Food Modelling Journal 2: e70008
doi: 10.3897/fmj.2.70008

FSKX Food Safety Knowledge

An FSKX compliant source attribution model for salmonellosis and a look at its major hidden pitfalls

Esther M. Sundermann‡, Guido Correia Carreira‡, Annemarie Käsbohrer‡

‡ German Federal Institute for Risk Assessment, Berlin, Germany

Corresponding author: Esther M. Sundermann (esther-maria.sundermann@bfr.bund.de), Guido Correia Carreira (guido.correia-carreira@bfr.bund.de), Annemarie Käsbohrer (annemarie.kaesbohrer@bfr.bund.de)

Academic editor: Matthias Filter

Received: 11 Jun 2021 | Accepted: 31 Aug 2021 | Published: 03 Nov 2021

Citation: Sundermann EM, Correia Carreira G, Käsbohrer A (2021) An FSKX compliant source attribution model for salmonellosis and a look at its major hidden pitfalls. Food Modelling Journal 2: e70008. https://doi.org/10.3897/fmj.2.70008

Abstract

To reduce the burden of human society that is caused by zoonotic diseases, it is important to attribute sources to human illnesses. One powerful approach in supporting any intervention decision is mathematical modelling. This paper presents a source attribution model which considers five sources (broilers, laying hens, pigs, turkeys) for salmonellosis and uses two datasets from Germany collected over two time periods; one from 2004 to 2007 and one from 2010 to 2011. The model uses a Bayesian modelling approach derived from the so-called Hald model and is based on microbial subtyping. In this case, *Salmonella* isolates from humans and animals were subtyped with respect to serovar and phage type. Based on that typing, the model estimates how many human salmonellosis cases can be attributed to each of the considered sources. A reference description of the model is available under DOI: 10.1111/zph.12645. Here, we present this model as a ready-to-use resource in the Food Safety Knowledge Exchange (FSKX) format. This open information exchange format allows to re-use, modify, and further develop the model and uses model metadata and controlled vocabulary to harmonise the annotation. In addition to the model, we discuss some technical pitfalls that might occur when running this Bayesian model based on Markov chain Monte Carlo calculations. As source attribution of zoonotic

disease is one useful tool for the One Health approach, our work facilitates the exchange, adjustment, and re-usage of this source attribution model by the international and multi-sectoral community.

Keywords

Salmonella, R programming language, mathematical modelling, Bayesian model, Markov chain Monte Carlo method, Food Safety Knowledge Exchange (FSKX) format

Introduction

Zoonotic diseases are a major burden for human society. The burden relates to two categories: 1) human health burden in form of mortality and morbidity (Taylor et al. 2001) and 2) economical burden, e.g., in form of losses due to health care costs (The World Bank 2012). In the European Union, over 320,000 human cases of zoonotic disease were reported in 2019 (European Food Safety Authority and European Centre for Disease Prevention and Control 2021).

Salmonella is the second most common zoonotic disease in Europe with a stable number of salmonellosis cases during 2014–2018 (European Food Safety Authority and European Centre for Disease Prevention and Control 2021). Although the salmonellosis burden is stagnating, the contribution of *Salmonella* serovars differs in prevalence and the source of human infection (European Food Safety Authority and European Centre for Disease Prevention and Control 2021, Jabin et al. 2019).

To reduce the human cases of zoonoses, it is important to understand the relationship of potential sources and human illness (Batz et al. 2005). In order to reduce consumer exposure and to optimize intervention measures, it is required to identify the different zoonotic sources of human infections and quantify their relative contribution (Batz et al. 2005, Pires et al. 2009). Both can be supported by source attribution methods. A powerful method is mathematical modelling. One approach for attributing foodborne illnesses is based on microbial subtyping. This approach includes various methods to distinguish bacterial and viral isolates from one another (Pires et al. 2009). A widely used microbial subtyping approach combines serotyping with phage typing that is based on phenotypic methods. In recent years further subtyping approaches based on molecular methods like plasmid analysis or whole-genome sequencing have been used (Boysen et al. 2014, Mather et al. 2015,Munck et al. 2020,Arnold et al. 2021). Whichever method is used, the data which describe the distribution of different subtypes in different sources can be used to do source attribution based on mathematical modelling.

One modelling approach for source attribution that is based on microbial subtyping is the Bayesian model. In the context of food safety, the models developed by Hald et al. (2004) and David et al. (2013) are of special interest. Jabin et al. (2019) used the approach of David et al. (2013) to attribute human salmonellosis to potential food sources in Germany

using two datasets from 2004–2007 and 2010/2011. Although the mathematical model has been published and described in detail in Jabin et al. (2019), it is not available in a ready-to-use format. Here, we present the Bayesian model referred as Bayes data-based model in Jabin et al. 2019 in the Food Safety Knowledge Exchange (FSKX) format. This open information exchange format uses model metadata and controlled vocabulary to harmonize annotations of risk assessment models (Haberbeck et al. 2018). Together with the model script, the visualization script, and simulation settings, the metadata are the key components of the format (de Alba Aparicio et al. 2018). Thus, FSKX format facilitates the model usage and re-usage.

The two datasets and the mathematical model by Jabin et al. (2019) are incorporated into a ready-to-use source attribution model which can be executed, developed further, and easily adapted to new data by the international risk assessment community. With our work, we facilitate the exchange, adjustment, and re-usage of this source attribution model.

Model metadata

The model metadata are part of the FSKX-file (see Suppl. material 1 for the FSKX-file). For details about the metadata schema and the used definitions see Haberbeck et al. (2018) and available on https://foodrisklabs.bfr.bund.de/rakip-harmonization-resources/ (we use the metadata schema Version 1.04).

General metadata

Source: PUBLISHED SCIENTIFIC STUDIES

Identifier: SourceAttributionBfRBayesDB

Rights: Creative Commons Attribution 4.0 (CC BY 4.0)

Availability: Open access

Language: English

Software: FSK-Lab 1.9.0

Language Written In: R 3

Objective: The model attributes human cases of the salmonellosis caused by various serovars of *Salmonella* from various sources (namely, broilers, laying hens, pigs, turkeys, and unknown). The model is parameterized using data from Germany.

Product/matrix

Name: Broilers

Description: Tons of broiler meat consumed

Unit: Tons

Origin Country: Germany

Name: Laying hens

Description: Number of eggs consumed

Unit: Number of eggs

Origin Country: Germany

Name: Pigs

Description: Tons of pork consumed

Unit: Tons

Origin Country: Germany

Name: Turkeys

Description: Tons of turkey meat consumed

Unit: Tons

Origin Country: Germany

Hazard

- **Type**: Microorganisms; **Name**: *Salmonella* Enteritidis; **Description**: The model considers the prevalence among isolates from animal samples (pig, broiler, laying hen, turkey) which are positive for *Salmonella enterica* subsp. *enterica* ser. Enteritidis (*Salmonella* Enteritidis or *S*.E. for short) which were further typed according to their phage type (Phage types: *S*.E. PT 1, *S*.E. PT 11, *S*.E. PT 14b, *S*. E. PT 19, *S*.E. PT 2, *S*.E. PT 21, *S*.E. PT 21c, *S*.E. PT 25, *S*.E. PT 35, *S*.E. PT 4, *S*.E. PT 4a, *S*.E. PT 4b, *S*.E. PT 5a, *S*.E. PT 6, *S*.E. PT 6a, *S*.E. PT 7, *S*.E. PT 7a, *S*.E. PT 8, other); **Unit**: %; **Adverse Effect**: The most common symptoms of salmonellosis are diarrhea, fever, abdominal cramps, and vomiting.
- **Type**: Microorganisms; **Name**: *Salmonella* Typhimurium; **Description**: The model considers the prevalence among isolates from animal samples (pig, broiler, laying hen, turkey) which are positive for *Salmonella enterica subsp. enterica* ser. Typhimurium (*Salmonella* Typhimurium or *S*.T. for short) which were further typed according to their phage type (Phage types: *S*.T. DT001, *S*.T. DT007, *S*.T. DT008, *S*.T. DT009, *S*.T. DT012, *S*.T. DT017, *S*.T. DT040, *S*.T. DT041, *S*.T. D066, *S*.T. DT099, *S*.T. DT104, *S*.T. DT120, *S*.T. DT126, *S*.T.. DT139, *S*.T. DT195, *S*.T. DT208, *S*.T. U302, *S*.T. U310, other); **Unit**: %; **Adverse Effect**: The most common symptoms of salmonellosis are diarrhea, fever, abdominal cramps, and vomiting.

