SAPIENZA
UNIVERSITÀ DI ROMA

Mathematical models for food safety in public health

Dottorato di Ricerca in Matematica - XXXIV Ciclo

Scuola di Dottorato in Scienze Astronomiche, Chimiche, Fisiche e Matematiche 'Vito Volterra'

Istituto Superiore di Sanità
Dipartimento di Sicurezza Alimentare, Nutrizione e Sanità Pubblica Veterinaria

Advisor
Prof. Marco Isopi
Dr. Gaia Scavia

Candidate
Ornella Moro
ID number 1324468

April 2022

Ornella Moro

Mathematical models for food safety in public health

PhD Thesis – University of Rome "La Sapienza"

moro@mat.uniroma1.it

If I had to choose, I would not choose you
If I had to stop, I would turn around and choose glue
If I had to hit the brakes, I'ma stop right here

Don't waste your mind
I'll be the one to say it all
To do what I am
To say what I am

Non è stata la sfortuna. Non è stata la congiura.
Non è stato il destino. Nemmeno gli alieni, l'Isis,
l'arbitro o quel burlone di Dio.
Sono stato, sono stato io.

Sono stato io. Sono stato io. Sono stato io.
(Sono stato io) Colpo di frusta sul colpo di mano.
Colpo di grazia chiesta al dio dell'Uragano.
(Sono stato io) Quando appare la noia in paranoia
ma, in quanto a pare, il vero condannato è il boia.
(Sono stato io) Non perdonatemi che so quello che faccio
e dentro al buio, a quanto pare, non c'è niente
a parte il rimorso di essere innocente.

(Siamo stati Noi) A domare il mare con la frusta,
scatenando la reazione opposta
(Siamo stati Noi) Ad affrontare la stessa tempesta
ma con una nave a testa.
Siamo stati Noi.

E come Damocle
non abbiamo mai ballato meglio
che con quella spada sulla testa.

Indice

Abstract

This project of applied science was undertaken to provide evidence to face complex and multifaceted issues posed by the control of zoonotic foodborne pathogens and the related challenges for public health, veterinary medicine, food and environmental safety.

The thesis is divided in two different parts, each addressing a specific zoonotic pathogens: Hepatitis E virus (HEV) and Shiga toxin-producing *E.coli* (STEC). The main goal was to produce estimations to help targeting intervention strategies in livestock and in food production chain.

In Part I, a bottom-up risk assessment approach has been followed using original and existing data, to analyse the risk of consumers exposure to HEV through the consumption of contaminated food, following the from farm to fork approach, i.e. considering the two main compartments of the food production chain: the primary production level and the consumer level. In the pre-harvest block the HEV transmission dynamic was modeled to identify risk factors influencing the prevalence of HEV positive pigs entering the food-production chain at slaughtering. In the post-harvest block we compared several foodstuff to evaluate the risk of transmission of HEV to humans and to underline possible risk factors in the primary production that can lead to higher contamination of food products.

In Part II, the objective of our quantitative assessment were Shiga toxin-producing E.coli (STEC). In this case the availability of subtyping characterisation data of STEC isolated from different non-human sources and the complexity of transmission cycle of the STEC led us to choose a top-down approach, from the reported cases of infection in humans, through the food chain up to the primary source of STEC infection. This project was carried out within an ongoing European project (DISCOVER) were a original dataset were put together. We were able to adapt classical source attribution model to STEC using a novel approach for pathogenicity definition that allowed us to weigh the importance of the different sources in causation of human infection.

Acknowledgments

First of all, I would like to thank my advisor prof. Marco Isopi for making this project possible and for teaching me how to look at things in a new and different way.

A huge thanks to my external advisor, Dr. Gaia Scavia who guided me step by step in this new world showing me the beauty of this profession. She taught me to look honestly to my strengths and weaknesses and with her priceless energy she led me towards the light of this long journey.

I also would like to thank all the colleagues that helped me on the way. Elisabetta Suffredini, Ilaria Di Bartolo, Maria Elena Tosti, Rosangela Tozzoli and Stefano Morabito from Istituto Superiore di Sanità; the colleagues from DTU, where I spent a lovely visiting period, Tine Hald and Ofosuhene Okofrobour Apenteng; the colleagues Lapo Mughini-Gras from RIVM and Sara Monteiro Pires from DTU for the big support in the source attribution part.
A special thanks to Eleonora Ventola and Michele Luca D'Errico that were not only colleagues but friends with whom I had the opportunity to share a wonderful period of my life.

Introduction

Public health challenges in infectious diseases in the global world

Infectious diseases are one of the major cause of human morbidity and mortality in many countries with between 13 and 15 million deaths worlwide annually [32]. While at the beginning of the XX century high-income countries dominated some of the most dangerous infectious diseases, in most of the medium and low income countries they are still one of the main cause of death, especially in pediatric age. In the last fifty years a number of emerging and re-emerging pathogens caused large outbreaks in Europe and North America.

Infectious disease epidemiology is characterised by the presence of at least one active player in addition to the human or animal population, namely the infectious agent or parasite. Transmission from one host to another is fundamental to the survival strategy of the infectious agent since any host will eventually either clear the infection or die. A consequence of transmission is that the occurrence of infectious disease events in individuals depends on the occurrence of that disease or infection in other members of the population. Sir Ronald Ross (1916) called this dependence of disease events in infectious diseases "dependent happenings".
Globalization, soil exploitation, demographic rise, deforestation and several other factors contributed to make this process of emergence, re-emergence and spread of pathogenic organisms to humans even more complex. In this context the dynamic of pathogen spread among species and ecological niches can be very complex and highly nonlinear.

Quantitative studies are to date one of the most efficient way to understand this complexity and to inform control policies to reduce the burden of infectious diseases on human health. The possibility to predict and simulate different kind of scenarios and/or to address data collection and analysis made mathematical modelling one of the most important phase in the infectious disease research activity.

Basic concepts in infectious disease epidemiology

Infectious diseases can be classified based on many different criteria. Among these, the most frequently adopted for this purpose include the etiological agent, that refers to the nature and biology of the causative agent (bacteria, viruses, parasites, fungi etc.),

the hosts implicated in the epidemiological transmission pathways (e.g. human diseases, zoonoses) and the main transmission routes: airborne disease (e.g. measles, influenza, etc.), sexual diseases (e.g. HIV), vectorborne (e.g. Lyme disease, dengue, malaria, etc.), foodborne and waterborne disease (Salmonella, Hepatitis A).

Box 1 (*Infectious disease epidemiology – basic concept and glossary*)

Infectious (transmissible, contagious) disease: *a disease due to a specific infectious agent or its toxic products that arises through transmission of that agent or its products from an infected host (person, animal or reservoir) to a susceptible host, either directly or indirectly through an intermediate plant or animal host, vector, or the inanimate environment.*

Exposed: *a person or animal coming into contact with a potential causal virus or bacteria or other infectious agent and placed in a situation where effective transmission could occur. The exposure does not necessarily succeeds in the transmission of the disease agent and the infection and the development of the disease.*

Infection: *Invasion by and multiplication of pathogenic microorganisms in a bodily part or tissue, which may produce subsequent tissue injury and progress to overt disease through a variety of cellular or toxic mechanisms.*

Reservoir: *any person, animal, plant, or environmental medium (soil, water) in which the microorganisms normally lives and multiply, on which depends primarily for survival, and where they multiply in such a manner that they can be transmitted to the susceptible host/s.*

Source of infection: *actual person, animal, or object from which the infection was acquired.*

Vehicle: *inanimate object that served to communicate disease.*

Vector: *live organism that serves to communicate disease.*

Latent period: *the time interval from infection to development of infectiousness.*

Period of infectiousness: *the time period during which the host can infect other host. The host can become non-infectious either by recovery from infection possibly developing immunity, or death. The host can be an asymptomatic carrier if the status of infectious is not associated with clinical symptoms.*

Incubation period: *the period from infection to the development of symptomatic disease and symptomatic period.*

Symptomatic period: *the time period during which the host show symptoms and signs caused by the infection. The host become an infectious carrier if he/she recovers from symptoms but remains infectious.*

Inapparent case or silent infection: *a successful infection that does not develop detected symptoms. Inapparent cases can be infectious.*

Transmission probability: *is the probability that given contact between an infective source and susceptible host, successfully transfer of the infectious agent will occur so that the susceptible host becomes infected.*

Contact: *the type of interaction (or situation) between a person acting as a source of an infectious agent and a person susceptible when the interaction may lead to transmission of the infectious agent.*

In epidemiology of infectious disease, fundamental concepts are needed to describe the

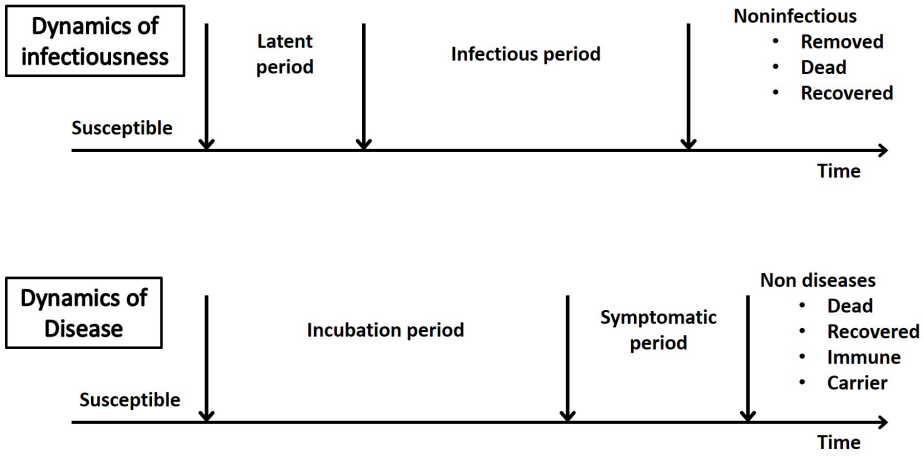

Figure 1. Timeline and dynamic for infection and disease [26]

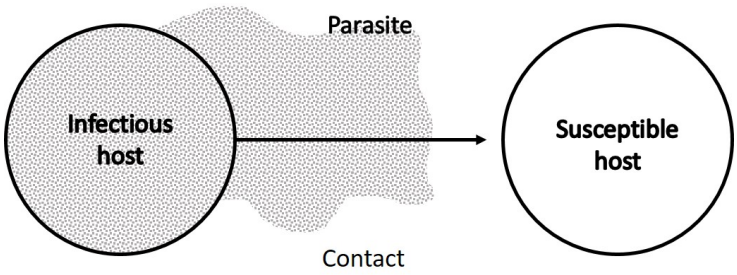

Figure 2. Transmission from an infective to a susceptible host during contact [26]

phenomena resulting from dependence of disease events. Exposure to infection plays a pivotal role because exposure to the infectious agent is necessary for infection and disease to occur. The components of exposure to infectious agent such as the contact of the infective and susceptible hosts as well as the degree and duration of infectiousness need to be taken into account in infectious disease epidemiology. We summarise in Box 1 a few concepts commonly used in infectious disease epidemiology to describe the pathogen/host relationship and the transmission pathways, we followed [8, 26, 29].

The basic paradigm for infection and disease events, following exposure to infectious agents, is represented in Figure 1. Within a population/group, the dependent happening structure of the transmission of infectious diseases is based on infectious agent passing from one host to another, as represented in Figure 2. This structure illustrates that the transmission probability depends on the characteristics (pathogenicity) of the infectious agent, the infectious host, the susceptible host and the contact.

This basic paradigm is not directly applicable to zoonoses, a group of infectious diseases shared between animals – including livestock, wildlife, and pets – caused by bacterial, viral, parasitic or unconventional agents (e.g. prions) that jump from non-human animals to humans. In this case the dependent happening structures of disease events may be complex or very complex with multiple transmission steps at the human-animal-environment interface – where people and animals interact with each other in their shared environment.

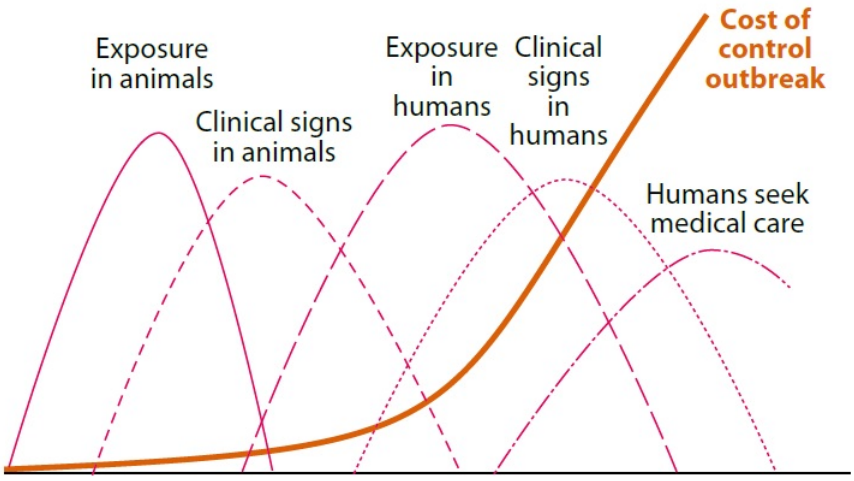

Figure 3. Timeline for zoonoses spread and exponential growth of outbreak cost [24]

For zoonoses, the dependent happening structure is peculiar for each disease, depending on various factors including the nature of the zoonotic agent and its pathogenicity, the number of implicated hosts (reservoir, intermediate and final host), the intensity and the mode through which humans come into contact with the agent (the transmission route) and eventually the possibility that secondary transmission cycle could be established in humans. Basically the spread of zoonotic agents from animal to humans occur through direct contact with animals and their excreta, ingestion of contaminated food and/or water (foodborne/ waterborne), bites of mosquitos, ticks etc. (vectorborne), or indirectly by fomites or environmental contamination. More rarely the transmission may be airborne. The dynamic of infection for zoonoses and timeline considering the whole cycle from animals to humans must therefore take into account the happening structures within the different components implicated in the transmission (Figure 3).

Zoonoses are indeed an important portion of human infectious diseases. The majority of human infectious diseases have animal-origin and 75% of emerging and new diseases are linked to spillover events from animal to humans [25]. In some cases, animals involved in emergence are wildlife, despite the majority are caused by domestic animals interfacing humans. Emerging diseases represent a huge problem for huaman health and, once emerged, can become epidemic and pandemic, as Covid19 taught us, or also endemic and remain in a specific region constantly and sometimes become neglected, meaning that they receive less attention and resources.

One category of diseases that largely affects humans, but generally underestimated, is represented by foodborne diseases. The burden of foodborne zoonoses was estimated from WHO to be of the same order of magnitude of each of the “big three” infectious diseases (malaria, HIV/AIDS and tuberculosis) and air pollution [11]. In high income countries food system represents the main transmission route for infectious disease to humans making this category of diseases a crucial research area.

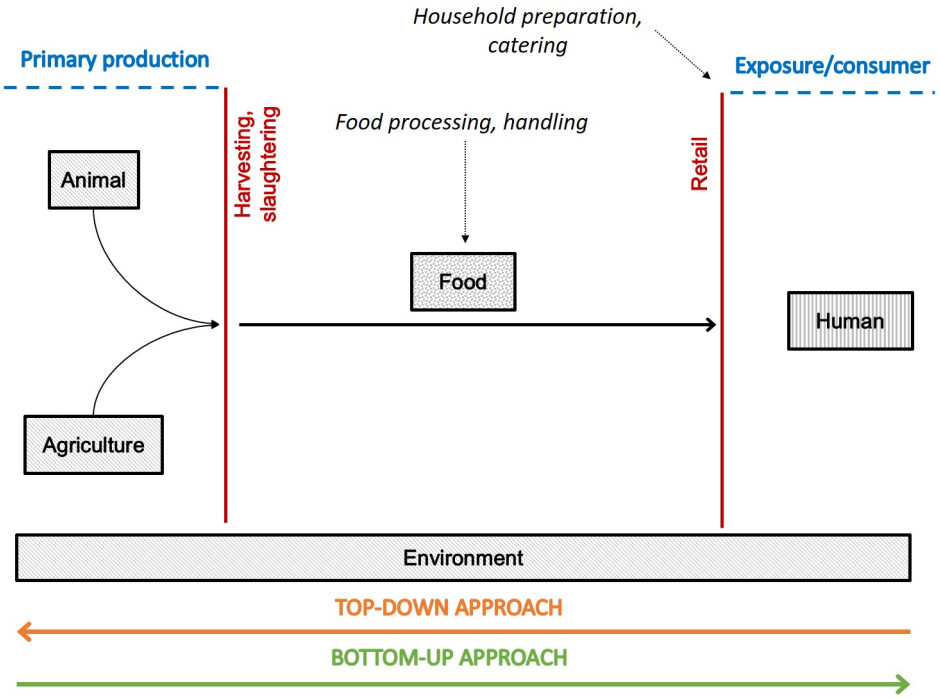

Figure 4. The foodborne transmission paradigm from the pre-harvesting compartment to the exposed consumer

Foodborne zoonoses and food safety

Foodborne zoonoses are an important public health threat. According to EFSA more than 230,000 human cases are confirmed in the EU each year, with a likely high under-reporting factor [3]. This type of diseases is caused by consumption of contaminated food or water. The contamination of food, as shown in Figure 4, may occur either upstream in the food production chain, i.e. at the primary level as a result of direct contamination of products by the infected animals (e.g. eggs, milk) and/or at the time of harvesting (e.g. fecal contamination of milk during the milking process) or downstream, during the manipulation and processing of food in the production chain. Thus, the basic paradigm for foodborne transmission of zoonotic agents to human involves a complex network of compartment in a flow that goes from the primary production in the livestock and agriculture compartment (pre-harvest level), to the final consumer (exposure level), passing through the food-processing and distribution chain (post-harvest level). Infection could be influenced by any of the steps involved and any type of production could have a different effect on the contamination process, given the specific features of the agent but also of produced foodstuff or commodities. Usually the contamination process starts in the primary production compartment through the infection of animals or, via environmental contamination, crops. During this phase the pathogen spreads into the population following, especially in animal populations, the dynamic of a classical infectious disease where transmission happens mainly by direct contact. The post-harvesting segment of the food chain can be a crucial phase for contamination of food. The hygienic condition and type of

food processing have a direct effect on the level of food contamination. Zoonotic bacteria (e.g. *Salmonella*) can growth and multiplicate in food matrices, if the food is manipulated and processed under poor hygienic condition or preserved in inadequate time/temperature condition. Virus do not growth in food but can only survive and viral contamination of food occurred at the pre-harvest level can only decrease in the post-harvest level or even be inactivated. Finally, the consumer can be exposed to the pathogen through food consumption or In the last steps, from retail points to consumer the variables influencing possible human infections are highly connected with the household handling and exposure level of the population (frequency of consumption, habits, average size of servings, etc.).

In late 90s Europe faced several food crises, starting from bovine spongiform encephalopathy (mad cow diseases), to dioxins contamination of food. These events triggered the development of a new food regulations at the EU level and program of scientific advice and communication of the risk in food safety. In particular, the general food law (Regulation 178/20021) conceptualised the new food safety framework and approach to consumer protection with a new paradigm for food safety which was named 'farm to fork'. This paradigm lies on two main crucial principles: i) all food business operators involved in the food production chain (from primary production to the retail level) are responsible for the safety of food consumed by consumers; ii) the adoption of control policies in food safety (risk management) must be supported by scientific evidence obtained through a formal process of risk analysis in food safety. This new conceptual framework based on the segregation of risk assessment, carried out by independent organisms, and risk management (responsibility of the competent authorities) enhanced enormously the implementation of research in the area of food safety and in particular of risk assessment at both the national and EU level. The European Food Safety Authority (EFSA), was founded in 2002 with the task of providing independent scientific evidence to the EU Commission and Member States on food safety issue. The authority was legally established by the EU under the General Food Law – Regulation 178/20021[1].

Several methods for *risk assessment* have been developed to support microbiological food safety and to inform control policies in the framework described above. There are three main possible approaches: "top-down", "bottom-up" or a combination of these two (see Figure 5). The "top-down" approach, or surveillance based, starts from infections and illnesses in humans to gain knowlege on source and transmission pathways and is informed by epidemiological system data. The "bottom-up" approach, conversely, follows the agent along the production chain steps or from a specific point of the production chain (e.g. retail) to the consumer. A combined approach of these two can be used, when information on both sides are available, to gain more information at once. The most classical approach is based on a four-step bottom-up *risk assessment* process that follows the foodborne infection paradigm as shown in Figure 4. The four steps of risk assessment bottom-up are: *hazard identification, exposure assessment, hazard characterization and risk characteriza-*

[1] https://www.efsa.europa.eu/en

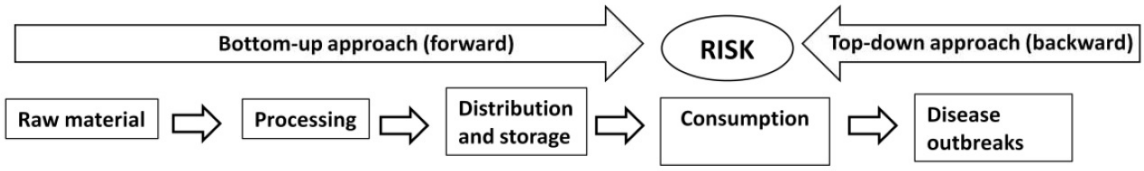

Figure 5. Top-down and bottom-up approached to risk assessment

tion (see Figure 6).

During the first step, the biological or chemical agent (hazard) is identified and specific

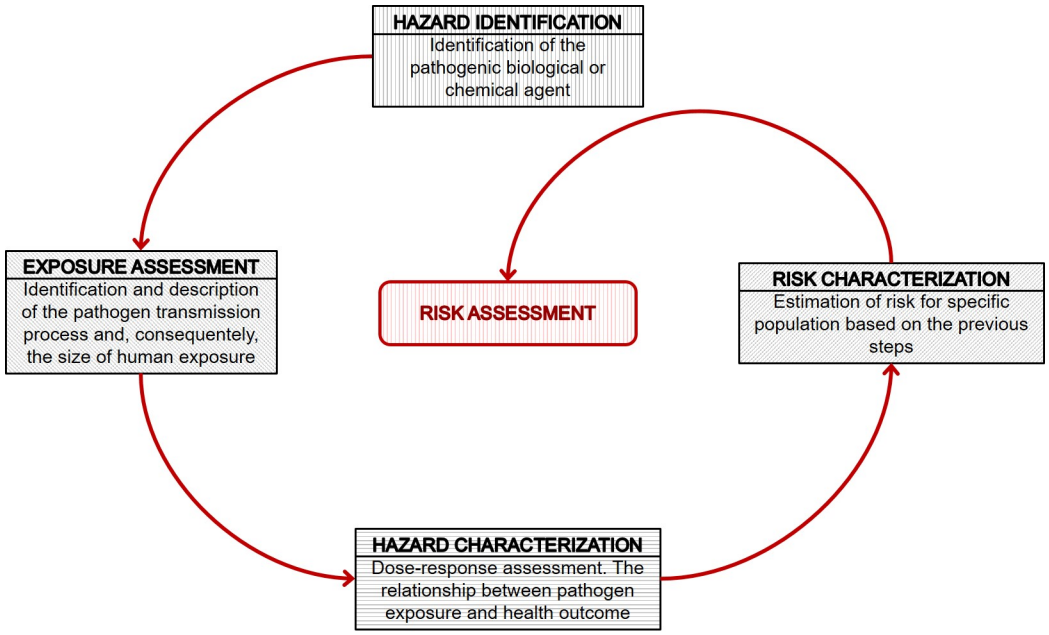

Figure 6. The four-step stricture of risk assessment

epidemiological and microbiological features information are collected. The exposure assessment is the part of the process where the pathogen transmission process is identified and described to allow the evaluation of the exposure of a given population. The hazard characterization, often known also as the dose-response assessment, is the step where the relationship between the pathogen exposure and the outcomes in the population are set. Finally, in the risk evaluation is carried out the actual estimation of the risk for the population chosen. The concept of *risk* is intended as a combination between probability of infection and probability of exposure for the population considered in the study, so it involves all the three previous steps. The tipology of study for this process is highly varied, going from the qualitative to the quantitative studies passing through the semi-quantitative. Qualitative-like risk assessment is used especially in early emergency phases or in absence of data to take fast preliminary actions, here the risk is expressed in terms of qualitative categories (e.g. low/medium/high, acceptable/non acceptable, etc.). See for example [7, 31]. The semi-quantitative approach are less frequent and it is used, for instance, when only partial quantitative data are available or a faster assessment is needed

(see for example [30]). In quantitative microbiological risk assessment studies (QMRA) risk is numerically quantified through quantitative data (e.g. probability of infection given certain conditions, incidence of a disease, etc.). See for example [12, 17, 18, 28]. Each of these methods have advantages and limitations but the golden standard is represented for sure by quantitative methods. Quantitative studies can include different types of techniques, as statistical models, network analysis, Bayesian modeling, deterministic or stochastic mechanistic models, etc.

The "top-down" approach studies starts instead from human cases data to gain knowledge on transmission pathways and sources. Among them are epidemiological methods or microbiological methods. Microbiological methods include a variety of different methods and models, including population genetic model or frequency-matching models that can use frquentist and Bayesian framework or, recently, machine learning techniques. See for example [10, 19, 21–23].

Aim and motivations

To reduce the health impact caused by zoonotic foodborne pathogens in humans, it is necessary to get knowledge about the animal sources and the transmission pathways implicated in human diseases. The availability of this evidence is important to support decision makers (i.e. health competent authorities) adopt and prioritize control policies, in livestock production and in food production chain. For longtime this process has relied just on experts' knowledge and beliefs. In more recent decades, qualitative and semi-quantitative methods progressively replaced such a poor evidence-based approach, in order to overcome limitations of opinion-based approaches and better support the decision-making process. The development of quantitative methods and approaches enabling to deal with the complexity of food-production chain and epidemiological cycles of zoonotic diseases, and incorporate the wealth of available monitoring and surveillance data, gave an important boost to the development of the risk assessment science in food-safety. Nowadays, quantitative methods represent the gold standard in food safety and animal health risk assessment (QMRA) science.

The aim of this project is to develop a framework of quantitative methods to support the risk assessment for two different zoonotic agents: Hepatitis E virus and Shiga toxin-producing *E.coli*. The goal was to produce estimations that help to target intervention strategies in livestock and food production chain. Both agents are considered emergent pathogens in the EU and consumers' protection against risk connected to foodborne exposure to these agents is regulated by the general food safety measures only, meaning that there are no specific dedicated management plans at EU level[2].

Hepatitis E Virus (HEV), is a viral pathogen that emerged very recently as zoonotic agent and spread rapidly in the EU. It is mainly considered a cause of mild or even asymptomatic infection in human population (Hepatitis E) even if severe cases, especially in immunocompromised patients, have been documented. Hepatitis E is subjected to passive reporting in humans, though it is well know that its occurrence is highly under-detected and underreported. Pigs and wild boars are considered the main animal HEV reservoirs. No compulsory monitoring of HEV exist in animal population and food chain. For these reasons, the available data are scarce and several blind spots on important pathogenic mechanisms and epidemiological aspects (i.e. environmental survival of the virus) exist. Despite the lack of scientific knowledge, public health and industry are committed to control HEV along the food production chain in order to protect public health. According to several literature studies, this pathogen is highly diffused in some European countries as France, Spain and Italy.

STEC *Shiga Toxin-producing Escherichia coli*, are a group of pathogenic bacteria of the *Escherichia coli* species. textitShiga Toxin-production Escherichia Coli (STEC). STEC are

[2]Only STEC in seeds and sprouts are subjected to food safety criteria according to Reg. 2073/2005/CE

considered among the priority zoonotic pathogens in the EU[3] because of their potential to cause severe and even fatal illnesses in humans, particularly in children. STEC also occur as large epidemic outbreaks such as the recent epidemic linked to the consumption of contaminated frozen pizza, which in winter 2022 caused tens of cases and deaths in small children. Surveillance of STEC in humans is well established in the EU even if not fully harmonized, with some countries only reporting severe cases and others having a more sensitive surveillance able to capture also mild symptomatic and asymptomatic cases. Tough, the very complex epidemiology of this pathogen and genetic plasticity brings other type of difficulties to the risk assessment analysis.

Approaching these two zoonotic agents means exploring two complementary approaches. In the first case, i.e. HEV, we followed a classical bottom-up risk assessment approach (see Figure 4) based on a from-farm-to-consumer approach, analyzing the two main compartments of the foodborne infection paradigm: the primary production and the exposure through food of the consumer. This approach allow us to built a model "around the existing data" and to characterize the production chain focusing on the main transmission pathways of the pathogen.

In the second case, i.e. STEC, the availability of subtyping characterisation data around the pathogens isolated in different sources and the complexity of transmission cycle of the STEC led us to choose a top-down approach, from the observed human cases, through the food chain up to the primary source of STEC infection.

[3] According to Dir. 99/2003 (EC)

Mathematical tools for epidemiology

Mathematical models are widely used tools in infectious disease epidemiology to investigate epidemics and to improve the capacity building to control the infectious disease spread intro the population. In the last 50 years this mathematical modelling has been largely employed in decision making, vaccination coverage assessment, intervention and so on. In the last decades, mathematical models resulted extremely important in early response to important outbreaks like HIV, Bovine Spongiform Encephalitis (BSE), Food and Mouth Disease (FMD), Severe Acute Respiratory Sindrome SARS and last but non least 2019 SARS-Cov2 pandemic. The stochastic nature and the high nonlinearity of infection dynamics are the two main factors that makes this interdisciplinary cooperation essential to address investigation and intervention in epidemic investigations.

The first recognized mathematical model for epidemiology was developed by Daniel Bernoulli in 1766 to analyse the benefits of smallpox vaccination [6]. The begin of the modern mathematical epidemiology though is referred to the publication of the model by the biochemist William Kermack and the physician Anderson McKendrick in 1927 [15]. This model and its variations are among the most used in this research field.

The Kermack-McKendrick models is also known as SIR model and is one of the basic but also fundamental epidemiological model. The individuals in the population are divided into compartments according to their epidemiological states. In the SIR model the states

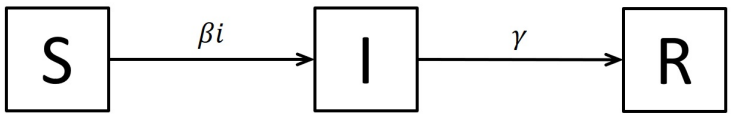

Figure 7. SIR model diagram. The capital letters represent the total number of individual in each epidemiological state, the lowercase letters are the proportion of individuals in the compartment over the entire population N.

are: Susceptibles, Infected and Recovered. Several extension of this basic framework are developed to include other states like Exposed, Asymptomatic and so on. In the SIR model, several assumptions are made: the population is assumed to be closed and homogeneously mixed and all the transitions rates are constant in time. Latent periods and changes in individual behaviour are not taken into account. The models is determined by the transition rates between the classes S, I and R. It is generally assumed that the infection per capita rate is proportional to the prevalence of infected individual in the population, following the law of mass action, and that the recovery rate is constant.
We call S, I and R the total number of individual in each class and $s = \frac{S}{N}$, $i = \frac{I}{N}$, and $r = \frac{R}{N}$, where $N = S + I + R$ is constant. The differential equation system for this specific

model is

$$\begin{cases} \frac{ds}{dt} &= -\beta si \\ \frac{di}{dt} &= \beta si - \gamma i \\ \frac{dr}{dt} &= \gamma i \end{cases} \tag{1}$$

Exahustive tractation of this argument are available in [1, 14, 20].

Stochastic epidemiological models

For many years, deterministic models were the only models used in epidemiological study of infectious diseases. Deterministic models are usually simple to analyse and allow complex and more realistic frameworks and parameterizations. On the other hand, stochastic models represents often a very natural way to describe transmission and spread of a disease. These models allow to quantify uncertainty and to analyse the possibility of extinction of diseases in finite time. In late 40s Bartlett published the first stochastic version of Kermack-McKendrick model [5] beginning the development of stochastic epidemic models field.

We present the standard SIR stochastic model as defined by Ball in 1995 [4] and by Andersson and Britton in 2012 andersson2012stochastic. We have m initial infected individuals and n susceptibles. The infectious periods are independent and identically distributed according to some random variable I with mean i and variance σ^2. Any infected individual has contact with others at the time point of a homogeneous Poisson process with intensity $\frac{\lambda}{n}$. As soon as a susceptible individual comes into contact with an infected individual, the susceptible becames immediately infected. All Poisson processes are independent between them and independent also of I.

Selke construction

A very interesting construction of the standard SIR stochastic model is due to Selke [27]. Any susceptible individual is associated with an individual threshold and is exposed to a total "infection pressure". As soon as the pressure reaches the threshold, the susceptible individual became infected.

The m initially infectives individuals are labelled as $-(m-1), -(m-2), \ldots, 0$ and n initial susceptibles with label $1, 2, \ldots, n$. Individuals have independent identically distributed infectious periods $I_{-(m-1)}, I_{-(m-2)}, \ldots, I_n$. Furthermore, let $Q_1, Q_2, \ldots, Q_n$ be the individual thresholds of the susceptible individuals and independent and identically distributed $Q_i \sim \exp(1)$. If we denote with $Y(t)$ the total infective individuals at time t, we can define the totale infection pressure as

$$A(t) = \frac{\lambda}{n} \int_0^t Y(u)du. \tag{2}$$

For $i = 1, 2, \ldots, n$ the susceptible labelled i becomes infected when $A(t)$ reaches Q_i and remains infectious for a time I_i and then is removed.

The Markovian case

In the Markovian version of the model we have $X(t)$ susceptible and $Y(t)$ infected individuals at time t and, of course, the $\Big(X, Y\Big) = \Big\{\Big(X(t), Y(t)\Big), t \geq 0\Big\}$ is a Markov process. If we suppose that I is exponentially distributed with intensity γ, we have the following transition's rate

$$\begin{aligned} (i,j) &\to (i-1, j+1) \quad \frac{\lambda}{n} ij \\ (i,j) &\to (i, j-1) \quad \gamma j \end{aligned} \tag{3}$$

Inference likelihood-based for Partially Observed Markov Processes (POMP)

Parameter estimation is a crucial step in model implementation. One of the most used approach is the likelihood-based inference. Here we present a useful method to implement likelihood-based inference for epidemiological models in a frequentist environment. We report the methodology as presented in the official documentation of R pomp package[4].

