Sensibilité du modèle aux priors informatifs

Quatrième article

Modèle bayésien d'attribution par typage microbiologique : sensibilité à l'information *a priori* informative et une proposition

En finalisation pour soumission à Foodborne Pathogens and Disease

Le modèle a été utilisé avec succès au Danemark, en Grande-Bretagne et aux Pays-Bas. Cependant, les résultats obtenus pour ces pays sont apparus comme divergents quant aux sources principales, hors cas liés à un voyage. De plus, pour la Pologne et l'Allemagne, la convergence du modèle n'a pu être atteinte. La raison invoquée repose sur le manque de qualité et de représentativité des données, ce qui semble peu réaliste pour les données allemandes.

Le modèle proposé par les Danois est un modèle surparamétré et nécessite donc l'introduction d'information *a priori* « informative » sur certains paramètres. Cette information informative peut porter sur le paramètre source-dépendant (qui mesure la différence entre les sources quant à leur capacité à véhiculer les salmonelles), et/ou sur le paramètre type-dépendant (qui mesure la différence entre les types quant à leur capacité à causer des infections). L'objectif des travaux présentés ci-après était d'évaluer l'impact de l'information informative introduite dans le modèle.

Pour ce faire, les analyses ont été conduites à partir du jeu de données français, pour lequel seules 4 sources potentielles ont été considérées (poules pondeuses, poulets de chair, dindes et porcs). Différentes variantes du modèle ont été considérées : un modèle simple déterministe, qui sert de base de référence, le modèle originel assorti de quatre types d'information différents, ainsi que l'adaptation du modèle proposée par Mullner *et al* (2009)[1]. Les quatre types d'information considérés correspondent à deux variantes de la paramétrisation utilisée dans la publication originelle (Hald, Vose et al. 2004) (le paramètre type-dépendant est fixé à une valeur arbitraire pour un type de référence), et deux variantes que nous proposons (les paramètres type-dépendants correspondants aux types spécifiques sont fixés soit à une valeur arbitraire soit à une valeur calculée à partir des données).

[1] Cf partie Objet et contexte de l'étude, § 5.2

Les résultats obtenus pour chacune de ces 6 variantes sont comparés en termes d'adéquation du modèle, de résultats d'attribution par source et de distribution *a posteriori* des paramètres source et type dépendants.

Le projet d'article suivant est prévu pour soumission à Foodborne Pathogens and Disease, après relecture par un traducteur.

Bayesian microbial subtyping attribution model: sensitivity to informative prior information and a proposition

J.M. David*[1,2,3,4], D. Guillemot[2,3,4], N. Bemrah[5], A. Thébault[5], C. Danan[6], M. Chemaly[7], FX. Weill[8], N. Jourdan[9], P. Sanders[1], L. Watier[2,3,4]

1 French Food Safety Agency, 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 Univ. Versailles Saint Quentin, Faculté de Médecine Paris Ile de France Ouest, F-78035, France

5 French Food Safety Agency, Maisons-Alfort, F-94701, France

6 French Food Safety Agency, Maisons-Alfort, F-94700, France

7 French Food Safety Agency, BP 53, Ploufragan, F-22440, France

8 Institut Pasteur, CNR des *Salmonella*, Laboratoire des Bactéries Pathogènes Entériques, Paris, F-75015, France

9 Institut de Veille Sanitaire, Saint-Maurice, F-94415, France

Corresponding author: Julie David, French Food Safety Agency, Institut Pasteur, BP 90203 35302 Fougères, France. Tel: +33 2 99 94 78 78; Fax: +33 2 99 94 78 99. Email: j.david@afssa.fr

Acknowledgements:

This study was supported by the Délégation Générale de l'Armement, Defence Ministry.

The authors wish to thank T. Hald and S.M. Pires for providing the code and for their support, all members of the MedVetNet WP28 Working Group for the useful discussions, all members of the CEB team at the French Food Safety Agency, Maisons Alfort, and of the *Salmonella* NRC at the Pasteur Institute for the subtyping analyses realized especially for this work and the members of the and of the Research Unit of Epidemiology and Welfare of Poultry and Rabbits at the French Food Safety Agency, Ploufragan. In addition, we wish to gratefully acknowledge Professor Lellouch for his constructive comments.

1 Introduction

Assessing the relative importance of the different reservoirs is a major issue regarding foodborne zoonoses such as salmonellosis or campylobacteriosis. As *Salmonella* is ubiquitous and present at a non negligible level in all the food-animal reservoirs and considering the fact that no fail-proof way to ensure the safety of food exists at any stage of the food chain (Allard 2002), controlling the presence of this pathogen from the farm level on seems critical (Sofos 2008). The attribution tool is thus a useful, let say indispensable tool to identify and prioritize interventions aiming at controlling it throughout the food chain (Batz, Doyle et al. 2005). Several methods are used worldwide to perform attribution: microbiological approaches (microbial subtyping approach and comparative exposure assessment), epidemiological approaches (case control studies), analyses of outbreak investigations as well as intervention studies and expert elicitation (Batz, Doyle et al. 2005; Pires, Evers et al. 2009). One of the most advanced models so far has been developed by Hald et al (Hald, Vose et al. 2004). This microbial subtyping approach tool allows attributing lab-confirmed human cases at the reservoir point to animal sources and is based on the serotyping and subtyping of *Salmonella* strains. Its principle is to compare the distribution of the types within the human cases and in the food sources, taking into account the consumption of each food source by the population. Its specificity is to take into account the differences between sources in their capacity to vehicle the pathogen and between the bacterial types in their capacity to induce infection, thanks to a bayesian framework with Markov Chain Monte Carlo simulations. However, as noticed by Mullner *et al*, the proposed model is not identifiable because of too many parameters to be estimated (Mullner, Jones et al. 2009). Thus, it is necessary to select *a priori* some parameters and fix them. In a Bayesian framework, including informative priors could have an important impact in the simulated joint posterior distribution of the parameters and thus on the marginal posterior distributions of interest (Gilks, Richardson et al. 1996). As a consequence, informative information included in a model has to be in agreement with the literature or validated by a sensitivity analysis (Binkowitz and Wartenberg 2001). With the informative priors (selected parameters and corresponding values) proposed by Hald *et al*, some countries met difficulties when applying the model to meet convergence, despite the availability of apparently reliable data on the main animal channels and on the human cases (Pires, Nichols et al. 2008). To overcome such problems, Mullner *et al* proposed a modification consisting in a hierarchical modelling, allowing reduction in the number of parameters to estimate. However, this way of doing entails a large increase in the posterior credibility intervals due to weak information on the introduced hierarchical level.