- **Type**: Microorganisms; **Name**: *Salmonella* Agama; **Description**: The model considers the prevalence among isolates from animal samples (pig, broiler, laying hen, turkey) which are positive for *Salmonella enterica* subsp. *enterica* ser. Agama (*Salmonella* Agama for short); **Unit**: %; **Adverse Effect**: The most common symptoms of salmonellosis are diarrhea, fever, abdominal cramps, and vomiting.
- **Type**: Microorganisms; **Name**: *Salmonella* Agona; **Description**: The model considers the prevalence among isolates from animal samples (pig, broiler, laying hen, turkey) which are positive for *Salmonella enterica* subsp. *enterica* ser. Agona (*Salmonella* Agona for short) ; **Unit**: %; **Adverse Effect**: The most common symptoms of salmonellosis are diarrhea, fever, abdominal cramps, and vomiting.
- **Type**: Microorganisms; **Name**: *Salmonella* Anatum; **Description**: The model considers the prevalence among isolates from animal samples (pig, broiler, laying hen, turkey) which are positive for *Salmonella enterica* subsp. *enterica* ser. Anatum (*Salmonella* Anatum for short) ; **Unit**: %; **Adverse Effect**: The most common symptoms of salmonellosis are diarrhea, fever, abdominal cramps, and vomiting.
- **Type**: Microorganisms; **Name**: *Salmonella* Blockley; **Description**: The model considers the prevalence among isolates from animal samples (pig, broiler, laying hen, turkey) which are positive for *Salmonella enterica* subsp. *enterica* ser. Blockley (*Salmonella* Blockley for short) ; **Unit**: %; **Adverse Effect**: The most common symptoms of salmonellosis are diarrhea, fever, abdominal cramps, and vomiting.
- **Type**: Microorganisms; **Name**: *Salmonella* Braenderup; **Description**: The model considers the prevalence among isolates from animal samples (pig, broiler, laying hen, turkey) which are positive for *Salmonella enterica* subsp. *enterica* ser. Braenderup (*Salmonella* Braenderup for short) ; **Unit**: %; **Adverse Effect**: The most common symptoms of salmonellosis are diarrhea, fever, abdominal cramps, and vomiting.
- **Type**: Microorganisms; **Name**: *Salmonella* Brandenburg; **Description**: The model considers the prevalence among isolates from animal samples (pig, broiler, laying hen, turkey) which are positive for *Salmonella enterica* subsp. *enterica* ser. Brandenburg (*Salmonella* Brandenburg for short) ; **Unit**: %; **Adverse Effect**: The most common symptoms of salmonellosis are diarrhea, fever, abdominal cramps, and vomiting.
- **Type**: Microorganisms; **Name**: *Salmonella* Coeln; **Description**: The model considers the prevalence among isolates from animal samples (pig, broiler, laying hen, turkey) which are positive for *Salmonella enterica* subsp. *enterica* ser. Coeln (*Salmonella* Coeln for short); **Unit**: %; **Adverse Effect**: The most common symptoms of salmonellosis are diarrhea, fever, abdominal cramps, and vomiting.
- **Type**: Microorganisms; **Name**: *Salmonella* Derby; **Description**: The model considers the prevalence among isolates from animal samples (pig, broiler, laying hen, turkey) which are positive for *Salmonella enterica* subsp. *enterica* ser. Derby (*Salmonella* Derby for short); **Unit**: %; **Adverse Effect**: The most common symptoms of salmonellosis are diarrhea, fever, abdominal cramps, and vomiting.
- **type**: Microorganisms; **Name**: *Salmonella* Eboko; **Description**: The model considers the prevalence among isolates from animal samples (pig, broiler, laying hen, turkey) which are positive for *Salmonella enterica* subsp. *enterica* ser. Eboko

(*Salmonella* Eboko for short); **unit**: %; **Adverse Effect**: The most common symptoms of salmonellosis are diarrhea, fever, abdominal cramps, and vomiting.

- **Type**: Microorganisms; **Name**: *Salmonella* Give; **Description**: The model considers the prevalence among isolates from animal samples (pig, broiler, laying hen, turkey) which are positive for *Salmonella enterica* subsp. *enterica* ser. Give (*Salmonella* Give for short); **Unit**: %; **Adverse Effect**: The most common symptoms of salmonellosis are diarrhea, fever, abdominal cramps, and vomiting.
- **Type**: Microorganisms; **Name**: *Salmonella* Heidelberg; **Description**: The model considers the prevalence among isolates from animal samples (pig, broiler, laying hen, turkey) which are positive for *Salmonella enterica* subsp. *enterica* ser. Heidelberg (*Salmonella* Heidelberg for short); **Unit**: %; **Adverse Effect**: The most common symptoms of salmonellosis are diarrhea, fever, abdominal cramps, and vomiting.
- **Type**: Microorganisms; **Name**: *Salmonella* Hessarek; **Description**: The model considers the prevalence among isolates from animal samples (pig, broiler, laying hen, turkey) which are positive for *Salmonella enterica* subsp. *enterica* ser. Hessarek (*Salmonella* Hessarek for short); **Unit**: %; **Adverse Effect**: The most common symptoms of salmonellosis are diarrhea, fever, abdominal cramps, and vomiting.
- **Type**: Microorganisms; **Name**: *Salmonella* Indiana; **Description**: The model considers the prevalence among isolates from animal samples (pig, broiler, laying hen, turkey) which are positive for *Salmonella enterica* subsp. *enterica* ser. Indiana (*Salmonella* Indiana for short); **Unit**: %; **Adverse Effect**: The most common symptoms of salmonellosis are diarrhea, fever, abdominal cramps, and vomiting.
- **Type**: Microorganisms; **Name**: *Salmonella* Infantis; **Description**: The model considers the prevalence among isolates from animal samples (pig, broiler, laying hen, turkey) which are positive for *Salmonella enterica* subsp. *enterica* ser. Infantis (*Salmonella* Infantis for short); **Unit**: %; **Adverse Effect**: The most common symptoms of salmonellosis are diarrhea, fever, abdominal cramps, and vomiting.
- **Type**: Microorganisms; **Name**: *Salmonella* Kedougou; **Description**: The model considers the prevalence among isolates from animal samples (pig, broiler, laying hen, turkey) which are positive for *Salmonella enterica* subsp. *enterica* ser. Kedougou (*Salmonella* Kedougou for short); **Unit**: %; **Adverse Effect**: The most common symptoms of salmonellosis are diarrhea, fever, abdominal cramps, and vomiting.
- **Type**: Microorganisms; **Name**: *Salmonella* Kottbus; **Description**: The model considers the prevalence among isolates from animal samples (pig, broiler, laying hen, turkey) which are positive for *Salmonella enterica* subsp. *enterica* ser. Kottbus (*Salmonella* Kottbus for short); **Unit**: %; **Adverse Effect**: The most common symptoms of salmonellosis are diarrhea, fever, abdominal cramps, and vomiting.
- **Type**: Microorganisms; **Name**: *Salmonella* Lexington; **Description**: The model considers the prevalence among isolates from animal samples (pig, broiler, laying hen, turkey) which are positive for *Salmonella enterica* subsp. *enterica* ser. Lexington (*Salmonella* Lexington for short); **Unit**: %; **Adverse Effect**: The most

common symptoms of salmonellosis are diarrhea, fever, abdominal cramps, and vomiting.