Suppose we have $y_1^*, \ldots, y_N^*$ noisy, incomplete and indirect observations of a Markov process $X(t)$ at time $t_1 < t_2 < \cdots < t_N$. Let's define Y_n the random variable modeling the observation or measurement process at time t_n and $X_n \coloneqq X(t_n)$. If we are able to describe the one-step transition density $f_{X_n|X_{n-1}}(x_n|x_{n-1};\theta)$, the measurement density $f_{Y_n|X_n}(y_n|x_n;\theta)$ and an initial density for the Markov process $f_{X_0}(x_0;\theta)$ we can specify the entire POMP model. In the following sections we will define a formal POMP model and its likelihood function and we will introduce a useful factorization for the likelihood and an iterative algorithm that allows to maximize the likelihood.

POMP models

Let's $\theta \in \mathbb{R}^{d_\theta}$ be a parameter vector and $\big(X(t;\theta)\big)_{t\in T}$ a Markov process that takes values in $\mathbb{R}^{d_x}$, with $T \subseteq \mathbb{R}$.
Now, suppose we are able to observe the process at given times $\{t_i \in T,\, i = 1, \ldots, N\}$, with $t_0 \in T$ the initial time and assume we have $t_0 \leq t_1 \leq \cdots \leq t_N$. We say that we have a partially observed Markov process if $X_{0:N} \coloneqq (X_0, \ldots, X_N)$ are observed only by way of another process $Y_{1:N} \coloneqq (Y_1, \ldots, Y_N)$ with values in $\mathbb{R}^{d_y}$.
Given that X_n is Markovian and Y_n depends only on the process at t_n time, if we assume that $Y_{1:N}$ are conditionally independent given $X_{0:N}$ we can write the joint density of the two processes as follow:

$$f_{X_{0:N},Y_{1:N}}(x_{0:N}, y_{1:N};\theta) = f_{X_0}(x_0;\theta) \prod_{n=1}^{N} f_{X_n|X_{n-1}}(x_n|x_{n-1};\theta) f_{Y_n|X_n}(y_n|x_n;\theta).$$

Hence, the marginal density for $Y_{1:N}$ evaluated at the observation data $y_1^*, \ldots, y_N^*$ is the likehood function for the model result to be

$$\mathcal{L}(\theta) = f_{Y_{1:N}}(y_{1:N}^*;\theta) = \int f_{X_{0:N},Y_{1:N}}(x_{0:N}, y_{0:N};\theta) dx_{0:N}. \tag{4}$$

[4] https://kingaa.github.io/pomp/

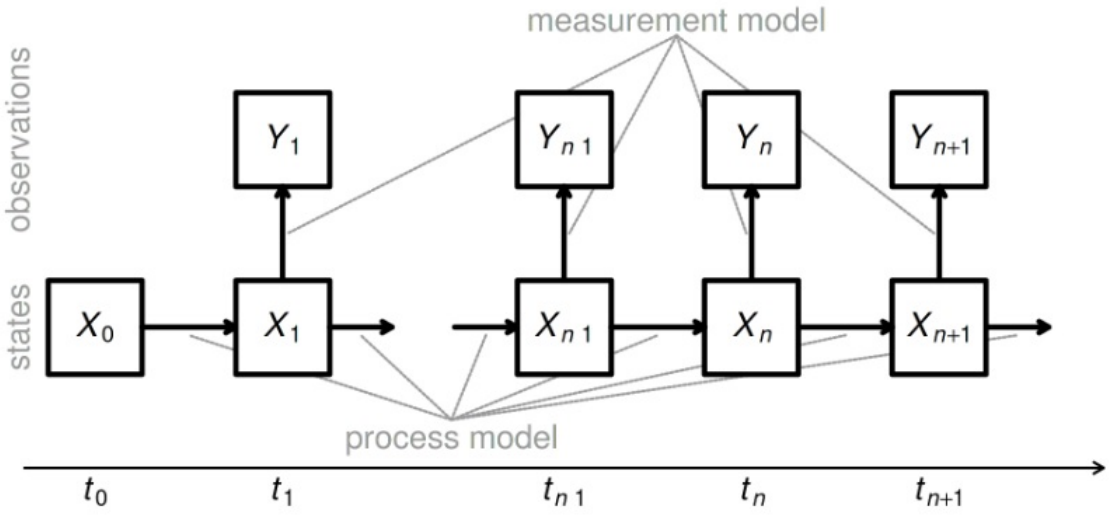

Figure 8. Schematic representation of a pomp model

Sometimes is more practicle to use the log transformed version of the likelihood, meaning the log-likelihood

$$\ell(\theta) = \log(\mathcal{L}(\theta)).$$

Now, we can rewrite the equation 4

$$\mathcal{L}(\theta) = \int \left\{ \prod_{n=1}^{N} f_{Y_n|X_n}(y_n^*|X_n;\theta) \right\} f_{X_{0:N}}(x_{0:N};\theta) dx_{0:N}$$

and it can therefore be written as an expected value

$$\mathcal{L}(\theta) = \mathbb{E}\left[\prod_{n=1}^{N} f_{Y_n|X_n}(y_n^*|X_n;\theta) \right] \tag{5}$$

Hence, thanks to the Law of Large Numbers (LLN) we could approximate likelihood with a Monte Carlo sampling as

$$\mathcal{L}(\theta) \approx \frac{1}{J} \sum_{j=1}^{J} \prod_{n=1}^{N} f_{Y_n|X_n}(y_n^*|X_n^j;\theta), \tag{6}$$

where X_n^j are sampled from $f_{X_{0:N}}(x_{0:N};\theta)$, for $j = 1, \ldots, J$.
Unfortunately, this classical approach is most of the time not reliable, especially if time-series data is very long. We present an approach where a different factorization of the likelihood is used, named the particle filtering [2, 9, 16].

Sequential Monte Carlo

The particle filtering technique is based on a specific factorization of the likelihood function, we begin writing

$$\begin{aligned}\mathcal{L}(\theta) = f_{Y_{1:N}}(y^*_{1:N};\theta) &= \prod_{n=1}^{N} f_{Y_n|Y_{1:n-1}}(y^*_n|y^*_{1:n-1};\theta) \\ &= \prod_{n=1}^{N} \int f_{Y_n|X_n}(y^*_n|X_n;\theta) f_{X_n|Y_{1:n-1}}(x_n|y^*_{1:n-1};\theta) dx_n,\end{aligned} \tag{7}$$

defining $f_{X_1|Y_{1:0}} = f_{X_1}$.
We can now use the Markov property and Bayes' theorem to obtain a recursive expression for $f_{X_n|Y_{1:n-1}}(x_n|y^*_{1:n-1};\theta)$:

$$\begin{aligned}&f_{X_n|Y_{1:n-1}}(x_n|y^*_{1:n-1};\theta) = \\ &\qquad = \int f_{X_n|X_{n-1}}(x_n|x_{n-1};\theta) f_{X_{n-1}|Y_{1:n-1}}(x_{n-1}|y^*_{1:n-1};\theta) dx_{n-1}.\end{aligned} \tag{8}$$

Bayes' theorem allows to rewrite $f_{X_{n-1}|Y_{1:n-1}}(x_{n-1}|y^*_{1:n-1};\theta)$ as follow:

$$\begin{aligned}&f_{X_n|Y_{1:n}}(x_n|y^*_{1:n};\theta) = \\ &\qquad = f_{X_n|Y_n,Y_{n-1}}(x_n|y^*_n, y^*_{1:n-1};\theta) \\ &\qquad = \frac{f_{Y_n|X_n}(y^*_n|x_n;\theta) f_{X_n|Y_{1:n-1}}(x_n|y^*_{1:n-1};\theta)}{\int f_{Y_n|X_n}(y^*_n|u_n;\theta) f_{X_n|Y_{1:n-1}}(u_n|y^*_{1:n-1};\theta) du_n}\end{aligned} \tag{9}$$

The two distributions $f_{X_n|Y_{1:n-1}}(x_n|y^*_{1:n-1};\theta)$ and $f_{X_n|Y_{1:n}}(x_n|y^*_{1:n};\theta)$ are called prediction and filtering distribution, respectively, and allow to produce a two step recursion to estimate the likelihood for each time step of the data time serie. The steps of a sequential Monte Carlo or particle filtering are the following:

1. We have J sample drawn from the filtering distribution at time t_{n-1} $X^F_{n-1,j}$, for $j = 1, \ldots, J$.

2. We draw a sample from the prediction model, meaning simulating the process model

$$X^P_{n,j} \sim process(X^F_{n-1,j};\theta), \quad j = 1, \ldots, J$$

3. We obtain a sample from the filtering distribution at time t_n resampling from $\{X^P_{n,j}, j = 1, \ldots, J\}$ with weights

$$w_{n,j} = f_{Y_n|X_n}(y^*_n|X^P_{n,j};\theta)$$

4. Given that $X^P_{n,j}$ is a drawn from $f_{X_n|Y_{1:n-1}}(x_n|Y^*_{1:n-1};\theta)$ can approximate the conditional likelihood as

$$\hat{\mathcal{L}}_n(\theta) \approx \frac{1}{J} \sum_j f_{Y_n|X_n}(y^*_n|X^P_{n,j};\theta)$$

We can iterate the procedure for all the N time steps and approximate the full likelihood

$$\ell(\theta) = \log \mathcal{L}(\theta) = \sum_n \log \mathcal{L}_n(\theta) \approx \sum_n \hat{\mathcal{L}}_n(\theta)$$

Iterated filtering algorithm

Iterated filtering is a procedure to maximize likelihood. The algorithm build and presented by Ionides and collegues [13] consists of an iteration of particle filtering procedure where the parameter vector does a random walk for each particle as the process had non constant parameters. After each iteration, the output parameter swarm is used as input in the next iteration. The variance of the random-walk decreases at each iteration going to zero. It is possible to prove that, this way, the algorithm converges to an area of the parameter space that maximizes likelihood. We report here the pseudocode as presented in [13].

input:
Simulator for $f_{X_0}(x_0;\theta)$
Simulator for $f_{X_n|X_{n-1}}(x_n|x_{n-1};\theta)$, $n = 1,\dots,N$
Evaluator for $f_{Y_n|X_n}(y_n|x_n;\theta)$, $n = 1,\dots,N$
Data $y_1^*,\dots,y_N^*$
algorithmic input:
Number of iterations, M
Number of particles, J
Initial parameter swarm, $\{\Theta_j^0,\ j = 1,\dots,J\}$
Perturbation density, $h_n(\theta|\phi;\sigma)$, $n = 1,\dots,N$
Perturbation sequence, $(\sigma_1,\dots,\sigma_M)$
output: Final parameter swarm, $\{\Theta_j^M,\ j = 1,\dots,J\}$

Algorithm (IF2 algorithm). *For* m *in* $1:M$

$\Theta_{0,j}^{F,m} \sim h_0(\theta|\Theta_j^{m-1};\sigma_m)$, *for* $j = 1,\dots,J\}$
$X_{0,j}^{F,m} \sim f_{X_0}(x_0;\Theta_{0,j}^{F,m})$, *for* $j = 1,\dots,J\}$
For n *in* $1:N$
$\Theta_{n,j}^{P,m} \sim h_n\left(\theta \mid \Theta_{n-1,j}^{F,m},\sigma_m\right)$, *for* $j = 1,\dots,J\}$
$X_{n,j}^{P,m} \sim f_{X_n|X_{n-1}}\left(x_n \mid X_{n-1,j}^{F,m};\Theta_j^{P,m}\right)$, *for* $j = 1,\dots,J\}$
$w_{n,j}^m = f_{Y_n|X_n}\left(y_n^* \mid X_{n,j}^{P,m};\Theta_{n,j}^{P,m}\right)$, *for* $j = 1,\dots,J\}$
Draw $k_{1,j}$ *with* $\mathbb{P}(k_j = i) = w_{n,i}^m / \sum_{u=1}^J w_{n,u}^m$
$\Theta_{n,j}^{F,m} = \Theta_{n,k_j}^{P,m}$ *and* $X_{n,j}^{F,m} = X_{n,k_j}^{P,m}$, *for* $j = 1,\dots,J\}$
End For
Set $\Theta_j^m = \Theta_{N,j}^{F,m}$ *for* $j = 1,\dots,J\}$

End For

We give in input simulator for initial state of the Markov chain and for the transition, then we need an evaluator for the measaurement process. Θ_j^0 is the swarm of initial parameters. h_n is the perturbation density, that will be Gaussian in this case. The outer loop

over m is for the number of iteration we want to perform of the algorithm, the inner loop is the particle filtering over the entire time series of dat $y_1^*, \ldots, y_N^*$.

References

[1] R. M. Anderson and R. M. May. *Infectious diseases of humans: dynamics and control*. Oxford university press, 1992.

[2] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *IEEE Transactions on signal processing*, 50(2):174–188, 2002.

[3] E. F. S. Authority, E. C. for Disease Prevention, and Control. The european union one health 2019 zoonoses report. *EFSA Journal*, 19(2):e06406, 2021. https://efsa.onlinelibrary.wiley.com/doi/10.2903/j.efsa.2021.6406.

[4] F. Ball and P. Donnelly. Strong approximations for epidemic models. *Stochastic processes and their applications*, 55(1):1–21, 1995.

[5] M. Bartlett. Some evolutionary stochastic processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 11(2):211–229, 1949.

[6] S. Blower and D. Bernoulli. An attempt at a new analysis of the mortality caused by smallpox and of the advantages of inoculation to prevent it. 1766. *Reviews in medical virology*, 14(5):275–288, 2004.

[7] E. de Freitas Costa, M. Cardoso, J. D. Kich, and L. G. Corbellini. A qualitative risk assessment approach to microbial foodborne hazards in brazilian intensive pork production: A step towards risk prioritization. *Microbial Risk Analysis*, 15:100105, 2020.

[8] A. Dictionary. The american heritage medical dictionary. *Med. Dict*, 32:909, 2007.

[9] A. Doucet, N. De Freitas, N. J. Gordon, et al. *Sequential Monte Carlo methods in practice*, volume 1. Springer, 2001.

[10] T. Hald, D. Vose, H. C. Wegener, and T. Koupeev. A bayesian approach to quantify the contribution of animal-food sources to human salmonellosis. *Risk Analysis: an International Journal*, 24(1):255–269, 2004.

[11] A. H. Havelaar, M. D. Kirk, P. R. Torgerson, H. J. Gibb, T. Hald, R. J. Lake, N. Praet, D. C. Bellinger, N. R. de Silva, N. Gargouri, N. Speybroeck, A. Cawthorne, C. Mathers, C. Stein, F. J. Angulo, B. Devleesschauwer, and on behalf of World Health Organization Foodborne Disease Burden Epidemiology Reference Group. World health

organization global estimates and regional comparisons of the burden of foodborne disease in 2010. *PLOS Medicine*, 12(12):1–23, 12 2015.

[12] A. A. Hill, R. R. Simons, L. Kelly, and E. L. Snary. A farm transmission model for salmonella in pigs, applicable to eu member states. *Risk Analysis*, 36(3):461–481, 2016.

[13] E. L. Ionides, D. Nguyen, Y. Atchadé, S. Stoev, and A. A. King. Inference for dynamic and latent variable models via iterated, perturbed bayes maps. *Proceedings of the National Academy of Sciences*, 112(3):719–724, 2015.

[14] M. J. Keeling and P. Rohani. *Modeling infectious diseases in humans and animals*. Princeton university press, 2011.

[15] W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115(772):700–721, 1927.

[16] G. Kitagawa. Non-gaussian state—space modeling of nonstationary time series. *Journal of the American statistical association*, 82(400):1032–1041, 1987.

[17] C. Lanzas, Z. Lu, and Y. T. Gröhn. Mathematical modeling of the transmission and control of foodborne pathogens and antimicrobial resistance at preharvest. *Foodborne pathogens and disease*, 8(1):1–10, 2011.

[18] C. Lanzas, L. D. Warnick, K. L. James, E. M. Wright, M. Wiedmann, and Y. T. Gröhn. Transmission dynamics of a multidrug-resistant salmonella typhimurium outbreak in a dairy farm. *Foodborne pathogens and disease*, 7(4):467–474, 2010.

[19] N. Lupolova, S. J. Lycett, and D. L. Gally. A guide to machine learning for bacterial host attribution using genome sequence data. *Microbial genomics*, 5(12), 2019.

[20] C. Mascia, A. Terracina, and E. Montefusco. *BioMat 1.0.* LaDotta, 2018.

[21] L. Mughini-Gras, P. Kooh, P. Fravalo, J.-C. Augustin, L. Guillier, J. David, A. Thébault, F. Carlin, A. Leclercq, N. Jourdan-Da-Silva, et al. Critical orientation in the jungle of currently available methods and types of data for source attribution of foodborne diseases. *Frontiers in microbiology*, 10:2578, 2019.

[22] P. Mullner, G. Jones, A. Noble, S. E. Spencer, S. Hathaway, and N. P. French. Source attribution of food-borne zoonoses in new zealand: A modified hald model. *Risk Analysis: An International Journal*, 29(7):970–984, 2009.

[23] N. Munck, P. M. K. Njage, P. Leekitcharoenphon, E. Litrup, and T. Hald. Application of whole-genome sequences and machine learning in source attribution of salmonella typhimurium. *Risk Analysis*, 40(9):1693–1705, 2020.

[24] D. G. Randolph et al. People, pathogens and our planet : The economics of one health. Technical report, World Bank, 2012. http://hdl.handle.net/10986/11892.

[25] D. G. Randolph et al. Preventing the next pandemic. zoonotic diseases and how to break the chain of transmission. Technical report, United Nations Environment Programme (UNEP), 2020. https://www.unep.org/resources/report/preventing-future-zoonotic-disease-outbreaks-protecting-environment-animals-and.

[26] K. J. Rothman, S. Greenland, T. L. Lash, et al. *Modern epidemiology*, volume 3. Wolters Kluwer Health/Lippincott Williams & Wilkins Philadelphia, 2008.

[27] T. Sellke. On the asymptotic distribution of the size of a stochastic epidemic. *Journal of Applied Probability*, 20(2):390–394, 1983.

[28] E. L. Snary, A. N. Swart, R. R. Simons, A. R. C. Domingues, H. Vigre, E. G. Evers, T. Hald, and A. A. Hill. A quantitative microbiological risk assessment for salmonella in pigs for the european union. *Risk Analysis*, 36(3):437–449, 2016.

[29] S. Straif-Bourgeois and R. Ratard. Infectious disease epidemiology. In *Handbook of Epidemiology*, pages 1327–1362. Springer, 2005.

[30] J. Sumner and T. Ross. A semi-quantitative seafood safety risk assessment. *International Journal of Food Microbiology*, 77(1-2):55–59, 2002.

[31] K. Walia, A. Kapoor, and J. Farber. Qualitative risk assessment of cricket powder to be used to treat undernutrition in infants and children in cambodia. *Food Control*, 92:169–182, 2018.

[32] W. H. O. (WHO). Mortality and global health estimates 2013, 2018. https://apps.who.int/gho/data/node.home.

Parte I

Quantitative bottom-up approach for hepatitis E in humans

Capitolo 1

Hepatitis E. Background and epidemiology

Background

Hepatitis E virus (HEV) is the causative agent of Hepatitis E, an emerging disease of worldwide occurrence causing 20 million human infection events yearly [87]. For many decades Hepatitis E was considered in Europe and in the USA a health problem limited to travellers coming back from area where this pathogen was endemic [50,56,69] but since the early 90s autochthonous cases have been increasingly reported. In Europe and other high-income countries Hepatitis E is considered a foodborne zoonosis causing mainly sporadic cases [56, 58, 69]. Clusters and small outbreaks of HEV infection were also occasionally reported in the European Union (EU) [33, 35, 66]. In the EU, however, surveillance of HEV infection is sparse and not harmonised hampering the possibility to adequately characterise the epidemiology of Hepatitis E, including the accurate identification of the food items implicated in the transmission of HEV to humans.

HEV is a RNA virus belonging to the species Orthohepevirus of the family of Hepeviridae, and was first described in 1978. This virus has a wide range of hosts including humans, domestic and wild animals (domestic swine, wild boar, deer, rabbit, mongoose, ferret, rat and chicken to bat and cutthroat trout). The Orthohepevirus A species includes seven genotypes (HEV1–HEV7) but only four HEV genotypes (HEV1, 2, 3, 4) are reported in humans. HEV-1 and HEV-2 affect primarily humans and are endemic in Asia, Africa and Latin America where large outbreaks have been reported [55]. HEV3 and HEV4 are more frequently reported in high-income countries and have a wider host range including both animals and humans (Figure 1.1). These genotypes include most of the zoonotic HEV variants which usually cause sporadic and chronic diseases in humans in high-income countries and also small outbreaks. HEV 3 and HEV4 affect a broad range of species but pigs and wild boars are considered the main HEV animal reservoirs.

Hepatitis E Virus (HEV) infection in humans

Infection with the HEV causes hepatitis E, which usually presents as an acute, self-

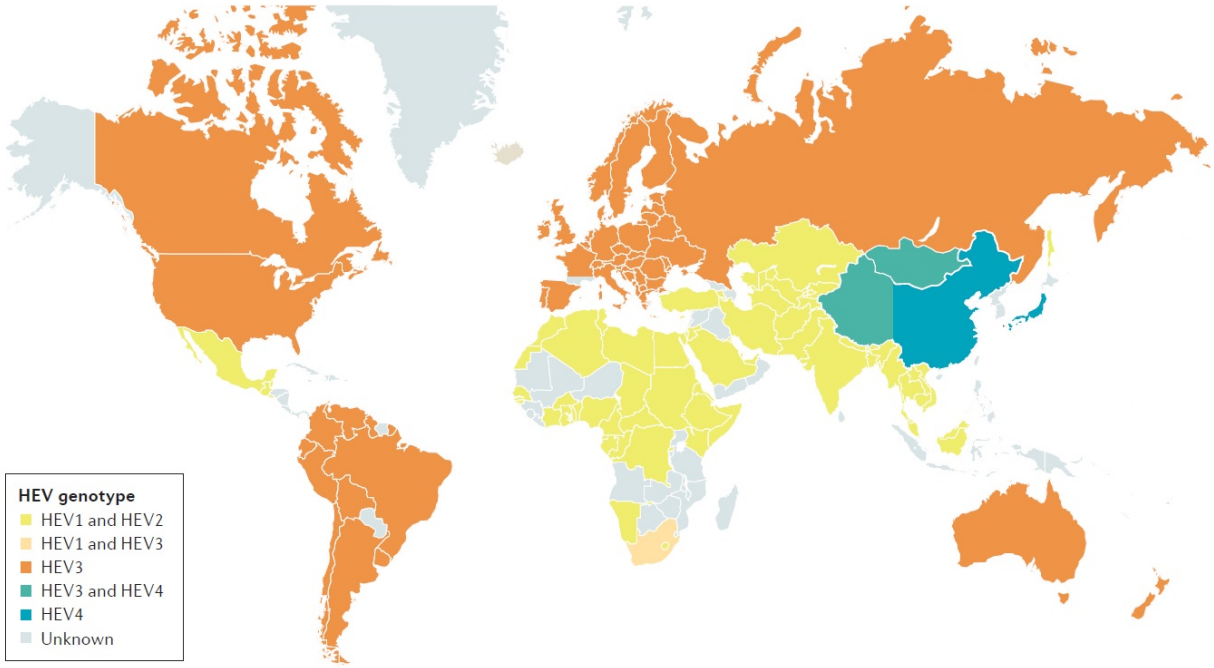

Figure 1.1. Worldwide presence of HEV genotypes. Schematic representation of the distribution of the different hepatitis E virus (HEV) genotypes. (Source [40])

limiting liver inflammation. Clinical signs include fever, anorexia, and jaundice. Extrahepatic manifestations and serious sequelae including chronic condition leading to cirrhosis, liver failure and death may occur especially in immunocompromised patients and in presence of comorbidity [39, 43, 64, 85, 86]. A major risk for chronic and fulminant hepatitis E is reported for pregnant women, with possibility of abortion or infant mortality [34].

The clinical course of Hepatitis E in humans is frequently asymptomatic especially in young adults and children. In Europe, the prevalence of hepatitis E differs between regions. Hepatitis E is hyperendemic in southwest France, with seroprevalence rates in population of >50%, it is endemic in northern France, the United Kingdom, Belgium, the Netherlands, Luxembourg and Germany where 10–30% of individuals have serological evidence of previous HEV exposure. However, the seroprevalence is only 2% in children aged 2–4 years living in southwest France despite the fact that this is a hyperendemic area [40].

In industrialised countries, the majority of human cases are due to zoonotic transmission and are attributable to the consumption of pork and wild boar meat and products thereof [30, 43, 78]. Contamination of surface and coastal waters can lead to a possible risk of HEV infection due to the consumption of HEV contaminated foodstuff of non-animal origin, as vegetables or shellfish [8, 24, 42, 47, 67]. An outline of the dependent happening structure for HEV is shown in Figure 1.2. Iatrogenic transmission of HEV through infected blood and blood products has also been documented. However, in England, transfusion-transmitted HEV infection was estimated to account for <1% of all human infections with HEV. Most iatrogenic transmissions remain asymptomatic [40].

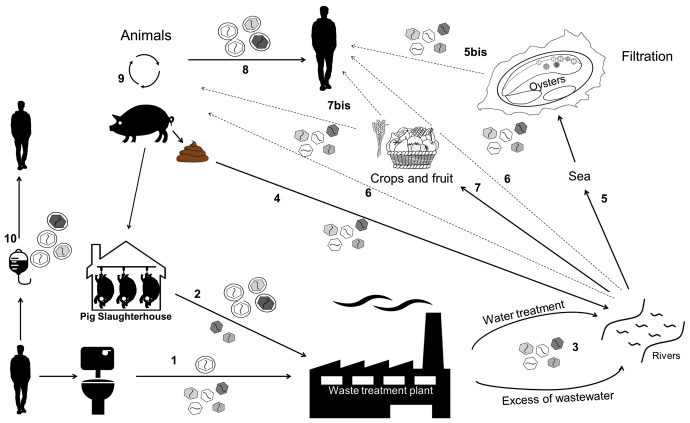

Figure 1.2. Dependent happening structure for HEV [29]

HEV in domestic animals and foods

HEV is highly prevalent in pig herds [39] where transmission occur primarily through the fecal-oral route (i.e. ingestion of contaminated food), direct contact with infected animals or contaminated water [1, 7]. Indeed, anti-HEV antibodies were detected in 46%–100% of swine farms from many countries. Clinical manifestations in animals are usually not detectable with microscopic liver lesions [50]. Pigs inoculated intravenously with swine HEV developed viremia prior to seroconversion, had histological evidence of hepatitis, but did not display clinical symptoms [25]. Infection in pigs has been proof to be highly age-dependent, with a peak of viral fecal shedding between 10 and 16 weeks of age [11, 49].

Similarly to any other virus, HEV requires the host cellular machinery to replicate meaning that outside the host, the virus can only survive. HEV survival in the environment or in food matrices is highly dependent on the time outside the host and on the environmental conditions [5, 88]. Hence, food handling and preparation are important factors to prevent a possible infection. Cooking at low-medium temperatures (56°) resulted insufficient to inactivate the virus [28], while Barnaud et al. [5] showed that HEV inactivation in food matrices requires at least 20 minutes cooking at 20°.

Unfortunately, HEV cannot be routinely cultivated on cells making really challenging to perform infectivity assay on the agent [55]. For this same reason, it is difficult to evaluate the potential infectivity of the virus when it is detected in any matrix, including food and water or to assess the efficacy of inactivation treatments, like cooking [5].

Epidemiology of HEV infections in Italy

Between 2007 and 2019, nearly 17,000 possible cases of acute Hepatitis E cases have been reported to the Italian national surveillance system (SEIEVA) [82] of which only 385 cases were confirmed. The 72.5% of these confirmed cases were autochthonous, meaning that the infection was acquired in Italy and was non-travel related. The human epidemio-

logy of the Hepatitis E along the territory is highly nonhomogeneous, with higher number of cases reported from the regions in central Italy, in particular Abruzzo, Lazio, The Marche, Umbria, and in Sardinia. A survey among blood donors conducted by Spada et al. indicated mean seroprevalence the 8.7% at the national level with a range among provinces from the 0% of Barletta-Andria-Trani province in Puglia to the 38.5% of Nuoro province in Sardinia. While the difference resulted non-statistically significant for the number of cases reported to the SEIEVA system [82], Spada et al. [75] reported that the prevalence of serum antibodies against HEV in the population >18 years living in these areas were statistically significantly higher compared with the estimated average HEV-antibodies seroprevalence at national level. Since the seroprevalence rate of anti-HEV antibodies can be used as a marker for present and past infections (IgM and IgG are markers of an acute or past infection, respectively), this finding suggests that many cases of Hepatitis E remains undetected in the population and corroborates the hypothesis that the population exposure to HEV differs importantly among regions mainly due to different food consumption habits.

Evidences from several studies indicate that the consumption of food containing raw or undercooked pork and wild boar meat are important risk factors [2,78,82] also in Italy. Alfonsi and collegues [2] estimated a 4.6 adjusted odds ratio for pork consumption as risk factor and 2.9 for undercooked sausages consumption. Similar results are reported in [82], where cured pork meat has a odds ration of 2 and undercooked or raw pork meat 3.1. Some studies also reported shellfish as vehicles potentially implicated in foodborne transmission of HEV [2,9,82]. The odds ratio are smaller for this foodstuff with 0.9 reported in [82]. Evidence is also available of a higher risk for professionally-exposed workers such as veterinarians, farmers and hunters [10,15,52,54].

Circulation of HEV in farmed pigs in Italy is widely documented [14,22,51,59,60]. Positivity of blood samples ranges from about 56% [57] to 93% found by Ponterio and collegues [60]. Fecal positivity is obviously lower with a range between 7.4% [14] and 42% [22]. Furthermore, Pavia et al. [57] reported a significative difference in seropositivity in smaller farms (<100 pigs) compared to other sizes (p-value < 0.001) and in fattening farms compared to farrow-to-finish (p-value=0.002).

In the food chain, HEV has been detected in pork foods such as dry and fresh sausages at retail level [12,20,21], but also in shellfish sampled in the production areas at retail [45,76], or in biomonitoring points [26]. Observational studies documented also HEV contamination of vegetables and fruits in Italy [74,80]. HEV RNA was found in sewage and surface water samples suggesting possible environmental contamination via recycled water [38,44,45].

The main goal of the study is to highlight the factors involved in the foodborne transmittion of hepatitits E to humans to help reduce circulation of the pathogen with suitable and effective interventions on both the primary production and/or the consumer level. To

do so, we worked to build a comprehensive model to understand the transmission dynamics along the whole production chain allowing to plan intervention and to exploit all the available data.

In chapter 2 we reported the results of the exposure study where we used original data at retail point to estimate the number of new infected people in a year over the Italian population, given the consumption of specific foodstuff. In chapter 3 we repotred the pre-harvest model, where a transmission model on farm has been developed to understand the transmission dynamics among pigs based on the different farming types present on the Italian territory.

Capitolo 2

Post-harvest model: the consumer level

In this part of the study, we developed a mathematical model to rank the importance of various types of food potentially implicated in the transmission of HEV to humans in the adult Italian population (around 50,000,000). The food categories considered in our study are pork products with liver (PL), pork products without liver (PNL), bivalve shellfish (SH), green leafy vegetables (GV), and raw milk (RM). These results were published here [53].

2.1 Model description

In order to obtain the ranking of food items most frequently implicated in HEV transmission in Italy, we developed a parametric stochastic model to estimate the expected number of newly infected persons who develop HEV infection in the Italian population ($\geq$18 years) through the consumption of the different foods, in one year period.

The analyses were carried out with the R software version 3. 6. 0 ([61]). For the heviest calculation the Gauss Cluster at the Turing Lab of Mathematics Department "Guido Castelnuovo" of Rome "La Sapienza" was used (http://centrocalcolo.mat.uniroma1.it, http://turinglab.mat.uniroma1.it).

2.1.1 The mathematical model

We modelled the individual infectious dose distribution S to HEV using a proxy of the infectious dose based on data available in the literature. We thus build the distribution C_i of HEV concentration in a food serving of category i, based on data on prevalence of HEV contamination of food at retail obtained from a recent sampling study in Italy. Using these two quantities we estimated the probability q_i for a single person to develop a HEV infection after the consumption of a single serving of food belonging to category i. The average number of portions of each food consumed per year per person and

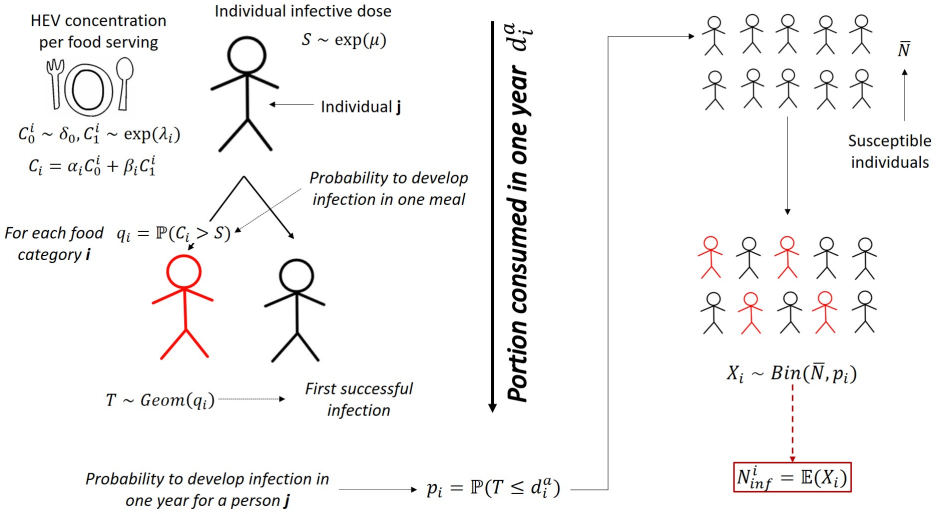

Figure 2.1. The model sketch

over the number of susceptible individuals in the Italian population (Figure 2.1) were then summed up to estimate the average number of newly infected cases in the Italian population in one year. All the parameters used in the model are summarised in Table 2.1.

The model framework is showed in Figure 2.1.

