Received: 18 September 2018 | Revised: 7 June 2019 | Accepted: 19 July 2019

DOI: 10.1111/zph.12645

ORIGINAL ARTICLE

WILEY

The role of parameterization in comparing source attribution models based on microbial subtyping for salmonellosis

Hannah Jabin | Guido Correia Carreira | Lars Valentin | Annemarie Käsbohrer

German Federal Institute for Risk Assessment, Berlin, Germany

Correspondence
Guido Correia Carreira, German Federal Institute for Risk Assessment, Max-Dohrn-Str. 8-10, 10589 Berlin, Germany.
Email: guido.correia-carreira@bfr.bund.de

Funding information
Bundesministerium für Bildung und Forschung, Grant/Award Number: 01KI1013A-H

Abstract

Source attribution methods attribute human cases of a zoonotic disease to a certain source putatively responsible for this disease. Identifying and quantifying the contribution of different sources to human infections is important for taking appropriate actions for reducing the exposure of the consumer to zoonotic pathogens. One widely used method is the microbial subtyping approach, whose principle is to compare the frequency of pathogen subtypes from different sources (e.g. animals or food) with the frequency of these subtypes in human cases. This paper studies the relationship between a Bayesian microbial subtyping approach described by Hald and coworkers subsequently modified by David and coworkers, here called the Hald model, and a frequentist approach known as the "Dutch model." The comparison between the Bayesian and frequentist model is done for two data sets on salmonellosis in Germany from different time periods (year 2004–2007 and 2010–2011). The results of both approaches are in good agreement with each other for the used data. It is shown here mathematically that a certain parameterization can be used to transform the probabilistic Hald model into a deterministic form, which is equivalent to the Dutch model. That certain parameterization secures independence of the model outcomes from the choice of so-called unique subtypes (which are unique in the sense that they are found exclusively in one of the sources). It is shown that deviating from that certain parameterization leads variations in the model outcome dependent on which unique subtypes are chosen in the process of modelling.

KEYWORDS
foodborne disease, Salmonella spp., source attribution model

1 | INTRODUCTION

Source attribution methods attribute cases of human foodborne infections to a potential source or vehicle, for example animals or food. Identifying and quantifying the contribution of the different zoonotic sources to human infections is important for reducing the exposure of the consumer to these zoonotic pathogens and for prioritization of intervention measures (Batz et al., 2005; Pires et al., 2009). Approaches for attributing foodborne illness to food sources include the analysis of outbreak data (Greig & Ravel, 2009; Pires, Vigre, Vigre, Makela, & Hald, 2010), epidemiological case–control studies of sporadic cases (Domingues, Pires, Halasa, & Hald, 2012a, 2012b), comparative microbial risk assessments (Evers, Fels-Klerx, Nauta, Schijven, & Havelaar, 2008; Kosmider et al., 2010), expert elicitation (Havelaar, Galindo, Kurowicka, & Cooke, 2008; Hoffmann, Fischbeck, Krupnick, & McWilliams, 2007) and microbial subtyping approaches (Hald, Vose, Wegener, & Koupeev, 2004).

Microbial subtyping approaches involve direct linking of human and animal data and are based on the assumption that the animal sources carry unique host-specific populations of microorganisms. There are several approaches to source attribution (for an overview see, EFSA, 2008a; Mughini-Gras, Franz, & Pelt, 2018) such as using exposure assessment (Pintar et al., 2017), expert elicitation (Hoffmann et al., 2017) or microbial subtyping (Glass et al., 2016). We will focus here on the microbial subtyping approach, which itself is based on different approaches what the typing is concerned. For an overview, see Barco, Barrucci, Olsen, and Ricci (2013). While classical approaches of typing used phenotypic techniques like phage typing (Hald et al., 2004) or a combination of phage types and antimicrobial resistance profiles (Hald, Wong, & Aarestrup, 2007), there is a recent rise in the use of molecular techniques (Boysen et al., 2014; Mather, Vaughan, & French, 2015) like whole-genome sequencing that continuously replace the phenotypic techniques. All of these different laboratory methods identify types of microbes (either by phage type, resistance profile or genetic fingerprint), and these types are then used in the source attribution models. The approach in our paper is based on phage typing data in order to estimate how many cases of human salmonellosis are attributable to which animal food source. But our results do not depend on the typing method since we focus on source attribution models and not on the way pathogens are typed. The mathematical source attribution models just need types. Whether these types were serotypes (as in Hald et al., 2004), resistance types (as in Hald et al., 2007) or genotypes (as in Miller, Marshall, French, & Jewell, 2017; Mullner et al., 2009) is secondary.

Although being very data-intensive (Mangen et al., 2010), during the last years a Bayesian microbial subtyping attribution model published by Hald et al. (2004) became increasingly popular and widely applied. This model, that we call 'Hald model' for short, was developed for attributing *Salmonella* cases in Denmark and has been adapted to attribute salmonellosis, for example in the United States (Guo et al., 2011), New Zealand (Mullner et al., 2009), Sweden (Wahlstrom, Andersson, Plym-Forshell, & Pires, 2011), France (David, Guillemot, et al., 2013), Italy (Mughini-Gras et al., 2014), the Netherlands (Mughini-Gras & van Pelt, 2014; Mughini-Gras et al., 2017), the European Union (Pires, Knegt, & Hald, 2011) and Australia (Glass et al., 2016). Similar attempts were conducted for attributing *Campylobacter* (Mullner et al., 2009; Ranta et al., 2011) and *Listeria monocytogenes* (Little, Pires, Gillespie, Grant, & Nichols, 2010) cases.

The original Hald model was built in 2004 for Denmark's extensive surveillance system. Its applications in other countries revealed that the model may show problems with identifiability and sparse data, which led to non-converging solution algorithms and therefore inconsistent model results or results with wide Bayesian credibility intervals (David, Guillemot, et al., 2013; Little et al., 2010; Mughini-Gras et al., 2014; Mullner et al., 2009). A variation of the Hald model, which deals with sparse data, was recently developed (Mikkela, Ranta, & Tuominen, 2019).

Several studies (Pires & Hald, 2010; Ranta et al., 2011) proposed an approach for solving the problem of identifiability by taking advantage of repeated data over several time steps (years) to provide independent information for the estimation. One essential requirement of this approach is the availability of consistent data sets for consecutive time steps. This precondition might not always be met.

Impacts

- We compared three source attribution models for foodborne salmonellosis and found them all to give essentially the same results for the used data on human salmonellosis in Germany. One model is the Bayesian David model (a variation of the Hald model with an alternative parameterization), the second model is the frequentist Dutch model, and the third is a variation of the Dutch model.
- We elucidated the relationship between the Hald/David model and the Dutch model by studying and supplementing the David parameterization and showed mathematically that the expected number of human cases in the Hald or David model becomes equivalent to that in the Dutch model.
- We study how different parameterizations affect the outcome of source attribution models for salmonellosis.

Other authors used repeated data and modified the model and/or its parameterization to overcome the difficulties of the original version (Ahlstrom et al., 2017; David, Guillemot, et al., 2013; Mullner et al., 2009). Furthermore, there is an approach, which sidesteps the parameterization issue by resorting to a non-parametric variation of the Hald model (Miller et al., 2017).

Another approach of microbial subtyping source attribution uses a frequentist approach and is known as the Dutch model. There have also been activities in order to modify the Dutch model in ways that its attributions do not seem to differ from the Hald model (Mughini-Gras et al., 2014).

This paper shows under which conditions the original Dutch model and the Hald model become mathematically equivalent. Overall, three models—a Bayesian one, a (frequentist) point estimating one and a (frequentist) interval estimating one—are compared as they were applied to two different data sets on salmonellosis in Germany. We call the three models "Bayes DB," "Deterministic" and "Deterministic UC," respectively. All models are used to estimate the expected number of human cases of salmonellosis attributable to potential food sources. The model 'Bayes DB' is based on a Bayesian source attribution model introduced by Hald et al. (2004) and modified by David, Guillemot, et al. (2013). The model we call "Deterministic" is also known as "Dutch model" (Barco et al., 2013; Van Pelt et al., 1999) and provides point estimates for the expected number of human cases. Via Monte Carlo Simulation, the results of the "Dutch model" were endowed with a distribution that models uncertainty and that was used for interval estimation in order to quantify uncertainty. This latter model version was named "Deterministic UC."

This paper compares two approaches considering microbial subtyping source attribution—the Hald model and the Dutch model—and then focusses on some aspects of one of the two approaches, namely the Dutch model. This does not reflect a preference for one of the two approaches—it just reflects the approach taken in this paper to discuss methodological aspects.

2 | MATERIAL AND METHODS

2.1 | Methods

2.1.1 | Hald attribution model and its characteristics

The Bayesian microbial subtyping source attribution model by Hald et al. (2004) compares the number of human cases, caused by different subtypes, with their prevalence in different food sources, weighted by the amount of food source consumed. Using the parameters defined in Table 1, the model can be reduced to two main equations:

$$o_i \sim \text{Poisson}\left(\sum_j \lambda_{ij}\right) \tag{1}$$

$$\lambda_{ij} = M_j \cdot p_{ij} \cdot q_i \cdot a_j \tag{2}$$

The model is based on the multiparameter prior defined in (2) with the two unknown parameters a_j and q_i that were parametrized in the following way.

$$a_j \sim \text{uniform}\,(0, 0.01) \tag{3}$$

$$q_i \sim \text{uniform}\,(0, 10) \tag{4}$$

Hald *et al.* propose that q_i reflects all subtype-dependent factors, for example the ability to survive during food processing or to cause disease in humans. The source-dependent parameter a_j, on the other hand, comprises differences between food types like differences in processing and preparation practices, or the capacity to act as a vehicle. The posterior distribution for the expected number of human cases (2) was calculated numerically using the Poisson likelihood function (1) based on the observed data and the prior distributions for the parameters a_j and q_i in Equations 3 and 4. The model was developed to estimate the contribution of animal food sources to human salmonellosis. It uses information from serotyping *Salmonella* strains and from phage typing the serotypes *S. enterica* serotype Enteritidis and *S. enterica* serotype Typhimurium. This microbial subtyping information leads to a subtype definition in the form of just "serotype" or "serotype + phage type," for example Enteritidis PT 4.