- **Type**: Microorganisms; **Name**: *Salmonella* Liverpool; **Description**: The model considers the prevalence among isolates from animal samples (pig, broiler, laying hen, turkey) which are positive for *Salmonella enterica* subsp. *enterica* ser. Liverpool (*Salmonella* Liverpool for short); **Unit**: %; **Adverse Effect**: The most common symptoms of salmonellosis are diarrhea, fever, abdominal cramps, and vomiting.
- **Type**: Microorganisms; **Name**: *Salmonella* Livingstone; **Description**: The model considers the prevalence among isolates from animal samples (pig, broiler, laying hen, turkey) which are positive for *Salmonella enterica* subsp. *enterica* ser. Livingstone (*Salmonella* Livingstone for short); **Unit**: %; **Adverse Effect**: The most common symptoms of salmonellosis are diarrhea, fever, abdominal cramps, and vomiting.
- **Type**: Microorganisms; **Name**: *Salmonella* London; **Description**: The model considers the prevalence among isolates from animal samples (pig, broiler, laying hen, turkey) which are positive for *Salmonella enterica* subsp. *enterica* ser. London (*Salmonella* London for short); **Unit**: %; **Adverse Effect**: The most common symptoms of salmonellosis are diarrhea, fever, abdominal cramps, and vomiting.
- **Type**: Microorganisms; **Name**: *Salmonella* Mbandaka; **Description**: The model considers the prevalence among isolates from animal samples (pig, broiler, laying hen, turkey) which are positive for *Salmonella enterica* subsp. *enterica* ser. Mbandaka (*Salmonella* Mbandaka for short); **Unit**: %; **Adverse Effect**: The most common symptoms of salmonellosis are diarrhea, fever, abdominal cramps, and vomiting.
- **Type**: Microorganisms; **Name**: *Salmonella* Montevideo; **Description**: The model considers the prevalence among isolates from animal samples (pig, broiler, laying hen, turkey) which are positive for *Salmonella enterica* subsp. *enterica* ser. Montevideo (*Salmonella* Montevideo for short); **Unit**: %; **Adverse Effect**: The most common symptoms of salmonellosis are diarrhea, fever, abdominal cramps, and vomiting.
- **Type**: Microorganisms; **Name**: *Salmonella* Newport; **Description**: The model considers the prevalence among isolates from animal samples (pig, broiler, laying hen, turkey) which are positive for *Salmonella enterica* subsp. *enterica* ser. Newport (*Salmonella* Newport for short); **Unit**: %; **Adverse Effect**: The most common symptoms of salmonellosis are diarrhea, fever, abdominal cramps, and vomiting.
- **Type**: Microorganisms; **Name**: *Salmonella* Ohio; **Description**: The model considers the prevalence among isolates from animal samples (pig, broiler, laying hen, turkey) which are positive for *Salmonella enterica* subsp. *enterica* ser. Ohio (*Salmonella* Ohio for short); **Unit**: %; **Adverse Effect**: The most common symptoms of salmonellosis are diarrhea, fever, abdominal cramps, and vomiting.
- **Type**: Microorganisms; **Name**: *Salmonella* Rissen; **Description**: The model considers the prevalence among isolates from animal samples (pig, broiler, laying hen, turkey) which are positive for *Salmonella enterica* subsp. *enterica* ser. Rissen

(*Salmonella* Rissen for short); **Unit**: %; **Adverse Effect**: The most common symptoms of salmonellosis are diarrhea, fever, abdominal cramps, and vomiting.

- **Type**: Microorganisms; **Name**: *Salmonella* Saintpaul; **Description**: The model considers the prevalence among isolates from animal samples (pig, broiler, laying hen, turkey) which are positive for *Salmonella enterica* subsp. *enterica* ser. Saintpaul (*Salmonella* Saintpaul for short); **Unit**: %; **Adverse Effect**: The most common symptoms of salmonellosis are diarrhea, fever, abdominal cramps, and vomiting.
- **Type**: Microorganisms; **Name**: *Salmonella* Senftenberg; **Description**: The model considers the prevalence among isolates from animal samples (pig, broiler, laying hen, turkey) which are positive for *Salmonella enterica* subsp. *enterica* ser. Senftenberg (*Salmonella* Senftenberg for short); **Unit**: %; **Adverse Effect**: The most common symptoms of salmonellosis are diarrhea, fever, abdominal cramps, and vomiting.
- **Type**: Microorganisms; **Name**: *Salmonella* Stanley; **Description**: The model considers the prevalence among isolates from animal samples (pig, broiler, laying hen, turkey) which are positive for *Salmonella enterica* subsp. *enterica* ser. Stanley (*Salmonella* Stanley for short); **Unit**: %; **Adverse Effect**: The most common symptoms of salmonellosis are diarrhea, fever, abdominal cramps, and vomiting.
- **Type**: Microorganisms; **Name**: *Salmonella* Tennessee; **Description**: The model considers the prevalence among isolates from animal samples (pig, broiler, laying hen, turkey) which are positive for *Salmonella enterica* subsp. *enterica* ser. Tennessee (*Salmonella* Tennessee for short); **Unit**: %; **Adverse Effect**: The most common symptoms of salmonellosis are diarrhea, fever, abdominal cramps, and vomiting.
- **Type**: Microorganisms; **Name**: *Salmonella* Virchow; **Description**: The model considers the prevalence among isolates from animal samples (pig, broiler, laying hen, turkey) which are positive for *Salmonella enterica* subsp. *enterica* ser. Virchow (*Salmonella* Virchow for short); **Unit**: %; **Adverse Effect**: The most common symptoms of salmonellosis are diarrhea, fever, abdominal cramps, and vomiting.
- **Type**: Microorganisms; **Name**: rough *Salmonella*; **Description**: The model considers the prevalence among isolates from animal samples (pig, broiler, laying hen, turkey) which contain rough strains of *Salmonella enterica* subsp. *enterica* with unspecified serovar (rough *Salmonella* for short); **Unit**: %; **Adverse Effect**: The most common symptoms of salmonellosis are diarrhea, fever, abdominal cramps, and vomiting.
- **Type**: Microorganisms; **Name**: *Salmonella - other serotypes*; **Description**: The model considers the prevalence among isolates from animal samples (pig, broiler, laying hen, turkey) which are positive for *Salmonella enterica* subsp. *enterica* with unspecified serotype (*Salmonella* - other serotypes for short); **Unit**: %; **Adverse Effect**: The most common symptoms of salmonellosis are diarrhea, fever, abdominal cramps, and vomiting.

Population

Name: People in Germany

Target Population: People in Germany that were identified by medical professionals to be suffering from salmonellosis

Scope

The model attributes human cases of the zoonotic disease salmonellosis to a certain source (namely, broilers, laying hens, pigs, turkeys, and unknown). It is based on a Bayesian microbial subtyping approach described by Hald et al. (2004) and subsequently modified by David et al. (2013). Data from two datasets using information from various years are used to parameterize the model. Both datasets are from Germany.

Temporal Information: In Jabin et al. 2019, data in Table 2 in Jabin et al. (2019) consider human salmonellosis in the years 2004-2007 and data in Table 3 in Jabin et al. (2019) consider the years 2010/2011. See Table 4 in Jabin et al. (2019) for details about the data sources for *Salmonella* in humans.

Data background

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

Study Description: Two datasets from active monitoring in Germany were available. The data comprise four potential animal sources: broilers, laying hens, pigs, and turkeys. For each considered salmonellosis case, the serotype was determined. For cases caused by *Salmonella* Enteritidis or *Salmonella* Typhimurium additionally the phage type was determined (see Tables 2 and 3 in Jabin et al. 2019). Both datasets cover multiple years (for details see "Temporal Information" in the Subsection "Scope" of the Section "Model metadata" and Jabin et al. 2019). The data are used to analyze the source attribution model see Section "Material and methods" and Jabin et al. (2019) for details.

Material and methods

Data

Datasets covering studies on *Salmonella* in different sources for two time periods were compiled and used for this analysis. For both time periods, reliable data from active monitoring on four potential animal sources were available: broilers, laying hens, pigs, and turkeys. Cattle were not included in any study or program and were therefore not included in this analysis. The datasets, which cover the years 2004–2007 and 2010/2011, are called baseline data and monitoring data, respectively. In addition, data on human salmonellosis cases were considered.

Baseline data

The first dataset on *Salmonella* in sources was generated by four baseline studies conducted during 2004 and 2007 in Germany (EFSA (2007b), EFSA (2007a), EFSA (2008b), EFSA (2008a)). These data describe the prevalence p_{ij} (in %) of each *Salmonella* subtype *i* in each source *j*. The prevalence for all strains of *Salmonella* serotypes Enteritidis and Typhimurium as well as the relevant phage types are considered (see Jabin et al. (2019) for details).

Monitoring data

The second dataset on *Salmonella* in sources was compiled from monitoring programs during 2010 and 2011 in Germany (Käsbohrer et al. 2012, Käsbohrer et al. 2013). Data for broilers, laying hens, and turkeys were obtained from the *Salmonella* control programs in poultry in 2010. Since national monitoring of *Salmonella* prevalence in pigs was conducted only in 2011, the data on pigs from 2011 were combined with the poultry data from 2010, assuming that the serotype distributions in the animals were equal in both years (see Jabin et al. (2019) for details).

To summarize, the baseline and the monitoring data are comparable, i.e., the data were compiled in a similar way and the intention measures in the years were the same, thus, no significant difference in the data is expected.

Human data

Data on human *Salmonella* cases came from the Robert Koch Institute (RKI). The serotype distribution was obtained via their online database SurvStat@RKI (https://survstat.rki.de/, data access: 07.02.2012). In addition, phage type information for S. Enteritidis and S. Typhimurium strains were provided via personal communication by Wolfgang Rabsch (RKI). Since only a subset of all strains isolated from humans had been phage typed, we assumed that the phage type distribution among the typed strains 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 sero- and phage types associated with human salmonellosis cases over that time period (see Jabin et al. (2019) for details).

Mathematical model

The presented Bayes data-based (DB) model is a source attribution model that is based on microbial subtyping (Jabin et al. 2019). It is derived from the model developed by Hald et al. 2004, the so-called Hald model. This model has been adopted widely (David et al. 2013, Pires et al. 2011, Ranta et al. 2011). The variation of the Hald model developed by David et al. 2013 reparameterizes the Hald model which leads to an improved robustness of source attribution estimates.