We build the distribution of concentration per serving C_i for each food category using HEV load data in food expressed in genome equivalent per gram of food (g.e./gram) (see section 2.1.2), and defining a mean serving size $serv_i$ (grams) for each food category, according to guidelines of Council for Agricultural Research and Analysis of the Agricultural Economy [48]. Multiplying these two quantities we obtained the distribution of the total genome equivalent HEV per serving (g.e./serving). The distribution was built as a mixture of random variables as follow

$$C_i = \alpha_i C_0^i + \beta_i C_1^i \quad \text{where } C_0^i \sim \delta_0, \quad C_1^i \sim exp(\lambda_i)$$

C_0^i is a Dirac delta with point mass in zero and represents the HEV negative food samples belonging to category i. The random variable C_1^i models the distribution of the viral concentration in the HEV positive samples belonging to category i and it is distributed according to an exponential distribution. The rate of the exponential distribution is the inverse of the mean viral concentration λ_i^{-1} of HEV positive samples that we estimated using the Maximum Likelihood Estimation (MLE). The weights α_i and β_i are the fraction of HEV negative and positive food samples, respectively.

To model the individual infectious dose distribution S we used data from outbreaks for which the HEV load in implicated food (g.e./gr) was documented [63, 65, 66, 79]. We fitted S on the estimated foodborne HEV intake (g.e.) of cases involved these outbreaks. The individual intake of HEV (g.e.) was estimated based on the viral concentration in the implicated food (HEV g.e./gr) for a mean serving size (gr) of the implicated food. The mean serving size ($serv_0$) was estimated according with the same data source used

Tabella 2.1: Model parameters

Parameter	Description
$i \in$ {PL, PNL, SH, GV, RM}	food category as listed in the Introduction
$\alpha_i = 1 - \beta_i$	contamination probability for food category i
$C_i \sim \alpha_i \exp(\lambda_i) + \beta_i \delta_0$	viral concentration per serving (g.e. HEV/serving)
λ_i^{-1}	average viral concentration per serving (g.e. HEV/serving)
$serv_i$	mean serving size for food category i (gr)
$(cons^i_{day})^{-1}$	mean daily intake of food category i in a year per person (gr/day)
$d^a_i = \frac{cons^i_{day} \cdot 365}{serv_i}$	total servings consumed per year per person (no. serving/year)
$S \sim \exp(\mu)$	individual HEV infectious dose distribution (g.e.)
μ^{-1}	mean individual infectious dose (g.e.) [63, 65, 66, 79]
$serv_{out}$	mean serving size (gr) of food implicated in outbreaks [63, 65, 66, 79]
N	Italian population 18+ (1st January 2021. ISTAT)
h	proportion of HEV seropositive population [75]
$\overline{N} = N \cdot (1 - h)$	susceptible population
$q_i = \mathbb{P}(C_i > S)$	probability of infection after consumption of one single serving
$T_i \sim \text{Geom}(q_i)$	number of failures before first successful exposure to HEV
$p_i = \mathbb{P}(T_i \leq d^a_i)$	probability of HEV infection per individual per year
$X_i \sim Bin(\overline{N}, p_i)$	distribution of new HEV infected individuals per year
$N^i_{inf} = \mathbb{E}(X_i)$	expected number of HEV infected individuals per year (*no.*)

for $serv_i$. We modelled S as an exponential distribution with parameter μ, which was estimated using MLE.

Any exposure to HEV, through the consumption of a single food serving, leading to an intake of HEV g.e. higher than the individual infectious threshold was defined as a new HEV infection event. We defined q_i as the probability of infection given the consumption of one single food serving. Each foodborne exposure to HEV was considered independent and no cumulative exposure to HEV in multiple meals was assumed possible.

Based on the available data on food consumption in Italy, we estimated the number of average servings consumed in one year per person d^i_a, for each food category i. Considering each meal as a Bernoulli trial of probability q_i, we built a geometrical random variable T of parameter q_i modelling the number of failures before the first successful exposure (i.e. infection). We estimated the probability p_i for an individual to become infected in one year as

$$p_i = P(T \leq d^i_a)$$

meaning the probability that the first successful exposure happens within one single year.

The number N of susceptible individuals in the Italian adult population was estimated subtracting to the total Italian population the fraction of HEV seropositive individual. These latter fraction was estimated based on data from a HEV seroprevalence survey among blood donors in Italy. We assumed a long-life immunization status against HEV of seropositive subjects (i.e. no reinfection possible). Moreover we assumed immune individuals homogeneously distributed all over the Italian population. We considered the total number of new infections per year X, given the consumption of food belonging to category i, to be Binomially distributed with parameters (N, p_i). Whence, we obtained the average number of new infected individuals in one year, as the expected value of X, meaning $N^i_{inf} = \mathbb{E}(X) = Np_i$.

2.1.2 Data and data sources

Data on HEV prevalence and concentration in food were generated from a sampling study carried out in Italy between 2016 and 2019 within the project "CCM 2016 – Hepatitis E, an emerging problem in food safety" (Suffredini, personal communication). Briefly, 730 samples were collected at retail level in different area of Italy: North, Centre and South (see Table 2.2). Samples were analysed using matrix-specific viral concentration procedure for pork products [77], bivalve shellfish, green leafy vegetables (ISO, 2017), and raw milk [35]. Detection and quantification of HEV in samples was performed by real-time RT-qPCR as detailed in Di Pasquale et al. [23].

Tabella 2.2: Survey sample size

Food category	**Sample size**
PNL	104
PL	92
RM	142
SH	204
GV	108

The number of individuals susceptible to HEV in the Italian population was estimated based on official demography data (as of 1st January 2021) (ISTAT (http://dati.istat.it/) and on a HEV seroprevalence study conducted among 10,011 blood donors' plasma unit samples ($\geq$ 18 years) in 2018 (0.02% of the adult Italian population as of 31st December 2018) [75].

The number of food servings consumed in one year by a single person was estimated based on food consumption data sourced from a nation-wide consumption survey conducted in Italy in 2005-2006 [46] available on FAO/WHO GIFT tool platform (http://www.fao.org/gift-individual-food-consumption/en/). Observations from the consumption survey database were filtered from the survey data to select the products

belonging to each food category using the "ingredient" variable. For RM and PL it was not possible to filter directly from data. For the former, we assumed that all the dairy caw milk consumed was raw, choosing a worst-case scenario. For the latter, data were not available for all the processed and/or seasoned food items. Therefore, we assumed the 20% of all pork sausages could contain liver.

Data collected during outbreak investigations with implicated food analysis were used to estimate the dose-response curve [63, 65, 66, 79].

Uncertainty and sensitivity

We performed uncertainty and sensitivity analysis to quantify the variability in the output due to the variability in the input parameters [31]. We followed a sampling-based method as described in Saltelli [72, 73] and generated 10,000 samples for each estimated parameters (i.e. $cons^{i}_{day}$, λ and μ) using parametric bootstrap [27, 84] and ran the model obtaining a sample of the same size for the output. This output sample was used to quantify the uncertainty and to perform the sensitivity steps with the aim to explore the effect of each parameter on the model output. We followed a two-step approach. We first analysed the scatter plots of input parameters versus the output and then calculated the standardized regression coefficient for each of the parameters by linear regression analysis. This latter step gave us the metric to rank the parameters. Details are described in Appendix A.

Evidence from the Italian surveillance system

Information on food consumption in cases of hepatitis E reported to the Italian surveillance system for acute viral hepatitis (SEIEVA) between 2016 and 2019 was used to discuss the outputs of the model. The SEIEVA is a voluntary system set up in 1985 by the Italian National Institute of Health now covering 83% of the Italian population [81, 82]. Since 2007 the local health units voluntary participating to the surveillance are required to perform and report HEV laboratory testing. Hepatitis E case definition is based on the positivity to IgM anti-HEV antibodies and elevate serum transaminases level (with or without clinical symptoms). Since the start of the SEIEVA activities, information on risk factors including food exposures were collected using a standardized questionnaire for all cases of acute viral hepatitis. The questionnaire was revised and released to include specific hepatitis E risk factors in late 2016. This activity was also completed within the national project CCM 2016 – Hepatitis E, an emerging problem in food safety.

2.2 Results

2.2.1 Parameter estimation

Food-specific parameter estimations are reported in Table 2.3. We reported also 95% interval confidence for the parameters estimated directly from data.

Tabella 2.3: Parameter estimations per food category. 95% CI are reported between brackets.

Category	Parameter	Estimation
PL-Liver pork products	α_{PL}	$0.11 \quad [0.06 - 0.20]$
	λ_{PL}	$8.55 \cdot 10^{-6} \quad [4.27 \cdot 10^{-6} - 1.43 \cdot 10^{-5}]$
	$serv_{PL}$	$150\ g$
	$cons^{PL}_{day}$	$0.73\ \frac{g}{day} \quad [0.70 - 0.752]$
PNL-No liver pork products	α_{PNL}	$0.028 \quad [0.006 - 0.082]$
	λ_{PNL}	$1.94 \cdot 10^{-5} \quad [4 \cdot 10^{-6} - 4.67 \cdot 10^{-5}]$
	$serv_{PNL}$	$150\ g$
	$cons^{PNL}_{day}$	$20\ \frac{g}{day} \quad [19.58 - 20.59]$
SH-Shellfish	α_{SH}	$0.0048 \quad [0.00012 - 0.027]$
	λ_{SH}	$1.15 \cdot 10^{-5} \quad [2.91 \cdot 10^{-7} - 4.24 \cdot 10^{-5}]$
	$serv_{SH}$	$150\ g$
	$cons^{SH}_{day}$	$10\ \frac{g}{day} \quad [9.99 - 11.7]$
GV-Leafy vegetables	α_{GV}	$0 \quad [0 - 0.033]$
	λ_{GV}	Inf
	$serv_{GV}$	$100\ g$
	$cons^{GV}_{day}$	$29\ \frac{g}{day} \quad [28.62 - 30.15]$
RM-Raw milk	α_{RM}	$0 \quad [0 - 0.025]$
	λ_{RM}	Inf
	$serv_{RM}$	$50\ g$
	$cons^{RM}_{day}$	$54.43\ \frac{g}{day} \quad [52.77 - 56.16]$

The results of food sampling survey indicated that the highest proportion of HEV positive samples (i.e. HEV food prevalence α_i) belongs to PL products (11%), followed by PNL (2.8%) and SH (0.48%). No positive samples were found for the GV and RM categories (0% prevalence) (Table 2.3).
The number of expected genome equivalent per 100gr size of serving (i.e. $\lambda_i^{-1} = \mathbb{E}(C_1^i)$) for PL and PNL products were $7.8 \cdot 10^4$ g.e./serving and $3.4 \cdot 10^4$ g.e./serving, respectively. The average servings consumed per person per year was 3 and 73, respectively for PL and PNL. The only positive sample for SH resulted in $8.7 \cdot 10^4$ g.e. per a 150g average serving size. The expected number of SH servings consumed yearly are 26. For GV and RM categories we did not obtain results for λ_i because no positive samples were detected. These estimates are shown in Table 2.4, where all the parameters directly involved in the ranking activity are also reported. We included in Appendix B a risk matrix that uses these parameters to profile also a qualitative ranking of the food categories.

Tabella 2.4: Mean viral load per serving (g.e./gr) $\mathbb{E}(C_1^i)$, meaning λ_i^{-1}, the average number of serving consumed in one year d_i^a, and the prevalence of HEV positive food samples for each category α_i.

Category	$\mathbb{E}(C_1^i)$ (g.e/serving)	d_i^a (no.serving/year)	Prevalence α_i
PL	78,000	3	0.11
PNL	34,000	49	0.028
SH	87,000	26	0.0048
GV	0	132	0
RM	0	107	0

The other parameter estimations are reported in Table 2.5. The average serving size consumed at outbreak events $serv_{out}$ resulted to be 100 gr, yielding to a mean individual infectious dose of $7.3 \cdot 10^6$ g.e. (i.e. μ^{-1}). The proportion of seropositive individuals of Italian population h was derived by Spada et al. [75] and resulted to be 8.3%. We subtract this proportion to the total Italian population to get the amount of susceptible individuals. The total population as of 1^{st} January 2021 was estimated to be 50,208,329 yielding to 45,840,205 susceptible individuals.

Tabella 2.5: General model parameters. 95% CI are reported between brackets.

Parameter	**Value**	**Reference/Comment**
μ	$1.37 \cdot 10^{-7}$ $[5.51 \cdot 10^{-8} - 2.55 \cdot 10^{-7}]$	[63, 65, 66, 79]
$serv_{out}$	100 g	Assumed
h	0.087	[75]
N	50,208,329	ISTAT data as of 1st January 2021
$\overline{N}$	45,840,205	$h \cdot N$

2.2.2 Model output

For each food category, the outputs of each step of the model are presented in Table 2.6, including individual probability of infection following the consumption of a single serving q_i, individual probability of infection in a year p_i, and the expected number of new infected per year $N_{inf}^i = \mathbb{E}(X_i)$.

We reported in addition density and cumulative probability sketches for new HEV infected individuals per year X_{PL}, X_{PNL}, and X_{SH} (Figure 2.2). Standard deviations of these three binomial random variables are 409 for PL, 668 for PNL, and 260 for SH.

Tabella 2.6: Probability of infection following the consumption of a single serving (q_i) and in one year (p_i) and expected number of new infected individuals per year per food category in the Italian population (N_{inf})

Category	q_i	p_i	N^i_{inf}
PL	$1.22 \cdot 10^{-3}$	$3.65 \cdot 10^{-3}$	167,200
PNL	$1.35 \cdot 10^{-4}$	$9.81 \cdot 10^{-3}$	449,917
SH	$5.67 \cdot 10^{-5}$	$1.47 \cdot 10^{-3}$	67,473
GV	0	0	0
RM	0	0	0

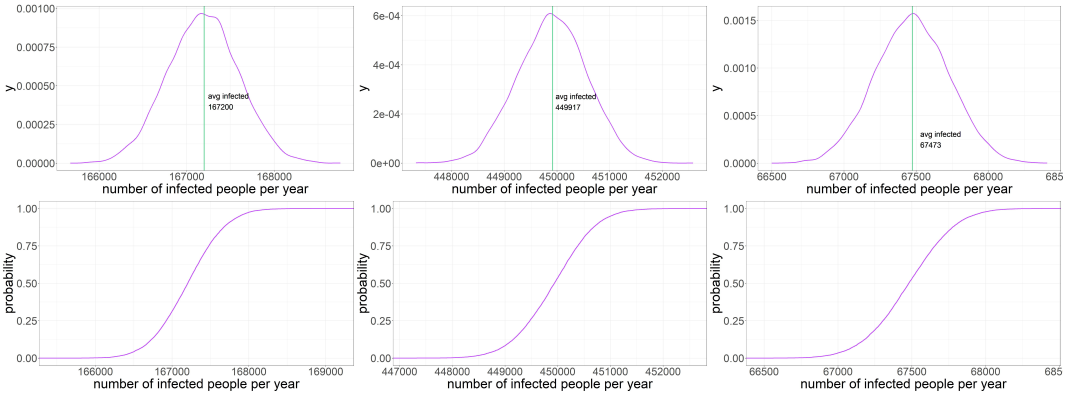

Figure 2.2. They are named from **a** to **f** from the top left corner to bottom right. (**a**) Density distribution of new HEV infected individuals per year X_{PL}. The green line indicates the mean value N^{PL}_{inf}. (**b**) Density distribution of new HEV infected individuals per year X_{PNL}. The green line indicates the mean value N^{PNL}_{inf}. (**c**) Density distribution of new HEV infected individuals per year X_{SH}. The green line indicates the mean value N^{SH}_{inf}. (**d**) Cumulative distribution of X_{PL}. (**e**) Cumulative distribution of X_{PNL}. (**f**) Cumulative distribution of X_{SH}.

2.2.3 Uncertainty analysis

We obtained bootstrap samples from input parameters and model output as described in section 2.1.2. We considered parameters uncorrelated given the Pearson correlation test results that are reported in Appendix A, where standard deviations of input parameter samples are also shown (see Table A.1).

In Table 2.7 we reported summary statistics of the output distribution for the three categories involved in the analysis. In Figure 2.3, 2.4, and 2.5 we displayed the histogram and the sample cumulative distribution of the output samples for category PL, PNL, and SH respectively. The output shown for this analysis is the individual infection probability in a year p_i.

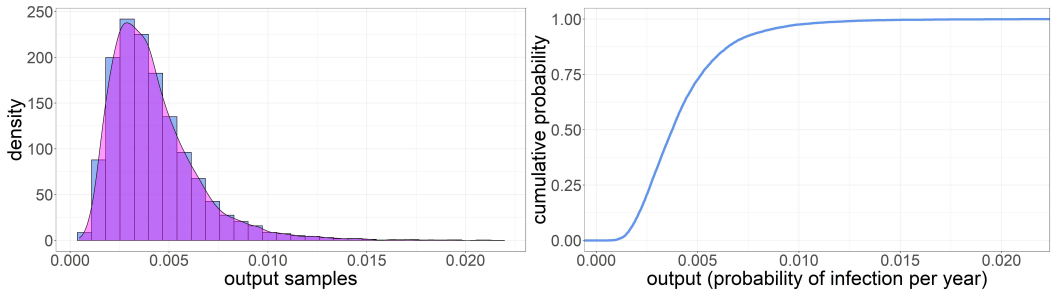

Figure 2.3. (**a**) Histogram of p_{PL} samples. (**b**) Sample cumulative distribution of p_{PL}.

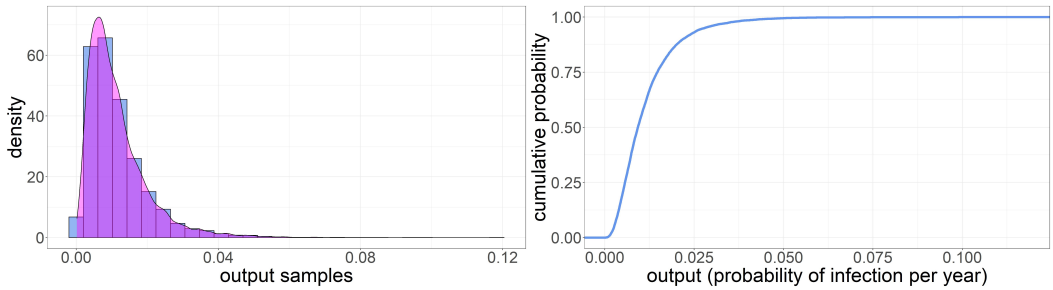

Figure 2.4. (**a**) Histogram of p_{PNL} samples. (**b**) Sample cumulative distribution of p_{PNL}.

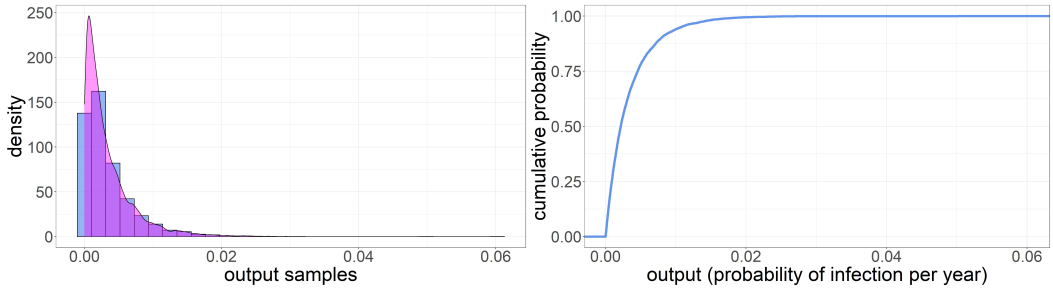

Figure 2.5. (**a**) Histogram of p_{SH} samples. (**b**) Sample cumulative distribution of p_{SH}.

Tabella 2.7: Summary statistics of output samples

Category	Parameter	Mean	1st Qu.	Median	3rd Qu.
PL	p_{PL}	$4.2 \cdot 10^{-3}$	$2.6 \cdot 10^{-3}$	$3.7 \cdot 10^{-3}$	$5.1 \cdot 10^{-3}$
	N_{inf}^{PL}	190,000	120,000	170,000	230,000
PNL	p_{PNL}	$1.1 \cdot 10^{-2}$	$5.6 \cdot 10^{-3}$	$9.2 \cdot 10^{-3}$	$1.4 \cdot 10^{-1}$
	N_{inf}^{PL}	520,000	250,000	420,000	670,000
SH	p_{SH}	$3.3 \cdot 10^{-3}$	$8.8 \cdot 10^{-4}$	$2.1 \cdot 10^{-3}$	$4.5 \cdot 10^{-3}$
	N_{inf}^{SH}	150,000	40,000	99,000	200,000

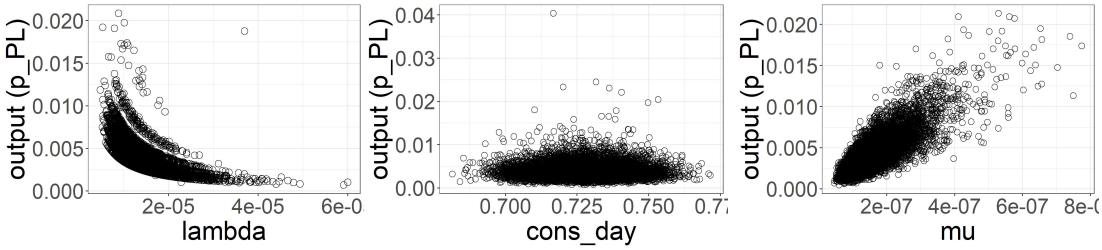

Figure 2.6. Scatterplot of input parameters for category PL versus the output considered (p_{PL}). (**a**) λ_{PL}. (**b**) $cons_{day}^{PL}$ (**c**) μ.

2.2.4 Sensitivity analysis

From uncertainty analysis, we obtained samples of input parameters and output used for perform the sensitivity analysis. The three scatter plots (Figures 2.6, 2.7, and 2.8) show the variations of the output samples versus the three parameters involved in the analysis, underlying the strength of the dependencies between them. The shape $cons_{day}^{i}$ plot is a bit flattened on the bottom with no strong structure for all the food categories, suggesting a very light dependency between the output and this parameter. λ seems to have a little more structured outline, especially for PL category, but the output exhibits the strongest dependency with the μ parameter whose plots shows a well defined shape, especially for PL and PNL.

To quantify the relative importance of input parameters we conducted a regression

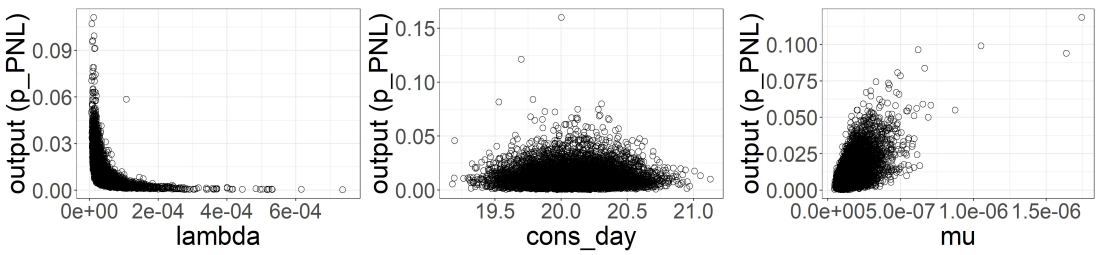

Figure 2.7. Scatterplot of input parameters for category PNL versus the output considered (p_{PNL}). (**a**) λ_{PNL}. (**b**) $cons_{day}^{PNL}$ (**c**) μ.

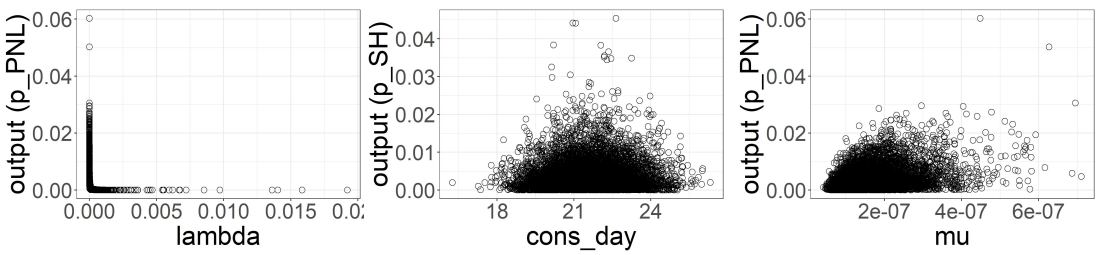

Figure 2.8. Scatterplot of input parameters for category SH versus the output considered (p_{SH}). (**a**) λ_{SH}. (**b**) $cons^{SH}_{day}$ (**c**) μ.

analysis using $\lambda_i, cons^i_{day}$, and μ as covariates. This, as explained in [73, Cap.1], allowed us to rank these three input parameters based on their impact on the output. Results are reported in Tables 2.8, 2.9, and 2.10.

Tabella 2.8: Regression analysis coefficients result for category PL

Coefficient	Estimate	Standardized estimate	Std. error	p-value
$\widehat{\lambda_{PL}}$	$-2.2 \cdot 10^2$	$-7.21 \cdot 10^{-1}$	1.39	$< 2 \cdot 10^{-16}$
$\widehat{cons^{PL}_{day}}$	$5 \cdot 10^{-3}$	$8.13 \cdot 10^{-1}$	$3.27 \cdot 10^{-5}$	$< 2 \cdot 10^{-16}$
$\hat{\mu}$	$2.26 \cdot 10^4$	$8.47 \cdot 10^{-1}$	$8 \cdot 10$	$< 2 \cdot 10^{-16}$

Tabella 2.9: Regression analysis coefficients result for category PNL

Coefficient	Estimate	Standardized estimate	Std. error	p-value
$\widehat{\lambda_{PNL}}$	$-9.25 \cdot 10$	$-3.93 \cdot 10^{-1}$	1.39	$< 2 \cdot 10^{-16}$
$\widehat{cons^{PNL}_{day}}$	$2 \cdot 10^{-4}$	$3 \cdot 10^{-1}$	$7.8 \cdot 10^{-6}$	$< 2 \cdot 10^{-16}$
$\hat{\mu}$	$7.14 \cdot 10^4$	$8.76 \cdot 10^{-1}$	$8.13 \cdot 10^2$	$< 2 \cdot 10^{-16}$

As already suggested by scatter plots, μ was the parameter that most influences the output for both the categories PL, PNL. The overall influence order for each category is the following:

$$|\hat{\mu}| > |\widehat{cons^{PL}_{day}}| > |\widehat{\lambda_{PL}}|$$

and

$$|\hat{\mu}| > |\widehat{\lambda_{PNL}}| > |\widehat{cons^{PNL}_{day}}|$$

From the value of the standardized estimates of the coefficients, we can evaluate the impact of each of them for perturbations equal to a fixed fraction of parameter's standard deviation [72, Cap. 6]. For the PL category, μ has an impact 4% higher than $cons^{PL}_{day}$ and 14% higher than λ_{PL}, while λ_{PL} has a impact 11% higher than $cons^{PL}_{day}$. For the PNL category, μ has a impact 55% higher than λ_{PNL} and 68% higher than $cons^{PNL}_{day}$. The impact of λ_{PNL} is about 29% higher than $cons^{PNL}_{day}$. For PL category it is clear that $cons^{PL}_{day}$ and λ_{PL} are very close and all the three parameters have an impact relatively

Tabella 2.10: Regression analysis coefficients result for category SH

Coefficient	Estimate	Standardized estimate	Std. error	p-value
$\widehat{\lambda_{SH}}$	$-1.57 \cdot 10^{-2}$	$-1.45 \cdot 10^{-2}$	$7.53 \cdot 10^{-3}$	$3.6 \cdot 10^{-2}$
$\widehat{cons_{day}^{SH}}$	$1.18 \cdot 10^{-5}$	$4.92 \cdot 10^{-2}$	$3.96 \cdot 10^{-6}$	$3.51 \cdot 10^{-3}$
$\hat{\mu}$	$1.99 \cdot 10^{4}$	$6.76 \cdot 10^{-1}$	$4.98 \cdot 10^{2}$	$< 2 \cdot 10^{-16}$

small.
The R^2 statistics for both categories is high, 0.97 for PL and 0.83 for PNL, meaning that these three parameters account for almost the entire uncertainty in the output.
Slightly different pattern is shown by SH category, where

$$|\hat{\mu}| > |\widehat{cons_{day}^{SH}}| > |\widehat{\lambda_{SH}}|.$$

We found that μ has an impact 98% higher that λ and 92% higher than $cons_{day}^{SH}$, for a perturbation equal to a fixed fraction of parameter's standard deviation. The R^2 for this parameter resulted much smaller with a 0.51, suggesting that other factors are contributing to the uncertainty. This is not surprising, given that we have only one positive sample for this category.

Last, in order to have a final quantification of the stability of the output we made a simple experiment on d_i^a parameter (so, indirectly on $cons_{day}^i$). We increased the d_{PL}^a to actually see the results in number of exposed people per year. We observed that with $d_{PL}^a = 6$ the number of infected people reached $335,446$ and with $d_{PL}^a = 8$ $N_{inf}^{PL} = 446,715$.

More details on results of these analysis are reported in Appendix A.

2.2.5 Food consumption in Italian hepatitis E cases

Between 2016 and 2019, a total of 213 autochthonous cases of hepatitis E were reported to the SEIEVA system. The availability of information on foods consumed by patients before the onset of the disease varied considerably among the different foods, depending on the type of questionnaire administered to hepatitis E patients. Information on shellfish consumption was available for a high proportion of cases (N=186; 87%) because the exposure to this foodstuff was investigated through both the general questionnaire for acute viral hepatitis and the specific questionnaire for hepatitis E, in place from late 2016. For all the other food items, which were investigated with the hepatitis E questionnaire only, this proportion did not exceed the 43% of the hepatitis E cases (Table 2.11) making the uncertainty around the exposure prevalence to these foods much higher. Pork meat and pork cured meat were by far the food items most frequently consumed by hepatitis E cases with more than 69% of the patients having consumed these foods, followed by pork liver, fruits, shellfish, vegetables and wild boar meat (Table 2.11).

Tabella 2.11: Information on food consumption among hepatitis E cases reported between 2016 and 2019 to the Italian surveillance system for acute viral hepatitis (SEIEVA)

Food	**Cases with information on consumption available**	**Cases reporting consumption of the food**	
	N	**N**	**% (of cases with information available)**
pork meat	90	69	76
pork cured meat	83	58	70
pork liver meat	51	14	29
fruit	36	10	28
shellfish	186	49	26
wild boar meat	77	17	22
vegetables	34	5	15
wild boar cured meat	55	8	14
offal	53	6	11
other game meat	53	4	7

2.3 Discussion

The model output shows that the consumption of PNL led to the greatest exposure to HEV in the Italian population and was associated with the highest number of new expected HEV infections per year, followed by the consumption of PL and SH. Based on these findings the risk posed by PNL is ranked first at the population level among foods implicated in the transmission of HEV, followed by PL and SH. For the other foods considered by our study (i.e. GV and RM) no expected cases of HEV infections were estimated by our model. The consumption of pork products has been frequently indicated as a risk factor for foodborne HEV infection [47]. This type of food has also been frequently linked to foodborne outbreaks of hepatitis E [13, 16, 63, 65]. The consumption of shellfish has also been pointed out as a possible risk factor in some studies [6, 63], although to our knowledge no outbreaks implicating the consumption of contaminated shellfish have ever been reported in Europe.
PNL are consumed much more frequently than PL and SH and by a larger proportion of population and this explains why the highest expected number of new cases in the population are associated with this food despite the mean prevalence of HEV contamination and the viral load per serving are higher for PL and SH.

The sensitivity analysis indicates that even slight increase in the consumption of PL servings at the individual level results in a remarkable increase in the expected number of new cases of HEV infection. As an example passing from three to eight portions of PL consumed per person per year, which is a realistic variation at individual level, the number

of new HEV infections in the population increases from about 168,000 to approximately 450,000 cases, revealing that the number of servings of PL at the individual level is a critical element to be taken into account for food risk ranking. Similar variations in the consumption of PNL do not result in comparable increases in the number of new HEV infection cases in the population. These findings provide evidence of the importance of collecting very accurate data on PL consumption at the population level, in order to strengthen the HEV food risk ranking and highlight the importance of PL consumption for the risk of HEV transmission, at the individual level.

The consumption of PL in Italy varies hugely among both individuals and population subgroups, depending not only on personal preferences but also on traditional differences in consumption habits. There is a wide geographic variability in the recipes and mode of preparation of pork products with important local peculiarities especially for products like cured meat and offal. This may lead to highly heterogeneous consumption of PL and exposure to HEV in specific population subgroups and geographical area. Unfortunately, the food consumption data source used in our study lacks of sufficient details to allow for a more accurate estimations of HEV exposure associated with the different types of PL

Food consumption data from hepatitis E cases reported to the SEIEVA (see section 2.2.5) are too scarce to support a formal validation process of our model. However, they are in line with the evidence obtained in our study about the importance of both PNL and PL as top risk foods for the transmission of HEV to humans. The consumption of pork meat and pork cured meat was reported by a high proportion of cases (>70%) while pork liver as well as shellfish by a much lower proportion of cases (i.e. each food group not exceeding the 30% of the cases). Although the SEIEVA data would indicate that a minor proportion of hepatitis E cases had consumed pork liver, it is necessary to consider that this food is not usually consumed as single food but it is much more frequently consumed as ingredient in mixed pork cured meat such as sausages, salami, mortadella etc., which were consumed by a high proportion of cases. Unfortunately, the lack of food consumption data from healthy controls hampers drawing more specific conclusions on the magnitude of the association between PNL and PL food consumption and hepatitis E, at the individual level.

Our model was build to support risk ranking. The sensitivity analysis shows that the parameter that brings the larger uncertainty is the mean μ^{-1} of the HEV individual infectious dose distribution. This is not surprising, given that the scarce availability of data to estimate μ affected the possibility to build a robust dose-response model similarly to many other exposure study in humans. In our study we used the total HEV load (g.e.) in food implicated in hepatitis E outbreaks as a proxy for the individual HEV infectious dose in humans. It is impossible, however, to evaluate to what degree the quantitative assessment of HEV in food differs from the true individual infectious dose. In addition, two different sources of uncertainty affect our dose-response model. On one hand, we only found four outbreak reports in the literature providing the information needed. On the other hand the uncertainty associated with the quantitative methods used for the assessment of HEV

in food should be also considered.
To obtain more reliable estimates of the actual number of HEV infections in humans a more robust estimation of parameter μ would be needed. Nonetheless, the food ranking does not appear to be influenced by the uncertainty introduced by this specific parameter since the dose-response model acts in the same way on all type of food considered in our study. The parameters potentially introducing differences in the ranking are the ones displayed in Table 2.4. To have a more direct view of how these parameters affect the ranking, we have provided a qualitative risk classification by building a risk matrix. The analysis is reported in the Appendix B.

