotypic method based on resistance traits). Also, such a method can be used extensively to type a greater number of strains each year and would not be an obstacle for the source attribution of resistant *Salmonella*. Regarding the potentially missing sources, we only considered foodborne transmission, thereby excluding sources such as pets (both domestic and exotic) which are known sources of human salmonellosis (Woodward et al., 1997; Bellido Blasco et al., 1998; Mermin et al., 2004; De Jong et al., 2005; Finley et al., 2006; Marcus et al., 2007), and person-to-person transmission (Todd et al., 2008). Moreover, among the potential food sources we focused on some animal food-sources which are the most frequent sources of human salmonellosis but not the only ones. Plants (Brandl, 2006; Elviss et al., 2009) and other non-animal products (Kirk et al., 2008) have been implicated as a source for salmonellosis. Finally, imported products, which represent from 15% (eggs) to 60% (lamb) of the national consumption according to the food-animals products are a potential major source for the human cases but could not be taken into account in this analysis. Because of this large category of cases with unknown sources, the attribution estimates, although valuable for studying the methodological differences between passive *versus* active data, have to be interpreted with caution regarding their representativeness and accuracy.

Direct comparison between the source-dependant parameters for the two datasets was difficult. The contamination indicator was prevalence in one case and proportion in the other case, so that the source-dependant parameter did not measure the same quantity, which is also true for the type-dependant parameter. The type-dependant parameters were almost systematically different when using one or the other dataset, even when (rarely) the ranking was coherent. However, the source-dependant parameter estimates were in good agreement between both datasets, except for cattle. Hald et al. (2004, p267) underlined that these factors were only multiplication factors that helped arrive at the most probable solution given

the observed data. This was actually the case in our study, although it was difficult to draw any firm conclusion on the posterior distributions, especially for the type-dependant parameter. The type- and source- parameters (q_i and a_j) are conceived as 'black boxes', and they allow differences between types and between sources to be taken into account, which is of primary importance (Blaser and Newman, 1982; D'Aoust, 1989; Sarwari et al., 2001; Coleman et al., 2004; Bollaerts et al., 2008; Jones et al., 2008). However, their nature is not defined. Thus, to better specify their prior distributions, it would be essential to give a definition to these parameters and to be able to use exogenous information, such as dose-response relationship, infective dose and pathogenicity (Bollaerts et al., 2008; Jones et al., 2008). It would also be of utmost importance to take the potential interactions between those factors into account. Interactions would reflect what is known about the specificity of the dose-illness relationship for a serotype-food matrix combination (Bollaerts et al., 2008). Bayesian models, as for all models, are conceptual simplifications of the real world and the modelling of latent traits (such as *a* and *q* parameters in this example) is questionable if experts do not even agree on the nature of these unobserved parameters. Our aim in this work was to study the behaviour of the Hald model according to the data quality. So we adapted the parameterization while keeping the original outlines of the model.

The parameterisation of the model relied on specific types for which the type-dependant parameter was set to a data-based value. Only a few of those types were common to both datasets. This raises the question of their definition. For now they are defined on an observational basis, but biological and ecological criteria could be used to refine the choice of those types and introduce complementary information to calculate the value of the associated type-dependant parameters and thus avoid any negative impact on the attribution results.

We may have taken into account the data generating process in the model by putting some priors on the prevalence of types, as has been recommended by Muellner *et al*,. The

sampling protocol of the active dataset was detailed enough to define those priors, but this would have introduced greater uncertainty on the results and would have led to a different structure of the model. Moreover, we couldn't have used this model structure for the passive dataset, as the sampling protocols are not known in this system.

In France, thanks to the European regulatory framework and to long running laboratory-based surveillance systems (the NRC began its activities in 1947 and the SN in 1997), all the data required for the chosen approach were available and the typing methods were harmonized for all datasets, as a consequence of the on-going cooperation for outbreaks investigations. Besides, we disposed for the sources both of active and passive surveillance data, an ideal situation to assess the relative advantages of using the data generated by those two surveillance systems in the Bayesian microbial subtyping attribution approach. Because the data were fragmented between several actors, gathering all those data required a large scope collaboration, which is a key element to ensure the success of such projects (Batz et al., 2005; ICMSF, 2006). However, we had to deal with some insufficiencies in the data. The human data were retrieved through voluntary participation of private and public laboratories, and thus were neither exhaustive nor perfectly representative of all cases. Moreover, the information on travel was scarce. The NRC relies on a stable and nationwide network of public and private laboratories, which is estimated to have good representativeness (a national survey conducted in 2008 concluded that the NRC covered 66.3% of all laboratory confirmed cases). Therefore, no major bias should have occurred. Another problem was that, due to logistical constraints, only a fraction of the human cases of serotype Enteritidis and Typhimurium were subtyped. As a consequence, it is likely that less prevalent subtypes have not been observed in the small sample of cases tested. Furthermore, around 96% of the Enteritidis and Typhimurium cases had to be reallocated (through Gamma distributions) in the model, which led to wide credibility intervals. This can be corrected in future studies by changing the subtyping method (for example a method

based on CRISPR polymorphisms that makes it logistically possible to routinely test a high number of strains). Regarding the sources contamination, the representativeness of the passive surveillance data cannot be assessed. In the active dataset the sampling point varies from one commodity to another. Poultry were tested at the farm, pigs were tested at the slaughterhouse, which could have led for the latter to underestimation of the prevalence but overestimation of the variety of serovars compared to what could have been observed at the farm level (Beloeil et al., 2004). Also for the active dataset, there is a lack of time consistency between the food-sources data and the human cases data, the sources data being posterior to the cases data that they should explain. We could not avoid this potential bias in this exploratory study. This is a further reason to consider the attribution results with caution. Such a problem should not occur in future studies because the main sources (layers, broilers, turkeys and pigs) are now monitored on a yearly basis as required by the European Zoonosis Directive (David et al., 2011). Finally, the food consumption data were recovered through interviews of people aged more than 3 years, whereas the cases were of all ages. Because it is known that the risk factors can be different for young children (Bellido Blasco et al., 1998) and because we could not exactly assess the number of cases under 3 year of age (age categories in the database were 0–1 year old and 1–5 years old, so that the number of 0-3 years old cases is between 200 and 1,590 out of the 9,076 cases included in the database), this could have caused a bias in our results, although probably minor. Including non-food sources (such as pets) might be particularly important in the case of children. Also, measuring exposure as a frequency of consumption rather than as quantities consumed would be more accurate. This could not be done here, though the frequency was available in the INCA dataset, because of the necessary update of the consumption data collected in 1999 based on availability of the market (tonnage) data.