In this work, we propose to study the impact of the chosen informative priors on the marginal posterior distributions of interest, for Hald's proposal and a data-based alternative. Results are also compared with Mullner's approach and with a proposed simple model. This work was applied to the 2005 French dataset.

2 Material and methods

2.1 Data

The microbial subtyping approach requires spatially and temporally related data on the distribution of *Salmonella* types in the human cases and the various food sources at the reservoir point (i.e. farm level or abattoir). Data on the consumption of the considered sources is also necessary.

The data on the human cases come from the *Salmonella* National Reference Centre (NRC) and the National Public Health Institute (InVS). The human cases included in the model are domestic sporadic cases registered in 2005. Known travellers, cases from the outseas territories and departments, and outbreak related cases are excluded on the basis of the information given by NRC and InVS. The most recent year for which all the necessary information on the cases was available was 2005, it is thus the one considered as reference year in this work.

As to dispose of national representative prevalence data per *Salmonella* serotype in the animal sources, we used the data collected by the Food Directorate of the French Agriculture Ministry in the frame of the European baseline studies (layers, broilers, turkeys and pigs) and of the national surveillance plan of antimicrobial resistance in indicator and zoonotic bacteria in cattle (David, Danan et al. submitted). The data were collected in 2005 for layers and cattle, 2006 for broilers and 2007 for turkeys and pigs, based on a national representative sample of farms or carcasses. No equivalent national representative data were available in France in 2005 for broilers, turkeys and pigs. The strains were collected at the farm level for layers, broilers and turkeys and at the abattoir level for pigs and cattle. The prevalences were adjusted to flock size for layers, broilers and turkeys.

To optimize the attribution, Enteritidis and Typhimurium strains which represent more than 30% of the human cases each, have been further subtyped. The subtypes were defined through Multiple Correspondence Analysis (MCA) and mixed classification (Berge, Atwill et al. 2003) applied on the antimicrobial resistance profiles (disk diffusion method, amoxicillin, chloramphenicol, ceftazidim, gentamicin, kanamycin, nalidixic acid, streptomycin, sulfamides, sulfamides-trimethoprim, tetracyclin) of all Enteritidis and Typhimurium strains for the animal sources and a subset of strains for the human cases (92 strains out of 3536 (2.6%) for Typhimurium ; 102 out of 3138 (3.3%) for Enteritidis).

To be as close as possible to the effective consumption of the different sources by the French population, the consumption data used, come from the national individual survey on food consumption (INCA study) conducted by the French Food Safety Agency (AFSSA) in 1999 on 3 003 representative subjects above 3 years old (Volatier 2000). These results were actualized on the basis of the data concerning the available amount of each source on the market published annually by the French Livestock Institute. This allows taking the evolution of the consumption from 1999 to 2005 into account.

2.2 Description of the models

2.2.1 Simple approach

A simple approach was used to estimate in a simple manner the number of human cases contaminated by a source. As mentioned by Zwietering, a simple model is well adapted to have an insight on the functionality of the model (Zwietering 2008) and thus, the attribution estimations obtained with the following simple model were used as reference.

This approach assumes that all the sources are equivalent vehicles for the pathogen and all the serotypes have the same capacity to induce infection. The expected number of human cases of type i (i=1,...,I) linked to a source j (j=1,..., J) (λ_{ij}) is proportional to the prevalence of type i in source j (p_{ij}) and to the exposition of the human population to source j, measured by the amount of source j consumed by the general population (M_j). The total observed number cases of type i (o_i) is distributed among the sources in which type i is present according to the relative weight:

$$\lambda_{ij} = \frac{p_{ij} \times M_j}{\sum_j p_{ij} \times M_j} \times o_i \qquad (1)$$

The expected number of cases due to source j is obtained by summing the λ_{ij} on i.

As mentioned earlier, only a percentage of Enteritidis and Typhimurium cases are subtyped It is thus necessary to introduce a reallocation step in the process. The types distribution observed within the cases that have been subtyped is used to allocate the cases with unknown subtype. This assumes similar subtype distribution for the cases subtyped and for the cases not subtyped. The obtained reallocated numbers of cases per type are used in place of o_i. This way of doing is referred to as "deterministic reallocation".

2.2.2 Bayesian approach

Hald model: With previous notations, the number of observed human cases (o_i) is assumed to be Poisson distributed:

$$o_i \sim \text{Poisson}\ (\Sigma_j\ \lambda_{ij})\ \text{with}\ \ \lambda_{ij}= M_j\ p_{ij}\ q_i\ a_j \qquad (2)$$

Thus, 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}), as well as the consumption of the source j in the general population (M_j) as for the simple model, but two new parameters are introduced: a source dependant parameter (a_j) and a type dependant parameter (q_i). As defined by Hald *et al*, the type dependant factor (q_i) summarizes the characteristics of the serotype (survivability, virulence, pathogenicity, …) which determine its capacity to cause an infection and the source dependant factor (a_j) summarizes the characteristics of the source (physical properties, preparation methods, processing procedures, …) which determine its capacity to act as a vehicle for *Salmonella* (Hald, Vose et al. 2004).

The expected number of cases due to source j is obtained by summing the λ_{ij} on i.

From the writing of the model, an overparameterization appears, which concerns J parameters. Thus, informative priors have to be introduced for at least J parameters, which consists in first selecting the parameters to be fixed and then determining the constant values to which they are fixed. This step will concern the type-dependant parameters, which can be justified by the fact that there are only J source-parameters and moreover that the type-dependant parameters, have a stronger impact on the results than the source dependent parameters (Sarwari, Magder et al. 2001).

The untyped Typhimurium and Enteritidis cases were here reallocated in a bayesian way, assuming, as for the simple model, that the subtypes distributions are similar for subtyped and not subtyped cases, and using Gamma distributions to reflect the uncertainty on the observed numbers of cases per type. For more details see Hald et al (2004).

Mullner model: To avoid including informative priors, the q_is are modelled as random variables which follow a log normal distribution with a fixed constant mean and a precision parameter to be estimated (in the place of the q_is). The gamma distribution was used for the precision parameter. From the simulated joint posterior distribution, marginal posterior distributions of the random variables q_is can be recovered and were used to draw comparisons with the other models, even if their trace plots didn't appear to stabilize. Prior information used was the same as authors' proposal (Mullner, Jones et al. 2009).