Other important factors affecting the model outcomes are the mean quantity of food consumed in a year $cons_{day}$ and the mean viral concentration in food λ^{-1} . While the λ parameter was estimated from data from a large national sampling study, the $cons_{day}$ was estimated from consumption dataset whose current reliability is difficult to assess for several reasons. First, the survey was conducted in the Italian population more than fifteen years ago and data might no longer reflect the current consumption in terms of type of food, frequency of consumption and quantity with the same level of accuracy. Second, the consumption data refer to general food categories and the lack of details on the food ingredients makes it difficult to extract the true quantities of food consumed for some categories, introducing uncertainty in particular for PL. Finally, estimates were only available at the national level and did not allow to take account for regional and local differences, which in the case of PL and PNL may be critical, as described above.

Another limitation of our study is the use of one single dataset for the estimation of the prevalence of HEV contamination in food. This methodological choice was driven by the lack of full comparability of the prevalence estimates among different studies, due to a poor harmonisation of laboratory methods used to detect HEV in food. In addition, the uncertainty analysis that we performed in our study would have been impossible using estimates from other studies. Due to the extremely high variety of foods and mode of preparation, estimating the prevalence of HEV contamination in food based one single survey may introduce a selection bias depending on the goodness of randomization of the samples. In our study, this type of bias may be suspected for GV and SH prevalence estimation. For both these foodstuffs the estimated prevalence was 0.5% and 0% respectively, representing a highly discrepant value compared with other similar studies conducted in Italy and abroad. In SH, Suffredini et al. and La Rosa et al. [45, 76] reported much higher prevalence of HEV contamination with values ranging up to 8.1%. The prevalence of GV contamination, although considerably lower than SH, was never estimated to be 0% in other studies [42, 74, 80]. These considerations suggest that the role of SH and GV for HEV transmission in the Italian population may be more important than our study showed. In terms of risk ranking, however, this does not appear to substantially change our results. Different is the case of RM. This food was included in our study because it was focused as a potential risk food for HEV transmission in China in 2016 where a high

prevalence of active HEV infection in cows was reported [37]. However, no further studies confirmed these findings [4, 32, 83].

Capitolo 3

Pre-harvest model: the farm level and pig demography

In this part of the study, we aimed at modelling the transmission dynamic of HEV at the pig farm level in Italy, in order to estimate the number of infected animals entering the food chain at slaughtering. To do that we build a SEIR stochastic model, which was applied to three different types of pig production farms. These farm types were defined on the basis of the animal husbandry practice adopted in the farm, the production type, the farm size and the demographic characteristics.
To the best of our knowledge, this is the first mathematical modeling study of hepatitis E in Italian farms.

3.1 Farming types and Italian pig population analysis

In Chapter 2 we developed a model to estimate the number of infected people in the population per year, given the consumption of specific food categories. The exposure study, conducted using point prevalence data of HEV contamination of food sampled at retail, allowed us to have a simple assessment of the risk posed by some food products. A limitation of this approach, however, is that the representativeness of point prevalence estimates despite accurate and valid (i.e. obtained from large or very large sampling study) is limited to the geographical and temporal context of the study. Estimating probability of HEV contamination for pork carcasses at slaughterhouse would allow to estimate the level of food contamination at retail, considering also that differently from bacteria viruses do not growth on meat and that HEV load in meat and offal could only progressively decrease during processing and seasoning of meat and meat products. HEV contamination of carcasses and offal at slaughter depends directly on the infectious status of pigs entering the slaughterhouse (pre-harvest) at the end of the production cycle, which in turn depends on the epidemiological situation of HEV in the farms that can be affected by other multiple factors connected to the primary production of pigs, i.e. animal husbandry practice, length of the production cycle, biosecurity measures etc.

To strengthen the assessment of HEV transmission to humans through food, it is crucial to consider all the possible factors connected to the primary production, from farm to slaughterhouse. Hence, a pre-harvest model has been developed to gain deeper knowledge of the possible factors affecting transmission of HEV among animals and viral persistence in pig farms.

As reported in Chapter 1, the geographical distribution of hepatitis E in Italy is quite heterogeneous. The variability of HEV seroprevalence in human population at regional level reflects a different risk of consumers exposure to HEV, through food. Although this could be explained by several factors, higher intake of HEV depends on either a greater consumption of food considered at risk for HEV (e.g. traditional recipes and regional consumption habits) or by higher level of HEV contamination of food at the consumer level (i.e. at retail). Pig industry in Italy is indeed characterised by a remarkable geographic variability with different types of production in the various regions. This heterogeneity suggests a possible correlation between circulation of the virus and farming practice (i.e. pigs production type). For these reasons, we performed a preliminary analysis of pig demographic distribution in Italian regions to characterise, compare, and finally define farm classes according to specific factors.

Characteristics of the Italian pig population

The total pig population in Italy accounts for 8,8 million pigs which are raised in 31,373 farms, as of 30 June 2021. In addition, there are 101,568 familiar farms (<15 pigs per farm) where animals are raised for domestic self-consumption, often not continuously over time.
The main pig productions are represented by:

1. Live breeding stock, including sows, breeding boars and young gilts. These animals are intended for the production of piglets for breeding return or fattening.
2. Weaner (7-30 Kg) or grower animals (50-100 Kg) intended for fattening farms.
3. Light fattener animals (up to 120 Kg corresponding to 7-8 months of age) intended for slaughter for fresh consumption and industrial processes.
4. Heavy fattener animals (up to 180 Kg corresponding to 9 months of age) intended for slaughtering mainly for production of seasoned or processed products, like hams, sausages or salami.

The management of production cycle consists of five main phases: mating, gestation/delivery, nursery, weaning and growing-finishing. Breeding is a continuous-cycle activity over the year.
Italian pig population is therefore characterized by three main types of farm:

Open-cycle breeding farms (Site 1) for the production of piglets up to the weaning stage. Mainly sows and boars are farmed in these farms;

Growing farms (Site 2): farms where pigs are growth up to the end of weaning and/or magronage, intended for fattening farms (Site 3) or breeding gilts production;

Fattening farms (Site 3): farms where pigs are growth up from the weaning and / or growing phase to the end of the production cycle, intended exclusively for slaughter;

Closed-cycle breeding farms : farms in which there are breeding sows and boars and pigs growing up to the fattening stage intended exclusively for slaughtering, and/or growing pigs intended to be transferred for farms from reproduction. These type of farms are also known as "farrow-to-finish".

While closed-cycle farms are usually derived from most traditional farms where all the phases are directly realised in the same structure, three-sites farms are considered the most intensive rearing facilities. In this intensive farm units the entire production cycle is segmented into specialised production sites. These units (sites) are very often, but not exclusively, owned by the same company which adopts this functional organisation (i.e. the so called "filiera"= chain) to increase the business volume and exploit scale economies.

Pork production chain

The pork production chain is composed by two main production steps, The first, includes the pre-harvest animal production cycle and the slaughtering. The second includes the post-harvest production steps, downstream of the previous one, represented from (meat and offal) processing of products (processed meats and cured meats) (see Figure 3.1). The central node of the production chain is represented by the slaughtering industry, which is generally supplied through direct purchase of pigs from farmers raised in Italy and other EU country (mainly Denmark, Netherlands, Poland). Slaughtering is mainly concentrated in industrial facilities with CE authorization and the activity of limited and marginal capacity plants is residual. In 2021, more than 2,000 slaughterhouses were present in the national Italian databank [19]. Despite this large number though, the activity is mainly concentrated in a very limited number of plants with considerable production capacity.

The second part of the pork production chain, the processing activity, is more fragmented. The slaughterhouse represents an important node of the supply chain. The majority of the production (about 65%) coming from the slaughterhouses is delivered to the processing industry (salumi), while the remaining portion for the fresh-meat market. The production of the processing industry continues through ham and sausages factories and, after the seasoning and the other processes are completed, reaches the consumer market through wholesalers, agents and dealers [17].

Pig population and farms in Italy

Data on Italian pig population and farms are regularly collected by the Veterinary Services of the local health authorities (ASL) and other organisations, according to the EU regulations, and are provided to the Istituto Zooprofilattico dell'Abruzzo e Molise

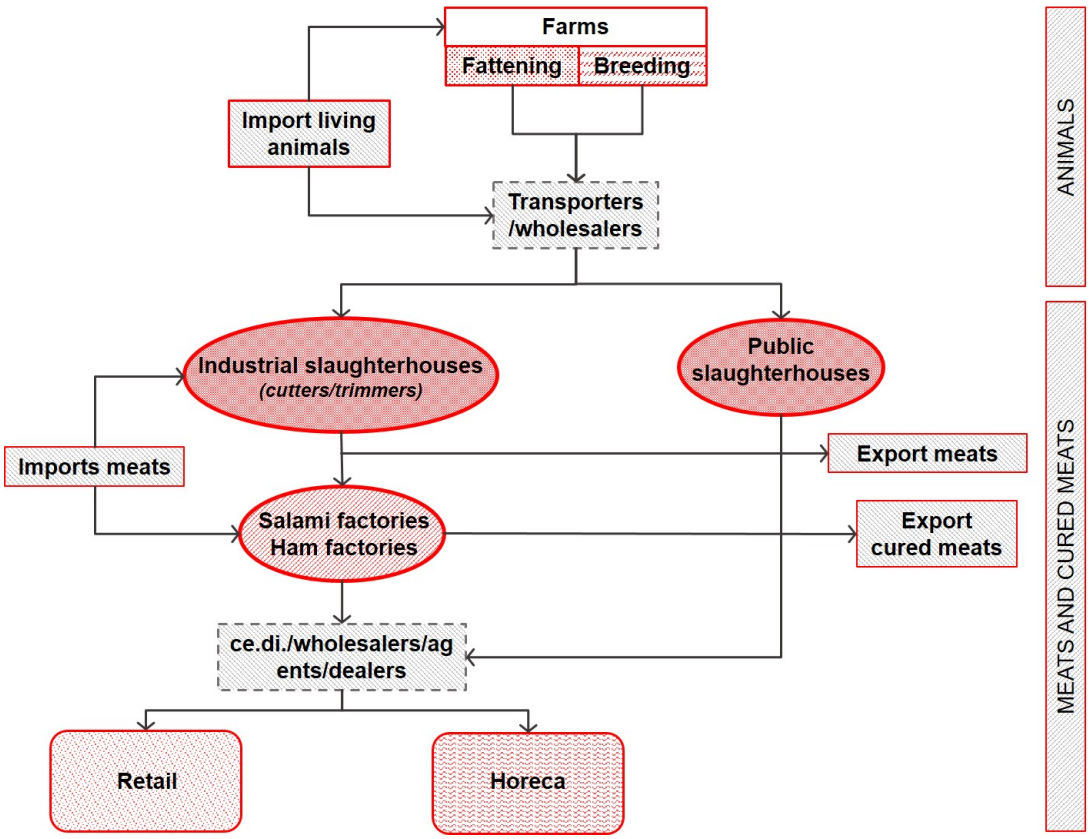

Figure 3.1. The main operators of the swine supply chain in Italy (Source [17])

that has been tasked by the Ministry of Health to manage the National Official Animal Registry. Pigs and pig farms are subjected to identification and registration. The basic data and summary statistics are public accessible on the web [19] and include a series of information on the location of the farm, the number of farmed pigs stratified by age/type, the main characteristics of the farms, the density of pig farms and heads by geographical area (ASL/ municipality/province/region), etc.

We analysed farm data from National Official Animal Registry, as of 31/12/2020. We observed a highly heterogeneous farming pattern between northern and central-southern regions, as shown in Figure 3.2. The top five Italian regions for total amount of heads of pig are northern regions as shown in Table 3.1.

Tabella 3.1: Top five Italian regions for the total amount of head of pig.

Region	Number of heads	Percentage on the total (%)
Lombardia	4,319,000	50.2
Piemonte	1,245,000	14.5
Emilia-Romagna	1,103,000	12.8
Veneto	641,000	7.5
Friuli-Venezia Giulia	240,000	2.8

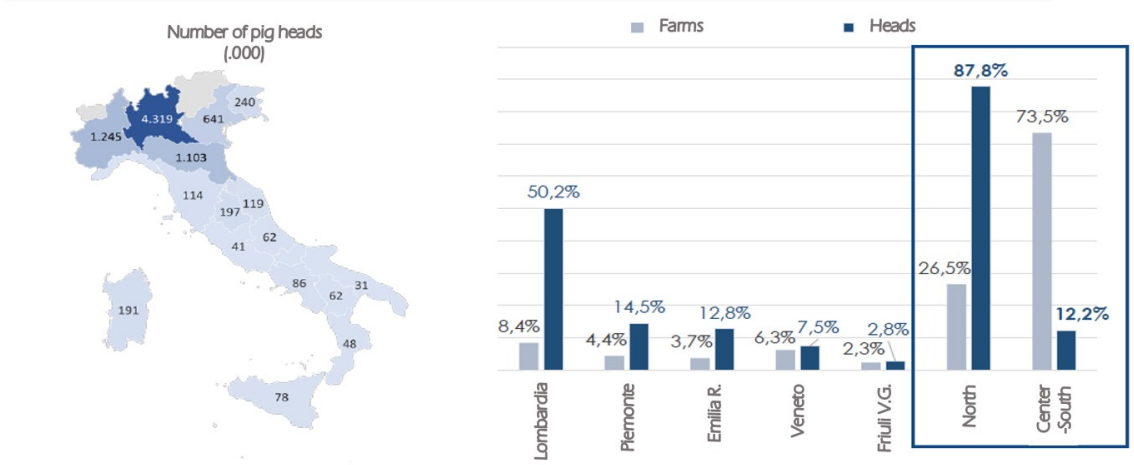

Figure 3.2. Territorial distribution of pig heads and farms (Source [18])

If we look at the total number of farms instead, the top five regions are southern and central, as shown in Table 3.2. Furthermore, more than 30% of farms in central and southern areas, except for islands, are classified as "familiar" farms. Excluding islands, the familiar farms are almost 90% of the regional farms for these areas.

Tabella 3.2: Top five Italian regions for the total amount of pig farms.

Region	Number of heads	Percentage on the total (%)
Campania	18,000	13.37
Sardinia	14,000	10.45
Lazio	13,700	10.33
Calabria	13,600	10.22
Abruzzo	13,300	10.02

The data show that pig population in the northern regions is characterised by much larger farms than in southern Italy where a large number of small sixe farms exist. Pigs farm in north Italy are mostly organized as intensive Site 1, Site 2, Site 3 farms while the farms located in central and south Italy are mostly closed-cycle breeding farms.
Another interesting feature is the mean age of pigs at slaughtering which reflect different production type. Pigs raised in Sardinia are usually slaughtered at a much younger age than in all other regions, at the end of the weaning phase (see Table 3.3), meaning around the 8th and the 10th week of age.

A deeper analysis of the demographic data of Italian pigs has been carried out and is available in Appendix C.
We analysed the demographical distribution of animal types in the different farms to define possible farm classes, based on specific factors including

- Farm size
- Relative distribution of reared animals by animal type

Tabella 3.3: Slaughtered animals in Sardinia by age class.

Age class	Total slaughtered	Percentage on the total (%)
Weaning	191,404	67.7
Growing	21,732	7.7
Light fattening	64,404	22.8
Heavy fattening	1,743	0.6
Other	3,227	1.2

- Farming activity and production stage
- Location of the farm

We defined fourteen classes that we pooled into nine macro-classes. All details of the data analysis carried out are reported in Appendix C. Here we isolated the most diffused farms in the country to define five representative farming types:

Type A Very large farms organized in the three the sites as described above (open-cycle breeding, growing and fattening farm) where mainly light pigs are reared.

Type B Medium/large size farms organised in the same three sites as for the type A with more heavy pigs reared.

Type C Closed-cycle farming type with at least 10% of sows of medium/small size.

Type D Small closed-cycle farms with less than 500 heads per farm.

Type SAR Sardinian farms with a high number of weaning and growing animals and mixed slaughtering age.

These types are heterogeneously diffused in the country, as shown in Figure 3.3 and already described in Table 3.2, with type A and B mostly diffused in the Northern regions and in part of the Center while C and D are more representative of the Southern area. The SAR type is of course exclusive of Sardinian region.

We will refer to the A and B types as “open-cycle farm” after its preeding compartment’s name and in opposition with the other three types that are all organized as closed-cycle.

3.2 Model description

We developed, for the five farm types defined in section 3.1, an individual-based discrete-time SEIR stochastic model [3] with three age classes for HEV transmission in pigs. The desired primary outcome of this model is the prevalence of HEV infected pigs entering at the slaughterhouse for the different farm types.

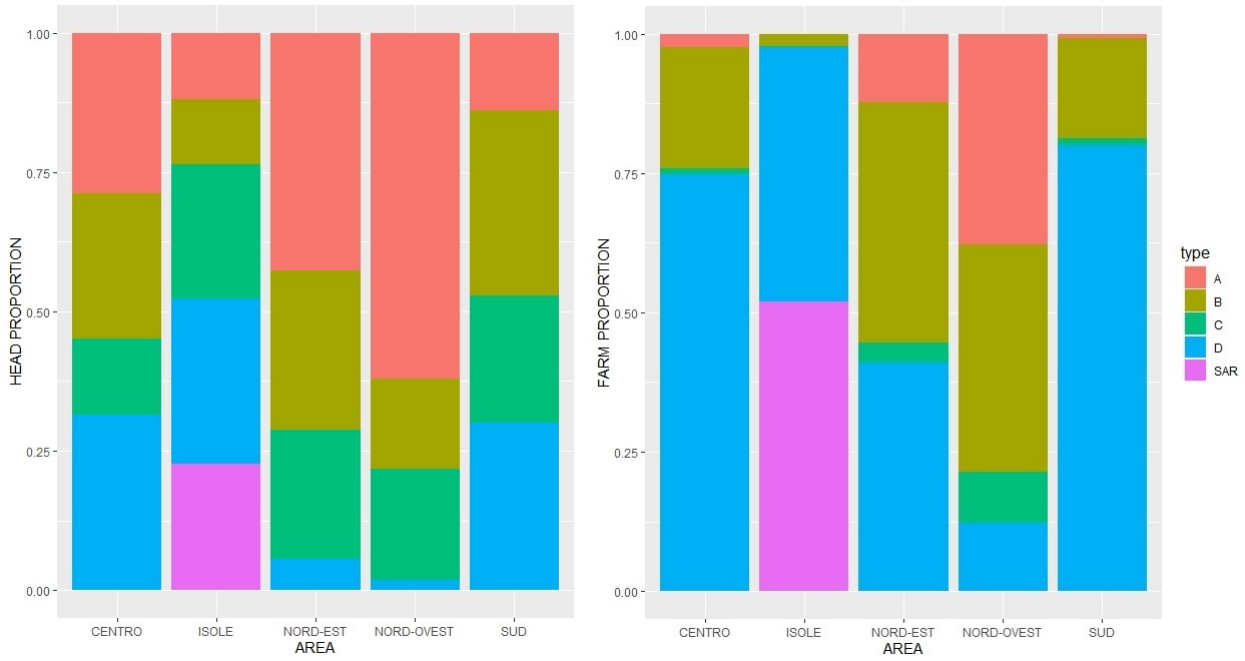

Figure 3.3. Distribution of heads (left) and farms (right) over the areas by types

We focused on the weaning-finishing production cycle, meaning that we are not considering in our model suckling pigs that are mostly covered by antibodies acquired by maternal colostrum. Hence, the three age classes analysed are weaning (W), growing (G) and fattening (F). We indicated with a superscript the variables referred to one of the three classes ($\bullet^W$, $\bullet^G$, $\bullet^F$).
We choose to represent four epidemiological status: susceptibles (S), exposed (E), infectious (I) and recovered (R). All animals are born susceptible and can became exposed after a successful contact event with an infectious animal. Even if infectivity can last for a longer time we considered only the shedding period as "infectious". Hence, an animal is considered infectious as long as they shed virus in faeces meaning when can spread the virus in shared environment. The exposed phase corresponds therefore with the latent period of the animal infection. Once an animal enters in the exposed compartment we keep track of the day post infection (DPI) a (in days). An infected animal experiences therefore two phases: a latent period from the exposure event to the beginning of the fecal shedding during which the animal is considered *exposed*, an *infectious* period until the end of the fecal shedding. After these two phases, when the animal is no longer able to infect others, it is considered *recovered*. We therefore will call infected any animal in the exposed or infectious compartment. The sequence of the epidemiological statuses is shown in Figure 3.4. We modelled infection and recovery transitions as explained below. All the model parameters are reported in Table 3.4.

Infection
We modelled the infection dynamics as a Poisson point process defined as explained below.

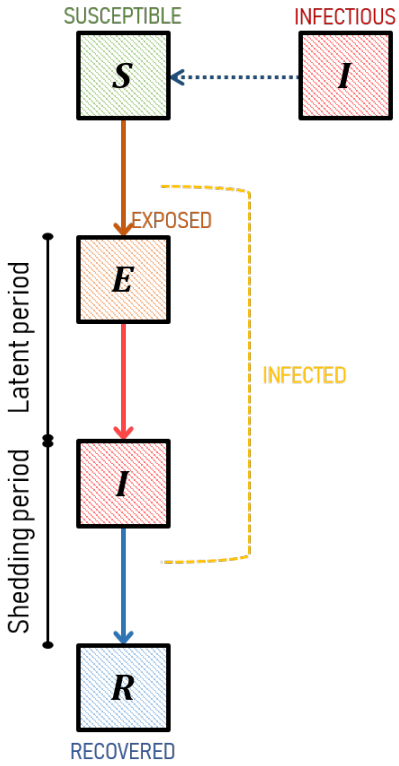

Figure 3.4. Epidemiological statuses sequence

We called the mean number of daily per capita contacts λ and with p the daily exposure probability, given an infectious contact.
The number of contacts happen at the time point of a homogeneous Poisson process of rate λ. Therefore, the number of daily per capita contacts are distributed as follow

$$Day_c \sim Poisson(\lambda).$$

The number of daily infectious contacts per capita depends on the density on infectious animals at that time and on the value of p. Hence, we obtain

$$Eff_c \sim Poisson\left(\lambda \cdot p\frac{I(t)}{N}\right)$$

where $I(t)$ are the total number of infectious individuals in the population. We can now calculate the daily per capita infection probability q_t as the probability to have at least an infectious contact per day

$$q_t = \mathbb{P}(Eff_c \geq 1) = 1 - e^{-\lambda\frac{I(t)}{N}p}$$

Let's now call $e_o(t)$ the number of newly exposed individuals at the end of day t. $e_0(t)$ results to be

$$e_0(t) \sim Poisson(q_t s(t)). \tag{3.1}$$

Following [71] we assumed a latent period distributed according to a Gamma distribution with shape α_{EI} and scale s_{EI}

$$T_{EI} \sim Gamma(\alpha_{EI}, s_{EI})$$

Parameter	Description
t	time (days)
a	day post infection – DPI (days)
$s(t)$	number of susceptible individuals at day t
$e(t,a)$	number of exposed individuals at day t which have been exposed for a days
$i(t,a)$	number of infectious individuals at day t which have been infected for a days
$r(t)$	number of recovered individuals at day t
$E(t) = \sum_{a\geq 1} e(t,a)$	number of exposed individuals at day t
$e_0(t)$	number of newly exposed individuals at the end of day t.
$I(t) = \sum_{a\geq 1} i(t,a)$	number of infectious individuals at day t
$i_0(t)$	number of newly infectious individuals at the end of day t
p	infection probability, given an infective contact, in a day
λ	mean number of contact per animal per day
$\beta = \lambda \cdot p$	mean number of infective contact per animal per day
$R_0(t)$	total number of newly recovered idividuals at the end of day t
$r_0(t,a)$	total number of newly recovered individuals at the end of day t with a day post infection
μ_{EI}	rate at which an exposed animal became infectious
μ_{IR}	recovery rate
ν	birth rate
δ	death rate
γ_{WG}	migration rate between W and G age class
γ_{GF}	migration rate between G and F age class
$\gamma_{F\bullet}$	slaughtering rate

Tabella 3.4: Model parameters

So, each individual has its own individual latent period modelled as an independent sample from T_{EI}. At the end of each day t we calculate the new infectious animals with a DPI as

$$i_0(t,a) = e(t-1, T_{EI} = a)$$

Recovery
For the recovery, we used a deterministic infectious period T_{IR}. As soon as the day post infection of the animal reaches the sum of its latent period and its shedding period the animal became recovered. At the end of each day we will have $r_0(t,a)$ new recovered animals, calculated as

$$r_0(t,a) = i(t-1, T_{IR} + T_{EI} = a).$$

The total number of newly recovered animal at day t will be indicated with $R_0(t)$

In Figures 3.5 and 3.6 are shown the diagram of the two farm types chosen. The first represents the open-cycle intensive production with the three sites (Site 1, 2 and 3) handling the three age classes in several smaller closed subpopulation representing the large herds mostly diffused in Northern and some central regions of the country. The second represents the weaning to finishing type, with a closed-cycle production for the smaller herds located in the central and especially southern areas. To the first macro-category

belong the first two different farm types A and B, in the second macro-category we have the other three farm types C, D and SAR.

Open-cycle intensive farms (type A and B)

We considered here a farming activity highly structured in a intensive farming orientation with high biosecurity in a all-in-all-out orgainisation form. In this kind of organisation we considered that several batches are reared together separated in subpopulations and that no inflow is possible nor cohexistence between animal of different ages. During the transition between one age class to the following the whole population is redistributed into the subpopulations, representing the movements of the animals between sites. No births or deaths are considered.

For this farm types we have three systems of equation for the three age classes that we considered. Each equation describes the evolution of individual of the same age reares in a single subpen.

The system of equation is the same for each age class and within each pens.

$$\begin{cases} s^X(t+1) = s^X(t) - e_0^X(t) \\ e^X(t+1,a) = \begin{cases} e^X(t,a-1) - i_0(t,a-1) & \text{for } a \geq 2 \\ e_0^X(t) & \text{for } a = 1 \end{cases} \\ i^X(t+1,a) = i^X(t,a-1) + i_0^X(t,a-1) - r_0^X(t,a-1) \\ r^X(t+1) = r^X(t) + R_0^X(t) \end{cases}$$

where $R_0^X(t) = \sum_a r_0^X(t,a)$ and $X \in \{W, G, F\}$.
Based on the age class considered we have:

$$1 \leq t \leq 30 \quad \text{for the weaning class (about 1 month);}$$

$$31 \leq t \leq 90 \quad \text{for the growing class (about 2 months);}$$

$$91 \leq t \leq 180 \quad \text{for light pigs in the fattening class (about 3 months);}$$

$$91 \leq t \leq 240 \quad \text{for heavy pigs in the fattening class (about 5 months);}$$

Furthermore, for all times t, we have

$$s^W(t)+e^W(t)+i^W(t)+r^W(t) = s^G(t)+e^G(t)+i^G(t)+r^G(t) = s^F(t)+e^F(t)+i^F(t)+r^F(t) = N$$

where N is the size of the subpopulations.

Closed-cycle farm (type C, D)

We assumed, for these types, a continuous-flow production from weaning to finishing where pigs are collected in larger groups, depening on the age class. We included here birth and deaths (slaughtering), whence in each age class animals of different ages are reared together. The diagram of the infection and growth dynamic is shown in Figure 3.7. The dynamics are very similar for the three age classes, so we have similar systems of

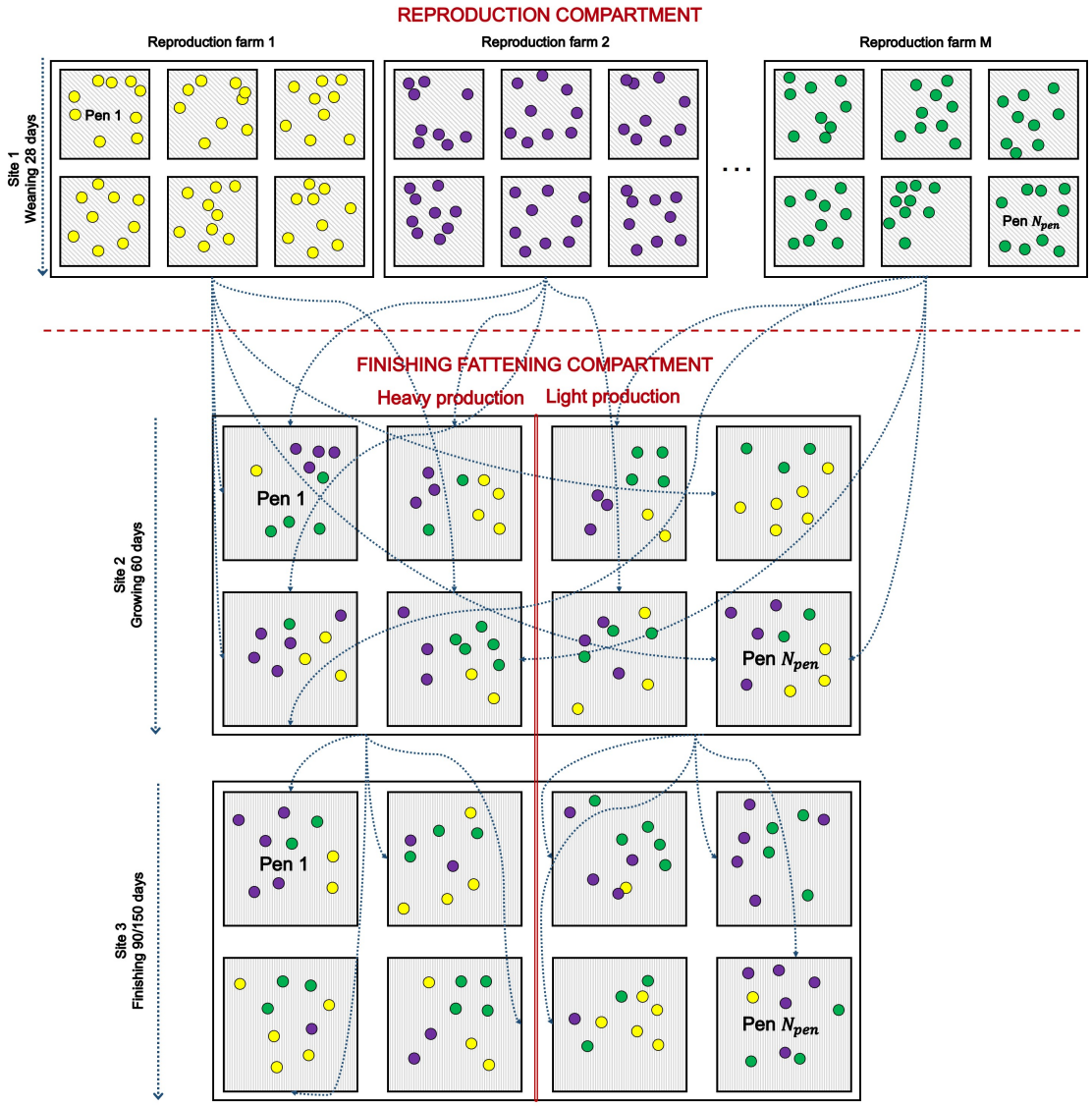

Figure 3.5. Flow of the A and B farm types – open-cycle three sites farm. The population is closed and segregated in smaller pens. The pens in each age class are populated picking at random from the previous age class based on their production type.

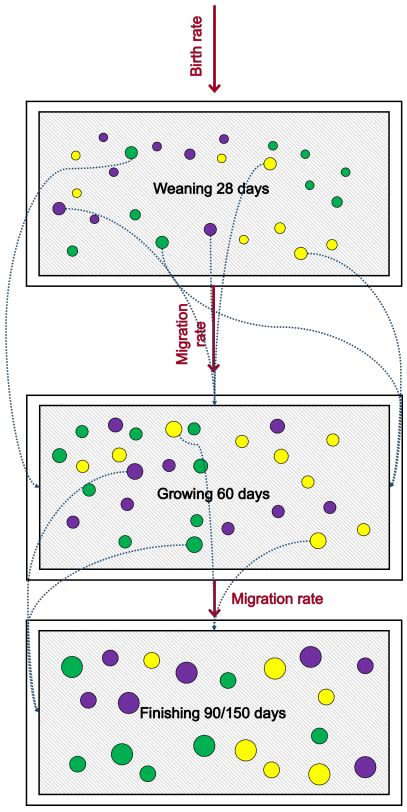

Figure 3.6. Flow of the C, D and SAR farm types – closed-cycle farm. Inflow and outflow were considered and within each age class animals of different ages are reared together.

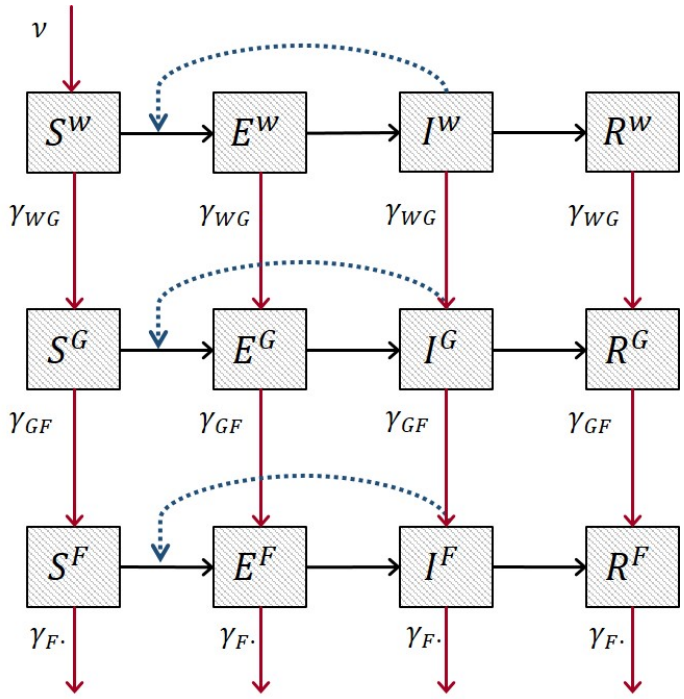

Figure 3.7. Diagram of infection and growth in closed-cycle model farms

equation for $t \geq 0$.

$$\begin{cases} s^W(t+1) = s^W(t) - e_0^W(t) + \nu - \gamma_{WG} s^W(t) \\ e^W(t+1,a) = \begin{cases} e^W(t,a-1) - i_0^W(t,a-1) - \gamma_{WG} e^W(t,a-1) & \text{for } a \geq 2 \\ e_0^W(t) & \text{for } a = 1 \end{cases} \\ i^W(t+1,a) = i^W(t,a-1) + i_0^W(t,a-1) - r_0^W(t,a-1) - \gamma_{WG} i^W(t,a-1) \\ r^W(t+1) = r^W(t) + R_0^W(t) - \gamma_{WG} r^W(t) \end{cases}$$

$$\begin{cases} s^G(t+1) = s^G(t) - e_0^G(t) - \gamma_{GF} s^G(t) + \gamma_{WG} s^W(t) \\ e^G(t+1,a) = \begin{cases} e^G(t,a-1) - i_0(t,a-1) - \gamma_{GF} e^G(t,a-1) + \gamma_{WG} e^W(t,a-1) & \text{for } a \geq 2 \\ e_0^G(t) & \text{for } a = 1 \end{cases} \\ i^G(t+1,a) = i^G(t,a-1) + i_0^G(t,a-1) - r_0^G(t,a-1) - \gamma_{GF} i^G(t,a-1) + \gamma_{WG} i^W(t,a-1) \\ r^G(t+1) = r^G(t) + R_0^G(t) - \gamma_{GF} r^G(t) + \gamma_{WG} r^W(t) \end{cases}$$

$$\begin{cases} s^F(t+1) = s^F(t) - e_0^F(t) - \gamma_{F\bullet} s^F(t) + \gamma_{GF} s^G(t) \\ e^F(t+1,a) = \begin{cases} e^F(t,a-1) - i_0(t,a-1) - \gamma_{F\bullet} e^F(t,a-1) + \gamma_{GF} e^G(t,a-1) & \text{for } a \geq 2 \\ e_0^F(t) & \text{for } a = 1 \end{cases} \\ i^F(t+1,a) = i^F(t,a-1) + i_0^F(t,a-1) - r_0^F(t,a-1) - \gamma_{F\bullet} i^F(t,a-1) + \gamma_{GF} i^G(t,a-1) \\ r^F(t+1) = r^F(t) + R_0^F(t) - \gamma_{F\bullet} r^F(t) + \gamma_{GF} r^G(t) \end{cases}$$

where $R_0^X(t)$ is defined as above $R_0^X(t) = \sum_a r_0(t,a)$.