TABLE 1 Parameters of the Hald model

Parameter	Description
i	Index of subtype, $i = 1 \ldots I$
j	Index of source, $j = 1 \ldots J$
λ_{ij}	Expected number of human cases of subtype i in source j
M_j	Consumption of source j
q_i	Subtype-dependent factor for subtype i
a_j	Source-dependent factor for source j
p_{ij}	Prevalence of subtype i in source j
o_i	Observed human cases of subtype i

The original Bayesian model is over-parametrized; that is, there are more parameters ($I + J$) than observed data points (I). Hald *et al.* circumvent this problem by assuming equal values for the type-dependent parameters within the serotypes Enteritidis and Typhimurium and therefore reducing the number of parameters:

- $q_{SE} = 1$ for all Enteritidis subtypes
- $q_{ST} \sim$ uniform (0, 10) for all Typhimurium subtypes

These assumptions reduce the number of parameters $I + J$ by the number of Enteritidis subtypes I_E and by the number of Typhimurium subtypes I_T. If $I_E + I_T \geq J$, the reduction solves the problem of over-parametrization.

The outcome of the Hald model is an estimate for the parameters a_j and q_i, with which (together with (2)) the expected number of observed human cases attributed to each considered source for an infinite population can be calculated:

$$\lambda_j = \sum_i \lambda_{ij} \tag{5}$$

2.1.2 | David model—A variation of the Hald model

Since some authors describe difficulties of convergence when applying the Hald model, David, Guillemot, et al. (2013) compared four different parameterizations for the Hald approach. They concluded that for their data only one parametrization led to a valid model and this parametrization was called "Specific-Type DB." "DB" stands for "data based" and should indicate that in this parametrization data-based values from, what (David, Guillemot, et al., 2013) call, "specific types" of bacteria are used to set the corresponding q_i to fixed values. A "specific type" is a type of bacteria that is specific for a food source in the sense that it appears only in that one source and not in the other sources under consideration. We will use here the term unique type (and "*ut*" as subscript in corresponding variable names) instead of specific type.

If there are unique types of bacteria for a source j in the data, then the type-dependent parameters $q_{ut,j}$ for these unique types are parametrized by David, Guillemot, et al. (2013)) in their "Specific-Type DB" approach according to Equation 6.

$$q_{ut,j} = \frac{o_{ut}}{\sum_i o_i} \cdot \frac{1}{p_{ut,j}} \tag{6}$$

But the approach of David, Guillemot, et al. (2013)) allows the parametrization of (6) only for unique types that do not belong

to the serotypes Enteritidis or Typhimurium. Meanwhile, type-dependent parameters that belong to these two serotypes were parametrized probabilistically by drawing values from uniform distributions. David, Guillemot, et al. (2013)) excluded Enteritidis and Typhimurium subtypes from parametrization described by (6) to prevent possible interaction between the parametrization and the subtype reallocation process. They used Gamma distributions to reflect the uncertainty about the observed number of cases per type.

Some authors pointed out that consumption data M_j is not essential for the approach (Mughini-Gras & van Pelt, 2014; Mullner et al., 2009; Wahlstrom et al., 2011). According to them, M_j serves as a scaling factor for a_j and could be omitted (as done in Mullner et al., 2009; Wahlstrom et al., 2011); ; ;). Jumping ahead we note that setting M_j to 1 did work for our data only if the interval of the prior distributions for the parameters a_j and q_i was chosen to be large enough [compared to intervals of the prior distributions as in Expression (8)]. If they were not large enough, the model ran into problems either in not converging or producing inconsistent results. With inconsistent, we mean here and later in the paper that the predicted number of human cases for a certain bacteria type was far off from the number of human cases found in the data for that bacteria type.

2.1.3 | Bayes DB—A variation of the David model

David, Guillemot, et al. (2013)) noted that only as many parameters as sources considered have to be set to a fixed value according to Equation 6. Following this thought, we adapted the approach described by David, Guillemot, et al. (2013)) to the following parametrization set-up:

- For each source choose freely one unique type
- Parametrize for the chosen unique types the type parameters $q_{ut,j}$ according to Equation 6
- For each source j = *nut* where no unique type is available, the corresponding parameter a_{nut} related to this source is according to:

$$a_{nut} = \frac{\sum_i o_i}{M_{nut}} \tag{7}$$

- This also applies in the case, that no source has a unique type.
- If no consumption data M_j is available, all M_j are set to appropriate constant values large enough in order to assure consistent model results. These constant values are found through trial and error and depend on the particular data set one uses (as described in the previous section).
- Assume as priors uniform distributions for all remaining non-fixed parameters:

$$\begin{aligned} a_j &\sim \text{uniform}\,(0, 0.2) \\ q_i &\sim \text{uniform}\,(0, 1) \end{aligned} \tag{8}$$

It follows from this set-up that if there are no unique types, then all a_j are parametrized according to Equation 7 and all q_i according to $q_i \sim$ uniform (0, 1).

The uniform distributions of q_i and a_j both have as lower limit the value zero. Thus, during a model run it might be that the values of q_i or a_j, drawn from the prior distributions, might be zero. In that case, the model assumes that the bacteria or the sources do not have the ability to cause disease or act as a vehicle for foodborne infections and the expected number of human cases of salmonellosis, λ_{ij}, is likewise zero. The upper limits were chosen as to be sufficiently wide in order not to cut off significant parts of the posterior distributions of q_i and a_j. This was assured by examining visually the plots of the posterior distributions of q_i and a_j and checking that they were not cut off (we followed here the procedure described by Hald et al., 2004). There were slight differences for the two data sets used. It turned out that for the prior distribution of a_j an upper limit of 0.2 and for the prior distribution of q_i an upper limit of 1 was sufficient to assure that the posterior distributions were not cut off for both data sets. After parametrizing the model as described above, the posterior distributions for the unknown parameters a_j and q_i are calculated via Markov Chain Monte Carlo (MCMC) simulation. This model version with condensed data-based (DB) parameter reduction (using (6) and (7)) will be designated here as "Bayes DB."

The Bayes DB model predicts how many of human salmonellosis cases are caused by each source in our data (i.e. broiler, laying hens, pig, turkey). However, there are types of *Salmonella enterica* serotype Enteritidis and serotype Typhimurium that are put into the residual categories of "other phage types of serotype Enteritidis/Typhimurium" in Tables 2 and 3, usually because they individually caused very few human cases and/or were not observed in the animal samples. Particularly, the latter point makes it impossible to know how prevalent these subtypes are (if at all) in each of the various sources. Consequently, we attribute all *Salmonella* in categories of "other" to a source we call "unknown." We combine a deterministic approach in estimating the cases due to unknown cases as used in David, Sanders, et al. (2013) with a probabilistic approach in order to provide an uncertainty measure.

The deterministic approach consists simply in estimating the number of human cases due to unknown sources by adding up all the cases from *Salmonella* in the categories "other phage types of serotype Enteritidis/Typhimurium" and "other serotypes." The probabilistic element comes in the form of adding a sum of residual values uk_i to the deterministic values for each of the MCMC iterations. The sum of the residuals is given by $\sum_{i=1}^{N} uk_i$ where N is the number of all types (i.e. all types that do not have 'other' in front of their designation) and the number of unknown cases is then calculated for each iteration to be

$$\text{unknown} = o_{\text{other SE}} + o_{\text{other ST}} + o_{\text{other serovars}} + \sum_{i=1}^{N} uk_i \tag{9}$$

with

$$uk_i = o_i - \lambda_i \tag{10}$$

TABLE 2 Number of human cases and prevalences (in %) of Salmonella subtypes in animal sources from the baseline studies 2004–2007 (unit of consumption is usually mass in tons, but for laying hens, it is number of eggs)

Serotype	Observed human cases	Broilers	Laying hens	Pigs	Turkeys
S. enterica serotype Enteritidis					
PT 1	4,881	0.00	0.37	0.00	0.00
PT 11	200	0.00	0.00	0.04	0.00
PT 14b	2,552	0.00	0.09	0.00	0.31
PT 19	44	0.00	0.09	0.00	0.00
PT 2	488	0.00	0.09	0.00	0.00
PT 21	16,198	0.22	0.18	0.00	0.00
PT 21c	1,509	0.00	0.09	0.00	0.00
PT 25	976	0.00	0.09	0.00	0.00
PT 35	133	0.00	0.18	0.00	0.00
PT 4	79,013	0.00	9.18	0.16	0.00
PT 4a	67	0.00	0.09	0.00	0.00
PT 4b	200	0.00	0.09	0.00	0.00
PT 5a	111	0.00	0.09	0.00	0.00
PT 6	3,950	0.00	0.18	0.00	0.00
PT 6a	732	0.00	0.18	0.00	0.00
PT 7	444	0.00	0.18	0.00	0.00
PT 7a	954	0.00	0.09	0.00	0.00
PT 8	18,505	0.22	1.47	0.16	0.00
Other phage types of serotype Enteritidis	4,637	0.43	1.10	0.04	0.00
S. enterica serotype Typhimurium					
DT001	315	0.00	0.00	0.04	0.31
DT007	1,868	0.00	0.09	0.04	0.00
DT008	575	0.00	0.00	0.04	0.00
DT009	43	0.00	0.09	0.00	0.00
DT012	1,162	0.22	0.00	0.27	0.00
DT017	271	0.00	0.00	0.04	0.00
DT040	358	0.00	0.00	0.04	0.00
DT041	163	0.00	0.00	0.08	0.00
DT066	43	0.00	0.00	0.04	0.00
DT099	22	0.00	0.00	0.04	0.31
DT104	11,000	0.22	0.46	2.81	0.62
DT120	12,748	0.00	0.18	0.31	0.00
DT126	22	0.00	0.00	0.04	0.00
DT193	7,557	0.00	0.00	0.23	0.00
DT195	43	0.00	0.09	0.00	0.00
DT208	999	0.22	0.00	0.20	0.00
U302	999	0.00	0.00	0.04	0.00
U310	499	0.00	0.00	0.12	0.31
Other phage types of serotype Typhimurium	6,482	0.86	0.37	2.66	0.93
S. Agama	23	0.00	0.00	0.04	0.00
S. Agona	206	0.00	0.09	0.12	0.62
S. Anatum	317	2.81	0.00	0.12	0.00
S. Blockley	118	0.00	0.00	0.00	0.62

(Continues)

TABLE 2 (Continued)

Serotype	Observed human cases	Broilers	Laying hens	Pigs	Turkeys
S. Braenderup	245	0.00	0.00	0.04	0.00
S. Brandenburg	416	0.00	0.00	0.16	0.00
S. Derby	692	0.00	0.00	1.13	0.31
S. Eboko	11	0.00	0.00	0.08	0.00
S. Give	221	0.00	0.00	0.04	0.00
S. Goldcoast	564	0.00	0.00	0.04	0.00
S. Hadar	606	0.00	0.18	0.00	1.24
S. Havana	22	0.00	0.18	0.00	0.00
S. Heidelberg	66	0.43	0.00	0.00	0.00
S. Indiana	77	0.43	0.09	0.00	0.00
S. Infantis	1,926	1.30	0.83	0.31	0.00
S. Kedougou	40	0.00	0.00	0.04	0.00
S. Kottbus	140	0.00	0.00	0.00	0.93
S. Lexington	8	0.00	0.00	0.04	0.00
S. Liverpool	6	0.00	0.09	0.00	0.00
S. Livingstone	270	0.22	0.28	0.08	0.31
S. London	184	0.00	0.00	0.16	0.00
S. Mbandaka	90	1.51	0.37	0.00	0.00
S. Montevideo	227	0.00	0.09	0.00	0.00
S. Newport	542	0.00	0.09	0.00	0.31
S. Ohio	87	0.65	0.00	0.08	0.00
S. Rissen	72	0.00	0.28	0.04	0.00
S. Saintpaul	184	0.00	0.00	0.00	1.54
S. Tennessee	120	0.22	0.28	0.00	0.00
S. Virchow	702	0.43	0.00	0.00	0.00
Rough strains	21	0.00	0.00	0.04	0.00
Other serotypes	39,490	7.13	5.41	2.74	1.85
Number of positive herds (sampled herds)		66 (378)	132 (563)	326 (2,569)	31 (300)
Herd prevalence (p_j)		17.5	23.4	12.7	10.3
Food consumption		3,308,022	4,238,098	17,992,539	1,960,826

where λ_i is the model estimate of cases attributed to a type i, and o_i the corresponding observed number of human cases in the data. In other words, the uncertainty of the cases attributed to unknown values is assumed to be comparable to and therefore estimated by the deviations of the model estimates from the observed number of cases.