As the active monitoring of food-sources is costly and not always developed for the production level (farm) (Batz et al., 2005; Kirk et al., 2008; Sofos, 2008), in some countries passively acquired surveillance data can be the only alternative available. Based on the

current study, passive data appear to be a potential complement for active datasets, even moreso when considering the costs of data acquisition. A systematic active surveillance system of food sources is costly and thus can only encompass a limited number of sources, whereas the costs of a passive system are less important and independent of the number of sources monitored (Havelaar et al., 2007), which are important features in the design of a public health tool. Including additional sources through a passive surveillance system could have several advantages. Such a system would encompass more potential sources and could thus avoid mis-attributions. Moreover, as the contamination of sources at the farm level evolves according to public health interventions, enforcement of regulations and changes in hygiene practices (Rostagno et al., 2005; Denagamage et al., 2007), and as the exposure of the population to the sources also changes (Combris, 1997; Allard, 2002; Desenclos et al., 2002), including more sources can help detect the emergence of newly important sources not institutionally monitored. Thus, when applied with both active and passive data, the model could be a useful tool to contribute to the adjustment of the institutional active surveillance of the potential food-sources, besides its potential role for targetting and evaluating public health interventions. If only passive surveillance data are available for the sources contamination, valuable results can still be obtained. Of course in this case, conclusions will have to be made with caution and rather on the sources ranking than on the estimated number of attributed cases, but nevertheless will help local risk assessors to communicate with risk managers and stakeholders.

With regard to an optimal use of the model for source attribution, key elements include the subtyping method and the number of sources included, as well as an extensive national coverage of both cases and sources. In addition and uppermost, some features of the model should be improved, principally the definition of the type- and source-parameters and the consideration of their interactions.

5. Conclusion

We here have addressed the problem of the impact of data quality on the attribution of human cases to food-animal sources of *Salmonella* when using the bayesian microbial subtyping approach. Our results indicate that it is possible to obtain reasonably reliable attribution estimates based on passively acquired data for the sources contamination. In addition, such cost-effective data can complement actively-acquired and representative prevalence data. The data we used have limitations which could be addressed by an improvement in human and animal datasets through enhancement of the subtyping system and the inclusion of other potential contamination sources, especially imported products. The Bayesian tool could also be improved with respect to the nature, dimension and interactions of the type- and source-parameters, which are of primary importance in the approach. Finally, the parallel use of active and passive data could contribute to a risk-based focus of active surveillance policy.

6. Acknowledgements

This study was supported by the Délégation Générale de l'Armement, Defence Ministry. The authors wish to thank all members of the MedVetNet WP28 Working Group for the useful discussions, the Office for foodborne zoonoses and food microbiology of the General directorate for food from the French Ministry of Agriculture, all members of the CEB team at the French Agency for Food, Environmental and Occupational Health and Safety, Maisons Alfort, and the Salmonella NRC at the Pasteur Institute for the subtyping analyses undertaken, the members of the HQPAP, EBEP and EBEAC units at the French Agency for Food, Environmental and Occupational Health and Safety, Ploufragan, N. Jourdan-Da Silva from the National Public Health Institute (InVS) for providing the outbreak data and advice on their use and Claire Chauvin from the French Agency for Food, Environmental and Occupational Health and Safety for guidance throughout this work.

7. Conflict of interest

None for any of the authors.

Role of the funding source : none

8. References

Allard, D.G., 2002. The 'farm to plate' approach to food safety - Everyone's business. Canadian Journal of Infectious Diseases 13, 185-190.

Anonymous, 2007a. Report of the Task Force on Zoonoses Data Collection on the Analysis of the baseline study on the prevalence of Salmonella in holdings of laying hen flocks of Gallus gallus. The EFSA Journal 97, 1-84.

Anonymous, 2007b. Report of the Task Force on Zoonoses Data Collection on the Analysis of the baseline survey on the prevalence of Salmonella in broiler flocks of Gallus gallus, in the EU, 2005-2006 - Part A: Salmonella prevalence estimates The EFSA Journal 98, 1-85.

Anonymous, 2008a. Report of the Task Force on Zoonoses Data Collection on the Analysis of the baseline survey on the prevalence of Salmonella in slaughter pigs, in the EU, 2006-2007 - Part A: Salmonella prevalence estimates. The EFSA Journal 135, 1-111.

Anonymous, 2008b. Report of the Task Force on Zoonoses Data Collection on the Analysis of the baseline survey on the prevalence of Salmonella in turkey flocks, in the EU, 2006-2007 - Part A: Salmonella prevalence estimates The EFSA Journal 134, 1-91.

Batz, M.B., Doyle, M.P., Morris Jr., J.G., Painter, J., Singh, R., Tauxe, R.V., Taylor, M.R., Wong, D.M.A.L.F., 2005. Attributing illness to food. Emerg. Infect. Dis. 11, 993-999.

Bellido Blasco, J.B., Gonzalez Cano, J.M., Galiano, J.V., Bernat, S., Arnedo, A., Gonzalez Moran, F., 1998. Factors associated with sporadic cases of salmonellosis in 1- to 7-year-old children. Study of cases and controls. Gac. Sanit. 12, 118-125.

Beloeil, P.A., Chauvin, C., Proux, K., Madec, F., Fravalo, P., Alioum, A., 2004. Impact of the Salmonella status of market-age pigs and the pre-slaughter process on Salmonella caecal contamination at slaughter. Vet. Res. 35, 513-530.

Berge, A.C.B., Atwill, E.R., Sischo, W.M., 2003. Assessing antibiotic resistance in fecal Escherichia coli in young calves using cluster analysis techniques. Prev. Vet. Med. 61, 91-102.

Blaser, M.J., Newman, L.S., 1982. A review of human salmonellosis: I. Infective dose. Rev. Infect. Dis. 4, 1096-1106.

Bollaerts, K., Aerts, M., Faes, C., Grijspeerdt, K., Dewulf, J., Mintiens, K., 2008. Human salmonellosis: estimation of dose-illness from outbreak data. Risk Anal. 28, 427-440.

Brandl, M.T., 2006. Fitness of human enteric pathogens on plants and implications for food safety. Annu Rev Phytopathol 44, 367-392.

Braoudaki, M., Hilton, A.C., 2004. Adaptive resistance to biocides in Salmonella enterica and Escherichia coli O157 and cross-resistance to antimicrobial agents. J. Clin. Microbiol. 42, 73-78.

Brooks, S.P., Gelman, A., 1998. General methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics 7, 434-455.

Brooks, S.P., Roberts, G.O., 1998b. Convergence assessment techniques for Markov chain Monte Carlo. Statistics and Computing 8, 319-335.

Coleman, M.E., Marks, H.M., Golden, N.J., Latimer, H.K., 2004. Discerning strain effects in microbial dose-response data. J. Toxicol. Environ. Health A 67, 667-685.

Combris, P., 1997. The consumption of animal products in France : current tendencies and future perspectives. Productions Animales 10, 267-274.

Condell, O., Iversen, C., Cooney, S., Power, K.A., Walsh, C., Burgess, C., Fanning, S., 2012. Efficacy of biocides used in the modern food industry to control salmonella enterica, and links between biocide tolerance and resistance to clinically relevant antimicrobial compounds. Appl. Environ. Microbiol. 78, 3087-3097.