Closed-cycle and early slaughtering age (type SAR)

To model the Sardinia farming type we used again a closed-cycle type farming with the same features of the previous two types (C and D). In this case, though, we have a early slaughtering age. We considered therefore only two age classes with a different slaughtering rates for weaning $\gamma_{W\bullet}$ and growing $\gamma_{G\bullet}$.

$$\begin{cases} s^W(t+1) = s^W(t) - e_0^W(t) + \nu - \gamma_{WG} s^W(t) - \gamma_{W\bullet} s^W(t) \\ e^W(t+1,a) = \begin{cases} e^W(t,a-1) - i_0^W(t,a-1) - \gamma_{WG} e^W(t,a-1) - \gamma_{W\bullet} e^W(t,a-1) & \text{for } a \geq 2 \\ e_0^W(t) & \text{for } a = 1 \end{cases} \\ i^W(t+1,a) = i^W(t,a-1) + i_0^W(t,a-1) - r_0^W(t,a-1) - \gamma_{WG} i^W(t,a-1) - \gamma_{W\bullet} i^W(t,a-1) \\ r^W(t+1) = r^W(t) + R_0^W(t) - \gamma_{WG} r^W(t) - \gamma_{W\bullet} r^W(t) \end{cases}$$

$$\begin{cases} s^G(t+1) = s^G(t) - e_0^G(t) - \gamma_{G\bullet} s^G(t) + \gamma_{WG} s^W(t) \\ e^G(t+1,a) = \begin{cases} e^G(t,a-1) - i_0(t,a) - \gamma_{GF} e^G(t,a-1) + \gamma_{W\bullet} e^W(t,a-1) & \text{for } a \geq 2 \\ e_0^G(t) & \text{for } a = 1 \end{cases} \\ i^G(t+1,a) = i^G(t,a-1) + i_0^G(t,a-1) - r_0^G(t,a-1) - \gamma_{G\bullet} i^G(t,a-1) + \gamma_{WG} i^W(t,a-1) \\ r^G(t+1) = r^G(t) + R_0^G(t) - \gamma_{G\bullet} r^G(t) + \gamma_{WG} r^W(t) \end{cases}$$

3.3 Parameter estimation

To estimate the disease-specific parameters, meaning p and λ and μ_{IR}, we performed a likelihood-based inference using the R pomp package [41]. To strenghten the estimation process we choose to estimate directly the product between λ and p, meaning $\beta = \lambda \cdot p$. To feed the inference process we looked at data on IgM prevalence from the study of [11], we used the average IgM prevalence among the all-in-all-out farms (see Table 3.5). We used the appearance of IgM as a proxy for the beginning of the shedding phase, meaning the transition between the exposed compartment and the infected one. The measurement made by Casas and collegues [11] are collected at weeks 3, 7, 13, 18 and 25. To fill the missing values we used a linear interpolation and then rounded the estimation to obtain an integer number (see Figure 3.9). Using these data we performed our parametric inference.

As described in the package documentation and in the literature, we first approximated our model dynamics. We used an SEIR model with constant rates as follow

$$\begin{cases} S(t) &= S(0) - N_{SE}(t) \\ E(t) &= E(0) + N_{SE}(t) - N_{EI}(t) \\ I(t) &= I(0) + N_{EI}(t) - N_{IR}(t) \\ R(t) &= R(0) + N_{IR}(t) \end{cases} \tag{3.2}$$

where $N_{**}(t)$ are the counting processes modelling the total amount of individual transitioned from one compartment to another at day t. The transition probabilities for these processes are

$$\begin{aligned} \mathbb{P}\Big[N_{SE}(t+\delta) = N_{SE}(t) + 1\Big] &= \mu_{SE}(t)S(t)\delta + o(\delta) \\ \mathbb{P}\Big[N_{EI}(t+\delta) = N_{EI}(t) + 1\Big] &= \mu_{EI}(t)E(t)\delta + o(\delta) \\ \mathbb{P}\Big[N_{IR}(t+\delta) = N_{IR}(t) + 1\Big] &= \mu_{IR}(t)I(t)\delta + o(\delta) \end{aligned} \tag{3.3}$$

where μ_{**} are the constant rates between compartments.

We used the Euler method to find the numerical solutions for the state variables $\tilde{S}(k\delta)$, $\tilde{E}(k\delta)$, $\tilde{I}(k\delta)$ and $\tilde{R}(k\delta)$ and we approximated the counting process using a binomial approximation with exponential probabilities, as follow

$$\begin{aligned} \tilde{N}_{SE}(t+\delta) &= \tilde{N}_{SE}(t) + Bin\Big[\tilde{S}(t), 1 - e^{-\mu_{SE}\left(\tilde{I}(t)\right)\delta}\Big] \\ \tilde{N}_{EI}(t+\delta) &= \tilde{N}_{EI}(t) + Bin\Big[\tilde{E}(t), 1 - e^{-\mu_{EI}\delta}\Big] \\ \tilde{N}_{IR}(t+\delta) &= \tilde{N}_{IR}(t) + Bin\Big[\tilde{I}(t), 1 - e^{-\mu_{IR}\delta}\Big] \end{aligned}$$

where $\mu_{SE}(\tilde{I}(t)) = \beta\frac{\tilde{I}(t)}{N}$. We assumed that the weekly reported seroprevalence in the dataset follows a negative binomial distribution with dispersion coefficient k and with a certain reporting rate $\rho \in [0, 1]$.

We performed a global search for the maximum of the log likelihood over the parameter space using the IF2 algorithm presented in the introduction. We picked 400 parameter

combinations from a subset of the parameter space and chosen following a preliminary local search. We used each of the 400 combination as stating value for the algorithm and set a 0.02 initial standard deviation for the random walk of the parameter vector. We iterated the IF2 algorithm for 150 iteration with 2000 particles each. On the results we repeat 10 other particle filtering with 5000 particles to estimate the likelihood of the original model, with constant parameters.
From the global search results we extracted an additional search box to build a profile likelihood of each parameter and calculate a 95% confidence integral via profile likelihood. Profile likelihood and confidence interval are defined as follow:

Definition 1. Let's $\theta = (\phi, \psi)$ be the parameter vector of a model. The profile likelihood for ϕ is a function of ϕ defined

$$\ell^{profile}(\phi) = \max_{\psi} \ell(\phi, \psi)$$

To build a 95% confidence interval for the profile likelihood we refer to Wilk's Theorem obtaining

$$\{\phi : \ell(\hat{\theta}) - \ell^{profile}(\phi) < 1.92\}$$

3.4 Model simulation

We developed an individual-based model where, for each animals, we kept track of

ID	Id progressive number
Age	Age (in days) initialized at 1.
Inf_state	Infectious state (S, E, I, R)
DPI	Age post infection (in days) initialized at 0
Pr_type	Production type ("l" low pig, "h" heavy pig)
Class	Age class ("w" weaning, "g" growing, "f" finishing, "s" slaughtered)
Lat_per	Latent period (i.i.d. samples from $T_{EI} \sim Gamma(\alpha_{EI}, \beta_{EI})$)
Inf_per	Infectious period ($\frac{1}{\mu_{IR}}$)

We initialised the population with starting epidemiological statuses as described in following paragraphs. The initial ages were set equals to 1 day for open-cycle farms and were uniformly chosen over the age range for the closed-cycle farms. Then for each day t we perform the following steps:

1. Calculate the number of animal in each compartment $I(t)$, $E(t)$, $S(t)$
2. Compute the number of new infected $E_0(t)$
3. From $S(t)$ pick at random $E_0(t)$ individuals and switch $Inf\_state \leftarrow$ "E''

4. If any exposed individual has $DPI > Lat\_period$: $Inf\_state \leftarrow$ "I"
5. If any infected individual has $DPI > Inf\_per + Lat\_period$: $Inf_state \leftarrow$ "R"
6. For each exposed or infected animal: $DPI = DPI + 1$
7. $Age = Age + 1$

Open-cycle three-sites farm

For the intensive farm model we have three separated phases corresponding to the three age classes: weaning, growing and finishing. The animals inside the farm are divided into subpens. We followed [36] to set the pen size pen_{size} to 40 animals. We simulated a number of pens equals to $N_{pen} = \frac{N}{pen_{size}}$, based on the average number of animals per farm type N.

We inizialised the pig population with $I^W(0) = \eta_I \cdot N$ infected and $R^W(0) = \eta_R \cdot N$ recovered animals. The total population was then distributed into the pens randomly in the weaning age class. For the type A we simulated only light pigs, while for the type B farm we choose the proportion of light and heavy pig to be reared based on the proportion of light and heavy pigs slaughtered in the area yearly. In the weaning compartment light and heavy pigs are mixed all together, while from the growing to the finishing phase they are separated (see Figure 3.5).

We ran the N_{pen} pens for 30 days for the weaning phase and then redistributed the animals at random to populate the growing compartment subpopulations. After 60 days we redistributed again the animals in finishing pens for a number of days depending on the production type, three other months for light pigs and five for heavy pigs. We simulated the all cycle five times to account for random variations. A representation of the flow for this model is shown in Figure 3.5.

Closed-cycle and Sardinian farm

For the other three farm types we structured a closed-cycle farm where the three age classes (weaning, growing and fattening) are reared in the same site. We assumed that the animals are grouped together based on the age classes. Within age classes the animals can contact each other animal (see Figure 3.6). We used a daily inflow for larger farms and weekly for the smaller, where the number of piglets per day resulted less than one.

We initialised the model with a mean population of N animals of ages distributed in the different age classes according to the average distribution indicated by national animal registry and slightly adjusted to mantain the poulation balanced. In particular, the distribution of the total population into the age classes were set to have balanced population, based on the calculated inflow. The proportion of initially infected and recovered animal per age class is reported in Table 3.9 and choosen according to literature data [49, 57]. The number of new piglets entering in the weaning phase daily is calculated based on the number of sow per farm n_{sow} and assuming a number of birth per sow per year $n_{delivery}$ equals to 2 and a mean litter size $litter_{size} = 10$, so we got the number of new piglets per

day or week

$$new_{piglets} = \frac{n_{delivery} \cdot n_{sow} \cdot litter_{size}}{t_{unit}}$$

where t_{unit} is equal 52 for weekly inflow and 365 per daily inflow. The birth rate therefore will be equals to

$$\nu = \frac{1}{new_{piglets}}.$$

The mean litter size is intended to reflect also the suckling mortality, it represents therefore the average number of animals weaned alive per sow.
For the Sardinian subtype the inflow was modelled differently. As soon as an animal is slaughtered, a weaner is introduced again in the farm.

As soon as an animal reaches the max number of day in an age class it is moved to the next age class or to slaughterhouse. The number of days spent in each class are the same for open and closed-cycle farm and are 30 days for weaners, 60 days for growers, 90 days for light fatteners and 150 for heavy fatteners. We simulated 700 days of farming activity for different simulations to account for randomicity. We simulated C type for 20 times and D and SAR for 50 times.

3.5 Statistical analysis

Statistical analysis were performed to evaluate differences between subtypes among the closed-cycle farms. We perform the ANOVA test on the total number of animal slaughtered while in the infectious or exposed compartment during the 700 days of simulation. We did preliminary test on the variance and distribution of the samples produced to decide which kind of ANOVA test to perform.

All the model simulations and statistical analysis have been carried out using the R software (version 4.1.2) [61]. Codes are available at https://github.com/Mezzanenne/HEV_Codes/tree/Farm-simulations.

3.6 Results

3.6.1 Parameter inference

The maximum likelihood search was performed for the $\beta = \lambda \cdot p$ and μ_{IR} parameters with starting point chosen at random from a $\mathbb{R}^2$ box $(\beta, \mu_{IR}) = [1, 80] \times [0.25, 1]$. The global search gave a maximum loglikelihood of -71 with a standard error of 0.000347 for the pair $(\beta, \mu_{IR}) = (151, 0.176)$. The simulated dynamics for the inference were set with weekly rates therefore we obtained a $\beta = 21$ and a $\mu_{IR} = 0.025$ corresponding to a mean infectious period of 40 days. The final results are plotted in Figure 3.8. The other parameters used for the model building were the same used for all the simulations. The

population size was set to 40 animals considering the size of a single pen and the mean latent period μ_{EI}^{-1} was set, according to [71] to 13 days, corresponding to $\mathbb{E}(T_{EI})$ as defined (see Table 3.6). Furthermore, we set an initial proportion of infected η_I equals to 0.02 and an initially proportion of recovered η_R equals to 0.06. The reporting rate ρ was set to 0.9 to account for possible measurement errors. The data available to perform the estimation were not enough to have a significant 95% confindence interval for the profile likelihood evaluated for the two parameters. The global search and profile likelihoods results are shown in Figure 3.8.

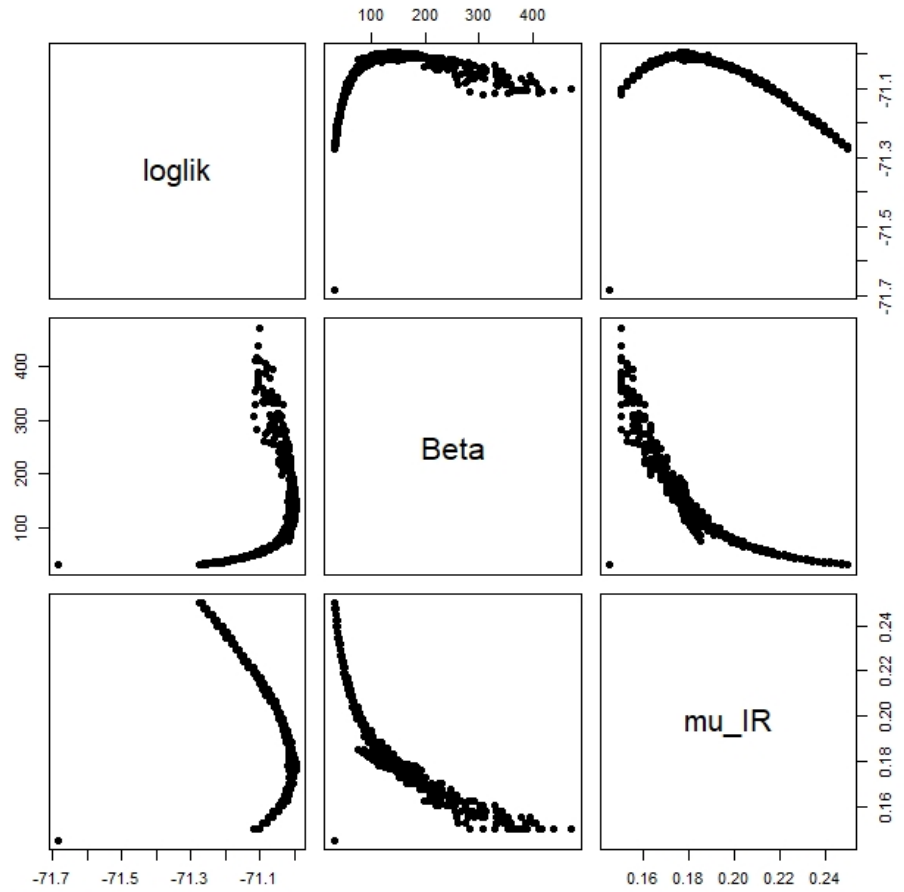

Figure 3.8. Final likelihood estimation for β and μ_{IR}

The original data from Casas [11] are reported in Table 3.5. The interpolated values used to feed the inference POMP model are shown in Figure 3.9.

3.6.2 Model simulations

The variables used for the simulations of each farm types are summarised in Table 3.7 and Table 3.8. For each type, we run several simulations to account for random variability, the number of simulations varied from a minimum of 20 to a maximum of 50 simulations.

Tabella 3.5: Mean IgM prevalence (%) over the all-in-all-out farms as reported in [11]

week	IgM prevalence (%)
3	0
7	25
13	89.3
18	61
25	68,7

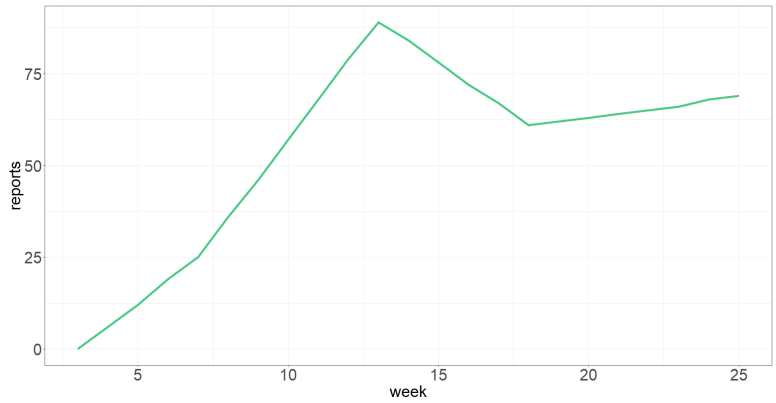

Figure 3.9. Interpolated values of IgM prevalence (%) from original data published by Casas and collegues [11]

Tabella 3.6: Model parameters

parameter	description	value	reference
$\beta = p\lambda$	no. infective contacts per day	21	section 3.6.1
μ_{IR}	recovery rate	0.025	
T_{EI}	latency period	$\sim Gamma(\alpha_{EI}, s_{EI})$	
α_{EI}	shape parameter of latency period	25.7	[71]
s_{EI}	scale parameter of latency period	0.5	

Open-cycle intensive farm

The specific parameters chosen for the simulations of the type A and B farms are reported in Table 3.7. In the type B farm we selected light and heavy pigs at random from the total population reared with probaility p_{light} and p_{heavy}, respectively.

In Figure 3.10 are shown the mean trajectories over the simulations of the total amount of animals per infective state up to day 150 of the simulation. The weeks on the x-axis represent the the simulation evolution and the age of the animals reared. We begin the simulations from week 5 that is the beginning of the weaning phase. The disease extincts in about 19 weeks. At the end of all the simulations (slaughter) the animals were all in the recovered compartment (for details see Figure 3.10).

The same pattern resulted from the type B farm. In Figure 3.11 are shown the tra-

Tabella 3.7: Parameter used for simulations of open-cycle farm types

farm type	parameter	description	value
A	N_{pens}	total number of pens simulated	85
	pr_type	farm production type	"h"
	η_I	fraction of initially infected animals	0.02
	η_R	fraction of initially recovered animals	0.06
B	N_{pens}	total number of pens simulated	50
	p_{light}	probability that a pen is for light production	0.5
	p_{heavy}	probability that a pen is for heavy production	0.5
	η_I	fraction of initially infected animals	0.02
	η_R	fraction of initially recovered animals	0.06

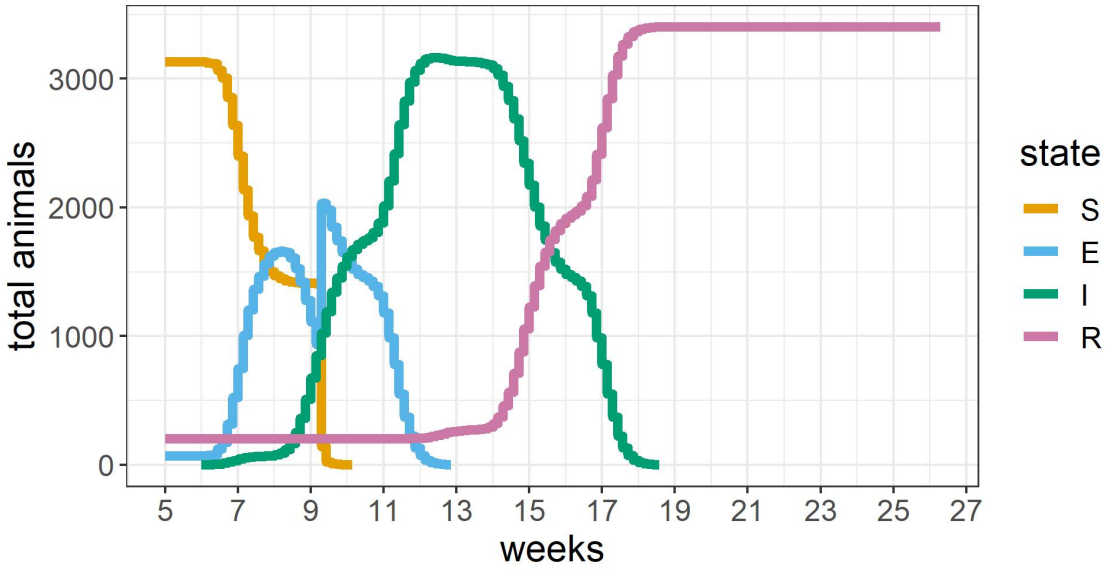

Figure 3.10. Trajectories for type A farm up to day 150. The weeks on the x-axis represent the the simulation evolution and the age of the animals reared

jectories for the type B farm. The lines are the average number of animal per week per epidemiological status while the bands represent minimum and maximum number of animal over the simulations performed. Even here the infection extincts around the 19th week therefore no differences resulted between the two production types (light and heavy).

To test how the β parameter affect the trajectories of the system, we repeated the one single simulation for the A type farm over different values of the parameter. In Figure 3.12 we reported only the lowest values tested. The transmission overall dynamics is affected only by values of β strictly under 0.4. The peak of infected increases with β until, for $\beta = 0.4$, the total number of infected animals is enough to extinct all the susceptibles.

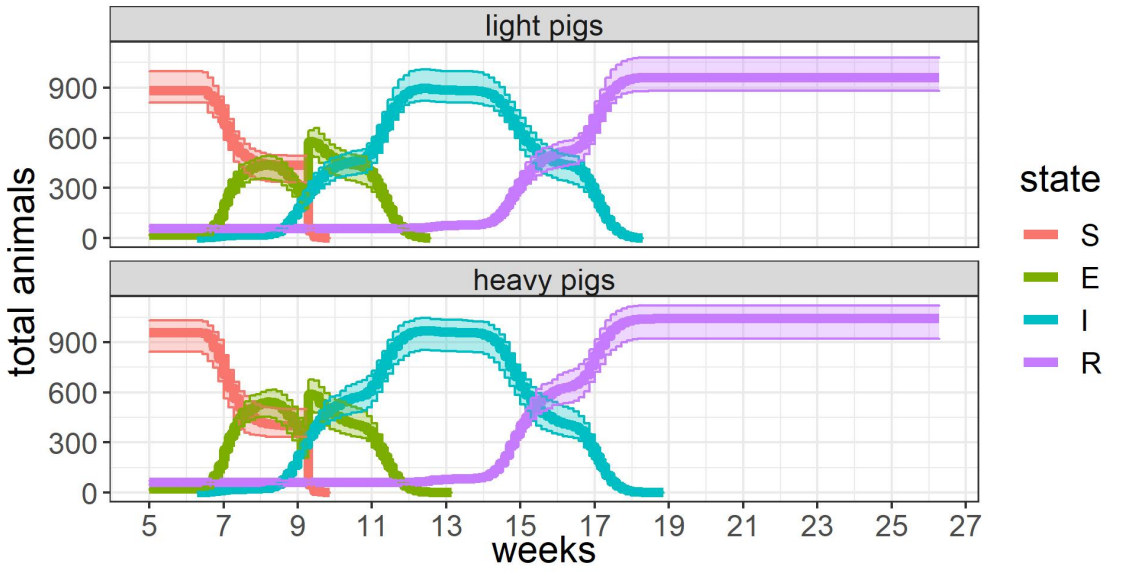

Figure 3.11. Trajectories for type B farm with minimum and maximum values over the five simulations up to day 150. The weeks on the x-axis represent the the simulation evolution and the age of the animals reared

Closed-cycle farm

The specific parameters used for the simulations of farms C, D and SAR are reported in Table 3.8, while in Table 3.9 are reported the fraction of initially infected and recovered animals per age class used to inizialise the different closed-cycle farm types. We refer to [49] and [57] to initialise infectious and recovered animals, respectively. To test the evolution of the process we repeated the simulations using different initial values for the epidemiological statuses.

We reported here for each farm type the smoothed trajectories for the 700 simulation days (see Figures 3.13, 3.15 and 3.17) and the average number of animals slaughtered while in the infectious or exposed compartment per day (see Figures 3.14, 3.16 and 3.18). Since for this type of farm on x-axis are represented the simulation days that do not correspond to the animal ages, we additionally reported the trajectories for each age class in Figures 3.19, 3.20 and 3.21.

The overall evolutions of the trajectories is quite similar for the three farm types, while the number of slaughtered animals varied a lot (see Figures 3.14, 3.16 and 3.18). For the first two types, the probability of slaughter an infected animal became very small after the 60th day, while for the SAR type this probability remains high for the entire simulation period. This difference does not reflect in the age class dynamic (see Figures 3.19, 3.20 and 3.21), suggests that the differences in the slaughtering of infected animals are driven mainly by the different slaughtering process modeled for the SAR farm type.

Furthermore, we reported the number of slaughtered animals in the infectious or exposed compartment per farm type per different values of $\beta > 1$ (see Figures 3.22, 3.23

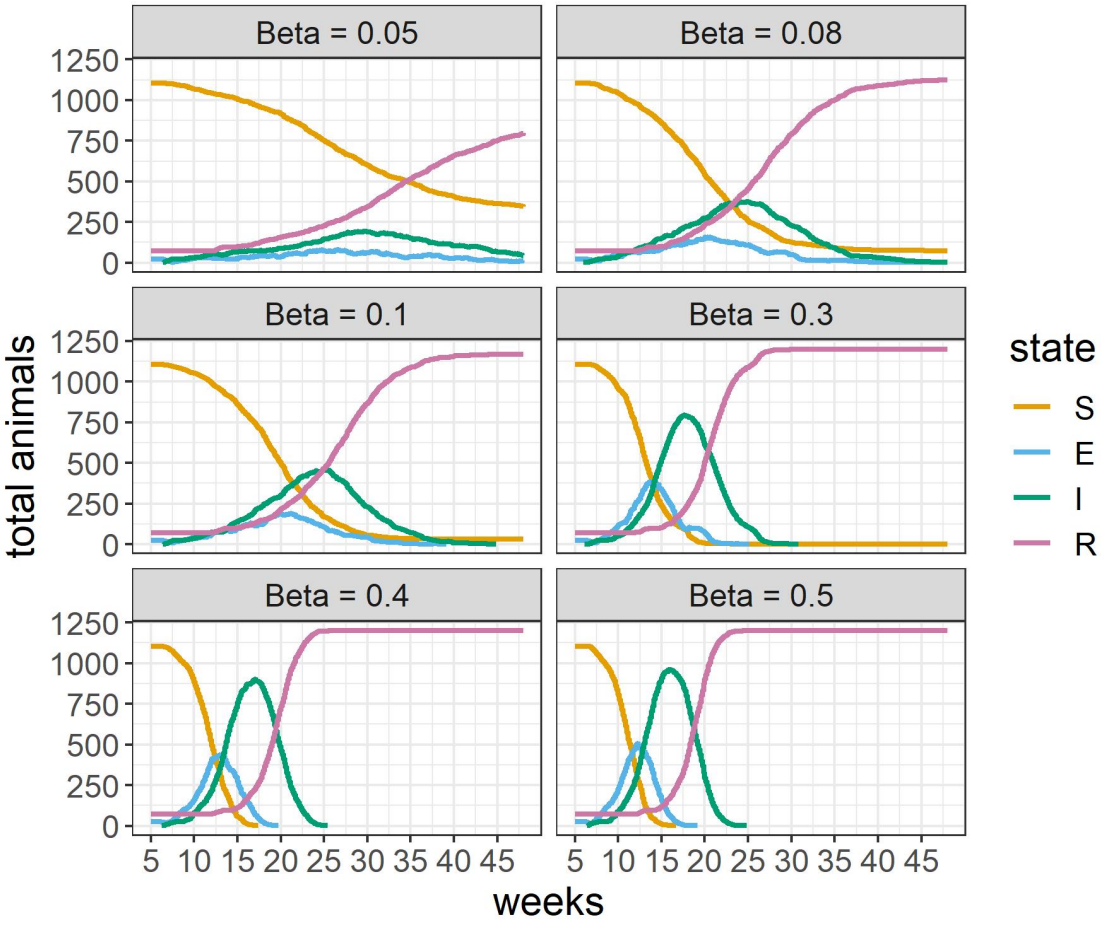

Figure 3.12. Transmission dynamics trajectories of the farm type A with varying $\beta < 1$

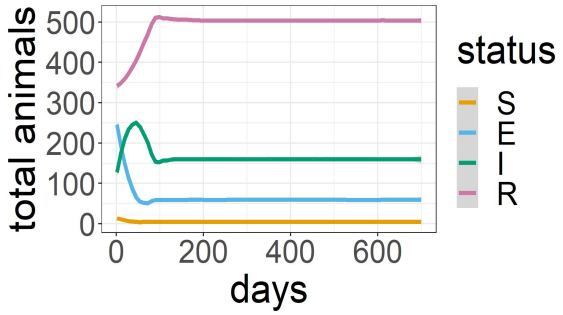

Figure 3.13. Smoothed trajectories for type C farm

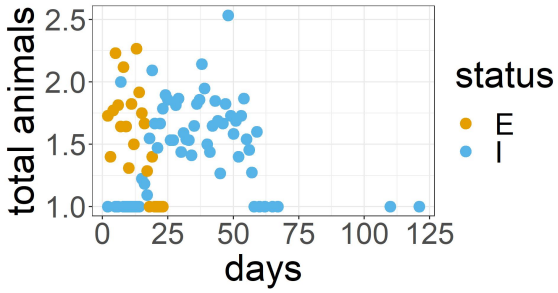

Figure 3.14. Number of infectious or exposed animal slaughtered in C farm type

and 3.24). The number of animals slaughtered while infected seems not to be highly dependent from the value of β. The higher variatons showed by the D farm type is possibly dependent on the effect of stochasticity on a smaller population size.

Tabella 3.8: Parameter used for simulations of closed-cycle farm types

farm type	parameter	description	value
C	N	total population simulated	728
	n_{sow}	total number of sow	76
	$new_{piglets}$	number of piglets introduced daily	4
	p_W	initial proportion of weaning pigs	0.33
	p_G	initial proportion of growing pigs	0.17
	p_F	initial proportion of fattening pigs	0.5
D	N	total population simulated	104
	n_{sow}	total number of sow	10
	$new_{piglets}$	number of piglets introduced weekly	4
	p_W	initial proportion of weaning pigs	0.17
	p_G	initial proportion of growing pigs	0.3
	p_F	initial proportion of fattening pigs	0.53
SAR	N	total population simulated	175
	n_{sow}	total number of sow	30
	$new_{piglets}$	number of piglets introduced daily	2
	p_W	initial proportion of weaning pigs	0.33
	p_G	initial proportion of growing pigs	0.27
	p_F	initial proportion of fattening pigs	0.5

Tabella 3.9: Initial values used for simulations of closed-cycle farm types

parameter	description	value	ref.
η_I^W	fraction of initially infected weaners	0.26	
η_I^G	fraction of initially infected growers	0.44	[49]
η_I^F	fraction of initially infected fatteners	0.09	
η_R^W	fraction of initially recovered weaners	0.24	
η_R^G	fraction of initially recovered growers	0.45	[57]
η_R^F	fraction of initially recovered fatteners	0.64	

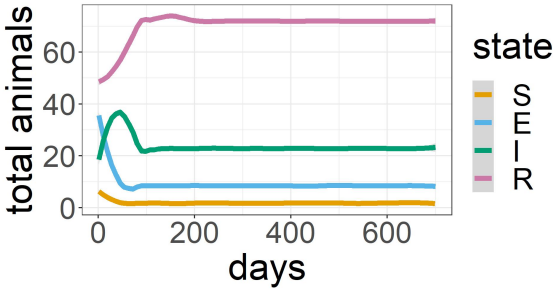

Figure 3.15. Smoothed trajectories for type D farm

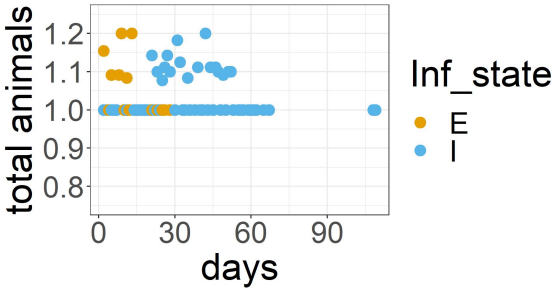

Figure 3.16. Number of infectious or exposed animal slaughtered in D farm type

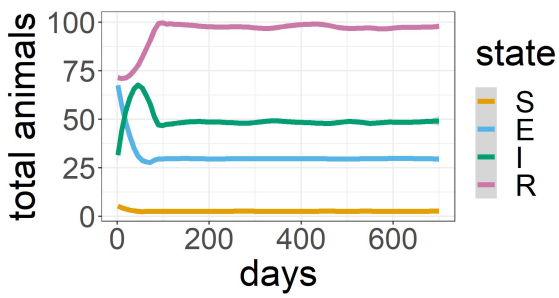

Figure 3.17. Smoothed trajectories for SAR farm type

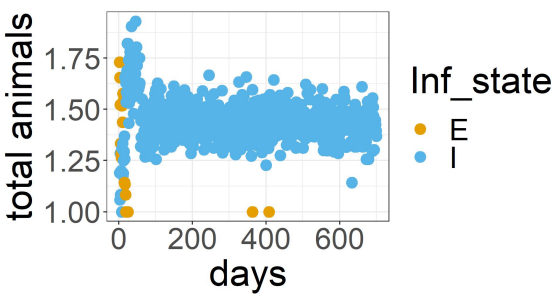

Figure 3.18. Number of infected or exposed animal slaughtered in SAR farm

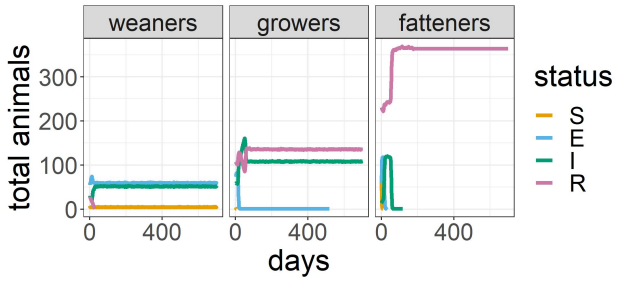

Figure 3.19. Trajectories for type C farm by age class

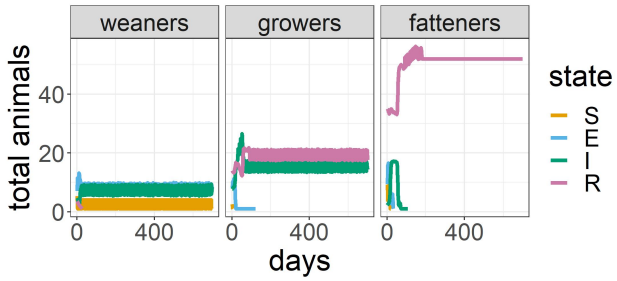

Figure 3.20. Trajectories for type D farm by age class

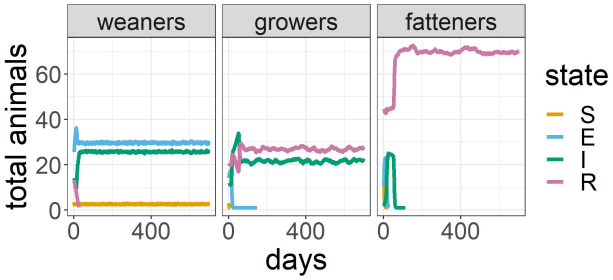

Figure 3.21. Trajectories for type SAR farm by age class

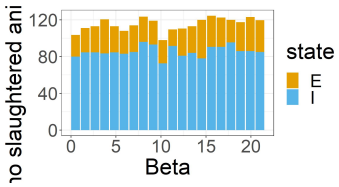

Figure 3.22. Total of slaughtered animal per farm type C with varied values of β.