2.1.4 | The 'Deterministic model'—From the Hald model to the Dutch model

Beside the Bayesian approach from Hald et al., there is a frequentist source attribution approach (Mughini-Gras & van Pelt, 2014; Mullner et al., 2009). The model is also known as "Dutch model" after its country of origin (Barco et al., 2013; Van Pelt et al., 1999). The Dutch model calculates a point estimate of λ_{ij} according to the expression in (11) (Mughini-Gras et al., 2014).

$$\lambda_{ij} = \frac{p_{ij}}{\sum_j p_{ij}} o_i \tag{11}$$

We will show that the Hald estimate for the expected human cases is equivalent to the Dutch model if one uses the parametrization of q_i that belong to unique types proposed by David, Guillemot, et al. (2013) (cf. (6)) and the parametrization for a_j analogous to (7).

The expected value of the Poisson distribution is its Poisson parameter. For Equation 1, this relation leads to:

$$o_i \sim \text{Poisson}\left(\sum_j \lambda_{ij}\right)$$

$$E\left(o_i\right) = \sum_j \lambda_{ij} \tag{12}$$

TABLE 3 Number of human cases and prevalences (in %) of Salmonella subtypes in animal sources from the monitoring data 2010/2011 (unit of consumption is usually mass in tons, but for laying hens, it is number of eggs)

Serotype	Observed human cases	Broilers	Laying hens	Pigs	Turkeys
S. enterica serotype Enteritidis					
PT 1	722	0.00	0.05	0.00	0.00
PT 11	31	0.00	0.02	0.00	0.00
PT 14b	659	0.00	0.04	0.00	0.00
PT 2	314	0.16	0.04	0.00	0.00
PT 21	973	0.16	0.04	0.00	0.00
PT 3	63	0.00	0.02	0.00	0.00
PT 4	3,610	0.00	0.81	0.11	0.00
PT 6	63	0.00	0.04	0.00	0.00
PT 8	2,355	0.00	0.38	0.00	0.00
Other phage types of serotype Enteritidis	691	0.00	0.13	0.00	0.00
S. enterica serotype Typhimurium					
DT001	109	0.00	0.02	0.00	0.00
DT008	62	0.00	0.02	0.00	0.00
DT009	16	0.00	0.02	0.00	0.00
DT012	140	0.00	0.05	0.00	0.00
DT104	1736	0.00	0.00	1.09	0.19
DT120	1,254	0.00	0.00	0.55	0.00
DT193	3,558	0.00	0.04	2.84	0.19
U311	187	0.00	0.00	0.11	0.00
Other phage types of serotype Typhimurium	1,300	0.16	0.09	2.95	0.19
S. Anatum	15	1.14	0.02	0.00	0.00
S. Coeln	4	0.00	0.02	0.00	0.00
S. Derby	169	0.00	0.00	0.87	0.00
S. Hadar	50	0.00	0.02	0.00	0.00
S. Heidelberg	11	0.00	0.07	0.00	0.00
S. Hessarek	2	0.00	0.04	0.00	0.00
S. Infantis	395	0.16	0.02	0.00	0.00
S. Livingstone	53	1.47	0.02	0.11	0.00
S. Mbandaka	26	0.16	0.05	0.00	0.00
S. Montevideo	64	0.00	0.07	0.00	0.00
S. Ohio	28	0.00	0.00	0.33	0.00
S. Rissen	18	0.00	0.02	0.11	0.00
S. Saintpaul	35	0.00	0.00	0.00	0.28
S. Senftenberg	55	0.16	0.00	0.00	0.00
S. Stanley	31	0.00	0.00	0.11	0.00
S. Virchow	98	0.00	0.04	0.00	0.00
Other serotypes	6,484	0.81	0.45	0.22	0.19
Number of positive herds (sampled herds)		193 (4,354)	112 (4,247)	90 (962)	10 (971)
Herd prevalence (p_i)		4.4	2.6	9.4	1.03
Food consumption		929,671	1,088,613	4,480,500	499,400

TABLE 4 Data sources for Salmonella in animals and humans

Data Set	Broilers	Laying hens	Pigs	Turkeys	Human Data
Baseline 2004–2007	BS Broilers 2005/2006[a]	BS Laying hens 2004/2005[b]	BS Pigs (lymph nodes) 2006/2007[c]	BS Turkeys 2006/2007[d]	RKI[e]
Monitoring 2010/2011	Control Program 2010[f]	Control Program 2010[f]	National Monitoring 2011[g]	Control Program 2010[f]	RKI[e]

Abbreviations: BS, Baseline Study; RKI, Robert Koch Institute.
[a]EFSA (2007b).
[b]EFSA (2007a).
[c]EFSA (2008b).
[d]EFSA (2008c).
[e]RKI (2012) and personal communication.
[f]Kaesbohrer et al. (2012).
[g]Kaesbohrer et al. (2013).

If one assumes that the observed value of human cases caused by bacteria of type i (i.e. o_i) can be estimated by its expected values $E(o_i)$, then we find.

$$o_i \approx E(o_i) = \sum_j \lambda_{ij} \tag{13}$$

For a unique type with index $i = ut$ in the individual source j, Equation 13 reduces to a single element of the sum $\sum_j \lambda_{ij}$:

$$\begin{aligned} o_{ut,j} &= \lambda_{ij} \\ &= M_j \cdot p_{ut,j} \cdot q_{ut,j} \cdot a_j \end{aligned} \tag{14}$$

inserting (6) in (14) and solving for a_j yields:

$$\begin{aligned} o_{ut,j} = M_j \cdot p_{ut,j} \cdot \frac{o_{ut,j}}{\sum_i o_i} \cdot \frac{1}{p_{ut,j}} \cdot a_j \\ \Leftrightarrow a_j = \frac{\sum_i o_i}{M_j} \end{aligned} \tag{15}$$

This sample calculation shows that by fixing the $q_{ut,j}$ to constant values for a unique type in source j as done in (6), a_j is automatically defined as in (15).

Although we derived this way of fixing values of a_j only for unique types, we will use it also for the cases that there are no unique types. Thus, we always use (15) for setting a_j to fixed values. A similar approach has been used above in the Bayesian approach by setting the values of a_{nut} in Equation 7. As for the q_i, we use Equation 6 for setting q_i to fixed values if q_i refers to a unique type. If q_i does not refer to a unique type, it is set to values according to the following derivation that starts from Equation 13.

$$\begin{aligned} o_i \approx \sum_j \lambda_{ij} = q_i \cdot \sum_j a_j M_j p_{ij} \\ \Leftrightarrow \quad q_i = \frac{o_i}{\sum_j a_j M_j p_{ij}} \end{aligned} \tag{16}$$

Now that all values for q_i and a_j for all types and sources are known, the contribution of each source to the number of human cases (λ_j) can be derived using Equations 2 and 5. Under these circumstances, the Hald model estimate for the expected human cases given by (2) is equivalent to the Dutch model, that is (11) as we will show now. In order to do so, we consider two cases.

First case: No unique types

Starting point is the estimate for the expected human cases proposed by Hald et al. (2004). Since in this case, there are no unique types we set the values of q_i according to Equation 16 and the values of a_j according to Equation 15.

$$\begin{aligned} \lambda_{ij} &= M_j \cdot p_{ij} \cdot q_i \cdot a_j \\ &= M_j \cdot p_{ij} \cdot \frac{o_i}{\sum_j a_j M_j p_{ij}} \cdot \frac{\sum_i o_i}{M_j} \\ &= p_{ij} \cdot \frac{o_i \cdot \sum_i o_i}{\sum_j \frac{\sum_i o_i}{M_j} M_j p_{ij}} \\ &= p_{ij} \cdot \frac{o_i \cdot \sum_i o_i}{\sum_i o_i \cdot \sum_j p_{ij}} \\ \Leftrightarrow \quad \lambda_{ij} &= \frac{p_{ij}}{\sum_j p_{ij}} o_i \end{aligned}$$

Hence, the first case leads to the Dutch model (cf. Equation 11).

Second case: There are unique types

If there are unique types and one follows the suggestion of David, Guillemot, et al. (2013) for setting as many q_i as there are sources to constant values according to Equation 6, then we get the following situation. For all types i, which are not unique types, we have the situation of the first case and hence have the Dutch model:

$$\lambda_{ij} = \frac{p_{ij}}{\sum_j p_{ij}} o_i$$

For all other types, which are unique types (i.e. $i = ut$) for the source j, we set the corresponding values of $q_i = q_{ut,j}$ according to Equation 6. The a_j is again set to values according to Equation 15.

$$\begin{aligned} \lambda_{ij} &= M_j p_{ij} q_i a_j \\ &= M_j p_{ij} \frac{o_i}{\sum_i o_i} \cdot \frac{1}{p_{ij}} \cdot \frac{\sum_i o_i}{M_j} \\ \lambda_{ij} &= o_i \end{aligned} \tag{17}$$

Equation 17 again is the Dutch model, namely for the case of unique types that are per definition types for which one has $\sum_j p_{ij} = p_{ij}$. That is, all p_{ij} are zero except for the one source j for which the bacteria type i is unique. In this case, the Dutch model in the form of Equation 11 reduces to the form of (17).