A note about terminology: the terms "subtype" and "type" are used interchangeably.

Hald model

The so-called Hald model (Hald et al. 2004) is a Bayesian modelling approach that uses microbial subtyping data to infer the sources for observed food-borne illnesses like salmonellosis. This model approach is based on inferring a posterior estimate of the considered outcome using prior assumptions and the use of data. In the Hald model, one assumption is that the number of human cases of salmonellosis is Poisson distributed:

$$o_i \sim Poisson\left(\sum_{j=1}^{J} \lambda_{ij}\right) \qquad (1)$$

where o_i is the number of observed cases for *Salmonella* of subtype *i*. The number of subtypes run from *i* = 1,2,..., *I*, where *I* is the total number of *Salmonella* subtypes present in the data. The number of sources in the data and thus considered in the model is *J*. Here, λ_{ij} is the number of expected cases caused by *Salmonella* subtype *i* in source *j* (with *j* running from *j* = 1,2,...,*J*). The Hald model defines λ_{ij} as follows:

$$\lambda_{ij} = M_j \cdot p_{ij} \cdot q_i \cdot a_j \qquad (2)$$

where M_j is the amount of source *j* consumed (in tons, except for laying hens where it is number of eggs). The values p_{ij} (in %) for the prevalence p_{ij} of *Salmonella* subtype *i* in the source *j*. The parameter q_i is a subtype-dependent factor which describes the ability of the *Salmonella* subtype *i* to cause illness. The parameter a_j is a source-dependent factor describing the ability of source *j* to serve as a vector for *Salmonella*. Equation 2 represents the multiparameter prior of the model with the two parameters a_j and q_i of unknown value. For the parameter a_j and q_i, uniform distributions where defined as prior distributions.

David model—a variation of the Hald model

Some authors describe difficulties with the convergence of the Hald model (Guo et al. 2011, Mullner et al. 2009). To address this issue, David et al. (2013) proposed a reparameterization of the Hald model based on unique types (or “specific types” as called in David et al. (2013)). A unique type is a subtype that is specific to a food-source and consequently is not found in another source under consideration. If there are one or multiple unique types for a source *j* in the considered data, then the corresponding unique subtype-dependent parameters $q_{ut,j}$ for these unique types are parameterized according to Equation 3 instead of Expression 6. The subscript, “ut” stands for unique type.

$$q_{ut,j} = \frac{o_{ut}}{\sum_i o_i} \cdot \frac{1}{p_{ut,j}} \qquad (3)$$

This reparameterization can only be done if all serotypes are phage typed. As not all the data of serotypes Enteritidis and Typhimurium considered by David et al. (2013) were phage typed, both serotypes were excluded from the reparameterization (see Section 2.2.5 in David et al. 2013).

Bayes data-based (DB) model—a variation of the David model

Following the idea of David et al. (2013) to use unique types for parameterizing the Hald model, the Bayes DB model uses the following parameterization setup:

1. Parameterization of the subtype-dependent parameter q_i

- For each source j, choose freely one unique type and call the corresponding subtype-dependent parameter $q_{ut,j}$ (if available)
- Parameterize $q_{ut,j}$ for the chosen unique type according to Equation 3
- If there are no unique types, parameterize all q_i according to Expression 6

2. Parameterization of the source-dependent parameter a_j

- For each source j where no unique type is available , the corresponding parameter a_{nut} is defined as ("nut" in subscripts stands for "no unique type"):

$$a_{nut} = \frac{\sum_i o_i}{M_{nut}} \qquad (4)$$

- This also applies to the case that no source has a unique type.

3. Parameterization of the consumption data M_j

- M_j is set according to consumption data (this is the case for the presented Bayes DB model).
- If no consumption data are available, all M_j are set to appropriate constant values. The values need to be large enough to assure consistent model results. These constant values are found through trial and error and depend on the considered dataset (for details see Section "The effect of consumption data on the consistency of source attribution estimates").

To estimate unknown parameters, uniform distributions are assumed as prior distributions for a_j and q_i (see Expressions 5 and 6, respectively). Unknown parameters are: 1) all q_i which belong to non-unique types, 2) unique q_i which have not been chosen according to the first step of the parameterization setup, 3)all a_j which correspond to sources j which have at least one unique type. Note that M_j is always set to a fixed value. Consequently, if there are no unique types, all a_j are parameterized according to Equation 4 and all q_i according to Expression 6.

In the model presented in this paper the following prior distributions were assumed:

$$a_j \sim uniform(0, 0.2) \qquad (j = 1, 2, .., J) \qquad (5)$$

$$q_i \sim uniform(0, 1) \qquad (i = 1, 2, ...I). \qquad (6)$$

The the limits of the prior distributions were chosen such that they produce complete posterior distributions for both datasets (baseline and monitoring data). Depending on the

data, one might have to adjust the limits of the distribution (see Section "The effect of prior distributions on completeness of posterior distributions" for details).

In the next section, we describe how to parameterize the model and run model simulations using FSKX format.

Simulations

All model parameters and their descriptions are presented in Table 1. Two simulation scenarios are provided in the fskx-model (see Table 2 for the parameter values of both scenarios and Suppl. material 1 for the fskx-model). The default simulation considers the data of the baseline study (see Section "Baseline data" for details). The second simulation setting is based on the monitoring data (see Section "Monitoring data" for details).

Table 1.

Description of the model parameters of the source attribution model. In the row that specifies the source, article always refers to the reference description of Jabin et al. (2019).

Id	list_sources
Classification	INPUT
Name	list_sources
Description	List all possible sources
Unit	[]
Data Type	INTEGER
Source	Article
Value	c('Broilers', 'Laying hens', 'Pigs', 'Turkeys')
Id	qfix_ind
Classification	INPUT
Name	qfix_ind
Description	Indices of subtype-dependent factor for subtype i (q_i), which will be set to fixed values. These are the four values for the human cases concerning the "unique types": S.Virchow, S.E. PT 1, S.T. DT 193, and S. Saintpaul
Unit	[]
Data Type	VECTOROFNUMBERS
Source	Data
Value	c(63,64,65,66)
Min Value	1

Max Value	Number of considered subtypes

Id	input_FileName
Classification	INPUT
Name	input_FileName
Description	Name of the file that contains the analysed data
Unit	[]
Data Type	STRING
Source	Article
Value	"Table2.csv"

Id	OpenBUGS_parameter
Classification	INPUT
Name	OpenBUGS_parameter
Description	The values that should be logged while running the OpenBUGS-model
Unit	[]
Data Type	STRING
Source	Article
Value	c("source", "unknown", "a", "q", "lambdaexp")

Id	OpenBUGS_niter
Classification	INPUT
Name	OpenBUGS_niter
Description	Number of total iterations per chain used in the OpenBUGS-model
Unit	[]
Data Type	INTEGER
Source	Article
Value	30000
Min Value	OpenBUGS_nburnin+1

Id	OpenBUGS_nburnin
Classification	INPUT
Name	OpenBUGS_nburnin
Description	Length of burn in, i.e. number of iterations to discard at the beginning.

Unit	[]
Data Type	INTEGER
Source	Article
Value	10000
Min Value	1
Id	aValue
Classification	INPUT
Name	aValue
Description	Values for the source-dependent factors (a_i) that are used to determine inital values for the OpenBUGS model
Unit	[]
Data Type	VECTOROFNUMBERS
Source	Data
Value	c(0.002,0.001,0.19, 0.18, 0.178)
Min Value	0
Id	qValue
Classification	INPUT
Name	qValue
Description	Values for the subtype-dependent factors (q_i) that are used to determine inital values for the OpenBUGS model
Unit	[]
Data Type	VECTOROFNUMBERS
Source	Data
Value	c(0.001,0.002, 0.199, 0.18, 0.175)
Min Value	0
Id	OpenBUGS_model
Classification	INPUT
Name	OpenBUGS_model
Description	The filename of the txt-file that contains the OpenBUGS-model
Unit	[]
Data Type	STRING

Source	The filename is freely chosen. The BUGS-model is descrided in the reference article.
Value	"BugsModel.txt"

Id	mean_res
Classification	OUTPUT
Name	mean_res
Description	Mean number of estimated human salmonellosiscases attribute to potential sources
Unit	Cases
Data Type	VECTOROFNUMBERS
Min Value	0
Max Value	1

Id	quantil_95
Classification	OUTPUT
Name	quantil_95
Description	95%-quantile of estimated human salmonellosiscases attributed to the potential sources
Unit	Cases
Data Type	VECTOROFNUMBERS
Min Value	0
Max Value	1

Id	quantil_05
Classification	OUTPUT
Name	quantil_05
Description	5%-quantile of estimated human salmonellosiscases attributed to the potential sources
Unit	Cases
Data Type	VECTOROFNUMBERS
Min Value	0
Max Value	1

Table 2.