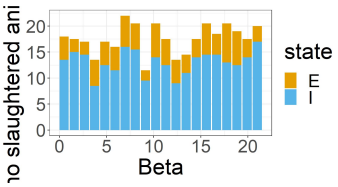

Figure 3.23. Total of slaughtered animal per farm type D with varied values of β.

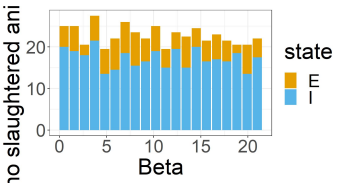

Figure 3.24. Total of slaughtered animal per farm type SAR with varied values of β.

3.6.3 Statistical analysis

To analyse differences between farm types we used the total number of slaughtered animals that were infected at the moment of slaughter per each simulation (data sheet is avai-

lable at https://github.com/Mezzanenne/HEV_Codes/tree/Statistical-analysis_closedcycle). Boxplots of the samples used are shown in Figure 3.25 where different scales were applied to allow better visualization. We first tested the normality of the samples using the Shapiro test. The

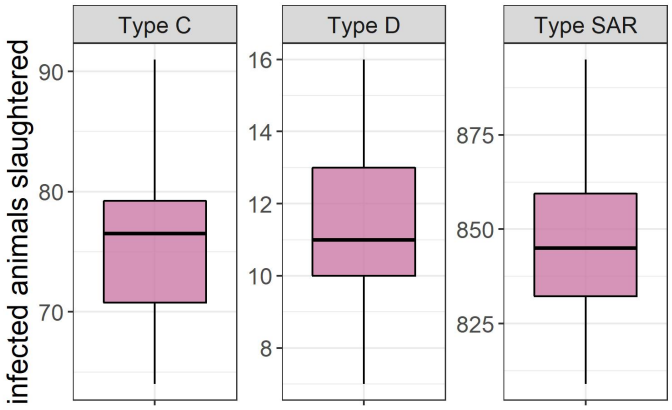

Figure 3.25. Boxplot representing the distributions of the number of infectious or exposed animals slaughtered per farm type for each simulation. Each plot has a different scale.

p-values obtained resulted all higher than 0.05 therefore we considered the samples normally distributed. We then used both the Bartlett and the Fligner tests to verify the homoscedasticity of the samples. Neither of the tests resulted significant, whence we used the ANOVA test for heteroscedastic samples.

We performed the Welch one way ANOVA test in R using the `welch_anova_test()` function from the `rstatix` package. The test resulted significant with very small p-values (see 3.10). Given

Tabella 3.10: Welch one way ANOVA test on the farm types C, D and SAR.

F statistic	DF	denom DF	p-value.
46410	2	39	$1.45 \cdot 10^{-66}$

the significance of the ANOVA test we perform a pairwise t test with Bonferroni family-error correction. The results show a significant difference between all the possible pairs of farm types, with all the p-values smaller than $2 \cdot 10^{-16}$.

3.7 Discussion

In our model, the HEV transmission dynamics resulted highly affected by the management process and the organization of the farm. These findings are, to some extent, in line with observational studies found in literature, although it is not always possible to compare data reported in literature with our findings when no details on the farm management or organization are reported. For both types of farm (open and closed cycle) the peak of HEV infected animals was reached in the growing phase, especially between the 8th and 15th weeks of age. This finding is consistent with several observational studies conducted in the Europe showing that the peak of HEV shedding is reached in animals aged between 3 and 4 month of age (see [70]) that exactly corresponds to the growing phase. In northern Italy's intensive farms, Caruso et al. [10] reported a seroprevalence in

weaning phase of 26.3% and a prevalence of HEV shedders in the same age class of 16%. Pavoni and colleagues [59] reported for growing animals reared in the same area, a prevalence of shedders of 27% for open-cycle farms. Other studies reported similar findings, with a seroprevalence of 12% for weaners, 30% for growers and 36% for fatteners. Especially for the first two age classes the magnitude orders seem to be comparable with our findings. A few details are available in the scientific literature for small farms located in center south of Italy. Costanzo and colleagues [14] reported of 75% of seroprevalence in animals < 6 months of age and of 96% for older animals and a prevalence of shedders equals to 6.8% in younger animals (<6 months of age) and 8% in older pigs.

Regarding the estimations on slaughtered animals, we showed that the probability to slaughter an infected animal in closed-cycle farms, more common in south Italy, is importantly higher respect to the open-cycle farms, more diffused in central and northern part of Italy, where we got no infected animal at the slaughterhouse. This result is corroborated by a recent study of Chelli and colleagues [12] conducted in several Italian slaughterhouses. According to this study, the location of the slaughterhouse and the origin of the animal represent a significant factor for the contamination of organic samples, especially for HEV RNA presence in faeces or liver. While positive samples of animals coming from north- and central-Italy farms are the 0.9 and 0.8%, respectively, animals coming from southern farms resulted positive for HEV RNA in liver and/or feces in 15.5% of the samples.

Parameter estimation

Given the lack of specific observational and experimental data on HEV pathogenesis and infection in pigs, the results of the parameter estimations are difficult to corroborate, which represent a limitation of our study. While the mean infectious period μ_{IR} has been estimated in different experimental studies, the β parameter reflects highly the model structure and it is indeed very difficult to compare it with others estimation for similar parameters. The infectious period estimated by our inference process equals to 40 days, which is very similar to the one estimated in [71]. In this study, pathogenesis of HEV in pigs were evaluated in presence of co-infection between hepatitis E and Porcine Reproductive and Respiratory Syndrome Virus (PRRSV), a very common virus in farmed pigs. In this study, Salines and colleagues estimated an infectious period of 48.6 days [95% CI 27.9 – 84.6 days], pointing out how co-infections could prolong the duration of HEV shedding. Since pathogens circulation in farms is quite common [21, 51] it seems reasonable to adopt a rate more similar to this one. It is important to underline that we did not take into account the variation of HEV quantity release during the period of shedding, since quantitative data are largely missed. Several studies on HEV shedding, based on detection of virus in fecal samples either at slaughterhouse in adult pigs [12, 21] or at farms in old sows [22] show that pathogenesis of HE in pigs is very complex and that the duration of shedding period could vary significantly. Unfortunately, the absence of strong quantitative data makes very challenging to study the shedding dynamics in pig population.

Open-cycle farms

The spread of HEV among pigs in open-cycle farms show the typical behavior observed in early phases of epidemics, when a new pathogen enters into a naïve population. In a first phase, the susceptible pigs progressively decrease while exposed and infectious animals increase. In a second phase, recovered animals start to increase when infected animals start to decrease. Overall, the main factor positively influencing the extinction of the infection is the lack of introduction of new animals (naïve) in the farm.

The infection rate per day is significantly influenced by the number of infectious individuals present because of the low size of the subpopulations. However, important modifications on β and on other characteristics of the farm seem not to be critical in the overall transmission dynamic. Significant changes in the dynamic are visible only for β below 1, exhibiting a good stability of the dynamics under perturbations. The use of a deterministic shedding period represents a possible additional factor of uncertainty in our study. As already pointed out, the duration of the shedding period seems to be highly variable and, given the short time simulated, it could represent a critical factor for the overall transmission dynamic.

Closed-cycle farms

For the closed-cycle farms simulation, the process seems to converge quickly to an equilibrium for all the three types of farms. We ran all the models from five different initial epidemiological statuses obtaining always the same long-term dynamic, with a larger variability in the D type farms that have the smallest population size.

The number of infected animal slaughtered per day is larger in the very first phase of the simulation, until the 60th day, especially for type C and D farms. After this period, as the trajectories became closer to the equilibrium, the number of HEV infected fatteners seems to reach zero, as slaughtered pigs also do. However, persistence of infected animals in growers and weaners compartment could represent a risk. Any slight change in the farm management, organization, etc. could indeed trigger the spread of the infection to the fatteners compartment.

In the Sardinian farming type, the number of animals slaughtered while infected is much higher compared to the other two types. Since the infection levels in the different age classes are comparable with the other two types, it seems reasonable that the main driver affecting the risk of slaughter infected animals is the slaughtering of pigs at a much younger age.

The statistical analysis results showed that the difference between the three farming types is significant. This possibly means that arrangements in the farming management could contribute importantly in the reduction of risk of HEV transmission in slaughterhouse and, consequently, in humans.

Limitations and future improvements

The study pointed out importantly how the differences in farming condition and pig production setting, housing and management of animals affect the transmission dynamics of the virus inside the farm and the probability to deliver HEV positive animals at slaughter. This is a very important finding because the level of HEV contamination of foodstuff at retail (i.e. at consumer level), depends only on the magnitude of the initial HEV contamination of pork meat and offal at slaughtering. Downstream of slaughterhouse HEV contamination in food chain can only be mitigated by the adoption of adequate food manufacturing practice that should be part of HACCP[1] programs of food business operators (FBOs). No information is available to us about how FBOs actually implement their HACCP programs to control the risk of HEV contamination of pork meat. The possibility to control the risk of HEV contamination of food at the consumer level through the adoption of good farming practice at primary level seems therefore a very interesting perspective from the public health point of view. Unfortunately, the lack of detailed information on internal farm organization, management practice and the high heterogeneity of pig farming system in Italy did not allow to also introducing these components in our model. However, understanding how animal management practices are effective in reducing the risk of HEV prevalence at slaughter in

[1] Hazard analysis and critical control points.

different type of farms is an attractive perspective for HEV control. In addition, slaughterhouse HEV contamination data could be used to fill the modelling gap between production stage and human exposure. This would be a crucial step to strengthen the estimation of the actual exposure of human population to hepatitis E. Data on HEV contamination at slaughterhouse including details on carcasses origin, location, characteristics, etc. are difficult to collect and, to date, are just a few [12].

Data on pathogenesis represented an additional challenge for the study, especially for the parameter estimation. To better differentiate the farm transmission dynamics, based on the farm types, it would be important to strongly estimate a different λ parameter at least for the two macro-types, representing the average number of contact per day. To this purpose, good longitudinal data for the specific farm types are needed. Due to the lack of these data we directly estimated $\beta = p \cdot \lambda$. We have neither the possibility to represent these two different quantities nor to specify λ for each farm category.

An expected improvement for the modelling that we proposed in this study would be to fine-tune and test possible within-farm intervention strategies, based on animal group segregation policies, manure managements, biosanitary measures in particular for light pigs and piglets productions.

Appendice A

Uncertainty analysis

A.1 Standard deviations of input parameters

Tabella A.1: Uncertainty analysis for input parameters

Parameter	**Standard deviation**
λ_{PL}	$7.76 \cdot 10^{-6}$
$cons_{day}^{PL}$	$2.14 \cdot 10^{-2}$
λ_{PNL}	$1 \cdot 10^{-4}$
$cons_{day}^{PNL}$	$2.5 \cdot 10^{-1}$
λ_{SH}	$3 \cdot 10^{-2}$
$cons_{day}^{SH}$	1.21
μ	$7 \cdot 10^{-8}$

A.2 Parameter correlations

We report the correlation test results and the scatterplots between them for the three categories involved in the analysis, meaning PL, PNL, and SH.

We can see from tables and plots that some of the parameters show a certain structure in plots especially for PL category, even if the numbers seem rather small. However, we considered parameter uncorrelated given the small correlation coefficients.

Tabella A.2: Parameter correlations for category PL

Parameters	Estimate	95%CI	p-value
$\lambda_{PL}\|\mu$	-0.018	$[-0.03 - 0.001]$	0.06
$\mu\|cons_{day}^{PL}$	0.01	$[-0.0095 - 0.03]$	0.3
$cons_{day}^{PL}\|\lambda_{PL}$	-0.0056	$[-0.025 - 0.014]$	0.5

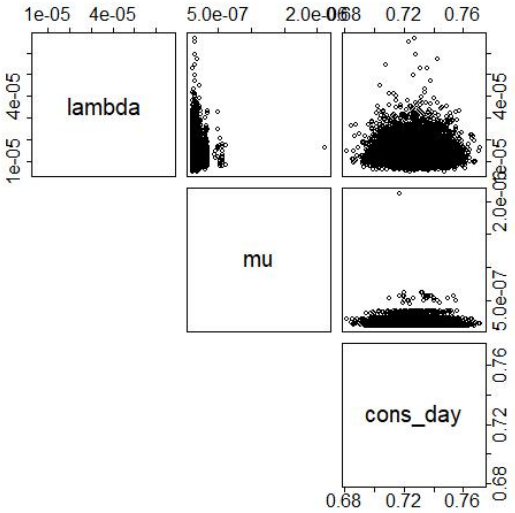

Figure A.1. Scatterplots paired between input parameters for category PL.

Tabella A.3: Parameter correlations for category PNL.

parameters	estimate	95%CI	p-value
$\lambda_{PNL}\|\mu$	-0.0043	$[-0.023 - 0.015]$	0.06
$\mu\|cons_{day}^{PNL}$	-0.02	$[-0.021 - 0.017]$	0.82
$cons_{day}^{PNL}\|\lambda_{PNL}$	-0.007	$[-0.026 - 0.012]$	0.46

Tabella A.4: Parameter correlations for category SH.

parameters	estimate	95%CI	p-value
$\lambda_{SH}\|\mu$	-0.014	$[-0.034 - 0.0048]$	0.13
$\mu\|cons_{day}^{SH}$	0.016	$[-0.002 - 0.03]$	0.09
$cons_{day}^{SH}\|\lambda_{SH}$	0.0001	$[-0.019 - 0.019]$	0.9

A.3 Parameter impact

As explained above, we can use the SRCs to rank the parameters used in the analysis based on their effect on the output. Our linear regression with SRCs resulted in the following expression for the two categories:

$$y_{PL} = -0.864\,\lambda_{PL} + 0.865\,cons_{day}^{PL} + 0.99\,\mu \tag{A.1}$$

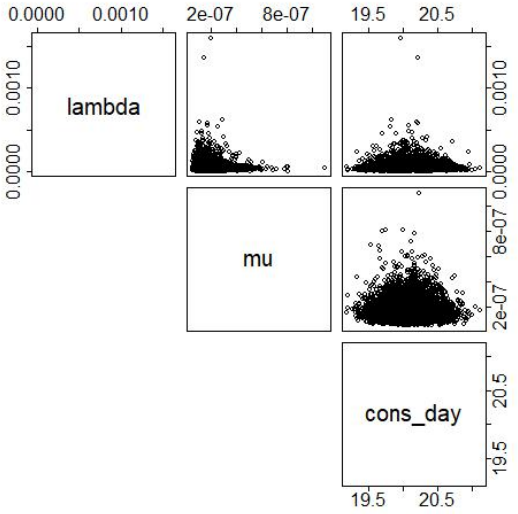

Figure A.2. Scatterplots paired between input parameters for category PNL.

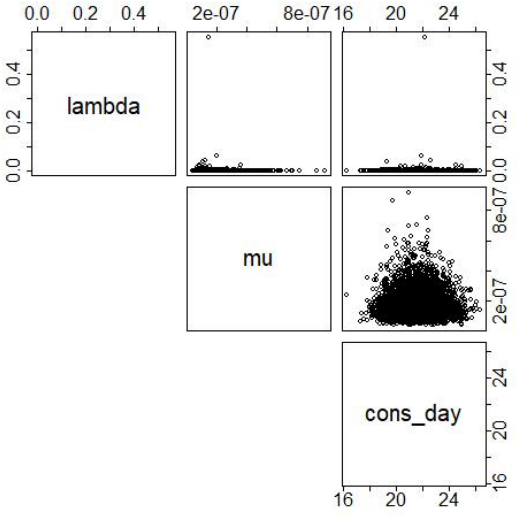

Figure A.3. Scatterplots paired between input parameters for category SH.

$$y_{PNL} = -0.9\,\lambda_{PNL} + 0.88\,cons_{day}^{PNL} + 0.995\,\mu \tag{A.2}$$

Appendice B

Risk matrix

We report here a risk matrix to qualitatively score the risk of each food category, based on the severity and likelihood of food contamination [62]. We are here considering the parameters reported in Table 2.4.

We defined the *severity* as the expected HEV genome equivalents per serving $\mathbb{E}(C_1^i) = \lambda_i^{-1}$, transformed in log_e. Severity scores definition are reported in Table B.1. The *likelihood* is defined as the fraction of contaminated servings consumed in one year. Meaning the number of servings consumed in a year d_i^a times the prevalence of HEV contaminated food samples α_i. Likelihood scores are displayed in Table B.1.

Tabella B.1: Severity (left side) and likelihood (right side) scores definition

$log_e\Big(\mathbb{E}(C_1^i)\Big)$	score	$d_i^a \cdot \alpha_i$	score
< 9	1	< 0.2	1
$9 - 11$	2	$0.2 - 0.5$	2
$11 - 13$	3	$0.5 - 1$	3
> 13	4	> 1	4

Based on the definitions and on the data sheet reported in Table B.2, we can calculate the risk for each of the food categories as *risk = likelihood x severity* (see Table B.2) and assign to each category a qualitative risk level, as illustrated in Figure B.1.

Tabella B.2: Data sheet and scores for each food category

Category	$log_e\Big(\mathbb{E}(C_1^i)\Big)$	$d_i^a \cdot \alpha_i$	Severity	Likelihood	**risk**
PL	11.2	0.33	3	2	**6 (medium)**
PNL	10.4	2	2	4	**8 (medium)**
SH	11.3	0.124	3	1	**3 (low)**
ML	0	0	0	0	**0**
VGT	0	0	0	0	**0**

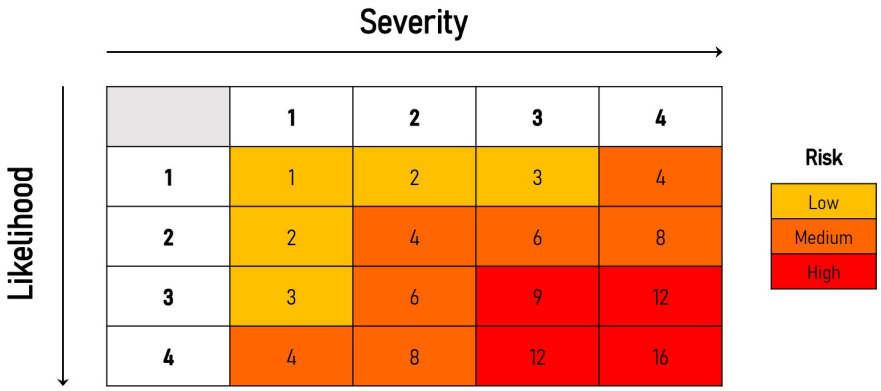

Figure B.1. Risk matrix

Appendice C

Farm class definition

Breeding for reproduction

The production cycle related to this type of breeding involves the following operations:

1. reception of gilts;
2. inseminations, gestation and delivery;
3. shipping-transfer to fattening.

Gilt receipt

The replacement gilts are purchased from farms specialized in the selection of breeding animals or are selected within the company. The gilts selected within the company are gathered in the heat waiting area ready for insemination.

Delivery

The delivery room is a room equipped with several cages capable of hosting the sow and the litter until the end of the suckling period (21 - 28 days). The sow usually gives birth without the need for assistance from the operator, who intervenes only in case of problematic farrowing. The sow in this phase is particularly aggressive and protective of the litter so any intervention by the operator to assist the sow or piglets can constitute a risk factor. It is clear that in this area of the farm it is necessary to implement a correct hygiene practice with cleaning and disinfection of the structures when the sows leave the farrowing room to return to the gestation sector (for the stimulation of a new heat).

Weaning of piglets

Piglets are removed from their mother around days 28 (7 kg) and are generally transferred to the weaning room or it is the farrowing room itself which, once the sow has been removed, fulfills this function. The weaned piglets, which have reached a weight of 30 - 40 kg depending on the farm management, are then ready to be shipped and / or transferred to the fattening sector.

Fattening breeding

Reception pigs

The pigs are transferred from the weaning sector to the fattening sector or arrive at the farm from other farms. Piglets are destined within the shelters where they will be reared up to the expected slaughter weight (160 kg for the production of ham or up to a weight of 100 - 120 kg for the butchery pig).

The pigs are divided into more or less homogeneous groups according to age and weight, and therefore confined to the pens. After unloading, the truck reaches the washing area where it is washed and disinfected before carrying out another load of animals.
Food management almost everywhere involves liquid feeding with automated distribution to troughs. During the fattening cycle, in some situations, the operators usually isolate the best subjects from the different boxes in order to group them in homogeneous boxes. This operation, known as equalization, is usually carried out a couple of times at the beginning of the first lean phase (weight loss 50 - 60 kg) and at the beginning of the fattening phase (80 - 100 kg body weight).
Once the expected slaughter weight has been reached, the pigs are loaded truck for transport to the slaughterhouse.

C.1 Demographic data analysis

We classified the farms starting from the official categorization and taking into account the production type, meaning open cycle fattening, open cycle reproduction, closed cycle. Pig farms with less than 20 pigs were excluded from the analysis. Factors considered for this purpose included:

- Farm size
- Relative distribution of reared animals by animal type
- Farming activity and production stage
- Location of the farm

A Fisher exact test were performed over the classes defined to look for significant differences between Italian areas.
All the analysis were carried out with R software (vers. 4.1.2).

The classes defined are nine and are described, base on the distribution of animal type, in the Figure C.1. The nine farm husbandry classes defined, as shown in Figure C.1, are summarizable as follow

A Growing only

B Mainly fattening

C Mainly breeding

D Other types

E Weaning only

F Breeding boar only

G Mainly sow

H Sows only

I Gilt only

We observed a heterogeneous distribution of pig population and type of farm across the country. In the northern area the majority of pigs are reared (50% only in region Lombardia) with a mean number of animals per farm of 1439 (median = 979, IQR = 1511). Here the production seems to be more oriented towards the fattening steps with a majority of husbandry farm from classes A, B and

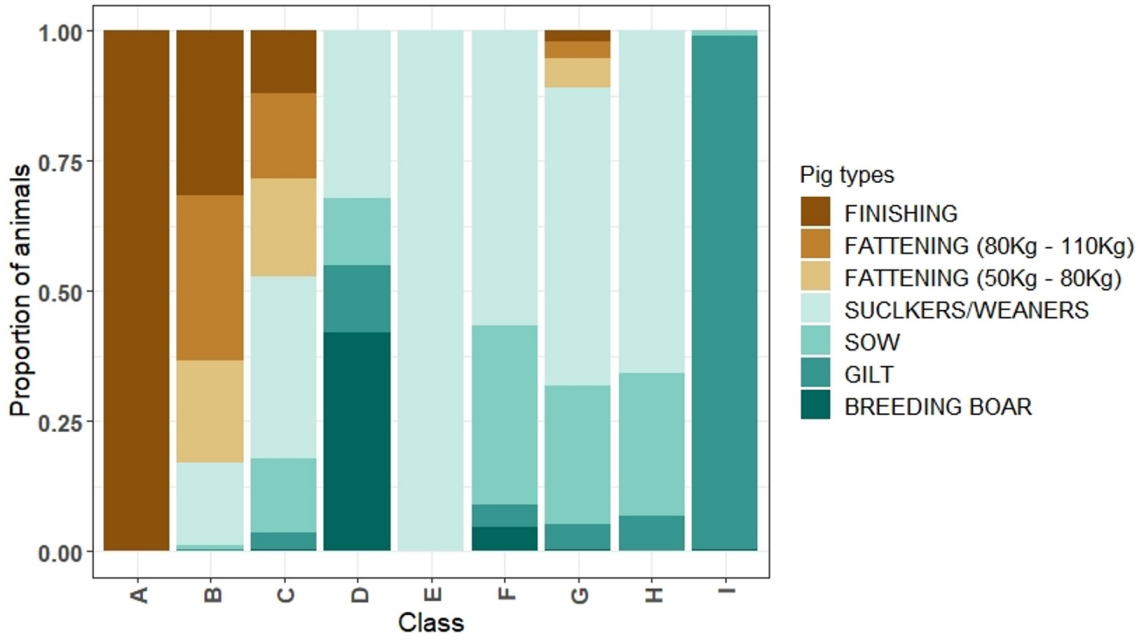

Figure C.1. Animal type distribution per farm class defined

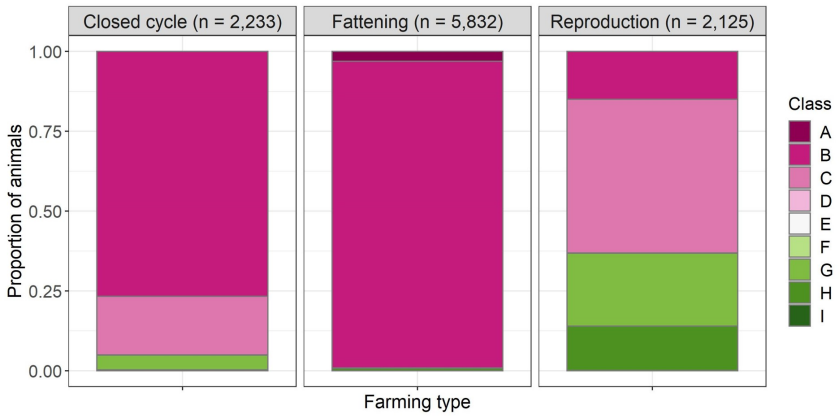

Figure C.2. Distribution of pig husbandry farm classes by production type (as defined in National Animal Registry)

C. In central and southern regions the distribution of animals is more uniform over all husbandry farm classes in smaller farms. Here mean reared animals per farm are 301 in the center (median = 106, IQR = 328) and 210 in south (median = 54, IQR 164). Islands seems to be the most peculiar areas with a consistent number of closed cycle farm distributed mostly over the husbandry farm classes C and G. Here the mean number of reared animals is just 71 per farm (med = 28, IQR = 22).

We performed a Fisher exact test on farm numbers over macro-areas in pairs and all of the tests result in a significant difference between areas with p-values < 0.0001

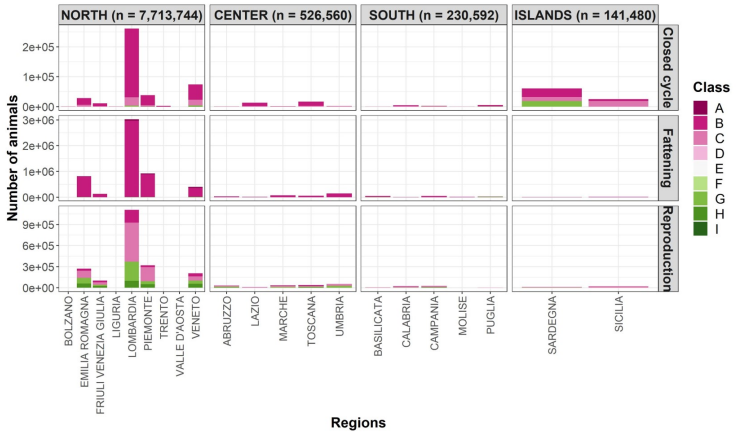

Figure C.3. Distribution of animals by pig husbandry farm classes, farm production type, macro-area and regions

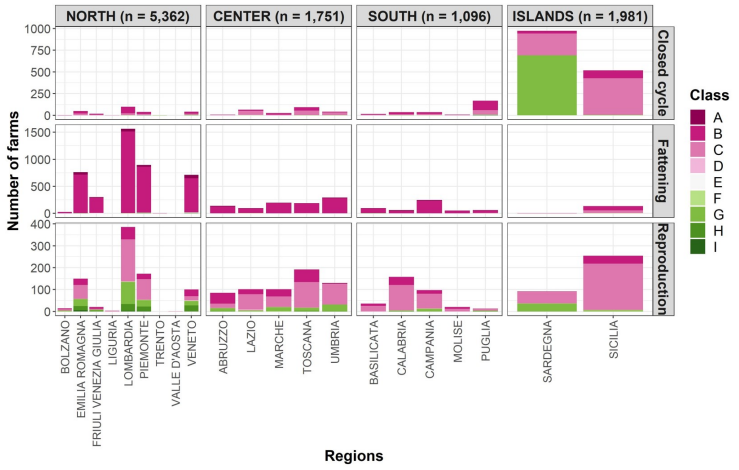

Figure C.4. Distribution of farm by pig husbandry farm classes, production type, macro-area and regions

References

[1] N. Albinana-Gimenez, M. P. Miagostovich, B. Calgua, J. M. Huguet, L. Matia, and R. Girones. Analysis of adenoviruses and polyomaviruses quantified by qpcr as indicators of water quality in source and drinking-water treatment plants. *Water research*, 43(7):2011–2019, 2009.

[2] V. Alfonsi, L. Romanò, A. R. Ciccaglione, G. La Rosa, R. Bruni, A. Zanetti, S. Della Libera, M. Iaconelli, P. Bagnarelli, M. R. Capobianchi, et al. Hepatitis e in italy: 5 years of national epidemiological, virological and environmental surveillance, 2012 to 2016. *Eurosurveillance*, 23(41):1700517, 2018.

[3] H. Andersson and T. Britton. *Stochastic epidemic models and their statistical analysis*, volume 151. Springer Science & Business Media, 2012.

[4] C. Baechlein and P. Bacher. No evidence for zoonotic hepatitis e virus infection through dairy milk in germany. *Hepatology*, 65(1):394–395, 2017.

[5] E. Barnaud, S. Rogée, P. Garry, N. Rose, and N. Pavio. Thermal inactivation of infectious hepatitis e virus in experimentally contaminated food. *Applied and environmental microbiology*, 78(15):5153–5159, 2012.

[6] K. Borgen, T. Herremans, E. Duizer, H. Vennema, S. Rutjes, A. Bosman, A. M. de Roda Husman, and M. Koopmans. Non-travel related hepatitis e virus genotype 3 infections in the netherlands; a case series 2004–2006. *BMC infectious diseases*, 8(1):1–10, 2008.

[7] M. Bouwknegt, K. Frankena, S. A. Rutjes, G. J. Wellenberg, A. M. de Roda Husman, W. H. van der Poel, and M. C. de Jong. Estimation of hepatitis e virus transmission among pigs due to contact-exposure. *Veterinary research*, 39(5):1, 2008.

[8] J. Brassard, M.-J. Gagné, M. Généreux, and C. Côté. Detection of human food-borne and zoonotic viruses on irrigated, field-grown strawberries. *Applied and environmental microbiology*, 78(10):3763–3766, 2012.

[9] B. Cacopardo, R. Russo, W. Preiser, F. Benanti, G. Brancati, and A. Nunnari. Acute hepatitis e in catania (eastern sicily) 1980–1994. the role of hepatitis e virus. *Infection*, 25(5):313–316, 1997.

[10] C. Caruso, S. Peletto, A. Rosamilia, P. Modesto, L. Chiavacci, B. Sona, F. Balsamelli, V. Ghisetti, P. Acutis, G. Pezzoni, et al. Hepatitis e virus: A cross-sectional serological and virological study in pigs and humans at zoonotic risk within a high-density pig farming area. *Transboundary and emerging diseases*, 64(5):1443–1453, 2017.

[11] M. Casas, R. Cortés, S. Pina, B. Peralta, A. Allepuz, M. Cortey, J. Casal, and M. Martín. Longitudinal study of hepatitis e virus infection in spanish farrow-to-finish swine herds. *Veterinary microbiology*, 148(1):27–34, 2011.

[12] E. Chelli, E. Suffredini, P. De Santis, D. De Medici, S. Di Bella, S. D'Amato, F. Gucciardi, A. Guercio, F. Ostanello, V. Perrone, et al. Hepatitis e virus occurrence in pigs slaughtered in italy. *Animals*, 11(2):277, 2021.

[13] P. Colson, M. Kaba, E. Bernit, A. Motte, and C. Tamalet. Hepatitis e associated with surgical training on pigs. *The Lancet*, 370(9591):935, 2007.