Common features

Unlike Hald's work, where three sources of uncertainty concerned the human data (travel's proportion, outbreak related cases, Enteritidis and Typhimurium subtyping) which led to a complex model, our application exclusively contains one source of uncertainty (Enteritidis and Typhimurium subtyping). Accordingly, with the French dataset, the model results are easiest to interpret for what concerns the impact of the informative priors on the attribution results and on the parameters estimations (Zwietering 2008).

For the three considered approaches, it is assumed that cases with a history of travel or living in the overseas departments and territories have not acquired their Salmonellosis on the national metropolitan territory and, the way around, that domestic cases have been infected on the metropolitan territory. Moreover, we consider that the human cases infected by a type included in the model shall have been contaminated by one of the sources considered. The exposition to the sources is assumed to be similar in the general population and in the cases. And finally, these approaches require the microbiological types to be heterogeneously distributed among the sources considered.

Prior information

Informative priors

Two different configurations of the Hald model, each declined in two versions were studied. The first one corresponds to the original proposal fixing a reference type, the second focuses on specific types (i.e. types present in only one food-animal source).

In their work, Hald *et al* assumes that the q_is are of equal value for the subtypes within the serotype Enteritidis (reference type) and within the serotype Typhimurium. For Enteritidis, the q_is are moreover fixed to an arbitrary constant value while for Typhimurium a uniform prior distribution is chosen (Table 1). This proposal is referred to as "Reference-Type 1". Enteritidis has been chosen by Hald *et al*, because it is the most frequent serotype within the human cases. Though, in the French dataset, the most frequent serotype is Typhimurium. Thus, another configuration has been studied, where Typhimurium and Enteritidis roles are inverted, Typhimurium becoming the reference type. This one is referred to as "Reference-Type 2". Both proposals assume that type dependant parameters for Enteritidis and for Typhimurium are independent of subtypes. Since Enteritidis and Typhimurium serotypes are both concerned by the reallocation process, an influence of the informative prior on this stage is to be expected.

We propose to fix specific types, not involved in the reallocation process as to prevent any possible interaction with this last one. For these serotypes, the link between the number of cases due to type i and its prevalence in the source is direct, which allows to determine a data-based value for q_i. However, as to study the impact of the way the constant values are defined, the type dependant parameters associated to specific types are either fixed to an arbitrary constant value (Specific-Types 1) or to data-based values (Specific-Types 2). These values were calculated as follow:

$$q_i = \frac{o_i}{\sum_i o_i} \times \frac{1}{p_{ij}} \qquad (3)$$

where o_i is the observed number of cases due to type i, p_{ij} is the prevalence of type i in the unique source j. We thus used the percentage of human cases divided by the prevalence in the source as an indicator of the capacity of type i to cause an infection.

Non informative prior distributions for other parameters

Prior distributions for the a_js and not fixed q_is are uniform distributions with 0 as lower value. To assess that the upper value set allows encompassing all the possible values for the parameter, the posterior distributions are visually checked. If these distributions seem to be arbitrarily cut off, the uniform distributions are widened as indicated by Hald *et al*.

Constant values and non informative prior distribution's parameters

Arbitrary constant and prior distribution's parameters were chosen to optimize the reallocation process and the convergence of marginal posterior distributions of the a_js et q_is unknown parameters (table 1). They were determined as to assure first the compatibility of the reallocation with the observed proportions of subtypes and, second, to ensure that a minimum of the a_js and q_is have marginal posterior distribution arbitrarily cut off.

Adequate posterior results were easily obtained for both Specific-Types based configurations but they were partially and particularly hard to achieve for both Reference-Type based configurations, and higher values than those proposed by Hald were necessary. For Reference-Type 1, 5 unknown q parameters (14% of the q's) and for reference-type 2, 7 unknown q parameters (19% of the q's) have a marginal posterior distribution cut off. However, using widely dispersed starting values didn't modify the posterior marginal distributions of interest.

2.3 Software used

Analyses were performed using Excel® for the simple approach and Winbugs 1.4® for the bayesian models (Lunn, Thomas et al. 2000).

3 Results

3.1 Data

A total of 9 076 human cases were included in the dataset, as well as the prevalence results for 519 layer farms, 371 broiler farms, 331 turkey farms, 1 166 pig carcasses and 334 cattle carcasses (Figure 1). However, because of the very low prevalence of Salmonella in cattle (only 2.4%), inducing an important asymmetric density of the marginal posterior distribution with a small number of attributed cases (around 50), cattle was excluded as a source for this comparative study. The types are heterogeneously distributed among the sources according to Fisher exact tests.

Serotypes considered were those who were the most frequent among human cases (> 30 cases) and in sources (> 15% of the strains within a source) and those corresponding to specific types. The serotypes which didn't fulfil those criteria were included in "others" category. For Enteritidis and Typhimurium, 9 subtypes were defined each. But only 5 subtypes for Typhimurium and 3 for Enteritidis were included in the model. Those subtypes are the ones observed simultaneously in the human cases and at least in one animal source. The other subtypes were grouped in categories "other Typhimurium" and "other Enteritidis".

Human cases belonging to the three "other" categories could not be considered for the attribution because, depending on the source, those categories have different types compositions, which makes it impossible to distribute the relative cases according to prevalences that are not comparable. Thus, as these cases cannot be attributed to a source, 5 938 cases spread between 28 serotypes, 5 Typhimurium subtypes and 3 Enteritidis subtypes were considered for attribution and referred to as attributable cases.

Our studied sample contains two reference types including 8 subtypes (2 of which are specific), 12 specific types not included in the reference types and 14 other serotypes. Excluding Enteritidis and Typhimurium subtypes, a total of 5 types are specific to pigs, 2 to layers and 2 to broilers and 2 to turkeys (Table 2).

Finally, on the basis of the updated results of the INCA study, national consumption for 2005 was 82 301 tons for layers (eggs), 84 842 tons for broilers, 18 967 tons for turkeys and 161 971 tons for pigs.

3.2 Attribution

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

We first performed comparisons corresponding to adequacy diagnosis. It concerned the number of cases per type and the total number of attributed cases, which were compared to the observed values. We then analyzed the expected numbers of cases per source with the simple model's results as reference. Finally, for models giving adequate results, the type and source dependant parameters were considered.