The simulation settings for the source attribution model. The settings specify the parameter names and the values (see Table 1 for details about the parameters).

defaultSimulation	
list_sources	c('Broilers', 'Laying hens', 'Pigs', 'Turkeys')
qfix_ind	c(63,64,65,66)
input_FileName	"Table2.csv"
OpenBUGS_parameter	c("source", "unknown", "a", "q", "lambdaexp")
OpenBUGS_niter	30000
OpenBUGS_nburnin	10000
aValue	c(0.002,0.001,0.19, 0.18, 0.178)
qValue	c(0.001,0.002, 0.199, 0.18, 0.175)
OpenBUGS_model	"BugsModel.txt"
SimulationTable3	
list_sources	c('Broilers', 'Laying hens', 'Pigs', 'Turkeys')
qfix_ind	c(30,31,32,33)
input_FileName	"Table3.csv"
OpenBUGS_parameter	c("source", "unknown", "a", "q", "lambdaexp")
OpenBUGS_niter	30000
OpenBUGS_nburnin	10000
aValue	c(0.01, 0.015, 0.099, 0.08,0.02)
qValue	c(0.001, 0.002, 0.9,0.85, 0.99)
OpenBUGS_model	"BugsModel.txt"

The Bayes DB model is implemented in the programming language R (R Core Team 2019). In addition to R, the open source software OpenBUGS is required to successfully execute the model (Neal 2009). The linkage between both software tools is done by the R package "R2OpenBUGS" (Sturtz et al. 2005 ; see file "packages.json" in the fskx-model).

The fskx-model can be executed, developed further, and easily adapted to new data on the local computer, e.g., using the KNIME extension FSK-Lab (see https://foodrisklabs.bfr.bund.de/fsk-lab/ and de Alba Aparicio et al. (2018)).

Executable model

In order to execute the model, please register at the virtual research environment "FMJ_Lab".

Execute with default simulation parameters: execute

The default simulation runs for 2 minutes 11 seconds on the virtuel research environment.

Execute another simulation scenario or create a personalized scenario: execute

Results

The main result is that the existing source attribution model previously published in Jabin et al. (2019) is available in the ready-to-use FSKX-format. The format is an open information exchange format and uses model metadata and controlled vocabulary to harmonise annotations. The transformation into an FSKX compliant model took about one day of work for a person already familiar with the format. In the FSKX compliant format the model predicts the same source attribution as in the original version as the R and the OpenBUGS code is nearly identical to the code used in Jabin et al. (2019).

To be able to successfully use the model, it is important to know how to set up and run the model as well as assess the appropriateness of the results. We present these practical issues since this is a purely technical paper it seems appropriate to provide this level of technicality here.

Successfully executing a Bayesian model using Markov chain Monte Carlo simulations

When running our Bayesian model using Markov Chain Monte Carlo (MCMC) methods, we studied three important aspects of model diagnostics. To ensure a high quality estimation of unknown parameters, we check the following aspects of a MCMC method: the convergence behaviour of the Markov chains, the completeness of posterior distributions, and the consistency of results.

The effect of prior distributions on completeness of posterior distributions

The limits for the uniform distribution have a strong influence on the completeness of the posterior distributions. The limits are incorporated into the OpenBUGS code of the model (see file "BugsModel.txt" in the fskx-model). In the Bayes DB model, the lower limit is 0 and the upper limits are 0.2 for a_j and 1 for q_i for both datasets (see Section "Bayes data-based (DB) model—a variation of the David model" and Expressions 5 and 6). The upper limits were chosen such that the model provides complete posterior distributions. This was assured by examining visually the plots of the posterior distributions of a_j and q_i (Hald et al. 2004; and Fig. 1 in this paper). If one changes the upper limit of the prior distribution of q_i to 0.2, incomplete posteriors were obtained. In Fig. 1, there is an example for a complete

and an incomplete posterior distribution (Subfigures A) and B), respectively). The incomplete posterior distribution is a trimmed version of the complete one. Please note that issues in the completeness of posterior distributions might only occur in some but not all of the model parameters.

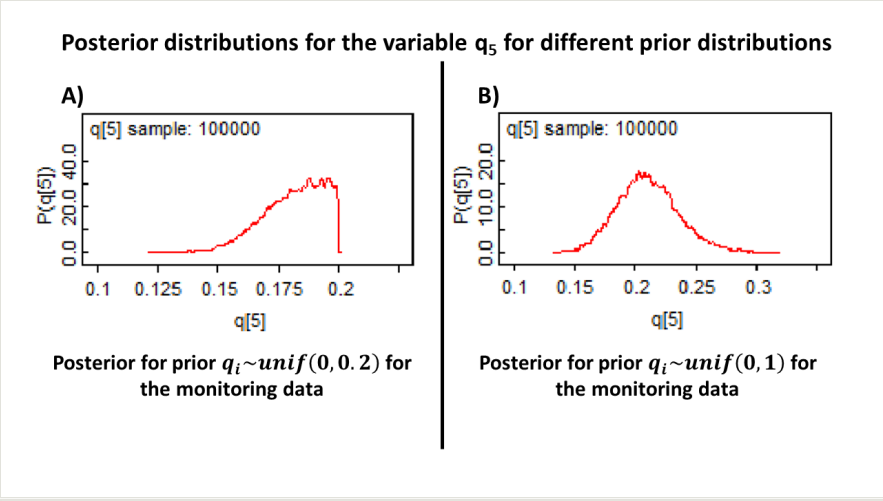

Figure 1. doi

The posterior distributions for the fifth entry in the list of *Salmonella* subtypes (q_5), which is *S. enterica* serotype Enteritidis PT 21, as a function of the possible values of q_5. The shown posterior distributions are calculated by the Gibbs-Sampler software OpenBUGS using the Bayes DB model presented in Jabin et al. (2019) for the monitoring data and a sample size of $1e^5$. Subfigure A) shows the posterior distribution calculated for $q_i \sim uniform(0, 0.2)$. The chosen limits of the uniform distribution lead to a cut off in the posterior distribution. When enlarging the interval defining the prior distribution to $q_i \sim uniform(0,1)$, then a complete posterior distribution is produced by OpenBUGS (see Subfigure B)).

The effect of initial values on convergence and uncertainty estimates

The choice of the starting values of the Markov chains (also known as initial values) has an impact on the convergence and uncertainty estimates of the model calculation. The model runs with five Markov chains. The default starting values for the five chains are listed in Table 1; the parameter names for a_j and q_i are "aValue" and "qValue", respectively. This means that the Markov chains for each unknown parameters start with five different, but predefined, starting points. The effect of initial values on the convergence and uncertainty estimates will be exemplified using the baseline data in three parameter scenarios. The parameter scenarios differ in their corresponding set of five starting points for their five Markov chains. The parameter scenarios exemplify the following effects: successful convergence (Parameter scenario 1), slow convergence (Parameter scenario 2), and no convergence (Parameter scenario 3). The parameter scenarios are represented graphically in Figs 2, 3, 4 which show the starting points for the Markov chains in scatter plots, convergence behaviour of the Markov chains in trace plots, and the result of the source

attribution estimates in bar plots. The bar plots include error bars which correspond to a 90 % equal-tailed interval (i.e. the interval between the 0.05-quantile and the 0.95-quantile of the posterior distribution of the number of human cases). The error bars represent the uncertainty in the model estimation; the bigger the bars the higher the uncertainty.

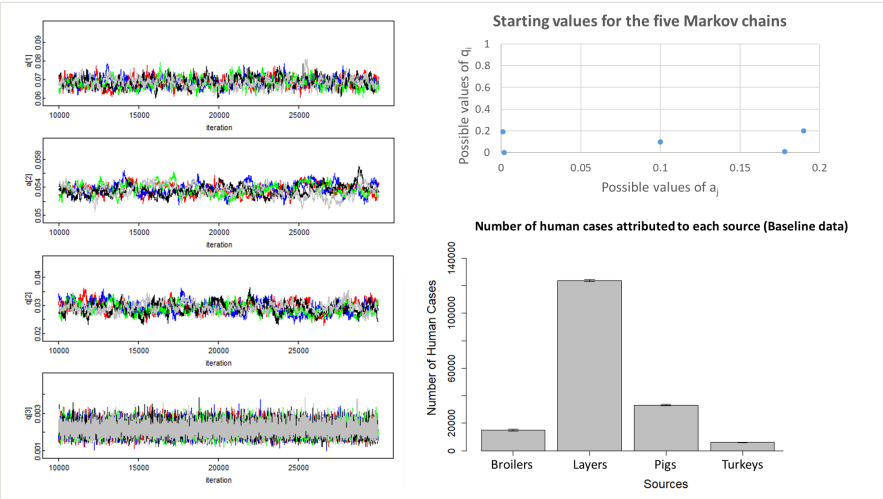

Figure 2. doi

Starting points for the Markov chains of Parameter scenario 1, their effects on the convergence behaviour and the model predictions. The starting points are evenly spaced in the lower fifth of the space of possible starting points (see the points in the scatter plot in the upper right corner). With this set of starting points, Markov chains converge quickly as can be seen in the four trace plots on the left hand side which show how the paramter values that the model estimates change through the iteration steps of the model calculations. Each of the four trace plots correspond to one model parameter (a_1, a_2, q_2 and q_3, where types 1, 2 and 3 correspond to *S. enterica* serotype Enteritidis PT 11, PT 14b, and PT 19, respectively). In each trace plot there are five traces, one trace for each Markov chain. Each Markov chain has its own colour. The predicted source attribution shows small error bars (see the bar plot).