Given that o_i is approximated by its expected value $E(o_i)$ (cf. Equation 13) and given that the type parameters q_i for unique types are fixed to values according to Equation 6 and given that in all other cases the q_i are fixed according to Equation 16 and given that the source parameters a_j are fixed to values according to Equation 15, the Hald estimate for the number of expected human cases (cf. (2)) becomes the same as in the Dutch model (cf. (11)).

We just demonstrated how a certain parameterization of q_i and a_j deterministically (instead by means of prior distributions) leads to a deterministic result that is in fact the Dutch model. However, one may choose a different deterministic parameterization of q_i and a_j and see what the consequences are. The result will always be a deterministic point estimate for λ_{ij}. Hence, we will designate the Hald formula from Equation 2 combined with any deterministic parameterization of q_i and a_j as "Deterministic model."

2.1.5 | Influence of parameterization of the Deterministic model

Influence of choice of the subtype-dependent parameter value

We studied how the results of the Deterministic model changed for two different parameterizations of unique subtypes $q_{ut,j}$. In particular, for each of the parameterizations, the effect of choosing a particular subtype was investigated.

In the first parameterization, $q_{ut,j}$ was set to the proportion of human cases that are caused by the unique subtype:

$$q_{ut,j} = \frac{o_{ut}}{\sum_i o_i}$$

For testing the influence of this parameterization on the estimated source contributions, we used the Deterministic model to calculate the results for all possible choices of unique types for a certain source, namely that of laying hens. We did our comparison of parameterizations using the Monitoring 2010/2011 data. That data contained 16 subtypes that were unique for the source laying hens. We ran the Deterministic model 16 times. Each time the same unique types for broilers, pigs and turkeys were used, while the unique type for laying hens was changed to be each of the 16 available ones.

In the second parameterization, $q_{ut,j}$ was set according to Equation 6 repeating the model runs for the 16 different available unique types for laying hens.

Influence of choice of the source-dependent parameter value

In our description of the Bayes DB model, we said that for each source j = *nut* where no unique type is available the corresponding source-dependent parameter a_{nut} should be fixed according to Equation 7. We explored what happened if one deviates from this procedure. Say, we set a_{nut} to 1:

$$a_{nut} = 1.$$

We assumed that all except one of sources had (at least) one unique type. The source-dependent parameter for the source lacking a unique type should be fixed setting the corresponding a_{nut} to 1. Each of the four sources was once chosen to be the source lacking a unique source, which resulted in four model runs. The same was done using the parameterization as defined in Equation 7. The results of the two parameterizations were subsequently compared. This comparison of parameterizations used again the Monitoring 2010/2011 data.

2.2 | Introducing uncertainty in prevalence data

Setting specific parameters to a data-based value and solving the model equations deterministically as described above leads to a straightforward opportunity of introducing uncertainty regarding the data into the modelling framework.

As an example, we introduce uncertainty regarding the herd prevalence for each source via a Monte Carlo simulation approach. The principle works as follows. The prevalences p_{ij} in Tables 2 and 3 are point estimates that are calculated by multiplying the type prevalence by the herd prevalence. Type prevalence here means the number of all typed samples that were identified as type i in source j [designated as $(NS)_{ij}$] divided by the total number of typed samples of source j, that is $\sum_i (NS)_{ij}$. Herd prevalence for source j here means the number of all herds (of animals of source j) that tested positive for *Salmonella* divided by the total number of tested herds for source j. This herd prevalence is designated here as p_j. The just outlined procedure for calculating the prevalence p_{ij} is expressed in mathematical terms in the following equation.

$$p_{ij} = \frac{(NS)_{ij}}{\sum_i (NS)_{ij}} \cdot p_j \tag{18}$$

Now we introduce uncertainty in the prevalence p_{ij} by drawing 2,000 values for the herd prevalence p_j from a beta distribution instead of using the point estimates of p_j (which are listed at the end of Tables 2 and 3). A similar approach to introduce uncertainty in the Dutch model was used by Mughini-Gras et al. (2014) and Mullner et al. (2009).

From the sampling scheme, the sample size n_j (i.e. the number herds tested for *Salmonella* in source j) and the number k_j of positive tested herds are known for each source j. The herd prevalences p_j are drawn from the following beta distribution, using these k_j positive tests out of n_j trials.

$$p_j = \text{beta}\left(k_j + 1, n_j - k_j + 1\right).$$

A prevalence p_j is sampled for each source j and the model equations of the attribution model are solved deterministically (using Equation 11) with the sampled prevalences. Using the

Monte Carlo method, the results are stored, and in a next iteration step, new prevalences are sampled. Going through this process 2000 times for each of the four sources, the calculated solutions take the form of a distribution. By analysing this distribution, the impact of uncertainty of prevalences on the results of the deterministic approach can be evaluated. This model version that incorporates uncertainty will be further designated as "Deterministic UC."

2.3 | Data

Two data sets, covering different studies on *Salmonella* in different time periods, were compiled and used for this analysis (see Table 4). For both time periods, reliable data from active monitoring on four potential animal reservoirs were available: broilers, laying hens, pigs and turkeys. Cattle were not included in any of these studies or programmes and were therefore not included in this analysis.

2.3.1 | *Salmonella* baseline studies 2004–2007

The first data set was generated by four baseline studies conducted during 2004 and 2007 in Germany (see Tables 2 and 4). These data describe for each animal reservoir the total number of samples, the number of samples positive with *Salmonella* and their serotypes. Additionally, for all strains belonging to serotype Enteritidis and Typhimurium, the phage type was determined.

The prevalence p_{ij} of each *Salmonella* subtype *i* in sources *j* was estimated by multiplying the relative frequency of the subtype in source *j* with the prevalence p_j of *Salmonella* in source *j* (cf. Equation 18). The prevalence p_j was also reported to the European Food Safety Authority (EFSA) by the Federal Institute for Risk Assessment (BfR) due to its obligations of reporting these statistics.

For the model Deterministic UC, we derived the uncertainty regarding the *Salmonella* prevalences in the particular sources from the known sample size and the number of positive findings (see Table 2).

2.3.2 | *Salmonella* monitoring data 2010/2011

The second data set was generated by studies conducted during 2010 and 2011 in Germany (see Tables 3 and 4). The data sets for broilers, laying hens and turkeys were obtained from *Salmonella* control programmes in poultry, coordinated by the BfR on behalf of the German government in year 2010. Since national monitoring of *Salmonella* prevalence in pigs was conducted only in 2011, the data on pigs from 2011 were combined with the data sets from 2010, assuming no significant change in serotype distribution between the 2 years.

Analogous to the baseline data set, the control programme and the national monitoring describe the number of samples taken in each animal reservoir, the number of samples positive for *Salmonella* and their serotype and phage type.

The prevalence of each subtype in all animal sources was estimated according to Equation 18 with p_j reported by BfR to EFSA. For estimating the uncertainty for the source prevalences, we followed the same approach as described for the baseline studies data set (see Table 3).

2.3.3 | Food consumption

For both data sets, food consumption data were obtained from the Federal Office for Agriculture and Food (BLE). Since the baseline studies cover several years, the consumption data for each animal food source were added up for the years 2004–2007 (BLE, 2008, 2005, 2006, 2007, 2011) (Table 2). Regarding the monitoring data set 2010/11, only consumption data from 2010 were considered (BLE, 2011), since most data were gathered during this year and no significant change in consumption patterns between 2010 and 2011 is assumed (Table 3).

2.3.4 | Human data

Data on human *Salmonella* cases were provided by the Robert Koch Institute (RKI). The serotype distribution was obtained via their online database SurvStat@RKI (http://www3.rki.de/SurvStat, data access: 07.02.2012). In addition, phage-type information for *S.* Enteritidis and *S.* Typhimurium strains was provided via personal communication by Wolfgang Rabsch (RKI).

SurvStat@RKI lists all *Salmonella* cases reported in a particular year without distinguishing between outbreak, sporadic, domestic or travel-related cases. We applied these data under two assumptions.

First, during the time period considered in Germany travel-related cases contributed <10% to the overall number of cases recorded and are therefore considered negligible compared with the domestically acquired cases.

The second assumption is that the outbreak-related cases do not alter the serotype distribution significantly. This is supported by the observation that most of the identified major outbreaks were linked to *S. enterica* ser. Enteritidis, which was also the dominating serotype in

TABLE 5 Estimated human Salmonella cases attributed to the four animal sources and an unknown source for the Baseline 2004–2007 data set: attribution results for the three different model versions

Source	Bayes DB Mean [$CI_{0.5}$, $CI_{0.95}$]	Deterministic (Dutch model)	Deterministic UC Mean [$CI_{0.5}$, $CI_{0.95}$]
Broilers	14,920 [14,280;15,550]	14,977	14,924 [13,198;16,425]
Laying hens	123,700 [122,800;124,700]	123,685	123,742 [121,428;126,068]
Pigs	33,190 [32,670;33,730]	33,118	33,209 [31,968;34,475]
Turkeys	5,986 [5,645;6,377]	6,067	5,972 [5,202;6,608]
Unknown	50,610 [49,770;51,430]	50,609	50,609 [50,609;50,609]

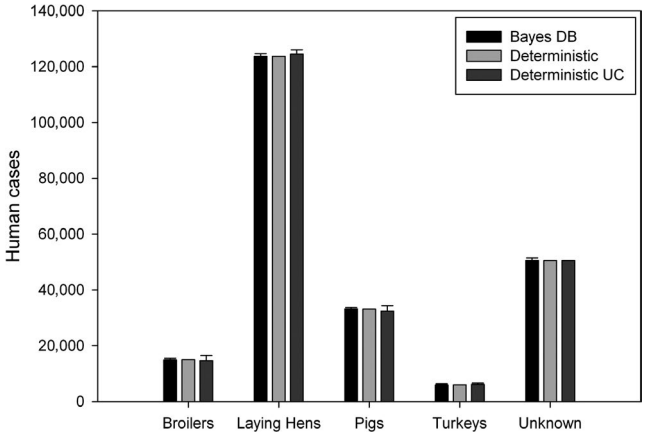

FIGURE 1 Estimated human Salmonella cases attributed to the four animal sources and an unknown source for the Baseline 2004–2007 data set: attribution results for the three different model versions

the overall number of cases. While—as mentioned—the Salmonellosis cases reported by SurvStat@RKI generally do not distinguish between outbreak-related, sporadic, domestic or travel-related cases, there are important exceptions. The phage-type information for the serovars Enteritidis and Typhimurium was adjusted by the RKI regarding the outbreak cases and only reflects the sporadic cases.

Since only a part of all strains was phage typed, we assumed that the phage-type distribution within the part of the strains that have been typed is also representative for the untyped strains.

To account for the four-year time period of the baseline studies (from 2004 to 2007), we summed up all the corresponding serotype and phage-type cases over that time period leading to the numbers of human cases in Table 2. For the monitoring data set, only the serotype and phage-type distributions from 2010 were considered (Table 3).