Cowles, M.K., Carlin, B.P., 1996. Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review. Journal of the American Statistical Association 91, 883-904.

D'Aoust, J., 1989. Salmonella. In: Doyle, M.P. (Ed.), Foodborne Bacterial Pathogens. Marcel Dekker, New York, 327-445.

Danan, C., Baroukh, T., Moury, F., Jourdan, D.A.S.N., Brisabois, A., Y, L.E.S., 2011. Automated early warning system for the surveillance of Salmonella isolated in the agro-food chain in France. Epidemiol. Infect. 139, 736-741.

David, J.M., Danan, C., Chauvin, C., Chazel, M., Souillard, R., Brisabois, A., Weill, F.X., Jourdan Da Silva, N., Picherot, M., Guillemot, D., Sanders, P., 2011. Structure of the French farm-to-table surveillance system for *Salmonella*. Revue Méd. Vét. 162, 489-500.

David, J.M., Guillemot, D., Bemrah, N., Thebault, A., Brisabois, A., Chemaly, M., Weill, F., Sanders, P., Watier, L., 2012. The Bayesian Microbial Subtyping Attribution Model: Robustness to Prior Information and a Proposition. Risk Anal. DOI 10.1111/j.1539-6924.2012.01877.x

De Jong, B., Andersson, Y., Ekdahl, K., 2005. Effect of regulation and education on reptile-associated salmonellosis. Emerg. Infect. Dis. 11, 398-403.

Denagamage, T.N., O'Connor, A.M., Sargeant, J.M., Rajic?, A., McKean, J.D., 2007. Efficacy of vaccination to reduce Salmonella prevalence in live and slaughtered swine: A systematic review of literature from 1979 to 2007. Foodborne Pathog. Dis. 4, 539-549.

Desenclos, J.C., Vaillant, V., De Valk, H., 2002. [Food-borne infections: do we need further data collection for public health action and research]. Rev. Epidemiol. Sante Publique 50, 67-79.

Edel, W., 1994. Salmonella enteritidis eradication programme in poultry breeder flocks in The Netherlands. Int. J. Food Microbiol. 21, 171-178.

Elviss, N.C., Little, C.L., Hucklesby, L., Sagoo, S., Surman-Lee, S., de Pinna, E., Threlfall, E.J., 2009. Microbiological study of fresh herbs from retail premises uncovers an international outbreak of salmonellosis. Int. J. Food Microbiol.

Fabre, L., Zhang, J., Guigon, G., Le Hello, S., Guibert, V., Accou-Demartin, M., de Romans, S., Lim, C., Roux, C., Passet, V., Diancourt, L., Guibourdenche, M., Issenhuth-Jeanjean, S., Achtman, M., Brisse, S., Sola, C., Weill, F.X., 2012. CRISPR Typing and Subtyping for Improved Laboratory Surveillance of Salmonella Infections. PLoS ONE 7, e36995.

Finley, R., Reid-Smith, R., Weese, J.S., 2006. Human health implications of Salmonella-contaminated natural pet treats and raw pet food. Clin. Infect. Dis. 42, 686-691.

Foley, S.L., Lynne, A.M., 2008. Food animal-associated Salmonella challenges: pathogenicity and antimicrobial resistance. J. Anim. Sci. 86, E173–E187 doi:110.2527/jas.2007-0447.

Hald, T., Vose, D., Wegener, H.C., Koupeev, T., 2004. A Bayesian approach to quantify the contribution of animal-food sources to human salmonellosis. Risk Anal. 24, 255-269.

Havelaar, A.H., Bräunig, J., Christiansen, K., Cornu, M., Hald, T., Mangen, M.-J.J., Mølbak, K., Pielaat, A., Snary, E., Van Pelt, W., Velthuis, A., Wahlstro?m, H., 2007. Towards an integrated approach in supporting microbiological food safety decisions. Zoonoses and Public Health 54, 103-117.

ICMSF, I.C.o.M.S.f.F., 2006. Use of epidemiologic data to measure the impact of food safety control programs. Food Control 17, 825-837.

InVS, 2004. Morbidité et mortalité dues aux maladies infectieuses d'origine alimentaire en France. 192.

Jones, T.F., Ingram, L.A., Cieslak, P.R., Vugia, D.J., Tobin-D'Angelo, M., Hurd, S., Medus, C., Cronquist, A., Angulo, F.J., 2008. Salmonellosis outcomes differ substantially by serotype. J. Infect. Dis. 198, 109-114.

Kirk, M.D., McKay, I., Hall, G.V., Dalton, C.B., Stafford, R., Unicomb, L., Gregory, J., 2008. Food safety: foodborne disease in Australia: the OzFoodNet experience. Clin Infect Dis 47, 392-400.

Lunn, D.J., Thomas, A., Best, N., Spiegelhalter, D., 2000. WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. Statistics and Computing 10, 325-337.

Marcus, R., Varma, J.K., Medus, C., Boothe, E.J., Anderson, B.J., Crume, T., Fullertone, K.E., Moore, M.R., White, P.L., Lyszkowicz, E., Voetsch, A.C., Angulo, F.J., 2007. Re-assessment of risk factors for sporadic Salmonella serotype Enteritidis infections: A case-control study in five FoodNet Sites, 2002-2003. Epidemiol. Infect. 135, 84-92.

Martinez, J.L., Baquero, F., 2002. Interactions among strategies associated with bacterial infection: pathogenicity, epidemicity, and antibiotic resistance. Clin. Microbiol. Rev. 15, 647-679.

Mather, A.E., Haydon, D.T., Matthews, L., Mellor, D.J., Reid-Smith, R.J., Boerlin, P., Dutil, L., McEwen, S.A., Reid, S.W.J., 2009. Comparing active and passive surveillance of antimicrobial reistance: Are we getting the whole story? In, 12th International Symposium on Veterinary Epidemiology and Economics, Durban.

Mermin, J., Hutwagner, L., Vugia, D., Shallow, S., Daily, P., Bender, J., Koehler, J., Marcus, R., Angulo, F.J., 2004. Reptiles, amphibians, and human Salmonella infection: A population-based, case-control study. Clin. Infect. Dis. 38.

Mølbak, K., 2004. Spread of resistant bacteria and resistance genes from animals to humans - The public health consequences. Journal of Veterinary Medicine Series B: Infectious Diseases and Veterinary Public Health 51, 364-369.

Mullner, P., Jones, G., Noble, A., Spencer, S.E.F., Hathaway, S., French, N.P., 2009. Source attribution of food-borne zoonoses in New Zealand: A modified hald model. Risk Anal. 29, 970-984.

Poirier, E., Watier, L., Espie, E., Weill, F.X., De Valk, H., Desenclos, J.C., 2008. Evaluation of the impact on human salmonellosis of control measures targeted to Salmonella Enteritidis and Typhimurium in poultry breeding using time-series analysis and intervention models in France. Epidemiol. Infect. 136, 1217-1224.