[14] N. Costanzo, E. Sarno, V. Peretti, L. Ciambrone, F. Casalinuovo, and A. Santoro. Serological and molecular investigation of swine hepatitis e virus in pigs raised in southern italy. *Journal of food protection*, 78(11):2099–2102, 2015.

[15] A. De Schryver, K. De Schrijver, G. François, R. Hambach, M. van Sprundel, R. Tabibi, and C. Colosio. Hepatitis e virus infection: an emerging occupational risk? *Occupational medicine*, 65(8):667–672, 2015.

[16] G. Deest, L. Zehner, E. Nicand, C. Gaudy-Graffin, A. Goudeau, and Y. Bacq. Autochthonous hepatitis e in france and consumption of raw pig meat. *Gastroenterologie clinique et biologique*, 31(12):1095–1097, 2007.

[17] I. dei Servizi per il Mercato Agricolo Alimentare (ISMEA). Allevamento suino. report economico finanziario [in italian]. Monetary report, Istituto dei Servizi per il Mercato Agricolo Alimentare (ISMEA), 2008. https://www.ismea.it/flex/files/D.b3aef742c589d5fca4f6/report_suino.pdf.

[18] I. dei Servizi per il Mercato Agricolo Alimentare (ISMEA). Scheda di settore. settore suinicolo [in italian]. Data sheet, Istituto dei Servizi per il Mercato Agricolo Alimentare (ISMEA), 2021. https://www.ismeamercati.it/carni/carne-suina-salumi.

[19] I. Z. S. dell'Abruzzo e del Molise «G. Caporale». National animal registry. https://www.vetinfo.it/j6_statistiche/index.html#/. Accessed: 16/04/2021.

[20] I. Di Bartolo, G. Angeloni, E. Ponterio, F. Ostanello, and F. M. Ruggeri. Detection of hepatitis e virus in pork liver sausages. *International journal of food microbiology*, 193:29–33, 2015.

[21] I. Di Bartolo, M. Diez-Valcarce, P. Vasickova, P. Kralik, M. Hernandez, G. Angeloni, F. Ostanello, M. Bouwknegt, D. Rodríguez-Lázaro, I. Pavlik, et al. Hepatitis e virus in pork production chain in czech republic, italy, and spain, 2010. *Emerging infectious diseases*, 18(8):1282, 2012.

[22] I. Di Bartolo, F. Martelli, N. Inglese, M. Pourshaban, A. Caprioli, F. Ostanello, and F. M. Ruggeri. Widespread diffusion of genotype 3 hepatitis e virus among farming swine in northern italy. *Veterinary microbiology*, 132(1-2):47–55, 2008.

[23] S. Di Pasquale, P. De Santis, G. La Rosa, K. Di Domenico, M. Iaconelli, G. Micarelli, E. Martini, S. Bilei, D. De Medici, and E. Suffredini. Quantification and genetic diversity of hepatitis e virus in wild boar (sus scrofa) hunted for domestic consumption in central italy. *Food microbiology*, 82:194–201, 2019.

[24] M. Diez-Valcarce, P. Kokkinos, K. Söderberg, M. Bouwknegt, K. Willems, A. M. de Roda-Husman, C.-H. von Bonsdorff, M. Bellou, M. Hernández, L. Maunula, et al. Occurrence of human enteric viruses in commercial mussels at retail level in three european countries. *Food and Environmental Virology*, 4(2):73–80, 2012.

[25] V. Doceul, E. Bagdassarian, A. Demange, and N. Pavio. Zoonotic hepatitis e virus: classification, animal reservoirs and transmission routes. *Viruses*, 8(10):270, 2016.

[26] D. Donia, M. C. Dell'Amico, A. R. Petrinca, I. Martinucci, M. Mazzei, F. Tolari, and M. Divizia. Presence of hepatitis e rna in mussels used as bio-monitors of viral marine pollution. *Journal of virological methods*, 186(1-2):198–202, 2012.

[27] B. Efron and R. J. Tibshirani. *An introduction to the bootstrap*. CRC press, 1994.

[28] A. Feagins, T. Opriessnig, D. Guenette, P. Halbur, and X. Meng. Inactivation of infectious hepatitis e virus present in commercial pig livers sold in local grocery stores in the united states. *International journal of food microbiology*, 123(1-2):32–37, 2008.

[29] H. Fenaux, M. Chassaing, S. Berger, C. Gantzer, I. Bertrand, and E. Schvoerer. Transmission of hepatitis e virus by water: An issue still pending in industrialized countries. *Water research*, 151:144–157, 2019.

[30] A. Garbuglia, A. Alessandrini, N. Pavio, S. Tesse, S. Grignolo, C. Viscoli, D. Lapa, and M. Capobianchi. Male patient with acute hepatitis e in genoa, italy: figatelli (pork liver sausage) as probable source of the infection. *Clinical Microbiology and Infection*, 21(1):e4–e6, 2015.

[31] C. Geffray, A. Gerschenfeld, P. Kudinov, I. Mickus, M. Jeltsov, K. Kööp, D. Grishchenko, and D. Pointer. 8 - verification and validation and uncertainty quantification. In F. Roelofs, editor, *Thermal Hydraulics Aspects of Liquid Metal Cooled Nuclear Reactors*, pages 383–405. Woodhead Publishing, 2019.

[32] Y. Geng, C. Zhao, W. Huang, X. Wang, Y. Xu, D. Wu, Y. Du, H. Liu, and Y. Wang. Hepatitis e virus was not detected in feces and milk of cows in hebei province of china: No evidence for hev prevalence in cows. *International journal of food microbiology*, 291:5–9, 2019.

[33] Y. Guillois, F. Abravanel, T. Miura, N. Pavio, V. Vaillant, S. Lhomme, F. S. Le Guyader, N. Rose, J.-C. Le Saux, L. A. King, et al. High proportion of asymptomatic infections in an outbreak of hepatitis e associated with a spit-roasted piglet, france, 2013. *Clinical Infectious Diseases*, 62(3):351–357, 2016.

[34] M. S. Hakim, W. Wang, W. M. Bramer, J. Geng, F. Huang, R. A. de Man, M. P. Peppelenbosch, and Q. Pan. The global burden of hepatitis e outbreaks: a systematic review. *Liver International*, 37(1):19–31, 2017.

[35] C. Henneùchart-Collette, S. Martin-Latil, A. Fraisse, and S. Perelle. Comparison of three extraction methods to detect noroviruses in dairy products. *Food Microbiology*, 61:113–119, 2017.

[36] A. A. Hill, R. R. Simons, L. Kelly, and E. L. Snary. A farm transmission model for salmonella in pigs, applicable to eu member states. *Risk Analysis*, 36(3):461–481, 2016.

[37] F. Huang, Y. Li, W. Yu, S. Jing, J. Wang, F. Long, Z. He, C. Yang, Y. Bi, W. Cao, et al. Excretion of infectious hepatitis e virus into milk in cows imposes high risks of zoonosis. *Hepatology*, 64(2):350–359, 2016.

[38] M. Iaconelli, G. Purpari, S. Della Libera, S. Petricca, A. Guercio, A. Ciccaglione, R. Bruni, S. Taffon, M. Equestre, M. Fratini, et al. Hepatitis a and e viruses in wastewaters, in river waters, and in bivalve molluscs in italy. *Food and Environmental Virology*, 7(4):316–324, 2015.

[39] N. Kamar, J. Izopet, and H. R. Dalton. Chronic hepatitis e virus infection and treatment. *Journal of clinical and experimental hepatology*, 3(2):134–140, 2013.

[40] N. Kamar, J. Izopet, N. Pavio, R. Aggarwal, A. Labrique, H. Wedemeyer, and H. R. Dalton. Hepatitis e virus infection. *Nature reviews Disease primers*, 3(1):1–16, 2017.

[41] A. A. King, D. Nguyen, and E. L. Ionides. Statistical inference for partially observed markov processes via the r package pomp. *arXiv preprint arXiv:1509.00503*, 2015.

[42] P. Kokkinos, I. Kozyra, S. Lazic, M. Bouwknegt, S. Rutjes, K. Willems, R. Moloney, A. M. de Roda Husman, A. Kaupke, E. Legaki, et al. Harmonised investigation of the occurrence of human enteric viruses in the leafy green vegetable supply chain in three european countries. *Food and environmental virology*, 4(4):179–191, 2012.

[43] S. Kumar, S. Subhadra, B. Singh, and B. Panda. Hepatitis e virus: the current scenario. *International Journal of Infectious Diseases*, 17(4):e228–e233, 2013.

[44] G. La Rosa, M. Pourshaban, M. Iaconelli, V. S. Vennarucci, and M. Muscillo. Molecular detection of hepatitis e virus in sewage samples. *Applied and environmental microbiology*, 76(17):5870–5873, 2010.

[45] G. La Rosa, Y. Proroga, D. De Medici, F. Capuano, M. Iaconelli, S. Della Libera, and E. Suffredini. First detection of hepatitis e virus in shellfish and in seawater from production areas in southern italy. *Food and environmental virology*, 10(1):127–131, 2018.

[46] C. Leclercq, D. Arcella, R. Piccinelli, S. Sette, C. Le Donne, et al. The italian national food consumption survey inran-scai 2005–06: main results in terms of food consumption. *Public health nutrition*, 12(12):2504–2532, 2009.

[47] H. Lewis, O. Wichmann, and E. Duizer. Transmission routes and risk factors for autochthonous hepatitis e virus infection in europe: a systematic review. *Epidemiology & Infection*, 138(2):145–166, 2010.

[48] L. Marletta, E. Camilli, G. Catasta, V. de Balzo, L. Lucchini, S. Marconi, D. Martone, M. Matter, R. Piccinelli, L. Scalfi, M. Ticca, E. Toti, and E. Troiano. Linee guida per una sana alimentazione. Technical report, Centro di Ricerca Alimenti e Nutrizione (CREA), 2017.

[49] C. McCreary, F. Martelli, S. Grierson, F. Ostanello, A. Nevel, and M. Banks. Excretion of hepatitis e virus by pigs of different ages and its presence in slurry stores in the united kingdom. *Veterinary Record*, 163(9):261–265, 2008.

[50] X.-J. Meng, R. H. Purcell, P. G. Halbur, J. R. Lehman, D. M. Webb, T. S. Tsareva, J. S. Haynes, B. J. Thacker, and S. U. Emerson. A novel virus in swine is closely related to the human hepatitis e virus. *Proceedings of the National Academy of Sciences*, 94(18):9860–9865, 1997.

[51] M. Monini, I. Di Bartolo, G. Ianiro, G. Angeloni, C. F. Magistrali, F. Ostanello, and F. M. Ruggeri. Detection and molecular characterization of zoonotic viruses in swine fecal samples in italian pig herds. *Archives of virology*, 160(10):2547–2556, 2015.

[52] S. Montagnaro, C. De Martinis, S. Sasso, R. Ciarcia, S. Damiano, L. Auletta, V. Iovane, T. Zottola, and U. Pagnini. Viral and antibody prevalence of hepatitis e in european wild boars (sus scrofa) and hunters at zoonotic risk in the latium region. *Journal of Comparative Pathology*, 153(1):1–8, 2015.

[53] O. Moro, E. Suffredini, M. Isopi, M. E. Tosti, P. Schembri, and G. Scavia. Quantitative methods for the prioritization of foods implicated in the transmission of hepatititis e to humans in italy. *Foods*, 11(1):87, 2021.

[54] L. Mughini-Gras, G. Angeloni, C. Salata, N. Vonesch, W. D'amico, G. Campagna, A. Natale, F. Zuliani, L. Ceglie, I. Monne, et al. Hepatitis e virus infection in north italy: high seroprevalence in swine herds and increased risk for swine workers. *Epidemiology & Infection*, 145(16):3375–3384, 2017.

[55] I. Nimgaonkar, Q. Ding, R. E. Schwartz, and A. Ploss. Hepatitis e virus: advances and challenges. *Nature Reviews Gastroenterology & Hepatology*, 15(2):96, 2018.

[56] W.-J. Park, B.-J. Park, H.-S. Ahn, J.-B. Lee, S.-Y. Park, C.-S. Song, S.-W. Lee, H.-S. Yoo, and I.-S. Choi. Hepatitis e virus as an emerging zoonotic pathogen. *Journal of veterinary science*, 17(1):1, 2016.

[57] G. Pavia, A. Gioffrè, M. Pirolo, D. Visaggio, M. T. Clausi, M. Gherardi, P. Samele, L. Ciambrone, R. Di Natale, G. Spatari, et al. Seroprevalence and phylogenetic characterization of hepatitis e virus in pig farms in southern italy. *Preventive Veterinary Medicine*, 194:105448, 2021.

[58] N. Pavio, X.-J. Meng, and C. Renou. Zoonotic hepatitis e: animal reservoirs and emerging risks. *Veterinary research*, 41(6):46, 2010.

[59] E. Pavoni, I. Barbieri, B. Bertasi, G. Lombardi, G. Giangrosso, P. Cordioli, and M. N. Losio. Detection and molecular characterisation of swine hepatitis e virus in brescia province, italy. *Italian journal of food safety*, 4(2), 2015.

[60] E. Ponterio, I. Di Bartolo, G. Orrù, M. Liciardi, F. Ostanello, and F. M. Ruggeri. Detection of serum antibodies to hepatitis e virus in domestic pigs in italy using a recombinant swine hev capsid protein. *BMC veterinary research*, 10(1):1–7, 2014.

[61] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2019.

[62] M. Rausand and S. Haugen. *Risk assessment: theory, methods, and applications*, volume 115. John Wiley & Sons, 2020.

[63] C. Renou, A.-M. R. Afonso, and N. Pavio. Foodborne transmission of hepatitis e virus from raw pork liver sausage, france. *Emerging infectious diseases*, 20(11):1945, 2014.

[64] P. Ripellino, E. Pasi, G. Melli, C. Staedler, M. Fraga, D. Moradpour, R. Sahli, V. Aubert, G. Martinetti, F. Bihl, et al. Neurologic complications of acute hepatitis e virus infection. *Neurology-Neuroimmunology Neuroinflammation*, 7(1), 2020.

[65] M. Riveiro-Barciela, B. Mínguez, R. Gironés, F. Rodriguez-Frías, J. Quer, and M. Buti. Phylogenetic demonstration of hepatitis e infection transmitted by pork meat ingestion. *Journal of clinical gastroenterology*, 49(2):165–168, 2015.

[66] A. Rivero-Juarez, M. Frias, A. Martinez-Peinado, M. Risalde, D. Rodriguez-Cano, A. Camacho, I. Garcia-Bocanegra, F. Cuenca-Lopez, J. Gomez-Villamandos, and A. Rivero. Familial hepatitis e outbreak linked to wild boar meat consumption. *Zoonoses and public health*, 64(7):561–565, 2017.

[67] D. Rodriguez-Lazaro, N. Cook, F. M. Ruggeri, J. Sellwood, A. Nasser, M. S. J. Nascimento, M. D'Agostino, R. Santos, J. C. Saiz, A. Rzeżutka, et al. Virus hazards from food, water and other contaminated environments. *FEMS microbiology reviews*, 36(4):786–814, 2012.

[68] S. M. Ross. *Introduction to probability and statistics for engineers and scientists*. Academic Press, 2020.

[69] F. M. Ruggeri, I. Di Bartolo, E. Ponterio, G. Angeloni, M. Trevisani, and F. Ostanello. Zoonotic transmission of hepatitis e virus in industrialized countries. *New Microbiol*, 36(4):331–44, 2013.

[70] M. Salines, M. Andraud, and N. Rose. From the epidemiology of hepatitis e virus (hev) within the swine reservoir to public health risk mitigation strategies: a comprehensive review. *Veterinary research*, 48(1):1–15, 2017.

[71] M. Salines, E. Barnaud, M. Andraud, F. Eono, P. Renson, O. Bourry, N. Pavio, and N. Rose. Hepatitis e virus chronic infection of swine co-infected with porcine reproductive and respiratory syndrome virus. *Veterinary research*, 46(1):1–10, 2015.

[72] A. Saltelli, K. Chan, and E. Scott. *Sensitivity Analysis: Gauging the Worth of Scientific Models*. Wiley Series in Probability and Statistics. Wiley, 2000.

[73] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola. *Global sensitivity analysis: the primer*. John Wiley & Sons, 2008.

[74] L. Serracca, I. Rossini, R. Battistini, M. Goria, S. Sant, C. Ercolini, et al. Potential risk of norovirus infection due to the consumption of "ready to eat" food. *Food and environmental virology*, 4(3):89–92, 2012.

[75] E. Spada, S. Pupella, G. Pisani, R. Bruni, P. Chionne, E. Madonna, U. Villano, M. Simeoni, S. Fabi, G. Marano, et al. A nationwide retrospective study on prevalence of hepatitis e virus infection in italian blood donors. *Blood Transfusion*, 16(5):413, 2018.

[76] E. Suffredini, F. Garofalo, C. Salzano, L. Cozzi, G. d. Marco, M. Romano, A. Pesce, et al. Evaluation of the presence of norovirus, hepatitis a, hepatitis e, astrovirus, adenovirus, aichivirus and rotavirus in bivalve molluscs [conference poster]. In *XVII Congresso Nazionale SI Di. LV, Pacengo di Lazise (VR), Italia, 28-30 settembre 2016*, pages 318–319. Società Italiana di Diagnostica di Laboratorio Veterinaria (SIDiLV), 2016.

[77] K. Szabo, E. Trojnar, H. Anheyer-Behmenburg, A. Binder, U. Schotte, L. Ellerbroek, G. Klein, and R. Johne. Detection of hepatitis e virus rna in raw sausages and liver sausages from retail in germany using an optimized method. *International Journal of Food Microbiology*, 215:149–156, 2015.

[78] G. Tarantino, P. Bagnarelli, M. Marzioni, K. Marinelli, G. Surace, S. Traini, G. S. Baroni, S. Menzo, and A. Benedetti. Hepatitis e in a region of italy: An emerging autochthonous infection? *Digestive and Liver Disease*, 48(11):1340–1345, 2016.

[79] S. Tei, N. Kitajima, K. Takahashi, and S. Mishiro. Zoonotic transmission of hepatitis e virus from deer to human beings. *The Lancet*, 362(9381):371–373, 2003.

[80] V. Terio, M. Bottaro, E. Pavoni, M. Losio, A. Serraino, F. Giacometti, V. Martella, A. Mottola, A. Di Pinto, and G. Tantillo. Occurrence of hepatitis a and e and norovirus gi and gii in ready-to-eat vegetables in italy. *International journal of food microbiology*, 249:61–65, 2017.

[81] M. Tosti, S. Longhi, C. De Waure, A. Mele, E. Franco, W. Ricciardi, and A. Filia. Assessment of timeliness, representativeness and quality of data reported to italy's national integrated surveillance system for acute viral hepatitis (seieva). *Public health*, 129(5):561–568, 2015.

[82] M. E. Tosti, L. Ferrigno, A. Mele, L. Romanò, D. Fiacchini, P. Bagnarelli, C. Zotti, M. Chironna, R. Prato, M. T. Giordani, et al. Epidemiology and surveillance of hepatitis e in italy: data from the seieva surveillance system 2007-2019. *Epidemiologia e Prevenzione*, 45(1-2):46–53, 2021.

[83] A.-S. Vercouter, I. M. Sayed, Z. Lipkens, K. De Bleecker, S. De Vliegher, R. Colman, M. Koppelman, K. Supré, and P. Meuleman. Absence of zoonotic hepatitis e virus infection in flemish dairy cows. *International journal of food microbiology*, 281:54–59, 2018.

[84] L. Wasserman. *All of statistics: a concise course in statistical inference*. Springer Science & Business Media, 2013.

[85] P. A. White, N. E. Netzler, and G. S. Hansman. *Foodborne viral pathogens*. CRC Press, 2016.

[86] K. Woolson, A. Forbes, L. Vine, L. Beynon, L. McElhinney, V. Panayi, J. Hunter, R. Madden, T. Glasgow, A. Kotecha, et al. Extra-hepatic manifestations of autochthonous hepatitis e infection. *Alimentary pharmacology & therapeutics*, 40(11-12):1282–1291, 2014.

[87] World Health Organisation. Hepatitis e, Fact Sheet, 2021. https://www.who.int/en/news-room/fact-sheets/detail/hepatitis-e (Last accessed on 2021-12-18).

[88] D. M. Yugo and X.-J. Meng. Hepatitis e virus: foodborne, waterborne and zoonotic transmission. *International journal of environmental research and public health*, 10(10):4507–4533, 2013.

Parte II

Top-down methods for source attribution of Shiga Toxin-producing Escherichia Coli (STEC)

Capitolo 1

Source attribution model for Shiga Toxin-producing Escherichia Coli (STEC)

In this chapter we report risk assessment activities carried out to quantify how much animal and other non-human sources of STEC weigh in the causation of human illnesses. This source attribution (SA) estimation has been carried out within the framework of the on-going research project of the One Health European Joint Program DiSCoVeR: *Discovering the sources of Salmonella, Campylobacter, VTEC and antimicrobial resistance.* The overall aim of thi project is to strenghten the capacity building at the EU level for the implementation of routine source attribution assessment, as a main component of the surveillance and control strategies for the priority zoonotic foodborne pathogens (i.e. Salmonella, Campylobacter, STEC) and antimicrobial resistant-bacteria (ESBL- Extended Spectrum Beta-lactamase Enterobatteriacee). This is done by mapping the availability of SA methods and gaps in the knowledge of pathogens population data, in particular of the genetic markers, as well as in the availability of surveillance and monitoring data to address classical and innovative SA methods. SA assessment is conducted within the DISCOVER by combining partners' know-how on the specific pathogens biology, epidemiology and ecology with methodological (modelling) skills and existing data, shared by DISCOVER partners. Zoonotic foodborne pathogens transmission to humans occurs mainly through food consumption but the dependent happening structure of a pathogen can include also other infection routes such as direct contact with infected animals or exposure through environmental matrices (for instance contaminated water). Weigh the relative importance of specific food sources (i.e. animal populations) or transmission routes to the overall burden of disease to humans is a crucial step to effectively address intervention strategies to reduce the health impact of foodborne pathogens in humans. To do that, several source attribution techniques have been developed. We can define human illness source attribution, quoting [13], as

> [...] the partitioning of the human disease burden of one or more foodborne infections to specific sources, where the term *source* includes animal reservoirs and vehicles.

There are several approaches used to attribute sources for foodborne pathogens. We focused here on approaches based on microbial subtyping where microbiological features of the pathogen are used as markers to compare the distribution of the pathogens' subtypes in humans and in the different

non-human sources. Microbial subtyping allows to classify the pathogen into subtypes based on the phenotypic and/or genotypic characteristics. Given the subtype distribution in observed human cases and in the potential sources it is possible to estimate the number of cases attributable to each source. The sources subtype frequencies are usually weighted with other factors representing the level of human exposure to sources or pathogen-dependent factors modelling virulence or survival capabilities of the pathogenic organism. In the area of STEC, which are the pathogens for which there was less experience of SA assessment available in the scientific community, the DISCOVER approach aimed to explore the feasibility of SA assessment using the classical SA approach developed for Salmonella. Additional information on the project are available at https://onehealthejp.eu/jrp-discover/.

1.1 Exploring classical source attribution approach for STEC

We refer to a specific Bayesian frequency-matching model developed by Hald and colleagues for the source attribution of *Salmonella* [5]. This model, as well as other modified similar models, have been widely applied to a number of pathogens, especially *Salmonella* [14] and *Campylobacter* [11]. The application of such models to other type of pathogens has not been extensively explored, among these *Shiga Toxin-producing Escherichia Coli (STEC)* [10], for a series of reasons. Beside the poor availability of surveillance and monitoring data on STEC, a main reason was that for many years there was not a common consensus on the scientific paradigm for the definition of pathogenicity of STEC, given the extremely genomic plasticity of these pathogens and the species E.coli, in general.

Recently, a Joint FAO/WHO initiative [2] as well as an European Food Safety Authority scientific opinion [12] solved the issue and agreed on a pathogenicity assessment scheme based on the presence of a series of virulence genes (stx1, stx2, eae/aggR). This approach allowed to make obsolete all paradigms based on E. coli O- and H- antigen characterization, which had been used to predict the pathogenicity of STEC for many decades. Anyway, E.coli O-group characterization continue to be used as a valid subtype epidemiological marker, also for STEC.

1.2 Shiga Toxin-Producing Escherichia Coli. Background and epidemiology

STEC are a specific pathogroup of E. coli species, capable of producing Shigatoxin (Stx) which is encoded by stx genes. Infections with Shiga-Toxin-producing *Escherichia Coli* (STEC) cause severe clinical manifestations in humans, with premature mortality and sequelae associated with chronic disability [16]. Young children (<5 years) more frequent develop hemorrhagic diarrhea and Hemolytic Uremic Syndrome (HUS) [3, 17], which is the most severe complication and is caused by the systemic effect and organ damage caused by the Shigatoxin, which is released by the STEC in the intestine and enter the blood stream. HUS occurs as a systemic complication of intestinal STEC infection. It is characterized by microangiopathic hemolytic anemia, thrombocytopenia and kidney damage. In pediatrics it is one of the leading cause of acute renal failure [3]. The disease initially manifests itself with a nonspecific clinical picture characterized by gastroenteric prodromal symptoms, with haemorrhagic colitis which is the most specific prodromal symptom of STEC infections. In about 15% of cases it evolves into HUS. The evolution and severity of manifestations

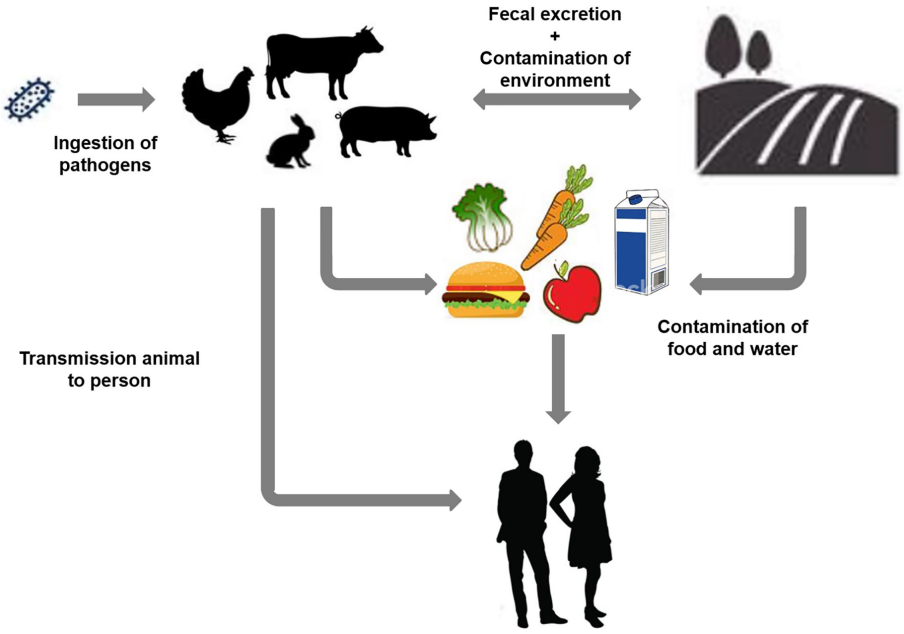

Figure 1.1. Dependent happening structure for STEC [6]

related to STEC infection are mainly determined by host-related factors in particular age, and by the virulence characteristics of the STEC. Acute lethality in HUS cases is between 3% and 5% [7].

The HUS is responsible for the highest burden of illness of STEC infections (see for instance [8]). In terms of the care burden, it should be remembered that almost all patients affected by HUS require dialysis and blood transfusions in the acute phase, which can only be performed in specialized centers. Long-term renal and extrarenal sequelae are frequent, particularly neurological (proteinuria, hypertension, chronic kidney disease requiring renal transplantation, permanent neurological damage, epilepsy) [7,16]. There is no specific etiological therapy for STEC infections. Antibiotic therapy is contraindicated as it promotes the release of Shigatoxin increasing the risk of developing HUS [7]. It is therefore very important to recognize HUS early to reduce the risk of permanent sequelae [1]. The epidemiological cycle of STEC infections is complex and characterized by multiple ways of transmission to humans (see Figure 1.1). Infection occurs through the ingestion of contaminated undercooked food (mainly minced meat, dairy products from raw milk, vegetables) and water. The infection can spread also through the person-person oral-fecal route, especially in the household and in community settings such as school, by direct contact with infected animals or contaminated environmental matrices (eg bathing water). Secondary person-to-person transmission is particularly important in community settings such as nursery schools, kindergartens, etc. [15]. The epidemic potential of STEC infections is high. Epidemic outbreaks have a strong media impact due to the clinical severity of the disease, the characteristics of patients affected by severe forms (usually children) and the fact that food transmission of the infection is perceived as poorly acceptable. Among the most recent outbreaks, the 2011 E. coli O104: H4 epidemic caused 3,842 cases of infection, 855 cases of HUS and 53 deaths over two months. In Italy, the 2013 E. coli O26 epidemic caused 20 cases of HUS [4].
In France, this year a large outbreak involving 51 sick children with two deaths was caused by consumption of contaminated frozen Buitoni Pizza produced by Nestlé.

Beside reasons explained above, the applicability of classical SA assessment to STEC has been poorly explored given its specific features. Among them, the high genomic variability of STEC makes difficult to identify the biomarker that better indicates virulence and pathogenicity

of the strains. Secondly, the wide range of non-humans reservoir of the organisms introduce a further challenge for such analysis. In particular, at European level, a source attribution study on STEC has never been performed before. This study aim is therefore to adapt existing SA models to the specific features of STEC, using a big dataset collected specifically for the OneHealth EJP DiSCoVeR project involving several European countries. This will allow us to stratified the analysis for several factors highlighting differences and similarities essential to gain new knowledge on the burden of the different sources to human STEC infections.

Capitolo 2

STEC database description

In our project we have access to a database of 15,588 STEC isolates coming from 11 European Member States countries collected between 1997 and 2021 (available at https://zenodo.org/record/5828412#.YlsG8dNByUk) collected from partners of the DiSCoVeR project of One Health European Joint Program.

The STEC isolates were obtained from human samples (14,196), animals (both livestock and wildlife) (1,139), food sample (224) or environment (29) samples. The isolates were classified into three hierarchical sublevels reported in Table 2.1. The level 3 subspecies are listed in Table 2.2. For each observation in the database we have

year,

country,

the three sub-levels,

virulence profile,

age-class and disease (for human cases only).

Distributions of the isolates per country and year are reported in Figure 2.1.

As explained by the level 2, human cases can be sporadic, outbreak- or travel-related. We considered as "sporadic" all the human cases not linked to an outbreak and "domestically acqui-

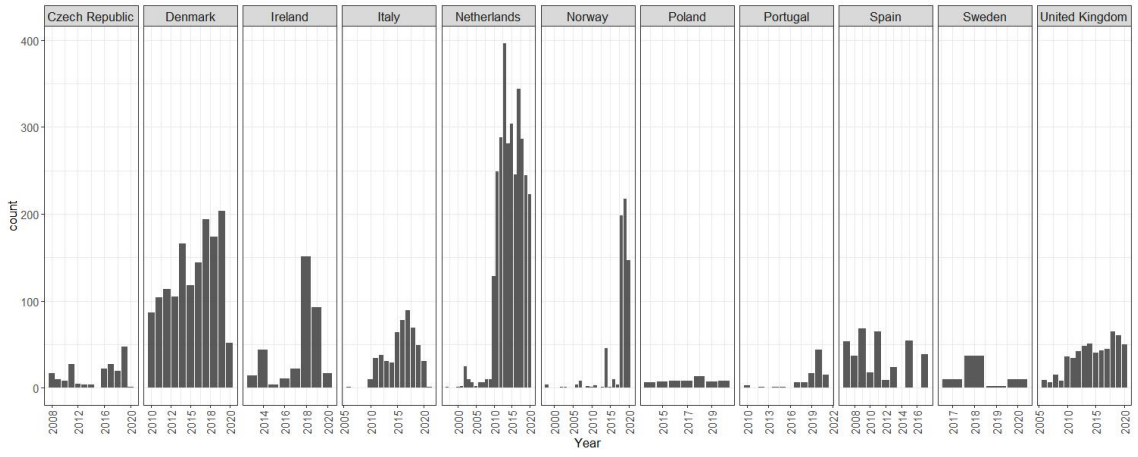

Figure 2.1. Isolates distribution per country and year

Level 1	Level 2	Level 3
Human	Sporadic Travel-related Outbreak related Unknown	Not Applicable
Animal	Livestock Zoo Pets Other	Animal Species (Table 2.2)
Food	Cheese/dairy/milk Meat Fruit/vegetables	Animal Species (Table 2.2) Vegetable Species (Table 2.2)
Environment	Water Other	Water types (Table 2.2)

Tabella 2.1: STEC database dimensions

red" (or simply "domestic") all the cases non travel-related, including the outbreak-related ones. Outbreak-related cases are very difficult to confirm, the number of underreported outbreak-related cases could be therefore very high. The number of sporadic isolates is the highest (40%) followed by travel-related cases. More than 35% of the human isolates though were classified as "Unknown".

Non-human source isolates belong to animal, food or environmental matrices. Most belonged to animal sources (80%), followed by food isolates. The environmental sources isolates are less than 2%. Distribution of non-human isolates per each country with details on level 2 and level 3 are displayed on Figure 2.3. In Figure 2.2 distribution of the same levels combination over the different O-groups are reported.

Beside the availability of O-group characterization, information on virulence genes regard, stx and eae. For the purpose of our analysis, we used as subtyping marker for STEC a combination of O-group (serogroup) and two virulence genes: stx1, stx2 and eae. In the single bacterial cell stx genes can be variably combined as stx1, stx2, stx1/stx2. eae gene, which encode for the intestinal adhesion factor intimin, can be present in STEC (eae+) or absent (eae-), although STEC responsible of severe illness in humans are usually eae+. STEC as any other E.coli may be classified according to the O-antigen of the lypopolisaccharide into 171 different O-groups. The most frequent STEC O-group reported in humans is STEC O157 that has been found in 22% of the isolates, followed by O26 (9%). The other O-groups have been detected in less than 5% of the isolates. More than 33% of isolates are classified as "Unknown". Disitrbutions of O-groups isolates by country are shown in Figure 2.4 meaning that characterization of STEC has not been completed.

The majority of isolates were stx2+ and eae positive, which is the virulence combination associated to more severe cases in human (see Figure 2.6), followed by stx1+ and eae positive (see Figure 2.5). For almost all the unknown stx isolates also the presence of eae gene was unknown, except for two eae positive isolates.