Human and animal data were thus matched in the following sense: we looked at the Salmonella types found in human salmonellosis cases. Then, we looked in the animal data where these human-related types appeared at least once. If a human-related type appeared at least in one of the four considered animal sources, this type was chosen as data to be included in Tables 2 and 3.

3 | RESULTS

The results presented for the Bayes models correspond to runs of five independent Markov chains of 50,000 iterations using MCMC simulation with the Gibbs sampler. The original Hald model and the Bayes DB model were set up in the OpenBUGS software (Lunn, Spiegelhalter, Thomas, & Best, 2009) and in R (R Core Team, 2013). Convergence was monitored applying Gelman Rubin convergence diagnostics (Toft, Innocent, Gettinby, & Reid, 2007).

The Deterministic models were set up in R, and the results of the Deterministic UC approach correspond to 2000 iterations using Monte Carlo simulation.

The source code for the models is available upon request.

The application of the original Hald model on both data sets results in inconsistent model estimations. The high proportion of unique subtypes (Baseline 2004–2007:42/66 = 0.64; Monitoring 2010/2011:23/33 = 0.70) in combination with the procedure to fix values proposed by Hald et al. (2004) leads to non-converging model solutions. We decided on discarding these model estimates, since the results were considered not reliable.

3.1 | Attribution results

3.1.1 | Total number of human cases

The predicted and observed total number of *Salmonella* cases were in good agreement for all three model versions (Bayes DB, Deterministic and Deterministic UC) and both data sets (Baseline 2004–2007, Monitoring 2010/2011). For the Baseline 2004–2007 data set, Bayes DB predicted the total number of *Salmonella* cases in Germany in 2004–2007 to n = 228,406. Deterministic and Deterministic UC both predicted them to n = 228,456, which corresponds with the number of the observed cases n = 228,456. For the Monitoring 2010/2011 data set, all three model versions predicted the total number of human cases to n = 25,380, which is in accordance with the number of the observed cases of 25,381.

3.1.2 | Role of food consumption weights

To check the Bayes DB approach for its sensitivity towards the consumption weights M_j, we assumed that the amounts of consumption for the four sources were unknown. Accordingly, all M_j were set to 1. Unless the prior distributions for the parameters a_j and q_i were adjusted, this led for both data sets to problems in the following way. Setting all M_j to 1 required choosing larger prior distributions than defined in Expression (8). When setting all M_j to 1 and using prior distributions as defined in Expression (8), OpenBUGS was not able to run the model at all, due to numerical problems (OpenBUGS reports an 'conjugate gamma updater error' for one of the q_i). This problem disappeared when the priors for q_i were changed to q_i ~ uniform (0, 2,600,000). At the same time, the priors for a_j had to be changed to a_j ~ uniform (0, 300,000) in order to avoid inconsistent results.

For the monitoring data setting, all M_j to 1 and using prior distributions as defined in Expression (8) led to inconsistent results. The model worked properly when setting priors were set to a_j ~ uniform (0, 30,000).

One way to interpret the need for enlarging the priors for q_i and a_j is the following: taking M_j effectively out of the model by setting all its values to 1 needs to be compensated by the model via corresponding changes in the parameters q_i and a_j. Since one may consider q_i and a_j as a complex prior, that is as a combined estimate of the potential to inflict salmonellosis, it makes sense that both prior distributions may need to be adjusted in the case that M_j becomes ineffective, for example when setting it to 1.

Source	Bayes DB Mean [$CI_{0.5}$, $CI_{0.95}$]	Deterministic	Deterministic UC Mean [$CI_{0.5}$, $CI_{0.95}$]
Broilers	1,520 [1,417; 1,625]	1,549	1,547 [1,502; 1,587]
Laying hens	7,977 [7,782; 8,180]	7,951	7,952 [7,828; 8,068]
Pigs	6,896 [6,670; 7,121]	6,900	6,896 [6,643; 7,139]
Turkeys	516 [371; 701]	506	510 [287; 735]
Unknown	8,471 [8210; 8725]	8,475	8,475 [8,475; 8,475]

TABLE 6 Estimated human Salmonella cases attributed to the four animal sources and an unknown source for the Monitoring 2010/2011 data set: attribution results for the three different model versions

3.1.3 | Source estimates

For both data sets, all three model versions resulted in very similar source. Convergence issues did not appear for Bayes DB on both sets. For the Baseline 2004–2007 data set, all models identified laying hens as the main source, followed by pigs, broilers and turkeys (Table 5 and Figure 1). All observed human cases that could not be assigned to one of the considered *Salmonella* subtypes (i.e. falling into one of the 'others' categories, see Table 2) are attributed to the source 'unknown', since they could not be connected to one of the considered animal sources. All three models assigned 22.2% of all cases to the unknown source.

The three model versions resulted in the same source ranking also for the Monitoring 2010/2011 data set (Table 6 and Figure 2). A third of all cases (33.4%) could not be attributed to one of the considered sources and was assigned to the source "unknown."

3.1.4 | Propagation of prevalence uncertainty

Assuming uncertainty regarding the prevalence values did not change the source ranking for both data sets. The mean source attribution values of the Deterministic UC approach correspond to the calculated source values of the Deterministic model and the posterior mean value of Bayes DB. The Deterministic UC and Bayes DB approaches yielded 90% confidence intervals for all sources of comparable size (see Tables 5 and 6).

3.2 | Influence of parameterization of the deterministic model

3.2.1 | Influence of choice of the subtype-dependent parameter value

We compared two different parameterizations of unique subtypes $q_{ut,j}$. The first parameterization:

$$q_{ut,j} = \frac{o_{ut}}{\sum_i o_i}$$

was used as described in the Material and Method section above. The corresponding results of these simulations are shown in Figure 3a. The estimates for the number of human cases caused from different sources vary with the chosen unique type in the "laying hens" source.

Repeating this approach, setting $q_{ut,j}$ to values as defined in Equation 6 leads to consistent model results that were independent of the choice of unique type for laying hens (see Figure 3b).

3.2.2 | Influence of choice of the source-dependent parameter value

Likewise, we compared two different parameterizations of the source-dependent parameter a_j.

In one parameterization, we set $a_{nut} = 1$ for one of the four sources (broilers, egg layers, pig and turkey). The results for the Monitoring 2010/2011 data set are shown in Figure 4a. The source attribution estimates are clearly dependent on the choice of which source was supposed to have no unique type.

In a second step, the same procedure was used except that the value of a_{nut} was fixed as defined in Equation 7. The results in Figure 4b show that in following this approach the model leads to consistent results, regardless of which source was chosen to lack a unique type. This underlines the suitableness of this parameterization.

4 | DISCUSSION

4.1 | Reducing the parameters with data-based values

David, Guillemot, et al. (2013) published a data-based approach for fixing values for the type-dependent parameters for all unique types, excluding Enteritidis and Typhimurium subtypes. This approach requires having at least as many unique types as food sources included

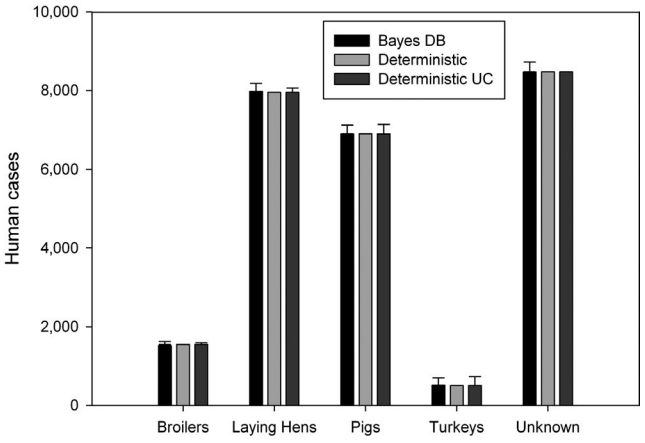

FIGURE 2 Estimated human Salmonella cases attributed to the four animal sources and an unknown source for the Monitoring 2010/2011 data set: attribution results for the three different model versions

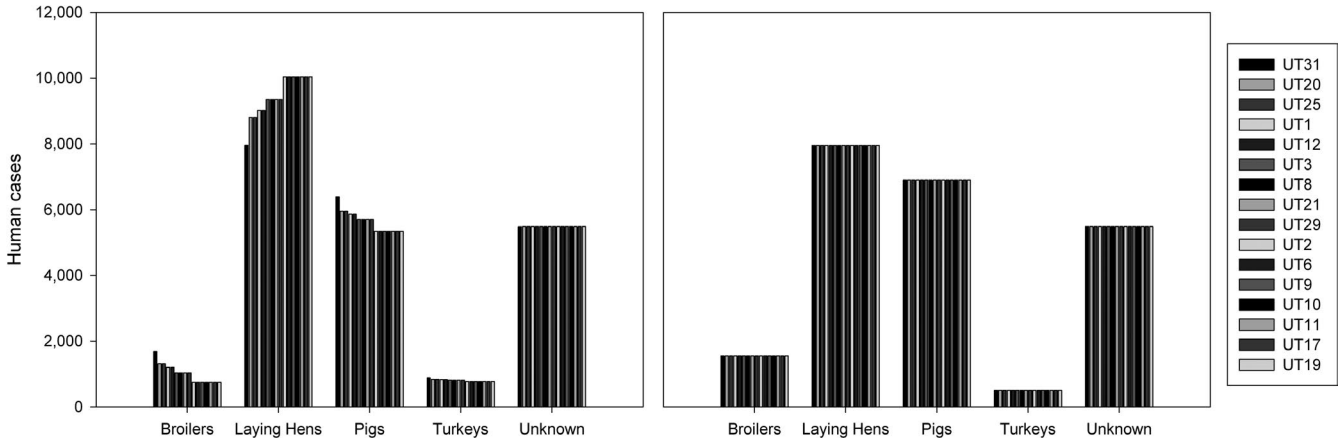

FIGURE 3 Attribution results of the Deterministic model for the Monitoring 2010/2011 data set. In each of the 16 model runs, a different unique type (UT) in laying hens is chosen for setting the values of $q_{ut,j}$: a) results for $q_{utj} = \frac{o_{ut}}{\sum_i o_i}$ and b) results for $q_{utj} = \frac{o_{ut}}{\sum_i o_i} \cdot \frac{1}{p_{utj}}$

in the data. But it is not necessary that there are unique types in each of the sources.

Therefore, we proposed an alternative approach that fixes one unique type per source and, in the case of a missing unique type, the setting of a fixed value for the source-dependent parameters. This scheme allows for the application on a much wider range of data sets with arbitrary numbers of unique subtypes without causing any problems with convergence.