Poole, K., 2012. Bacterial stress responses as determinants of antimicrobial resistance. J. Antimicrob. Chemother. 67, 2069-2089.

Rostagno, M.H., Hurd, H.S., McKean, J.D., 2005. Resting pigs on transport trailers as an intervention strategy to reduce Salmonella enterica prevalence at slaughter. J. Food Prot. 68, 1720-1723.

Sarwari, A.R., Magder, L.S., Levine, P., McNamara, A.M., Knower, S., Armstrong, G.L., Etzel, R., Hollingsworth, J., Morris J.G., J., 2001. Serotype distribution of Salmonella isolates from food animals after slaughter differs from that of isolates found in humans. J. Infect. Dis. 183, 1295-1299.

Sofos, J.N., 2008. Challenges to meat safety in the 21st century. Meat Sci. 78, 3-13.

Soussy, C.J., Carret, G., Cavallo, J.D., Chardon, H., Chidiac, C., Choutet, P., Courvalin, P., Dabernat, H., Drugeon, H., Dubreuil, L., Goldstein, F., Jarlier, V., Leclercq, R., Nicolas-Chanoine, M.H., Philippon, A., Quentin, C., Rouveix, B., Sirot, J., 2000. [Antibiogram Committee of the French Microbiology Society. Report 2000-2001]. Pathol. Biol. (Paris) 48, 832-871.

Todd, E.C., Greig, J.D., Bartleson, C.A., Michaels, B.S., 2008. Outbreaks where food workers have been implicated in the spread of foodborne disease. Part 5. Sources of contamination and pathogen excretion from infected persons. J. Food Prot. 71, 2582-2595.

Volatier, J.L., 2000. Enquête nationale sur les consommations alimentaires. In: Tec&Doc (Ed.) CREDOC-AFSSA-DGAL.

Wegener, H.C., Hald, T., Wong, D.L.F., Madsen, M., Korsgaard, H., Bager, F., Gerner-Smidt, P., Mølbak, K., 2003. Salmonella control programs in Denmark. Emerg. Infect. Dis. 9, 774-780.

Woodward, D.L., Khakhria, R., Johnson, W.M., 1997. Human salmonellosis associated with exotic pets. J. Clin. Microbiol. 35, 2786-2790.

Table 1. Characteristics of the Enteritidis subtypes defined on the basis of the antimicrobial profiles, passive and active datasets for the surveillance of *Salmonella* in France, 2005 (2004-2007 for the sources contamination in the active surveillance)

Subtype	Main charactersitics	Origin*	Number of observed human cases**
SE multiS	Pansusceptible strains	Cases, sources	2092
SE1	Resistance to nalidixic acid only	Cases, sources	615
SE2	Resistance to nalidixic acid associated with another resistance	Cases, sources	92
SE3	Resistance to amoxicillin	Cases, sources	307
SE4	Resistance to tetracycline	Sources	-
SE5	Multiresistance	Cases	31
SE6	Resistance to sulfonamides	Sources	-
SE7	Resistance to streptomycin	Sources	-
SE8	Resistance to chloramphenicol	Sources	-

* Cases refers to strains isolated from human cases and sources to strains isolated from food-animal sources (passive or active dataset)
** For Enteritidis and Typhimurium subtypes, deterministically reallocated numbers of cases are used in place of the observed numbers of cases.

Table 2. Characteristics of the Typhimurium subtypes defined on the basis of the antimicrobial profiles, passive and active datasets for the surveillance of *Salmonella* in France, 2005 (2004-2007 for the sources contamination in the active surveillance)

Subtype	Main characteristics	Origin*	Number of observed human cases**
ST multiS	pansusceptible	Cases	1115
ST1	Pentaresistance AmxCSSssTe	Cases, sources	1038
ST2	AmxCSSssTe associated with nalidixic acid resistance	Cases, sources	154
ST3	AmxCSSssTe associated with sulfamethoxazole-trimethoprim resistance	Cases, sources	154
ST4	AmxCSSssTe associated with kanamycin resistance	Sources	-
ST5	susceptible to streptomycine	Cases	615
ST6	Resistance to streptomcyine and tetracycline	Cases, sources	384
ST7	Resistance to streptomycine	Cases, sources	77
ST8	Résistance to streptomycine and nalidixic acid	Sources	-

* Cases refers to strains isolated from human cases and sources to strains isolated from food-animal sources (passive or active dataset)
** For Enteritidis and Typhimurium subtypes, deterministically reallocated numbers of cases are used in place of the observed numbers of cases.

Table 3a. Number of cases and prevalence in the sources of *Salmonella* Sero- and Subtypes in the different food-animal Species and amount of foodstuffs available for consumption in France in 2005 (2004-2007 for the sources contamination), active surveillance dataset.

Serotype**	Number of observed human cases*	Active dataset Prevalence in sources (%)				
		Layers	Broilers	Turkeys	Pigs	Cattle
ST1	1038	1.25%	0.00%	0.00%	3.60%	0.00%
ST2	154	0.00%	0.00%	0.00%	0.09%	0.00%
ST3	154	0.00%	0.00%	1.88%	0.17%	0.00%
ST4	384	1.88%	0.00%	0.00%	0.77%	0.00%
ST5	77	2.82%	0.07%	0.00%	2.40%	0.00%
Other-Typhimurium	1730	0.31%	0.00%	0.75%	0.51%	0.00%
SE-multiS	2092	2.66%	0.25%	0.41%	0.09%	0.00%
SE1	615	0.41%	0.00%	0.20%	0.00%	0.00%
SE2	92	0.20%	0.00%	0.00%	0.00%	0.00%
Other-Enteritidis	338	0.82%	0.12%	0.10%	0.09%	0.00%
Infantis	168	5.49%	0.07%	0.00%	0.43%	0.00%
Derby	144	0.00%	0.00%	3.55%	6.52%	0.00%
Hadar	115	0.08%	1.55%	0.76%	0.09%	0.00%
Newport	105	0.00%	0.49%	0.77%	0.17%	0.00%
Virchow	100	0.86%	1.53%	0.00%	0.00%	0.00%
Napoli	90	0.00%	0.07%	0.42%	0.00%	0.00%
Agona	85	0.98%	0.00%	0.82%	0.43%	0.00%
Indiana	75	0.00%	0.34%	0.75%	0.09%	0.30%
Brandenburg	71	0.00%	0.00%	0.00%	0.34%	0.00%
Stourbridge	42	0.00%	0.00%	0.00%	0.09%	0.00%
Bovismorbificans	40	0.00%	0.00%	0.00%	0.09%	0.00%
Coeln	36	0.73%	0.00%	0.11%	0.09%	0.00%
Anatum	33	5.25%	0.21%	0.00%	0.17%	0.00%
Montevideo	32	1.92%	1.31%	0.00%	0.09%	0.00%
Oranienburg	31	0.09%	0.00%	0.00%	0.00%	0.00%
Bredeney	31	0.00%	0.00%	0.72%	0.26%	0.00%
Heidelberg	30	0.00%	0.02%	0.00%	0.00%	0.00%
S IIIa 48:z4,z23:-	24	0.00%	0.00%	0.00%	0.18%	0.00%
Senftenberg	17	0.27%	0.07%	0.00%	0.00%	0.00%
Havana	16	0.04%	0.00%	0.00%	0.00%	0.00%
Ohio	12	0.00%	0.07%	0.00%	0.00%	0.00%
Mbandaka	9	2.81%	0.88%	1.58%	0.09%	1.50%
Goldcoast	9	0.00%	0.00%	0.00%	0.09%	0.00%
Bareilly	8	2.95%	0.00%	0.00%	0.00%	0.00%
Muenster	7	0.00%	0.00%	0.38%	0.00%	0.00%
Ajiobo	2	0.00%	0.00%	0.04%	0.00%	0.00%
Others	1070	9.74%	1.72%	0.81%	1.27%	0.60%