3.2.1 Adequacy of the model

Predicted and observed numbers of cases per type are in good agreement for Specific-Types 2 and for Mullner models, unlike both Reference-Type configurations and Specific-Types 1 (figure 2). Indeed, for both Reference-Type configurations, 5 subtypes are not well predicted. Regarding Typhimurium, subtypes 1, 2, 3 and 4 are concerned. The expected number of cases for Typhimurium globally is 113 for Reference-Type and 1 for Reference-Type 2, where 1 807 Typhimurium cases are observed. The discrepancy is less spectacular but still significant for Enteritidis, with 2 concerned subtypes, SE multiS and SE1. For Reference-Type 1, 1 796 and 869 cases are expected respectively for SE-multiS and SE1 vs 2 092 and 615 observed cases. For Reference-Type 2, 315 cases are expected for SE1 vs 615 observed. These weak fits are not related to the cut off of the marginal posteriors distributions of the q's, except for SE1 in the Reference-Type 2 configuration.

For Specific-Types 1 configuration, almost all specific types are not well predicted (Stourbridge, Bovismorbificans, Oranienburg, Heidelberg, S 48:z4,z23:-, Havana, Ohio, Goldcoast, Bareilly, Aijobo). Though, in this last case, the convergence is good and the mismatch concerns only serotypes associated to a small number of cases (5% of attributable cases overall).

As a consequence, the total expected number of cases with Specific-Types 2 (n= 5 745) and Mullner (n = 6 000) models are in accordance with the 5 938 observed attributable domestic sporadic cases. This is also true for Specific-Types 1 (n = 5 745) despite the discrepancy for some of the specific types. Finally, for the Reference-Type configurations, a decrease of around 30% in attributed cases is observed (4 122 expected cases for Reference-Type 1 and 4 079, for Reference-Type 2).

3.2.2 Number of cases per source

According to the model, posterior attribution means per food source can be very different (figure 3). Only two proposals are in accordance with the proposed simple model, that is Specific-Type 2 and Mullner's models. In those cases, considering marginal posterior means or percentages, ranking are similar for the most important sources: layers as the main source for human cases and pigs at the second place (Tableau 3). The ranking for turkeys and broilers is not so clear. However, in both models these sources are not significantly different. Though, with Mullner's model, no firm conclusion on the relative importance of the sources can be made, because of very wide 95% CI (95% Credibility Intervals).

For the other models, the results are quite heterogeneous since presenting contrasting marginal posterior means, some of which seem unlikely. For example, when applying Reference-Type 1, less than 10% of the cases should be attributed to layers which is not compatible with literature (Mølbak and Neimann 2002; Hald, Vose et al. 2004; Hennessy, Cheng et al. 2004; Greig and Ravel 2009), and with the high proportion of Enteritidis strains observed in this source (Figure 1).

3.2.3 Type and source dependant parameters

The results for those parameters are presented only for Specific-Types 2 and Mullner models which have a good adequacy and attribution results in accordance with the simple model.

The ranking and even the posterior means of all type-dependant parameters are quite similar for both models, except for subtype ST2 (figure 4). Though, the 95% CI are so wide for this subtype that the difference of the posterior means for the two models isn't significant. In fact, wide credibility intervals related to the reallocation process are observed for all the subtypes within Enteritidis and to a less extent within Typhimurium.

The global agreement of q values between both models and especially for the specific types, strengthens the appropriateness of the proposed data-based informative values of the type dependant parameters for the specific serotypes. What's more, we compared the posterior means obtained with the Specific-Types 2 model for the 2 specific types not fixed because involved in the reallocation (ST2 and SE2), with the constant value that would have been calculated as defined in equation (3) (for the subtypes, the proportion of observed cases is obtained using the deterministic reallocation). The calculated value and posterior means were respectively 19.78 versus 14.01 (95%CI 0.68-37.17) for ST2 and 4.96 versus 4.30 (95%CI 0.17-12.76) for SE2, which shows a good agreement between the calculated and estimated values.

Several serotypes appear to have high posterior means, or calculated values. The Enteritidis subtypes, especially SE-multiS and SE1, ST2 and some of the specific types, i.e. Stourbridge (Pigs-related), Bovismorbificans (Pigs-related), Oranienburg (Layers-related), Heidelberg (Broilers-related) and Havana (Layers-related).

The source dependant parameters present, for both models, highest posterior means for broilers (Mullner: 1.11 [0.21 – 3.49], Reference-Types 2: 0.41 [0.30 – 0.55]). As the 95% CI for Mullner is very wide, the source-dependant parameter for broilers is not significantly different between both models. For what concerns the other sources. They are smallest than the broilers-dependant factors and near from each other (Mullner: 0.08 to 0.11, Specific-Types 2: 0.05 to 0.09). Within Specific-Types 2 model, the relative a factors are significantly smallest than the broilers-dependant one. Thus no firm conclusion can be drawn from the Mullner estimations, but with the Specific-Types 2 configuration, the source-dependant factor relative to broilers appears significantly higher than the three others.

4 Discussion

The dataset gathered for this work suffers from some insufficiencies. The human cases are collected on a passive way and on a voluntary basis, as sporadic Salmonellosis isn’t submitted to mandatory declaration. Though, the coverage of the NRC database is estimated to 40% of the confirmed cases and its representativeness has been estimated as good (David, Danan et al. submitted). It thus constitutes a reasonable basis for such a study.

The data gathered for the sources are not timely consistent with the human ones, but they are representative and give access to prevalences per serotype and subtype, which wasn’t possible with other sources of data on food-animals. As the aim of this work was to understand the functionality of the model and not to produce attribution estimates for France, we still used them. Though, the estimations provided by the models must be considered with caution in view of the quality of the dataset, the exclusion of cattle as a source and the underlying assumptions.

The overparameterization of Hald model has been originally treated by assuming equality in some type dependant parameters, while fixing a reference serotype to an arbitrary constant value. Recently, Mullner proposed to reduce the number of parameters to be estimated by introducing a hierarchical structure for the type dependant parameters despite of their low number. Our proposal targets the informative prior as did Hald *et al*, but is based on specific types rather than on a chosen reference type. We use the fact that, as specific types are present in a unique source, the link between the prevalence and the proportion of the human cases induced is direct. The values, to which they are fixed, are data-based and shall reflect their capacity to induce infection measured in percentage of cases per prevalence rate. These values are in agreement with marginal posterior means of Mullner’s model which comforts our proposal. For simplicity, all specific types were fixed, but, fixing only J specific types among which at least one in each source, led to similar results.