In a second step, we evaluated the dependence of the model results from the approach for fixing the values for $q_{ut,j}$ and a_{nut}. The results showed model outputs can be very sensitive to the values one chooses for $q_{ut,j}$ and a_{nut}. Parametrization of $q_{ut,j}$ according to Equation 6 shows no dependence on the choice of unique subtypes. Parametrizing the source-dependent parameter a_{nut} according to Equation 7 follows directly from the model equations, and the results obtained are independent of the source that is fixed.

A potential limitation of this approach is that it partly loses its direct interpretation of the parameters q_i and a_j as described by Hald et al. (2004). The fixing value for $q_{ut,j}$ is directly proportional to the fraction of human cases caused by the corresponding subtype and inversely proportional to its prevalence. The higher the fraction of human cases and the smaller its prevalence, the higher its type-dependent value becomes. These relations reflect the interpretation of the type-dependent parameter as mirroring the ability of the subtype to cause disease in humans. From this interpretative parameter definition, all other values for a_{nut}, q_i and a_j are derived. As a consequence, their values are less directly interpretable than in the original Hald model.

4.2 | Comparing Bayes DB and deterministic model

The conversion of the Bayes DB model into a deterministic equation system led to the same source attribution estimates. That is, we found for our data sets that the Deterministic model results equal the posterior means of the Bayes DB model.

This equality in results may have different sources. It might be that the parameterization in setting specific types to data-based values is in part responsible for the convergence of the results between Bayes DB and the Deterministic model. Similar effects were reported by David, Guillemot, et al. (2013). They found that setting specific parameters to data-based values led to attribution results of their Bayesian approach comparable to their simple Deterministic model, which has some similarity with the Dutch model.

It might also be that the converging results are in part a result of the specific data analysed here. A report in the grey literature (Hald, 2002) produced during the Food and Agriculture COST Action 920 in the Year 2001 compares the classic Hald model with the Dutch model for two data sets on possible human salmonellosis from various food sources. One data set came from Denmark another from the Netherlands. While the Hald and the Dutch Model largely agreed in the estimated number of human cases of salmonellosis attributable to certain food sources when using the Danish data, they differed when using the Dutch data. The author of the COST Action report suspects that the difference in data quality might be an issue here. Namely that the Dutch data for example might lack a clear distinction between egg layers and slaughter poultry. Hald (2002) suggests that this "may explain why both models had difficulties in determining, which of the sources—egg or poultry—were the more important" for the Dutch data. Returning to the converging results we found, one could reason that the convergence hints at comparable data quality between the two data sets (baseline and monitoring).

4.3 | Source attribution estimates

Taking the results at face value, the results of all model versions shown in Tables 5 and 6 identified laying hens (i.e. eggs) as the main source for human salmonellosis in Germany, followed by pigs, broilers and turkeys. This ranking confirms the results of a study conducted for EFSA by Pires et al. (2011). Their study provided estimates on the relative contribution of different food and animal sources to human *Salmonella* infections in the European Union and its Member States. Data used in the models are based on serotype information only and covered the period from 2006 to 2009. For Germany, they

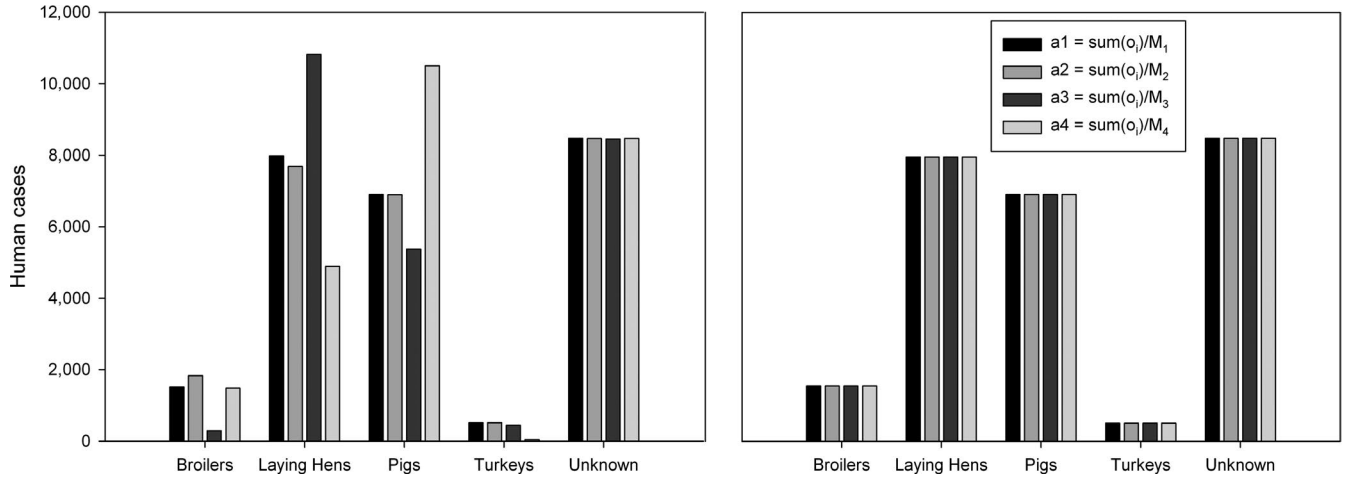

FIGURE 4 Attribution results of the Deterministic model for the Monitoring 2010/2011 data set. In each of the 4 model runs, a different source is fixed: a) results for $a_{nut} = 1$ and b) results for $a_{nut} = \frac{\sum_i o_i}{M_{nut}}$

reported source attribution estimates that are in good accordance with our results: laying hens (51.1%), pigs (32.5%), turkeys (1.3%) and broilers (0.5%).

Although the ranking of the sources stayed the same for both considered time periods, a decrease in the relative proportion attributed to laying hens (from 54.1% in 2004–2007 to 31.3% in 2010/11) and an increase in the attributed proportion for pigs (from 14.5% in 2004–2007 to 27.2% in 2010/11) could be observed. This shift in the relative importance between the sources could be a direct consequence of different *Salmonella* control programmes in poultry that were initiated by the German federal government between 2007 and 2010. Since less *Salmonella* is found in poultry and especially in laying hens, the importance of pigs as a potential source for human salmonellosis infections becomes more evident.

A limitation of our study is that other sources, including cattle, could not be considered in this approach. Regular monitoring of beef continuously reflects that, although beef might be consumed raw, due to the very low *Salmonella* prevalence, the probability of exposure with Salmonella is quite low (Kaesbohrer et al., 2012). Exposure through direct animal contact, for example with reptiles or other pets, was described as an important source for *Salmonella* infection, but the serotypes involved are different from those where cases had been attributed to livestock species (Pees et al., 2013; Sting, Ackermann, Blazey, Rabsch, & Szabo, 2013). Nonetheless, the omission of cattle and other possible sources like non-animal foods or companion animals makes it likely that the absolute numbers in attributed cases are overestimating the impact of the four sources considered here. Another limitation is that we could not fully remove all outbreak-related cases. The impact on the allocation of cases to sources should be limited as subtype distributions of *Salmonella enterica* ser. Enteritidis and ser. Typhimurium cases were adjusted for this, and other foodborne outbreaks were rarely identified.

A drawback from the method used in the Bayes DB model of setting the parameter values of $q_{ut,j}$ and a_{nut} according to Equations 6 and 7 to constant values is that the credible intervals for the posterior become quite narrow. In running the model with a different combination of parameterizations, we found that the more parameter values we fixed (according to Equations 6 and 7), the smaller the credible intervals became (results not shown). This is not surprising given that fixing a parameter value to a constant value constrains the possible models outcomes more than drawing the parameter values from a prior distribution. Consequently, constrains on possible model outcomes lead to the credible interval becoming smaller. Therefore, one should be cautious in fixing parameter values of unique types and check the effect of the procedure at every step of the analysis in order to prevent that one overestimates the certainty of the model prediction. We consequently interpret the results of our Bayes DB model as probably too optimistic concerning the uncertainty that it displays. But since we are mainly interested in the relative differences in cases attributed to sources for the purposes of a rough ranking we find that the Bayes DB model with parameter fixing is robust enough to come to comparable qualitative results as the Bayes DB model where no parameters are fixed. Thus, our cautious interpretation then says that laying hens or eggs are for both data sets and therefore both periods of time one if not the major source of human Salmonellosis. However, the monitoring data indicate that pigs might have been a comparable important source in the years 2010/2011. Broiler meat is much less important and turkey plays an even smaller role. In both data sets, all models consider the number of cases due to unknown serotype salmonella and therefore not attributable (i.e. unknown) sources as a major source of human salmonellosis. This underscores how much more improvement can be made in the uncertainty of source attribution if only a more complete typing of samples would be achieved.

In conclusion, we found that the David variation of the Hald Model comes to very similar results as the Dutch model concerning the ranking of possible animal food sources of human salmonellosis for the used data. Adding uncertainty to the Dutch Model approach as described here is one way to come to uncertainty estimates of comparable size.

4.4 | Bayes DB model parameter

The parameters q_i and a_j of the Bayes DB model go back to the original Hald model (Hald et al., 2004). As described above, the subtype-specific parameter q_i should model all subtype-dependent factors, for example the ability to survive during food processing or to cause disease in humans. The source-dependent parameter a_j should represent differences between food types like differences in processing and preparation practices, or the capacity to act as a vehicle. Other authors have looked at these parameters for differently structured versions of the Hald model in order to check how robust the parameter values are against a change of model structure (Ravel et al., 2017). They found evidence for the robustness of subtype-specific parameters q_i that seems to confirm the interpretation that these parameters are representing in a reliable fashion subtype-specific properties. Within our study, we do not have different types of Hald models and are therefore more limited in our possibilities to gauge how reliably q_i may represent microbial or epidemiological characteristics of the real world. What can be done with our results is to compare the estimated values for the q_i coming from the BaysDB model with epidemiological data, in this case the number of observed human cases. The idea is to look at the data and compare the number of observed human cases in which a certain *Salmonella* type was found with corresponding q_i values estimated by the Bayes DB model. One would expect that *Salmonella* types involved in large numbers of human cases should also resemble, at least to a certain extent, their subtype-specific ability to cause salmonellosis in humans. Our results point in the direction that there is no clear linear relationship as we will discuss with the following examples. For the baseline data, the largest q_i is q_5 = 0.178 and belongs to Enteritidis with phage-type PT21. This Salmonella is with 16,198 human cases one of the most frequently found types in human cases, but there are phage types as PT 8 with 18,505 and PT 4 with 79,013 observed human cases. This last phage type is the one with the highest number of cases and has a q_9 = 0.037. On the other hand, the smallest q_i in the baseline data was estimated to be q_{56} = 0.0002 for serovar Mbandaka that was found in 90 observed human salmonellosis cases. The type with the smallest number of observed cases (i.e. 6 cases) in the baseline data was serovar Liverpool and Bayes DB estimated the corresponding type-dependent parameter value to be q_{53} = 0.0003. One sees that there is in fact a trend for type-dependent parameter to be larger if the corresponding types caused more cases but the relationship is far from straightforward.