Total number / prevalence	9076	30.82%	8.23%	13.78%	18.18%	2.40%
Food amount		82 301	84 842	18 967	161 971	144 504

*the number of human cases corresponds to observed cases for all serotypes except Enteritidis and Typhimurium for which deterministically allocated numbers of cases are shown.
**the serotypes that are present in the sources for the passive dataset only are bolded (i.e. they were not observed with the active surveillance in the sources)

Table 3b: Number of cases and prevalence in the sources of *Salmonella* Sero- and Subtypes in the different food-animal Species and amount of foodstuffs available for consumption in France in 2005, active surveillance dataset.

Serotype**	Number of observed human cases*	Passive dataset Proportion in sources (%)								
		Layers	Broilers	Turkeys	Pigs	Cattle	Sheep	Sea products	Ducks	Other poultry
ST1	1038	3.24%	1.34%	4.77%	4.23%	22.54%	2.88%	7.50%	0.00%	1.66%
ST2	154	0.00%	1.57%	0.00%	1.41%	0.00%	0.00%	7.50%	0.00%	14.54%
ST3	154	0.00%	0.31%	4.90%	3.29%	0.00%	0.00%	0.00%	0.00%	6.09%
ST4	384	4.05%	1.34%	0.00%	14.55%	0.00%	1.92%	0.00%	1.00%	5.12%
ST5	77	4.05%	1.26%	2.89%	0.00%	0.00%	1.92%	0.00%	9.12%	3.32%
Other-Typhimurium	1730	1.30%	3.85%	2.89%	0.00%	0.00%	0.00%	0.00%	13.49%	7.76%
SE-multiS	2092	16.86%	16.58%	0.88%	0.00%	2.82%	0.00%	0.00%	1.54%	2.22%
SE1	615	2.11%	0.16%	1.88%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
SE2	92	4.05%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
SE3	307	1.30%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Other-Enteritidis	31	2.92%	1.02%	0.00%	0.00%	0.00%	3.85%	0.00%	0.00%	0.00%
Infantis	168	1.46%	3.93%	1.01%	6.57%	2.82%	0.00%	5.00%	4.85%	0.55%
Derby	144	0.65%	2.20%	9.05%	16.43%	19.72%	3.85%	5.00%	0.13%	1.94%
Hadar	115	0.81%	5.50%	8.17%	1.41%	0.00%	0.00%	0.00%	1.70%	2.49%
Newport	105	0.00%	0.63%	1.63%	0.00%	0.00%	1.92%	0.00%	0.29%	2.77%
Virchow	100	1.46%	5.89%	0.25%	0.00%	0.00%	0.00%	2.50%	0.10%	0.14%
Napoli	90	0.16%	0.16%	3.77%	0.00%	0.00%	0.00%	0.00%	0.67%	0.55%
Agona	85	0.81%	2.59%	8.29%	2.35%	1.41%	0.00%	0.00%	0.26%	1.11%
Manhattan	81	0.00%	0.00%	0.00%	7.51%	0.00%	0.00%	0.00%	0.00%	0.00%
Indiana	75	2.27%	10.37%	9.55%	0.47%	0.00%	0.00%	2.50%	23.10%	31.02%
Brandenburg	71	0.16%	0.24%	0.88%	2.82%	4.23%	0.00%	0.00%	0.03%	0.00%
Dublin	67	0.16%	0.00%	0.00%	0.00%	2.82%	0.00%	0.00%	0.00%	0.00%
Kentucky	43	0.16%	0.55%	0.00%	0.00%	7.04%	0.00%	0.00%	0.00%	0.28%
Stourbridge	42	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.03%	0.14%
Bovismorbificans	40	0.00%	0.00%	0.50%	0.94%	0.00%	0.00%	0.00%	0.00%	0.00%
Kottbus	36	0.32%	2.75%	3.77%	0.47%	0.00%	0.00%	0.00%	23.19%	1.80%
Coeln	36	0.49%	0.16%	2.39%	0.00%	1.41%	0.00%	0.00%	0.03%	0.42%
Anatum	33	0.97%	1.34%	0.50%	0.00%	2.82%	0.00%	5.00%	0.80%	0.55%
Montevideo	32	3.73%	2.44%	1.01%	0.47%	14.08%	0.00%	7.50%	7.04%	2.91%
Oranienburg	31	0.16%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Bredeney	31	0.32%	1.02%	6.91%	0.94%	0.00%	0.00%	0.00%	0.06%	2.49%
Heidelberg	30	0.16%	2.99%	0.63%	0.00%	0.00%	0.00%	0.00%	0.00%	0.28%
Blockley	29	0.00%	0.08%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Corvallis	24	0.00%	0.00%	0.00%	0.94%	0.00%	0.00%	0.00%	0.00%	0.00%
Poona	21	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.50%	0.00%	0.00%
Senftenberg	17	19.77%	13.04%	14.70%	0.00%	0.00%	0.00%	0.00%	1.35%	2.77%
Havana	16	0.16%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Weltevreden	9	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	25.00%	0.00%	0.00%
Mbandaka	9	3.08%	0.86%	0.50%	0.94%	1.41%	0.00%	0.00%	0.13%	0.97%

Goldcoast	9		0.00%	0.00%	0.00%	0.94%	0.00%	0.00%	0.00%	0.00%	0.00%
Sandiego	8		0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.06%	0.00%
Muenster	7		0.16%	0.08%	0.75%	0.00%	0.00%	0.00%	2.50%	0.00%	0.00%
S 4,12:-:-	4		0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.14%
Essen	2		0.00%	0.00%	0.13%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Aberdeen	2		0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.14%
S 61:k:1,5,7	1		0.00%	0.00%	0.00%	0.00%	0.00%	57.69%	0.00%	0.00%	0.00%
Others	789		22.69%	15.79%	7.41%	33.33%	16.90%	25.96%	27.50%	11.02%	5.82%
Total number / prevalence	9076		617	1273	796	213	71	104	40	3113	722
Food amount			82 301	84 842	18 967	161 971	144 504	20 232	111 764	5 309	4 580

*the number of human cases corresponds to observed cases for all serotypes except Enteritidis and Typhimurium for which deterministically allocated numbers of cases are shown.