In Parameter sScenario 1, the starting points are evenly spaced in the lower fifth of the plane of possible starting values (see Fig. 2). The error bars in the source attribution results are small (see bar plot in Fig. 2).

In Parameter scenario 2, the starting points are concentrated near two points: one point is (0, 0) the other (0.18, 0.18) (see Fig. 3). The Markov chains converge slowly for some parameters, e.g., a_2 and q_3 in the trace plot of Fig. 3. The error bars of the model results are large (much larger than in the previous parameter scenario) (see bar plot in Fig. 3).

Finally, the starting points cluster near two points: one point is (0, 0.18) the other (0.18, 0.18) (see Fig. 4). In this parameter scenario, the Markov chains do not converge at all within the 30,000 iterations for some parameters, e.g., for a_2 and q_3 (the trace plot of Fig. 4). Consequently, the error bars of the model results are larger than in the previous two parameter scenarios (see bar plot in Fig. 4).

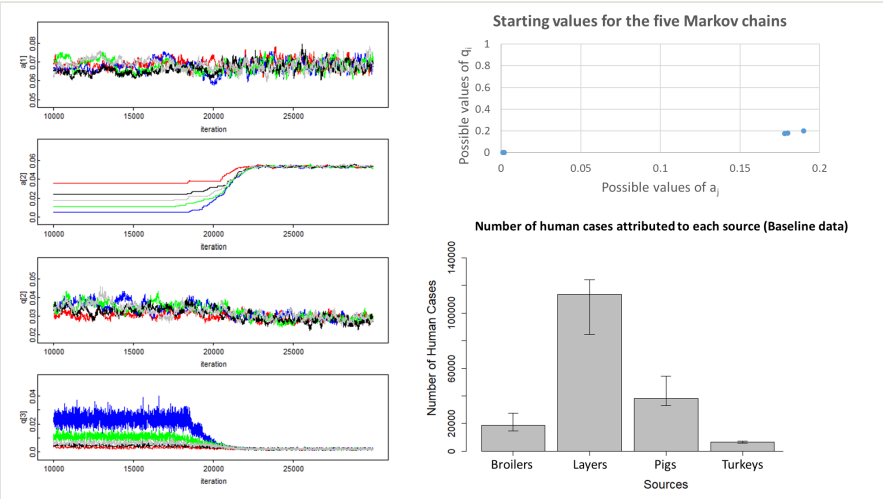

Figure 3. doi

Starting points for the Markov chains of Parameter scenario 2 and their effects on the convergence behaviour and the model predictions. The starting points are concentrated near the points (0, 0) and (0.18, 0.18) (see the points in the scatter plot in the upper right corner). With this set of starting points, Markov chains converge slowly as can be seen in the four trace plots on the left hand side which show how the paramter values that the model estimates change through the iteration steps of the model calculations. Each of the four trace plots correspond to one model parameter (a_1, a_2, q_2 and q_3, where types 1, 2 and 3 correspond to *S. enterica* serotype Enteritidis PT 11, PT 14b, and PT 19, respectively). In each trace plot there are five traces, one trace for each Markov chain. Each Markov chain has its own colour. The predicted source attribution shows small error bars (see the bar plot).

The effect of consumption data on the consistency of source attribution estimates

Some authors pointed out that the parameter for consumption data, M_j, are not essential for the approach (Mughini-Gras and 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)). The approach of setting M_j to 1, caused problems for the Bayes DB model. Problems are either numerical issues or inconsistent results. Inconsistency means that the predicted number of human cases for a certain subtype is not in the same order of magnitude of the number of human cases found in the data (in such cases we found that the number of observed cases could be up to a factor of 2000 larger than the model estimates).

Simplifying the Bayes DB model for the baseline data by setting all M_j to 1 and keeping the prior distributions as they were defined in Expression 5 and Expression 6 caused problems. OpenBUGS was not able to successfully execute the model, due to numerical problems (OpenBUGS reports an "conjugate gamma updater error" for one of the q_i). This problem disappeared when the prior distributions for q_i were changed to

q_i ~ *uniform*(0,2600000 but the model results remained inconsistent. Changing the prior distribution for a_j to a_j ~ *uniform*(0,30000) led to consistent results.

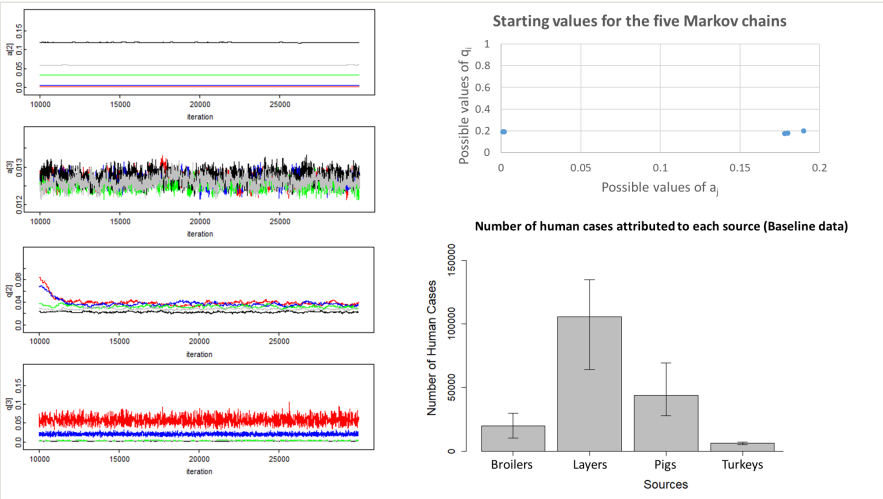

Figure 4. doi

Starting points for the Markov chains of Parameter scenario 3 and their effect on the convergence behaviour and the model predictions. The starting points are concentrated near the points (0, 0.18) and (0.18, 0.18) (see the points in the scatter plot in the upper right corner). With this set of starting points the Markov chains do not converge within 30,000 iterations for the parameters a_2 or q_3 as can be seen in the four trace plots on the left hand side which show how the paramter values that the model estimates change through the iteration steps of the model calculations. Each of the four trace plots correspond to one model parameter (a_2, a_3, q_2 and q_3, where types 2 and 3 correspond to *S. enterica* serotype Enteritidis PT 14b and PT 19, respectively). In each trace plot there are five traces, one trace for each Markov chain. Each Markov chain has its own colour). The error bars of the predicted source attribution are large (see the bar plot).

For the monitoring data, setting all M_j to 1 and using prior distributions as defined in Expression 5 and 6 led to inconsistent results. The model worked properly when priors distributions were set to a_j ~ *uniform*(0,30000) while the priors for q_i remained the same as in Expression 6 (cf. Fig. 5).

One way to interpret the need for enlarging the priors for a_j and q_i is that parameters a_j and q_i must compensate for the restrictions applied to M_j. One may consider a_j and q_i as complex priors distributions that combine estimates of the potential of the *Salmonella* of type *i* and the source *j* to cause salmonellosis. In summary, if M_j is simplified, the prior distributions may need to be adjusted.