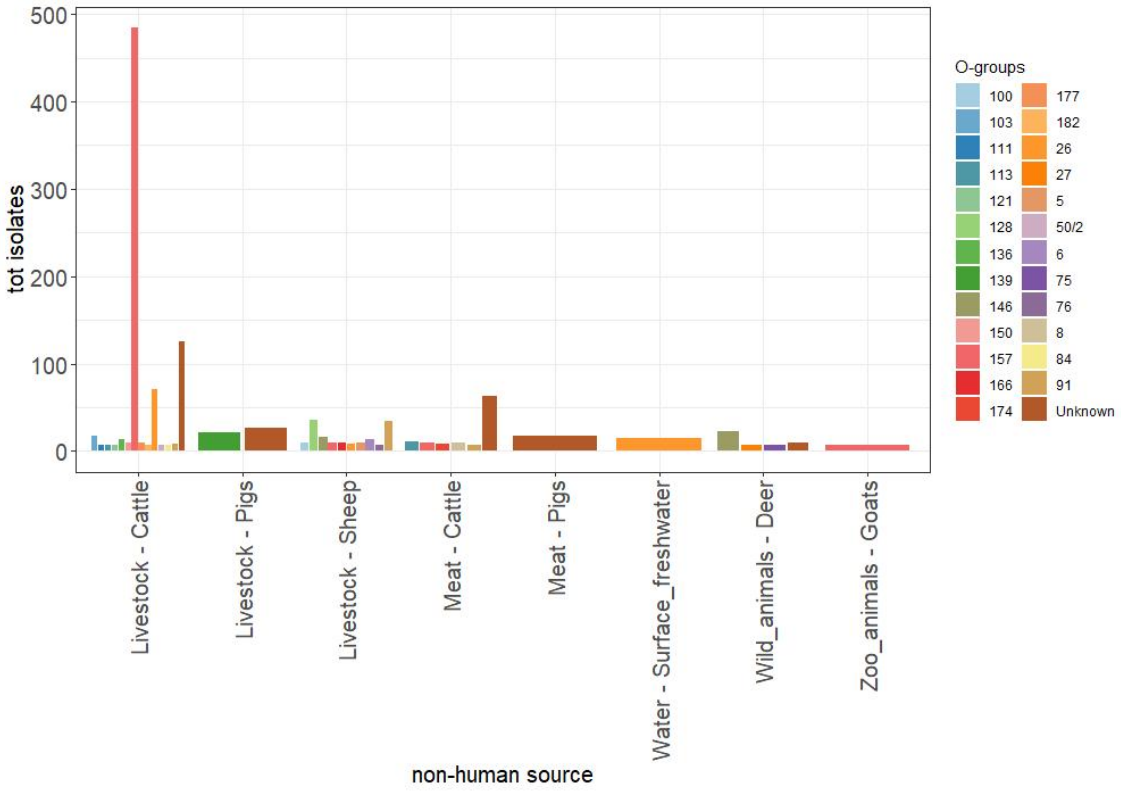

Figure 2.2. O-group distribution in non-human sources

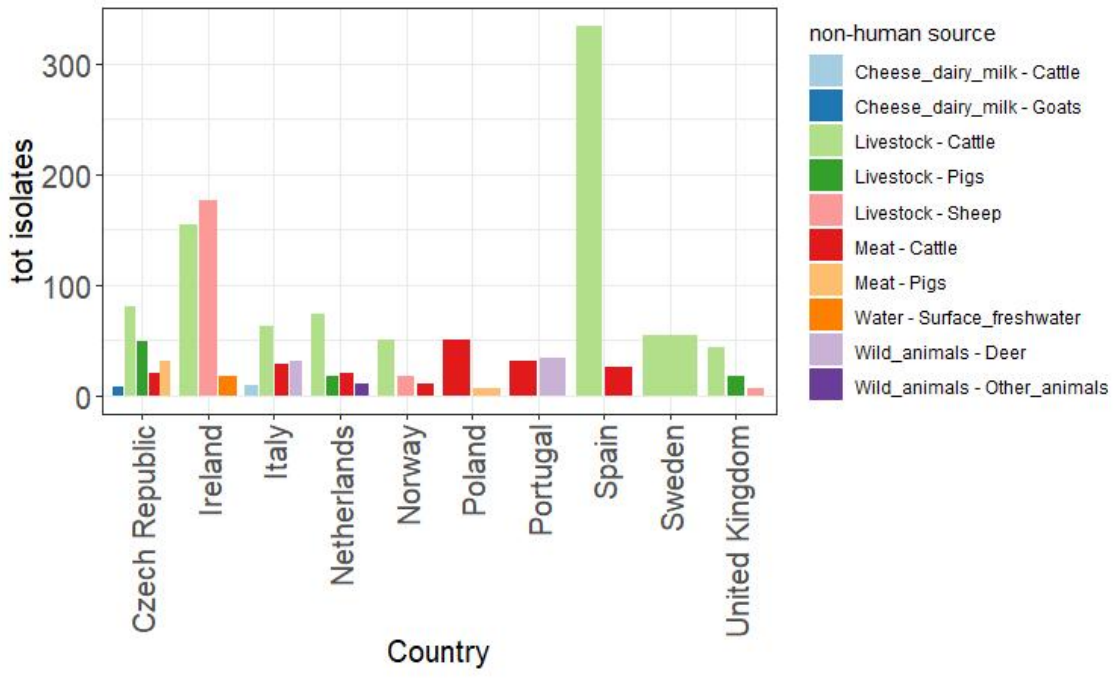

Figure 2.3. Non-human isolates distirbution over country

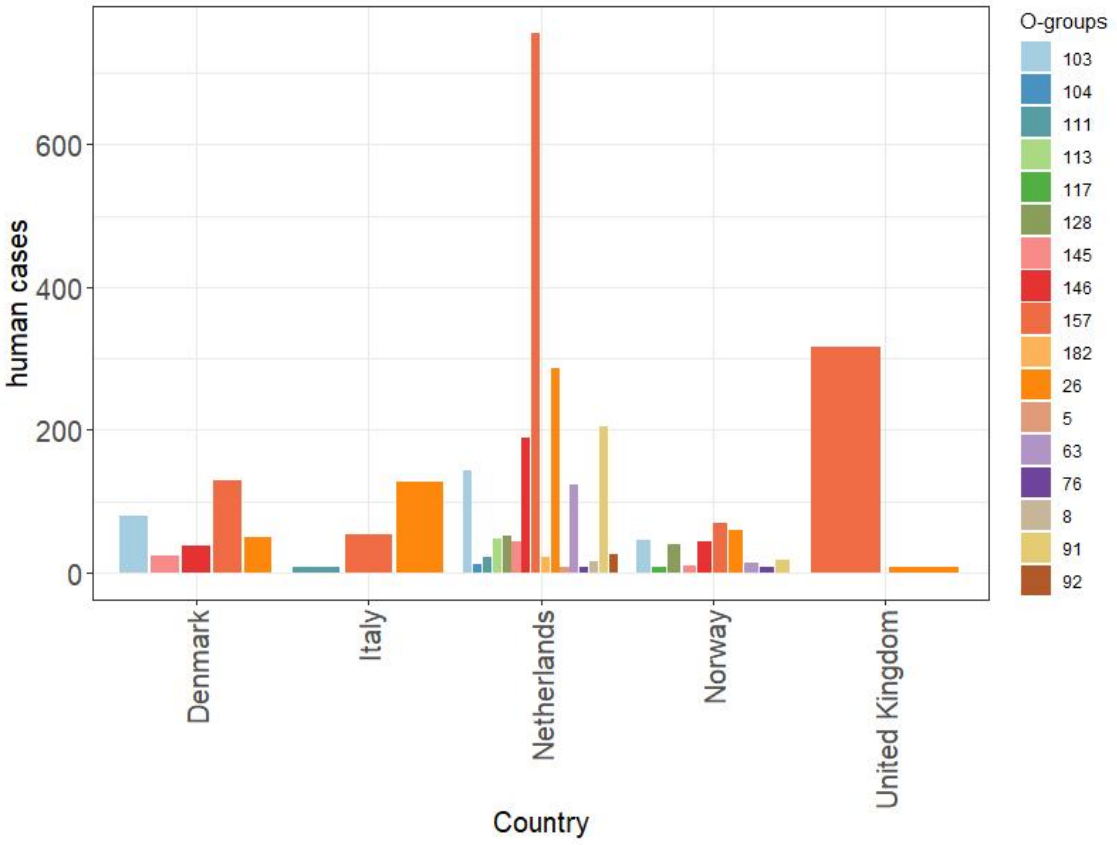

Figure 2.4. Distribution of O-groups isolates per country

Animal Species	Birds
	Cattle
	Cats
	Deer
	Dogs
	Goat/sheep
	Horses/donkeys
	Pigs
	Poultry
	Wild boar
	Other animal
	Unknown animal
Vegetables Species	Green leafy vegetables
	Sprout
	Other vegetables
Water types	Groundwater
	Sewage wastewater
	Surface freshwater
	Other

Tabella 2.2: Subspecies of Level 3 in Table 2.1

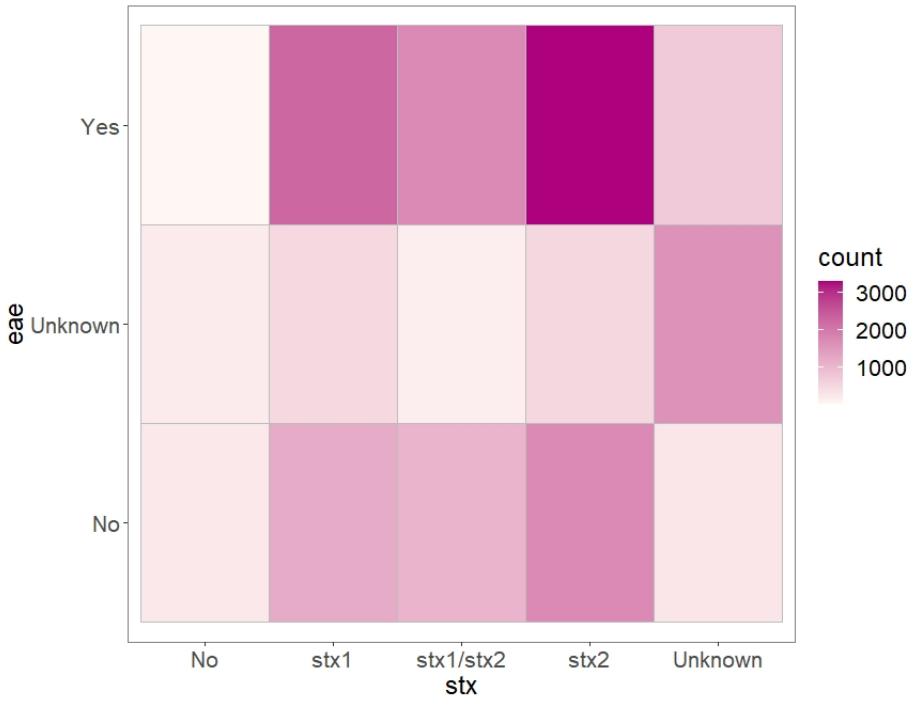

Figure 2.5. Heatmap that shows the count of isolates with specific eae and stx combination.

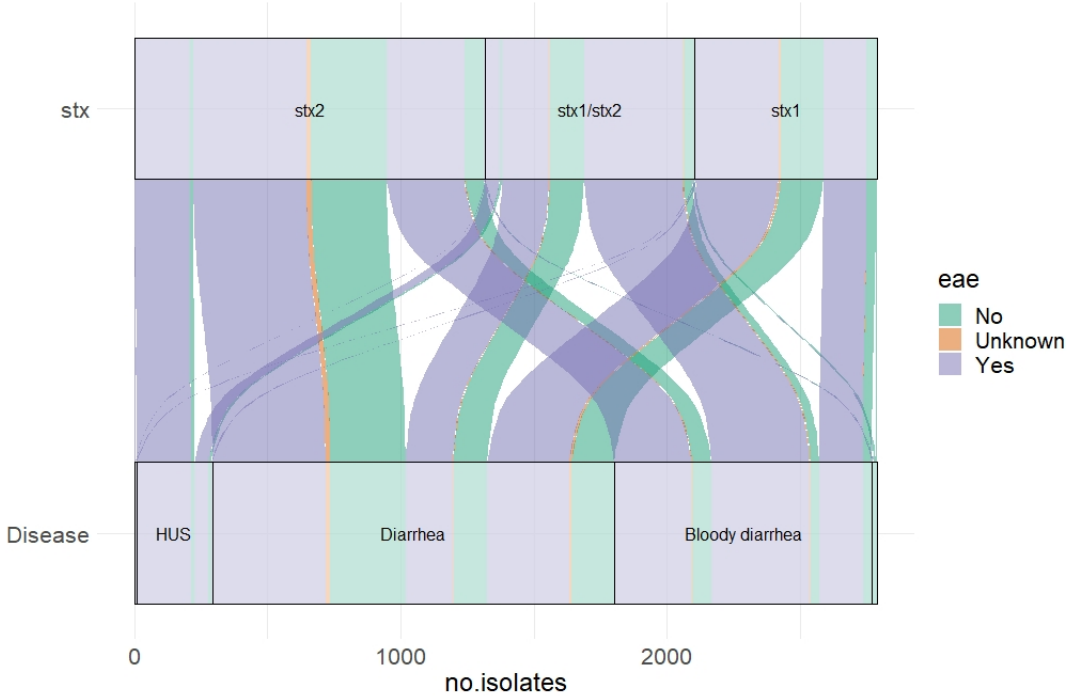

Figure 2.6. Alluvial diagram of stx/eae combination versus clinical outcomes.

Capitolo 3

Source attribution model for STEC

The approach we chose for our SA assessment in the DISCOVER was developed on a Bayesian frequency-matching model based on microbial subtyping. In this approach we used a Bayesian model to compare the distribution of STEC subtypes in STEC isolates from human cases with isolates from non-human source samples. We used a framework developed by Hald et al. in 2004 originally appllied for the attribution of *Salmonella* [5,9,14] and successfully applied also for other pathogen [10,11]. The model is built to estimate the expected number of human sporadic STEC infections occurred in a specific country and year, that are attributable to the different non-human reservoirs or pathways sources of the pathogen. To estimate posterior distribution we used Markov Chain Monte Carlo simulation, in particular the Gibbs sampler. The simulations were run with R software version 4.1.2 and WinBUGS 14 through `R2WinBUGS` package.

3.1 Model description

We modified the original Hald model [5] to include isolates with information on four different dimensions, that are: year t, country c, STEC subtype i and non-human source j. As described in Chapter 2, we have isolates from several year and countries. Unfortunately, as already mentioned, some observations in the database had missing information. Therefore, similarily as done in [5], we modelled uncertainty on that observations using Beta-Binomial or Dirichlet-Multinomial models to estimate the additional isolates for each dimension's combination (see Table 3.1). All the detail of the uncertainty modelization are reported in Appendix D.

We assumed that the expected number of cases λ_i^{tc} observed in year t and country c of subtype i that are attributable to source j depend on three different factors:

- the prevalence p_{ij}^{tc} of subtype i in source j in year t in country c;
- a bacteria-dependent factor q_i expressing differences in STEC subtypes, as ability to infect humans or survival capability
- a source dependent factor a_j^{tc} modelling differences between source in acting as vehicle for human transmission

parameter	description	value
t	time in years	–
c	country	–
i	STEC subtype, combination of O-group-stx-eae	–
j	food-animal source	–
o_i^{tc}	number of observed cases with subtype i in year t and country c	data
ao_{ij}^{tc}	non-human isolates of source j with subtype i collected in year t and country c	data
p_{ij}^{tc}	prevalence of type i in source j in year t and country c	$\sim Dir(ao_{ij}^{tc}+1)$
λ_i^{tc}	expected number of sporadid domestic cases infected with subtype i in year t and in country c	$\sum_j \lambda_{ij}^{tc}$
λ_{ij}^{tc}	expected number of sporadic domestic cases per subtype i and source j observed in year t and country c	$p_{ij}^{tc} \cdot a_j^{tc} \cdot q_i$
a_j^{tc}	source-dependent factor for source j in year t in country c	$\sim exp(\mu)$
q_i	subtype-dependent factor for type i	$\sim unif(\alpha, \beta)$
tr_i^{tc}	number of observed cases with subtype i in year t and country c known to be travel-related	data
$estr_i^{tc}$	estimated additional number of travel-realted cases of type i observed in year t and country c	Appendix D

Tabella 3.1: Description and definition of the parameters used to estimate the number of cases reported per each year and country per STEC subtype

Therefore, λ_i^{tc} is defined

$$\lambda_i^{tc} = \sum_j \lambda_{ij}^{tc}$$

where λ_{ij}^{tc} are the expected number of human cases of subtype i from source j in year t and country c.

We are, similarly as done by Pires and collegues in [14], letting a depends on year and on country to account for possible variability in consumption, preparation or handling of the sources. All the parameters are shown in Table 3.1.
To get the expected number of cases Λ_i^{tc} reported in a certain year t in country c per subtype i, as done by Hald [5], we added to $\lambda_i^{tc} = \sum_j \lambda_{ij}^{tc}$ the total travel-related cases $tr_i^{tc} + estr_i^{tc}$ and the total outbreak-realted cases, meaning the cases originally travel- or outbreak-related plus the additional cases estimated.

The model is defined as follow:

$$o_i^{tc} \sim Poisson(\Lambda_i^{tc}) \tag{3.1}$$

where Λ_i^{tc} is defined

$$\Lambda_i^{tc} = \lambda_i^{tc} + tyto_i^{tc} + tob_i^{tc}.$$

We used uninformative uniform prior for q_i. To improve idenfiability of the model, following [11], for a_{ij}^{tc} we used an exponential prior.

Appendice D

Data uncertainty modelling

In the following section we describe how we modelled the uncertainty about human source data. For non-human isolates we proceeded in the exact same way but in this cases only for subtype information that were the only missing.

For human isolates we defined the subtype i as the combination of three different subdimensions that are

- the O-group k
- the stx type h
- the eae presence or absence l

We called $o_{is}^{tc} = o_{khls}^{tc}$ the observed human cases for which all the information were available, where s is an additional dimension indicating travel information (domestic or travel-related). We indicated with $_{i}^{tc} = o_{khl}^{tc}$ the number of observed human cases both domestic and travel-related. To indicate the observations with missing information on some dimensions, we use capital letters and write $U * O$, where asterisk stands for the missing dimensions, as described in Table D.1. Additionally, in Figure D.1 is schematically represented the summary of the variable names based on the known dimension, where t and c dimensions were omitted.

variable name	description
USO_{khl}^{tc}	observed cases with unknown travel information but known O-group, stx, eae
UKO_{hls}^{tc}	observed cases with unknown O-group but known stx, eae, travel information
ULO_{khs}^{tc}	observed cases with unknown eae but known O-group, stx, travel informatione
$ULSO_{kh}^{tc}$	observed cases with unknown eae and travel information, but known O-group and stx
$UKSO_{hl}^{tc}$	observed cases with unknown O-group and travel information, but known stx and eae
$UKLO_{hs}^{tc}$	observed cases with unknown O-group and eae but known stx and travel information
$UKLSO_{h}^{tc}$	observed cases with unknown O-group, eae and travel information but known stx
$UKHSO_{l}^{tc}$	observed cases with unknown O-group, stx and travel information but known eae

Tabella D.1: Variable names for the missing dimensions observations for human isolates

To attribute missing information on the isolates we used a Bayesian framework, depending on the number of missing dimensions of the isolate and on the number of level of the specific dimenion.

The additional cases attributed to each of the group described in Table D.1 are named the same way of the missing information group with an "A" in place of the "U". For instance, ASO_{khls*}^{Htc}

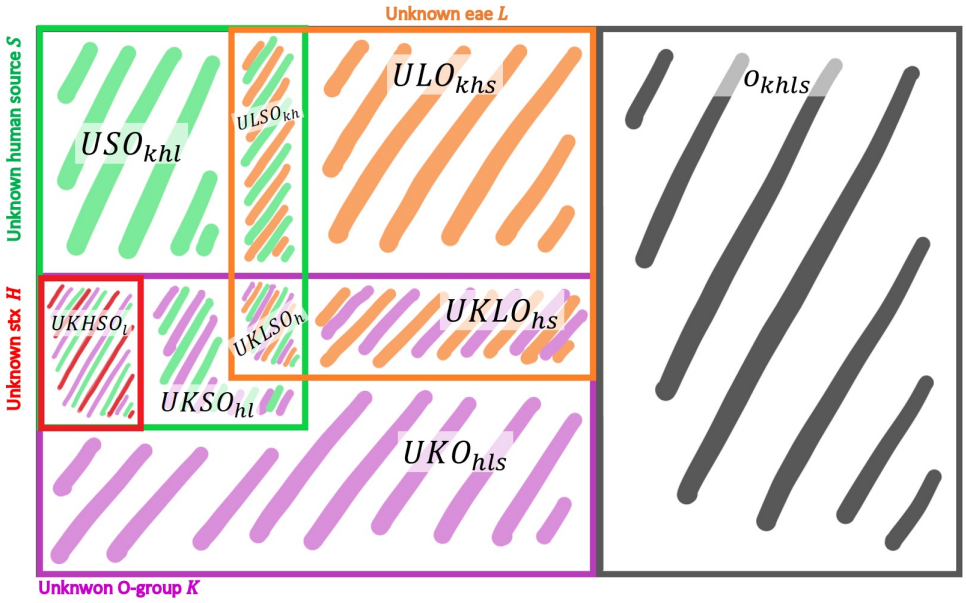

Figure D.1. Variable names for the missing dimensions observations for human isolates (tc superscipt was ignored for simplicity).

are the additional cases initially without travel information information (USO^{Htc}_{khl}) with attributed travel information s^* (domestic or travel-related). To estimate the additional cases for each dimensions we used the distribution observed in the other isolates, where that dimension is not missing. Whence, we are assuming that the distribution of the missing dimension is the same for the isolates where we known all the dimensions and where some of the dimensions are missing. For instance to estimate the additional number of cases with a certain O-group, we assumed that the distribution of O-groups in the isolate with missing O-group information is the same as in the isolated where O-group is indicated.

We used, similarly as in [5], we used Beta and Dirichlet distribution as conjugate prior for Binomial or Multinomial distribution, depending on the number of levels of the dimension. For instance, we used Beta-Binomial model for isolates with missing information on *eae* presence since eae is a bivariate variable while we used Dirichlet-Multinomial model for multivariate dimensions like O-group or *stx*. For isolates with more than one missing dimensions we used a combination of these models, as described in Tables D.3 and D.4. We indicated the with an superscript the sum over one specific dimension, for instance $o^{tc;k}_{hls} := \sum_k o^{tc}_{khls}$.

Therefore we obtained the total number of cases per subtype as reported in Table 3.1. Recalling that $i = (k, h, l)$, we got the total number of domestic cases $spdo^{tc}_i + espdo^{tc}_i$, where $spdo^{tc}_i$ are the observed domestic cases and $espdo^{tc}_i$ are the estimated additional domestic cases. In the same way we got the total travel-related cases $tr^{tc}_i + estr^{tc}_i$, where tr^{tc}_i are the observed travel-realted cases and $estr^{tc}_i$ are the estimated additional travel-related cases.

parameter	description	definition
$puko^{tc}_{k^*hls}$	probability that a case with unknown O-group and stx h, eae l and travel information s belongs to O-group k^*	$Dirichlet(o^{tc}_{k^*hls}+1)$
$AKO^{tc}_{k^*hls}$	estimated additional number of cases with stx h, eae l and travel information s that had O-group k^*	$UKO^{tc}_{hls} \cdot puko^{tc}_{k^*hls}$
$puho^{tc}_{kh^*ls}$	probability that a case with unknown stx and eae l, O-group k and travel information s has stx h^*	$Dirichlet(o^{tc}_{kh^*ls}+1)$
$AHO^{tc}_{kh^*ls}$	estimated additional number of cases with eae l, O-group k and travel information s that had stx h^*	$UHO^{tc}_{kls} \cdot puho^{tc}_{kh^*ls}$
$pulo^{tc}_{khl^*s}$	probability that a case with unknown eae and O-group k, stx h and travel information s has eae l^*	$Dirichlet(o^{tc}_{khl^*s}+1)$
$ALO^{tc}_{khl^*s}$	estimated additional number of cases with O-group k, stx h and travel information s that had eae l^*	$ULO^{tc}_{khs} \cdot pulo^{tc}_{khl^*s}$
$puso^{tc}_{khls^*}$	probability that a case with unknown travel information and O-group k, stx h and eae l has travel information s^*	$Dirichlet(o^{tc}_{khls^*}+1)$
$ASO^{tc}_{khls^*}$	estimated additional number of cases with O-group k, stx h and eae l that had travel information s^*	$USO^{tc}_{khl} \cdot puso^{tc}_{khls^*}$

Tabella D.2: Parameters used to model the missing information on cases with one dimension missing

parameter	description	definition
$u^{tc}_{k^*hs}$	proportion of cases with unknown O-group and eae and known stx h and travel information s to allocate to O-group k^*	$\frac{Gamma(o^{tc;l}_{k^*hs};1)}{\sum_k Gamma(o^{tc;l}_{khs};1)}$
$AUKLO^{tc}_{k^*hl^*s}$	estimated additional cases with stx h and travel information s that had O-group k^* and eae l^*	$Bin(u^{Htc}_{k^*hs} \cdot UKLO^{tc}_{hs}; pulo^{tc}_{k^*hl^*s})$
$u^{tc}_{khl^*}$	proportion of cases with unknown travel information and eae and known stx h and O-group k to allocate to eae l^*	$\frac{Gamma(o^{tc;s}_{khl^*};1)}{\sum_l Gamma(o^{tc;s}_{khl};1)}$
$AULSO^{tc}_{khl^*s^*}$	estimated additional cases with stx h and O-group k that had travel information s^* and eae l^*	$(u^{tc}_{khl^*} \cdot ULSO^{tc}_{kh}) \cdot puso^{tc}_{khl^*s^*}$
$u^{tc}_{k^*hl}$	proportion of cases with unknown travel information and O-group and known stx h and eae l to allocate to O-group k^*	$\frac{Gamma(o^{tc;s}_{k^*hl};1)}{\sum_k Gamma(o^{tc;s}_{khl};1)}$
$AUKSO^{tc}_{k^*hls^*}$	estimated additional cases with stx h and eae l that had travel information s^* and O-group k^*	$(u^{tc}_{k^*hl} \cdot UKSO^{tc}_{hl}) \cdot puso^{tc}_{k^*hls^*})$

Tabella D.3: Parameters used to model the missing information on cases with two dimension missing

parameter	description	definition
$u^{tc}_{k^*h^*l}$	proportion of cases with unknown O-group, travel information and stx and known eae l to allocate to the combination O-group k^* and stx h^*	$\frac{Gamma(o^{tc;s}_{k^*h^*l};1)}{\sum_{k,h} Gamma(o^{tc;s}_{khl};1)}$
$AKHSO^{H}_{k^*h^*l^*s}$	estimated additional cases with eae l that had stx h^*, travel information and s^* O-group k^*	$(u^{tc}_{k^*h^*l} \cdot UKHSO^{tc}_{h}; puso^{tc}_{k^*h^*ls^*})$
$u^{tc}_{k^*hl^*}$	proportion of cases with unknown travel information, O-group and eae and known stx h to allocate to the combination O-group k^* and eae l^*	$\frac{Gamma(o^{Htc;s}_{k^*hl^*};1)}{\sum_{k,l} Gamma(o^{tc;s}_{khl};1)}$
$AKLSO^{tc}_{k^*hl^*s^*}$	estimated additional cases with stx h that had travel information s^*, O-group k^* and eae l^*	$(u^{tc}_{k^*hl^*} \cdot UKLSO^{tc}_{kh}) \cdot puso^{tc}_{k^*hl^*s^*}$

Tabella D.4: Parameters used to model the missing information on cases with three dimensions missing

References

[1] G. Ardissino, F. Tel, I. Possenti, S. Testa, D. Consonni, F. Paglialonga, S. Salardi, N. Borsa-Ghiringhelli, P. Salice, S. Tedeschi, et al. Early volume expansion and outcomes of hemolytic uremic syndrome. *Pediatrics*, 137(1), 2016.

[2] V. authors. Joint fao/who expert meeting on microbiological risk assessment (jemra) on shiga toxin-producing escherichia coli (stec) associated with meat and dairy products. Technical report, FAO/WHO, 2020.

[3] F. Fakhouri, J. Zuber, V. Frémeaux-Bacchi, and C. Loirat. Haemolytic uraemic syndrome. *The Lancet*, 390(10095):681–696, 2017.

[4] C. Germinario, A. Caprioli, M. Giordano, M. Chironna, M. S. Gallone, S. Tafuri, F. Minelli, A. Maugliani, V. Michelacci, L. Santangelo, et al. Community-wide outbreak of haemolytic uraemic syndrome associated with shiga toxin 2-producing escherichia coli o26: H11 in southern italy, summer 2013. *Eurosurveillance*, 21(38):30343, 2016.

[5] T. Hald, D. Vose, H. C. Wegener, and T. Koupeev. A bayesian approach to quantify the contribution of animal-food sources to human salmonellosis. *Risk Analysis: an International Journal*, 24(1):255–269, 2004.

[6] S.-b. Hwang, R. Chelliah, J. E. Kang, M. Rubab, E. Banan-Mwine Daliri, F. Elahi, and D.-H. Oh. Role of recent therapeutic applications and the infection strategies of shiga toxin-producing escherichia coli. *Frontiers in cellular and infection microbiology*, 11:450, 2021.

[7] D. Karpman, S. Loos, R. Tati, and I. Arvidsson. Haemolytic uraemic syndrome. *Journal of internal medicine*, 281(2):123–148, 2017.

[8] S. E. Majowicz, E. Scallan, A. Jones-Bitton, J. M. Sargeant, J. Stapleton, F. J. Angulo, D. H. Yeung, and M. D. Kirk. Global incidence of human shiga toxin–producing escherichia coli infections and deaths: a systematic review and knowledge synthesis. *Foodborne pathogens and disease*, 11(6):447–455, 2014.

[9] L. Mughini-Gras, F. Barrucci, J. Smid, C. Graziani, I. Luzzi, A. Ricci, L. Barco, R. Rosmini, A. Havelaar, W. Van Pelt, et al. 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 & Infection*, 142(5):1070–1082, 2014.

[10] L. Mughini-Gras, W. Van Pelt, M. Van der Voort, M. Heck, I. Friesema, and E. Franz. 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 and Public Health*, 65(1):e8–e22, 2018.

[11] P. Mullner, G. Jones, A. Noble, S. E. Spencer, S. Hathaway, and N. P. French. Source attribution of food-borne zoonoses in new zealand: A modified hald model. *Risk Analysis: An International Journal*, 29(7):970–984, 2009.

[12] E. B. Panel, K. Koutsoumanis, A. Allende, A. Alvarez-Ordóñez, S. Bover-Cid, M. Chemaly, R. Davies, A. De Cesare, L. Herman, F. Hilbert, et al. Pathogenicity assessment of shiga toxin-producing escherichia coli (stec) and the public health risk posed by contamination of food with stec. *EFSA Journal*, 18(1):e05967, 2020.

[13] S. M. Pires, E. G. Evers, W. Van Pelt, T. Ayers, E. Scallan, F. J. Angulo, A. Havelaar, and T. Hald. Attributing the human disease burden of foodborne infections to specific sources. *Foodborne pathogens and disease*, 6(4):417–424, 2009.

[14] S. M. Pires and T. Hald. 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, 2010.

[15] G. Scavia, A. Gianviti, V. Labriola, P. Chiani, A. Maugliani, V. Michelacci, F. Minelli, R. Tozzoli, A. Caprioli, and S. Morabito. A case of haemolytic uraemic syndrome (hus) revealed an outbreak of shiga toxin-2-producing escherichia coli o26: H11 infection in a nursery, with long-lasting shedders and person-to-person transmission, italy 2015. *Journal of Medical Microbiology*, 67(6):775–782, 2018.

[16] J. M. Spinale, R. L. Ruebner, L. Copelovitch, and B. S. Kaplan. Long-term outcomes of shiga toxin hemolytic uremic syndrome. *Pediatric nephrology*, 28(11):2097–2105, 2013.

[17] P. I. Tarr, C. A. Gordon, and W. L. Chandler. Shiga-toxin-producing escherichia coli and haemolytic uraemic syndrome. *The lancet*, 365(9464):1073–1086, 2005.

Conclusions

Mathematical models represent an essential tool in applied science to face complex and multi-faceted issues such as the control of zoonotic foodborne pathogens and the related challenges for public health, veterinary medicine, food and environmental safety. The aim of this work was to address needs of public health sector taking advantage of quantitative methodologies in a comprehensive One Health framework, meaning with a specific attention to the strong correlation between animals, human and environmental health. Aside from this, our findings underline that non-health related factors and actors may play key role in influencing the food safety and the risk of consumer exposure to foodborne pathogens, suggesting the need to enlarge the One-Health vision well beyond the multisectoral collaboration across sanitary sectors. Conclusions from the pre-harvest model for HEV indicated that pig production model as well as farm management have a direct impact on the risk of population exposure to HEV. These non-sanitary factors together with the consumers traditional eating habits and behaviors may represent a valid explanation for the significant higher HEV sero-prevalence observed by Spada and colleagues in human population in both central regions of Italy and Sardinia. These findings highlight the importance of engaging actors such as farmers and consumers in the strategies for HEV control.

Overall, our study produced two main achievements. From one side, we were able to obtain scientific evidence based on quantitative risk assessment which could directly inform control policy and support decision makers facing emerging problems and needs of public health sectors in an innovative fashion. An important added value of this approach is that uncertainty around results is clearly estimated, which a crucial element to be taken into account in decision making-process. On the other side, the use of quantitative methods allow to analyse the existing gaps in knowledge and data, which is extremely important in addressing research and monitoring and surveillance policies. The identification of these gaps was an important part of our work. Documenting clearly the challenges and the source of uncertainty affecting results gave us the chance to highlight the gaps affecting the knowledge for HEV, in particular pathogenic aspects in pigs, and the lack of robust data on STEC in non-human sources. In conclusion, we strongly believe that the empowerment of mathematical models for public health and for real life problems in general are one of the most effective way to maximize a sustainable use of existing data and information, avoiding the constant need to produce novel and original data. This applied quantitative study engaged deeply inter-disciplinary and multidisciplinary expertise, knowledge and professional skills, implicating a continuous effort to transfer knowledge from one area to the others. An important challenge was represented by the need to overcome language and conceptual barriers across discipline and to implement a cooperative common environment allowing mutual understanding to find the better way to achieve the study aims.