**the serotypes that are present in the sources for the passive dataset only are bolded (i.e. they were not observed with the active surveillance in the sources)

Table 4. Characteristics of the specific types in the active and passive food-animal datasets - *Salmonella* cases in France in 2005 (2004-2007 for the sources contamination in the active surveillance).

Type	Active dataset		Passive dataset	
	Prevalence (%)	Number of observed cases	Proportion (%)	Number of observed cases
Source contamination: Broilers				
Blockley	-	-	0.08	29
Heidelberg	0.02	30	-	-
Ohio	0.07	12	-	-
Source contamination: Layers				
Bareilly	2.95	8	-	-
Havana	0.04	16	0.16	16
Oranienburg	0.09	31	0.16	31
SE2	0.20	92*	4.05	92*
SE3	-	-	1.30	307*
Source contamination: Pigs				
Bovismorbificans	0.09	40	-	-
Brandenburg	0.34	71	-	-
Corvallis	-	-	0.94	24
Goldcoast	0.09	9	0.94	9
Manhattan	-	-	7.51	81
S 48:z4,z23:-	0.18	24	-	-
ST2	0.09	154*	-	-
Stourbridge	0.09	42	-	-
Source contamination: Turkeys				
Aijobo	0.04	2	-	-
Essen	0.13	2	-	-
Muenster	0.38	7	-	-
Source contamination: Other poultry				
Aberdeen	-	-	0.14	2
S 4,12:-:-	-	-	0.14	4
Source contamination: Sea products				
Poona	-	-	2.50	21
Weltevreden	-	-	25.0	9
Source contamination: Ducks				
Sandiego	-	-	0.06	8
Source contamination: Sheep				
S 61:k:1,5,7	-	-	57.7	1

* For Enteritidis and Typhimurium subtypes, the type distribution observed for subtypes cases were deterministically applied to allocate the cases with unknown subtype.

Table 5: Posterior means and 95% CI* of the source-dependant parameters (a_i) according to the dataset - *Salmonella* cases in France in 2005 (2004-2007 for the sources contamination in the active surveillance).

	Active dataset		Passive dataset	
Source	Posterior mean	95% CI*	Posterior mean	95% CI*
Layers	0.094	0.070-0.12	0.097	0.070-0.13
Broilers	0.092	0.066-0.12	0.090	0.059-0.13
Turkeys	0.32	0.14-0.59	0.11	0.02-0.31
Pigs	0.054	0.046-0.062	0.055	0.045-0.065
Cattle	13.5	0.025-83.64	0.0039	6.57E-05-0.018
Sheep	-	-	0.54	0.056-1.62
Sea products	-	-	0.076	0.050-0.11
Ducks	-	-	1.37	0.59-2.51
Other poultry	-	-	1.06	0.38-2.16

*95% credibility interval.

Table 6. Posterior means and 95% CI of the top 10 type-dependant parameters (q_i) according to the dataset - *Salmonella* cases in France in 2005 (2004-2007 for the sources contamination in the active surveillance).

	Active dataset			Passive dataset		
Type	Posterior mean	95% CI*	Ranking	Posterior mean	95% CI	Ranking
Heidelberg	19.24	-	1	-	-	(26)**
SE1	14.16	8.45-21.67	2	2.88	1.61-4.56	5
ST2	14.01	0.65-36.75	3	-	-	(33)
SE-multiS	8.18	6.17-10.56	4	0.73	0.55-0.94	9
Stourbridge	5.40	-	5	5.41	2.42-10.72	1
Bovismorbificans	5.14	-	6	-	-	(13)
Havana	4.41	-	7	1.09	-	8
SE2	4.26	0.17-12.82	8	-	-	(19)
Blockley	-	-	(NR)***	4.07	-	2
Orianenburg	3.89	-	9	2.11	-	6
Napoli	3.23	1.65-5.80	10	0.56	0.26-0.98	10
Dublin	-	-	(NR)	3.19	0.77-6.26	3
SE3	-	-	(NR)	3.03	1.45-5.41	4
Sandiego	-	-	(NR)	1.37	-	7

*95% credibility interval;
** numbers in brackets correspond to the ranking of types not included in the top 10;
*** not relevant, corresponds to types not observed in the dataset.