As expected, the bayesian attribution model appears very sensitive to informative priors and the way the priors are defined. Even when complying with Hald's proposal, that is fixing Enteritidis (Reference Type 1) or Typhimurium (Reference Type 2) which are almost equally frequent within the human cases, the attribution results are spectacularly different.

The Specific-Types based information has thus several advantages. Namely, in regard to the Reference-Type solution which fixes the most frequent serotypes (over 30% of the cases in our dataset), when fixing specific serotypes, only a small proportion of the human cases (5% in our dataset) is concerned. With no certainty on the accuracy of the constant values chosen to fix the relative q parameters, the potential direct influence is thus minimized. This also allows avoiding any interaction with the reallocation process in which the reference types are implied. Namely, when fixing a reference type dependant parameter, the reallocation process was shown to be dependant of the combination: constant value / upper value of the uniform distributions of (q_i). Moreover, when fixing parameters to arbitrary values, the repartition of the cases among the sources isn't realistic, the total expected number of cases isn't in accordance with the number of attributable cases and the expected numbers of cases per type aren't in agreement with the observations for the fixed types, whichever those are (reference types or specific types). On the contrary, introducing data-based informative priors on specific types not involved in the reallocation seems a good alternative. Another point is that the Specific-Types based solution doesn't require to make any assumption on the equality of the type-dependant parameters for the subtypes within Enteritidis and Typhimurium. As we used the antimicrobial profiles as subtyping tool and as some resistance traits have been shown to be linked to virulence factors especially in Typhimurium (Martinez and Baquero 2002; Mølbak 2004; Foley and Lynne 2008), this is of primarily importance. Regarding Mullner approach, the obtained results are comparable, but the convergence is easier to meet with the Specific-Types based solution and the credibility intervals are much nearer?, allowing concluding on the relative importance of the sources, which wasn't possible with Mullner model for our dataset.

Though, this approach requires disposing of as many specific types other than Enteritidis and Typhimurium as food-sources included, which for example was not the case in the Danish dataset (Hald, Vose et al. 2004), but maybe because specific serotypes were among the less frequent types categorized as "others". If no such specific serotypes can be found even in the less frequent ones, then a solution would be to apply the simple model, except that including the source-dependant and the type-dependant parameters is a key point to enhance the source attribution estimates, as differences between the sources in their capacity to vehicle the pathogen and the differences between types to induce infection are described in the literature (Blaser and Newman 1982; D'Aoust 1989; Sarwari, Magder et al. 2001; Coleman, Marks et al. 2004; Bollaerts, Aerts et al. 2008; Jones, Ingram et al. 2008). Another limit of this approach was common with Hald's and Mullner's proposals and linked to the dataset. The subset of Enteritidis and Typhimurium strains tested for antimicrobial resistance in the human dataset represents less than 4% of the total number of Enteritidis and Typhimurium cases. This leads to wide credibility intervals for the posterior distribution of the relative type dependant parameters and for the predicted numbers of cases per type and source, whichever the model was. Testing more strains for antimicrobial resistance could greatly enhance the predictions of all models.

As a conclusion, we proposed alternative informative priors, based on specific types, to be used in the Hald model. For datasets comprising specific types not involved in a reallocation process, our proposal allows enhancing the convergence and avoiding biases in the results of the model. The obtained results are consistent with the simple model and allow drawing significant differences between the sources. We thus would recommend using the Specific-Types based informative prior when possible.

However, this approach could be enhanced in several ways. First of all, the bayesian framework allows to introduce uncertainty on the parameters. Thus, low flat distributions centred on the proposed data-based values could be introduced to replace the constant values to which the parameters are fixed. A further step in this approach would be to find exogenous information reflecting the infectious capacity of the concerned types (dose-response relationship, virulence or pathogenicity) to define those values.

Another path to enhance the quality of the model's predictions would be to use proportionality parameters in the informative priors set on the specific-types dependant parameters associated to the same source. At last, when running the model, the source and the type dependant parameters appeared inter-dependant, which is consistent with the known specificity of the dose-illness relationship for a serotype-food matrix combination (Bollaerts, Aerts et al. 2008). It thus would be interesting to initiate a reflection on how to introduce interactions between both parameters in the model.

Finally, a recent paper emphasized the possibility to obtain convergent results with inappropriate priors without warning from the BUGS software (Lunn, Spiegelhalter et al. 2009). Thus, it appeared in this work that the sensitivity analysis was essential to judge of the appropriateness of the approaches for a given dataset and that some apparently logical assumptions led to inappropriate results.

References

Allard, D. G. (2002). "The 'farm to plate' approach to food safety - Everyone's business." Canadian Journal of Infectious Diseases 13(3): 185-190.

Batz, M. B., M. P. Doyle, et al. (2005). "Attributing illness to food." Emerging Infectious Diseases 11(7): 993-999.

Berge, A. C. B., E. R. Atwill, et al. (2003). "Assessing antibiotic resistance in fecal Escherichia coli in young calves using cluster analysis techniques." Preventive Veterinary Medicine 61(2): 91-102.

Binkowitz, B. S. and D. Wartenberg (2001). "Disparity in quantitative risk assessment: a review of input distributions." Risk Anal 21(1): 75-90.

Blaser, M. J. and L. S. Newman (1982). "A review of human salmonellosis: I. Infective dose." Reviews of Infectious Diseases 4(6): 1096-1106.

Bollaerts, K., M. Aerts, et al. (2008). "Human salmonellosis: estimation of dose-illness from outbreak data." Risk Anal 28(2): 427-40.

Brooks, S. P. and A. Gelman (1998). "General methods for monitoring convergence of iterative simulations." Journal of Computational and Graphical Statistics 7(4): 434-455.

Brooks, S. P. and G. O. Roberts (1998b). "Convergence assessment techniques for Markov chain Monte Carlo." Statistics and Computing 8(4): 319-335.

Coleman, M. E., H. M. Marks, et al. (2004). "Discerning strain effects in microbial dose-response data." J Toxicol Environ Health A 67(8-10): 667-85.

Cowles, M. K. and B. P. Carlin (1996). "Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review." Journal of the American Statistical Association 91(434): 883-904.

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

David, J. M., C. Danan, et al. (submitted). "Structure of the French farm to table surveillance system for *Salmonella*."