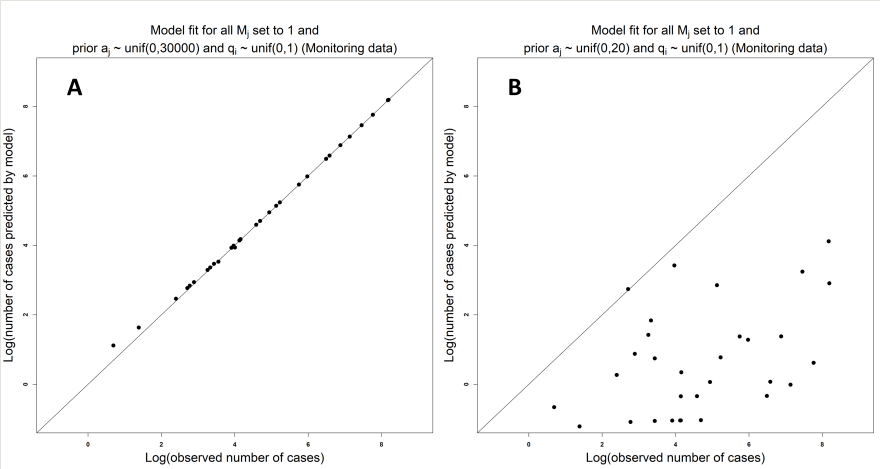

Figure 5. doi

Model-data fit when setting M_j=1 and using different parameterizations. Each point in the figure corresponds to one bacterial subtype. Subfigure A shows a consistent model fit due to using the prior a_j ~ *uniform*(0,30000). I.e. the logarithm of the number of cases as found in the data corresponds well to the logarithm of predicted number of cases. Subfigure B shows an inconsistent model fit due to using the prior a ~ *uniform*(0,20). Here, the model systematically underestimates the number of cases for the subtypes as the points gather well below the identity line.

Source attribution determination

Source attribution methods aim to identify and quantify the contribution of different sources to disease burdens like salmonellosis (Jabin et al. 2019). The human salmonellosis cases are attributed to different sources (namely broilers, laying hens, pigs, turkeys, and unknown). In Fig. 6A, the number of human cases of *Salmonella* subtypes in animal sources from the baseline studies 2004–2007 are presented. The source that causes the majority of salmonellosis cases is laying hens while turkeys cause the smallest burden of the considered sources. The results for the monitoring data (2010/2011) are presented in Fig. 6B. The majority of cases here results from unknown sources. Closely followed by laying hens and pigs. A relatively low burden results from turkeys.

The presented results allow to analyse the quantity of the burden assignable to each source and provide the basis to compare different datasets. Although the baseline and the monitoring data are comparable and no significant difference between the datasets is expected (see Section "Data" for details), it provides the basis for comparison. There is much more to say about the model and its results but we focus here on the technical aspects of making the model FSKX compliant and some of the model mechanics. For a more detailed discussion of the model and its results see Jabin et al. (2019).

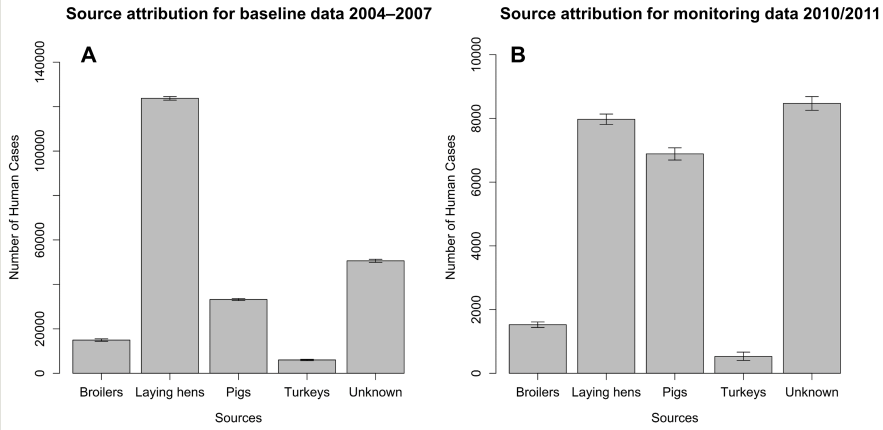

Figure 6. doi

Bar plot of number of human cases of *Salmonella* infection attributed to different sources. Subfigure A shows the result for the baseline data 2004–2007 (the so-called defaultSimulation in Table 2) and Subfigure B the results for themonitoring data 2010/2011 (the so-called SimulationTable3 in Table 2).

Discussion

Modelling of source attribution is a powerful approach that can contribute to the reduction of human zoonotic cases, in particular salmonellosis. However, model results are highly sensitive to changes of multiple parameters that can differ for each model. In the presented model, these parameters include the initial values for observed *Salmonella* cases and the assumption about the consumption data. If someone aims to reproduce the model results, this is only possible if the parameter settings are identical to the original settings. In other words, slight changes in a model parameter might result in a big change in the model prediction and thus, the results presented in an article or report cannot be reproduced. The issue of reproducible results is a general challenge in science (Baker 2016, Goodman et al. 2016). It might seem that in computational work it is in principle easy to re-use and share the used data and the used code in order to reproduce results (except for variations when the model calculations include probabilistic elements). Nonetheless, there is a problem with reproducibility in this area as well (Waltemath and Wolkenhauer 2016, Stodden et al. 2018, Tiwari et al. 2021,Miłkowski et al. 2018). It can be hard to re-use one own modestly documented models; it can be particularly challenging to re-use models developed by other authors (see Topalidou et al. 2015 for an illustrative example). Problems arise from the limited amount of documentation and versioning of the code and/or insufficient information about the model scope (Stodden et al. 2018, Waltemath and Wolkenhauer 2016). The consequence of the reproducibility/sharing problem is that models are re-invented and re-implemented; a time-consuming and/or error-prone process. Several approaches have been discussed in the literature to remedy these problems (Grimm et al. 2014, Wilson et al.

2017, Schölzel et al. 2021, Kim et al. 2018, McDougal et al. 2016, Tiwari et al. 2021). All approaches come down to a combination of standardized way to document the model and to choose ways to store and share computational resources like data and model code platform independently. FSKX format is an open information exchange format that provides a way to create well-documented and reproducible mathematical risk assessment models that are annotated in a harmonised way using model metadata and controlled vocabulary (de Alba Aparicio et al. 2018).

The implementation of a model in a standardized and annotated exchange format like FSKX-format is a way that focuses on long-term usability and understandability of the model. The community as well as the creators would benefit from such an approach. One example where a creator developed a model with an FSKX conform end-product in mind is the work of Plaza-Rodríguez et al. (2019).

Much time-consuming and/or error-prone work can be saved in the future if model development is done with a mind-set of long-term usability, reproducibility, and understandability. The FSKX format enables sharing model code reliably and reproducibly and thus paves the way for successful collaboration and further development of models.

Conclusion

In this work, we demonstrated that it is straight forward to take a Bayesian source attribution model running under R and OpenBUGS originally published in Jabin et al. (2019), and translate it into the Food Safety Knowledge Exchange (FSKX) format. This standardized format provides an annotated model together with relevant simulation settings. The ready-to-use model can be executed in this "executable paper" and on the local computer, e.g., using software like the KNIME extension FSK-Lab (de Alba Aparicio et al. 2018). In addition, it is easy to re-use the model code and interpret simulation results. In conclusion, we provide an annotated, ready-to-use source attribution model and the considered *Salmonella* datasets and by that facilitate model exchange, adjustment, and re-use by the international and multi-sectoral One Health community.

Acknowledgements

We like to acknowledge the original work of Hannah Jabin and Lars Valentin and furthermore thank the colleagues from the Robert Koch Institute (W. Rabsch) and German Federal Institute for Risk Assessment (C. Dorn, A. Schroeter, R. Helmuth, A. Friedrich and I. Szabo) for providing the data on *Salmonella* isolates from humans and animals, respectively.

Funding program

EMS is funded by the JIP MATRIX within the One Health EJP. One Health EJP has received funding from the European Union Horizon 2020 research and innovation program

under grant agreement No 773830. Gathering the data and analyzing it with source attribution models was initially done as a part of the project RESET which was financially supported by the German Federal Ministry of Education and Research (BMBF) through the German Aerospace Center grant number 01KI1013A-H.