Of course, a straightforward relationship is not to be expected since the number of cases caused by a certain Salmonella type is also moderated by the matrix in which it appears. For example, a highly pathogenic Salmonella type may actually cause a relatively low number of human cases since it only has a low prevalence in a food that is rarely consumed compared to a less pathogenic type, which is highly prevalent in a popular food that is often consumed raw.

4.5 | Which approach should one use?

Based on the two data sets presented here, there seems to be not much of a difference in terms of source attribution estimates between the Dutch and Hald approach. Other studies in part show similar results. For example, David, Guillemot, et al. (2013) found that for French data a model similar to the Dutch model (which David et al. call "simple attribution model") may produce results for source attribution comparable to the ones produced by variations of the Hald model (namely the variations that David et al. called "Reference-type ST," "Specific-Type DB" and "Mullner" in Figure 2 in David, Guillemot, et al., 2013). Something similar can be found in a study that used Italian data (Mughini-Gras et al., 2014). There the comparison of a modified Dutch model with a modified Hald model again finds that at least for using food data and a combination of food and farm data the results of both models are quite comparable (cf. Figure 2 in Mughini-Gras et al., 2014).

Hence, the natural question arises whether it makes a difference to use one or the other approach? There is not one easy answer to this question. If, for example, one wants a rough first source attribution estimate, then the Dutch model can be a viable option. However, one should not rely on the Dutch model as an in-depth analysis of the data. Based on the literature on the comparison of both approaches, one can identify the following points: assuming one has very good data quality, meaning here numerous and unbiased data points that allow for a robust estimation of the prevalence of the different types in the different sources; then, it is reasonable to expect that the Dutch model and the Hald model will show similar results. This means that in this optimal situation for source attribution, one might as well resort just to the easy to implement Dutch model. However, in most practical setting the data will not meet such high standards. This was already mentioned above where (Hald, 2002) suspected somewhat poorer data quality being behind the divergence she observed between the Dutch and the Hald model. This was also discussed in section 5.9 in Mullner et al. (2009). Mullner et al. found that for their New Zeeland data their modified Hald model identified pork as the major source for human salmonellosis. This result was at odds with the rare instances where *Salmonella* is isolated from domestic pork and living animals in New Zeeland. Mullner et al. suspected that since their data on pork were much sparser than for the other sources under consideration and since the few pork isolates contained types that are also found in a large number of human salmonellosis cases this may have led to an overestimation of the importance of the pork source. We add here that the Dutch model under such circumstances is also prone to overestimate the role of pork since the prevalence p_{ij} would also be high and a correspondingly high proportion of cases would be attributed to pork.

We suggest that in most cases data quality will not be excellent, and in this case, it makes sense to use both models and see how they converge and diverge. One should keep in mind the general differences in the two approaches. The Dutch model gives a quick and

direct attribution result that straightforwardly uses the prevalences for a given *Salmonella* type in each source to gauge the importance of each source for cases of human salmonellosis caused by that given *Salmonella* type. But that simple point estimate does not incorporate any uncertainty measure. One might use bootstrapping methods like in Ravel et al. (2017) or a Monte Carlo approach as we did here to introduce uncertainty 'by hand' into the Dutch model. Still, the Dutch model lacks flexibility that would account for varying pathogenicity among the different *Salmonella* types or the effect of the matrix (here eggs, or the meat of broiler, pig or turkey) to harbour the different types. This additional flexibility is provided by the Hald model and its many variants. As a Bayesian approach, it also comes equipped with an in-build mechanism of introducing uncertainty in the shape of likelihood functions and prior distributions. Now, the Bayesian models are much less straightforward in producing 'one' result since there are many possibilities in choosing the prior distributions for parameters or of fixing individual parameters as in this paper or in David, Guillemot, et al. (2013). Depending on the choices thus made the results will change to a larger or lesser extent. We consider this useful since it makes transparent that real-world data always contains some uncertainty and variability. Consequently, one should not expect that real-world data determine exactly one result but one should expect that real-world data can lead to somewhat different results if different models are employed to analyse it. In studying the underlying assumptions of the various model components, one gets a fuller picture of the range of reasonable interpretations of the data.

At the end, we would like to repeat a note from the introduction. This paper compared two approaches considering microbial subtyping source attribution—the Hald model and the Dutch model—and then focusses on some aspects of one of the two approaches, namely the Dutch model. This does not reflect a preference for one of the two approaches—it just reflects the approach taken in this paper to discuss methodological aspects.

ACKNOWLEDGEMENT

The authors would like to thank the colleagues in RKI (W. Rabsch) and BfR (C. Dorn, A. Schroeter, R. Helmuth, A. Friedrich and I. Szabo) for providing the data. This study is a part of the RESET Consortium financially supported by the German Federal Ministry of Education and Research (BMBF) through the German Aerospace Center (01KI1013A-H).

CONFLICT OF INTEREST

The authors declare that they are not aware of any conflict of interest with respect to the work presented in this paper.

ORCID

Guido Correia Carreira https://orcid.org/0000-0002-0012-3735

REFERENCES

Ahlstrom, C., Muellner, P., Spencer, S. E. F., Hong, S., Saupe, A., Rovira, A., … Alvarez, J. (2017). Inferring source attribution from a multiyear multisource data set of *Salmonella* in Minnesota. *Zoonoses Public Health*, *64*(8), 589–598. https://doi.org/10.1111/zph.12351

Barco, L., Barrucci, F., Olsen, J. E., & Ricci, A. (2013). Salmonella source attribution based on microbial subtyping. *International Journal of Food Microbiology*, *163*(2–3), 193–203. https://doi.org/10.1016/j.ijfoodmicro.2013.03.005

Batz, M. B., Doyle, M. P., Morris, G. Jr, Painter, J., Singh, R., Tauxe, R. V., … Food Attribution Working, G. (2005). Attributing illness to food. *Emerging Infectious Diseases*, *11*(7), 993–999. https://doi.org/10.3201/eid1107.040634

BLE (2005). *Statistisches Jahrbuch ueber Ernaehrung, Landwirtschaft und Forsten der Bundesrepublik Deutschland 2005*. Münster-Hiltrup, Germany: Landwirtschaftsverlag Muenster-Hiltrup.

BLE (2006). *Statistisches Jahrbuch ueber Ernaehrung, Landwirtschaft und Forsten der Bundesrepublik Deutschland 2006*. Münster-Hiltrup, Germany: Landwirtschaftsverlag Muenster-Hiltrup.

BLE (2007). *Statistisches Jahrbuch ueber Ernaehrung, Landwirtschaft und Forsten der Bundesrepublik Deutschland 2007*. Münster-Hiltrup, Germany: Landwirtschaftsverlag Muenster-Hiltrup.

BLE (2008). *Statistisches Jahrbuch ueber Ernaehrung, Landwirtschaft und Forsten der Bundesrepublik Deutschland 2008*. Bremerhaven, Germany:Wirtschaftsverlag NW GmbH Bremerhaven.

BLE (2011). *Statistisches Jahrbuch ueber Ernaehrung, Landwirtschaft und Forsten der Bundesrepublik Deutschland 2011*. Münster-Hiltrup, Germany: Landwirtschaftsverlag Muenster-Hiltrup.

Boysen, L., Rosenquist, H., Larsson, J. T., Nielsen, E. M., Sorensen, G., Nordentoft, S., & Hald, T. (2014). Source attribution of human campylobacteriosis in Denmark. *Epidemiology and Infection*, *142*(8), 1599–1608. https://doi.org/10.1017/S0950268813002719

David, J. M., Guillemot, D., Bemrah, N., Thébault, A., Brisabois, A., Chemaly, M., … Watier, L. (2013). How important are the cognitive skills of teenagers in predicting subsequent earnings? *Journal of Policy Analysis and Management*, *19*(4), 547–568. https://doi.org/10.1111/j.1539-6924.2012.01877.x

David, J. M., Sanders, P., Bemrah, N., Granier, S. A., Denis, M., Weill, F.-X., … Watier, L. (2013). Attribution of the French human Salmonellosis cases to the main food-sources according to the type of surveillance data. *Preventive Veterinary Medicine*, *110*(1), 12–27. https://doi.org/10.1016/j.prevetmed.2013.02.002

Domingues, A. R., Pires, S. M., Halasa, T., & Hald, T. (2012a). Source attribution of human campylobacteriosis using a meta-analysis of case-control studies of sporadic infections. *Epidemiology and Infection*, *140*(6), 970–981. https://doi.org/10.1017/S0950268811002676

Domingues, A. R., Pires, S. M., Halasa, T., & Hald, T. (2012b). Source attribution of human salmonellosis using a meta-analysis of case-control studies of sporadic infections. *Epidemiology and Infection*, *140*(6), 959–969. https://doi.org/10.1017/S0950268811002172

EFSA (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*. *EFSA Journal*, *5*(2).

EFSA (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. *EFSA Journal*, *5*(3).

EFSA (2008a). Overview of methods for source attribution for human illness from food-borne microbiological hazards-Scientific Opinion of the Panel on Biological Hazards. *EFSA Journal*, *6*(7), 764.

EFSA (2008b). Report of the Task Force on Zoonoses Data Collection on the analysis of the baseline survey on the prevalence of Salmonella in slaughter pigs, Part A. *The EFSA Journal*, *135*, 1–111.

EFSA (2008c). Report of the Task Force on Zoonoses Data Collection on the Analysis of the baseline survey on the prevalence of Salmonella in turkey flocks, Part A. *The EFSA Journal*, *134*, 1–91.