Foley, S. L. and A. M. Lynne (2008). "Food animal-associated Salmonella challenges: pathogenicity and antimicrobial resistance." Journal of animal science 86(14 Suppl).

Gilks, W. R., S. Richardson, et al. (1996). Markov Chain Monte Carlo in practice: Interdisciplinary Statistics. London, Chapman & Hall/CRC.

Greig, J. D. and A. Ravel (2009). "Analysis of foodborne outbreak data reported internationally for source attribution." International Journal of Food Microbiology 130(2): 77-87.

Hald, T., D. Vose, et al. (2004). "A Bayesian approach to quantify the contribution of animal-food sources to human salmonellosis." Risk Anal 24(1): 255-69.

Hennessy, T. W., L. H. Cheng, et al. (2004). "Egg consumption is the principal risk factor for sporadic Salmonella serotype heidelberg infections: A case-control study in foodnet sites." Clinical Infectious Diseases 38(SUPPL. 3).

Jones, T. F., L. A. Ingram, et al. (2008). "Salmonellosis outcomes differ substantially by serotype." J Infect Dis 198(1): 109-14.

Lunn, D., D. Spiegelhalter, et al. (2009). "The BUGS project: Evolution, critique and future directions." Stat Med.(28): 3049-3067.

Lunn, D. J., A. Thomas, et al. (2000). "WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility." Statistics and Computing 10(4): 325-337.

Martinez, J. L. and F. Baquero (2002). "Interactions among strategies associated with bacterial infection: pathogenicity, epidemicity, and antibiotic resistance." Clin Microbiol Rev 15(4): 647-79.

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(8-9): 364-369.

Mølbak, K. and J. Neimann (2002). "Risk factors for sporadic infection with salmonella enteritidis, Denmark, 1997-1999." American Journal of Epidemiology 156(7): 654-661.

Mullner, P., G. Jones, et al. (2009). "Source attribution of food-borne zoonoses in New Zealand: A modified hald model." Risk Analysis 29(7): 970-984.

Pires, S. M., E. G. Evers, et al. (2009). "Attributing the human disease burden of foodborne infections to specific sources." Foodborne Pathog Dis 6(4): 417-24.

Pires, S. M., G. Nichols, et al. (2008). *Salmonella* source attribution in different European countries. 21st International ICFMH Symposium: "Evolving Microbial Food Quality and Safety". Aberdeen.

Sarwari, A. R., L. S. Magder, et al. (2001). "Serotype distribution of Salmonella isolates from food animals after slaughter differs from that of isolates found in humans." Journal of Infectious Diseases 183(8): 1295-1299.

Sofos, J. N. (2008). "Challenges to meat safety in the 21st century." Meat Science 78(1-2): 3-13.

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

Zwietering, M. H. (2008). "Quantitative risk assessment: Is more complex always better? Simple is not stupid and complex is not always more correct." Int J Food Microbiol.

Model name	Prior distribution for a	Informative and non informative distribution for q*	
		Types	Prior
Reference Type 1	Uniform distribution: U(0,10)	**Enteritidis**	**Constant value : 10**
		Typhimurium	**Subtypes equals** Uniform distribution: U(0,1000)
		Others	Uniform distribution: U(0,1000)
Reference Type 2	Uniform distribution: U(0,10)	**Typhimurium**	**Constant value: 10**
		Enteritidis	**Subtypes equals** Uniform distribution: U(0,10 000)
		Others	Uniform distribution: U(0,10 000)
Specific Types 1	Uniform distribution: U(0,100)	**Specific types other than Enteritidis and Typhimurium**	**Constant value: 1**
		Others	Uniform distribution: U(0,100)
Specific Types 2	Uniform distribution: U(0,100)	**Specific types other than Enteritidis and Typhimurium**	**Specific value :** $q_i = \frac{o_i}{\sum_i o_i} \times \frac{1}{p_{ij}}$
		Others	Uniform distribution: U(0,100)

Table 1: Informative and non informative prior information for Hald's model
bold indicates informative information.

Type	Source	p	Number of cases	Specific value**
Heidelberg	Broilers	0.02%	30	19.240
Ohio	Broilers	0.07%	12	1.889
SE2	Layers	0.20%	92*	-
Oranienburg	Layers	0.09%	31	3.892
Havana	Layers	0.04%	16	4.407
Bareilly	Layers	2.95%	8	0.030
ST2	Pigs	0.09%	154*	-
Brandenburg	Pigs	0.34%	71	2.280
Stourbridge	Pigs	0.09%	42	5.396
Bovismorbificans	Pigs	0.09%	40	5.139
S 48:z4,z23:-	Pigs	0.18%	24	1.469
Goldcoast	Pigs	0.09%	9	1.102
Muenster	Turkeys	0.38%	7	0.201
Aijobo	Turkeys	0.04%	2	0.551

Table 2: Specific types characteristics
*reallocated number of cases for the subtypes, using the deterministic reallocation
** definition in table 1
p : prevalence

Source	Simple model	Specific-Types 2		Mullner	
	Percentage	Posterior percentage	95% CI	Posterior percentage	95% CI
Layers	53.3	53.5	46.0 - 60.0	40.0	17.6 - 60.0
Pigs	32.2	25.8	20.7 - 31.1	34.3	21.5 - 45.9
Broilers	8.7	8.0	5.6 - 9.5	13.6	4.7 - 27.5
Turkeys	5.8	12.7	6.9 - 19.2	12.1	2.0 - 45.9

Table 3: Posterior percentages of expected cases per source for the Specific-Types 2 and Mullner models in comparison to the simple approach