Author contributions

Esther M. Sundermann: Conceptualization, Data Curation, Project administration, Software, Visualization, Writing - Original Draft, Writing - Review & Editing. Guido Correia Carreira: Conceptualization, Formal analysis, Writing - Original Draft, Visualization, Writing - Review & Editing. Annemarie Käsbohrer: Data Curation, Writing - Review & Editing. The author contributions are taken from https://www.elsevier.com/authors/policies-and-guidelines/credit-author-statement

References

- Arnold M, Smith RP, Tang Y, Guzinski J, Petrovska L (2021) Bayesian Source Attribution of Salmonella Typhimurium Isolates From Human Patients and Farm Animals in England and Wales. Frontiers in microbiology 12: 579888. https://doi.org/10.3389/fmicb.2021.579888
- Baker M (2016) 1,500 scientists lift the lid on reproducibility. Nature 533 (7604): 452-4. https://doi.org/10.1038/533452a
- Batz MB, Doyle MP, Morris Jr JG, Painter J, Singh R, Tauxe RV, Taylor MR, Wong DMLF, Food Attribution Working Group (2005) Attributing illness to food. Emerging Infectious Diseases 11 (7): 993.
- Boysen L, Rosenquist H, Larsson JT, Nielsen EM, Sorensen G, Nordentoft S, Hald T (2014) Source attribution of human campylobacteriosis in Denmark. Epidemiology and Infection 142 (8): 1599-608. https://doi.org/10.1017/S0950268813002719
- David JM, Guillemot D, Bemrah N, Thebault A, Brisabois A, Chemaly M, Weill FX, Sanders P, Watier L (2013) The Bayesian microbial subtyping attribution model: robustness to prior information and a proposition. Risk Analysis 33 (3): 397-408. https://doi.org/10.1111/j.1539-6924.2012.01877.x
- de Alba Aparicio M, Buschhardt T, Swaid A, Valentin L, Mesa-Varona O, Günther T, Plaza-Rodriguez C, Filter M (2018) FSK-Lab—An open source food safety model integration tool. Microbial Risk Analysis 10: 13-19. https://doi.org/10.1016/j.mran.2018.09.001
- 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). URL: https://doi.org/10.2903/j.efsa.2007.97r
- 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). URL: https://doi.org/10.2903/j.efsa.2007.98r

- EFSA (2008a) 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. EFSA Journal 134: 1-91.
- 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. EFSA Journal 135: 1-111.
- European Food Safety Authority, European Centre for Disease Prevention and Control (2021) The European Union One Health 2019 Zoonoses Report. EFSA Journal 19 (2): 6406. https://doi.org/10.2903/j.efsa.2021.6406
- Goodman SN, Fanelli D, Ioannidis JPA (2016) What does research reproducibility mean? Science translational medicine 8 (341): 341ps12. https://doi.org/10.1126/scitranslmed.aaf5027
- Grimm V, Augusiak J, Focks A, Frank B, Gabsi F, Johnston AA, Liu C, Martin B, Meli M, Radchuk V, Thorbek P, Railsback S (2014) Towards better modelling and decision support: Documenting model development, testing, and analysis using TRACE. Ecological Modelling 280: 129-139. https://doi.org/10.1016/j.ecolmodel.2014.01.018
- Guo C, Hoekstra R, Schroeder C, Pires SM, Ong KL, Hartnett E, Naugle A, Harman J, Bennett P, Cieslak P, Scallan E, Rose B, Holt K, Kissler B, Mbandi E, Roodsari R, Angulo F, 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
- Haberbeck LU, Plaza-Rodrguez C, Desvignes V, Dalgaard P, Sanaa M, Guillier L, Nauta M, Filter M (2018) Harmonized terms, concepts and metadata for microbiological risk assessment models: The basis for knowledge integration and exchange. Microbial Risk Analysis 10: 3-12. https://doi.org/10.1016/j.mran.2018.06.001
- Hald T, Vose D, Wegener HC, 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
- Jabin H, Correia Carreira G, Valentin L, Käsbohrer A (2019) The role of parameterization in comparing source attribution models based on microbial subtyping for salmonellosis. Zoonoses and Public Health 66 (8): 943-960.
- Käsbohrer A, Tenhagen BA, Alt K, Schroeter A, Szabo I, Hartung M (2012) BfR Wissenschaft: Erreger von Zoonosen in Deutschland im Jahr 2010. In: Hartung M, Käsbohrer A (Eds) BfR Wissenschaft: Erreger von Zoonosen in Deutschland im Jahr 2010. BfR Wissenschaft, 103 pp.
- Käsbohrer A, Tenhagen BA, Alt K, Schroeter A, Szabo I, Hartung M (2013) BfR Wissenschaft: Erreger von Zoonosen in Deutschland im Jahr 2011. In: Hartung M, Käsbohrer A (Eds) BfR Wissenschaft: Erreger von Zoonosen in Deutschland im Jahr 2011. BfR Wissenschaft, 109 pp.
- Kim Y, Poline J, Dumas G (2018) Experimenting with reproducibility: a case study of robustness in bioinformatics. GigaScience 7 (7). https://doi.org/10.1093/gigascience/giy077
- Mather AE, Vaughan TG, French NP (2015) Molecular approaches to understanding transmission and source attribution in nontyphoidal *Salmonella* and their application in Africa. Clinical Infectious Diseases 61 https://doi.org/10.1093/cid/civ727

- McDougal RA, Bulanova AS, Lytton WW (2016) Reproducibility in computational neuroscience models and simulations. IEEE transactions on bio-medical engineering 63 (10): 2021-35. https://doi.org/10.1109/TBME.2016.2539602
- Miłkowski M, Hensel WM, Hohol M (2018) Replicability or reproducibility? On the replication crisis in computational neuroscience and sharing only relevant detail. Journal of computational neuroscience 45 (3): 163-172. https://doi.org/10.1007/s10827-018-0702-z
- 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
- Mullner P, Jones G, Noble A, Spencer SE, Hathaway S, French NP (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
- Munck N, Leekitcharoenphon P, Litrup E, Kaas R, Meinen A, Guillier L, Tang Y, Malorny B, Palma F, Borowiak M, Gourmelon M, Simon S, Banerji S, Petrovska L, Dallman TJ, Hald T (2020) Four European Salmonella Typhimurium datasets collected to develop WGS-based source attribution methods. Scientific data 7 (1): 75. https://doi.org/10.1038/s41597-020-0417-7
- Neal T (2009) Overview. http://openbugs.net/w/FrontPage. Accessed on: 2021-2-05.
- Pires S, Evers E, van Pelt W, Ayers T, Scallan E, Angulo F, Havelaar A, Hald T (2009) Attributing the human disease burden of foodborne infections to specific sources. Foodborne pathogens and disease 6 (4): 417-424.
- Pires S, 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. EFSA Supporting Publications 8 (8). https://doi.org/10.2903/sp.efsa.2011.en-184
- Plaza-Rodríguez C, Kaesbohrer A, Tenhagen B (2019) Probabilistic model for the estimation of the consumer exposure to methicillin-resistant Staphylococcus aureus due to cross-contamination and recontamination. MicrobiologyOpen 8 (11): e900. https://doi.org/10.1002/mbo3.900
- 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
- R Core Team (2019) R: A language and environment for statistical computing. https://www.R-project.org
- Schölzel C, Blesius V, Ernst G, Goesmann A, Dominik A (2021) Countering reproducibility issues in mathematical models with software engineering techniques: A case study using a one-dimensional mathematical model of the atrioventricular node [Preprint , Posted February 19, 2021 and retrieved 3rd of May 2021]. bioRxiv 2021.02.19.431951 https://doi.org/10.1101/2021.02.19.431951
- Stodden V, Seiler J, Ma Z (2018) An empirical analysis of journal policy effectiveness for computational reproducibility. Proceedings of the National Academy of Sciences of the United States of America 115 (11): 2584-2589. https://doi.org/10.1073/pnas.1708290115
- Sturtz S, Ligges U, Gelman A (2005) R2WinBUGS: A package for running WinBUGS from R. Journal of Statistical Software 12 (3).

- Taylor LH, Latham SM, Woolhouse ME (2001) Risk factors for disease emergence. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 356: 983-989.
- The World Bank (2012) People, pathogens and our planet. Volume 2: The economics of One Health. The World Bank, Washington, DC, US.
- Tiwari K, Kananathan S, Roberts MG, Meyer JP, Sharif Shohan MU, Xavier A, Maire M, Zyoud A, Men J, Ng S, Nguyen TVN, Glont M, Hermjakob H, Malik-Sheriff RS (2021) Reproducibility in systems biology modelling. Molecular systems biology 17 (2): e9982. https://doi.org/10.15252/msb.20209982
- Topalidou M, Leblois A, Boraud T, Rougier NP (2015) A long journey into reproducible computational neuroscience. Frontiers in computational neuroscience 9: 30. https://doi.org/10.3389/fncom.2015.00030
- Wahlstrom H, Andersson Y, Plym-Forshell L, Pires SM (2011) Source attribution of human *Salmonella* cases in Sweden. Epidemiology and Infection 139 (8): 1246-1253. https://doi.org/10.1017/S0950268810002293
- Waltemath D, Wolkenhauer O (2016) How modeling standards, software, and initiatives support reproducibility in systems biology and systems medicine. IEEE transactions on bio-medical engineering 63 (10): 1999-2006. https://doi.org/10.1109/TBME.2016.2555481
- Wilson G, Bryan J, Cranston K, Kitzes J, Nederbragt L, Teal TK (2017) Good enough practices in scientific computing. PLoS computational biology 13 (6): e1005510. https://doi.org/10.1371/journal.pcbi.1005510

Supplementary material

Suppl. material 1: SourceAttributionModel.fskx doi

Authors: Esther M. Sundermann
Data type: fskx-model
Download file (1.44 MB)