Evers, E. G., Van Der Fels-Klerx, H. J., Nauta, M. J., Schijven, J. F., & Havelaar, A. H. (2008). Campylobacter source attribution by exposure assessment. *International Journal of Risk Assessment and Management*, *8*(1), 174–190. https://doi.org/10.1504/IJRAM.2008.016151

Glass, K., Fearnley, E., Hocking, H., Raupach, J., Veitch, M., Ford, L., & Kirk, M. D. (2016). Bayesian source attribution of salmonellosis in South Australia. *Risk Analysis*, *36*(3), 561–570. https://doi.org/10.1111/risa.12444

Greig, J. D., & Ravel, A. (2009). Analysis of foodborne outbreak data reported internationally for source attribution. *International Journal of Food Microbiology*, *130*(2), 77–87. https://doi.org/10.1016/j.ijfoodmicro.2008.12.031

Guo, C., Hoekstra, R. M., Schroeder, C. M., Pires, S. M., Ong, K. L., Hartnett, E., … Cole, D. (2011). Application of Bayesian techniques to model the burden of human salmonellosis attributable to U.S. Food Commodities at the Point of Processing: Adaptation of a Danish model. *Foodborne Pathogens and Disease*, *8*(4), 509–516. https://doi.org/10.1089/fpd.2010.0714

Hald, T. (2002). *Quantifying sources of human salmonellosis in the Netherlands and in Denmark: Two approaches*, *same results?* Retrieved from

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 Analysis*, *24*(1), 255–269. https://doi.org/10.1111/j.0272-4332.2004.00427.x

Hald, T., Wong, D. M. A. L., & Aarestrup, F. M. (2007). The attribution of human infections with antimicrobial resistant Salmonella bacteria in Denmark to sources of animal origin. *Foodborne Pathogens and Disease*, *4*(3), 313–326. https://doi.org/10.1089/fpd.2007.0002

Havelaar, A. H., Galindo, A. V., Kurowicka, D., & Cooke, R. M. (2008). Attribution of foodborne pathogens using structured expert elicitation. *Foodborne Pathogens and Disease*, *5*(5), 649–659. https://doi.org/10.1089/fpd.2008.0115

Hoffmann, S., Devleesschauwer, B., Aspinall, W., Cooke, R., Corrigan, T., Havelaar, A., … Hald, T. (2017). Attribution of global foodborne disease to specific foods: Findings from a World Health Organization structured expert elicitation. *PLoS ONE*, *12*(9), e0183641. https://doi.org/10.1371/journal.pone.0183641

Hoffmann, S., Fischbeck, P., Krupnick, A., & McWilliams, M. (2007). Using expert elicitation to link foodborne illnesses in the United States to foods. *Journal of Food Protection*, *70*(5), 1220–1229. https://doi.org/10.4315/0362-028x-70.5.1220

Kaesbohrer, A., Tenhagen, B.-A., Alt, K., Schroeter, A., Szabo, I., & Hartung, M. (2012). BfR Wissenschaft: Erreger von Zoonosen in Deutschland im Jahr 2010. In M. Hartung & A. Kaesbohrer (Eds.), *BfR Wissenschaft: Erreger von Zoonosen in Deutschland im Jahr 2010* (pp. 29–132). Berlin, Germany: BfR Wissenschaft.

Kaesbohrer, A., Tenhagen, B.-A., Alt, K., Schroeter, A., Szabo, I., & Hartung, M. (2013). BfR Wissenschaft: Erreger von Zoonosen in Deutschland im Jahr 2011. In M. Hartung & A. Kaesbohrer (Eds.), *BfR Wissenschaft: Erreger von Zoonosen in Deutschland im Jahr 2011* (pp. 35–144). Berlin, Germany: BfR Wissenschaft.

Kosmider, R. D., Nally, P., Simons, R. R., Brouwer, A., Cheung, S., Snary, E. L., & Wooldridge, M. (2010). Attribution of human VTEC O157 infection from meat products: A quantitative risk assessment approach. *Risk Analysis*, *30*(5), 753–765. https://doi.org/10.1111/j.1539-6924.2009.01317.x

Little, C. L., Pires, S. M., Gillespie, I. A., Grant, K., & Nichols, G. L. (2010). Attribution of human Listeria monocytogenes infections in England and Wales to ready-to-eat food sources placed on the market: Adaptation of the Hald Salmonella source attribution model. *Foodborne Pathogens and Disease*, *7*(7), 749–756. https://doi.org/10.1089/fpd.2009.0439

Lunn, D., Spiegelhalter, D., Thomas, A., & Best, N. (2009). The BUGS project: Evolution, critique and future directions. *Statistics in Medicine*, *28*(25), 3049–3067. https://doi.org/10.1002/sim.3680

Mangen, M. J., Batz, M. B., Kasbohrer, A., Hald, T., Morris, J. G. Jr, Taylor, M., & Havelaar, A. H. (2010). Integrated approaches for the public health prioritization of foodborne and zoonotic pathogens. *Risk Analysis*, *30*(5), 782–797. https://doi.org/10.1111/j.1539-6924.2009.01291.x

Mather, A. E., Vaughan, T. G., & French, N. P. (2015). Molecular approaches to understanding transmission and source attribution in nontyphoidal Salmonella and their application in Africa. *Clinical Infectious Diseases*, *61*, S259–S265. https://doi.org/10.1093/cid/civ727

Mikkela, A., Ranta, J., & Tuominen, P. (2019). A Modular Bayesian Salmonella source attribution model for sparse data. *Risk Analysis*, *39*(8), 1796–1811. https://doi.org/10.1111/risa.13310

Miller, P., Marshall, J., French, N., & Jewell, C. (2017). sourceR: Classification and source attribution of infectious agents among heterogeneous populations. *Plos Computational Biology*, *13*(5), e1005564. https://doi.org/10.1371/journal.pcbi.1005564

Mughini-Gras, L., Barrucci, F., Smid, J. H., Graziani, C., Luzzi, I., Ricci, A., … Busani, L. (2014). Attribution of human Salmonella infections to animal and food sources in Italy (2002–2010): Adaptations of the Dutch and modified Hald source attribution models. *Epidemiology and Infection*, *142*(5), 1070–1082. https://doi.org/10.1017/S0950268813001829

Mughini-Gras, L., Franz, E., & van Pelt, W. (2018). New paradigms for Salmonella source attribution based on microbial subtyping. *Food Microbiology*, *71*, 60–67. https://doi.org/10.1016/j.fm.2017.03.002

Mughini-Gras, L., & van Pelt, W. (2014). Salmonella source attribution based on microbial subtyping: Does including data on food consumption matter? *International Journal of Food Microbiology*, *191*, 109–115. https://doi.org/10.1016/j.ijfoodmicro.2014.09.010

Mughini-Gras, L., van Pelt, W., van der Voort, M., Heck, M., Friesema, I., & Franz, E. (2017). Attribution of human infections with Shiga toxin-producing *Escherichia coli* (STEC) to livestock sources and identification of source-specific risk factors, The Netherlands (2010–2014). *Zoonoses Public Health*, *65*, e8–e22. https://doi.org/10.1111/zph.12403

Mullner, P., Jones, G., Noble, A., Spencer, S. E., Hathaway, S., & French, N. P. (2009). Source attribution of food-borne zoonoses in New Zealand: A modified Hald model. *Risk Analysis*, *29*(7), 970–984. https://doi.org/10.1111/j.1539-6924.2009.01224.x

Pees, M., Rabsch, W., Plenz, B., Fruth, A., Prager, R., Simon, S., … Braun, P.. (2013). Evidence for the transmission of Salmonella from reptiles to children in Germany, July 2010 to October 2011. *Eurosurveillance*, *18*(46), 1–10.

Pintar, K. D. M., Thomas, K. M., Christidis, T., Otten, A., Nesbitt, A., Marshall, B., … Ravel, A. (2017). A comparative exposure assessment of campylobacter in Ontario, Canada. *Risk Analysis*, *37*(4), 677–715. https://doi.org/10.1111/risa.12653

Pires, S. M., de Knegt, L., & Hald, T. (2011). *Estimation of the relative contribution of different food and animal sources to human Salmonella infections in the European Union*. Retrieved from https://orbit.dtu.dk/files/52965188/jul11%20SA%20biomon.pdf

Pires, S. M., Evers, E. G., van Pelt, W., Ayers, T., Scallan, E., Angulo, F. J., … Med-Vet-Net Workpackage 28 Working, G. (2009). Attributing the human disease burden of foodborne infections to specific sources. *Foodborne Pathogens and Disease*, *6*(4), 417–424. https://doi.org/10.1089/fpd.2008.0208

Pires, S. M., & Hald, T. (2010). Assessing the differences in public health impact of Salmonella subtypes using a Bayesian microbial subtyping approach for source attribution. *Foodborne Pathogens and Disease*, *7*(2), 143–151. https://doi.org/10.1089/fpd.2009.0369

Pires, S. M., Vigre, H., Makela, P., & Hald, T. (2010). Using outbreak data for source attribution of human salmonellosis and campylobacteriosis in Europe. *Foodborne Pathogens and Disease*, *7*(11), 1351–1361. https://doi.org/10.1089/fpd.2010.0564

R Core Team (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Ranta, J., Matjushin, D., Virtanen, T., Kuusi, M., Viljugrein, H., Hofshagen, M., & Hakkinen, M. (2011). Bayesian temporal source attribution of foodborne zoonoses: Campylobacter in Finland and Norway. *Risk Analysis*, *31*(7), 1156–1171. https://doi.org/10.1111/j.1539-6924.2010.01558.x

Ravel, A., Hurst, M., Petrica, N., David, J., Mutschall, S. K., Pintar, K., … Pollari, F. (2017). Source attribution of human campylobacteriosis at the point of exposure by combining comparative exposure assessment and subtype comparison based on comparative genomic fingerprinting. *PLoS ONE*, *12*(8), e0183790. https://doi.org/10.1371/journal.pone.0183790

RKI (2012). *SURVSTAT@RKI*. Retrieved from http://www3.rki.de/SurvStat

Sting, R., Ackermann, D., Blazey, B., Rabsch, W., & Szabo, I. (2013). Salmonella infections in reptiles–prevalence, serovar spectrum and impact on animal health. *Berliner und Munchener Tierarztliche Wochenschrift*, *126*(5–6), 202–208.

Toft, N., Innocent, G. T., Gettinby, G., & Reid, S. W. (2007). Assessing the convergence of Markov Chain Monte Carlo methods: An example from evaluation of diagnostic tests in absence of a gold standard. *Preventive Veterinary Medicine*, *79*(2–4), 244–256. https://doi.org/10.1016/j.prevetmed.2007.01.003

Van Pelt, W., Van De Giessen, A., Van Leeuwen, W., Wannet, W., Henken, A., Evers, E., … Van Duynhoven, Y. (1999). Oorsprong, omvang en kosten van humane salmonellose. Deel 1. Oorsprong van humane salmonellose met betrekking tot varken, rund, kip, ei en overige bronnen. *Infectieziekten Bulletin*, *10*(12), 240–243.

Wahlstrom, H., Andersson, Y., Plym-Forshell, L., & Pires, S. M. (2011). Source attribution of human Salmonella cases in Sweden. *Epidemiology and Infection*, *139*(8), 1246–1253. https://doi.org/10.1017/S0950268810002293

How to cite this article: Jabin H, Correia Carreira G, Valentin L, Käsbohrer A. The role of parameterization in comparing source attribution models based on microbial subtyping for salmonellosis. *Zoonoses Public Health*. 2019;66:943–960. https://doi.org/10.1111/zph.12645