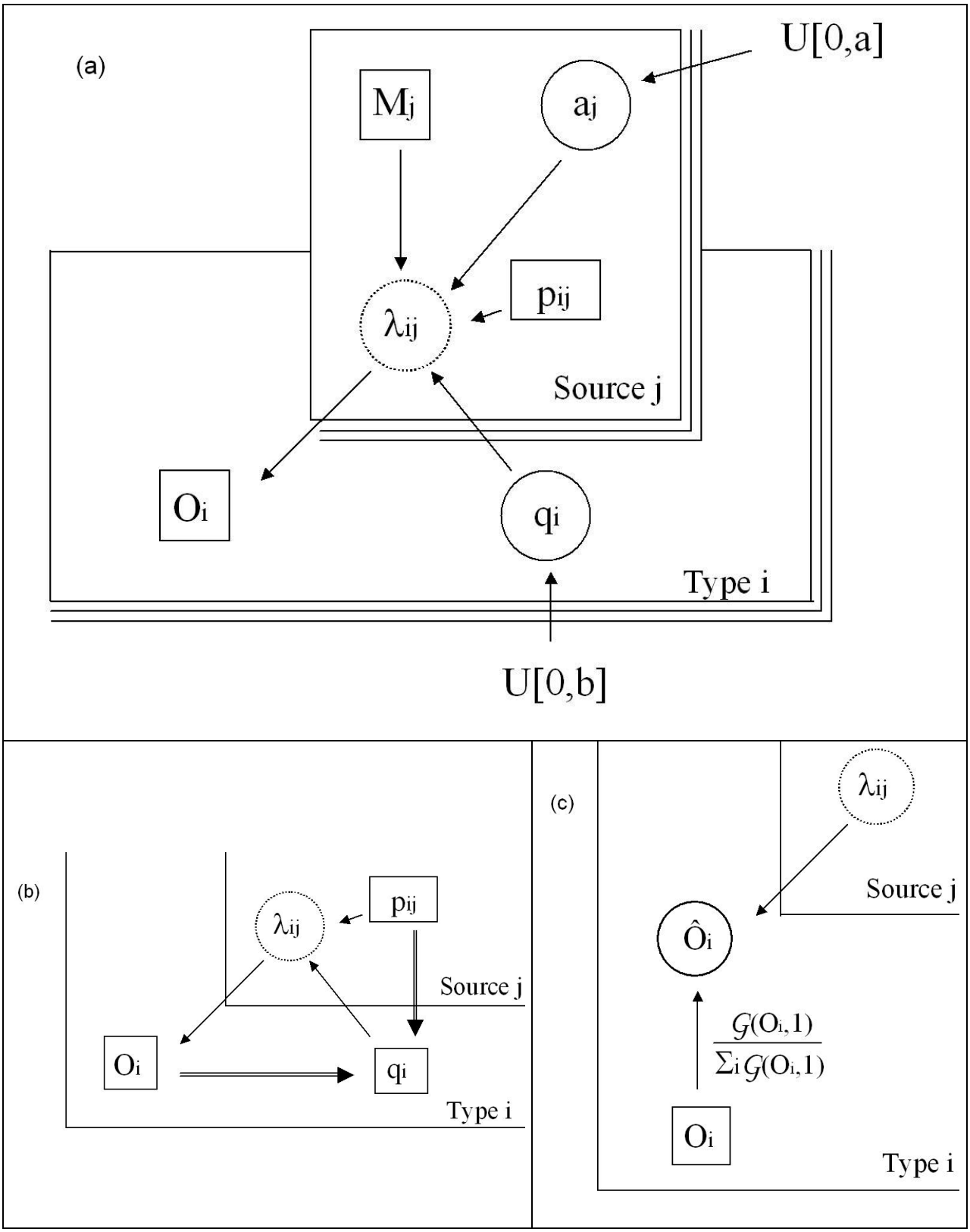

Figure 1. Directed Acyclic Graph of the bayesian model: (a) general DAG for non specific types and non SE and ST types, the double arrows indicate a deterministic link ; (b) setting specific types non SE and ST with a data-based value; (c) reallocation step for SE and ST types. Observations are represented in squares, parameters are represented in circles.
i = sero- or sub-type index;

j = source index;
o_i = observed number of human cases for sero- or sub-type i;
λ_{ij} = expected number of human cases due to type i in source j;
M_j = observed consumption in the general population of the source j in tons;
p_{ij} = observed prevalence of type i in source j;
q_i = unknown type-dependant parameter of type i;
a_j = unknown source-dependant parameter of source j.
U stands for uniform distribution and G for gamma distribution

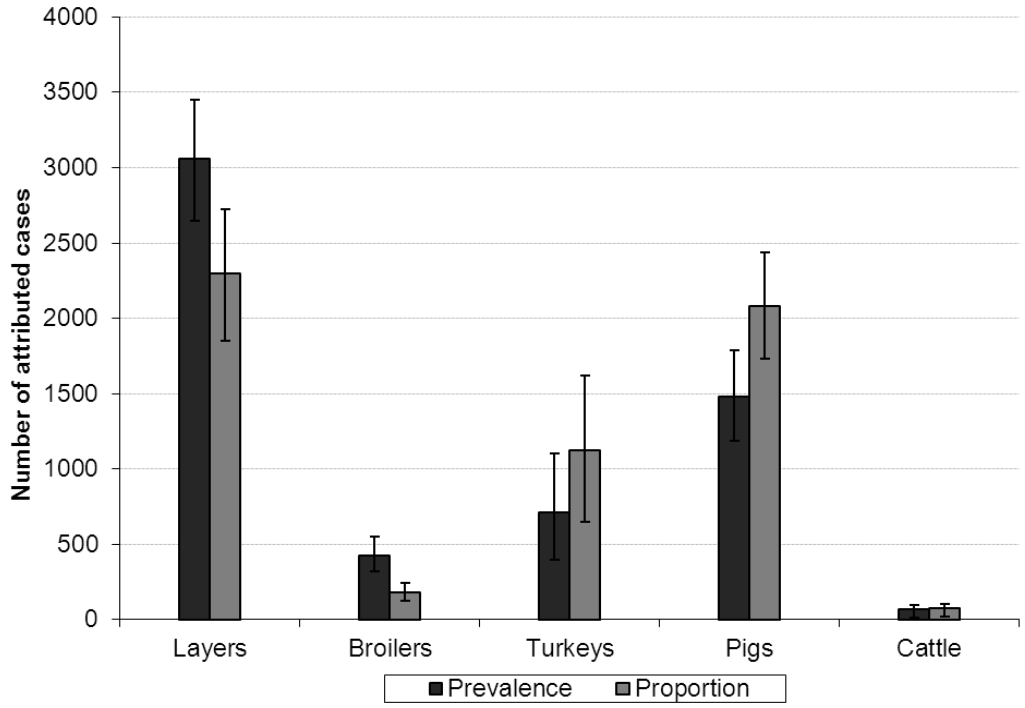

Figure 2. Attribution estimates (posterior means and 95% CI) for the active dataset according to the sources contamination indicator (proportion *vs* prevalence), *Salmonella* - France - 2005 (2004-2007 for the sources contamination)

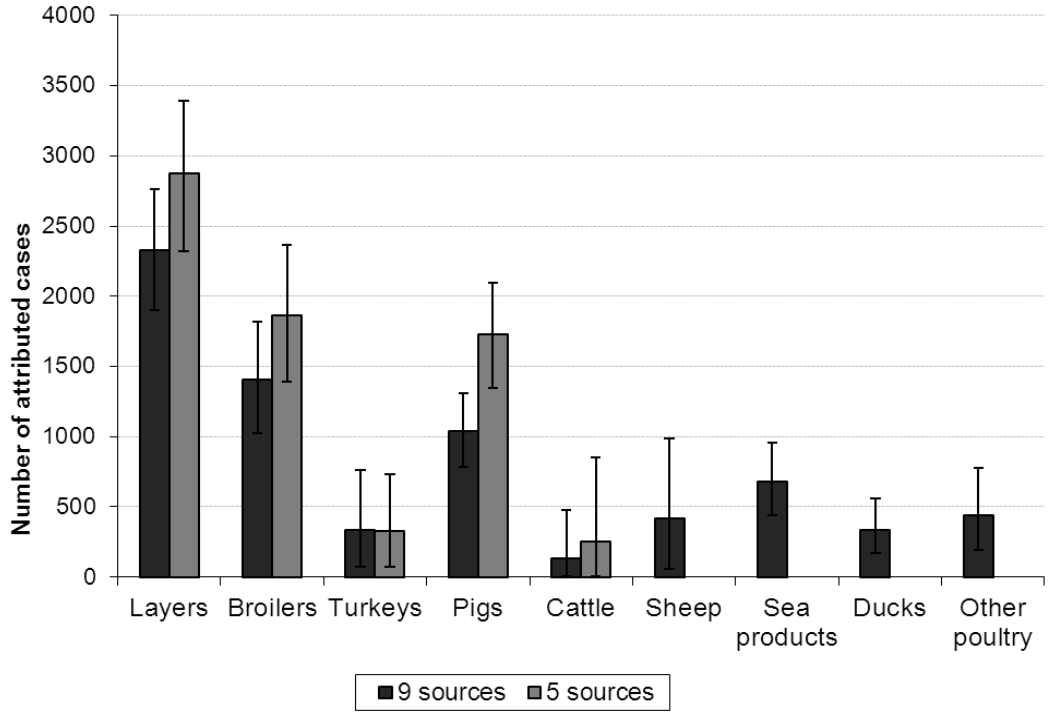

Figure 3. Attribution estimates (posterior means and 95% CI) for the passive dataset according to the number of sources included in the model, *Salmonella* - France - 2005