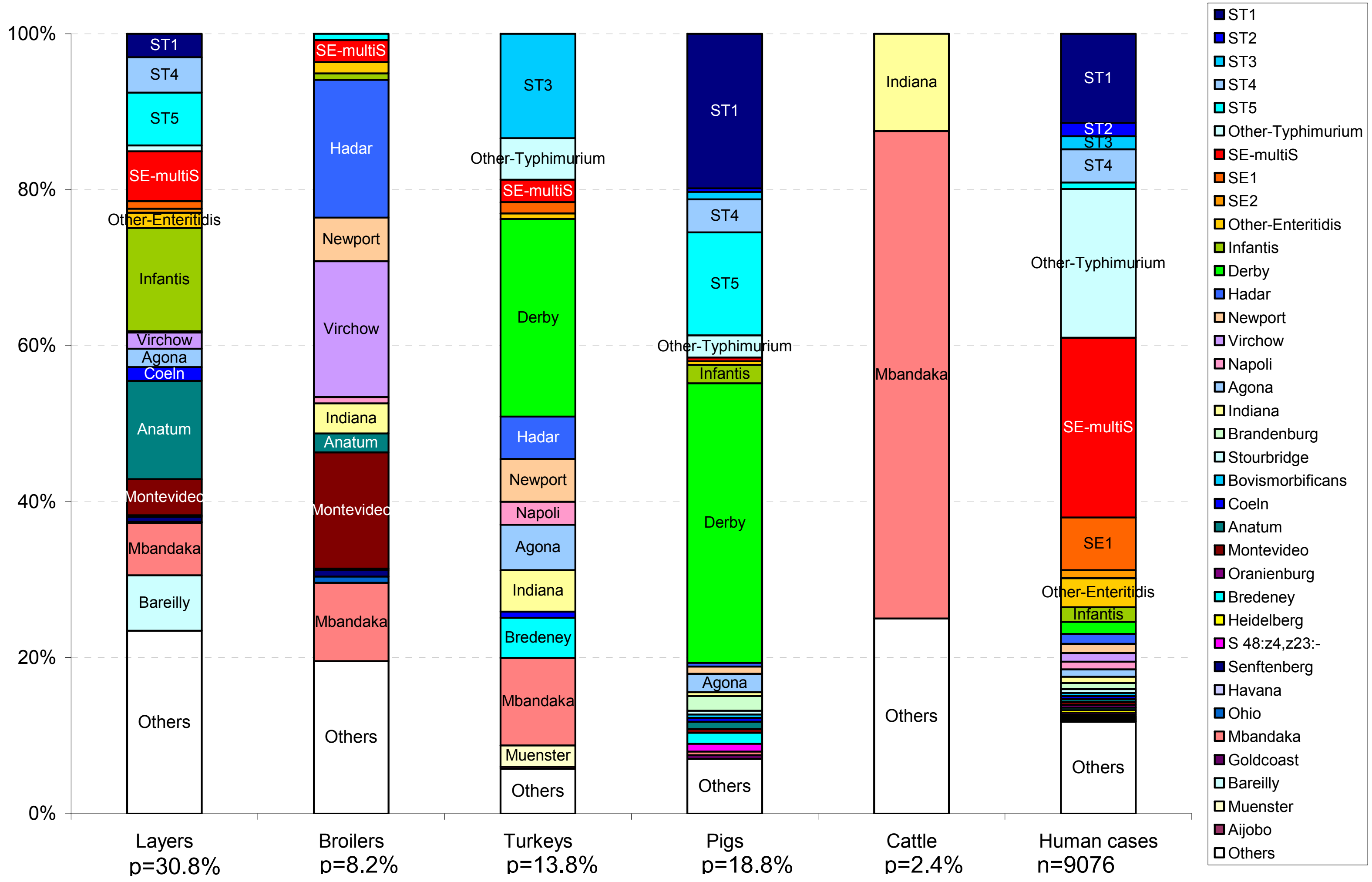

Figure 1: Surveillance results, types repartition in the sources and among the human cases, p indicated prevalence of the sources, n indicated the total number of human cases.

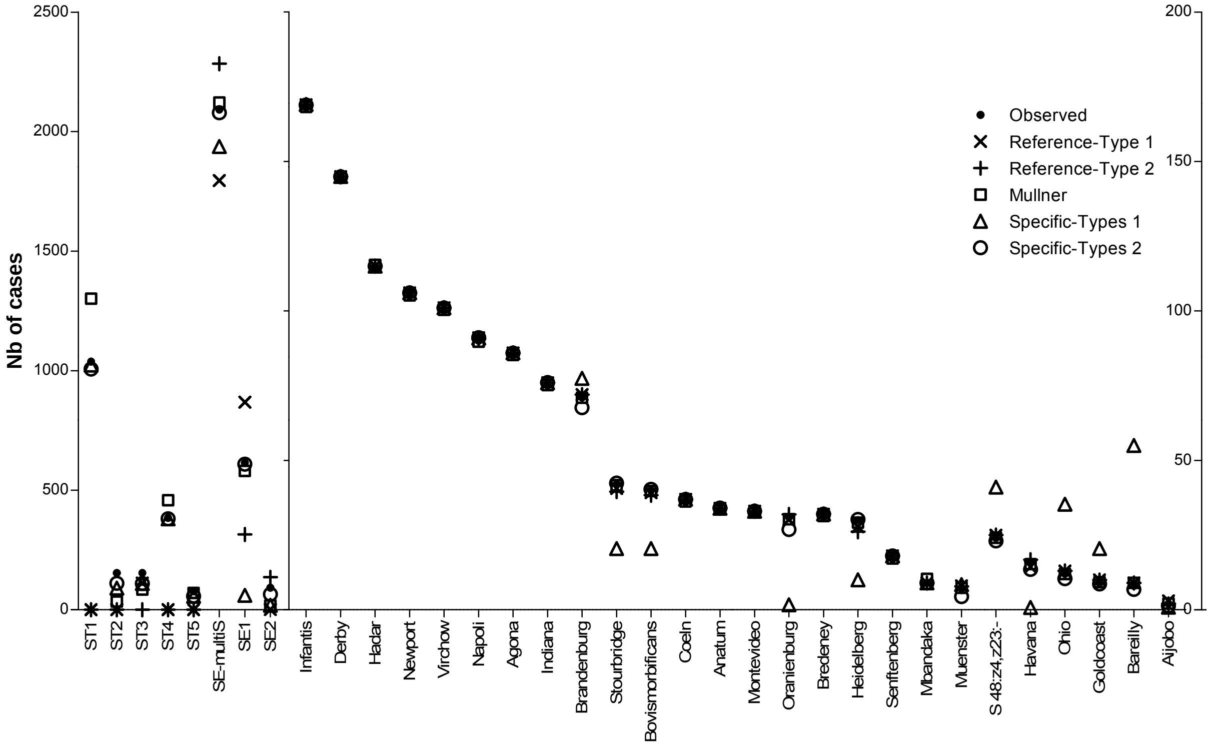

Figure 2: Comparison of posterior predicted cases per type and observed ones
The observed numbers of cases for the subtypes are those obtained with the deterministic reallocation

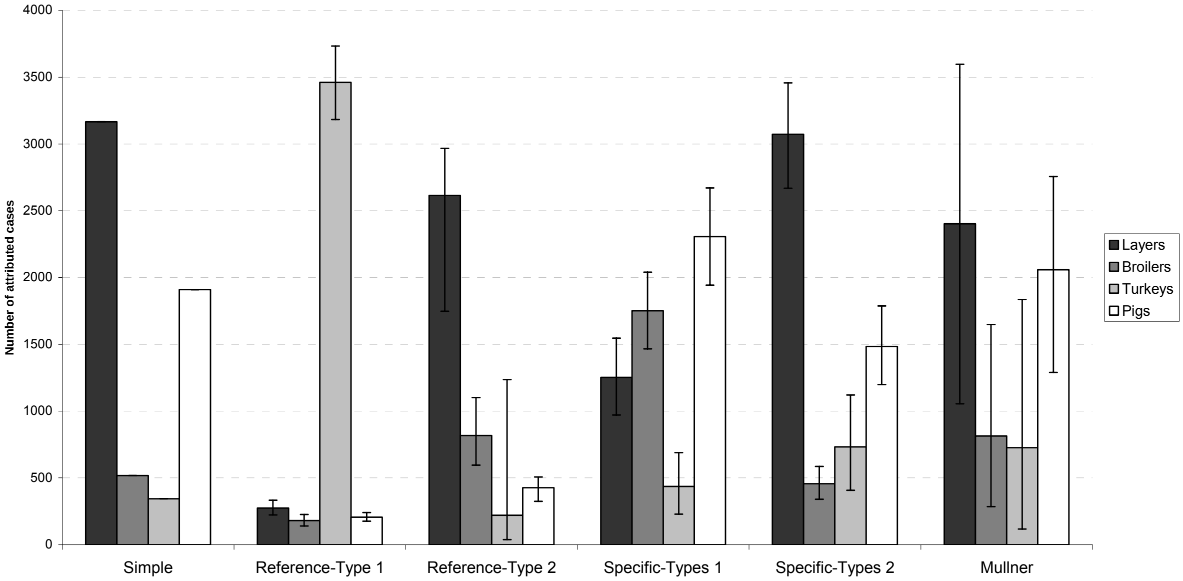

Figure 3: Attribution results for the simple model, the four parameterizations and Mullner's model

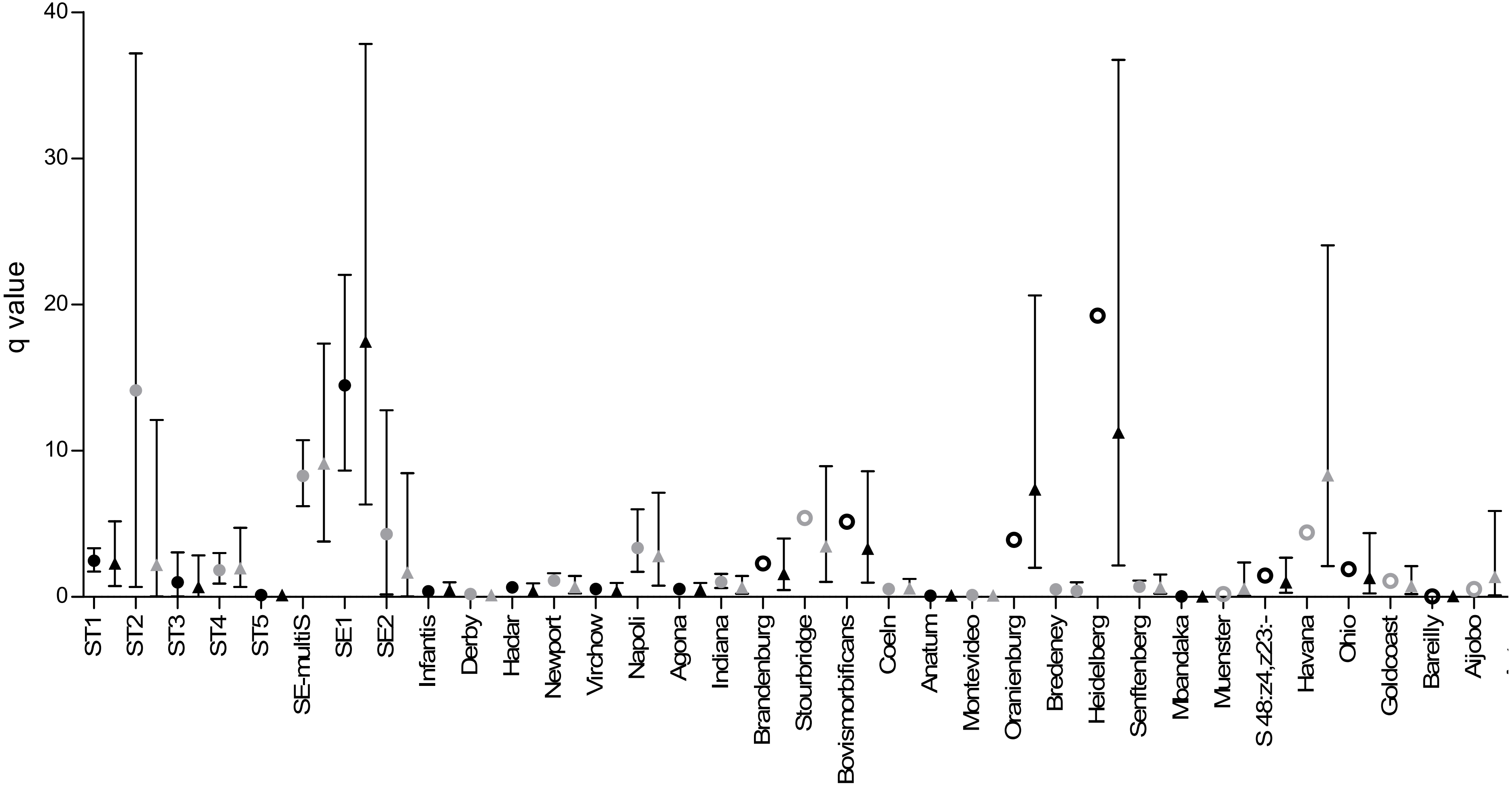

Figure 4: Posterior means and 95% CI of bacteria dependant parameters
Circles are for Specific-types 2 posterior means; empty circles indicate the fixed values; Triangles are set for Mullner's model posterior means.