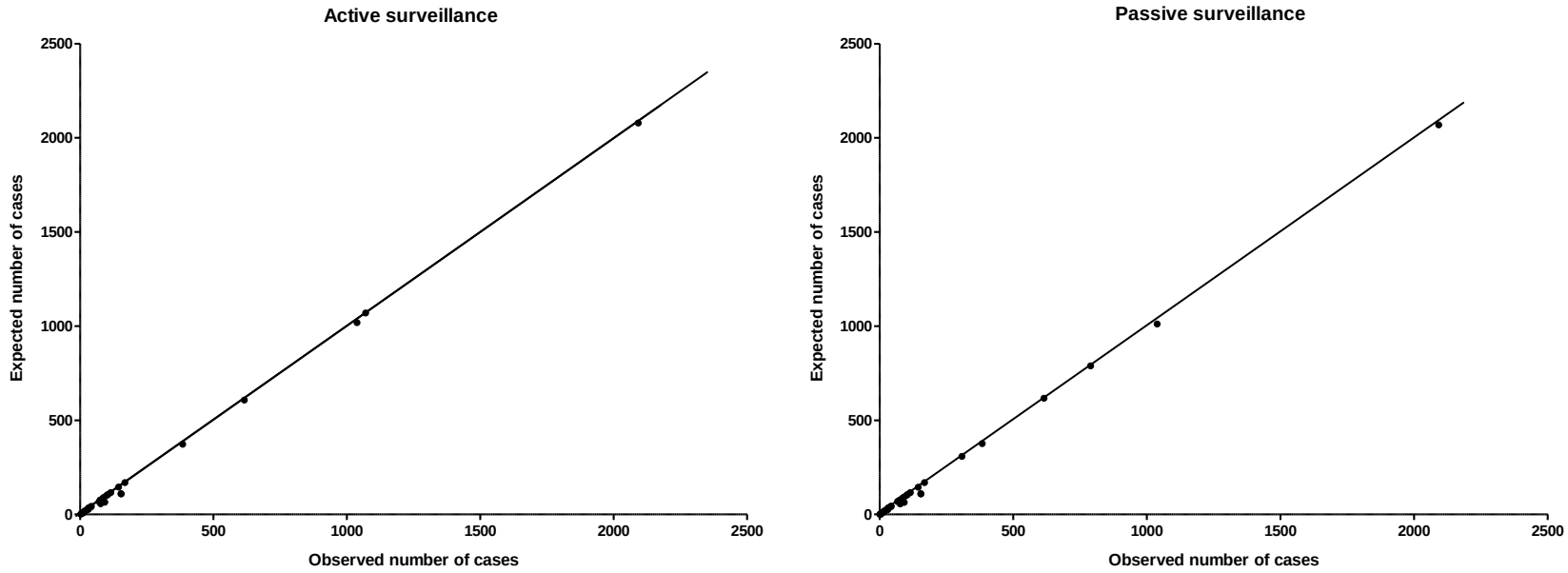

Figure 4. Adequacy of the attribution model (estimated over observed human cases per type) for the active and passive dataset - *Salmonella*, France 2005 (2004-2007 for the sources contamination in the active surveillance)

For Enteritidis and Typhimurium subtypes, deterministically reallocated numbers of cases are used in place of the observed numbers of cases

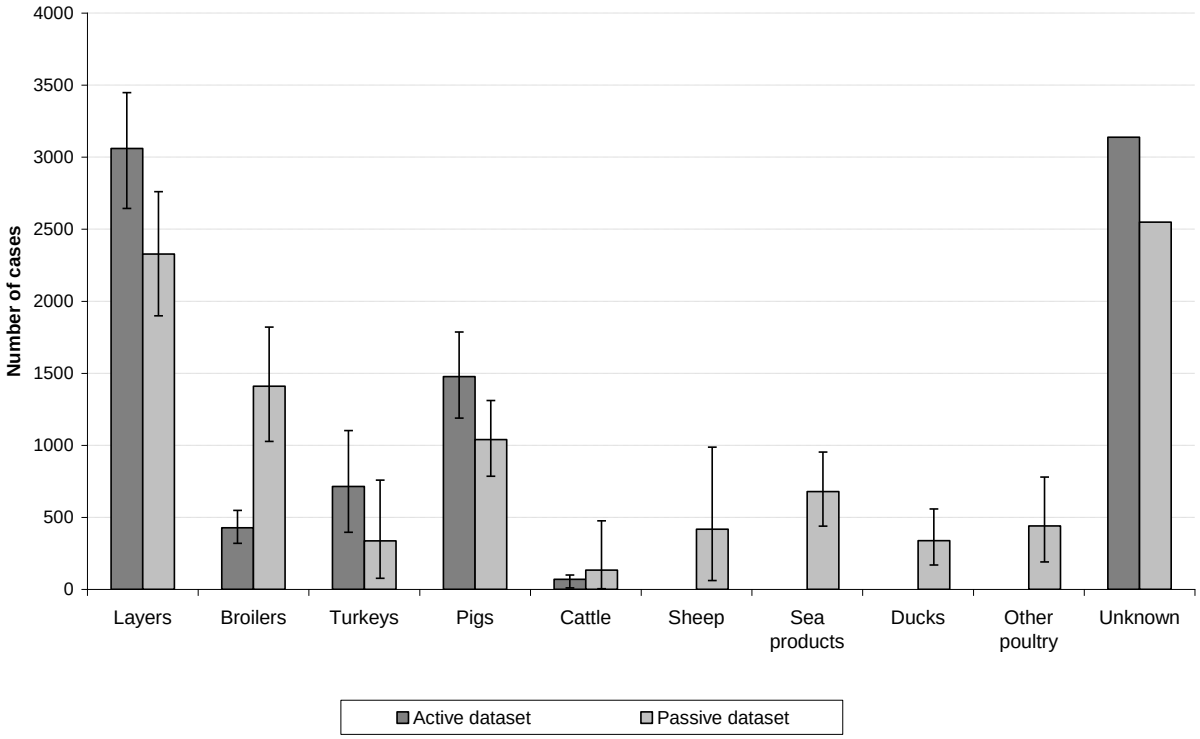

Figure 5. Attribution estimates for the active and passive datasets (posterior means and 95% CI) – *Salmonella*, France 2005 (2004-2007 for the sources contamination in the active surveillance)